Properties of optimal regression designs under the second-order least squares estimator

Chi-Kuang Yeh, Julie Zhou

January 2019

Properties of optimal regression designs under the second-order least squares estimator

Chi-Kuang Yeh and Julie Zhou[1]

Department of Mathematics and Statistics

University of Victoria

Victoria, BC, Canada V8W 2Y2

ABSTRACT

We investigate properties of optimal designs under the second-order least squares estimator (SLSE) for linear and nonlinear regression models. First we derive equivalence theorems for optimal designs under the SLSE. We then obtain the number of support points in A-, c- and D-optimal designs analytically for several models. Using a generalized scale invariance concept we also study the scale invariance property of D-optimal designs. In addition, numerical algorithms are discussed for finding optimal designs. The results are quite general and can be applied for various linear and nonlinear models. Several applications are presented, including results for fractional polynomial, spline regression and trigonometric regression models.

Key words and phrases: A-optimal design, convex optimization, D-optimal design, fractional polynomial, generalized scale invariance, Peleg model, spline regression, number of support points.

MSC 2010: 62K05, 62K20.

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

1 Introduction

Consider a general regression model,

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

where $y_i \in R$ is the response at design point $\mathbf{x}_i \in S \subset R^p$, S is the design space, $\boldsymbol{\theta} \in R^q$ is an unknown regression 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. A class of second-order least squares estimators of $\boldsymbol{\theta}$ is proposed and studied in Wang and Leblanc (2008). Let $\boldsymbol{\theta_0}$ and σ_0 be the true parameter values of $\boldsymbol{\theta}$ and σ, respectively. Define $\mu_3 = E(\epsilon_1^3)$ and $\mu_4 = E(\epsilon_1^4)$. Assuming $\sigma_0^2(\mu_4 - \sigma_0^4) - \mu_3^2 \neq 0$, Wang and Leblanc (2008) derived the most efficient second-order least squares estimator (SLSE), denoted by $\hat{\boldsymbol{\theta}}_{SLS}$. If the error distribution is asymmetric, then the SLSE is asymptotically more efficient than the ordinary least squares estimator (LSE). From Wang and Leblanc (2008), the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{SLS}$ is given by

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

where $t = \frac{\mu_3^2}{\sigma_0^2(\mu_4 - \sigma_0^4)}$,

$$\boldsymbol{g}_1(\xi, \boldsymbol{\theta_0}) = E\left[\frac{\partial g(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \,|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right], \; \boldsymbol{G}_2(\xi, \boldsymbol{\theta_0}) = E\left[\frac{\partial g(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial g(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \,|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}\right]. \tag{3}$$

The expectation in (3) is taken with respect to the distribution $\xi(\mathbf{x})$ of $\mathbf{x}$. Parameter t is related to the skewness of the error distribution. It has been shown that $0 \leq t < 1$ for any error distribution, and $t = 0$ for a symmetric error distribution (Gao and Zhou, 2014).

Gao and Zhou (2014) proposed and investigated optimal regression design criteria based on the SLSE and obtained several theoretical results, which include the symmetry and transformation invariance properties of D-optimal designs. Bose and Mukerjee (2015) and Gao and Zhou (2017) made further developments for optimal designs under the SLSE including the convexity results of the design criteria and numerical algorithms. Bose and Mukerjee (2015) discussed applications for factorial designs with two levels. Gao and Zhou (2017) applied the CVX program in MATLAB (Grant and Boyd, 2013) for finding the D-optimal designs through the moments of the distribution of $\mathbf{x}$, and their methods can be applied for univariate polynomial and trigonometric regression models. Yin and Zhou (2017) studied and characterized A-optimal designs under the SLSE and developed a numerical algorithm via semidefinite programming (Sturm, 1999).

In this paper we will derive equivalence theorems for optimal designs under the SLSE, obtain the number of support points in A-, c- and D-optimal designs for various models analytically, and use a generalized scale invariance concept to study D-optimal designs under scale transformation for nonlinear models. We also discuss a numerical algorithm for finding optimal designs for any linear and nonlinear regression models. The algorithm is based on the semi-definite programming in convex optimization and can find optimal designs on discrete design spaces. It is powerful to solve optimal design problems with a very large number of points in the design space.

The rest of the paper is organized as follows. In Section 2 we discuss A-, c- and D-optimality criteria under the SLSE and derive equivalence results. In Section 3 we study the number of support points in A-, c- and D-optimal designs for various regression models. In Section 4 we discuss a numerical algorithm for finding optimal designs and present applications. Bayesian optimal designs are also discussed briefly. In Section 5 we use a generalized scale invariance concept to study scale invariance property of D-optimal designs. Concluding remarks are in Section 6. All proofs are given in the Appendix, and the MATLAB codes of Example 2 are given in the supplementary material.

2 Optimality criteria under the SLSE

Approximate designs are studied in this paper. Let Ξ denote the class of probability measures on S including all discrete measures. We first give the characterization of A-, c- and D-optimality under the SLSE, and then obtain equivalence results for verifying optimal designs.

2.1 Optimality criteria

Define matrix $\mathbf{A}(\xi, \boldsymbol{\theta_0}) = \boldsymbol{G}_2(\xi, \boldsymbol{\theta_0}) - t\, \boldsymbol{g}_1(\xi, \boldsymbol{\theta_0}) \boldsymbol{g}_1(\xi, \boldsymbol{\theta_0})^\top$. From (2), it is clear that $Cov(\hat{\boldsymbol{\theta}}_{SLS})$ is proportional to $\mathbf{A}^{-1}(\xi, \boldsymbol{\theta_0})$, the inverse of $\mathbf{A}(\xi, \boldsymbol{\theta_0})$. We minimize the following loss functions,

$$
\begin{aligned}
\phi_A(\xi, \boldsymbol{\theta_0}) &= \mathrm{tr}\left(\mathbf{A}^{-1}(\xi, \boldsymbol{\theta_0})\right), \\
\phi_D(\xi, \boldsymbol{\theta_0}) &= \det\left(\mathbf{A}^{-1}(\xi, \boldsymbol{\theta_0})\right), \\
\phi_c(\xi, \boldsymbol{\theta_0}) &= \mathbf{c}^\top \mathbf{A}^{-1}(\xi, \boldsymbol{\theta_0}) \mathbf{c},
\end{aligned}
$$

over $\xi \in \Xi$, to get A-, D-, and c-optimal designs under the SLSE, respectively, where $\mathbf{c} \in R^q$ is a given vector. When $\mathbf{A}(\xi, \boldsymbol{\theta_0})$ is singular, $\phi_A(\xi, \boldsymbol{\theta_0})$, $\phi_D(\xi, \boldsymbol{\theta_0})$ and $\phi_c(\xi, \boldsymbol{\theta_0})$ are defined to be ∞. Notice that c-optimal designs under the SLSE have not been studied before. Let ξ_A^*, ξ_D^* and ξ_c^* denote the optimal designs for A-, D-, and c-optimal designs, respectively.

Since all the optimal designs have nonsingular $\mathbf{A}(\xi, \boldsymbol{\theta_0})$, the discussion below is for nonsingular $\mathbf{A}(\xi, \boldsymbol{\theta_0})$. By the definition of $\mathbf{A}(\xi, \boldsymbol{\theta_0})$, when $\mathbf{A}(\xi, \boldsymbol{\theta_0})$ is nonsingular, $\boldsymbol{G}_2(\xi, \boldsymbol{\theta_0})$ is also nonsingular. For two given designs (distributions) ξ_1 and ξ_2 of $\mathbf{x}$, let $\xi_\alpha = (1-\alpha)\xi_1 + \alpha\xi_2$ be a convex combination of ξ_1 and ξ_2 with $0 \le \alpha \le 1$. Since $\mathbf{A}(\xi_\alpha, \boldsymbol{\theta_0})$ is not linear in α, it is hard to use it to develop numerical algorithms and obtain theoretical properties. Gao and Zhou (2017) introduced a useful matrix,

$$
\mathbf{B}(\xi, \boldsymbol{\theta_0}) = \begin{pmatrix} 1 & \sqrt{t}\, \boldsymbol{g}_1(\xi, \boldsymbol{\theta_0})^\top \\ \sqrt{t}\, \boldsymbol{g}_1(\xi, \boldsymbol{\theta_0}) & \boldsymbol{G}_2(\xi, \boldsymbol{\theta_0}) \end{pmatrix}, \tag{4}
$$

to characterize the loss functions for A- and D-optimality criteria. It is easy to show that $\det(\mathbf{A}(\xi, \boldsymbol{\theta_0})) = \det(\mathbf{B}(\xi, \boldsymbol{\theta_0}))$ and

$$\mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0}) = \begin{pmatrix} 1/r & -\frac{\sqrt{t}}{r}\, \boldsymbol{g}_1(\xi, \boldsymbol{\theta_0})^\top \boldsymbol{G}_2^{-1}(\xi, \boldsymbol{\theta_0}) \\ -\frac{\sqrt{t}}{r}\, \boldsymbol{G}_2^{-1}(\xi, \boldsymbol{\theta_0})\boldsymbol{g}_1(\xi, \boldsymbol{\theta_0}) & \mathbf{A}^{-1}(\xi, \boldsymbol{\theta_0}) \end{pmatrix}, \tag{5}$$

with $r = 1 - t\, \boldsymbol{g}_1(\xi, \boldsymbol{\theta_0})^\top \boldsymbol{G}_2^{-1}(\xi, \boldsymbol{\theta_0})\boldsymbol{g}_1(\xi, \boldsymbol{\theta_0}) > 0$. The characterization of $\phi_A(\xi, \boldsymbol{\theta_0})$ is given in Yin and Zhou (2017) while Gao and Zhou (2017) gives a characterization of $\phi_D(\xi, \boldsymbol{\theta_0})$. Those results are stated in Lemmas 1 and 2 below. Notice that we have obtained additional results for $\phi_A(\xi, \boldsymbol{\theta_0})$ and $\phi_c(\xi, \boldsymbol{\theta_0})$ in Lemma 1, and the proof of Lemma 1 is in the Appendix.

Lemma 1. *For a nonsingular matrix* $\mathbf{A}(\xi, \boldsymbol{\theta_0})$, *we have*

$$\phi_A(\xi, \boldsymbol{\theta_0}) = tr\left(\mathbf{A}^{-1}(\xi, \boldsymbol{\theta_0})\right) = tr\left(\mathbf{C}\ \mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\right) = tr\left(\mathbf{C}\mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{C}\right), \tag{6}$$

$$\phi_c(\xi, \boldsymbol{\theta_0}) = \mathbf{c}^\top \mathbf{A}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{c} = \mathbf{c}_1^\top \mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{c}_1, \tag{7}$$

where $\mathbf{C} = 0 \oplus \mathbf{I}_q$ *with* $\mathbf{I}_q$ *being the identity matrix of order* q, $\oplus$ *denotes the matrix direct sum, and* $\mathbf{c}_1^\top = (0, \mathbf{c}^\top)$ *is a vector in* R^{q+1}.

Lemma 2. *For a nonsingular matrix* $\mathbf{A}(\xi, \boldsymbol{\theta_0})$, *we have*

$$\phi_D(\xi, \boldsymbol{\theta_0}) = det\left(\mathbf{A}^{-1}(\xi, \boldsymbol{\theta_0})\right) = det\left(\mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\right).$$

By (3) and (4) it is obvious that $\mathbf{B}(\xi_\alpha, \boldsymbol{\theta_0})$ is linear in α, i.e.,

$$\mathbf{B}(\xi_\alpha, \boldsymbol{\theta_0}) = (1-\alpha)\mathbf{B}(\xi_1, \boldsymbol{\theta_0}) + \alpha\mathbf{B}(\xi_2, \boldsymbol{\theta_0}).$$

From Lemmas 1 and 2, $\phi_A(\xi_\alpha, \boldsymbol{\theta_0})$, $\phi_c(\xi_\alpha, \boldsymbol{\theta_0})$, $-\left(\phi_D(\xi_\alpha, \boldsymbol{\theta_0})\right)^{-1/(q+1)}$ and $\log\left(\phi_D(\xi_\alpha, \boldsymbol{\theta_0})\right)$ are convex functions of α (Boyd and Vandenberghe, 2004). Similar convexity results are given in Bose and Mukerjee (2015). The above results clearly indicate that matrix $\mathbf{B}(\xi, \boldsymbol{\theta_0})$ plays an important role for us to explore various theoretical results and to develop efficient numerical algorithms for finding optimal designs under the SLSE.

2.2 Equivalence results for optimal designs under SLSE

Define vector $\mathbf{f}(\mathbf{x}, \boldsymbol{\theta_0}) = \frac{\partial g(\mathbf{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mid_{\boldsymbol{\theta}=\boldsymbol{\theta_0}}$ and $(q+1) \times (q+1)$ matrix

$$\mathbf{M}(\mathbf{x}, \boldsymbol{\theta_0}) = \begin{pmatrix} 1 & \sqrt{t}\, \mathbf{f}^\top(\mathbf{x}, \boldsymbol{\theta_0}) \\ \sqrt{t}\, \mathbf{f}(\mathbf{x}, \boldsymbol{\theta_0}) & \mathbf{f}(\mathbf{x}, \boldsymbol{\theta_0})\, \mathbf{f}^\top(\mathbf{x}, \boldsymbol{\theta_0}) \end{pmatrix}, \quad \text{for } \mathbf{x} \in S. \tag{8}$$

Then we have $\mathbf{B}(\xi, \boldsymbol{\theta_0}) = E\left(\mathbf{M}(\mathbf{x}, \boldsymbol{\theta_0})\right)$ and the expectation is taken with respect to $\xi(\mathbf{x})$. Following the derivations in Kiefer and Wolfowitz (1959), Kiefer (1974) and Silvey (1980), we obtain the equivalence results for optimal designs under the SLSE. To present the results, define functions

$$d_A(\mathbf{x}, \xi) = \operatorname{tr}\left(\mathbf{M}(\mathbf{x}, \boldsymbol{\theta_0})\mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{C}\mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\right) - \operatorname{tr}\left(\mathbf{C}\mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{C}\right), \tag{9}$$

$$d_D(\mathbf{x}, \xi) = \operatorname{tr}\left(\mathbf{M}(\mathbf{x}, \boldsymbol{\theta_0})\mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\right) - (q+1), \tag{10}$$

$$d_c(\mathbf{x}, \xi) = \mathbf{c}_1^\top \mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{M}(\mathbf{x}, \boldsymbol{\theta_0})\mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{c}_1 - \mathbf{c}_1^\top \mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{c}_1. \tag{11}$$

Theorem 1. *Consider optimal designs under the SLSE for model (1) with design space S.*

(i) ξ_A^* is an A-optimal design if and only if it satisfies that $\mathbf{B}(\xi_A^*, \boldsymbol{\theta_0})$ is nonsingular and $d_A(\mathbf{x}, \xi_A^*) \le 0$ for all $\mathbf{x} \in S$.

(ii) ξ_D^* is a D-optimal design if and only if it satisfies that $\mathbf{B}(\xi_D^*, \boldsymbol{\theta_0})$ is nonsingular and $d_D(\mathbf{x}, \xi_D^*) \le 0$ for all $\mathbf{x} \in S$.

(iii) ξ_c^* is a c-optimal design if and only if it satisfies that $\mathbf{B}(\xi_c^*, \boldsymbol{\theta_0})$ is nonsingular and $d_c(\mathbf{x}, \xi_c^*) \le 0$ for all $\mathbf{x} \in S$.

For each case, the equality holds at the support points of the optimal design.

The proof of Theorem 1 is similar to the derivation in Silvey (1980) and is omitted. Theorem 1 is helpful to derive the minimal number of support points in various optimal designs. It is also useful to verify optimal designs computed from the numerical algorithm in Section 4.

3 Number of support points in optimal designs

It is always interesting to study the minimal number of support points in optimal designs. For instance, knowing the number of support points in optimal designs can simplify the numerical determination of optimal designs. Gao and Zhou (2017) and Yin and Zhou (2017) have commented the number of support points in optimal designs under SLSE for various models based on their numerical results. Let n_A, n_c and n_D denote the minimal number of support points in ξ_A^*, ξ_c^* and ξ_D^*, respectively. Here we derive theoretical results for n_A, n_c and n_D for several linear and nonlinear models, including polynomial, fractional polynomial, Michaelis-Menten, Peleg, and trigonometric models.

3.1 Polynomial models

Consider a polynomial model of degree q ($q \geq 1$) without intercept,

$$y_i = \theta_1 x_i + \cdots + \theta_q x_i^q + \epsilon_i, \; x_i \in S = [-1, 1], \quad i = 1, \cdots, n. \tag{12}$$

Gao and Zhou (2017) showed that the D-optimal design for (12) is symmetric on S, while Yin and Zhou (2017) proved that the A-optimal design is also symmetric. If there is an intercept term in (12), then the A- and D-optimal designs under the SLSE are the same as those under the LSE (Gao and Zhou, 2014). For (12), $\mathbf{f}(x, \boldsymbol{\theta_0}) = (x, \cdots, x^q)^\top$ and

$$\mathbf{M}(x, \boldsymbol{\theta_0}) = \begin{pmatrix} 1 & \sqrt{t}\, x & \cdots & \sqrt{t}\, x^q \\ \sqrt{t}\, x & x^2 & \cdots & x^{q+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{t}\, x^q & x^{q+1} & \cdots & x^{2q} \end{pmatrix}. \tag{13}$$

Theorem 2. *For model (12), both $d_A(x, \xi_A^*)$ and $d_D(x, \xi_D^*)$ are polynomial functions of $x \in S$ with degree $2q$ and are symmetric about zero. Furthermore, for $0 \leq t < 1$*

we have

$$n_A = \begin{cases} q+1, & \textit{for odd } q, \\ q \textit{ or } q+1, & \textit{for even } q, \end{cases}$$

and

$$n_D = \begin{cases} q+1, & \textit{for odd } q, \\ q \textit{ or } q+1, & \textit{for even } q. \end{cases}$$

The proof of Theorem 2 is in the Appendix. The results in Theorem 2 are consistent with the numerical results in Gao and Zhou (2017) and Yin and Zhou (2017). Notice that Gao and Zhou (2017) gave the results about the number of support points for D-optimal designs based on the moment theory, but Theorem 2 generalizes the results for both A- and D-optimal designs based on the equivalence results. For c-optimal designs it is also easy to show that $n_c \leq q+1$. Example 1 below illustrates the results in Theorems 1 and 2 for an even q.

Example 1. Consider model (12) with $q = 2$. Write a design ξ with support points $x_1^*, \cdots, x_m^*$ as

$$\xi = \begin{pmatrix} x_1^* & \cdots & x_m^* \\ p_1 & \cdots & p_m \end{pmatrix},$$

where $p_1, \cdots, p_m$ are the positive weights for the support points, respectively. It is easy to show that A-optimal and D-optimal designs are given by

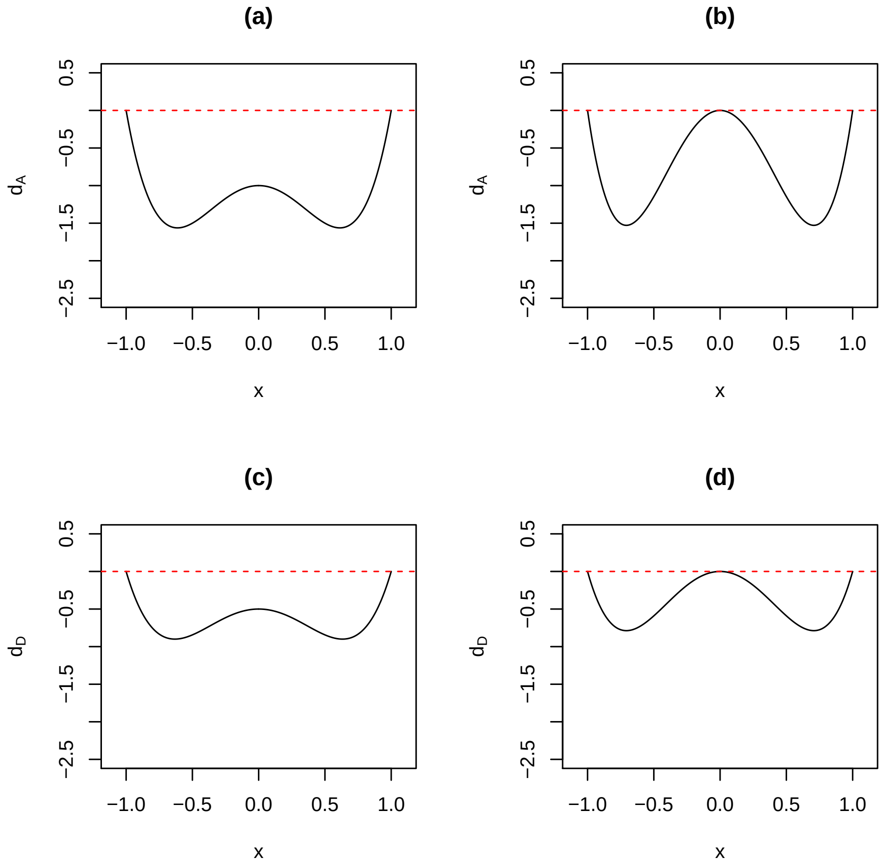
$$\xi_A^* = \begin{pmatrix} -1 & +1 \\ 0.5 & 0.5 \end{pmatrix}, \quad \text{for } 0 \leq t \leq 2-\sqrt{2};$$

$$\xi_A^* = \begin{pmatrix} -1 & 0 & +1 \\ \frac{2-\sqrt{2}}{2t} & \frac{t-2+\sqrt{2}}{t} & \frac{2-\sqrt{2}}{2t} \end{pmatrix}, \quad \text{for } 2-\sqrt{2} < t < 1;$$

$$\xi_D^* = \begin{pmatrix} -1 & +1 \\ 0.5 & 0.5 \end{pmatrix}, \quad \text{for } 0 \leq t \leq 2/3;$$

$$\xi_D^* = \begin{pmatrix} -1 & 0 & +1 \\ \frac{1}{3t} & \frac{3t-2}{3t} & \frac{1}{3t} \end{pmatrix}, \quad \text{for } 2/3 < t < 1.$$

Figure 1: Plots of $d_A(x, \xi_A^*)$ and $d_D(x, \xi_D^*)$ versus x: (a) A-optimality and $t = 0.5$, (b) A-optimality and $t = 0.6$, (c) D-optimality and $t = 0.6$, (d) D-optimality and $t = 0.7$.



It is clear that the optimal designs have either 2 or 3 support points, which is consistent with Theorem 2. The result of ξ_D^* is obtained in Gao and Zhou (2014), while the result of ξ_A^* is new. Figure 1 shows the plots of $d_A(x, \xi_A^*)$ versus x and $d_D(x, \xi_D^*)$ versus x for several t values. It is obvious that $d_A(x, \xi_A^*) \leq 0$ and $d_D(x, \xi_D^*) \leq 0$ for all $x \in S = [-1, 1]$, and they are consistent with Theorem 1. □

If the design space in (12) is asymmetric, say $S = [a, 1]$ with $-1 < a < 1$, then n_A is either q or $q + 1$ and n_D is either q or $q + 1$ for all q. The derivation of these results is similar to the proof of Theorem 2 and is omitted.

Fractional polynomial models are useful for various practical applications. For example, Torsney and Alahmadi (1995) constructed optimal designs for a chemical experiment which studies the relationship between the viscosity y and the concentration x of a chemical solution, and a fractional polynomial model is given by

$$y = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2 + \epsilon, \quad x \in (0.0, 0.2]. \tag{14}$$

Royston et al. (1999) discussed the use of fractional polynomials to model continuous risk variables in epidemiology. Model (14) is also studied in Mandal et al. (2017). By applying the results for model (12), we can easily obtain the number of support points in A- and D-optimal designs under SLSE for fractional polynomial models. Here we use (14) for illustration. Define variable $z = x^{1/2}$, so model (14) is a polynomial model in z with degree 4 and the design space becomes $(0.0, \sqrt{0.2}\ ]$. Following the discussion for polynomial models, we can show that n_A and n_D are at most 4 or 5. Since there are only three unknown parameters in the model, $\boldsymbol{G}_2(\xi, \boldsymbol{\theta}_0)$ is a 3×3 matrix and there should be at least 3 support points in optimal designs to have a nonsingular $\boldsymbol{G}_2(\xi, \boldsymbol{\theta}_0)$. Thus, the possible numbers of n_A and n_D for model (14) are 3, 4, or 5.

3.2 Michaelis-Menten and Peleg models

The Michaelis-Menten model is one of the simplest and best-known approaches to enzyme kinetics, which is given by

$$y_i = \frac{\theta_1 x_i}{\theta_2 + x_i} + \epsilon_i, \qquad \theta_1 > 0,\ \theta_2 > 0, \quad \text{and } x_i \in S = [0, b]. \tag{15}$$

Optimal designs have been studied by several authors using various criteria. For example, see Dette et al. (2005). For this model we have $\mathbf{f}(x, \boldsymbol{\theta_0}) = \left(\frac{x}{\theta_{20}+x}, \ \frac{-\theta_{10}x}{(\theta_{20}+x)^2}\right)^\top$, where $\boldsymbol{\theta_0} = (\theta_{10}, \theta_{20})^\top$, and

$$\mathbf{M}(x, \boldsymbol{\theta_0}) = \begin{pmatrix} 1 & \sqrt{t}\ \mathbf{f}^\top(x, \boldsymbol{\theta_0}) \\ \sqrt{t}\ \mathbf{f}(x, \boldsymbol{\theta_0}) & \mathbf{f}(x, \boldsymbol{\theta_0})\mathbf{f}^\top(x, \boldsymbol{\theta_0}) \end{pmatrix}_{3\times 3}.$$

If we define $z = 1/(\theta_{20} + x)$ for $x \in S = [0, b]$, then we can write $\mathbf{f}(x, \boldsymbol{\theta_0}) = (1 - \theta_{20}z, \; -\theta_{10}z(1 - \theta_{20}z))^\top$. Thus, all the elements of $\mathbf{f}(x, \boldsymbol{\theta_0})$ and $\mathbf{M}(x, \boldsymbol{\theta_0})$ are polynomial functions of z and $z \in S_z = [1/(\theta_{20} + b), 1/\theta_{20}]$. This implies that $d_A(x, \xi_A^*)$ and $d_D(x, \xi_D^*)$ can be viewed as polynomial functions of z with degree 4. By analyzing the roots of $d_A(x, \xi_A^*) = 0$ and $d_D(x, \xi_D^*) = 0$ in terms of $z \in S_z$, similar to those for polynomial model (12), we obtain the result for n_A and n_D for model (15) as follows.

Theorem 3. *For model (15) n_A and n_D are either 2 or 3 for $0 \leq t < 1$, and $n_D = 2$ for $t = 0$.*

The proof of Theorem 3 is omitted since it is similar to that of Theorem 2. For $t = 0$, it is easy to show that $d_D(0, \xi_D^*) < 0$, which implies that the boundary point zero is not a support point in ξ_D^*. Thus, we have $n_D = 2$. We also have $n_c \leq 3$ for $0 \leq t < 1$.

Peleg model is used to study water absorption kinetics in many substances and food ingredients, and optimal designs under LSE are constructed, for instance, in Paquet-Durand et al. (2015). The model is given by

$$y = m_0 + \frac{x}{\theta_1 + \theta_2 x} + \epsilon, \quad 0 \leq x \leq b, \; \theta_1 > 0, \; \theta_2 > 0, \tag{16}$$

where y denotes the water content at the time x, m_0 is the initial water content measured at time zero, and θ_1 and θ_2 are the unknown parameters. For this model we have $\mathbf{f}(x, \boldsymbol{\theta}) = \left(\frac{-x}{(\theta_1+\theta_2 x)^2}, \frac{-x^2}{(\theta_1+\theta_2 x)^2}\right)^\top$. Define $z = 1/(\theta_1 + \theta_2 x)$. For $x \in [0, b]$, we get $z \in [1/(\theta_1 + \theta_2 b), 1/\theta_1]$. Then we can write $\mathbf{f}(x, \boldsymbol{\theta}) = (-z(1 - \theta_1 z)/\theta_2, \; -(1 - \theta_1 z)^2/\theta_2^2)^\top$. Thus, all the elements in $\mathbf{M}(x, \boldsymbol{\theta})$ are polynomial functions of z, with highest degree 4. Similar to the discussion for Michaelis-Menten model, we have $2 \leq n_A \leq 3$, $2 \leq n_D \leq 3$, and $n_c \leq 3$ for all $0 \leq t < 1$.

3.3 Trigonometric regression models

Now let us consider trigonometric regression models, for $i = 1, \cdots, n$,

$$y_i = \sum_{j=1}^{k} \theta_{1j} \cos(jx_i) + \sum_{j=1}^{k} \theta_{2j} \sin(jx_i) + \epsilon_i, \; x_i \in S_b = [-b, b], \; 0 < b \leq \pi, \tag{17}$$

and let parameter vector $\boldsymbol{\theta} = (\theta_{11}, \cdots, \theta_{1k}, \theta_{21}, \cdots, \theta_{2k})^\top$ and $q = 2k$. For (17) we have $\mathbf{f}(x, \boldsymbol{\theta_0}) = (\cos(x), \cdots, \cos(kx), \sin(x), \cdots, \sin(kx))^\top$, and

$$\mathbf{M}(x, \boldsymbol{\theta_0}) = \begin{pmatrix} 1 & \sqrt{t}\ \cos(x) & \cdots & \sqrt{t}\ \sin(kx) \\ \sqrt{t}\ \cos(x) & \cos^2(x) & \cdots & \cos(x)\sin(kx) \\ \vdots & \vdots & \ddots & \vdots \\ \sqrt{t}\ \sin(kx) & \cos(x)\sin(kx) & \cdots & \sin^2(kx) \end{pmatrix}_{(2k+1)\times(2k+1)}. \tag{18}$$

We study n_A and n_D for two cases: (i) $b = \pi$, (ii) $0 < b < \pi$.

Consider case (i): $b = \pi$ and $S_b = [-\pi, \pi]$. We notice that, for $j, \; l = 1, 2, \cdots, k$,

$$\begin{aligned} &\int_{-\pi}^{\pi} \cos(jx)dx = \int_{-\pi}^{\pi} \sin(jx)dx = 0, \quad \int_{-\pi}^{\pi} \cos(jx)\ \sin(lx)dx = 0, \\ &\int_{-\pi}^{\pi} \cos^2(jx)dx = \int_{-\pi}^{\pi} \sin^2(jx)dx = \pi, \\ &\int_{-\pi}^{\pi} \cos(jx)\ \cos(lx)dx = \int_{-\pi}^{\pi} \sin(jx)\ \sin(lx)dx = 0, \quad \text{for } j \neq l. \end{aligned} \tag{19}$$

The A-optimal and D-optimal designs are given in the following theorem, and its proof is in the Appendix.

Theorem 4. *For model (17) with $S_b = [-\pi, \pi]$, the ξ_A^* and ξ_D^* are the uniform distribution on S_b for all $0 \leq t < 1$.*

For case (ii), $0 < b < \pi$ and $S_b = [-b, b]$, define $z = \cos(x)$. We discuss the number of support points for $x \in S_b$ in optimal designs through the number of distinct points for $z \in [\cos(b), 1]$. Since $\cos(x)$ is an even function of x, each point $z \in [\cos(b), 1]$ corresponds to two symmetric points $\pm\, x \in S_b$. From Gao and Zhou (2017) all the elements of matrix $\mathbf{M}(x, \boldsymbol{\theta}_0)$ in (18) can be written as polynomial functions of z with the highest degree $2k$. By (9), (10) and (11), functions d_A, d_D and d_c are polynomial functions of z with the highest degree $2k$. From Theorem 1 and the discussion in Section 3.1, the number of support points in optimal designs in terms of $z \in [\cos(b), 1]$ are at most $k + 1$.

4 Numerical algorithm and applications

Although we have derived the equivalence results in Theorem 1 and have obtained the number of support points in optimal designs for various models, it may be still challenging to find optimal designs ξ_A^*, ξ_c^* and ξ_D^* analytically. In addition, there are no theoretical results about n_A, n_c and n_D for many complicated linear/nonlinear models. When we use numerical algorithms to compute optimal designs, the theoretical results in Lemmas 1 and 2 and Theorem 1 are helpful and the details are explained in this section. A couple of algorithms are effective and efficient for finding optimal designs on discretized design space, say S_N, with N points. Here we will apply the CVX program in MATLAB to solve optimal design problems on S_N and the algorithm is described as follows.

The CVX program is developed to solve convex optimization problems (Boyd and Vandenberghe, 2004). If the design space S is not discrete, we first discretize it and form a discrete design space S_N. Often we use equally spaced grid points to get S_N, where N is a user selected number and can be very large. Let $S_N = \{\mathbf{u}_1, \cdots, \mathbf{u}_N\} \subset S \subset R^p$. Let Ξ_N denote the class of probability distributions on S_N and each distribution is written as

$$\xi = \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_N \\ w_1 & \cdots & w_N \end{pmatrix},$$

where the weights, $w_1, \cdots, w_N$, satisfy $w_i \geq 0$ for all i and $\sum_{i=1}^N w_i = 1$. For any $\xi \in \Xi_N$, we have

$$\mathbf{B}(\xi, \boldsymbol{\theta_0}) = E\left(\mathbf{M}(\mathbf{x}, \boldsymbol{\theta_0})\right) = \sum_{i=1}^N w_i \mathbf{M}(\mathbf{u}_i, \boldsymbol{\theta_0}),$$

where $\mathbf{M}(\mathbf{x}, \boldsymbol{\theta_0})$ is defined in (8). It is obvious that, for a given parameter vector $\boldsymbol{\theta_0}$, $\mathbf{B}(\xi, \boldsymbol{\theta_0})$ is linear in $\mathbf{w}$, where $\mathbf{w} = (w_1, \cdots, w_N)$ is called a weight vector. From Lemmas 1 and 2, $-\left(\phi_D(\xi, \boldsymbol{\theta_0})\right)^{-1/(q+1)}$, $\phi_A(\xi, \boldsymbol{\theta_0})$ and $\phi_c(\xi, \boldsymbol{\theta_0})$ are all convex functions of $\mathbf{w}$. Thus, the optimal design problems on S_N can be written as convex optimization problems as follows

$$\begin{cases} \quad\quad \min_{\mathbf{w}} & \phi(\xi, \boldsymbol{\theta_0}) \\ \text{subject to:} & w_i \geq 0, \ i = 1, \cdots, N, \quad \sum_{i=1}^{N} w_i = 1, \end{cases} \tag{20}$$

where $\phi(\xi, \boldsymbol{\theta_0})$ can be $-\left(\phi_D(\xi, \boldsymbol{\theta_0})\right)^{-1/(q+1)}$, $\phi_A(\xi, \boldsymbol{\theta_0})$ or $\phi_c(\xi, \boldsymbol{\theta_0})$.

As $N \to \infty$, the optimal designs on S_N usually converge to those on S. See the results for the optimal designs under LSE in Ye et al. (2017). Gao and Zhou (2017) applied the CVX program to compute the optimal moments in D-optimal designs. Similarly, the CVX program can be applied for finding the solutions of problem (20) including A-, c- and D-optimal designs, and it is effective and efficient. Here is an algorithm with the detailed steps using the CVX program.

Algorithm I: Computing optimal designs via CVX

Step (i): Input true parameter value $\boldsymbol{\theta}_0$ and design points $\mathbf{u}_1, \cdots, \mathbf{u}_N$.

Step (ii): Compute $(q+1) \times (q+1)$ matrix $\mathbf{M}(\mathbf{u}_i, \boldsymbol{\theta}_0)$ at each $\mathbf{u}_i$ for $i = 1, \cdots, N$.

Step (iii): Use CVX in MATLAB to solve problem (20) for $\mathbf{w}$ and denote the result as $\hat{\mathbf{w}}$.

Step (iv): Check for optimality condition for $\hat{\mathbf{w}}$.

The algorithm is simple and can be applied to any linear/nonlinear model and A-, c- and D-optimality criteria. We do not need to know n_A, n_c and n_D to use the algorithm. Steps (i) and (ii) depend on the model and the design space, while the optimality criterion is specified in Step (iii) through the objective function $\phi(\xi, \boldsymbol{\theta_0})$ in problem (20). In Step (iv), we verify if the numerical result $\hat{\mathbf{w}}$ is an optimal

Table 1: Optimal designs for Peleg model with $\boldsymbol{\theta_0} = (0.5, 0.05)^\top$ and $S = [0, 100]$

t		optimal design ξ^*: support points with weights in parentheses	$\phi(\xi^*, \boldsymbol{\theta_0})$	d_{max}
0	A-	6.10 (0.850), 100 (0.150)	0.01770	2.36×10^{-7}
	c-	6.00 (0.875), 100 (0.125)	0.01649	2.10×10^{-8}
	D-	8.30 (0.500), 100 (0.500)	-131.18975	1.37×10^{-6}
0.3	A-	6.80 (0.833), 100 (0.167)	0.02128	3.04×10^{-10}
	c-	6.80 (0.854), 100 (0.146)	0.02023	3.10×10^{-8}
	D-	8.30 (0.500), 100 (0.500)	-116.48391	1.32×10^{-6}
0.7	A-	0.0 (0.108), 8.30 (0.713), 100 (0.179)	0.03395	5.49×10^{-8}
	c-	0.0 (0.128), 8.30 (0.714), 100 (0.158)	0.03321	2.72×10^{-8}
	D-	0.0 (0.048), 8.30 (0.476), 100 (0.476)	-88.05076	2.75×10^{-5}

design. The optimality conditions are obtained in Theorem 1. In practice, those conditions are implemented as follows,

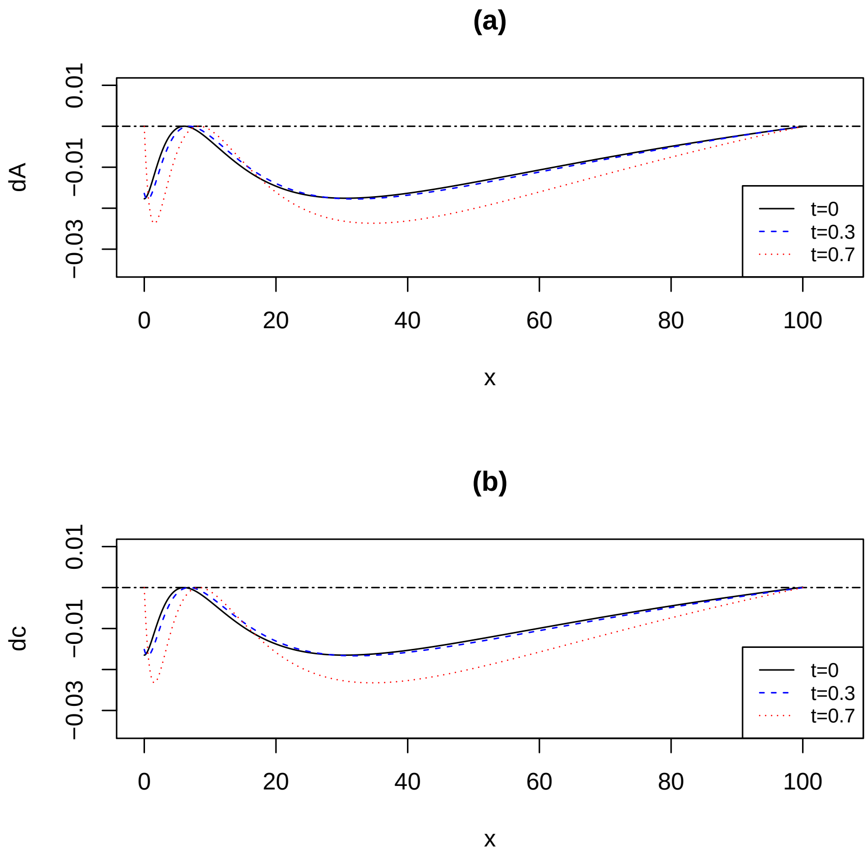
$$d_A(\mathbf{x}, \xi_A^*) \leq \delta, \quad d_D(\mathbf{x}, \xi_D^*) \leq \delta, \quad d_c(\mathbf{x}, \xi_c^*) \leq \delta, \quad \text{for all } \; \mathbf{x} \in S_N,$$

where δ is a small positive number, say 10^{-4}.

Example 2 below illustrates the use of Algorithm I, and the MATLAB codes are available from an online supplement of this paper. All the computation is done on a PC (Intel(R) Core(TM)2 Quad CPU Q9550@2.83GHz).

Example 2. Consider Peleg model in (16) with $\boldsymbol{\theta_0} = (0.5, 0.05)^\top$ and design space $S = [0, 100]$. To apply Algorithm I, we first discretize S into S_N consisting of N equally spaced grid points $u_i = 100(i-1)/(N-1)$, $i = 1, \cdots, N$. In Step (ii) of Algorithm I, we can evaluate 3×3 matrix $\mathbf{M}(u_i, \boldsymbol{\theta}_0)$ easily for this model. We compute A-, c- and D-optimal designs for various values of t, where $\mathbf{c} = (1, 1)^\top$ is used for c-optimality. Let $d_{max} = \max d_A(\mathbf{x}, \xi_A^*)$ ($\max d_D(\mathbf{x}, \xi_D^*)$, or $\max d_c(\mathbf{x}, \xi_c^*)$) for A-optimality (D-optimality or c-optimality). Algorithm I is fast, and it takes less than 3 seconds to find the optimal designs for $N = 1001$. Representative results

Figure 2: Plots of $d_A(x, \xi_A^*)$ and $d_c(x, \xi_c^*)$ versus x for Peleg model: (a) A-optimality with $t = 0,\ 0.3$ and 0.7, (b) c-optimality with $t = 0,\ 0.3$ and 0.7.



are given in Table 1. For all the optimal designs in Table 1, we have $d_{max} < 10^{-4}$, so the optimality conditions in Theorem 1 are satisfied. We have also plotted functions $d_A(x, \xi_A^*)$ and $d_c(x, \xi_c^*)$ versus x in Figure 2 to check for the support points in the optimal designs and the optimality conditions. Function $d_D(x, \xi_D^*)$ is similar and is omitted. It is clear that the optimality conditions are satisfied. The number of support points is consistent with the results in Section 3.2. There are 2 or 3 support points in all the optimal designs. When t is small, there are 2 support points. When t is large, there are 3 support points. All the optimal designs include the boundary point 100 as a support point. $\square$

Table 2: Relative efficiencies for the optimal designs in Example 1, where Bayesian* denotes the Bayesian optimal designs with $t_1 = 0.5$ and $t_2 = 0.8$

t_0	$Eff_A(t_0, t_*)$				$Eff_D(t_0, t_*)$			
	t_*				t_*			
	0.3	0.5	0.7	0.9	0.3	0.5	0.7	0.9
0.4	1.000	1.000	0.941	0.477	1.000	1.000	0.996	0.739
0.6	0.982	0.991	0.958	0.554	1.000	1.000	0.996	0.739
0.8	0.779	0.852	0.979	0.977	0.863	0.900	0.978	0.974
Bayesian*	0.853	0.910	0.999	0.898	0.973	0.982	0.999	0.816

For practical applications, t is often unknown. We now study the efficiency of optimal designs if t is misspecified. Define the following relative efficiency measures for A-, c- and D-optimal designs, respectively,

$$Eff_A(t_0, t_*) = \frac{\phi_A(\xi_A^*, \boldsymbol{\theta}_0)}{\phi_A(\xi_A^0, \boldsymbol{\theta}_0)}, \quad Eff_c(t_0, t_*) = \frac{\phi_c(\xi_c^*, \boldsymbol{\theta}_0)}{\phi_c(\xi_c^0, \boldsymbol{\theta}_0)}, \quad Eff_D(t_0, t_*) = \left(\frac{\phi_D(\xi_D^*, \boldsymbol{\theta}_0)}{\phi_D(\xi_D^0, \boldsymbol{\theta}_0)}\right)^{1/q},$$

where t_* is the true value of t and the covariance matrix of the SLSE is evaluated at $t = t_*$, ξ_A^0, ξ_c^0 and ξ_D^0 are the corresponding optimal designs obtained by using $t = t_0$.

We have computed the relative efficiencies using various values of t_0 and t_* for the A- and D-optimal designs in Example 1, and the representative results are given in Table 2. The results indicate that (i) the efficiency is high if there is a small misspecification of t, (ii) when we compare the two cases: (a) t_0 is overestimate t_*, (b) t_0 is underestimate t_*, the efficiency is higher for case (a), (iii) D-optimal designs are less sensitive to t than A-optimal designs. Thus, we suggest to use large t to construct optimal designs for practical applications with asymmetric errors. This also agrees with the reason that we use the SLSE. In Example 2, the efficiencies for $t_0 = 0.7$ and $t_* = 0.3$ are 0.886, 0.872 and 0.962 for A-, c- and D-optimal designs, respectively.

An alternative way to deal with unknown parameter t is to construct Bayesian optimal designs. To focus on the idea, we use an uniform discrete prior distribution for t with two possible values t_1 and t_2. Bayesian A-, c- and D-optimal designs on S_N, denoted by $\xi^*_{A,[t_1,t_2]}$, $\xi^*_{c,[t_1,t_2]}$, and $\xi^*_{D,[t_1,t_2]}$, minimize loss function $0.5\phi(\xi, \boldsymbol{\theta}_0)|_{t=t_1} + 0.5\phi(\xi, \boldsymbol{\theta}_0)|_{t=t_2}$, where function ϕ is defined in (20). The optimal design problems are similar to those in (20). Since $0.5\phi(\xi, \boldsymbol{\theta}_0)|_{t=t_1} + 0.5\phi(\xi, \boldsymbol{\theta}_0)|_{t=t_2}$ is still convex function of $\mathbf{w}$, we can apply Algorithm I to find the Bayesian optimal designs. In Example 1, using $t_1 = 0.5$ and $t_2 = 0.8$, we obtain the following Bayesian optimal designs,

$$\xi^*_{A,[0.5,0.8]} = \begin{pmatrix} -1 & 0 & +1 \\ 0.408 & 0.184 & 0.408 \end{pmatrix}, \quad \xi^*_{D,[0.5,0.8]} = \begin{pmatrix} -1 & 0 & +1 \\ 0.483 & 0.034 & 0.483 \end{pmatrix}.$$

The efficiencies of the Bayesian optimal designs are presented in Table 2, and it is clear that the Bayesian optimal designs have high efficiency over a range of true values of t. However, the optimal choice of the prior distribution for t depends on regression models and needs further research.

5 Generalized scale invariance of D-optimal designs

D-optimal designs under LSE are scale invariant for some linear regression models, which allows us to scale the design space and present D-optimal designs on the scaled design space. However, for nonlinear models we usually do not have the scale invariance property and the (local) optimal designs also depend on the true parameter value $\boldsymbol{\theta_0}$. For a given nonlinear model D-optimal designs have to be constructed for each design space and each true parameter value. Here a generalized scale invariance concept is studied for D-optimal designs under the SLSE such that the D-optimal designs can also be constructed on the scaled design space.

For a given model, let $\xi^*_D(S, \boldsymbol{\theta_0})$ be the D-optimal design on S with true parameter vector $\boldsymbol{\theta_0}$. Let S^V be a scaled design space from S, defined by a diagonal matrix $V = diag\{v_1, \cdots, v_p\}$ with positive diagonal elements v_i and $S^V = \{\ V\mathbf{x}\ |\mathbf{x} \in S\}$. Transformation from $\mathbf{x}$ to $V\mathbf{x}$ is called a scale transformation, and V is called a scale matrix. If there exists a parameter vector $\tilde{\boldsymbol{\theta_0}}$ such that the D-optimal design on S^V with $\tilde{\boldsymbol{\theta_0}}$, denoted by $\xi_D^*(S^V, \tilde{\boldsymbol{\theta_0}})$, can be obtained from $\xi_D^*(S, \boldsymbol{\theta_0})$ by the scale transformation, then we say that $\xi_D^*(S, \boldsymbol{\theta_0})$ is scale invariant for the model. This property of $\xi_D^*(S, \boldsymbol{\theta_0})$ is defined as a generalized scale invariance. Note that $\tilde{\boldsymbol{\theta_0}}$ is often related to $\boldsymbol{\theta_0}$ through $v_1, \cdots, v_p$ in scale matrix V. For linear models, since $\xi_D^*(S, \boldsymbol{\theta_0})$ does not depend on $\boldsymbol{\theta_0}$, the generalized scale invariance of $\xi_D^*(S, \boldsymbol{\theta_0})$ becomes the traditional scale invariance. The generalized scale invariance property holds for D-optimal designs under both LSE and SLSE for various nonlinear models, and the following lemma provides a condition to check for this property.

Lemma 3. *For a given model and scale matrix V, if there exists a parameter vector $\tilde{\boldsymbol{\theta_0}}$ and a nonsingular diagonal matrix $\mathbf{Q}$ (not depending on $\mathbf{x}$) such that $\mathbf{f}(V\mathbf{x}, \tilde{\boldsymbol{\theta_0}}) = \mathbf{Q}\,\mathbf{f}(\mathbf{x}, \boldsymbol{\theta_0})$ for all $\mathbf{x} \in S$, then $\xi_D^*(S, \boldsymbol{\theta_0})$ has the generalized scale invariance property.*

The proof of Lemma 3 is in the Appendix. With the generalized scale invariance property we can scale the design space to present D-optimal designs. In Example 2, we can verify that the D-optimal designs for Peleg model are scale invariant. Since $p = 1$, $V = v_1 > 0$ and $\boldsymbol{\theta_0} = (\theta_1, \theta_2)^\top$, we can set $\tilde{\boldsymbol{\theta_0}} = (\theta_1, \theta_2/v_1)^\top$. Then we have

$$\mathbf{f}(V\mathbf{x}, \tilde{\boldsymbol{\theta_0}}) = \begin{pmatrix} \frac{-v_1 x}{\left(\theta_1 + \frac{\theta_2}{v_1}\ v_1 x\right)^2} \\ \frac{-(v_1 x)^2}{\left(\theta_1 + \frac{\theta_2}{v_1}\ v_1 x\right)^2} \end{pmatrix} = \mathbf{Q} \begin{pmatrix} \frac{-x}{(\theta_1+\theta_2 x)^2} \\ \frac{-x^2}{(\theta_1+\theta_2 x)^2} \end{pmatrix} = \mathbf{Q}\,\mathbf{f}(\mathbf{x}, \boldsymbol{\theta_0}), \text{ with } \mathbf{Q} = diag\{v_1, v_1^2\},$$

which satisfies the condition in Lemma 3. It is also easy to verify the result numerically by computing D-optimal designs on S and S^V with corresponding true parameter values using Algorithm I.

If matrix $\mathbf{B}(\xi, \boldsymbol{\theta_0})$ is ill conditioned, then Algorithm I may fail in Step (iii). In some situations, we can apply the generalized scale invariance property to work on a scaled design space such that $\mathbf{B}(\xi, \tilde{\boldsymbol{\theta}_0})$ is not ill conditioned and Algorithm I can find D-optimal designs successfully. Here is one example.

Example 3. Dette et al. (2008) investigated optimal designs under LSE for spline regression models with unknown knots, and the number of support points in D-optimal designs and various properties were obtained. We use one of their models to illustrate the generalized scale invariance property and its usefulness in Algorithm I. The model is given by

$$y = \theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3 + \theta_5 (x - \lambda)_+^3 + \epsilon, \quad x \in [0, b],$$

where function $(x-\lambda)_+ = \max\{0, (x-\lambda)\}$ and parameter vector $\boldsymbol{\theta_0} = (\theta_1, \cdots, \theta_5, \lambda)^\top$. It is a nonlinear regression model and

$$\mathbf{f}(x, \boldsymbol{\theta_0}) = \left(1, x, x^2, x^3, (x-\lambda)_+^3, -3\theta_5(x-\lambda)_+^2 \right)^\top.$$

Consider a scale transformation $V = v_1 > 0$ and let $\tilde{\boldsymbol{\theta}_0} = (\theta_1, \cdots, \theta_5, v_1\lambda)^\top$. Similar to the discussion for Peleg model above, we have $\mathbf{f}(V\mathbf{x}, \tilde{\boldsymbol{\theta}_0}) = \mathbf{Q}\ \mathbf{f}(\mathbf{x}, \boldsymbol{\theta_0})$ with $\mathbf{Q} = diag\{1, v_1, v_1^2, v_1^3, v_1^3, v_1^2\}$. Thus, the D-optimal designs have the generalized scale invariance property. In addition, since the model is linear in $\theta_1, \cdots, \theta_5$, the D-optimal designs under LSE and SLSE do not depend on $\theta_1, \cdots, \theta_5$ (Yin and Zhou, 2017). Therefore, the D-optimal designs on $S^V = [0, v_1 b]$ with parameters, $\theta_1, \cdots, \theta_5$ and $v_1\lambda$, can be easily obtained from the D-optimal designs on $S = [0, b]$ with parameters, $\theta_1, \cdots, \theta_5$ and λ. This property can be used to avoid ill conditioned matrix $\mathbf{B}(\xi, \boldsymbol{\theta_0})$ in Algorithm I. For instance, we want to find the D-optimal design for design space $S = [0, 10]$ with $\lambda = 8$, which gives ill conditioned matrix $\mathbf{B}(\xi, \boldsymbol{\theta_0})$ and Algorithm I fails to find the D-optimal design. However, we can easily find the D-optimal design for design space $S = [0, 1]$ with $\lambda = 0.80$, and the optimal design has 6 support points: 0, 0.225, 0.590, 0.820, 0.935, 1.0, with equal weights: $1/6, \cdots, 1/6$. The optimal design is consistent with that in Dette et al. (2008). In the computation, we use $N = 1001$ equally spaced points in S_N in Algorithm I. By the scale invariance property, the D-optimal design for $S = [0, 10]$ with $\lambda = 8$ has 6 support points: 0, 2.25, 5.90, 8.20, 9.35, 10.0, with equal weights: $1/6, \cdots, 1/6$. $\square$

Note that the model in Example 3 contains an intercept, so the D-optimal designs under LSE and SLSE are the same (Gao and Zhou, 2014). In addition, Algorithm I can find the D-optimal designs easily for spline regression models with several unknown knots.

6 Conclusion

We have derived the equivalence theorems for optimal designs under the SLSE, which are helpful for analyzing the number of support points in optimal designs and for verifying numerical results as optimal designs. For several linear and nonlinear regression models we have obtained the number of support points analytically. A numerical algorithm is also discussed to find approximate optimal designs on discrete design spaces. It is very efficient and effective, and we have computed optimal designs with N as large as $20,000$. In addition, a generalized scale invariance concept is studied for D-optimal designs, which can be applied for both linear and nonlinear models. This scale invariance property allows us to scale design spaces for finding D-optimal designs and to avoid some computational issues in Algorithm I.

Appendix: Proofs

Proof of Lemma 1: From the fact that $\mathbf{C} = \mathbf{CC}$ and $\phi_A(\xi, \boldsymbol{\theta_0}) = \mathrm{tr}\,(\mathbf{C}\ \mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0}))$ (Yin and Zhou, 2017), we have

$$\phi_A(\xi, \boldsymbol{\theta_0}) = \mathrm{tr}\left(\mathbf{C}\ \mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\right) = \mathrm{tr}\left(\mathbf{CC}\ \mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\right) = \mathrm{tr}\left(\mathbf{C}\ \mathbf{B}^{-1}(\xi, \boldsymbol{\theta_0})\mathbf{C}\right),$$

which gives the result in (6). The proof of (7) is similar and is omitted. □

Proof of Theorem 2: Here we prove the result for D-optimality. The proof of the result for A-optimality is similar and is omitted.

From (10) and (13), it is clear that $d_D(x, \xi_D^*)$ is a polynomial function of x with degree $2q$, and from Theorem 1 the support points of ξ_D^* satisfy $d_D(x, \xi_D^*) = 0$. Since the distribution of ξ_D^* is symmetric about zero on S by Gao and Zhou (2017), function $d_D(x, \xi_D^*)$ must be symmetric about zero too.

In order to have a nonsingular $\mathbf{B}(\xi_D^*, \boldsymbol{\theta_0})$, we need a nonsingular $\boldsymbol{G}_2(\xi_D^*, \boldsymbol{\theta_0})$ by (4), which implies that we must have at least q nonzero support points for model (12). Since $d_D(x, \xi_D^*)$ is a polynomial of degree $2q$, $d_D(x, \xi_D^*) = 0$ has at most $2q$ real roots in S. Gao and Zhou (2017) showed that two boundary points -1 and $+1$ are always support points in ξ_D^* for $q \geq 1$, so there are at most $2q - 2$ real roots of $d_D(x, \xi_D^*) = 0$ in $(-1, 1)$. By Theorem 1, $d_D(x, \xi_D^*) \leq 0$ for all $x \in [-1, 1]$, the roots in $(-1, 1)$ must have multiplicity at least two. Thus, there are at most $2 + (2q - 2)/2 = q + 1$ distinct roots in $[-1, 1]$. Based on the above discussions, we must have $q \leq n_D \leq (q + 1)$.

Notice that function $d_D(x, \xi_D^*)$ is symmetric about zero. If $x_0 \in (-1, 1)$ is a root of $d_D(x, \xi_D^*) = 0$ then $-x_0$ is also a root. When q is odd, all the possible $(q - 1)$ roots in $(-1, 1)$ must be nonzero. Therefore, there are $q + 1$ support points in ξ_D^*.

When q is even, one of the possible $(q - 1)$ roots in $(-1, 1)$ is zero. When zero is a root, there are $(q + 1)$ support points in ξ_D^*. When zero is not a root, there are q support points in ξ_D^*. To determine if zero is a root, we just need to evaluate $d_D(0, \xi_D^*)$. By (5), (10) and (13), we get $d_D(0, \xi_D^*) = \frac{1}{r} - (q + 1)$ with $r = 1 - t\, \boldsymbol{g}_1(\xi_D^*, \boldsymbol{\theta_0})^\top \boldsymbol{G}_2^{-1}(\xi_D^*, \boldsymbol{\theta_0}) \boldsymbol{g}_1(\xi_D^*, \boldsymbol{\theta_0})$. Clearly when $t = 0$, we have $d_D(0, \xi_D^*) = -q < 0$ for all q, which implies that zero is not a root and there are only q support points in ξ_D^*. □

Proof of Theorem 4: Using (19), we can easily verify that,

$$\mathbf{B}(\xi_A^*, \boldsymbol{\theta_0}) = \mathbf{B}(\xi_D^*, \boldsymbol{\theta_0}) = 1 \oplus 0.5\mathbf{I}_{2k},$$

when ξ_A^* and ξ_D^* are the uniform distribution on S_b. Notice that $q = 2k$ here. By

(9), (10) and (18), we obtain

$$\begin{aligned} d_A(x,\xi_A^*) &= \operatorname{tr}\left(\mathbf{M}(x,\boldsymbol{\theta_0})(1\oplus 2\mathbf{I}_{2k})(0\oplus \mathbf{I}_{2k})(1\oplus 2\mathbf{I}_{2k})\right) - \operatorname{tr}\left((0\oplus 2\mathbf{I}_{2k})\right) \\ &= 4\left(\sum_{j=1}^{k}\cos^2(jx)+\sum_{j=1}^{k}\sin^2(jx)\right)-4k \\ &= 0, \quad \text{for all} \;\; x\in S_b=[-\pi,\pi]. \end{aligned}$$

Similarly we have $d_D(x,\xi_D^*) = 0$ for all $x \in S_b = [-\pi,\pi]$. Thus, the uniform distribution on S_b is an A-optimal and D-optimal design for all t. □

Proof of Lemma 3: For design space S and parameter $\boldsymbol{\theta_0}$, $\xi_D^*(S,\boldsymbol{\theta_0})$ minimizes

$$\phi_D(\xi,\boldsymbol{\theta_0}) = \det\left(\mathbf{B}^{-1}(\xi,\boldsymbol{\theta_0})\right),$$

where $\mathbf{B}(\xi,\boldsymbol{\theta_0}) = E\left(\mathbf{M}(\mathbf{x},\boldsymbol{\theta_0})\right)$ and $\mathbf{M}(\mathbf{x},\boldsymbol{\theta_0})$ is given by (8), and the expectation is taken with respect to $\xi(\mathbf{x})$ on S.

For design space $S^V = \{\, V\mathbf{x} \;|\mathbf{x} \in S\}$ and parameter $\tilde{\boldsymbol{\theta_0}}$, we minimize the following function to find $\xi_D^*(S^V,\tilde{\boldsymbol{\theta_0}})$,

$$\phi_D(\xi,\tilde{\boldsymbol{\theta_0}}) = \det\left(\mathbf{B}^{-1}(\xi,\tilde{\boldsymbol{\theta_0}})\right),$$

where $\mathbf{B}(\xi,\tilde{\boldsymbol{\theta_0}}) = E\left(\mathbf{M}(\mathbf{z},\tilde{\boldsymbol{\theta_0}})\right)$, and the expectation is taken with respect to $\xi(\mathbf{z})$ on S^V. By the definition of S^V, each $\mathbf{z}\in S^V$ can be written as $V\mathbf{x}$, where $\mathbf{x}\in S$, so we have

$$\mathbf{M}(\mathbf{z},\tilde{\boldsymbol{\theta_0}}) = \mathbf{M}(V\mathbf{x},\tilde{\boldsymbol{\theta_0}}).$$

By (8) and the assumption in Lemma 3, we obtain

$$\mathbf{M}(V\mathbf{x},\tilde{\boldsymbol{\theta_0}}) = (1\oplus\mathbf{Q})\,\mathbf{M}(\mathbf{x},\boldsymbol{\theta_0})\,(1\oplus\mathbf{Q}), \quad \text{for all} \;\; \mathbf{x}\in S.$$

Thus, $\mathbf{B}(\xi,\tilde{\boldsymbol{\theta_0}}) = (1\oplus\mathbf{Q})\,\mathbf{B}(\xi,\boldsymbol{\theta_0})\,(1\oplus\mathbf{Q})$ and $\phi_D(\xi,\tilde{\boldsymbol{\theta_0}}) = \phi_D(\xi,\boldsymbol{\theta_0})/\left(\det(\mathbf{Q})\right)^2$. This implies that we can minimize $\phi_D(\xi,\boldsymbol{\theta_0})$ to get $\xi_D^*(S^V,\tilde{\boldsymbol{\theta_0}})$, and $\xi_D^*(S^V,\tilde{\boldsymbol{\theta_0}})$ is the scale transformation from $\xi_D^*(S,\boldsymbol{\theta_0})$. □

Supplementary material

We have provided the MATLAB codes of Example 2 for computing A-, c- and D-optimal designs in the supplementary material. These codes can be modified for finding optimal designs for other models and design spaces.

Acknowledgements

This research work is supported by Discovery Grants from the Natural Science and Engineering Research Council of Canada.

References

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.

Dette, H., Melas, V.B. and Pepelyshev, A. (2008). Optimal designs for free knot least squares splines. *Statistica Sinica*, 18, 1047-1062.

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

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. (2017). D-optimal designs based on the second-order least squares estimator. *Statistical Papers*, 58, 77-94.

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.)

Kiefer, J. (1974). General equivalence theorem for optimum designs (approximate theory). *The Annals of Statistics*, 2, 849-879.

Kiefer, J. and Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 30, 271-294.

Mandal, S., Torsney, B. and Chowdhury, M. (2017). Optimal designs for minimising covariances among parameter estimators in a linear model. *Australian & New Zealand Journal of Statistics*, 59, 255 - 273.

Paquet-Durand, O., Zettel, V. and Hitzmann, B. (2015). Optimal experimental design for parameter estimation of the Peleg model. *Chemometrics and Intelligent Laboratory Systems*, 140, 36-42.

Royston, P., Ambler, G. and Sauerbrei, W. (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology*, 28, 964-974.

Silvey, S.D. (1980). *Optimal Designs: An Introduction to the Theory for Parameter Estimation.* Chapman and Hall, London.

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

Torsney, B. and Alahmadi, A.M. (1995). Designing for minimally dependent observations. *Statistica Sinica*, 5, 499-514.

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

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. and Zhou, J. (2017). Optimal designs for regression models using the second-order least squares estimator. *Statistica Sinica*, 27, 1841-1856.