Statistica Sinica Preprint No: SS-2015-0285R2	
Title	Optimal designs for regression models using the second-order least squares estimator
Manuscript ID	SS-2016-0285R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202015.0285

Complete List of Authors	Yue Yin and Julie Zhou
Corresponding Author	Julie Zhou
E-mail	jzhou@uvic.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

Optimal designs for regression models using the second-order least squares estimator

Yue Yin and Julie Zhou[1]

Department of Mathematics and Statistics

University of Victoria

Victoria, BC, Canada V8W 2Y2

Key words and phrases: A-optimal design, convex optimization, D-optimal design, multiplicative algorithm, nonlinear model, SeDuMi, transformation invariance.

MSC 2010: 62K05, 62K20.

ABSTRACT

We investigate properties and numerical algorithms for A- and D-optimal regression designs based on the second-order least squares estimator (SLSE). Several results are derived, including a characterization of the A-optimality criterion. We can formulate the optimal design problems under SLSE as semidefinite programming or convex optimization problems and we show that the resulting algorithms can be faster than more conventional multiplicative algorithms, especially in nonlinear models. Our results also indicate that the optimal designs based on the SLSE are more efficient than those based on the ordinary least squares estimator, provided the error distribution is highly skewed.

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

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

1 Introduction

Consider a regression model to study the relationship between a response variable y and a vector of independent variables $\mathbf{x} \in R^p$,

$$y_i = g(\mathbf{x}_i; \boldsymbol{\theta}) + \epsilon_i, \ i = 1, \cdots, n, \tag{1}$$

where y_i is the observation at $\mathbf{x}_i$, $\boldsymbol{\theta} \in R^q$ is an unknown parameter vector, $g(\mathbf{x}; \boldsymbol{\theta})$ can be a linear or nonlinear function of $\boldsymbol{\theta}$, and the errors ϵ_i's are i.i.d. having mean 0 and variance σ^2. Various optimality criteria have been investigated to construct optimal regression designs. The criteria and optimal designs depend on the response function, design space S, model assumptions and the estimation method of $\boldsymbol{\theta}$; for example, Pukelsheim (1993), Berger and Wong (2009), and Dean *et al.* (2015).

The ordinary least squares estimator (OLSE) is usually used to estimate $\boldsymbol{\theta}$, and optimal designs have been constructed for various models based on it. However, if the error distribution is asymmetric, the second-order least squares estimator (SLSE) in Wang and Leblanc (2008) is more efficient than the OLSE. Using this result, Gao and Zhou (2014) proposed new optimality criteria under SLSE and obtained several results. Bose and Mukerjee (2015) and Gao and Zhou (2015) made further developments, including the convexity results for the criteria and numerical algorithms. Bose and Mukerjee (2015) applied the multiplicative algorithms in Zhang and Mukerjee (2013) for computing the optimal designs, while Gao and Zhou (2015) used the CVX program in MATLAB (Grant and Boyd (2013)).

In this paper we investigate the optimal designs based on the SLSE. Since there are fewer results for A-optimal designs in previous work, this paper fills a gap. Let $\mathbf{u}_1, \cdots, \mathbf{u}_N \in S$ be N possible distinct levels of $\mathbf{x}$. Define a discrete design space $S_N = \{\mathbf{u}_1, \cdots, \mathbf{u}_N\}$, and let Ξ_N be the set of all discrete distributions on S_N. We construct discrete optimal designs from Ξ_N using the optimality criteria in Gao and Zhou (2014). Several results are derived for A- and D-optimal designs. The A-optimality criterion is characterized in a new way, and this leads to an efficient algorithm for computing the designs. The algorithm uses the SeDuMi (Self-Dual-

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

Minimization) program in MATLAB for semidefinite programming (SDP) problems (Boyd and Vandenberghe (2004)). The A-efficiency and D-efficiency of designs are studied to compare the SLSE with the OLSE, and our results indicate that the optimal designs based on the SLSE can be much more efficient than those based on the OLSE, provided the error distribution is highly skewed.

The rest of the paper is organized as follows. In Section 2 we derive properties of optimal designs under SLSE and an expression of the A-optimality criterion. In Section 3 we develop numerical algorithms for computing optimal designs. In Section 4 applications are presented, and numerical algorithms and efficiencies of optimal designs are compared. Concluding remarks are in Section 5. Proofs are given in the Appendix, or in the Supplementary Material.

2 Optimal designs based on the SLSE

In model (1), the SLSE $\hat{\boldsymbol{\gamma}}_{SLS}$ of $\boldsymbol{\gamma} = (\boldsymbol{\theta}^\top, \sigma^2)^\top$ minimizes

$$Q(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \rho_i^\top(\boldsymbol{\gamma}) \mathbf{W}_i \rho_i(\boldsymbol{\gamma}),$$

where vector $\rho_i(\boldsymbol{\gamma}) = (y_i - g(\mathbf{x}_i; \boldsymbol{\theta}), y_i^2 - g^2(\mathbf{x}_i; \boldsymbol{\theta}) - \sigma^2)^\top$ and $\mathbf{W}_i = \mathbf{W}(\mathbf{x}_i)$ is a 2×2 positive semidefinite matrix that may depend on $\mathbf{x}_i$. The most efficient SLSE is obtained by choosing an optimal matrix $\mathbf{W}_i$ to minimize the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}}_{SLS}$, as derived in Wang and Leblanc (2008). For the rest of the paper, the discussion is about the most efficient SLSE. Suppose $\boldsymbol{\theta_0}$ and σ_0^2 are the true values of $\boldsymbol{\theta}$ and σ^2, respectively. Let $\mu_3 = E(\epsilon_1^3 \mid \mathbf{x})$, $\mu_4 = E(\epsilon_1^4 \mid \mathbf{x})$, and $t = \mu_3^2/(\sigma_0^2(\mu_4 - \sigma_0^4))$. Under some regularity conditions (Wang and Leblanc (2008)), the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}}_{SLS}$ is

$$Cov(\hat{\boldsymbol{\gamma}}_{SLS}) = \begin{pmatrix} Cov(\hat{\boldsymbol{\theta}}_{SLS}) & \frac{\mu_3}{\mu_4 - \sigma_0^4} V(\hat{\sigma}_{SLS}^2) \boldsymbol{G}_2^{-1} \boldsymbol{g}_1 \\ \frac{\mu_3}{\mu_4 - \sigma_0^4} V(\hat{\sigma}_{SLS}^2) \boldsymbol{g}_1^\top \boldsymbol{G}_2^{-1} & V(\hat{\sigma}_{SLS}^2) \end{pmatrix}, \tag{2}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

where

$$Cov(\hat{\boldsymbol{\theta}}_{SLS}) = (1-t)\,\sigma_0^2\left(\boldsymbol{G}_2 - t\boldsymbol{g}_1\boldsymbol{g}_1^\top\right)^{-1}, \qquad V(\hat{\sigma}^2_{SLS}) = \frac{(\mu_4-\sigma_0^4)(1-t)}{1-t\boldsymbol{g}_1^\top\boldsymbol{G}_2^{-1}\boldsymbol{g}_1}, \tag{3}$$

$$\boldsymbol{g}_1 = E\left[\frac{\partial g(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right],\ \boldsymbol{G}_2 = E\left[\frac{\partial g(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial g(\mathbf{x};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^\top}\,|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]. \tag{4}$$

The expectation in (4) is taken with respect to the distribution of $\mathbf{x}$. The asymptotic covariance matrix of the OLSE, $\hat{\boldsymbol{\gamma}}_{OLS} = (\hat{\boldsymbol{\theta}}^\top_{OLS}, \hat{\sigma}^2_{OLS})^\top$, is

$$Cov(\hat{\boldsymbol{\gamma}}_{OLS}) = \begin{pmatrix} Cov(\hat{\boldsymbol{\theta}}_{OLS}) & \mu_3\boldsymbol{G}_2^{-1}\boldsymbol{g}_1 \\ \mu_3\boldsymbol{g}_1^\top\boldsymbol{G}_2^{-1} & V(\hat{\sigma}^2_{OLS}) \end{pmatrix} = \begin{pmatrix} \sigma_0^2\boldsymbol{G}_2^{-1} & \mu_3\boldsymbol{G}_2^{-1}\boldsymbol{g}_1 \\ \mu_3\boldsymbol{g}_1^\top\boldsymbol{G}_2^{-1} & \mu_4-\sigma_0^4 \end{pmatrix}. \tag{5}$$

If the error distribution is symmetric, then $\mu_3 = 0$, $t = 0$, and the covariance matrices in (2) and (5) are the same. For asymmetric errors, we have $0 < t < 1$ (Gao and Zhou, (2014)) and $Cov(\hat{\boldsymbol{\gamma}}_{OLS}) - Cov(\hat{\boldsymbol{\gamma}}_{SLS}) \succeq 0$ (positive semidefinite) from Wang and Leblanc (2008), so the SLSE is more efficient than the OLSE.

2.1 A- and D-optimality criteria

In Gao and Zhou (2014), the A- and D-optimal designs based on the SLSE are defined to minimize $\mathrm{tr}(Cov(\hat{\boldsymbol{\theta}}_{SLS}))$ and $\det(Cov(\hat{\boldsymbol{\theta}}_{SLS}))$, respectively, where tr() and det() are the matrix trace and determinant functions. For any distribution $\xi(\mathbf{x}) \in \Xi_N$ of $\mathbf{x}$, let $\xi(\mathbf{x}) = \{(\mathbf{u}_i, w_i) \mid w_i = P(\mathbf{x} = \mathbf{u}_i), \mathbf{u}_i \in S_N, i = 1, \cdots, N\}$, where

$$\sum_{i=1}^N w_i = 1, \quad \text{and} \quad w_i \geq 0, \quad \text{for} \quad i = 1, \cdots, N. \tag{6}$$

Define $\mathbf{f}(\mathbf{x};\boldsymbol{\theta}) = \partial g(\mathbf{x};\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, and write $\boldsymbol{g}_1$ and $\boldsymbol{G}_2$ in (4) as

$$\boldsymbol{g}_1(\mathbf{w}) = \boldsymbol{g}_1(\mathbf{w};\boldsymbol{\theta}_0) = \sum_{i=1}^N w_i\mathbf{f}(\mathbf{u}_i;\boldsymbol{\theta}_0), \boldsymbol{G}_2(\mathbf{w}) = \boldsymbol{G}_2(\mathbf{w};\boldsymbol{\theta}_0) = \sum_{i=1}^N w_i\ \mathbf{f}(\mathbf{u}_i;\boldsymbol{\theta}_0)\mathbf{f}^\top(\mathbf{u}_i;\boldsymbol{\theta}_0), \tag{7}$$

where weight vector $\mathbf{w} = (w_1, \cdots, w_N)^\top$. Let $\mathbf{A}(\mathbf{w}) = \mathbf{A}(\mathbf{w};\boldsymbol{\theta}_0) = \boldsymbol{G}_2(\mathbf{w}) - t\boldsymbol{g}_1(\mathbf{w})\boldsymbol{g}_1^\top(\mathbf{w})$. By (3), the A- and D-optimal designs minimize loss functions

$$\phi_1(\mathbf{w}) = \mathrm{tr}\left((\mathbf{A}(\mathbf{w}))^{-1}\right) \quad \text{and} \quad \phi_2(\mathbf{w}) = \det\left((\mathbf{A}(\mathbf{w}))^{-1}\right) \tag{8}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

over $\mathbf{w}$ satisfying the conditions in (6), respectively. If $\mathbf{A}(\mathbf{w})$ is singular, $\phi_1(\mathbf{w})$ and $\phi_2(\mathbf{w})$ are defined to be $+\infty$. The A- and D-optimal designs are denoted by $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$, respectively. For nonlinear models, since optimal designs often depend on the unknown parameter $\boldsymbol{\theta}_0$, they are called locally optimal designs. For simplicity, we write optimal designs instead of locally optimal designs. Since all the elements of $\boldsymbol{g}_1$ and $\boldsymbol{G}_2$ in (7) are linear functions of $\mathbf{w}$, we have the following from Bose and Mukerjee (2015).

Lemma 1. *$\phi_1(\mathbf{w})$ and $\log(\phi_2(\mathbf{w}))$ are convex functions of $\mathbf{w}$.*

For a discrete distribution on S_N, define

$$\mathbf{B}(\mathbf{w}) = \begin{pmatrix} 1 & \sqrt{t}\ \boldsymbol{g}_1^\top(\mathbf{w}) \\ \sqrt{t}\ \boldsymbol{g}_1(\mathbf{w}) & \boldsymbol{G}_2(\mathbf{w}) \end{pmatrix}, \tag{9}$$

so all the elements of $\mathbf{B}(\mathbf{w})$ are linear functions of $\mathbf{w}$. Gao and Zhou (2015) derived an alternative expression for the D-optimality criterion in Lemma 2.

Lemma 2. *The D-optimal design based on the SLSE minimizes $1/\det(\mathbf{B}(\mathbf{w}))$, and $-log\,(\det(\mathbf{B}(\mathbf{w})))$ and $-(\det(\mathbf{B}(\mathbf{w})))^{1/(q+1)}$ are convex functions of $\mathbf{w}$.*

From (8) and Lemma 2, $\xi^D(\mathbf{x})$ minimizes $1/\det(\mathbf{A}(\mathbf{w}))$ or $1/\det(\mathbf{B}(\mathbf{w}))$. In fact, $\det(\mathbf{A}(\mathbf{w})) = \det(\mathbf{B}(\mathbf{w}))$. It is easier to use $\mathbf{B}(\mathbf{w})$ to develop numerical algorithms for computing the optimal designs in Section 3.

2.2 Properties of $\xi^A(\mathbf{x})$

For some regression models, the transformation invariance property implies the symmetry of $\xi^A(\mathbf{x})$. We derive a characterization of the A-optimality criterion.

Let T be an one-to-one transformation defined on S_N with $T^2\mathbf{u} = \mathbf{u}$ for any $\mathbf{u} \in S_N$. We say that the design space S_N is invariant under transformation T or that S_N is T-invariant. If a distribution $\xi(\mathbf{x})$ on a T-invariant S_N satisfies $P(\mathbf{x} = \mathbf{u}_i) = P(\mathbf{x} = T\mathbf{u}_i)$, for $i = 1, \cdots, N$, we say the distribution is invariant under the

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

transformation T, or T-invariant. To derive transformation invariance properties of $\xi^A(\mathbf{x})$, we order the points in S_N such that $T\mathbf{u}_i = \mathbf{u}_{N-i+1}$ for $i = 1, \cdots, m$ and, if $m < N/2$, $T\mathbf{u}_i = \mathbf{u}_i$ for $i = m+1, \cdots, N-m$. Here the points $\mathbf{u}_{m+1}, \cdots, \mathbf{u}_{N-m}$ are fixed under T. If $m = N/2$, then there are no fixed points in S_N. For T-invariant $\xi(\mathbf{x})$, the weights satisfy $w_i = w_{N-i+1}$ for $i = 1, \cdots, m$. To partially reverse the order of the elements in $\mathbf{w}$, set

$$\text{rev}(\mathbf{w}) = (\underbrace{w_N, \cdots, w_{N-m+1}}, \underbrace{w_{m+1}, \cdots, w_{N-m}}, \underbrace{w_m, \cdots, w_1})^\top.$$

Theorem 1. *Suppose $\xi^A(\mathbf{x})$ is an A-optimal design for a regression model on a T-invariant S_N. If the weight vector of $\xi^A(\mathbf{x})$, $\mathbf{w}^A = (w_1^A, \cdots, w_N^A)$, satisfies*

$$tr\left(\left(\mathbf{A}(\mathbf{w}^A)\right)^{-1}\right) = tr\left(\left(\mathbf{A}(rev(\mathbf{w}^A))\right)^{-1}\right), \tag{10}$$

then there exists an A-optimal design that is invariant under the transformation T.

The proof of Theorem 1 is in the Appendix. The condition in (10) requires that one know the weights of an A-optimal design that may be hard to derive analytically. The next theorem gives two sufficient conditions to check for the condition.

Theorem 2. *The condition at (10) holds if one of the following conditions holds:*

(i) there exists a $q \times q$ constant matrix $\mathbf{Q}$ with $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_q$ (identity matrix) such that $\mathbf{f}(T\mathbf{x}; \boldsymbol{\theta}_0) = \mathbf{Q}\ \mathbf{f}(\mathbf{x}; \boldsymbol{\theta}_0)$ for all $\mathbf{x} \in S_N$;

(ii) there exists a $q \times q$ matrix $\mathbf{U}$ satisfying $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_q$ such that $\boldsymbol{g}_1(rev(\mathbf{w})) = \mathbf{U}\ \boldsymbol{g}_1(\mathbf{w})$ and $\boldsymbol{G}_2(rev(\mathbf{w})) = \mathbf{U}\ \boldsymbol{G}_2(\mathbf{w})\ \mathbf{U}^\top$ for any $\mathbf{w}$.

The proof of Theorem 2 is in the Appendix. The conditions in Theorem 2 are easy to verify, especially condition (i). The results in Theorems 1 and 2 can be applied to both linear and nonlinear models.

Example 1. For the second-order regression model with independent variables x_1 and x_2, $y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2 + \epsilon$, we study the symmetry of A-optimal designs for the design spaces

$$S_{9,1} = \{(1,0), (-1,0), (0,1), (0,-1), (1,1), (-1,1), (1,-1), (-1,-1), (0,0)\},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

$$S_{9,2} = \{(\sqrt{2},0),(-\sqrt{2},0),(0,\sqrt{2}),(0,-\sqrt{2}),(1,1),(-1,1),(1,-1),(-1,-1),(0,0)\}.$$

Except for the center point $(0,0)$, the points in $S_{9,1}$ are located on the edges of a square while the points in $S_{9,2}$ are on a circle with radius $\sqrt{2}$. These spaces are invariant under several transformations, including

$$T_1\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_1 \\ x_2 \end{pmatrix}, \quad T_2\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix},$$

$$T_3\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -x_1 \\ -x_2 \end{pmatrix}, \quad T_4\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}.$$

It is easy to show that there exists an A-optimal design that is invariant under T_1, T_2, or T_3. For T_4, we have $\mathbf{f}(T_4\mathbf{x};\boldsymbol{\theta}) = \mathbf{Q}\ \mathbf{f}(\mathbf{x};\boldsymbol{\theta})$ with

$$\mathbf{Q} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus 1,$$

where $\oplus$ is the matrix direct sum. It is clear that $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_5$. Thus, there exists an A-optimal design that is invariant under T_4. If we apply the four transformations sequentially and use the results in Theorems 1 and 2, there exists an A-optimal design that is invariant under all the four transformations. This implies that there exists an A-optimal design $\xi^A(\mathbf{x})$ on $S_{9,1}$ (or $S_{9,2}$) having $w_1^A = w_2^A = w_3^A = w_4^A$ and $w_5^A = w_6^A = w_7^A = w_8^A$. □

Example 2. Consider a nonlinear model, $y_i = \theta_1 x/(x^2+\theta_2) + \epsilon,\ \theta_1 \neq 0,\ \theta_2 \neq 0$, on the design space $S_N \subset [-a,a]$, invariant under transformation $Tx = -x$. Here $\mathbf{f}(x;\boldsymbol{\theta}) = (x/(x^2+\theta_2), -\theta_1 x/((x^2+\theta_2)^2))^\top$, and it is easy to verify that $\mathbf{f}(Tx;\boldsymbol{\theta}_0) = \mathbf{Q}\ \mathbf{f}(x;\boldsymbol{\theta}_0)$ with $\mathbf{Q} = \mathrm{diag}(-1,-1)$ and $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_2$. Thus, there exists an A-optimal design that is symmetric about zero. □

The results in Theorems 1 and 2 can be extended easily to D-optimal designs $\xi^D(\mathbf{x})$ by changing tr() to det() in (10). By applying the result in Theorem 1, we can reduce the number of unknown weights in the loss functions $\phi_1(\mathbf{w})$ and $\phi_2(\mathbf{w})$ in (8). For instance, in Example 1, the number of unknown weights is reduced to 3.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

From Lemma 2, an alternative expression for $\phi_2(\mathbf{w})$ is $\phi_2(\mathbf{w}) = \det\left((\mathbf{B}(\mathbf{w}))^{-\mathbf{1}}\right)$, since $\det\left((\mathbf{A}(\mathbf{w}))^{-\mathbf{1}}\right) = \det\left((\mathbf{B}(\mathbf{w}))^{-\mathbf{1}}\right)$. For $\phi_1(\mathbf{w})$, we do not have $\mathrm{tr}\left((\mathbf{A}(\mathbf{w}))^{-\mathbf{1}}\right) = \mathrm{tr}\left((\mathbf{B}(\mathbf{w}))^{-\mathbf{1}}\right)$, but we can also characterize the A-optimality criterion using $\mathbf{B}(\mathbf{w})$.

Theorem 3. *If $\boldsymbol{G}_2(\mathbf{w})$ in (9) is nonsingular, then $\phi_1(\mathbf{w}) = tr\left((\mathbf{A}(\mathbf{w}))^{-\mathbf{1}}\right) = tr\left(\mathbf{C}(\mathbf{B}(\mathbf{w}))^{-\mathbf{1}}\right)$, where $\mathbf{C} = 0 \oplus \mathbf{I}_q$ is a $(q+1)\times(q+1)$ matrix.*

The proof of Theorem 3 is in the Supplementary Material. This characterization of the A-optimality criterion is useful for developing an efficient algorithm for computing A-optimal designs. If we are interested in a subset of the model parameters, the criterion can be easily modified. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, where $\boldsymbol{\theta}_1 \in R^{q_1}$ and $\boldsymbol{\theta}_2 \in R^{q_2}$ with $q_1 + q_2 = q$. The A-optimal design based on the SLSE of $\boldsymbol{\theta}_2$ minimizes $\phi_3(\mathbf{w}) = \mathrm{tr}\left(\mathbf{C}_1(\mathbf{B}(\mathbf{w}))^{-\mathbf{1}}\right)$, where $\mathbf{C}_1 = \mathbf{0}_{q_1+1} \oplus \mathbf{I}_{q_2}$ and $\mathbf{0}_{q_1+1}$ is a $(q_1+1)\times(q_1+1)$ matrix of zeros.

3 Numerical algorithms

For some regression models, $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$ can be constructed analytically. Examples are given in Gao and Zhou (2014) and Bose and Mukerjee (2015). In general, it is hard to find the optimal designs analytically, so numerical algorithms are developed. After reviewing the algorithms in Bose and Mukerjee (2015), we propose efficient algorithms for computing $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$. These algorithms do not use the derivatives of the loss functions. Yang *et al.* (2013) proposed another efficient algorithm for computing optimal designs using the derivatives.

3.1 Multiplicative algorithms

Bose and Mukerjee (2015) proposed multiplicative algorithms to compute $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$. For simplicity, we write $\mathbf{f}(\mathbf{u})$ for $\mathbf{f}(\mathbf{u};\boldsymbol{\theta}_0)$. Define, for $i = 1, \cdots, N$,

$$\psi_{Ai}(\mathbf{w}) = (1-t)\mathbf{f}^\top(\mathbf{u}_i)\mathbf{A}^{-2}\mathbf{f}(\mathbf{u}_i) + t\left(\mathbf{f}(\mathbf{u}_i) - \boldsymbol{g}_1(\mathbf{w})\right)^\top \mathbf{A}^{-2}\left(\mathbf{f}(\mathbf{u}_i) - \boldsymbol{g}_1(\mathbf{w})\right),$$
$$\psi_{Di}(\mathbf{w}) = (1-t)\mathbf{f}^\top(\mathbf{u}_i)\mathbf{A}^{-1}\mathbf{f}(\mathbf{u}_i) + t\left(\mathbf{f}(\mathbf{u}_i) - \boldsymbol{g}_1(\mathbf{w})\right)^\top \mathbf{A}^{-1}\left(\mathbf{f}(\mathbf{u}_i) - \boldsymbol{g}_1(\mathbf{w})\right), \quad (11)$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

where $\mathbf{A}^{-1} = (\mathbf{A}(\mathbf{w}))^{-1}$ and $\mathbf{A}^{-2} = (\mathbf{A}(\mathbf{w}))^{-1}(\mathbf{A}(\mathbf{w}))^{-1}$. Start with the uniform weight vector, $\mathbf{w}^{(0)} = (1/N, \cdots, 1/N)^\top$. For $\xi^A(\mathbf{x})$, the multiplicative algorithm finds $\mathbf{w}^{(j)}$, $j = 1, 2, \cdots$, iteratively as $w_i^{(j)} = w_i^{(j-1)}\psi_{Ai}(\mathbf{w}^{(j-1)})/\mathrm{tr}\left(\mathbf{A}(\mathbf{w}^{(j-1)})\right)^{-1}$, for $i = 1, \cdots, N$, till $\mathbf{w}^{(j)}$ satisfies

$$\psi_{Ai}(\mathbf{w}^{(j)}) - \mathrm{tr}\left(\mathbf{A}(\mathbf{w}^{(j)})\right)^{-1} \leq \delta, \quad \text{for} \quad i = 1, \cdots, N, \tag{12}$$

for some prespecified small δ (> 0). Similarly, for $\xi^D(\mathbf{x})$, the algorithm finds $\mathbf{w}^{(j)}$ iteratively as $w_i^{(j)} = w_i^{(j-1)}\psi_{Di}(\mathbf{w}^{(j-1)})/q$, for $i = 1, \cdots, N$, till $\mathbf{w}^{(j)}$ satisfies

$$\psi_{Di}(\mathbf{w}^{(j)}) - q \leq \delta, \quad \text{for} \quad i = 1, \cdots, N. \tag{13}$$

Conditions in (12) and (13) are approximated from necessary and sufficient conditions for the $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$ in Bose and Mukerjee (2015).

Lemma 3. *The weight vector* $\mathbf{w}$ *is*

(a) A-optimal if and only if $\mathbf{A}(\mathbf{w})$ *is nonsingular and* $\psi_{Ai}(\mathbf{w}) \leq tr(\mathbf{A}(\mathbf{w}))^{-1}$*, for* $i = 1, \cdots, N$*,*

(b) D-optimal if and only if $\mathbf{A}(\mathbf{w})$ *is nonsingular and* $\psi_{Di}(\mathbf{w}) \leq q$*, for* $i = 1, \cdots, N$*.*

These algorithms can preserve the transformation invariance property for the weights at each iteration, if there exist transformation invariant $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$.

Theorem 4. *Suppose the design space* S_N *is invariant under a transformation* T*. If there exists a* $q \times q$ *constant matrix* $\mathbf{Q}$ *with* $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_q$ *such that* $\mathbf{f}(T\mathbf{x}; \boldsymbol{\theta}_0) = \mathbf{Q}\,\mathbf{f}(\mathbf{x}; \boldsymbol{\theta}_0)$ *for all* $\mathbf{x} \in S_N$*, then the weights from the multiplicative algorithms satisfy* $rev(\mathbf{w}^{(j)}) = \mathbf{w}^{(j)}$ *for all* $j = 0, 1, 2, \cdots$ *.*

The proof of Theorem 4 is in the Supplementary Material. This result depends on the fact that the initial weight vector satisfies $rev(\mathbf{w}^{(0)}) = \mathbf{w}^{(0)}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

3.2 Convex optimization algorithms

The CVX program in MATLAB (Grant and Boyd (2013)) is powerful and widely used to solve convex optimization problems. Gao and Zhou (2015) applied the CVX program to find the D-optimal designs based on the SLSE through the moments of distribution $\xi(\mathbf{x})$. The optimal design problems for $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$ on a discrete design space can be formulated as convex optimization problems, differently from those in Gao and Zhou (2015). Using $w_N = 1 - \sum_{i=1}^{N-1} w_i$, we define a weight vector having $N-1$ weights as $\tilde{\mathbf{w}} = \left(w_1, w_2, \cdots, w_{N-1}, 1 - \sum_{i=1}^{N-1} w_i\right)^\top$. Let $\mathbf{D}(\tilde{\mathbf{w}}) = \text{diag}\left(w_1, w_2, \cdots, w_{N-1}, 1 - \sum_{i=1}^{N-1} w_i\right)$ be a diagonal matrix. One has $\phi_1(\mathbf{w}) = \phi_1(\tilde{\mathbf{w}})$ and $\phi_2(\mathbf{w}) = \phi_2(\tilde{\mathbf{w}})$, and $\phi_1(\tilde{\mathbf{w}})$ and $\log(\phi_2(\tilde{\mathbf{w}}))$ are convex functions of $\tilde{\mathbf{w}}$. The conditions in (6) are equivalent to that $\mathbf{D}(\tilde{\mathbf{w}}) \succeq 0$. Thus, the A- and D-optimal design problems become, respectively,

$$\begin{cases} \min_{\tilde{\mathbf{w}}} \phi_1(\tilde{\mathbf{w}}), \\ \text{subject to: } \mathbf{D}(\tilde{\mathbf{w}}) \succeq 0, \end{cases} \tag{14}$$

$$\begin{cases} \min_{\tilde{\mathbf{w}}} \log(\phi_2(\tilde{\mathbf{w}})), \\ \text{subject to: } \mathbf{D}(\tilde{\mathbf{w}}) \succeq 0. \end{cases} \tag{15}$$

The CVX program in MATLAB has some technical issues. In (15), we need to use $\mathbf{B}(\tilde{\mathbf{w}})$ in $\phi_2(\tilde{\mathbf{w}})$, and the CVX program works well to solve $\min_{\tilde{\mathbf{w}}} -\log(\det(\mathbf{B}(\tilde{\mathbf{w}}))$ or $\min_{\tilde{\mathbf{w}}} -(\det(\mathbf{B}(\tilde{\mathbf{w}}))^{1/(q+1)}$. In (14), however, it does not work to use $\mathbf{A}(\tilde{\mathbf{w}})$ in $\phi_1(\tilde{\mathbf{w}})$, and it is not straightforward to use $\mathbf{B}(\tilde{\mathbf{w}})$. We develop a novel formulation of the A-optimal design problem with a linear objective function and linear matrix inequality constraints that is an SDP problem.

Let $\mathbf{e}_i$ be the ith unit vector in R^{q+1}, $i = 1, \cdots, q+1$, $\mathbf{v} = (v_2, \cdots, v_{q+1})^\top$, and

$$\begin{aligned} \mathbf{B}_i &= \begin{pmatrix} \mathbf{B}(\tilde{\mathbf{w}}) & \mathbf{e}_i \\ \mathbf{e}_i^\top & v_i \end{pmatrix}, \quad \text{for} \quad i = 2, \cdots, q+1, \\ \mathbf{H}(\tilde{\mathbf{w}}, \mathbf{v}) &= \mathbf{B}_2 \oplus \cdots \oplus \mathbf{B}_{q+1} \oplus \mathbf{D}(\tilde{\mathbf{w}}). \end{aligned} \tag{16}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

Since $\mathbf{B}(\tilde{\mathbf{w}})$ and $\mathbf{D}(\tilde{\mathbf{w}})$ are linear matrices in $\tilde{\mathbf{w}}$, $\mathbf{H}(\tilde{\mathbf{w}}, \mathbf{v})$ is a linear matrix in $\tilde{\mathbf{w}}$ and $\mathbf{v}$. Then $\xi^A(\mathbf{x})$ can be solved through

$$\begin{cases} \min_{\tilde{\mathbf{w}},\mathbf{v}} \sum_{i=2}^{q+1} v_i, \\ \text{subject to: } \mathbf{H}(\tilde{\mathbf{w}}, \mathbf{v}) \succeq 0, \end{cases} \tag{17}$$

Theorem 5. *The solutions to the optimization problems (14) and (17) satisfy*

(i) if $\tilde{\mathbf{w}}^*$ is a solution to (14), then $(\tilde{\mathbf{w}}^*, \mathbf{v}^*)$ is a solution to (17) with

$$\mathbf{v}^* = (\mathbf{e}_2^\top (\mathbf{B}(\tilde{\mathbf{w}}^*))^{-1} \mathbf{e}_2, \cdots, \mathbf{e}_{q+1}^\top (\mathbf{B}(\tilde{\mathbf{w}}^*))^{-1} \mathbf{e}_{q+1})^\top,$$

(ii) if $(\tilde{\mathbf{w}}^*, \mathbf{v}^*)$ is a solution to (17), then $\tilde{\mathbf{w}}^*$ is a solution to (14).

The proof of Theorem 5 is in the Appendix. To solve (17), the SeDuMi program in MATLAB can be used. See Sturm (1999) for a user's guide. There is a MATLAB program in Ye *et al.* (2015) that applies the SeDuMi for solving a different SDP problem.

4 Applications and efficiencies

We compute $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$ for various linear and nonlinear models and give representative results. The A-optimal designs are computed by the multiplicative algorithm and the SeDuMi program, while the D-optimal designs are computed by the multiplicative algorithm and the CVX program. Conditions in (12) and (13) are used to verify that the numerical solutions are A- and D-optimal designs, respectively. Numerical algorithms are compared, and efficiencies of the SLSE and its optimal designs are discussed. A property of locally optimal designs is also derived for nonlinear models.

4.1 Examples

Example 3. Consider the regression model and design spaces in Example 1 and compute $\xi^A(\mathbf{x})$ and $\xi^D(\mathbf{x})$ for various values of t. Since the number of points in

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

Table 1: A- and D-optimal weights, w_1^A, w_5^A, w_9^A, w_1^D, w_5^D, w_9^D, in Example 3

t	w_1^A	w_5^A	w_9^A	w_1^D	w_5^D	w_9^D
Design space $S_{9,1}$						
0	0.131	0.119	0.000	0.071	0.179	0.000
0.3	0.130	0.120	0.000	0.072	0.178	0.000
0.5	0.128	0.122	0.000	0.074	0.176	0.000
0.9	0.118	0.121	0.044	0.088	0.162	0.000
Design space $S_{9,2}$						
0	0.104	0.146	0.000	0.125	0.125	0.000
0.3	0.104	0.146	0.000	0.125	0.125	0.000
0.5	0.104	0.146	0.000	0.125	0.125	0.000
0.9	0.088	0.125	0.148	0.116	0.116	0.072

design spaces $S_{9,1}$ and $S_{9,2}$ is small, all the algorithms work well and quickly. We computed the weights, $w_1, \cdots, w_9$. The results from the multiplicative algorithms are the same as those from the CVX and SeDuMi programs, and the weights have the transformation invariance property discussed in Example 1. Representative results are given in Table 1, where only three weights, w_1, w_5, w_9, are listed due to the invariance property. The results indicate that the optimal designs depend on the value of t. For small t the center point has weight zero for all the optimal designs, but for $t = 0.9$ the center point has a positive weight for three optimal designs. $\square$

For linear models, the optimal designs do not depend on $\boldsymbol{\theta}$. If there is an intercept term in the model, the optimal designs ξ^A and ξ^D are the same as those based on the OLSE (Gao and Zhou (2014)). For nonlinear models, the optimal designs usually depend on the true value, $\boldsymbol{\theta}_0$, and are called locally optimal designs. In practice an estimate of $\boldsymbol{\theta}_0$ is used to construct the optimal designs. However, if a nonlinear model is linear in a subset of parameters, then optimal design ξ^D does not depend

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

on the true value of the subset.

Theorem 6. *Let* $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$*, where* $\boldsymbol{\alpha} \in R^a$ *and* $\boldsymbol{\beta} \in R^b$ *with* $a + b = q$*. For a nonlinear model*

$$g(\mathbf{x}; \boldsymbol{\theta}) = \sum_{i=1}^{a} \alpha_i h_i(\mathbf{x}; \boldsymbol{\beta}_i), \tag{18}$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_a)^\top$*,* $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \cdots, \boldsymbol{\beta}_a^\top)^\top$ *with* $\boldsymbol{\beta}_i \in R^{q_i}$*, and* $h_i(\mathbf{x}; \boldsymbol{\beta}_i)$ *can be linear or nonlinear in* $\boldsymbol{\beta}_i$*, then* $\xi^D(\mathbf{x})$ *does not depend on* $\boldsymbol{\alpha}$*.*

The proof of Theorem 6 is in the Supplementary Material. A similar result for D-optimal designs based on the OLSE is in Dette *et al.* (2006). For $\xi^A(\mathbf{x})$, this result is not true in general. More discussion about other approaches for locally optimal designs can be found in Yang and Stufken (2012).

Example 4. The Michaelis-Menten model (Michaelis and Menten (1913)), one of the best-known model in biochemistry, is used to study enzyme kinetics. Enzyme-kinetics studies the chemical reactions for substrate that are catalyzed by enzymes. The relationship between the reaction rate and the concentration of the substrate can be described as $y = \alpha x/(\beta + x) + \epsilon,\ x \geq 0$, where y represents the speed of reaction, and x is the substrate concentration. Optimal designs for this model have been studied by many authors, including Dette *et al.* (2005) and Yang and Stufken (2009). Table 2 lists representative results of ξ^A and ξ^D for various values of t and N for the model with $\alpha = 1$, $\beta = 1$, and $S_N = \{4(i-1)/(N-1) : \ i = 1, \cdots, N\} \subset [0, 4]$. By Theorem 6 the D-optimal designs do not depend on the value of α, but the A-optimal designs depend on α from the numerical results. For large N, the CVX and SeDuMi programs were much faster than the multiplicative algorithms. In fact, the A-optimal designs in Table 2 were all calculated by the SeDuMi program. The A-optimal and D-optimal designs depend on t. For small t there are two support points, while for large t there are three support points. When $t = 0$, the results give the optimal designs based on the OLSE and they match the results shown in the website (http://optimal-design.biostat.ucla.edu/optimal/OptimalDesign.aspx).

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

Table 2: A- and D-optimal design points and their weights (in parentheses) for the Michaelis-Menten model with $\alpha = 1$ and $\beta = 1$

$N = 101$				
$t = 0,\ 0.3$	ξ^A:	0.520 (0.666)	4.000 (0.334)	
	ξ^D:	0.680 (0.500)	4.000 (0.500)	
$t = 0.7$	ξ^A:	0.640 (0.641)	4.000 (0.359)	
	ξ^D:	0.000 (0.048)	0.680 (0.476)	4.000 (0.476)
$t = 0.9$	ξ^A:	0.000 (0.154)	0.680 (0.536)	4.000 (0.310)
	ξ^D:	0.000 (0.260)	0.680 (0.370)	4.000 (0.370)
$N = 201$				
$t = 0$	ξ^A:	0.500 (0.671)	4.000 (0.329)	
	ξ^D:	0.660 (0.500)	4.000 (0.500)	
$t = 0.3$	ξ^A:	0.540 (0.661)	4.000 (0.339)	
	ξ^D:	0.660 (0.500)	4.000 (0.500)	
$t = 0.7$	ξ^A:	0.640 (0.641)	4.000 (0.359)	
	ξ^D:	0.000 (0.048)	0.660 (0.476)	4.000 (0.476)
$t = 0.9$	ξ^A:	0.000 (0.159)	0.660 (0.536)	4.000 (0.305)
	ξ^D:	0.000 (0.260)	0.660 (0.370)	4.000 (0.370)
$N = 501$				
$t = 0$	ξ^A:	0.504 (0.670)	4.000 (0.330)	
	ξ^D:	0.664 (0.500)	4.000 (0.500)	
$t = 0.3$	ξ^A:	0.536 (0.662)	4.000 (0.338)	
	ξ^D:	0.664 (0.500)	4.000 (0.500)	
$t = 0.7$	ξ^A:	0.632 (0.642)	4.000 (0.358)	
	ξ^D:	0.000 (0.048)	0.664 (0.476)	4.000 (0.476)
$t = 0.9$	ξ^A:	0.000 (0.158)	0.664 (0.536)	4.000 (0.306)
	ξ^D:	0.000 (0.260)	0.664 (0.370)	4.000 (0.370)

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

The number of support points for $t = 0$ also agrees with the result in Yang and Stufken (2009). □

4.2 A-efficiency and D-efficiency

To compare optimal designs based on the SLSE and the OLSE, we define A-efficiency and D-efficiency measures as follows. Let

$$\begin{aligned} a_1(\xi) &= \operatorname{tr}\left(Cov(\hat{\boldsymbol{\theta}}_{SLS})\right) = \operatorname{tr}\left((1-t)\,\sigma_0^2\left(\boldsymbol{G}_2 - t\boldsymbol{g}_1\boldsymbol{g}_1^\top\right)^{-1}\right), \\ d_1(\xi) &= \det\left(Cov(\hat{\boldsymbol{\theta}}_{SLS})\right) = \det\left((1-t)\,\sigma_0^2\left(\boldsymbol{G}_2 - t\boldsymbol{g}_1\boldsymbol{g}_1^\top\right)^{-1}\right), \end{aligned}$$

and take the A-efficiency and D-efficiency measures as,

$$\text{Eff}_A = \frac{a_1(\xi_{SLS}^A)}{a_1(\xi_{OLS}^A)}, \qquad \text{Eff}_D = \left(\frac{d_1(\xi_{SLS}^D)}{d_1(\xi_{OLS}^D)}\right)^{1/q},$$

where ξ_{SLS}^A and ξ_{SLS}^D are the A- and D-optimal designs based on the SLSE, and ξ_{OLS}^A and ξ_{OLS}^D are based on the OLSE. Since the SLSE is more efficient than the OLSE, all the measures are evaluated using the covariance of the SLSE. If $\text{Eff}_A < 1$ ($\text{Eff}_D < 1$), then ξ_{SLS}^A (ξ_{SLS}^D) is more A-efficient (D-efficient) than ξ_{OLS}^A (ξ_{OLS}^D).

We computed the efficiency measures for the examples in Section 4.1, and representative results are given in Table 3. For small t the optimal designs based on the SLSE and the OLSE are similar, which is consistent with the results in Tables 1 and 2. For large t, the optimal designs based on the SLSE can be much more efficient than those based on the OLSE for some models.

5 Discussion

If the design space S is not discrete, such as a closed interval, we can discretize it and then use the methods in this paper to construct optimal designs. Gao and Zhou (2014, 2015) obtained some results for the D-optimal designs on closed interval design spaces. Several results here for A-optimal designs can also be extended to

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

Table 3: A-efficiency and D-efficiency for the optimal designs in Examples 3 and 4.

	Example 3 with design space $S_{9,2}$		Example 4 with $N = 501$	
t	Eff_A	Eff_D	Eff_A	Eff_D
0.0	1.000	1.000	1.000	1.000
0.3	1.000	1.000	0.997	0.999
0.7	1.000	1.000	0.963	0.996
0.9	0.836	0.975	0.704	0.739

closed interval design spaces. For any $\xi(\mathbf{x})$ on a closed interval, we can define matrix $\mathbf{B}(\xi)$ similarly to the one in (9), but using $\boldsymbol{g}_1$ and $\boldsymbol{G}_2$ in (4). The expression for the A-optimality criterion in Theorem 3 can be easily modified using $\mathbf{B}(\xi)$. The definition of transformation invariance can be changed slightly by including all discrete and continuous distributions, and the invariance property can be studied for the A-optimal designs. Furthermore, the number of support points in optimal designs can be investigated analytically, a future research topic.

6 Supplementary Materials

Supplementary Materials are posted on the journal's website, where two more examples of A- and D-optimal designs are presented. Example 5 is for compartmental models, and Example 6 is for an Emax model. The results show that SeDuMi and CVX can work faster than the multiplicative algorithm for large N. The proofs of Theorems 3 , 4, and 6 are there.

Acknowledgements

The authors would like to thank an associate editor and two referees for their helpful comments to improve the presentation of this article. This research work is

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

supported by Discovery Grants from the Natural Science and Engineering Research Council of Canada.

Appendix: Proofs

Let $I_1 = \{1, \cdots, m, (N-m+1), \cdots, N\}$ and $I_2 = \{(m+1), \cdots, (N-m)\}$. If $m = N/2$, I_2 is an empty set.

Proof of Theorem 1: Using $\xi^A(\mathbf{x})$, we define a distribution $\xi^\lambda(\mathbf{x})$ having weight vector $\mathbf{w}(\lambda)$ with elements $w_i(\lambda) = (1-\lambda)w_i^A + \lambda w_{N+1-i}^A$ for $i \in I_1$ and $w_i(\lambda) = w_i^A$ for $i \in I_2$, for $\lambda \in [0,1]$. Since S_N is T-invariant, it is obvious that distribution $\xi^{0.5}(\mathbf{x})$ is T-invariant.

We show that $\phi_1(\mathbf{w}(\lambda)) \leq \phi_1(\mathbf{w}^A)$, where ϕ_1 is defined in (8). For fixed weight $\mathbf{w}^A$, the elements of $\mathbf{w}(\lambda)$ are linear functions of λ. From Lemma 1, $\phi_1(\mathbf{w})$ is a convex function of $\mathbf{w}$, so $\phi_1(\mathbf{w}(\lambda))$ is a convex function of λ. Notice that $\mathbf{w}(0) = \mathbf{w}^A$ and $\mathbf{w}(1) = \text{rev}(\mathbf{w}^A)$. By (8) and (10), we have $\phi_1(\mathbf{w}(0)) = \phi_1(\mathbf{w}(1))$. Using the convex property, we get

$$\phi_1(\mathbf{w}(\lambda)) \leq (1-\lambda)\phi_1(\mathbf{w}(0)) + \lambda\phi_1(\mathbf{w}(1)) = \phi_1(\mathbf{w}(0)) = \phi_1(\mathbf{w}^A).$$

Since $\xi^A(\mathbf{x})$ minimizes $\phi_1(\mathbf{w})$, we must have $\phi_1(\mathbf{w}(\lambda)) = \phi_1(\mathbf{w}^A)$, for all $\lambda \in [0,1]$. This implies that $\xi^\lambda(\mathbf{x})$ is also an A-optimal design. Thus, there exists an A-optimal design $\xi^{0.5}(\mathbf{x})$ that is T-invariant. $\square$

Proof of Theorem 2: (i) For T-invariant S_N, we have $T\mathbf{u}_i = \mathbf{u}_{N+1-i}$ for $i \in I_1$ and $T\mathbf{u}_i = \mathbf{u}_i$ for $i \in I_2$. If there exists a $q \times q$ constant matrix $\mathbf{Q}$ with $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_q$ such that $\mathbf{f}(T\mathbf{x}; \boldsymbol{\theta}_0) = \mathbf{Q}\ \mathbf{f}(\mathbf{x}; \boldsymbol{\theta}_0)$ for all $\mathbf{x} \in S_N$, we have, from (7),

$$\begin{aligned}
\boldsymbol{g}_1(\text{rev}(\mathbf{w})) &= \sum_{i \in I_1} w_{N+1-i}\mathbf{f}(\mathbf{u}_i; \boldsymbol{\theta}_0) + \sum_{i \in I_2} w_i\mathbf{f}(\mathbf{u}_i; \boldsymbol{\theta}_0) \\
&= \sum_{i \in I_1} w_i\mathbf{f}(\mathbf{u}_{N+1-i}; \boldsymbol{\theta}_0) + \sum_{i \in I_2} w_i\mathbf{f}(\mathbf{u}_i; \boldsymbol{\theta}_0) \\
&= \sum_{i=1}^{N} w_i\mathbf{f}(T\mathbf{u}_i; \boldsymbol{\theta}_0) = \sum_{i=1}^{N} w_i\mathbf{Q}\ \mathbf{f}(\mathbf{u}_i; \boldsymbol{\theta}_0) = \mathbf{Q}\ \boldsymbol{g}_1(\mathbf{w}),
\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

and similarly, $\boldsymbol{G}_2(\text{rev}(\mathbf{w})) = \mathbf{Q}\ \boldsymbol{G}_2(\mathbf{w})\ \mathbf{Q}^\top$. Since $\mathbf{A}(\mathbf{w}) = \boldsymbol{G}_2(\mathbf{w}) - t\boldsymbol{g}_1(\mathbf{w})\boldsymbol{g}_1^\top(\mathbf{w})$, it is clear that $\mathbf{A}(\text{rev}(\mathbf{w})) = \mathbf{Q}\ \mathbf{A}(\mathbf{w})\ \mathbf{Q}^\top$. Thus,

$$\begin{aligned}
\text{tr}\left((\mathbf{A}(\text{rev}(\mathbf{w})))^{-1}\right) &= \text{tr}\left(\left(\mathbf{Q}\ \mathbf{A}(\mathbf{w})\ \mathbf{Q}^\top\right)^{-1}\right) = \text{tr}\left(\left(\mathbf{Q}^\top\right)^{-1}(\mathbf{A}(\mathbf{w}))^{-1}\mathbf{Q}^{-1}\right) \\
&= \text{tr}\left((\mathbf{A}(\mathbf{w}))^{-1}\right), \quad \text{from} \quad \mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_q,
\end{aligned}$$

which implies that the condition in (10) holds.

(ii) The proof is similar to that in part (i) and is omitted. □

Proof of Theorem 5: (i) If $\tilde{\mathbf{w}}^*$ is a solution to (14), then $\mathbf{A}(\tilde{\mathbf{w}}^*) \succ 0$ (positive definite) and $\mathbf{B}(\tilde{\mathbf{w}}^*) \succ 0$ by (8) and Theorem 3. Let

$$\mathbf{v}^* = (v_2^*, \cdots, v_{q+1}^*)^\top = (\mathbf{e}_2^\top(\mathbf{B}(\tilde{\mathbf{w}}^*))^{-1}\mathbf{e}_2, \cdots, \mathbf{e}_{q+1}^\top(\mathbf{B}(\tilde{\mathbf{w}}^*))^{-1}\mathbf{e}_{q+1})^\top.$$

Then, from (16) $\mathbf{B}_i \succeq 0$, for $i = 2, \cdots, q+1$ and the constraint in (17) is satisfied. For any $\tilde{\mathbf{w}}$ satisfying $\mathbf{D}(\tilde{\mathbf{w}}) \succeq 0$ and $\mathbf{B}(\tilde{\mathbf{w}}) \succ 0$, we get

$$\begin{aligned}
\sum_{i=2}^{q+1} v_i^* &= \sum_{i=2}^{q+1} \mathbf{e}_i^\top(\mathbf{B}(\tilde{\mathbf{w}}^*))^{-1}\mathbf{e}_i \\
&= \text{tr}\left(\mathbf{C}\,(\mathbf{B}(\tilde{\mathbf{w}}^*))^{-1}\right) \\
&= \phi_1(\tilde{\mathbf{w}}^*), \quad \text{from Theorem 3,} \\
&\le \phi_1(\tilde{\mathbf{w}}), \quad \text{from } \tilde{\mathbf{w}}^* \text{ being a solution to problem (14),} \\
&= \sum_{i=2}^{q+1} \mathbf{e}_i^\top(\mathbf{B}(\tilde{\mathbf{w}}))^{-1}\mathbf{e}_i \\
&\le \sum_{i=2}^{q+1} v_i, \quad \text{from } \mathbf{B}_i \succeq 0,
\end{aligned}$$

which implies that $(\tilde{\mathbf{w}}^*, \mathbf{v}^*)$ is a solution to (17).

(ii) Suppose that $(\tilde{\mathbf{w}}^*, \mathbf{v}^*)$ is a solution to (17). Since $\mathbf{B}_i \succeq 0$, we must have

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

$\mathbf{B}(\tilde{\mathbf{w}}^*) \succ 0$. For any $\tilde{\mathbf{w}}$ satisfying $\mathbf{D}(\tilde{\mathbf{w}}) \succeq 0$ and $\mathbf{B}(\tilde{\mathbf{w}}) \succ 0$, we have

$$\begin{aligned}
\phi_1(\tilde{\mathbf{w}}^*) &= \operatorname{tr}\left(\mathbf{C}\left(\mathbf{B}(\tilde{\mathbf{w}}^*)\right)^{-1}\right), \text{from Theorem 3,} \\
&= \sum_{i=2}^{q+1} \mathbf{e}_i^\top \left(\mathbf{B}(\tilde{\mathbf{w}}^*)\right)^{-1} \mathbf{e}_i \\
&\leq \sum_{i=2}^{q+1} v_i^*, \quad \text{from } \mathbf{B}_i \succeq 0, \\
&\leq \sum_{i=2}^{q+1} v_i, \quad \text{from } \tilde{\mathbf{v}}^* \text{ being a solution to problem (17),} \\
&= \sum_{i=2}^{q+1} \mathbf{e}_i^\top \left(\mathbf{B}(\tilde{\mathbf{w}})\right)^{-1} \mathbf{e}_i, \quad \text{by choosing } \ v_i = \mathbf{e}_i^\top \left(\mathbf{B}(\tilde{\mathbf{w}})\right)^{-1} \mathbf{e}_i, \\
&= \phi_1(\tilde{\mathbf{w}}).
\end{aligned}$$

Thus, $\tilde{\mathbf{w}}^*$ is a solution to (14). □

References

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

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.

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

Dette, H., Melas, V.B., and Wong, W.K. (2005). Optimal design for goodness of fit of the Michaelis-Menten enzyme kinetic function. *Journal of the American Statistical Association* **100**, 1370-1381.

Dette, H., Melas, V.B., and Wong, W.K. (2006). Locally D-optimal designs for exponential regression models. *Statistica Sinica* **16**, 789-803.

Gao, L.L. and Zhou, J. (2014). New optimal design criteria for regression models with asymmetric errors. *Journal of Statistical Planning and Inference* **149**, 140-151.

Gao, L.L. and Zhou, J. (2015). D-optimal designs based on the second-order least squares estimator. *Statistical Papers*, to appear.

Statistica Sinica: Preprint
doi:10.5705/ss.202015.0285

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.)

Michaelis, L. and Menten, M.L. (1913). Die Kinetik der invertinwirkung. *Biochemische Zeitschrift* **49**, 333-369.

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

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

Wang, L. and Leblanc, A. (2008). Second-order nonlinear least squares estimation. *Annals of the Institute of Statistical Mathematics* **60**, 883-900.

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.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *The Annals of Statistics* **37**, 518-541.

Yang, M. and Stufken, J. (2012). Identifying locally optimal designs for nonlinear models: a simple extension with profound consequences. *The Annals of Statistics* **40**, 1665-1681.

Ye, J.J., Zhou, J., and Zhou, W. (2015). Computing A-optimal and E-optimal designs for regression models via semidefinite programming. *Communications in Statistics - Simulation and Computation*, to appear.

Zhang, R. and Mukerjee, R. (2013). Highly efficient factorial designs for cDNA microarray experiments: use of approximate theory together with a step-up step-down procedure. *Statistical Applications in Genetics and Molecular Biology* **12**, 489-503.