Optimal Designs for Multi-Response Nonlinear Regression Models With Several Factors via Semidefinite Programming

Weng Kee Wong, Yue Yin & Julie Zhou

August 2018

Optimal designs for multi-response nonlinear regression models with several factors via semi-definite programming

Weng Kee Wong

Department of Biostatistics, University of California, Los Angeles

Los Angeles, CA 90095-1772, USA

Yue Yin and Julie Zhou[1]

Department of Mathematics and Statistics, University of Victoria

Victoria, BC, Canada V8W 2Y2

ABSTRACT

We use semi-definite programming (SDP) to find a variety of optimal designs for multi-response linear models with multiple factors, and for the first time, extend the methodology to find optimal designs for multi-response nonlinear models and generalized linear models with multiple factors. We construct transformations that (i) facilitate improved formulation of the optimal design problems into SDP problems, (ii) enable us to extend SDP methodology to find optimal designs from linear models to nonlinear multi-response models with multiple factors and (iii) correct erroneously reported optimal designs in the literature caused by formulation issues. We also derive invariance properties of optimal designs and their dependence on the covariance matrix of the correlated errors, which are helpful for reducing the computation time for finding optimal designs. Our applications include finding A-, A_s-, c- and D-optimal designs for multi-response multi-factor polynomial models, locally c- and D-optimal designs for a bivariate E_{max} response model and for a bivariate Probit model useful in the biosciences.

Key words and phrases: A-optimality, c-optimality, generalized linear model, invariance property, multi-response model, semi-definite programming.

MSC 2010: 62K05, 62K20.

[1]Corresponding author, email: jzhou@uvic.ca, phone: 250-721-7470.

1 Introduction

Different types of optimal designs have been extensively constructed and studied for linear and nonlinear models, see for example, the design monographs by Fedorov (1972), Pukelsheim (1993), Berger and Wong (2009), and Dean et al. (2015). Applications of optimal design ideas are found across broad disciplines, such as in education testing studies, food science, agriculture, pharmaceutical industry, toxicology, engineering, and manufacturing sectors. As cost of experiments rises, the need of having a well designed study that saves costs provides a powerful impetus for the continuing search for more effective ways to compute efficient designs.

The bulk of optimal regression designs concerns models with one response variable and one or two factors. When there are multiple factors, the models are usually assumed to be linear and the mean response is adequately modeled by a low degree multi-factor polynomial or an additive non-polynomial model. For example, Chang et al. (2001), Huang et al. (2006) and Montgomery (2013, p496) considered design issues for experiments with two or three response variables and up to three factors. Optimal designs for nonlinear models with several factors are rare.

Multi-response models are increasingly common and design issues are decidedly more complicated than finding optimal designs for one-response models. There is relatively little work in this area even for polynomial homoscedastic models. A classic reference is Krafft and Schaefer (1992), who found D-optimal designs for a single covariate model with multiple responses. Bischoff (1993, 1995) developed determinant formulae from a factorization lemma to help find such optimal designs. In the last two decades, modest advances have been made in terms of theory, new criteria and new applications. For example, Chang (1997) developed an algorithm to generate near D-optimal designs, and the models include mainly the first-order or second-order polynomial model for each response. Imhof (2000) found exact optimal designs for a bivariate model where each mean response is modeled by a linear and a quadratic model and the joint responses may be correlated. Wang (2000) discussed exact and approximate D-optimal designs for multi-response polynomial models. Using iterated estimators and convex analysis, Fedorov et al. (2001) found optimal designs for a model with bivariate responses when the covariance matrix depends on unknown parameters. Chang et al. (2001) derived D- and D_s-optimal designs for polynomial regression models with two response variables and one independent variable. More recent results include Yue and Liu (2010) and Liu et al. (2011) where they proposed a class of standardized optimality criteria, including design criteria based on predictive ellipsoids for linear models, and constructed various optimal designs for multi-factor linear models. Yue et al. (2014) constructed D-optimal designs for multi-response linear models with quantitative and qualitative factors.

Optimal designs for multi-response nonlinear models are less studied and there is increasing interest in this area, especially in the pharmaceutical industry. For example, in dose-finding studies, multiple endpoints comprising efficacy and toxicity endpoints can be studied more efficiently when the design is optimally constructed. Heise and Myers (1996) and Kpamegan (1998) are early examples where they found optimal designs for bivariate logistic and Probit models, respectively. Dragalin et al. (2008) constructed two-stage design for monitoring efficacy and safety in an early phase clinical trial, where the goal is to find a target dose that most likely produces the desired probabilities that balance toxicity and efficacy effects of a drug. Fedorov and Wu (2007) and Fedorov et al. (2012) also developed optimal designs for measuring patients' progress using continuous outcomes turned binary measures for studying efficacy and toxicity. Gueorguieva et al. (2006) found locally D-optimal designs for multivariate response pharmacokinetic models assuming that observations from different patients were independent and observations from the same patient at different time points were correlated. Mukhopadhyay and Khuri (2008) compared various optimal designs for multi-response generalized linear models (GLMs), and Magnusdottir (2013) derived results for c-optimal designs for estimating a given linear function of parameters in a bivariate Emax model with one independent variable. Dette et al. (2013) constructed Bayesian D-optimal designs for a specific type of multi-response generalized linear models. Most recently, Kim (2014) found D- and c-optimal designs for the bivariate and trivariate Weibull model with applications to the electronic manufacturing industry.

An overarching issue for finding optimal designs for multi-response nonlinear models with several covariates is computational. Published work in the area do not discuss numerical methods well; there are either brief or ad-hoc algorithms that apply only to specific design problems. Semi-definite programming (SDP) as an optimization tool has been widely used in engineering and mathematics. Even though it is well known that SDP solvers can solve SDP problems with hundreds or more variables efficiently (Papp, 2012), it is rarely used for solving optimal design problems. Atashgah and Seifi (2007, 2009) was an early user of SDP to find A-, D- and E-optimal designs for linear models and Papp (2012) was the first to provide a theory based approach for solving a class of optimal design problems for regression models, where the mean response is a polynomial or a rational function with heteroscedastic noise modeled also by a polynomial or rational weight function. In either case, the models are linear, except for the rational polynomial models briefly considered in Papp (2012). Recent work using SDP to find different types of optimal designs for nonlinear models has only one response variable, see for example, Duarte et al. (2016*), Duarte and Wong (2015), and Duarte et al. (2016). Elfving's Theorem is a useful tool for geometrically characterizing a c-optimal design and Sagnol (2011) extended Elfving's celebrated theorem for finding a c-optimal design geometrically in multi-response experiments and used second-order cone programming for computing various optimal designs.

Our main contributions in this work are to extend SDP methodology to find various A-, A_s-, c-, and D-optimal designs for multi-response linear/nonlinear models with multi-factors and correct some results in Atashgah and Seifi (2009) due to formulation issues. We derive several theoretical properties of optimal designs that facilitate our SDP-based algorithms for finding them when we have multi-response linear or nonlinear models with several independent variables. In addition, we develop equivalence theorems for different design criteria to verify optimality of a design for multi-response models with multiple independent variables and present optimal designs found from our algorithm useful for biomedical applications.

Section 2 discusses the best linear unbiased estimator for multi-response linear models and design optimality criteria. In Section 3 we transform A-, A$_s$-, c- and D-optimal design problems into SDP problems and use SDP algorithms to find optimal designs. Section 4 presents optimal designs from our algorithm and compares them with those found by other methods. Section 5 finds optimal designs for multi-response GLMs. Section 6 derives several theoretical properties for A-, A$_s$- and D-optimal designs that are useful for the SDP algorithms, and concluding remarks are in Section 7. All proofs are in the Appendix.

2 Optimality criteria for multi-response models

Suppose we have r response variables, $y_1, \cdots, y_r$ and p design variables $x_1, \cdots, x_p$. To fix ideas, we first consider linear models and note that the approach can be applied to nonlinear models and GLMs with details in Sections 4 and 5. The linear models are given by

$$y_{ij} = \mathbf{f}_i^\top(\mathbf{x}_j)\boldsymbol{\theta}_i + \epsilon_{ij}, \quad j = 1, \cdots, n, \quad i = 1, \cdots, r, \tag{1}$$

where y_{ij} is the jth observation on the response variable y_i, $\mathbf{x}$ denotes the vector of the p design variables, i.e., $\mathbf{x} = (x_1, \cdots, x_p)^\top$, $\mathbf{x}_j = (x_{1j}, \cdots, x_{pj})^\top$ is the jth run input (design point), $\mathbf{f}_i(\mathbf{x})$ is a known vector of regressors for variable y_i, $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \cdots, \theta_{iq_i})^\top$ is the vector of regression parameters for y_i, and ϵ_{ij} are random errors with mean zero. Further, we assume that $Cov(\epsilon_{ij}, \epsilon_{lk}) = 0$ for $j \neq k$, and $Cov(\epsilon_{ij}, \epsilon_{lj}) = \sigma_{il}$. This means that the observations for the r response variables are correlated in the same run and observations from different runs are independent.

Let $q = q_1 + \cdots + q_r$ be the total number of regression parameters in (1) and for

$j = 1, \cdots, n$, define

$$\mathbf{y}_j = \begin{pmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{rj} \end{pmatrix}_{r\times 1}, \quad \boldsymbol{\epsilon}_j = \begin{pmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \vdots \\ \epsilon_{rj} \end{pmatrix}_{r\times 1}, \quad \mathbf{Z}_j = \begin{pmatrix} \mathbf{f}_1^\top(\mathbf{x}_j) & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{f}_2^\top(\mathbf{x}_j) & 0 & \cdots & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & 0 & \cdots & \mathbf{f}_r^\top(\mathbf{x}_j) \end{pmatrix}_{r\times q}.$$

Put

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}_{nr\times 1}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_n \end{pmatrix}_{nr\times 1}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_n \end{pmatrix}_{nr\times q}, \quad \boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_r \end{pmatrix}_{q\times 1}.$$

Manifestly, in matrix form the model becomes

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{2}$$

where the covariance matrix of the error term $\boldsymbol{\epsilon}$ is

$$\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{\epsilon}) = \Sigma_0 \oplus \Sigma_0 \oplus \cdots \oplus \Sigma_0, \quad \text{with} \quad \Sigma_0 = \mathrm{Cov}(\boldsymbol{\epsilon}_1) = (\sigma_{il})_{r\times r}. \tag{3}$$

Here $\oplus$ is the matrix direct sum and Σ_0 is assumed to be a known positive definite matrix. The best linear unbiased estimator (BLUE) of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \left(\mathbf{Z}^\top \boldsymbol{\Sigma}^{-1}\mathbf{Z}\right)^{-1} \mathbf{Z}^\top \boldsymbol{\Sigma}^{-1}\mathbf{Y}$, and its covariance matrix is

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}) = \left(\mathbf{Z}^\top \boldsymbol{\Sigma}^{-1}\mathbf{Z}\right)^{-1} = \left(\sum_{j=1}^{n} \mathbf{Z}_j^\top {\Sigma_0}^{-1} \mathbf{Z}_j\right)^{-1}. \tag{4}$$

For a predetermined sample size n, we want to choose "optimal" design points $\mathbf{x}_j$, $j = 1, \cdots, n$, from the given design space to minimize some loss function $\phi\left(\mathrm{Cov}(\hat{\boldsymbol{\theta}})\right)$ represented by a scalar function ϕ. Various optimality criteria can be used, such as A-, A_s-, c-, and D-optimality criteria. Throughout, we assume that we are willing to discretize the design space into N user-selected points. Let this discretized design space be $S_N \subset R^p$, let its points be $\mathbf{u}_1, \cdots, \mathbf{u}_N$ and let $\mathbf{U}_j$ be the $r \times q$ matrix $\mathbf{Z}_j$ evaluated at $\mathbf{u}_j, j = 1, \cdots, N$. A design measure $\xi(\mathbf{x})$ of $\mathbf{x}$ is denoted by

$$\xi(\mathbf{x}) = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_N \\ w_1 & w_2 & \cdots & w_N \end{pmatrix}, \tag{5}$$

where w_j is the proportion of times that $\mathbf{u}_j$ is selected as a design point of $\xi(\mathbf{x})$, with $w_j \geq 0$ and $\sum_{j=1}^N w_j = 1$. If $w_j > 0$, then $\mathbf{u}_j$ is called a support point of $\xi(\mathbf{x})$. Let $\mathbf{w} = (w_1, \cdots, w_N)$ be the weight vector and let

$$\mathbf{A}(\mathbf{w}) = \sum_{j=1}^{N} w_j \mathbf{U}_j^\top {\Sigma_0}^{-1} \mathbf{U}_j, \tag{6}$$

so that the covariance matrix in (4) is proportional to $\mathbf{C}(\mathbf{w}) = \mathbf{A}^{-1}(\mathbf{w})$.

An optimal exact design problem is defined by the optimization problem

$$\begin{cases} \min_{\mathbf{w}} \phi\left(\mathbf{C}(\mathbf{w})\right) \\ \text{s.t.} \;\; w_j \in \{0, 1/n, 2/n, \cdots, 1\}, \; j = 1, \cdots, N, \quad \sum_{j=1}^N w_j = 1. \end{cases}$$

When $\mathbf{A}(\mathbf{w})$ is singular, $\phi\left(\mathbf{C}(\mathbf{w})\right)$ is defined to be $+\infty$. Since it is hard to solve exact design problems in general, we relax the constraints on $\mathbf{w}$ and find optimal approximate designs defined by:

$$\begin{cases} \min_{\mathbf{w}} \phi\left(\mathbf{C}(\mathbf{w})\right) \\ \text{s.t.} \;\; w_j \geq 0, \quad j = 1, \cdots, N, \quad \sum_{j=1}^N w_j = 1. \end{cases} \tag{7}$$

Optimal exact designs are much more difficult to find and study than optimal approximate designs. The former are constrained integer-valued optimization problems and they defy analytical solutions. For each sample size n allocated for the study, the optimal exact design varies and can depend sensitively on the value n, the criterion and the model. In contrast, optimal approximate designs do not depend on n and frequently, analytical formulae and algorithms can generate the optimal designs. Consequently, optimal approximate designs are easier to find and study and are more widely used in practice as approximations to optimal exact designs.

When the design criterion is convex, which is the case in our paper, a particularly appealing advantage of working with approximate designs is that an equivalence theorem is available to verify the optimality of any design among all designs on the given design space. Each convex criterion results in a different equivalence theorem but they all have the same form, i.e. each is an inequality with an upper bound of

0 and becomes an equality at the design points of the optimal design. Equivalence theorems are widely discussed in design monographs (Fedorov, 1972, Pukelsheim, 1993, Berger and Wong, 2009). Most of the discussion concerns univariate response models. Heise and Myers (1996) is an early example, where design issues for bivariate response model were considered. In what is to follow, we focus only on approximate designs for multi-response models.

For the D-optimality criterion, we have $\phi\left(\mathbf{C}(\mathbf{w})\right) = \det\left(\mathbf{C}(\mathbf{w})\right)$. For A-, A$_s$-, and c-optimality, the loss function $\phi\left(\mathbf{C}(\mathbf{w})\right)$ has a general form given by

$$h(\mathbf{w}, \mathbf{L}) = \text{trace}\left(\mathbf{L}^\top \mathbf{C}(\mathbf{w})\mathbf{L}\right) = \text{trace}\left(\mathbf{L}^\top \mathbf{A}^{-1}(\mathbf{w})\mathbf{L}\right), \tag{8}$$

where

$$\mathbf{L} = \begin{cases} \mathbf{I}_q, & \text{for A-optimality,} \\ \text{diag}(\mathbf{a}), & \text{for A}_s\text{-optimality,} \\ \mathbf{c}_{q\times 1}, & \text{for c-optimality.} \end{cases} \tag{9}$$

Here $\mathbf{I}_q$ is the $q \times q$ identity matrix, diag($\mathbf{a}$) is a user-selected diagonal matrix with components of the vector $\mathbf{a}$ equal to 0 or 1 on the diagonal and $\mathbf{c}$ is the given vector of coefficients for estimating $\mathbf{c}^\top\boldsymbol{\theta}$. Since $\mathbf{A}(\mathbf{w})$ is linear in $\mathbf{w}$, $h(\mathbf{w}, \mathbf{L})$ in (8) is a convex function of $\mathbf{w}$ for each fixed $\mathbf{L}$. Consequently, using convex analysis results, we can check optimality of a design for multi-response models using Lemma 1 as follows. We omit its proof because it is similar to that in Bose and Mukerjee (2015).

Lemma 1. *Let* $h_j(\mathbf{w}, \mathbf{L}) = \textit{trace}\left(\Sigma_0{}^{-1}\mathbf{U}_j\mathbf{A}^{-1}(\mathbf{w})\mathbf{L}\mathbf{L}^\top\mathbf{A}^{-1}(\mathbf{w})\mathbf{U}_j^\top\right)$ *and let* $d_j(\mathbf{w}) = \textit{trace}\left(\mathbf{U}_j^\top\Sigma_0{}^{-1}\mathbf{U}_j\mathbf{A}^{-1}(\mathbf{w})\right), \quad j = 1, \cdots, N.$ *Then*

(i) $\hat{\mathbf{w}}$ *solves problem (7) with the loss function in (8) if and only if* $h_j(\hat{\mathbf{w}}, \mathbf{L}) \leq h(\hat{\mathbf{w}}, \mathbf{L}),$ *for all* $j = 1, \cdots, N,$

(ii) $\hat{\mathbf{w}}$ *solves problem (7) for the D-optimality if and only if* $d_j(\hat{\mathbf{w}}) \leq q,$ *for all* $j = 1, \cdots, N.$

Here is an interpretation of the equivalence theorem/optimality condition for the A- and D-optimality in Lemma 1. For each $j = 1, \cdots, N$, let us write $\mathbf{U}_j^\top \Sigma_0^{-1/2} = (\mathbf{b}_{1j}, \cdots, \mathbf{b}_{rj})$, where all $\mathbf{b}_{ij} \in R^q$. The r vectors, $\mathbf{b}_{1j}, \cdots, \mathbf{b}_{rj}$, are all transformed from design point $\mathbf{u}_j \in S_N$, and the transformation is linear in the mean functions $\mathbf{f}_1, \cdots, \mathbf{f}_r$ in the multi-response model. Then the condition for the D-optimality in Lemma 1 can be stated as

$$d_j(\hat{\mathbf{w}}) = \sum_{i=1}^{r} \mathbf{b}_{ij}^\top \mathbf{A}^{-1}(\hat{\mathbf{w}}) \mathbf{b}_{ij} \leq q, \; j = 1, \cdots, N.$$

This means that the sum of the squared distance (normalized by $\mathbf{A}(\hat{\mathbf{w}})$) for the r vectors, $\mathbf{b}_{1j}, \cdots, \mathbf{b}_{rj}$, is less than or equal to q, for all $\mathbf{u}_j \in S_N$. When $r = 1$, this can be interpreted as that all the points $\mathbf{u}_j$ or the transformed points are within an ellipsoid centered at the origin (Boyd and Vandenberghe, 2004, p388). A similar interpretation for the A-optimality in Lemma 1 can be obtained by writing $h_j(\hat{\mathbf{w}}, \mathbf{I}_q) = \sum_{i=1}^{r} \mathbf{b}_{ij}^\top \mathbf{A}^{-2}(\hat{\mathbf{w}}) \mathbf{b}_{ij}$, so the optimality condition controls the sum of the squared distance (normalized by $\mathbf{A}^2(\hat{\mathbf{w}})$) for the r vectors.

3 SDP for optimal design problems

It is hard to derive analytical solutions for optimal designs in general, and numerical methods are often applied to construct optimal designs in practice. One effective method is to use convex optimization algorithms, SeDuMi and CVX in MATLAB software, for finding optimal designs. SeDuMi and CVX are very efficient to solve SDP problems. After introducing SDP problems, we investigate and characterize the relationship between SDP and various optimal design problems.

SDP problems are a special class of convex optimization problems. They have a linear objective function and have linear matrix inequality as the constraints. Boyd and Vandenberghe (2004) has many applications and results for SDP problems. Here we give two simple SDP problems in Example 1, which are very important to understand the SDP problem for A-optimal design problems later in this section.

Example 1. Define matrices

$$\mathbf{M}_1 = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 8 & 0 \\ 1 & 0 & v_1 \end{pmatrix}, \ \mathbf{M}_2 = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 8 & 1 \\ 0 & 1 & v_2 \end{pmatrix}, \ \mathbf{M}_3 = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 8 & 0 & 1 \\ 1 & 0 & v_1 & 0 \\ 0 & 1 & 0 & v_2 \end{pmatrix},$$

where v_1 and v_2 are two variables. Consider the following two SDP problems,

$$\begin{cases} \min_{v_1,v_2} v_1 + v_2 \\ \text{s.t.} \ \ \mathbf{M}_1 \oplus \mathbf{M}_2 \succeq 0 \ , \end{cases} \tag{10}$$

and

$$\begin{cases} \min_{v_1,v_2} v_1 + v_2 \\ \text{s.t.} \ \ \mathbf{M}_3 \succeq 0 \ , \end{cases} \tag{11}$$

where $\mathbf{M}_1 \oplus \mathbf{M}_2 \succeq 0$ means that $\mathbf{M}_1 \oplus \mathbf{M}_2$ is positive semidefinite (PSD). Both problems have a linear objective function of v_1 and v_2 and have linear matrix inequalities as the constraints. Each constraint defines a convex set of (v_1, v_2). It is easy to show that the solution to (10) is $(v_1^*, v_2^*) = (2.0, 0.25)$, while the solution to (11) is $(v_1^*, v_2^*) = (2.5, 0.75)$. Figure 1 shows the convex sets and solutions. □

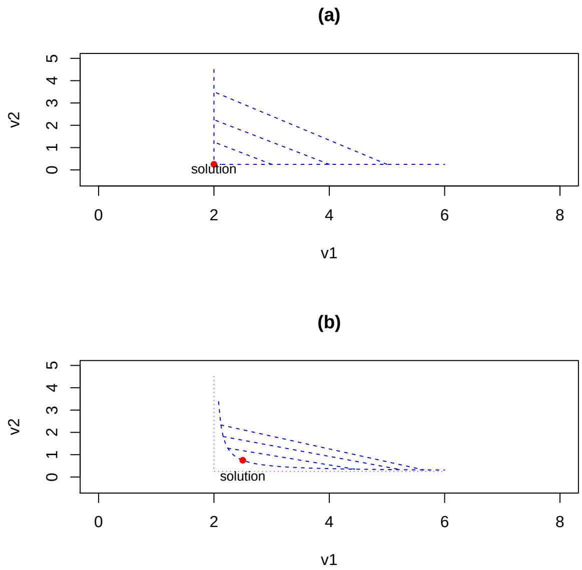
In Example 1, we note that the matrices $\mathbf{M}_1, \mathbf{M}_2$ and $\mathbf{M}_3$ have a common principal submatrix formed by the first two rows and the first two columns. More generally, the SDP problem formulations in (10) and (11) can be generalized to find optimal designs as follows. Let $\mathbf{M}$ be a $q \times q$ symmetric positive definite matrix and let $\mathbf{e}_i$ be the ith unit vector in R^q, $i = 1, \cdots, q$. Define matrices

$$\mathbf{M}_i = \begin{pmatrix} \mathbf{M} & \mathbf{e}_i \\ \mathbf{e}_i^\top & v_i \end{pmatrix}, \quad i = 1, \cdots, q, \quad \mathbf{M}_{q+1} = \begin{pmatrix} \mathbf{M} & \mathbf{I}_q \\ \mathbf{I}_q & \mathbf{V} \end{pmatrix},$$

where $\mathbf{V} = \text{diag}(v_1, \cdots, v_q)$ is a diagonal matrix. Generalize (10) and (11) as

$$\begin{cases} \min_{v_1,\cdots,v_q} v_1 + \cdots + v_q \\ \text{s.t.} \ \ \mathbf{M}_1 \oplus \cdots \oplus \mathbf{M}_q \succeq 0 \ , \end{cases} \tag{12}$$

Figure 1: The plots of convex sets (shaded areas) defined by the constraints in SDP problems and solutions: (a) SDP problem in (10), (b) SDP problem in (11).



and

$$\begin{cases} \min_{v_1,\cdots,v_q} v_1 + \cdots + v_q \\ \text{s.t.} \quad \mathbf{M}_{q+1} \succeq 0 \ . \end{cases} \tag{13}$$

The solution to problem (12) is in Theorem 1 while the solution to problem (13) is discussed in Theorem 2. Both proofs are deferred to the Appendix.

Theorem 1. *Suppose $b_{11}, \cdots, b_{qq}$ are diagonal elements of $\mathbf{M}^{-1}$. Problem (12) has a unique solution given by $v_1^* = b_{11}, \cdots, v_q^* = b_{qq}$, and $v_1^* + \cdots + v_q^* = trace\,(\mathbf{M}^{-1})$.*

Theorem 2. *If $v_1^*, \cdots, v_q^*$ is a solution to problem (13), then $v_1^* + \cdots + v_q^* \geq trace\,(\mathbf{M}^{-1})$ and the equality holds if and only if $\mathbf{M}$ is a diagonal matrix.*

The two theorems play important roles to discuss the transformation of A-optimal design problems into SDP problems. If we follow problem (12), we get the correct transformation. However, following problem (13) leads to a wrong transformation. From the proof of Theorem 2, we note that the convex set of $(v_1, \cdots, v_q)$

defined by the constraint in (13) is smaller than that in (12) in general, so the minimum value of $v_1 + \cdots + v_q$ in (13) is larger than that in (12). Example 1 confirms these results for a case of $q = 2$.

Many optimal design problems can be transformed to SDP problems. We characterize transformations for various criteria for multi-response models below, including new transformations for A$_s$- and c-optimality.

3.1 A-optimality

The loss function is $h(\mathbf{w}, \mathbf{I}_q) = \text{trace}\,(\mathbf{A}^{-1}(\mathbf{w}))$ by (8) and (9). Define matrices

$$\mathbf{D}_i = \begin{pmatrix} \mathbf{A}(\mathbf{w}) & \mathbf{e}_i \\ \mathbf{e}_i^\top & v_i \end{pmatrix}, \quad i = 1, \cdots, q, \quad \mathbf{W} = \text{diag}\left(w_1, \cdots, w_{N-1}, 1 - \sum_{j=1}^{N-1} w_j \right),$$

where $v_1, \cdots, v_q$ are real variables. Then the constraints in (7) are equivalent to $\mathbf{W} \succeq 0$. By Theorem 1, we transform problem (7) into a SDP problem as follows:

$$\begin{cases} \min_{v_1,\cdots,v_q,w_1,\cdots,w_{N-1}} & v_1 + \cdots + v_q \\ \text{s.t.} \ \ \mathbf{D}_1 \oplus \cdots \oplus \mathbf{D}_q \oplus \mathbf{W} \succeq 0 \ . \end{cases} \tag{14}$$

This transformation is an extension from one-response models in Boyd and Vandenberghe (2004) and Ye et al. (2017). The SDP problem for multi-response models in Atashgah and Seifi (2009) is similar to (13). From Theorem 2, the solution to their problem usually does not minimize $\text{trace}\,(\mathbf{A}^{-1}(\mathbf{w}))$.

3.2 A$_s$-optimality

The loss function is $h(\mathbf{w}, \text{diag}(\mathbf{a})) = \text{trace}\,(\text{diag}(\mathbf{a})\ \mathbf{A}^{-1}(\mathbf{w})\ \text{diag}(\mathbf{a}))$ by (8) and (9). Let $\mathbf{a} = (a_1, \cdots, a_q)^\top$, and a_i is either 0 or 1. A SDP problem for (7) is given by

$$\begin{cases} \min_{v_1,\cdots,v_q,w_1,\cdots,w_{N-1}} & a_1 v_1 + \cdots + a_q v_q \\ \text{s.t.} \ \ \mathbf{D}_1 \oplus \cdots \oplus \mathbf{D}_q \oplus \mathbf{W} \succeq 0 \ . \end{cases} \tag{15}$$

The constraints in (14) and (15) are the same, but the objective functions are different. In (15), a subset of the variances is included in the objective function, and the subset is defined by vector $\mathbf{a}$. This new transformation for the A_s-optimality also leads to a new one for c-optimality.

3.3 c-optimality

The loss function is $h(\mathbf{w}, \mathbf{c}) = \mathbf{c}^\top \mathbf{A}^{-1}(\mathbf{w})\mathbf{c}$. Let $\mathbf{c} = (c_1, \cdots, c_q)^\top$ and k is the smallest integer such that $c_k \neq 0$. Let $\mathbf{I}_{(-k)}$ be the $q \times (q-1)$ matrix obtained by deleting column k from $\mathbf{I}_q$. Define matrices

$$\mathbf{E} = (\mathbf{c},\ \mathbf{I}_{(-k)})_{q\times q}, \quad \tilde{\mathbf{D}}_i = \begin{pmatrix} \mathbf{E}^{-1}\mathbf{A}(\mathbf{w})\mathbf{E}^{-\top} & \mathbf{e}_i \\ \mathbf{e}_i^\top & v_i \end{pmatrix}, \quad i = 1, \cdots, q,$$

where $\mathbf{E}^{-\top} = \left(\mathbf{E}^\top\right)^{-1}$ and $\mathbf{E}$ is non-singular. It is easy to verify that

$$h(\mathbf{w}, \mathbf{c}) = \text{trace}\left(\text{diag}(1, 0, \cdots, 0)\left(\mathbf{E}^{-1}\mathbf{A}(\mathbf{w})\mathbf{E}^{-\top}\right)^{-1}\text{diag}(1, 0, \cdots, 0)\right).$$

Since $\mathbf{E}$ is a constant matrix, $\mathbf{E}^{-1}\mathbf{A}(\mathbf{w})\mathbf{E}^{-\top}$ is also linear in $w_1, \cdots, w_{N-1}$. Thus, we follow (15) with $\mathbf{a} = (1, 0, \cdots, 0)^\top$ to get the SDP problem for c-optimality as

$$\begin{cases} \min_{v_1,\cdots,v_q,w_1,\cdots,w_{N-1}} \quad v_1 \\ \text{s.t.} \ \ \tilde{\mathbf{D}}_1 \oplus \cdots \oplus \tilde{\mathbf{D}}_q \oplus \mathbf{W} \succeq 0\ . \end{cases} \tag{16}$$

3.4 D-optimality

We define a $q \times q$ upper triangular matrix $\mathbf{R}$ with diagonal elements $R_{ii},\ i = 1, \cdots q$. The SDP problem for D-optimality is as follows,

$$\begin{cases} \min_{v,\mathbf{R},w_1,\cdots,w_{N-1}} \quad -v \\ \text{s.t.} \ \ \mathbf{W} \succeq 0, \quad \begin{pmatrix} \mathbf{A}(\mathbf{w}) & \mathbf{R}^\top \\ \mathbf{R} & \mathbf{I}_q \end{pmatrix} \succeq 0, \\ \qquad v \leq \left(R_{11} \cdots R_{qq}\right)^{1/q}, \quad R_{ii} \geq 0,\ i = 1, \cdots, q. \end{cases} \tag{17}$$

The constraints of $v \leq \left(R_{11} \cdots R_{qq}\right)^{1/q}, \quad R_{ii} \geq 0,\ i = 1, \cdots, q$, can be written as linear matrix inequality by $\begin{pmatrix} R_{11} & v \\ v & R_{22} \end{pmatrix} \succeq 0$ for $q = 2$, and for $q > 2$ we can use the inequality repeatedly.

SeDuMi or CVX in MATLAB can be applied for finding solutions to (14), (15), (16) and (17). There is a sample MATLAB program in Ye et al. (2017) for solving a SDP problem for one-response model. In practice the following condition to check for optimal designs from Lemma 1 is applied, as it is done similarly in Bose and Mukerjee (2015),

$$h_j(\hat{\mathbf{w}}, \mathbf{L}) - h(\hat{\mathbf{w}}, \mathbf{L}) \leq \delta, \quad \text{for all } j = 1, \cdots, N, \tag{18}$$

with a small positive δ. For D-optimality, we use $d_j(\hat{\mathbf{w}}) - q \leq \delta$, for all $j = 1, \cdots, N$.

Like with most optimization problems, there is seldom one unique way of solving them. For example, in our problems, two other potential numerical approaches are (i) the state-of-the-art algorithm proposed recently by Yang et al. (2013) and (ii) nature-inspired meta-heuristic algorithms such as particle swarm optimization (PSO). The algorithm in Yang et al. (2013) is an exchange based algorithm where the optimal weights are updated iteratively and it is fast and flexible. However, it needs to compute the second derivatives of $\phi(\mathbf{C}(\mathbf{w}))$ with respect to $\mathbf{w}$, and sometimes it produces negative weights in the iteration. Thus, it may be hard for users to apply it and make adjustments to find optimal weights. PSO is easy to program and implement and it can solve optimization problems with hundreds of variables. An issue with PSO is that its convergence frequently depends on the tuning parameters the user selects for the algorithm. The recommended values for the tuning parameters may or may not work well, depending on the problem at hand.

The convex optimization algorithms based on the SeDuMi and CVX are fast and they always produce non-negative weights. It is very easy to apply these algorithms in MATLAB, and they do not need the second derivatives of $\phi(\mathbf{C}(\mathbf{w}))$. Users do not need to choose any tuning parameters for the convergence of the algorithms. In Sections 4 and 5 we give several applications to find optimal designs using SeDuMi and CVX.

4 Applications

We present three examples to show how our approach can find optimal designs for multi-response models, including one nonlinear and two linear models. The computation time reported is from a PC (Intel(R) Core(TM)2 Quad CPU Q9550@2.83GHz), and the equivalence results with $\delta = 10^{-5}$ are satisfied for all examples. To present the results we have rounded optimal weights to 4 decimal places, but condition (18) should be checked using the numerical results in MATLAB. Example 2 shows that our algorithms are very powerful to deal with large q and N and various design spaces. Example 3 compares our results with those in Atashgah and Seifi (2007, 2009), and we get better results for A-optimal designs. Example 4 shows the flexibility of our algorithms for finding optimal designs for nonlinear models.

Example 2. Consider a three-response model with six design variables $x_1, \cdots x_6$,

$$\mathbf{f}_1^\top(\mathbf{x}) = (1, x_1, x_2, x_3, x_4, x_5, x_6, x_1x_2, x_1x_3, x_1x_4, x_3x_4),$$
$$\mathbf{f}_2^\top(\mathbf{x}) = (1, x_1, x_2, x_3, x_4, x_5, x_6, x_1x_2, x_1x_3, x_2x_3, x_1x_4),$$
$$\mathbf{f}_3^\top(\mathbf{x}) = (1, x_1, x_2, x_3, x_4, x_5, x_6, x_5x_6),$$

where $x_1 \in [-1, 1], x_2 \in [0, 1], x_3 \in [-1, 1], x_4 \in [-0.5, 0.5], x_5 \in [-8, 8], x_6 \in [0, 2]$. Let N_i be the number of user selected grid points for x_i, $i = 1, \cdots, 6$, and S_N is formed by the Cartesian product of these grid points with $N = N_1N_2 \cdots N_6$. The covariance matrix is given by $\Sigma_0 = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$. This model has a large number of regression parameters with $q = 30$. We compute A-optimal designs for various values of N including $N = 19{,}600$. Representative results are in Table 1, which gives the information of N and Σ_0, the computation time, and optimality condition value $\delta_A = \max_j(h_j(\hat{\mathbf{w}}, \mathbf{L}) - h(\hat{\mathbf{w}}, \mathbf{L}))$. The results show that the algorithms based on the SDP are effective and powerful for finding optimal designs. For $q = 30$ and $N = 3{,}375$, it takes less than 151 seconds. For $N = 10{,}000$ it takes less than 20 minutes, while for $N = 19{,}600$ it takes about 64 minutes. For large q

and $N > 20,000$ the computation time can be long but can be reduced significantly by using the properties derived in Section 6, which will be discussed there. All the numerical results of δ_A indicate that condition (18) with $\delta = 10^{-5}$ is satisfied. $\square$

Table 1: Results for A-optimal designs in Example 2 with $q = 30$. A fast algorithm is used for the cases indicated by $**$, which is discussed in Section 6.

$N = N_1 \cdots N_6$	$(\rho_{12}, \rho_{13}, \rho_{23})$	computation time (sec.)	δ_A
$N = 5 \times 5 \times 3 \times 3 \times 3 \times 5 = 3,375$	(0.1, 0.1, 0.1)	143.3	9.31×10^{-6}
$N = 5 \times 5 \times 3 \times 3 \times 3 \times 5 = 3,375$	(0.2, 0.3, 0.5)	150.9	1.18×10^{-9}
$N = 5 \times 5 \times 3 \times 3 \times 5 \times 5 = 5,625$	(0.1, 0.1, 0.1)	345.2	6.30×10^{-6}
$N = 5 \times 5 \times 3 \times 3 \times 5 \times 5 = 5,625$	(0.5, 0.3, 0.2)	356.1	8.29×10^{-8}
$N = 5 \times 5 \times 4 \times 4 \times 5 \times 5 = 10,000$	(0.1, 0.1, 0.1)	1100.4	1.28×10^{-8}
$N = 5 \times 5 \times 4 \times 4 \times 5 \times 5 = 10,000 **$	(0.1, 0.1, 0.1)	49.6 $**$	1.28×10^{-8}
$N = 5 \times 5 \times 4 \times 4 \times 5 \times 5 = 10,000$	(0.5, 0.5, 0.5)	1038.7	1.81×10^{-6}
$N = 7 \times 5 \times 4 \times 4 \times 7 \times 5 = 19,600$	(0.1, 0.1, 0.1)	3836.1	8.04×10^{-7}
$N = 7 \times 5 \times 4 \times 4 \times 7 \times 5 = 19,600 **$	(0.1, 0.1, 0.1)	166.0 $**$	8.04×10^{-7}

Example 3. Consider a two-response model with three independent variables investigated in Atashgah and Seifi (2007, 2009), where the mean functions have $\mathbf{f}_1^\top(\mathbf{x}) = (1, x_1, x_2, x_3, x_1x_2, x_1x_3, x_1^2, x_3^2)$ and $\mathbf{f}_2^\top(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$. We take a design space S_N containing 19 possible points in R^3, which are listed in Table 2. For $\Sigma_0 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, we compute A- and D-optimal designs for various values of ρ, using CVX program in MATLAB (Grant and Boyd, 2013). Representative results in Table 2 indicate that the optimal designs depend on Σ_0 only through $|\rho|$ and are not sensitive to small changes in $|\rho|$. From Table 2 the computation is very fast since $N = 19$ is small. We also computed D-optimal designs for a design space S_N with 16 points in Atashgah and Seifi (2007) and our results are the same as theirs.

Atashgah and Seifi (2009) computed an A-optimal design when $\Sigma_0 = \begin{pmatrix} 2 & 0.4 \\ 0.4 & 1 \end{pmatrix}$. Their design is shown in Table 3, along with the A-optimal design found from solving the SDP problem in (14). The latter design has a smaller value for trace($\mathbf{A}^{-1}(\mathbf{w})$), which confirms the results in Theorems 1 and 2 and our correct SDP formulation in (14) for finding A-optimal designs. $\square$

Table 2: A- and D-optimal designs in Example 3 for various values of ρ

	Design points in S_N			A-optimal design weights			D-optimal design weights		
	x_1	x_2	x_3	$\rho = 0$	$\rho = \pm 0.1$	$\rho = \pm 0.5$	$\rho = 0$	$\rho = \pm 0.1$	$\rho = \pm 0.5$
$\mathbf{u}_1$:	1.6800	0	0	0.0616	0.0610	0.0441	0.0599	0.0593	0.0469
$\mathbf{u}_2$:	0	1.6800	0	0.0000	0.0000	0.0276	0.0000	0.0000	0.0009
$\mathbf{u}_3$:	0	0	0	0.3773	0.3773	0.3640	0.0851	0.0850	0.0822
$\mathbf{u}_4$:	1.7290	1.7270	-1.7030	0.0000	0.0000	0.0020	0.0000	0.0000	0.0000
$\mathbf{u}_5$:	1.7280	-1.7290	-1.7200	0.0487	0.0484	0.0401	0.0805	0.0803	0.0757
$\mathbf{u}_6$:	1.7290	1.7290	1.7290	0.0530	0.0530	0.0527	0.0890	0.0891	0.0896
$\mathbf{u}_7$:	-1.7250	-1.7230	1.7150	0.0150	0.0146	0.0077	0.0671	0.0670	0.0662
$\mathbf{u}_8$:	-1.7300	1.7210	1.7290	0.0271	0.0269	0.0269	0.0715	0.0713	0.0674
$\mathbf{u}_9$:	1.7300	-1.7290	1.7290	0.0369	0.0367	0.0304	0.0748	0.0746	0.0712
$\mathbf{u}_{10}$:	-1.7300	1.7300	0.0260	0.0578	0.0580	0.0589	0.0805	0.0806	0.0837
$\mathbf{u}_{11}$:	1.7300	-1.7300	-0.0450	0.0064	0.0070	0.0246	0.0163	0.0169	0.0300
$\mathbf{u}_{12}$:	-1.7290	-1.7300	-1.7280	0.0649	0.0647	0.0599	0.1056	0.1056	0.1056
$\mathbf{u}_{13}$:	-1.7300	-0.0960	1.7300	0.0474	0.0478	0.0499	0.0354	0.0359	0.0460
$\mathbf{u}_{14}$:	1.7290	1.7240	-1.7290	0.0377	0.0378	0.0384	0.0758	0.0759	0.0774
$\mathbf{u}_{15}$:	-0.1540	1.7300	-1.7300	0.0822	0.0820	0.0669	0.0883	0.0882	0.0860
$\mathbf{u}_{16}$:	-0.1010	-1.7300	1.7300	0.0694	0.0694	0.0709	0.0702	0.0703	0.0712
$\mathbf{u}_{17}$:	1.7290	1.7290	1.7220	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\mathbf{u}_{18}$:	-1.5168	-1.6182	0.6520	0.0146	0.0154	0.0350	0.0000	0.0000	0.0000
$\mathbf{u}_{19}$:	0.1158	1.6289	1.5256	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Computation time (sec.)				2.23	3.30	3.01	3.26	4.13	4.01

Table 3: A-optimal weights computed from SDP in (14) and Atashgah and Seifi (2009) (denoted by ATSE) with trace$(\mathbf{A}^{-1}(\mathbf{w})) = 17.546$ and 18.012, respectively. The computation time is 1.95 sec.

Design points	$\mathbf{u}_1$	$\mathbf{u}_2$	$\mathbf{u}_3$	$\mathbf{u}_4$	$\mathbf{u}_5$	$\mathbf{u}_6$	$\mathbf{u}_7$	$\mathbf{u}_8$	$\mathbf{u}_9$	$\mathbf{u}_{10}$
SDP in (14)	0.0504	0.0124	0.3634	0.0000	0.0460	0.0544	0.0147	0.0323	0.0343	0.0575
ATSE	0.0536	0.0000	0.4080	0.0318	0.0456	0.0000	0.0000	0.0455	0.0243	0.0498
Design points	$\mathbf{u}_{11}$	$\mathbf{u}_{12}$	$\mathbf{u}_{13}$	$\mathbf{u}_{14}$	$\mathbf{u}_{15}$	$\mathbf{u}_{16}$	$\mathbf{u}_{17}$	$\mathbf{u}_{18}$	$\mathbf{u}_{19}$	
SDP in (14)	0.0174	0.0642	0.0374	0.0405	0.0769	0.0702	0.0000	0.0280	0.0000	
ATSE	0.0066	0.0796	0.0238	0.0000	0.0656	0.0687	0.0427	0.0544	0.0000	

Example 4. Magnusdottir (2013) used a bivariate Emax model to investigate the efficacy and side-effects of a drug given by

$$y_1 = E_{\max}\frac{x}{x + ED_{50}} + \epsilon_1, \qquad y_2 = S_{\max}\frac{x}{x + SD_{50}} + \epsilon_2,$$

respectively. Here $E_{\max}$ is the maximal achievable effect from the drug, $S_{\max}$ is the maximal realizable side-effect and $x \geq 0$ is the dose level of the drug. The interesting parameters are ED_{50} and SD_{50} which are the dose levels that give half maximal effect and side-effect, respectively. For this bivariate nonlinear model, we let $\boldsymbol{\theta}_1^\top = (E_{\max}, ED_{50})$, $\boldsymbol{\theta}_2^\top = (S_{\max}, SD_{50})$, $\sigma_1^2 = Var(\epsilon_1)$, $\sigma_2^2 = Var(\epsilon_2)$ and $\rho = Cov(\epsilon_1, \epsilon_2)$. The range of dose of interest is $[0, 500]$.

We construct locally D-optimal designs for estimating the model parameters and a c-optimal design for estimating utility of a dose. Following Magnusdottir (2013), let the nominal values for the model parameters $E_{\max}, S_{\max}, ED_{50}, SD_{50}$ be $E^*_{\max}, S^*_{\max}, ED^*_{50}, SD^*_{50}$, respectively. To apply our methods in Sections 2 and 3, we first substitute $\mathbf{f}_1(x)$ and $\mathbf{f}_2(x)$ by $\mathbf{z}_1(x)$ and $\mathbf{z}_2(x)$, respectively, where

$$\mathbf{z}_1(x) = \frac{\partial}{\partial \boldsymbol{\theta}_1}\left(E_{\max}\frac{x}{x + ED_{50}}\right)\Big|_{\boldsymbol{\theta}_1 = (E^*_{\max}, ED^*_{50})^\top},$$

$$\mathbf{z}_2(x) = \frac{\partial}{\partial \boldsymbol{\theta}_2}\left(S_{\max}\frac{x}{x + SD_{50}}\right)\Big|_{\boldsymbol{\theta}_2 = (S^*_{\max}, SD^*_{50})^\top}.$$

To implement our strategy, we next discretize the dose interval to form a discrete design space by $S_N = \{500(j-1)/(N-1), j = 1, \cdots, N\}$ with $N = 10001$. Table 4 displays locally D- and c-optimal designs for various parameter values. We observe that (i) the support points of the D-optimal designs are clustered around two or three points, (ii) they all include the extreme dose 500 as a support point, and (iii) the correlation ρ between the two responses seems to have influence on the optimal designs. When $\rho = 0$, the optimal designs are not affected by the ratio σ_2^2/σ_1^2. We also observe that as SD^*_{50}/ED^*_{50} increases, signifying that the difference between the two models increases, the number of support points tends to increase.

Magnusdottir (2013) proposed with justification a composite utility index (CUI) $\text{CUI}(x) = k_1 y_1 - k_2 y_2$ to measure the utility of using the dose x. As before y_1

Table 4: Locally D- and c-optimal designs for the bivariate Emax model when the nominal values are $ED^*_{50} = 1, E^*_{\max} = S^*_{\max} = 1$, $k_1 = k_2 = 1$ and selected values of SD^*_{50}. The support points are x_1, x_2, x_3 and 500 with corresponding weights w_1, w_2, w_3 and $1 - w_1 - w_2 - w_3$. If a point has zero weight, then it is not specified in the table. The computation times are given for the cases of $SD^*_{50} = 5$.

D-optimal		ρ $(\frac{\sigma_2^2}{\sigma_1^2} = 1)$			$\frac{\sigma_2^2}{\sigma_1^2}$ $(\rho = 0)$		
SD^*_{50}		0	±0.5	±0.7	0.5	1.5	5.0
2	x_1	1.4000	1.4000	1.4000	1.4000	1.4000	1.4000
	w_1	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000
3	x_1	1.7000	1.7000	1.7000	1.7000	1.7000	1.7000
	w_1	0.3618	0.3390	0.3142	0.3618	0.3618	0.3618
	x_2	1.7500	1.7500	1.7500	1.7500	1.7500	1.7500
	w_2	0.1382	0.1610	0.1858	0.1382	0.1382	0.1382
5	x_1	2.2000	1.3500	1.0500	2.2000	2.2000	2.2000
	w_1	0.3755	0.2757	0.2611	0.3755	0.3755	0.3755
	x_2	2.2500	4.3500	5.4500	2.2500	2.2500	2.2500
	w_2	0.1245	0.2465	0.2472	0.1245	0.1245	0.1245
	x_3			5.5000			
	w_3			0.0367			
Computation time (sec.)		22.58	33.86	33.31	22.69	22.96	22.82
c-optimal		ρ $(\frac{\sigma_2^2}{\sigma_1^2} = 1)$			$\frac{\sigma_2^2}{\sigma_1^2}$ $(\rho = 0)$		
SD^*_{50}		−0.5	0	0.5	0.5	1.5	5
2	x_1	1.0500	1.1000	1.1000	0.9500	1.2000	1.7500
	w_1	0.0863	0.3948	0.5116	0.3801	0.0700	0.0933
	x_2	1.1000				1.2500	1.8000
	w_2	0.1935				0.3343	0.3494
3	x_1	1.2500	1.2500	0.8500	1.0500	1.4000	1.8500
	w_1	0.2325	0.1331	0.4744	0.1431	0.1617	0.4908
	x_2	1.3000	1.3000	6.0000	1.1000	1.4500	4.8500
	w_2	0.1400	0.3570	0.2194	0.3269	0.3425	0.0791
5	x_1	1.3500	0.9500	0.7000	0.9500	0.9500	0.9500
	w_1	0.3951	0.4363	0.4745	0.4620	0.4198	0.3679
	x_2	7.2500	8.7500	8.7000	8.7000	8.8500	9.0500
	w_2	0.0857	0.0524	0.3837	0.1561	0.3144	0.2976
	x_3		8.8000				9.1000
	w_3		0.1995				0.2196
Computation time (sec.)		32.94	22.83	30.15	22.15	22.79	22.93

represents efficacy and y_2 represents side effect and, k_1 and k_2 are two user-specified positive constants. The goal is to find the most desirable dose that maximizes $E[\text{CUI}(x)]$. Let $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)$ and let $h_c(\boldsymbol{\theta}) = \max_x E[\text{CUI}(x)]$. Magnusdottir (2013) provided selected c-optimal designs that minimize the variance of $h_c(\hat{\boldsymbol{\theta}})$. Table 4 shows the c-optimal designs from our algorithm, where we observe that their support points are also clustered around two or three points. However, the c-optimal designs are more sensitive to the true parameter values than the D-optimal designs. In addition, the D-optimal designs depend only on $|\rho|$, but the c-optimal designs depend on ρ. The results are consistent with those in Magnusdottir (2013) with small differences due to the discretized design space. The designs in Magnusdottir (2013) were derived by informative guessing, which can be difficult for practitioners. Our SDP-based algorithms can provide an effective and systematic way for finding optimal designs, which remain very efficient even for N as large as 10001 in this example. The computation time is about $20 \sim 35$ seconds for each case. □

5 Multi-response generalized linear models

SDPs can also be used to generate optimal designs for GLMs. Consider the maximum likelihood estimator (MLE) to estimate unknown parameters in GLMs. Let $l(\boldsymbol{\theta}|\mathbf{y}, \xi)$ denote the log-likelihood function for a GLM, where $\boldsymbol{\theta} \in R^q$ is the unknown parameter vector, $\mathbf{y}$ is the vector of multi-response variables. If we have resources to take n observations and ξ is the distribution of design points $\mathbf{x}_1, \cdots, \mathbf{x}_n \in R^p$, the $(i,k)^{th}$ element of the Fisher information matrix of the MLE is $A_{i,k} = -E\left(\frac{\partial^2 l(\boldsymbol{\theta}|\mathbf{y},\xi)}{\partial\theta_i \partial\theta_k}\right)$, $i, k = 1, \cdots, q$, and the expectation is taken over the distribution of $\mathbf{y}$. Using the discrete design space S_N for $\mathbf{x}$ and the distribution ξ in (5), we can write all the elements $A_{i,k}$ in $\mathbf{A}(\xi, \boldsymbol{\theta})$ as linear functions of $\mathbf{w}$. Locally optimal designs minimize some scalar functions of $\mathbf{A}(\xi, \boldsymbol{\theta})$ over $\mathbf{w}$. For instance, locally D-optimal designs minimize $-\det(\mathbf{A}(\xi, \boldsymbol{\theta}))$. We now use CVX and SeDuMi to find locally D-optimal, A-optimal and other optimal designs for GLMs and discuss one such application in some detail.

Kpamegan (1998) found locally D-optimal designs for a bivariate probit response model to study the relationship between two toxic responses y_1 and y_2 and the levels of two drugs, x_1 and x_2. The two binary response variables are y_1 and y_2, which take on values 1 for toxicity and 0 for non-toxicity responses. These bivariate responses $F_{ij}(x_1, x_2) = P(y_1 = i, y_2 = j | (x_1, x_2))$, $i,\ j = 0,\ 1$, are modeled using the bivariate normal probability density function (pdf) $\phi(u, v)$ with parameter vector $\boldsymbol{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)^\top$, where μ_1 and μ_2 are the location parameters, σ_1 and σ_2 are the scale parameters and ρ is the correlation parameter for the two response variables. For instance, we have $F_{11}(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} \phi(u, v) du dv$. Kpamegan (1998) found D-optimal designs for the following three cases:
Case (i): Assume $\rho = 0$ and the parameter vector is reduced to $\boldsymbol{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2)^\top$.
Case (ii): Assume $\rho = 0$ and $\sigma_1 = \sigma_2 = \sigma$, and $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma)^\top$.
Case (iii): Assume $\rho \neq 0$ and the parameter vector is $\boldsymbol{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2, \rho)^\top$.

We construct D-optimal designs for the three cases below and compare the results with those in Kpamegan (1998). Let $S_N = \{(x_{1j}, x_{2j}), j = 1, \cdots, N\}$ contain N possible combinations of the levels of the two drugs. Define transformed variables $z_{ij} = (x_{ij} - \mu_i)/\sigma_i$, $j = 1, \cdots, N$, $i = 1, 2$. In Case (i), the Fisher information matrix is given by

$$\mathbf{A}_1(\xi, \boldsymbol{\theta}) = N \sum_{j=1}^{N} w_j \begin{pmatrix} \frac{1}{\sigma_1^2} m(z_{1j}) & \frac{1}{\sigma_1^2} z_{1j} m(z_{1j}) & 0 & 0 \\ \frac{1}{\sigma_1^2} z_{1j} m(z_{1j}) & \frac{1}{\sigma_1^2} z_{1j}^2 m(z_{1j}) & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_2^2} m(z_{2j}) & \frac{1}{\sigma_2^2} z_{2j} m(z_{2j}) \\ 0 & 0 & \frac{1}{\sigma_2^2} z_{2j} m(z_{2j}) & \frac{1}{\sigma_2^2} z_{2j}^2 m(z_{2j}) \end{pmatrix},$$

where $m(\cdot) = \phi^2(\cdot)/[\Phi(\cdot)(1 - \Phi(\cdot))]$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cumulative distribution function of the univariate standard normal random variable. In Case (ii), the Fisher information matrix is given by

$$\mathbf{A}_2(\xi, \boldsymbol{\theta}) = \frac{N}{\sigma^2} \sum_{j=1}^{N} w_j \begin{pmatrix} m(z_{1j}) & 0 & z_{1j} m(z_{1j}) \\ 0 & m(z_{2j}) & z_{2j} m(z_{2j}) \\ z_{1j} m(z_{1j}) & z_{2j} m(z_{2j}) & z_{1j}^2 m(z_{1j}) + z_{2j}^2 m(z_{2j}) \end{pmatrix}.$$

Both $\mathbf{A}_1(\xi, \boldsymbol{\theta})$ and $\mathbf{A}_2(\xi, \boldsymbol{\theta})$ are linear in $\mathbf{w}$. As in Kpamegan (1998), we find optimal designs in terms of standardized variables z_1 and z_2, where $z_i = (x_i - \mu_i)/\sigma_i$, $i = 1, 2$. Let $z_{il} = -3 + 6(l-1)/(N_0 - 1)$, $l = 1, \cdots, N_0$, $i = 1, 2$. The design space $S_N \subset [-3, 3] \times [-3, 3]$ contains $N = N_0^2$ grid points formed by z_{1l} and z_{2l}. Using CVX program we compute the D-optimal designs for Case (i) with $N = 101^2 = 10,201$ and Case (ii) with $N = 201^2 = 40,401$ and obtain the results in Table 5. It took less than 25 seconds for Case (i) and less than 4 minutes for Case (ii). Our results are similar to those in Kpamegan (1998), which are also listed in Table 5. The search is reduced to one dimensional in Kpamegan (1998) after analyzing the information matrix and using the symmetry of the D-optimal designs, which worked well for Cases (i) and (ii). However, if the information matrix is complicated, the method discussed in Kpamegan (1998) may not work and we will show this for Case (iii).

Table 5: D-optimal designs for the bivariate probit response model. The computation times are 23.1 and 215.2 seconds for Cases (i) and (ii), respectively.

Case/Method	Support points [weights]	function value
Case (i): $\sigma_1 = \sigma_2 = 1$		
CVX ($N = 101^2$)	$(-1.140, -1.140)$, $(-1.140, 1.140)$, $(1.140, -1.140)$, $(1.140, 1.140)$ [0.25], [0.25], [0.25], [0.25]	$\det(\mathbf{A}_1/N) = 0.0394748$
Kpamegan (1998)	$(-1.138, -1.138)$, $(-1.138, 1.138)$, $(1.138, -1.138)$, $(1.138, 1.138)$ [0.25], [0.25], [0.25], [0.25]	$\det(\mathbf{A}_1/N) = 0.0394752$
Case (ii): $\sigma = 1$		
CVX ($N = 201^2$)	$(-0.940, -0.940)$, $(-0.940, 0.940)$, $(0.940, -0.940)$, $(0.940, 0.940)$ [0.25], [0.25], [0.25], [0.25]	$\det(\mathbf{A}_2/N) = 0.1703124$
Kpamegan (1998)	$(-0.937, -0.937)$, $(-0.937, 0.937)$, $(0.937, -0.937)$, $(0.937, 0.937)$ [0.25], [0.25], [0.25], [0.25]	$\det(\mathbf{A}_2/N) = 0.1703150$

In Case (iii), the Fisher information matrix $\mathbf{A}_3(\xi, \boldsymbol{\theta})$ (5×5 matrix) is also linear in $\mathbf{w}$. Since $\mathbf{A}_3(\xi, \boldsymbol{\theta})$ contains many complicated expressions involving integrals and derivatives and it takes 2 pages to display it, we do not show it here. Notice that there are typos in expressions R, U and V for $\mathbf{A}_3(\xi, \boldsymbol{\theta})$ in Kpamegan (1998, page 70), and each expression is missing a factor. Expressions of R, U and V should include factors $\frac{\partial F_{11}}{\partial \rho}$, $\frac{\partial F_{11}}{\partial \rho}$ and $\left(\frac{\partial F_{11}}{\partial \rho}\right)^2$, respectively, and detailed expressions are given in Yin (2017, p76). $\mathbf{A}_3(\xi, \boldsymbol{\theta})$ is much more complicated than $\mathbf{A}_1(\xi, \boldsymbol{\theta})$ or $\mathbf{A}_2(\xi, \boldsymbol{\theta})$, and it is extremely hard to find the D-optimal design for Case (iii). Because of computational difficulties, Kpamegan (1998) only considered symmetric designs, which have equal weights on four support points $(-a, -a)$, $(-a, a)$, $(a, -a)$, (a, a) for some positive number a, to find the D-optimal design. Using CVX, we can find the D-optimal design without any assumption on the design. Some representative results are given in Table 6, which shows that the optimal designs found by CVX are better than those in Kpamegan (1998). For $\rho > 0$, the D-optimal designs do not have equal weights on support points $(-a, -a)$, $(-a, a)$, $(a, -a)$, (a, a). In fact, they have equal weights on points $(-a, -a)$ and (a, a). A theoretical result about symmetry of optimal designs is derived in Section 6. This application shows that our algorithms can find optimal designs for very complicated nonlinear/GLM models.

Table 6: D-optimal designs for different values of ρ for the GLM

D-optimal designs by CVX					D-optimal designs in Kpamegan (1998)	
ρ	N	support points	weights	det($\mathbf{A}_3/N$)	support points (weighs)	det($\mathbf{A}_3/N$)
0	61^2	(-1.0000, -1.0000)	0.0838	0.0072		0.0061
		(-1.0000, 1.0000)	0.0838			
		(-0.9000, -0.9000)	0.1662		$(1.138, 1.138)$ (0.2500)	
		(-0.9000, 0.9000)	0.1662		$(-1.138, 1.138)$ (0.2500)	
		(0.9000, -0.9000)	0.1662		$(1.138, -1.138)$ (0.2500)	
		(0.9000, 0.9000)	0.1662		$(-1.138, -1.138)$ (0.2500)	
		(1.0000, -1.0000)	0.0838			
		(1.0000, 1.0000)	0.0838			
0.1	101^2	(-0.9600, -0.9600)	0.4390	0.0077	$(0.94, 0.94)$ (0.2500)	0.0074
		(-0.9000, -0.9000)	0.0610		$(-0.94, 0.94)$ (0.2500)	
		(0.9000, 0.9000)	0.0610		$(0.94, -0.94)$ (0.2500)	
		(0.9600, 0.9600)	0.4390		$(-0.94, -0.94)$ (0.2500)	
0.5	11^2	(-1.2000,-1.2000)	0.3264	0.0134	$(0.89, 0.89)$ (0.2500)	0.0119
		(-0.6000,-0.6000)	0.1736		$(-0.89, 0.89)$ (0.2500)	
		(0.6000, 0.6000)	0.1736		$(0.89, -0.89)$ (0.2500)	
		(1.2000, 1.2000)	0.3264		$(-0.89, -0.89)$ (0.2500)	

6 Properties of optimal designs

In this section we derive several theoretical results for optimal designs for multi-response models. These results are helpful for computing optimal designs, and in some situations we can reduce the computation time by reducing q and/or N. Theorem 3 can be used to reduce q, Theorem 6 can be applied to reduce N, and Theorems 4 and 5 are related to the covariance matrix Σ_0.

Let $\Sigma_0^{-1} = (s_{il})_{r\times r}$ and let $\mathbf{A}_{il}(\mathbf{w}) = \sum_{j=1}^{N} w_j \mathbf{f}_i(\mathbf{u}_j)\mathbf{f}_l^\top(\mathbf{u}_j)$, $i, l = 1, \cdots, r$. By (6), we obtain

$$\mathbf{A}(\mathbf{w}) = \begin{pmatrix} s_{11}\mathbf{A}_{11}(\mathbf{w}) & \cdots & s_{1r}\mathbf{A}_{1r}(\mathbf{w}) \\ s_{21}\mathbf{A}_{21}(\mathbf{w}) & \cdots & s_{2r}\mathbf{A}_{2r}(\mathbf{w}) \\ \vdots & \cdots & \vdots \\ s_{r1}\mathbf{A}_{r1}(\mathbf{w}) & \cdots & s_{rr}\mathbf{A}_{rr}(\mathbf{w}) \end{pmatrix}, \qquad (19)$$

which implies the result in Lemma 2 and leads to the result in Theorem 3.

Lemma 2. *If the vector of regression functions are all the same for the r response variables, i.e., $\mathbf{f}_1(\mathbf{x}) = \mathbf{f}_2(\mathbf{x}) = \cdots = \mathbf{f}_r(\mathbf{x})$, the matrix $\mathbf{A}(\mathbf{w})$ in (19) becomes*

$$\mathbf{A}(\mathbf{w}) = \Sigma_0^{-1} \otimes \mathbf{A}_{11}(\mathbf{w}), \qquad (20)$$

where $\otimes$ denotes Kronecker product and $\mathbf{A}_{11}(\mathbf{w}) = \sum_{j=1}^{N} w_j \mathbf{f}_1(\mathbf{u}_j)\mathbf{f}_1^\top(\mathbf{u}_j)$.

By properties of Kronecker product of matrices, it is immediate that (i) $\mathbf{A}^{-1}(\mathbf{w}) = \Sigma_0 \otimes \mathbf{A}_{11}^{-1}(\mathbf{w})$, (ii) $det\,(\mathbf{A}^{-1}(\mathbf{w})) = (det\,(\Sigma_0))^{q_1}\,(det\,(\mathbf{A}_{11}(\mathbf{w})))^{-r}$ and (iii) trace $(\mathbf{A}^{-1}(\mathbf{w})) =$ trace$(\Sigma_0) \cdot$ trace $\left(\mathbf{A}_{11}^{-1}(\mathbf{w})\right)$, which directly leads to the following result.

Theorem 3. *Under the assumption in Lemma 2, (i) the A- and D-optimal designs for the multi-response model do not depend on Σ_0, and (ii) the A- and D-optimal designs for the multi-response model are the same as those for one-response model with response function $\mathbf{f}_1^\top(\mathbf{x})\boldsymbol{\theta}_1$.*

The above results are more general than those in Krafft and Schaefer (1992) and Chang et al. (2001), where they considered models with one independent variable only. Using the result in Theorem 3 (ii), we can reduce q to q_1 for computing A- and D-optimal designs when $\mathbf{f}_1(\mathbf{x}) = \mathbf{f}_2(\mathbf{x}) = \cdots = \mathbf{f}_r(\mathbf{x})$.

If the r vectors of regressors $\mathbf{f}_1(\mathbf{x}), \cdots, \mathbf{f}_r(\mathbf{x})$ are not the same, optimal designs may depend on Σ_0. See the numerical results in Example 3. When $r = 2$ and

$$\Sigma_0 = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad -1 < \rho < 1, \tag{21}$$

we have $s_{11} = s_{22} = 1/(1-\rho^2)$ and $s_{12} = s_{21} = -\rho/(1-\rho^2)$. For this situation, various optimal designs have been studied, see for example, Chang et al. (2001) and Atashgah and Seifi (2007, 2009). Theorem 4 and the results that follow provide further insights into such optimal designs. Their proofs are in the Appendix.

Theorem 4. *Suppose $r = 2$ and Σ_0 is given by (21). The A-, A_s- and D-optimal designs depend on Σ_0 only through $|\rho|$. In addition, if the two response functions are nested, say $\mathbf{f}_2^\top(\mathbf{x}) = (\mathbf{f}_1^\top(\mathbf{x}), \mathbf{g}_1^\top(\mathbf{x}))$ for some vector function $\mathbf{g}_1^\top(\mathbf{x})$, then the D-optimal design does not depend on Σ_0.*

If Σ_0 depends only on a correlation coefficient ρ, we write it as $\Sigma_0(\rho)$. Also note that if $\mathbf{Q}$ is a $q \times q$ diagonal matrix with diagonal elements equal to 1 or -1, we have (i) $\mathbf{Q}^{-1} = \mathbf{Q}$, (ii) det $(\mathbf{Q}\mathbf{A}^{-1}(\mathbf{w})\mathbf{Q})$ =det$(\mathbf{A}^{-1}(\mathbf{w}))$ and (iii) trace $(\mathbf{Q}\mathbf{A}^{-1}(\mathbf{w})\mathbf{Q})$ = trace $(\mathbf{A}^{-1}(\mathbf{w}))$, which gives the following result.

Theorem 5. *In (6), if $\Sigma_0^{-1}(\rho)$ satisfies*

$$\mathbf{A}(\mathbf{w}) = \sum_{j=1}^{N} w_j \mathbf{U}_j^\top {\Sigma_0}^{-1}(\rho)\mathbf{U}_j = \mathbf{Q} \sum_{j=1}^{N} w_j \mathbf{U}_j^\top {\Sigma_0}^{-1}(|\rho|)\mathbf{U}_j \mathbf{Q},$$

where $\mathbf{Q} = \mathbf{Q}_1 \oplus \mathbf{Q}_2 \oplus \cdots \oplus \mathbf{Q}_r$ with $\mathbf{Q}_i = \pm \mathbf{I}_{q_i}$, then A-, A_s- and D-optimal designs depend on Σ_0 only through $|\rho|$.

The numerical results in Examples 3 and 4 are consistent with Theorems 4 and 5, which may be used to reduce the computation by having fewer choices of ρ.

Now we explore invariance properties of optimal designs which includes the symmetric property. Let T be an one-to-one function: $S_N \mapsto S_N$ and T^2 is an identity map. For a distribution $\xi(\mathbf{x})$ defined in (5), let

$$\xi(T(\mathbf{x})) = \begin{pmatrix} T(\mathbf{u}_1) & T(\mathbf{u}_2) & \cdots & T(\mathbf{u}_N) \\ w_1 & w_2 & \cdots & w_N \end{pmatrix}. \tag{22}$$

The distribution ξ is called $T-$invariant if $\xi(\mathbf{x}) = \xi(T(\mathbf{x}))$. A $T-$invariant distribution ξ implies that $w_i = w_j$ if $T(\mathbf{u}_i) = \mathbf{u}_j$.

Theorem 6. *Suppose model (2) is defined on a given discretized design space S_N and T is an one-to-one function with T^2 being an identity map. If there exists r diagonal matrices $\mathbf{Q}_1, \cdots, \mathbf{Q}_r$ with diagonal elements taking values 1 or -1 and $\mathbf{f}_i(T(\mathbf{x})) = \mathbf{Q}_i\mathbf{f}_i(\mathbf{x})$ for all $\mathbf{x} \in S_N$, $i = 1, \cdots, r$, then there exists $T-$invariant A-, A_s- and D-optimal designs for any Σ_0.*

From the proof of Theorem 6 in the Appendix for A- and D-optimality, one can relax the requirement that $\mathbf{Q}_i$'s be diagonal. For a function T, if the $\mathbf{Q}_i$'s satisfy $\mathbf{Q}\mathbf{Q}^\top = \mathbf{I}$, where $\mathbf{Q} = \mathbf{Q}_1 \oplus \cdots \oplus \mathbf{Q}_r$, one can show that there exists $T-$invariant A- and D-optimal designs. Example 5 below shows how the result in Theorem 6 can be applied to reduce the computation time.

Example 5. Consider a three-response model with two independent variables x_1 and x_2, $\mathbf{f}_1^\top(\mathbf{x}) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$, $\mathbf{f}_2(\mathbf{x}) = \mathbf{f}_1(\mathbf{x}), \mathbf{f}_3^\top(\mathbf{x}) = (1, x_1, x_2)$, and

$$S_9 = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -\sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -\sqrt{2} \end{pmatrix}, \begin{pmatrix} 0 \\ \sqrt{2} \end{pmatrix} \right\}.$$

Define three functions

$$T_1(\mathbf{x}) = -\mathbf{x}, \quad T_2(\mathbf{x}) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} \quad \text{and} \quad T_3(\mathbf{x}) = \begin{pmatrix} -x_1 \\ x_2 \end{pmatrix}.$$

By Theorem 6 for T_1 and T_3 and the comments below Theorem 6 for T_2, we obtain an A-optimal design with equal weights at points $\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$ and $\mathbf{u}_5$ and equal weights at $\mathbf{u}_6, \mathbf{u}_7, \mathbf{u}_8$ and $\mathbf{u}_9$. Similarly, the same result holds for a D-optimal design. In the computation, the number of unknown weights w_i's can be reduced from 9 to

$(1 + (N-1)/4) = 3$. This property is very useful to reduce the computation time for huge N. $\square$

In Example 2, we have $N = 10,000$ and $19,600$. By Theorem 6, we can reduce N to $10,000/4 = 2,500$ and $19,600/4 = 4,900$, respectively, if we use the property that the optimal designs are symmetric about the variables x_3 and x_4. In Example 2, x_3 takes 4 grid points $-1, -0.5, 0.5, 1$, and x_4 takes 4 grid points $-0.5, -0.25, 0.25, 0.5$. There are 16 grid points for x_3 and x_4. Using the invariance property in Theorem 6, we only need to include 4 grid points for x_3 and x_4 for this model, so the total number of points in the design space is reduced by a factor of 1/4. In the computation we need to combine the terms with equal weights in matrix $\mathbf{A}(\mathbf{w})$ in (6) so that there are only $N/4$ unknown weights w_i's, and $\sum w_i = 1/4$. The computation times are indicated by the cases with ** in Table 1, which show significant reduction in computation times using the invariance property. Note that there are other ways to use the invariance property for this model and we can reduce N further. Thus, Theorem 6 is very helpful for finding optimal designs for huge N in this example.

For GLMs, we can use the following result to check for invariance property. Let $\mathbf{A}(\xi, \boldsymbol{\theta}) = \sum_{j=1}^{N} w_j \mathbf{B}(\mathbf{u}_j; \boldsymbol{\theta})$, for some symmetric and PSD matrices $\mathbf{B}(\mathbf{u}_j; \boldsymbol{\theta})$.

Theorem 7. *Let T be an one-to-one function defined on a given design space S_N and T^2 is an identity map. If there exists a diagonal matrix $\mathbf{Q}$ with diagonal elements taking values -1 or 1 such that $\mathbf{B}(T(\mathbf{u}_j); \boldsymbol{\theta}) = \mathbf{Q}\mathbf{B}(\mathbf{u}_j; \boldsymbol{\theta})\mathbf{Q}$, for all $\mathbf{u}_j \in S_N$, $j = 1, \cdots, N$, then there exists $T-$invariant A- and D-optimal designs.*

The proof of Theorem 7 is similar to that of Theorem 6 and is omitted. Theorem 7 can be applied for the optimal designs in Table 6.

7 Conclusion

We have studied SDP transformations for various optimal design problems for multi-response regression models and applied CVX and SeDuMi algorithms for finding A-,

A_s- and c-optimal designs by solving SDP problems. The algorithms are effective and efficient, and they can be used for finding optimal designs systematically for any discrete design spaces. They can also find I-optimal and L-optimal designs.

We have also investigated A-, A_s- and D-optimal designs for multi-response regression models using the invariance property and dependence on the error covariance matrix Σ_0. Symmetry of optimal designs can be examined through the $T-$invariance property obtained in the paper, and it can be applied for various models and design spaces. Those theoretical results are useful to reduce the computation time for finding optimal designs for various models.

In this paper we have assumed that Σ_0 is known and the BLUE is used to estimate the regression parameter vector. When Σ_0 is unknown, we need an estimate of Σ_0 to compute the BLUE. However, it may be interesting to study robust regression designs against possible misspecification in Σ_0. Wiens (2015) reviewed robustness of design, which includes several approaches for constructing robust designs for one response models. Some of the approaches can be developed for robust designs for multi-response models, and the resulting designs should be robust against possible misspecification in Σ_0 and still maintain high efficiency for the estimator of $\boldsymbol{\theta}$.

Appendix: Proofs

Proof of Theorem 1: The constraint in problem (12) is true if and only if $\mathbf{M}_i \succeq 0$ for all $i = 1, \cdots, q$. Since $\mathbf{M} \succ 0$ (positive definite), $\mathbf{M}_i \succeq 0$ is true if and only if $v_i - \mathbf{e}_i^\top \mathbf{M}^{-1}\mathbf{e}_i \geq 0$ (Horn and Johnson, 2009, p472). This gives $v_i - b_{ii} \geq 0$, $i = 1, \cdots, q$, where b_{ii} is the ith diagonal element of $\mathbf{M}^{-1}$ and so $v_1 + \cdots + v_q$ is minimized by $v_1^* = b_{11}, \cdots, v_q^* = b_{qq}$, and $v_1^* + \cdots + v_q^* = b_{11} + \cdots + b_{qq} = \text{trace}\,(\mathbf{M}^{-1})$. □

Proof of Theorem 2: From Horn and Johnson (2009, p472), the constraint in problem (13) is true if and only if $\mathbf{V} - \mathbf{I}_q\mathbf{M}^{-1}\mathbf{I}_q \succeq 0$, which is

$$\mathbf{V} - \mathbf{M}^{-1} \succeq 0. \tag{23}$$

Therefore $v_1 + \cdots + v_q = \text{trace}\,(\mathbf{V}) \geq \text{trace}\,(\mathbf{M}^{-1})$ and the minimizer of $v_1 + \cdots + v_q$ must satisfy $v_1^* + \cdots + v_q^* \geq \text{trace}\,(\mathbf{M}^{-1})$. Let $\mathbf{M}^{-1} = (b_{ij})_{q\times q}$. If $\mathbf{M}$ is a diagonal matrix, so is $\mathbf{M}^{-1}$ and by (23), $v_i - b_{ii} \geq 0$, $i = 1, \cdots, q$ so that the minimizer of $v_1 + \cdots + v_q$ is $v_1^* = b_{11}, \cdots, v_q^* = b_{qq}$ and $v_1^* + \cdots + v_q^* = \text{trace}\,(\mathbf{M}^{-1})$. If $\mathbf{M}$ is not a diagonal matrix, $\mathbf{M}^{-1}$ is still symmetric but not diagonal. This means that there exists at least two off-diagonal nonzero elements that are nonzero. Without loss of generality we assume $b_{12} = b_{21} \neq 0$. By (23), we still have $v_i - b_{ii} \geq 0$, $i = 1, \cdots, q$. However, $\mathbf{V} = \text{diag}(b_{11}, \cdots, b_{qq})$ does not satisfy (23) since $\mathbf{a}^\top (\mathbf{V} - \mathbf{M}^{-1})\,\mathbf{a} = -2|b_{12}| < 0$, with $\mathbf{a} = (1, \text{sign}(b_{12}), 0, \cdots, 0)^\top \in R^q$. Thus, the minimizer of $v_1 + \cdots + v_q$ must satisfy $v_1^* + \cdots + v_q^* > b_{11} + \cdots + b_{qq} = \text{trace}\,(\mathbf{M}^{-1})$. □

Proof of Theorem 4: By (19), when $r = 2$ we have

$$\mathbf{A}(\mathbf{w}) = \begin{pmatrix} s_{11}\mathbf{A}_{11}(\mathbf{w}) & s_{12}\mathbf{A}_{12}(\mathbf{w}) \\ s_{21}\mathbf{A}_{21}(\mathbf{w}) & s_{22}\mathbf{A}_{22}(\mathbf{w}) \end{pmatrix},$$

where for simplicity, we write $\mathbf{A}_{ij}(\mathbf{w})$ as $\mathbf{A}_{ij}$. Let $\mathbf{G} = \mathbf{A}_{22} - \rho^2 \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, let $s_{11} = s_{22} = 1/(1-\rho^2)$ and let $s_{12} = s_{21} = -\rho/(1-\rho^2)$. One verifies that

$$\det(\mathbf{A}(\mathbf{w})) = \left(\frac{1}{1-\rho^2}\right)^q \cdot \det(\mathbf{A}_{11}) \cdot \det(\mathbf{G}),$$

$$\mathbf{A}^{-1}(\mathbf{w}) = (1-\rho^2) \cdot \begin{pmatrix} \mathbf{A}_{11}^{-1} + \rho^2 \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{G}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \rho\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{G}^{-1} \\ \rho\mathbf{G}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{G}^{-1} \end{pmatrix}, \quad \text{and}$$

$$\text{trace}\left(\mathbf{A}^{-1}(\mathbf{w})\right) = (1-\rho^2) \cdot \text{trace}\left(\mathbf{A}_{11}^{-1} + \rho^2 \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{G}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} + \mathbf{G}^{-1}\right).$$

It follows that A- and D-optimal designs depend on Σ_0 only through $|\rho|$, and the same is true for any A_s-optimal design.

If the two response models are nested, say $\mathbf{f}_2^\top(\mathbf{x}) = (\mathbf{f}_1^\top(\mathbf{x}), \mathbf{g}_1^\top(\mathbf{x}))$, we may write

$$\mathbf{A}_{21} = \begin{pmatrix} \mathbf{A}_{11} \\ \tilde{\mathbf{A}}_{21} \end{pmatrix}, \quad \mathbf{A}_{12} = \mathbf{A}_{21}^\top, \quad \mathbf{A}_{22} = \begin{pmatrix} \mathbf{A}_{11} & \tilde{\mathbf{A}}_{21}^\top \\ \tilde{\mathbf{A}}_{21} & \tilde{\mathbf{A}}_{22} \end{pmatrix},$$

where $\tilde{\mathbf{A}}_{21} = \sum_{j=1}^N w_j \mathbf{g}_1(\mathbf{u}_j)\mathbf{f}_1^\top(\mathbf{u}_j)$ and $\tilde{\mathbf{A}}_{22} = \sum_{j=1}^N w_j \mathbf{g}_1(\mathbf{u}_j)\mathbf{g}_1^\top(\mathbf{u}_j)$. Consequently,

$$\det(\mathbf{G}) = \det\left((1-\rho^2)\mathbf{A}_{11}\right) \cdot \det\left(\tilde{\mathbf{A}}_{22} - \tilde{\mathbf{A}}_{21}\mathbf{A}_{11}^{-1}\tilde{\mathbf{A}}_{21}^\top\right),$$

which implies that the D-optimal design does not depend on ρ or Σ_0. $\square$

Proof of Theorem 6: For any distribution $\xi(\mathbf{x})$, let $\xi_1(\mathbf{x}) = \xi(T(\mathbf{x}))$ as in (22). After rearranging the columns, we write

$$\xi_1(\mathbf{x}) = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_N \\ \tilde{w}_1 & \tilde{w}_2 & \cdots & \tilde{w}_N \end{pmatrix}$$

and note that $\tilde{w}_j = w_i$ and $\tilde{w}_i = w_j$ if $T(\mathbf{u}_i) = \mathbf{u}_j$. Let $\tilde{\mathbf{w}} = (\tilde{w}_1, \cdots, \tilde{w}_N)$ and let $\xi_{0.5}(\mathbf{x}) = 0.5\xi(\mathbf{x}) + 0.5\xi_1(\mathbf{x})$, i.e.,

$$\xi_{0.5}(\mathbf{x}) = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_N \\ 0.5(w_1 + \tilde{w}_1) & 0.5(w_2 + \tilde{w}_2) & \cdots & 0.5(w_N + \tilde{w}_N) \end{pmatrix}.$$

If $T(\mathbf{u}_i) = \mathbf{u}_j$, $\xi_{0.5}(\mathbf{x})$ has the same weight at $\mathbf{u}_i$ and $\mathbf{u}_j$ and so $\xi_{0.5}(\mathbf{x})$ is $T-$invariant.

Let $\mathbf{Q} = \mathbf{Q}_1 \oplus \cdots \oplus \mathbf{Q}_r$. Since $\mathbf{f}_i(T(\mathbf{x})) = \mathbf{Q}_i \mathbf{f}_i(\mathbf{x})$, we have

$$\mathbf{A}(\tilde{\mathbf{w}}) \;=\; \sum_{j=1}^{N} \tilde{w}_j \mathbf{U}_j^\top {\Sigma_0}^{-1} \mathbf{U}_j = \sum_{j=1}^{N} w_j \mathbf{Q} \mathbf{U}_j^\top {\Sigma_0}^{-1} \mathbf{U}_j \mathbf{Q}^\top = \mathbf{Q}\mathbf{A}(\mathbf{w})\mathbf{Q}^\top.$$

Since $\det(\mathbf{A}(\tilde{\mathbf{w}})) = \det(\mathbf{A}(\mathbf{w}))$, $\text{trace}(\mathbf{A}^{-1}(\tilde{\mathbf{w}})) = \text{trace}(\mathbf{A}^{-1}(\mathbf{w}))$ and $\text{trace}(\mathbf{L}^\top \mathbf{A}^{-1}(\tilde{\mathbf{w}})\mathbf{L}) = \text{trace}(\mathbf{L}^\top \mathbf{A}^{-1}(\mathbf{w})\mathbf{L})$, where $\mathbf{L}$ is defined in (9) and they are all convex in $\mathbf{w}$, we have $\det\,(\mathbf{A}(0.5(\mathbf{w}+\tilde{\mathbf{w}}))) \leq \det\,(\mathbf{A}(\mathbf{w}))$, $\text{trace}\,(\mathbf{A}^{-1}(0.5(\mathbf{w}+\tilde{\mathbf{w}}))) \leq \text{trace}\,(\mathbf{A}^{-1}(\mathbf{w}))$ and $\text{trace}\,(\mathbf{L}^\top \mathbf{A}^{-1}(0.5(\mathbf{w}+\tilde{\mathbf{w}}))\mathbf{L}) \leq \text{trace}\,(\mathbf{L}^\top \mathbf{A}^{-1}(\mathbf{w})\mathbf{L})$ for any $\xi(\mathbf{x})$. It follows that the weight vector $0.5(\mathbf{w} + \tilde{\mathbf{w}})$ corresponds to the $T-$invariant distribution $\xi_{0.5}(\mathbf{x})$, which implies that there exist $T-$invariant A-, A_s- and D-optimal designs. $\square$

Acknowledgements

This research work was partially supported by Discovery Grants from the Natural Science and Engineering Research Council of Canada. The research of Wong was partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639.

References

Atashgah, A.B. and Seifi, A. (2007). Application of semi-definite programming to the design of multi-response experiments. *IIE Transactions*, 39, 763-769.

Atashgah, A.B. and Seifi, A. (2009). Optimal design of multi-response experiments using semi-definite programming. *Optimization and Engineering*, 10, 75-90.

Berger, M.P.F. and Wong, W.K. (2009). *An Introduction to Optimal Designs for Social and Biomedical Research.* Wiley, Chichester, UK.

Bischoff, W. (1993). On D-optimal designs for linear models under correlated observations with an application to a linear model with multiple response. *Journal of Statistical Planning and Inference*, 37, 69-80.

Bischoff, W. (1995). Determinant formulas with applications to designing when the observations are correlated. *Annals of the Institute of Statistical Mathematics*, 47, 385-399.

Bose, M. and Mukerjee, R. (2015). Optimal design measures under asymmetric errors, with application to binary design points. *Journal of Statistical Planning and Inference*, 159, 28-36.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization.* Cambridge University Press, New York.

Chang, S. (1997). An algorithm to generate near D-optimal designs for multiple response surface models. *IIE Transactions*, 29, 1073-1081.

Chang, F.C., Huang, M.N.L., Lin, D.K.J. and Yang, H.C. (2001). Optimal designs for dual response polynomial regression models. *Journal of Statistical Planning and Inference*, 93, 309-322.

Dean, A., Morris, M., Stufken, J. and Bingham, D. (2015). *Handbook of Design and Analysis of Experiments.* CRC Press, New York.

Dette, H., Hoyden, L., Kuhnt, S. and Schorning, K. (2013). Optimal designs for multi-response generalized linear models with applications in thermal spraying. *http://arxiv.org/pdf/1312.4472.pdf*

Dragalin, V., Fedorov, V.V. and Wu, Y. (2008). Two-stage design for dose-finding that accounts for both efficacy and safety. *Statistics in Medicine*, 27, 5156-5176.

Duarte, B.P.M. and Wong, W.K. (2015). Finding Bayesian optimal designs for nonlinear models: a semidefinite programming-based approach. *International Statistical Review*, 83, 239-262.

Duarte, B.P.M., Wong, W.K. and Oliveira, N.M.C. (2016). Model-based optimal design of experiments - semidefinite and nonlinear programming formulations. *Chemometrics and Intelligent Laboratory Systems*, 151, 153-163.

Duarte, B.P.M., Sagnol, G. and Wong, W.K. (2016*). An algorithm based on semidefinite programming for finding minimax optimal designs. *Statistics and Computing*, in press.

Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

Fedorov, V., Gagnon, R. and Leonov, S. (2001). Optimal design for multiple responses with variance depending on unknown parameters. GlaxoSmithKline Pharmaceuticals, GSK BDS Technical Report, August 2001.

Fedorov, V.V. and Wu, Y. (2007). Dose finding designs for continuous responses and binary utility. *Journal of Biopharmaceutical Statistics*, 17, 1085-1096.

Fedorov, V.V., Wu, Y. and Zhang, R. (2012). Optimal dose-finding designs with correlated continuous and discrete responses. *Statistics in Medicine*, 31, 217-234.

Grant, M.C. and Boyd, S.P. (2013). *The CVX Users Guide*. Release 2.0 (beta), CVX Research, Inc. (http://cvxr.com/cvx/doc/CVX.pdf, October 14, 2013.)

Gueorguieva, I., Aarons, L., Ogungbenro, K., Jorga, K.M., Rodgers, T., Rowland, M. (2006). Optimal design for multivariate response pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamics*, 33, 97-124.

Heise, M.A. and Myers, R.M. (1996). Optimal designs for bivariate logistic regression. *Biometrics*, 52, 613-624.

Horn, R.A. and Johnson, C.R. (2009). *Matrix Analysis*. Cambridge University Press, New York.

Huang, M.L., Chen, R.B., Lin, C.S. and Wong, W.K. (2006). Optimal designs for parallel models with correlated responses. *Statistica Sinica*, 16, 121-133.

Imhof, L. (2000) Optimum designs for a multiresponse regression model. *Journal of Multivariate Analysis*, 72, 120-131.

Kim, S. (2014). Optimal experimental design under a new multivariate Weibull regression function. PhD Thesis, Department of Statistics, University of Missouri.

Kpamegan, E.P. (1998). D-optimal designs given a bivariate probit response function, in *New Developments and Applications in Experimental Design*, eds, Flournoy, N., Rosenberger, W.F., and Wong, W.K., 62-72, Institute of Mathematical Statistics, Hayward, California.

Krafft, O. and Schaefer, M. (1992). D-optimal designs for a multivariate regression model. *Journal of Multivariate Analysis*, 42, 130-140.

Liu, X., Yue, R.X. and Hickernell, F.J. (2011). Optimality criteria for multiresponse linear models based on predictive ellipsoids. *Statistica Sinica*, 21, 421-432.

Magnusdottir, B.T. (2013). c-Optimal designs for the bivariate Emax model. In *mODa 10-Advances in Model-Oriented Design and Analysis*, edited by Uciński, D., Atkinson, A.C., and Patan, M., page 153-161. Springer, Switzerland.

Montgomery, D.C. (2013). *Design and Analysis of Experiments*, Eighth Edition. Wiley, New York.

Mukhopadhyay, S., Khuri, A.I. (2008). Comparison of designs for multivariate generalized linear models. *Journal of Statistical Planning and Inference*, 138, 169-183.

Papp, D. (2012). Optimal designs for rational function regression. *Journal of the American Statistical Association*, 107, 400-411.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Sagnol, G. (2011). Computing optimal designs of multiresponse experiments reduces to second-order cone programming. *Journal of Statistical Planning and Inference*, 141, 1684-1708.

Sturm, J.F. (1999). Using SeDuMi 1.02, A Matlab toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11, 625-653.

Wang, R.H. (2000). *Approximate and Exact D-optimal Designs for Multi-response Polynomial Regression Models*. Master Thesis, Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan.

Wiens, D.P. (2015). Robustness of Design, Chapter 20 of *Handbook of Design and Analysis of Experiments*, edited by Dean, A., Morris, M., Stufken, J. and Bingham, D., CRC Press, New York, 719-753.

Yang, M., Biedermann, S. and Tang, E. (2013). On optimal designs for nonlinear models: a general and efficient algorithm. *Journal of the American Statistical Association*, 108, 1411-1420.

Ye, J.J., Zhou, J. and Zhou, W. (2017). Computing A-optimal and E-optimal designs for regression models via semidefinite programming. *Communications in Statistics - Simulation and Computation*, 46, 2011-2024.

Yin, Y. (2017). *Theory and Algorithms for Finding Optimal Regression Designs*. PhD Thesis, Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada.

Yue, R.X. and Liu, X. (2010). I_L^r-optimal designs for a hierarchically ordered system of regression models. *Computational Statistics and Data Analysis*, 54, 3458-3465.

Yue, R.X., Liu, X. and Chatterjee, K. (2014). D-optimal designs for multiresponse linear models with a qualitative factor. *Journal of Multivariate Analysis*, 124, 57-69.