SIAM J. OPTIM.
Vol. 23, No. 1, pp. 666–686

MINIMIZING THE CONDITION NUMBER TO CONSTRUCT DESIGN POINTS FOR POLYNOMIAL REGRESSION MODELS*

JANE J. YE† AND JULIE ZHOU†

Abstract. In this paper we study a new optimality criterion, the K-optimality criterion, for constructing optimal experimental designs for polynomial regression models. We focus on the pth order polynomial regression model with symmetric design space $[-1, 1]$. For this model, we show that there is always a symmetric K-optimal design with exactly $p + 1$ support points including the boundary points -1 and 1. It is well known that the condition number for a positive definite matrix as the ratio of the maximum eigenvalue to the minimum eigenvalue is usually nonsmooth. We show that for our model, the condition number of the information matrix is continuously differentiable. Theoretical K-optimal designs are derived for $p = 1$ and 2. Numerical results are presented for $3 \le p \le 10$.

Key words. condition number, information matrix, K-optimal design, polynomial regression model, semidefinite programming, symmetric design

AMS subject classifications. 62K05, 90C30, 65F35

DOI. 10.1137/110850268

1. Introduction. Many practical and theoretical problems in science and engineering consider the relationship between a response variable and a predictor as a pth order polynomial regression model:

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots + \theta_p x^p + \epsilon,$$

where y is the response variable observed at design point $x \in [-1, 1]$, $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)^T$ is the unknown regression parameter vector, and ϵ is a random error with mean 0 and variance σ^2. The functions $1, x, x^2, \ldots, x^p$ are referred to as regressors.

Suppose that $x_1, \ldots, x_n$ are distinctive design points (or support points) taken from the design space $\mathcal{U} := [-1, 1]$, and r_i independent experiment runs have been carried out at each design point $x_i, i = 1, \ldots, n$. Suppose that observations $y_{ij}, j = 1 \ldots, r_i$ were made at the design point x_i. The collection of design points and their probabilities,

$$\xi_N = \left\{ \begin{array}{c} x_1, \ldots, x_n \\ p_1, \ldots, p_n \end{array} \right\},$$

where $p_i = r_i/N$ and $N = \sum_{i=1}^n r_i$, is called a design (of the experiment) with sample size N.

Based on the observations

$$y_{11}, \ldots, y_{1r_1}, \ldots, y_{n1}, \ldots, y_{nr_n},$$

the least squares estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is defined to be

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \sum_{j=1}^{r_i} (y_{ij} - \boldsymbol{\theta}^T \mathbf{f}(x_i))^2,$$

*Received by the editors October 4, 2011; accepted for publication (in revised form) January 29, 2013; published electronically March 28, 2013. This research was partially supported by NSERC.
http://www.siam.org/journals/siopt/23-1/85026.html

†Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8W 3R4, Canada (janeye@uvic.ca, jzhou@uvic.ca).

where $\mathbf{f}(x)^T = (1, x, \ldots, x^p)$. Taking the derivative with respect to variable $\boldsymbol{\theta}$ and setting it to zero gives the normal equations

$$\mathbf{A}(\xi_N)\hat{\boldsymbol{\theta}} = \mathbf{Y}, \tag{1}$$

where

$$\mathbf{A}(\xi_N) := \sum_{i=1}^{n} p_i \mathbf{f}(x_i)\mathbf{f}(x_i)^T, \qquad \mathbf{Y} := \sum_{i=1}^{n} p_i \bar{y}_i \mathbf{f}(x_i), \qquad \bar{y}_i := \frac{1}{r_i}\sum_{j=1}^{r_i} y_{ij}.$$

The matrix $\mathbf{A}(\xi_N)$ is usually called the "information matrix." If the inverse matrix $\mathbf{A}^{-1}(\xi_N)$ exists, then the normal equations (1) yield the unique solution

$$\hat{\boldsymbol{\theta}} = \mathbf{A}^{-1}(\xi_N)\mathbf{Y},$$

which gives an explicit expression for the least squares estimator.

Assume that the random errors are uncorrelated. By using the calculation rules for expected values and variance from probability theory, it is easy to see that the least squares estimator $\hat{\boldsymbol{\theta}}$ is unbiased and has covariance matrix

$$Cov(\hat{\boldsymbol{\theta}}) = \frac{\sigma^2}{N}\mathbf{A}^{-1}(\xi_N).$$

Optimal regression design problems in statistics are to find designs such that some scalar function of the covariance matrix of the least squares estimator is minimized. Several such minimization criteria have been studied in the optimal design literature; see, e.g., [8, 9, 22, 24]. For example, D-optimal designs minimize the determinant of the matrix, $\det(Cov(\hat{\boldsymbol{\theta}}))$ and A-optimal designs minimize the trace of the matrix, $\mathrm{trace}(Cov(\hat{\boldsymbol{\theta}}))$. In essence, D-optimal designs minimize the volumes of confidence regions of $\boldsymbol{\theta}$ and A-optimal designs minimize the average of the variances of $\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_p$.

In this paper, we use an alternative design criterion to construct regression designs based on the condition number of the information matrix. Such an optimal design criterion was proposed and the motivation of studying this criterion was discussed in [13, 14] but not much work was done. The condition number of a matrix is a fundamental quantity in the perturbation theory of finite dimensional linear systems; it measures the sensitivity of a solution to changes in data. Suppose that we wish to solve a linear system $\mathbf{Ax} = \mathbf{b}$, and due to the error we can only have a vector $\mathbf{x}'$ and a vector $\mathbf{b}'$ for which $\mathbf{Ax}' = \mathbf{b}'$. Then the relative error in taking $\mathbf{x}'$ in lieu of $\mathbf{x}$ can be estimated by

$$\frac{\|\mathbf{x} - \mathbf{x}'\|}{\|\mathbf{x}\|} \le \kappa(\mathbf{A})\frac{\|\mathbf{b} - \mathbf{b}'\|}{\|\mathbf{b}\|},$$

where $\kappa(\mathbf{A})$ denotes the condition number of matrix $\mathbf{A}$. For a given design ξ_N, the least squares estimator $\hat{\boldsymbol{\theta}}$ is a solution of the normal equations (1). Therefore the condition number of the information matrix $\mathbf{A}(\xi_N)$ represents the maximum amount by which a perturbation in an experimental measurement y_{ij} will be transmitted to the unknown regression parameter vector $\boldsymbol{\theta}$. In order to minimize the error sensitivity, it is desirable to find a design ξ_N which minimizes the condition number of the information matrix $\kappa(\mathbf{A}(\xi_N))$. In regression analysis, multicollinearity refers to a situation in which two or more regressors are highly linearly related. Multicollinearity is a serious problem in regression analysis. The effects of multicollinearity include (1) the estimated regression coefficients tend to have large sampling variability, which implies that the estimated regression coefficients tend to vary widely from one sample to another; (2) the estimated individual regression coefficients may not be statistically significant even though a definite statistical relation exists; (3) the interpretation of regression coefficients is often altered. The condition number is a good measure of multicollinearity. A large condition number indicates severe multicollinearity (see, e.g., [19]). Thus it is desirable to minimize the condition number to choose the levels of regressors to observe the response variable, when we design experiments.

Since the numbers of the experiment runs $r_i, i = 1, \ldots, n$, in a design ξ_N are positive integers, experimental designs for finite sample size lead to, often intractable, integer optimization problems. Some of these difficulties could be avoided if one considers the continuous (or asymptotic) optimal designs to be defined below. Let ξ be any probability measure on $\mathcal{U}$ and we consider a design as a probability measure ξ. Given a design ξ, define

$$\mathbf{A}(\xi) = \int_{\mathcal{U}} \mathbf{f}(x)\mathbf{f}^T(x) d\xi(x)$$

as the information matrix of ξ. If the probability measure is a discrete one concentrating on a finite number of support points $x_1, \ldots, x_n \in \mathcal{U}$ with positive probabilities $p_1, \ldots, p_n$, then the information matrix is reduced to

$$\mathbf{A}(\xi) = \sum_{i=1}^{n} p_i \mathbf{f}(x_i)\mathbf{f}^T(x_i). \tag{2}$$

Let $\mu_j(\xi)$ be the jth moment of ξ, i.e.,

$$\mu_j(\xi) := \int_{\mathcal{U}} x^j d\xi(x).$$

Using the moments, the information matrix $\mathbf{A}(\xi)$ can be written as

$$\mathbf{A}(\xi) = (\ \mu_{i+j-2}(\xi)\ )_{(p+1)\times(p+1)} = \begin{pmatrix} 1 & \mu_1(\xi) & \cdots & \mu_p(\xi) \\ \mu_1(\xi) & \mu_2(\xi) & \cdots & \mu_{p+1}(\xi) \\ \vdots & \vdots & & \vdots \\ \mu_p(\xi) & \mu_{p+1}(\xi) & \cdots & \mu_{2p}(\xi) \end{pmatrix}. \tag{3}$$

In this paper we consider the optimization problem of finding an optimal design among all designs such that the condition number is minimized. We call such a design a K-optimal design. We will show that the K-optimal design can be chosen to be symmetric and have exactly $p+1$ support points which include the boundary points -1 and 1. Such a K-optimal design ξ can be used to approximate the exact design ξ_N and the approximation improves as the sample size N increases.

It is obvious that an information matrix is symmetric and positive semidefinite. One can then define the condition number of the information matrix (see, e.g., [12]) as

$$\kappa(\mathbf{A}) = \begin{cases} \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})} & \text{if } \lambda_{\min}(\mathbf{A}) > 0, \\ \infty & \text{if } \lambda_{\min}(\mathbf{A}) = 0, \end{cases}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and the smallest eigenvalues, respectively. Although not very systematically, the topics of minimizing condition number have been studied in the literature; see [4, 5, 11, 17, 18, 20] with various models. It is well known that the condition number of a positive definite matrix may be nonsmooth when either the largest or the smallest eigenvalue is not simple. It is a quasi-convex and pseudoconvex function of the matrix variables but not convex (see [18]). Hence the problem of minimizing the condition number is usually a nonsmooth and nonconvex optimization problem and nonsmooth optimization techniques are usually needed to solve the problem (see, e.g., [5]). Therefore computationally it is a hard optimization problem. In the case of minimizing the condition number over a convex subset of symmetric positive definite matrices, it is possible to approximate the problem by a sequence of convex optimization problems (see [18]), or to transform a linearly and positive homogeneously parametrized condition number optimization problem with a positive semidefinite representable constraint set to a convex optimization problem in the semidefinite programming framework (see [2, 3, 17]). Such convexifying techniques may be used in our model to find the best moments since the information matrix $\mathbf{A}(\xi)$ depends linearly and positive homogeneously on the moments $\mu := (\mu_1, \mu_2, \ldots, \mu_{2p})$. However, it can not be used to find the optimal design points x_i and corresponding probabilities p_i directly since the information matrix is nonlinearly parametrized in variables x_i and p_i.

In this paper, we show that the condition number of the information matrix for the polynomial regression model with the interval $[-1, 1]$ as the design space is smooth provided the number of support points is larger than or equal to $p + 1$. Consequently we can reduce our K-optimal design problem to a smooth optimization problem with p variables and one linear equality constraint.

Note that numerically there are a few alternative ways to find the least squares estimator. Suppose n independent experimental runs have been carried out and $x_1, \ldots, x_n$ are the design points (which may or may not be distinct). Based on the observations $y_1, \ldots, y_n$ the least squares estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is defined to be

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} (y_i - \boldsymbol{\theta}^T \mathbf{f}(x_i))^2 = \arg\min_{\boldsymbol{\theta}} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})$$

where

$$\mathbf{X} := \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{pmatrix}, \quad \mathbf{y} := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}. \tag{4}$$

Taking the derivative with respect to variable $\boldsymbol{\theta}$ and setting it to zero gives the normal equations

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\theta} = \mathbf{X}\mathbf{y}, \tag{5}$$

or equivalently

$$\mathbf{A}\boldsymbol{\theta} = \frac{1}{n} \mathbf{X}\mathbf{y},$$

where the information matrix $\mathbf{A} := \frac{1}{n}\mathbf{X}^T\mathbf{X}$. Instead of forming the matrix $\mathbf{A}$ and solving the normal equations, one can use the QR factorization or the singular value decomposition (SVD) to find the least squares estimator. Let $\|\cdot\|$ denote the Euclidean vector norm and matrix norm. The Euclidean condition number of the retangular matrix $\mathbf{X}$ is defined by [10]

$$\kappa(\mathbf{X}) = \max_{y \neq 0} \frac{\|y\|}{\|\mathbf{X}y\|} \max_{z \neq 0} \frac{\|\mathbf{X}z\|}{\|z\|} = \|\mathbf{X}\|\|\mathbf{X}^\dagger\| = \sqrt{\kappa(\mathbf{A})} = \sqrt{\frac{\lambda_1(\mathbf{A})}{\lambda_{p+1}(\mathbf{A})}},$$

where $\mathbf{X}^\dagger = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is the Moore–Penrose generalized inverse of $\mathbf{X}$. According to the above relationship between the condition numbers of $\mathbf{X}$ and $\mathbf{A}$, although it may be numerically more attractive to find the least squares estimator by using QR factorization or SVD of X, minimizing the condition number of $\mathbf{X}$ is the same as minimizing the condition number of $\mathbf{A}$. Thus, for the rest of the paper, we will minimize $\kappa(\mathbf{A})$ to find K-optimal designs, where $\mathbf{A}$ is defined as in (3).

The rest of the paper is organized as follows. In section 2, the symmetry of the K-optimal designs for polynomial regression models is investigated, and in section 3, K-optimal designs are derived analytically for $p = 1$ and 2. In section 4, we show that it is always possible to find discrete symmetric K-optimal designs. In section 5 we discuss the smoothness of the condition number. In section 6, we show that a symmetric K-optimal design can be chosen to have exactly $p + 1$ support points including the boundary points -1 and 1. In section 7, numerical methods to find symmetric K-optimal designs are discussed and the results are presented for $3 \leq p \leq 10$. Section 8 contains the concluding remarks.

2. Symmetry of the *K*-optimal designs. To study the symmetry of K-optimal designs, we define two probability distribution functions, $\xi_1(x)$ and $\xi_t(x)$ ($0 \leq t \leq 1$), as follows. Distribution function $\xi_1(x)$ is an image distribution of $\xi(x)$: for a discrete distribution $\xi(x)$ with support points $x_1, \ldots, x_n$ and probabilities $p_1, \ldots, p_n$, $\xi_1(x)$ is defined to have support points $-x_1, \ldots, -x_n$ and probabilities $p_1, \ldots, p_n$; for a continuous distribution $\xi(x)$ with density function $v(x)$, $\xi_1(x)$ is defined to have density function $v_1(x) = v(-x)$. If the two distributions $\xi(x)$ and $\xi_1(x)$ are the same, then $\xi(x)$ is a symmetric distribution on $[-1, 1]$. Distribution function $\xi_t(x)$ is a convex combination of $\xi(x)$ and $\xi_1(x)$, i.e., $\xi_t(x) = (1 - t)\xi(x) + t\xi_1(x)$. It is easy to verify that $\xi_{0.5}(x)$ is a symmetric distribution on $[-1, 1]$.

Now the moments of distributions ξ, ξ_t, and ξ_1 have the following relationships:

$$\mu_j(\xi_1) = (-1)^j \mu_j(\xi), \quad \mu_j(\xi_t) = (1 - t)\mu_j(\xi) + t\mu_j(\xi_1), \quad j \geq 1.$$

Hence the information matrices satisfy

$$\mathbf{A}(\xi_1) = \mathbf{Q}\ \mathbf{A}(\xi)\ \mathbf{Q}, \tag{6}$$

$$\mathbf{A}(\xi_t) = (1 - t)\mathbf{A}(\xi) + t\mathbf{A}(\xi_1), \tag{7}$$

where matrix $\mathbf{Q}$ is a $(p + 1) \times (p + 1)$ diagonal matrix with diagonal elements

$$(-1)^0, (-1)^1, \ldots, (-1)^p.$$

The following theorem on the properties of the condition number is the key result to study the symmetry of K-optimal designs. For a symmetric matrix $\mathbf{A}$ of order $p + 1$, we use $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \cdots \geq \lambda_{p+1}(\mathbf{A})$ to denote the real eigenvalues of $\mathbf{A}$ being arranged in nonincreasing order.

THEOREM 2.1. *If matrix $\mathbf{A}(\xi)$ is positive definite, then we have*

(i) $\kappa(\mathbf{A}(\xi)) = \kappa(\mathbf{A}(\xi_1))$,

(ii) $\kappa(\mathbf{A}(\xi_t)) \leq \kappa(\mathbf{A}(\xi))$ *for all* $0 < t < 1$.

Proof. (i) Since $\mathbf{A}(\xi_1)$ is diagonally similar to $\mathbf{A}(\xi)$ by (6), the eigenvalues of $\mathbf{A}(\xi_1)$ and $\mathbf{A}(\xi)$ are the same. This implies that the largest and smallest eigenvalues are the same, i.e., $\lambda_1(\mathbf{A}(\xi_1)) = \lambda_1(\mathbf{A}(\xi))$ and $\lambda_{p+1}(\mathbf{A}(\xi_1)) = \lambda_{p+1}(\mathbf{A}(\xi)) > 0$, and the result follows.

(ii) From (7), $\mathbf{A}(\xi_t) = (1-t)\mathbf{A}(\xi) + t\mathbf{A}(\xi_1)$. Since the largest and the smallest eigenvalues are convex and concave, respectively, it follows that for all $0 < t < 1$,

$$\begin{aligned} \lambda_1(\mathbf{A}(\xi_t)) &\le (1-t)\lambda_1(\mathbf{A}(\xi)) + t\lambda_1(\mathbf{A}(\xi_1)) = \lambda_1(\mathbf{A}(\xi)),\\ \lambda_{p+1}(\mathbf{A}(\xi_t)) &\ge (1-t)\lambda_{p+1}(\mathbf{A}(\xi)) + t\lambda_{p+1}(\mathbf{A}(\xi_1)) = \lambda_{p+1}(\mathbf{A}(\xi)) > 0. \end{aligned}$$

Thus

$$\kappa(\mathbf{A}(\xi_t)) = \frac{\lambda_1(\mathbf{A}(\xi_t))}{\lambda_{p+1}(\mathbf{A}(\xi_t))} \le \frac{\lambda_1(\mathbf{A}(\xi))}{\lambda_{p+1}(\mathbf{A}(\xi))} = \kappa(\mathbf{A}(\xi)). \qquad \square$$

From Theorem 2.1, the condition number of the information matrix for the symmetric distribution (design) $\xi_{0.5}$ is always less than or equal to the condition number of the information matrix for distribution ξ, which implies that we can just focus on symmetric designs to construct K-optimal designs. Thus, for the rest of the paper, we assume that ξ is symmetric. For a symmetric design, the odd moments are all zero, i.e., $\mu_j(\xi) = 0$ for odd j.

3. *K*-optimal designs for $p = 1$ and 2. In this section we derive the analytical solutions for K-optimal designs for polynomial regression models with $p = 1$ and 2 and compare the results with other classical optimal designs such as D-optimal and A-optimal designs. For simplicity, we use μ_j for $\mu_j(\xi)$.

For any distribution on $\mathcal{U}$, it is easy to see that the even moments satisfy

$$1 \ge \mu_2 \ge \mu_4 \ge \mu_6 \ge \cdots \ge 0. \tag{8}$$

For $p = 1$, the information matrix $\mathbf{A}(\xi) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \mu_2 \end{smallmatrix}\right)$ is diagonal, and the two eigenvalues are $\lambda_1(\mathbf{A}(\xi)) = 1$ and $\lambda_{p+1}(\mathbf{A}(\xi)) = \mu_2$. Hence minimizing the condition number $\kappa(\mathbf{A}(\xi)) = 1/\mu_2$ is equivalent to maximizing the second moment μ_2. Since $\mu_2 \le 1$ by (8), the maximum value of μ_2 is 1 and can be reached by this $\xi^{[1]}(x)$ having two support points $x_1 = +1$ and $x_2 = -1$ with $p_1 = p_2 = 0.5$. This K-optimal design $\xi^{[1]}(x)$ is also D-optimal and A-optimal for $p = 1$; see [22].

For $p = 2$, the information matrix is

$$\mathbf{A}(\xi) = \begin{pmatrix} 1 & 0 & \mu_2 \\ 0 & \mu_2 & 0 \\ \mu_2 & 0 & \mu_4 \end{pmatrix},$$

and the three eigenvalues are

$$\begin{aligned} \lambda_1(\mathbf{A}(\xi)) &= \frac{1}{2}\left(1 + \mu_4 + \sqrt{(1-\mu_4)^2 + 4\mu_2^2}\right),\\ \lambda_2(\mathbf{A}(\xi)) &= \mu_2,\\ \lambda_3(\mathbf{A}(\xi)) &= \frac{1}{2}\left(1 + \mu_4 - \sqrt{(1-\mu_4)^2 + 4\mu_2^2}\right). \end{aligned}$$

Notice that if (μ_2, μ_4) is a minimizer then $\mu_2 \ne 0$ since if $\mu_2 = 0$, then $\mu_4 = 0$ and

$\kappa(\mathbf{A}(\xi)) = \infty$, which is not a minimal condition number. Hence

$$\lambda_1(\mathbf{A}(\xi)) > \frac{1}{2}\left(1 + \mu_4 + \sqrt{(1-\mu_4)^2}\right) = 1,$$
$$\mu_4 \le \lambda_2(\mathbf{A}(\xi)) \le 1,$$
$$\lambda_3(\mathbf{A}(\xi)) < \frac{1}{2}\left(1 + \mu_4 - \sqrt{(1-\mu_4)^2}\right) = \mu_4.$$

Therefore $\lambda_1(\mathbf{A}(\xi)) > \lambda_2(\mathbf{A}(\xi)) > \lambda_3(\mathbf{A}(\xi))$, which implies that the condition number is

$$\kappa(\mathbf{A}(\xi)) = \frac{\lambda_1(\mathbf{A}(\xi))}{\lambda_3(\mathbf{A}(\xi))} = \frac{1 + \mu_4 + \sqrt{(1-\mu_4)^2 + 4\mu_2^2}}{1 + \mu_4 - \sqrt{(1-\mu_4)^2 + 4\mu_2^2}}. \tag{9}$$

Minimizing $\kappa(\mathbf{A}(\xi))$ over μ_2 and μ_4 gives the K-optimal design for $p = 2$ in the following theorem.

THEOREM 3.1. *The condition number in* (9) *is minimized by design* $\xi^{[2]}(x)$ *having three support points* $x_1 = -1$*,* $x_2 = 0$*, and* $x_3 = +1$ *with probabilities* $p_1 = 1/6$*,* $p_2 = 4/6$ *and* $p_3 = 1/6$*, respectively.*

Proof. Define two functions

$$g(\mu_2, \mu_4) = (1-\mu_4)^2 + 4\mu_2^2,$$
$$h(\mu_2, \mu_4) = \kappa(\mathbf{A}(\xi)) = \frac{1 + \mu_4 + \sqrt{g(\mu_2,\mu_4)}}{1 + \mu_4 - \sqrt{g(\mu_2,\mu_4)}}.$$

Taking the partial derivative with respect to μ_4 for function $h(\mu_2, \mu_4)$ gives

$$\frac{\partial h(\mu_2,\mu_4)}{\partial \mu_4} = \frac{-2g(\mu_2,\mu_4) - 2(1-\mu_4^2)}{\sqrt{g(\mu_2,\mu_4)}\left(1 + \mu_4 - \sqrt{g(\mu_2,\mu_4)}\right)^2} < 0,$$

since $0 \le \mu_4 \le \mu_2 \le 1$. Therefore for any fixed μ_2 the condition number is a decreasing function of μ_4. Thus for any fixed μ_2, the condition number is minimized at $\mu_4 = \mu_2$. Now we will find the value of μ_2 to minimize $h(\mu_2, \mu_2)$. Let $\varphi(\mu_2) := h(\mu_2, \mu_2)$. Then from

$$\varphi'(\mu_2) = \frac{12\mu_2 - 4}{\sqrt{g(\mu_2,\mu_2)}\left(1 + \mu_2 - \sqrt{g(\mu_2,\mu_2)}\right)^2} = 0, \tag{10}$$

we get $\mu_2 = 1/3$, and it is easy to check from (10) that $\mu_2 = 1/3$ minimizes $h(\mu_2, \mu_2)$. Therefore $\mu_2 = \mu_4 = 1/3$ minimizes the condition number $\kappa(\mathbf{A}(\xi))$. For $\xi^{[2]}(x)$ having three support points $x_1 = -1$, $x_2 = 0$, and $x_3 = +1$ with probabilities $p_1 = 1/6$, $p_2 = 4/6$, and $p_3 = 1/6$, respectively, it is easy to verify that $\mu_2(\xi^{[2]}) = \mu_4(\xi^{[2]}) = 1/3$. Thus $\xi^{[2]}(x)$ minimizes the condition number. $\square$

The three support points of the K-optimal design in Theorem 3.1 are the same as those of D-optimal and A-optimal designs [22], but the corresponding probabilities are different. The D-optimal design has $p_1 = p_2 = p_3 = 1/3$, while the A-optimal design has $p_1 = 1/4$, $p_2 = 1/2$, and $p_3 = 1/4$. We compare $\kappa(\mathbf{A})$, $\det(\mathbf{A})$, and $\mathrm{trace}(\mathbf{A}^{-1})$ in Table 1. Numerical results are rounded to 3 decimal point. Note that the D-optimal design maximizes $\det(\mathbf{A})$ (equivalently minimizing $\det(\mathbf{A}^{-1})$) and the

TABLE 1
Optimal design criteria.

Types of optimal designs	$\kappa(\mathbf{A})$	$det(\mathbf{A})$	trace$(\mathbf{A}^{-1})$
K-optimal design	5.828	0.074	9
D-optimal design	10.404	0.148	9
A-optimal design	6.854	0.125	8

A-optimal design minimizes trace$(\mathbf{A}^{-1})$. The numerical results are consistent with the analytical results.

4. Existence of discrete *K*-optimal designs. For the case of $p = 1, 2$ in the previous section, one can find discrete symmetric K-optimal designs. Is this true for the general case? The answer is given in the following theorem.

THEOREM 4.1. *One can always find a discrete symmetric K-optimal design which minimizes the condition number of the information matrix among all probability measures. Moreover the number of support points is between $p+1$ and $2p+2$.*

Proof. Step 1. First we prove the existence of a K-optimal design. Since the design space $\mathcal{U} = [-1, 1]$ is compact, by the Prohorov theorem [1], the class of all probability measures on $\mathcal{U}$ equipped with weak topology is compact. We denote it by Θ. Since the map $\xi \to \mathbf{A}(\xi)$ is linear and bounded and hence continuous on Θ and the condition number $\kappa(\mathbf{A})$ is lower semicontinuous on the set of all positive semidefinite matrices, the map $\xi \to \kappa(\mathbf{A}(\xi))$ is lower semicontinuous on Θ. Therefore by the Weierstrass theorem, there exists a K-optimal design. Second, we show that the number of support points for a discrete K-optimal design must be at least $p+1$. Indeed from the definition of a condition number, the condition number for a degenerate semidefinite matrix is positive infinity. Hence for ξ to be a K-optimal design, the matrix $\mathbf{A}(\xi)$ must be positive definite and consequently, the rank of the matrix $\mathbf{A}(\xi)$ must be equal to $p+1$. By (2), since for a discrete design the matrix $\mathbf{A}(\xi)$ is a sum of rank one matrices of size $p+1$, the number of support points must be at least $p+1$ in order for the rank to be equal to $p+1$.

Step 2. We prove the existence of a discrete symmetric K-optimal design with the number of support points not more than $2p+2$. Define a $(p+1) \times (p+1)$ matrix,

$$\mathbf{W}(x) = \begin{pmatrix} 1 & 0 & x^2 & 0 & \cdots & & \\ 0 & x^2 & 0 & \cdots & & & \\ x^2 & 0 & \cdots & & & & \\ 0 & \cdots & & & & & \\ \vdots & & & & & & 0 \\ & & & & & 0 & x^{2p} \end{pmatrix}. \tag{11}$$

Let ξ denote a symmetric design. Then by the definition of a probability measure and moments, it is easy to see that

$$\begin{aligned} \mathbf{A}(\xi) &\in \overline{co}\,\{W(x) : -1 \le x \le 1\} \\ &= co\,\{W(x) : -1 \le x \le 1\}, \end{aligned}$$

where $co\{S\}$ denotes the convex hull of set S and $\overline{A}$ denotes the closure of the set A, and the second equality holds because the set

$$S = \{W(x) : -1 \le x \le 1\}$$

is compact.

On the other hand, for any symmetric design ξ, the patterns in the matrix $\mathbf{A}(\xi)$ show that any matrix $\mathbf{A}(\xi)$ is completely described by p elements. Consequently, the set S can be considered as a subset of a p dimensional space. By Caratheodory's theorem (see, e.g., [8, Theorem 2.1.1]), any $\mathbf{A}(\xi)$ can be represented by a convex combination of $p+1$ elements in the set S and hence there exist $p_i \geq 0, i = 1, \ldots, p+1$, satisfying $\sum_{i=1}^{p+1} p_i = 1$ and $x_i, i = 1, \ldots, p+1$, such that

$$\mathbf{A}(\xi) = \sum_{i=1}^{p+1} p_i W(x_i). \tag{12}$$

In particular if ξ is a discrete K-optimal design, then by Step 1, the number of support points must be at least $p+1$ and hence all points $x_1, \ldots, x_{p+1}$ must be distinct and all $p_i, i = 1, \ldots, p+1$, must be positive in the above representation.

Let ξ be a discrete K-optimal design. Suppose that the $p+1$ distinctive points $x_1, \ldots, x_{p+1}$ and the positive numbers $p_1, \ldots, p_{p+1}$ selected by using Caratheodory's theorem are symmetric, i.e.,

$$\xi = \left\{ \begin{array}{l} x_1, x_2, \ldots, x_m, x_{m+1}, -x_m, \ldots, -x_2, -x_1 \\ p_1, p_2, \ldots, p_m, p_{m+1}, \quad p_m, \ldots, \quad p_2, \quad p_1 \end{array} \right\},$$

where $x_i \in [-1, 0)$, $p_i \in (0, 1]$, $i = 1, \ldots, m$, $m = \lfloor \frac{p+1}{2} \rfloor$ (integer part of $(p+1)/2$), $x_{m+1} = 0$, and $2\sum_{i=1}^{m} p_i + p_{m+1} = 1$. For even p, $p_{m+1} > 0$, while for odd p, $p_{m+1} = 0$. Then ξ is a symmetric K-optimal design with $p+1$ support points. Now suppose that the $p+1$ points $x_1, \ldots, x_{p+1}$ and the positive numbers $p_1, \ldots, p_{p+1}$ selected by using Caratheodory's theorem are not symmetric. Then by section 2, one can always find a symmetric K-optimal design by taking $\xi_{0.5}$. It is obvious that $\xi_{0.5}$ is a discrete symmetric K-optimal design with at most $2p+2$ support points. $\square$

5. Smoothness of the condition number. In this section we aim at showing that the condition number of the information matrix is continuously differentiable at $(\mu_2, \mu_4, \ldots, \mu_{2p})$ which corresponds to a symmetric design with at least $p+1$ design points. As a result, the condition number of the information matrix as a composition of two smooth mappings is a smooth function of x_i and p_i in the feasible region of our optimization problem (P) to be studied numerically in section 7.

Consider the following information matrix for a symmetric design ξ:

$$\mathbf{A}_{p+1} := \mathbf{A}_{p+1}(\xi) = \begin{pmatrix} 1 & 0 & \mu_2 & \cdots & \\ 0 & \mu_2 & 0 & \cdots & \\ \mu_2 & 0 & \mu_4 & \cdots & \\ \vdots & \vdots & & \ddots & \\ & & & & \mu_{2p} \end{pmatrix}_{(p+1)\times(p+1)},$$

where $\mu_j = \int_{\mathcal{U}} x^j d\xi(x), j = 2, 4, \ldots, 2p$ are the even moments. By exchanging rows and columns of $\mathbf{A}_{p+1}$ at the same time, we obtain a matrix $\mathbf{B}_{p+1}$ which is a direct sum of two matrices:

$$\mathbf{B}_{p+1} := \mathbf{C}_k \oplus \mathbf{D}_m = \begin{pmatrix} \mathbf{C}_k & 0 \\ 0 & \mathbf{D}_m \end{pmatrix},$$

where

$$\mathbf{C}_k := \left(\mu_{2(i+j-2)}\right)_{k\times k} = \begin{pmatrix} 1 & \mu_2 & \cdots & \mu_{2(k-1)} \\ \mu_2 & \mu_4 & \cdots & \mu_{2k} \\ \vdots & \vdots & & \vdots \\ \mu_{2(k-1)} & \mu_{2k} & \cdots & \mu_{4(k-1)} \end{pmatrix}_{k\times k},$$

$$\mathbf{D}_m := \left(\mu_{2(i+j-1)}\right)_{m\times m} = \begin{pmatrix} \mu_2 & \mu_4 & \cdots & \mu_{2m} \\ \mu_4 & \mu_6 & \cdots & \mu_{2m+2} \\ \vdots & \vdots & & \vdots \\ \mu_{2m} & \mu_{2m+2} & \cdots & \mu_{4m-2} \end{pmatrix}_{m\times m},$$

and $m = \lfloor \frac{p+1}{2} \rfloor$, $k := p+1-m$. If p is odd, then $k = m = \frac{p+1}{2}$. If p is even, then $k = m+1$. It is easy to verify that

$$\mathbf{B}_{p+1} = \mathbf{Q}_{p+1}^T \mathbf{A}_{p+1} \mathbf{Q}_{p+1},$$

where matrix

$$\mathbf{Q}_{p+1} = (\mathbf{e}_1, \mathbf{e}_3, \ldots, \mathbf{e}_{(2k-1)}, \mathbf{e}_2, \mathbf{e}_4, \ldots, \mathbf{e}_{2m}),$$

and $\mathbf{e}_i$ is a $(p+1)\times 1$ unit vector with the ith element being 1. Thus we have the following result for the eigenvalues of $\mathbf{A}_{p+1}$ and $\mathbf{B}_{p+1}$.

PROPOSITION 5.1. *Matrices $\mathbf{A}_{p+1}$ and $\mathbf{B}_{p+1}$ have the same eigenvalues, i.e.,*

$$\lambda_i(\mathbf{A}_{p+1}) = \lambda_i(\mathbf{B}_{p+1}), \qquad i = 1, \ldots, p+1.$$

We say that a square matrix is strictly totally positive if all its minors are positive. We say that a matrix $\mathbf{A}$ is a Hankel matrix if it can be represented as $\mathbf{A} = (b_{i+j})_{i,j=0}^n$ for real numbers $\{b_0, \ldots, b_{2n}\}$. The following result characterizes the strictly total positivity of a Hankel matrix.

LEMMA 5.1 (see [21, Theorem 4.4]). *A Hankel matrix $\mathbf{A}$ is strictly totally positive if and only if matrix $\mathbf{A}$ and its submatrix $\mathbf{B}$ obtained by deleting the first column and the last row are both positive definite.*

PROPOSITION 5.2. *Let ξ be any given symmetric design with n support points and*

$$\{1, \mu_2, \ldots, \mu_{2p}\}$$

be its even moments. If $n \geq p+1$, then the matrices $\mathbf{C}_k$ and $\mathbf{D}_m$ are strictly totally positive, where $m = \lfloor \frac{p+1}{2} \rfloor$ and $k = p+1-m$.

Proof. Since $\xi(x)$ is a symmetric design with n support points,

$$\mathbf{C}_k = \sum_{i=1}^n p_i \mathbf{x}_i \mathbf{x}_i^T, \tag{13}$$

where $\mathbf{x}_i = (1, x_i^2, x_i^4, \ldots, x_i^{2(k-1)})^T, p_i > 0, \sum_{i=1}^n p_i = 1$. Since $\mathbf{x}_i\mathbf{x}_i^T, i = 1, \ldots, n$, are positive semidefinite matrices and p_i are positive, $\mathbf{C}_k$ must be positive semidefinite. In fact we can show that the matrix $\mathbf{C}_k$ has full rank k and hence it is positive definite. Indeed by the symmetry of the design ξ, there are $\lfloor \frac{n+1}{2} \rfloor$ distinctive points x_i^2 in $[-1,1]$. Since $n \geq p+1$, we have

$$\left\lfloor \frac{n+1}{2} \right\rfloor \geq \left\lfloor \frac{p+2}{2} \right\rfloor = \left\lfloor \frac{p}{2} \right\rfloor + 1 = k.$$

That is, there are no less than k distinctive points x_i^2 in $[-1,1]$. Notice that $\mathbf{x}_i\mathbf{x}_i^T, i = 1,\ldots,n$ are rank one matrices and $p_i's$ are positive. Consequently, $p_i\mathbf{x}_i\mathbf{x}_i^T, i = 1,\ldots,n$ are rank one matrices. Hence by expression (13), matrix $\mathbf{C}_k$ is a sum of no less than k distinctive rank one matrices. Therefore the rank of the matrix is equal to k. We now show that the matrix $\mathbf{D}_{k-1}$ is also positive definite. Since $\xi(x)$ is a symmetric design with n support points, it is easy to verify that

$$\mathbf{D}_{k-1} = \sum_{i=1}^{n} p_i x_i^2 \mathbf{x}_i \mathbf{x}_i^T, \tag{14}$$

where $\mathbf{x}_i = (1, x_i^2, x_i^4, \ldots, x_i^{2(k-2)})^T, p_i > 0, \sum_{i=1}^n p_i = 1$. Since $\mathbf{x}_i\mathbf{x}_i^T, i = 1,\ldots,n$ are positive semidefinite matrices and p_i are positive, $\mathbf{D}_{k-1}$ must be positive semidefinite. In fact we can show that the matrix $\mathbf{D}_{k-1}$ has full rank $k-1$ and hence is positive definite. Indeed by the symmetry of the design ξ, there are at least $\lfloor \frac{n}{2} \rfloor$ nonzero distinctive points x_i^2 in $[-1,1]$. Since $n \geq p+1$,

$$\left\lfloor \frac{n}{2} \right\rfloor \geq \left\lfloor \frac{p+1}{2} \right\rfloor = m \geq k-1.$$

Notice that $\mathbf{x}_i\mathbf{x}_i^T, i = 1,\ldots,n$ are rank one matrices and $p_i's$ are positive. Consequently, $p_i x_i^2 \mathbf{x}_i\mathbf{x}_i^T$ are rank one matrices for those i such that $x_i^2 > 0$. Hence by expression (14), matrix $\mathbf{D}_{k-1}$ is a sum of no less than $k-1$ distinctive rank one matrices. Therefore the rank of the matrix $\mathbf{D}_{k-1}$ is equal to $k-1$ and hence the matrix $\mathbf{D}_{k-1}$ is positive definite. $\mathbf{D}_{k-1}$ is the submatrix of $\mathbf{C}_k$ deleting the first column and the last row. Since both matrices $\mathbf{C}_k$ and $\mathbf{D}_{k-1}$ are positive definite, by Lemma 5.1 $\mathbf{C}_k$ is strictly totally positive.

It remains to show that the matrix $\mathbf{D}_m$ is strictly totally positive. Let

$$\mathbf{E}_{m-1} := \left(\mu_{2(i+j)}\right)_{(m-1)\times(m-1)} = \begin{pmatrix} \mu_4 & \mu_6 & \cdots & \mu_{2m} \\ \mu_6 & \mu_8 & \cdots & \mu_{2(m+1)} \\ \vdots & \vdots & & \vdots \\ \mu_{2m} & \mu_{2(m+2)} & \cdots & \mu_{4m-4} \end{pmatrix}_{(m-1)\times(m-1)}.$$

When p is even, $k = m+1$ and when p is odd, $k = m$. We have shown that when p is even, $\mathbf{C}_{m+1}$ is positive definite and when p is odd, $\mathbf{C}_m$ is positive definite. Since $\mathbf{E}_{m-1}$ is the principal submatrix of $\mathbf{C}_{m+1}$ deleting the first and the last rows and columns and is the principal submatrix of $\mathbf{C}_m$ deleting the first row and the first column, it is positive definite. Since $\mathbf{E}_{m-1}$ is the submatrix of $\mathbf{D}_m$ deleting the first column and the last row and both matrices $\mathbf{D}_m$ and $\mathbf{E}_{m-1}$ are positive definite, $\mathbf{D}_m$ is strictly totally positive by Lemma 5.1. $\square$

PROPOSITION 5.3. *Let n be the number of support points of the symmetric design measure ξ. If $n \geq p+1$, then we have*

$$\lambda_1(\mathbf{A}_{p+1}) = \lambda_1(\mathbf{C}_k),$$
$$\lambda_{p+1}(\mathbf{A}_{p+1}) = \begin{cases} \lambda_m(\mathbf{D}_m) & \text{if } p \text{ is odd,} \\ \lambda_{m+1}(\mathbf{C}_{m+1}) & \text{if } p \text{ is even.} \end{cases}$$

Proof. It is clear that

$$\lambda_1(\mathbf{A}_{p+1}) = \lambda_1(\mathbf{B}_{p+1}) = \max(\lambda_1(\mathbf{C}_k), \lambda_1(\mathbf{D}_m)),$$
$$\lambda_{p+1}(\mathbf{A}_{p+1}) = \lambda_{p+1}(\mathbf{B}_{p+1}) = \min(\lambda_k(\mathbf{C}_k), \lambda_m(\mathbf{D}_m)).$$

By Proposition 5.2, both $\mathbf{C}_k$ and $\mathbf{D}_m$ are positive definite.

Since $\mathbf{C}_m - \mathbf{D}_m = \sum_{i=1}^n p_i(1-x_i^2)\mathbf{x}_i\mathbf{x}_i^T$ with $\mathbf{x}_i = (1, x_i^2, \ldots, x_i^{2(m-1)})^T$, it is clear that $\mathbf{C}_m - \mathbf{D}_m$ is positive semidefinite. Consequently by [12, Corollary 7.7.4] we have

$$\begin{aligned} &\lambda_1(\mathbf{C}_m) \geq \lambda_1(\mathbf{D}_m), \\ &\lambda_m(\mathbf{C}_m) \geq \lambda_m(\mathbf{D}_m). \end{aligned} \tag{15}$$

Moreover since $k \geq m$, by the interlacing theorem [12, Theorem 4.3.8] we have

$$\lambda_1(\mathbf{C}_k) \geq \lambda_1(\mathbf{C}_m),$$

which implies that $\lambda_1(\mathbf{A}_{p+1}) = \lambda_1(\mathbf{C}_k)$.

When p is an odd number, $k = m = \frac{p+1}{2}$. Then we have

$$\lambda_{p+1}(\mathbf{A}_{p+1}) = \lambda_{p+1}(\mathbf{B}_{p+1}) = \min(\lambda_m(\mathbf{C}_m), \lambda_m(\mathbf{D}_m)) = \lambda_m(\mathbf{D}_m),$$

where the third equation follows by (15).

When p is an even number, $k = m + 1$, and we have $\mathbf{B}_{p+1} = \mathbf{C}_{m+1} \oplus \mathbf{D}_m$. Since $\mathbf{E}_m$ is the principle submatrix of $\mathbf{C}_{m+1}$ by deleting the first row and the first column,

$$\lambda_{m+1}(\mathbf{C}_{m+1}) \leq \lambda_m(\mathbf{E}_m),$$

by the interlacing theorem [12, Theorem 4.3.8]. In addition, it is easy to verify that $\mathbf{D}_m - \mathbf{E}_m$ is positive semidefinite, which implies

$$\lambda_m(\mathbf{D}_m) \geq \lambda_m(\mathbf{E}_m).$$

Thus $\lambda_{p+1}(\mathbf{A}_{p+1}) = \lambda_{p+1}(\mathbf{B}_{p+1}) = \min\{\lambda_{m+1}(\mathbf{C}_{m+1}), \lambda_m(\mathbf{D}_m)\} = \lambda_{m+1}(\mathbf{C}_{m+1})$. ∎

We now prove the main result of this section.

THEOREM 5.1. *The largest eigenvalue and the smallest eigenvalue of $\mathbf{A}_{p+1}$ are smooth at $(\mu_2, \ldots, \mu_{2p})$ which correspond to a symmetric design ξ with at least $p+1$ support points. Consequently the condition number $\kappa(\mathbf{A}_{p+1})$ is smooth at $(\mu_2, \ldots, \mu_{2p})$ which corresponds to a symmetric design ξ with at least $p+1$ support points.*

Proof. Let $\mu_2, \ldots, \mu_{2p}$ correspond to a symmetric design ξ with at least $p+1$ support points. Since $\mathbf{C}_k$ and $\mathbf{D}_m$ are totally positive matrices by Proposition 5.2, all eigenvalues are positive and simple. Therefore $\lambda_1(\mathbf{C}_k)$, $\lambda_k(\mathbf{C}_k)$, and $\lambda_m(\mathbf{D}_m)$ are continuously differentiable at $(\mu_2, \ldots, \mu_{2p})$. Consequently by Proposition 5.3, the largest eigenvalue and the smallest eigenvalue of $\mathbf{A}_{p+1}$ are continuously differentiable at $(\mu_2, \ldots, \mu_{2p})$ as well, which implies that the condition number $\kappa(\mathbf{A}_{p+1})$ is continuously differentiable at $(\mu_2, \ldots, \mu_{2p})$. ∎

6. Support points of *K*-optimal designs. For the cases of $p = 1$ and 2 in section 3, one can find symmetric K-optimal designs that have exactly $p+1$ support points and the boundary points -1 and $+1$ are included. Is this true for the general case? The positive answer is given in this section. First we need the following technical result.

LEMMA 6.1. *Suppose that the number of support points $n \geq p+1$. Then $\kappa(\mathbf{A}_{p+1})$ is a nonincreasing function of μ_{2p} when all other moments are fixed.*

Proof. *Case* 1: p is odd. In this case since $k = m = \frac{p+1}{2}$, we have $\mathbf{B}_{p+1} = \mathbf{C}_m \oplus \mathbf{D}_m$. By Proposition 5.3 we have

$$\kappa(\mathbf{A}_{p+1}) = \frac{\lambda_1(\mathbf{C}_m)}{\lambda_m(\mathbf{D}_m)}$$

and μ_{2p} is only in matrix $\mathbf{D}_m$. By Proposition 5.2, $\mathbf{D}_m$ is a strictly total positive matrix and hence all eigenvalues are simple and positive. Consequently the function $\lambda_m(\mathbf{D}_m)$ is a smooth function of $\mathbf{D}_m$ and hence a smooth function of variable μ_{2p}. Denote by $\lambda_{\min}(\mu_{2p})$ the smallest eigenvalue of $\mathbf{D}_m$ with all moments other than μ_{2p} fixed. Since by the Rayleigh–Ritz theorem [12, Theorem 4.22],

$$\lambda_{\min}(\mu_{2p}) = \lambda_m(\mathbf{D}_m) = \min_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{D}_m \mathbf{x}$$

and the minimum is achieved by the unit eigenvector for $\lambda_m(\mathbf{D}_m)$, by the Danskin's theorem stated on p. 99 of [6] (also see [7]), we have

$$\lambda'_{\min}(\mu_{2p}) = (x_m^{\min})^2,$$

where $x_m^{\min}$ denotes the last component of the unit eigenvector $\mathbf{x}^{\min}$ for the smallest eigenvalue $\lambda_m(\mathbf{D}_m)$. Let $\kappa(\mu_{2p}) := \frac{\lambda_1(\mathbf{C}_m)}{\lambda_m(\mathbf{D}_m)}$ with all moments other than μ_{2p} fixed. Then

$$\kappa'(\mu_{2p}) = -\frac{\lambda_1(\mathbf{C}_m)}{\lambda^2_{\min}(\mu_{2p})} \lambda'_{\min}(\mu_{2p}) \leq 0.$$

Case 2: p is even. In this case since $m = \lfloor \frac{p+1}{2} \rfloor = \frac{p}{2}$ and $k = m + 1$, we have $\mathbf{B}_{p+1} = \mathbf{C}_{m+1} \oplus \mathbf{D}_m$. By Proposition 5.3 we have

$$\kappa(\mathbf{A}_{p+1}) = \kappa(\mathbf{C}_{m+1}) = \frac{\lambda_1(\mathbf{C}_{m+1})}{\lambda_{m+1}(\mathbf{C}_{m+1})}.$$

and μ_{2p} is included in the matrix $\mathbf{C}_{m+1}$. By Proposition 5.2, $\mathbf{C}_{m+1}$ is a strictly total positive matrix and hence all eigenvalues are simple and positive, i.e.,

$$\lambda_1(\mathbf{C}_{m+1}) > \lambda_2(\mathbf{C}_{m+1}) > \cdots > \lambda_{m+1}(\mathbf{C}_{m+1}).$$

Hence the functions

$$\lambda_{\max}(\mu_{2p}) := \lambda_1(\mathbf{C}_{m+1}), \quad \lambda_{\min}(\mu_{2p}) := \lambda_{m+1}(\mathbf{C}_{m+1}), \quad \kappa(\mu_{2p}) := \kappa(\mathbf{C}_{m+1})$$

(with all moments other than μ_{2p} fixed) are smooth functions of μ_{2p}. By the quotient rule, we have

$$\kappa'(\mu_{2p}) = \frac{\lambda'_{\max}(\mu_{2p})\lambda_{\min}(\mu_{2p}) - \lambda_{\max}(\mu_{2p})\lambda'_{\min}(\mu_{2p})}{\lambda^2_{\min}(\mu_{2p})}.$$

Since by the Rayleigh–Ritz yheorem [12, Theorem 4.22],

$$\lambda_{\max}(\mu_{2p}) = \max_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{C}_{m+1} \mathbf{x}, \quad \lambda_{\min}(\mu_{2p}) = \min_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{C}_{m+1} \mathbf{x},$$

and the maximum and the minimum is achieved by the unit eigenvector for $\lambda_1(\mathbf{C}_{m+1})$ and $\lambda_{m+1}(\mathbf{C}_{m+1})$, respectively, by Danskin's theorem we have

$$\lambda'_{\max}(\mu_{2p}) = (x_{m+1}^{\max})^2, \qquad \lambda'_{\min}(\mu_{2p}) = (x_{m+1}^{\min})^2,$$

where $x_{m+1}^{\min}$ and $x_{m+1}^{\min}$ denote the last component of the unit eigenvector $\mathbf{x}^{\max}$ for the largest eigenvalue $\lambda_1(\mathbf{C}_{m+1})$ and $\mathbf{x}^{\min}$ for the smallest eigenvalue $\lambda_{m+1}(\mathbf{C}_{m+1})$, respectively. To prove the result it suffices to show that $(x_{m+1}^{\max})^2 \leq (x_{m+1}^{\min})^2$. Since the matrix $\mathbf{C}_{m+1}$ is a positive matrix, by the Perron–Frobenius theorem, the eigenvector corresponding to the largest eigenvalue has strictly positive components and hence $x_i^{\max} > 0, i = 1, \ldots, m+1$. Since $\mathbf{C}_{m+1}\mathbf{x} = \lambda_{\max}\mathbf{x}$ and

$$1 > \mu_2 > \mu_4 > \cdots > \mu_{4m},$$

it is obvious that we have

$$\lambda_{\max} x_1^{\max} > \lambda_{\max} x_2^{\max} > \cdots > \lambda_{\max} x_{m+1}^{\max}$$

which implies that

$$x_1^{\max} > x_2^{\max} > \cdots > x_{m+1}^{\max}.$$

Moreover since $\mathbf{x}^{\max}$ is a unit vector, we have $(x_{m+1}^{\max})^2 < \frac{1}{m+1}$. It remains to show that $(x_{m+1}^{\min})^2 > \frac{1}{m+1}$. The equation

$$\mathbf{x}^T \mathbf{C}_{m+1} \mathbf{x} = \lambda_{\min}$$

represents an $m+1$ dimensional ellipsoid and the unit eigenvector corresponding to $\lambda_{\min}$ is a solution to the above equation. Moreover since $\mathbf{x}^{\min}$ is a minimizer for the problem

$$\min_{\|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{C}_{m+1} \mathbf{x},$$

it is the vertex with the longest radius of the ellipsoid. Since the intercepts of the ellipsoid with the $x_1, \ldots, x_{m+1}$ axes are

$$\pm \begin{pmatrix} \sqrt{\lambda_{\min}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \pm \begin{pmatrix} 0 \\ \frac{\sqrt{\lambda_{\min}}}{\mu_4} \\ \vdots \\ 0 \end{pmatrix} \ldots, \pm \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \frac{\sqrt{\lambda_{\min}}}{\mu_{4m}} \end{pmatrix}$$

and $1 > \mu_4 > \cdots > \mu_{4m}$, the intercept of the ellipsoid with the x_{m+1} axis is the largest one. Hence $|x_{m+1}^{\min}|$ is larger than any of the quantities $|x_m^{\min}|, \ldots, |x_1^{\min}|$ which implies that $(x_{m+1}^{\min})^2 > \frac{1}{m+1}$ since $\mathbf{x}^{\min}$ is a unit vector. Therefore we have shown that $(x_{m+1}^{\max})^2 < (x_{m+1}^{\min})^2$. Consequently we have

$$x_{\max}^2 \lambda_{\min}(\mu_{2p}) < x_{\min}^2 \lambda_{\max}(\mu_{2p})$$

which implies that

$$\lambda'_{\max}(\mu_{2p})\lambda_{\min}(\mu_{2p}) - \lambda_{\max}(\mu_{2p})\lambda'_{\min}(\mu_{2p}) < 0.$$

Therefore $\kappa'(\mu_{2p}) < 0$. ∎

THEOREM 6.1. *One can always find a symmetric K-optimal design with $p+1$ support points which include the boundary points -1 and 1.*

Proof. The case $p = 1, 2$ was proved in section 3. We now show the result for the case $p \geq 3$. Consider the moment space generated by the power functions $u_i(x) = x^{2i}, i = 0, 1, \ldots, p$ on the interval $[0, 1]$:

$$\mathcal{M}_{p+1} := \left\{ \mathbf{c} = (c_0, c_1, c_2, \ldots, c_p) : c_j = \int_0^1 x^{2j} d\sigma(x) \quad j = 0, 1, \ldots, p, \ \sigma \in \Xi \right\},$$

where Ξ denotes the set of all nondecreasing right continuous functions of bounded variation on the interval $[0, 1]$. By Theorem 4.1, one can find a symmetric K-optimal design with at least $p + 1$ support points. Let $\hat{\xi}$ be a symmetric K-optimal design with $q \geq p + 1$ support points and $\hat{\mu} = (1, \hat{\mu}_2, \ldots, \hat{\mu}_{2p})$ be the even moments of order up to $2p$ corresponding to $\hat{\xi}$ and $\hat{\mathbf{c}} := \frac{\hat{\mu}}{2}$.

Case 1: $\hat{\mathbf{c}} = (\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_p)$ is a boundary point of $\mathcal{M}_{p+1}$. It is easy to verify that the moment space is a closed convex cone (see, e.g., [15, Theorem II 1.1]). By the convex separation theorem, there is a supporting hyperplane passing through the origin. Hence one can find scalars $a_0, a_1, \ldots, a_p$ not all zero such that

$$\sum_{j=0}^{p} a_j \hat{c}_j = 0, \tag{16}$$

$$\sum_{j=0}^{p} a_j c_j \geq 0 \quad \text{for all } \mathbf{c} \in \mathcal{M}_{p+1}. \tag{17}$$

Define $u(x) := \sum_{j=0}^{p} a_j x^{2j}$. Suppose that q is even. Then since 0 is not a support point, the symmetric design $\hat{\xi}$ has $l = \frac{q}{2}$ nonnegative support points $0 < \hat{x}_1 < \cdots < \hat{x}_l$ with positive probabilities $\hat{p}_1, \ldots, \hat{p}_l$. By (16), we have

$$0 = \sum_{j=0}^{p} a_j \left(\sum_{i=1}^{l} \hat{p}_i \hat{x}_i^{2j} \right) = \sum_{i=1}^{l} \hat{p}_i \left(\sum_{j=0}^{p} a_j \hat{x}_i^{2j} \right) = \sum_{i=1}^{l} \hat{p}_i u(\hat{x}_i). \tag{18}$$

Since for any $0 \leq x \leq 1$,

$$(1, x^2, x^4, \ldots, x^{2p}) \in \mathcal{M}_{p+1}$$

by (17) we have $u(x) \geq 0$ for all $0 \leq x \leq 1$ and hence (18) implies that $u(\hat{x}_i) = 0, i = 1, \ldots, l$. Let $y = x^2$ and consider the function, $\phi(y) := \sum_{j=0}^{p} a_j y^j$. Since $\phi(y)$ is a polynomial of order less than or equal to p there are at most p zeros. Since $u(x) = \phi(x^2)$, we conclude that there are at most p nonnegative zeros for $u(x)$. However since $u(x) \geq 0, x \in [0, 1]$, all zeros for $u(x)$ in the interval $(0, 1)$ must have multiplicity at least two and hence $u(x)$ has at most $\lfloor \frac{p}{2} \rfloor$ distinctive zeros in the open interval $(0, 1)$. Since $\frac{q}{2} > \lfloor \frac{p}{2} \rfloor$, the boundary 1 must be included as a support point. Moreover since there are at most $\lfloor \frac{p}{2} \rfloor$ distinctive zeros for $u(x)$ in $(0, 1)$ and the support points must be roots of $u(x)$, the number of support points in $(0, 1]$ must be equal to $\frac{q}{2} = \lfloor \frac{p}{2} \rfloor + 1$. Furthermore, since the degree p has to be at least the number of roots (counting multiplicity), we must have $p \geq 2\lfloor \frac{p}{2} \rfloor + 1$, meaning that p has to be odd. Suppose now that q is odd. Then 0 must be a support point and hence the symmetric design $\hat{\xi}$ has $l = \frac{q-1}{2}$ support points on $(0, 1]$. Moreover since 0, as a support point, is now a root of $u(x)$, counting multiplicity, $u(x)$ has at most $\lfloor \frac{p-1}{2} \rfloor$ distinctive roots in $(0, 1)$. Since $\frac{q-1}{2} > \lfloor \frac{p-1}{2} \rfloor$, the boundary 1 must be included as a support point.

Moreover since there are at most $\lfloor \frac{p-1}{2} \rfloor$ distinctive zeros for $u(x)$ in $(0,1)$ and the support points must be roots of $u(x)$, the number of support points in $(0,1]$ must be equal to $\frac{q-1}{2} = \lfloor \frac{p-1}{2} \rfloor + 1$. Furthermore, since the degree p has to be at least the number of roots (counting multiplicity), we must have $p \geq 2\lfloor \frac{p-1}{2} \rfloor + 2$, meaning that p has to be even. Combining the even and the odd cases we have shown that if $\hat{\mathbf{c}}$ is a boundary point of $\mathcal{M}_{p+1}$ then the support points of the design $\hat{\xi}$ are exactly equal to $p+1$ and the boundary points $-1, 1$ are included.

Case 2: $\hat{\mathbf{c}}$ is an interior point of $\mathcal{M}_{p+1}$. In this case, $\hat{\xi}$ itself may not have the desired property. We now show that we can always find another symmetric K-optimal design that has the desired property. Since $\hat{\mathbf{c}}$ is an interior point of $\mathcal{M}_{p+1}$, $(\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_{p-1})$ must be an interior point of $\mathcal{M}_p$. Consider the set which has the same moments with the moment $\hat{\mathbf{c}}$ up to the order $p-1$, i.e.,

$$V(\hat{\mathbf{c}}) := \left\{ \sigma \in \Xi : \hat{c}_j = \int_0^1 x^{2j} d\sigma(x) \quad j = 0, 1, 2, \ldots, p-1 \right\}.$$

Let

$$\bar{\gamma} := \max_{\sigma \in V(\hat{\mathbf{c}})} \int_0^1 x^{2p} d\sigma(x)$$

and $\bar{\sigma}$ denote the maximizer. Then the corresponding moment point

$$\bar{\mathbf{c}} = (\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_{p-1}, \bar{\gamma})$$

is a boundary point of $\mathcal{M}_{p+1}$. Suppose that $\bar{\xi}$ is the symmetric measure on $[-1,1]$ corresponding to $\bar{\sigma}$. Since the matrix $\mathbf{A}_{p+1}(\hat{\xi})$ is positive definite (due to the fact that $\hat{\xi}$ is a K-optimal design), the matrix $\mathbf{A}_{p+1}(\bar{\xi})$ is positive definite as well. Consequently, the number of support points for $\bar{\xi}$ is at least $p+1$. Let the index $I(\bar{c})$ of a point $\bar{c}$ in $\mathcal{M}_{p+1}$ be the minimal number of points in

$$\{(1, x^2, \ldots, x^{2p}) | 0 \leq x \leq 1\}$$

that span $\bar{c}$ under the special convention that $(1, x^2, x^4, \ldots, x^{2p})$ for $x = 0$ and $(1, x^2, x^4, \ldots, x^{2p})$ for $x = 1$ counted as half points, while $(1, x^2, x^4, \ldots, x^{2p})$ for $0 < x < 1$ received a full count. Since $\bar{\mathbf{c}}$ is a boundary point of $\mathcal{M}_{p+1}$, by [15, Theorem II 2.1]) $I(\bar{\mathbf{c}}) < \frac{p+1}{2}$, i.e., $I(\bar{\mathbf{c}}) \leq \frac{p}{2}$. On the other hand since $(\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_{p-1})$ is an interior point of $\mathcal{M}_p$, by [15, Theorem II 2.1]) $I(\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_{p-1}) \geq \frac{p}{2}$. Consequently $I(\bar{c}) = \frac{p}{2}$. When p is even $I(\bar{c}) = \frac{p}{2}$ is an integer and hence either both 0 and 1 are included or not included. But since the number of support points for $\bar{\xi}$ is at least $p+1$ and the index is $\frac{p}{2}$, both 0 and 1 must be included and the number of support points must be $p+1$. When p is odd, $I(\bar{c})$ is not an integer and hence only one of the end points $0, 1$ is included. Since the number of support points for $\bar{\xi}$ is at least $p+1$, 1 must be included and the number of support points must be $p+1$. Hence combining the even and the odd cases we conclude that the number of support points of $\bar{\xi}$ is exactly equal to $p+1$ and the boundary points $-1, 1$ are included in the support points. By Lemma 6.1, $\kappa(\mathbf{A}_{p+1})$ is a nonincreasing function of μ_{2p} and when the other moments are fixed we have

$$\kappa(\mathbf{A}_{p+1}(\bar{\xi})) \leq \kappa(\mathbf{A}_{p+1}(\hat{\xi})).$$

That is, $\bar{\xi}$ is a symmetric K-optimal design with $p+1$ support points which include $-1, 1$. $\square$

7. Numerical algorithms and results. It is difficult to derive K-optimal designs analytically for general p. In this section, we explain how the results in the previous sections can be used in finding numerical solutions, and the solutions for symmetric K-optimal designs for $p = 3, \dots, 10$ are presented and compared with D-optimal and A-optimal designs. Two numerical algorithms are presented and discussed here. Algorithm I computes the support points and their probabilities directly, while Algorithm II computes the moments $\mu_2, \dots, \mu_{2p}$ first and then finds the support points and their probabilities.

7.1. Algorithm I. By Theorem 6.1, any K-optimal design can be realized by a symmetric K-optimal design with exactly $p+1$ support points including the boundary points. By Theorem 5.1, the condition number is smooth if there are at least $p+1$ support points. These results allow us to simplify the problem with the minimal number of unknown variables.

A symmetric design with $p+1$ support points including the boundary points -1 and 1 can be represented by

$$\xi_K = \left\{ \begin{array}{l} 1, x_2, \dots, x_m, x_{m+1}, -x_m, \dots, -x_2, -1 \\ p_1, p_2, \dots, p_m, p_{m+1}, \quad p_m, \dots, \quad p_2, \quad p_1 \end{array} \right\},$$

where $x_i \in (0,1)$ $(i = 2, \dots, m)$, $p_i \in (0,1]$ $(i = 1, \dots, m)$, $m = \lfloor \frac{p+1}{2} \rfloor$, $x_{m+1} = 0$, and $2\sum_{i=1}^m p_i + p_{m+1} = 1$. For even p, $p_{m+1} > 0$, while for odd p, $p_{m+1} = 0$.

Let $w = (x_2, \dots, x_m, p_1, \dots, p_{m+1})$. Then the information matrix can be written as a function of w:

$$\mathbf{A}(w) = 2\left[p_1 \mathbf{W}(1) + \sum_{i=2}^m p_i \mathbf{W}(x_i) \right] + p_{m+1}\mathbf{W}(0),$$

where matrix $\mathbf{W}(x)$ is defined in (11). Our optimization problem becomes

$$\begin{aligned} (P) \qquad & \min_w \; z(w) := \kappa(\mathbf{A}(w)) \\ & \text{s.t. } x_2, \dots, x_m \in (0,1), \\ & \qquad p_1, \dots, p_{m+1} \in (0,1], \\ & \qquad 2\sum_{i=1}^m p_i + p_{m+1} = 1, \end{aligned}$$

where for odd p, $p_{m+1} = 0$. Note that since when the support points are fewer than $p+1$ the condition number will become infinity, the optimal design points of the problem (P) will be always distinct. By Theorem 5.1, $\kappa(\mathbf{A})$ is a smooth function of the moments $\mu_2, \dots, \mu_{2p}$ when the number of support points is equal to $p+1$. Since the moments are smooth functions of w, the objective function $z(w)$ is a smooth function of w on the feasible region of the problem (P) as a composition of two smooth functions $w \to \mu$ and $\mu \to \kappa(\mathbf{A})$. Many software programs have an optimization toolbox that can be used for the above problem when p is small. In the computation, we find that it is easier to use objective function $\kappa(\mathbf{A}(w))^{1/(p+1)}$ than to use $\kappa(\mathbf{A}(w))$, since the search algorithm converges faster.

7.2. Algorithm II. As suggested by one of the referees, we can also use SeDuMi in [23] to solve the K-optimal design problem in the following two steps.

Step 1: Minimize $\kappa(\mathbf{A_{p+1}})$ over $\mu_2, \ldots, \mu_{2p}$ subject to a constraint on the moments $\mu_2, \ldots, \mu_{2p}$. The minimization problem can be transformed to a semidefinite programming problem which can be solved by the SeDuMi algorithm in MATLAB software. Let

$$s = \frac{\lambda_1(\mathbf{A}_{p+1})}{\lambda_{p+1}(\mathbf{A}_{p+1})}, \quad u_j = \frac{\mu_j}{\lambda_{p+1}(\mathbf{A}_{p+1})}, \quad j = 0, 1, \ldots, 2p,$$
$$\mathbf{U}_{p+1} = (u_{i+j-2})_{(p+1)\times(p+1)} = \frac{1}{\lambda_{p+1}(\mathbf{A}_{p+1})}\mathbf{A}_{p+1}, \quad \mathbf{V}_p = (u_{i+j})_{p\times p},$$

where $\mu_0 = 1$ and $\mu_j = 0$ for odd j. It is obvious that $\lambda_1(\mathbf{U}_{p+1}) = s$ and $\lambda_{p+1}(\mathbf{U}_{p+1}) = 1$. The constraint on the moments $\mu_2, \ldots, \mu_{2p}$ is, from [16, Theorem 5.39], $\mathbf{U}_p - \mathbf{V}_p \succeq 0$ (positive semidefinite). Then the minimization problem becomes a semidefinite programming problem as follows:

$$\min_{s, u_0, u_2, \ldots, u_{2p}} \quad s$$
$$s.t. \quad \begin{pmatrix} \mathbf{U}_p - \mathbf{V}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & s\,\mathbf{I} - \mathbf{U}_{p+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}_{p+1} - \mathbf{I} \end{pmatrix} \succeq 0.$$

Use the SeDuMi algorithm to find the minimizer: $\hat{s}, \hat{u}_0, \hat{u}_2, \ldots, \hat{u}_{2p}$, and the corresponding moments are given by $\hat{\mu}_j = \hat{u}_j/\hat{u}_0$, $j = 2, \ldots, 2p$.

Step 2: Find design points $x_2, \ldots, x_m$ and probabilities $p_1, p_2, \ldots, p_m, p_{m+1}$ to match the moments $\hat{\mu}_2, \ldots, \hat{\mu}_{2p}$ obtained in Step 1. Notice the relationship between the distribution and the moments,

$$2p_1 + 2\sum_{i=2}^{m} x_i^j p_i = \hat{\mu}_j, \quad j = 2, \ldots, 2p,$$
$$2\sum_{i=1}^{m} p_i + p_{m+1} = 1.$$

7.3. Numerical results. Using Algorithm I, we can easily obtain the numerical results for the K-optimal designs for small p, and some representative results are presented in Table 2. Numerical results from Algorithm II are consistent with those in Table 2. For example, when $p = 3$, using the SeDuMi algorithm we get $\hat{\mu}_2 = 0.3626$, $\hat{\mu}_4 = 0.2287$, and $\hat{\mu}_6 = 0.2006$, and the design points and probabilities in Table 2 give the same values for these moments. The numerical results for $3 \le p \le 10$ and the theoretical results for $p = 1$ and 2 indicate that the K-optimal designs are unique with respect to the moments.

D-optimal and A-optimal designs can be found in [22]. To compare the K-optimal designs with D-optimal and A-optimal designs, we can plot the support points and their probabilities. A representative plot is given in Figure 1 for $p = 4$. It is very interesting to notice that the support points are almost the same for the three optimal designs, but the probabilities are different. The probabilities of K-optimal designs are similar to those of A-optimal designs, while the probabilities of D-optimal designs are constant over the support points.

Although the distributions of D-optimal, A-optimal, and K-optimal designs are different, they are similar to each other in the following ways:

TABLE 2

Numerical results for K-optimal designs for pth order polynomial regression models, $3 \le p \le 10$. Since the designs are symmetric, only nonnegative support points (rounded to 3 decimal points) are listed here with probabilities in the brackets.

p	K-optimal design						$\kappa(\mathbf{A})^{1/(p+1)}$
3	0.458 (0.403)	1.0 (0.097)					2.327671
4	0.0 (0.405)	0.663 (0.240)	1.0 (0.057)				2.76018
5	0.286 (0.303)	0.779 (0.153)	1.0 (0.044)				3.073318
6	0.0 (0.292)	0.469 (0.219)	0.841 (0.104)	1.0 (0.031)			3.341160
7	0.209 (0.241)	0.594 (0.156)	0.884 (0.077)	1.0 (0.026)			3.549046
8	0.0 (0.229)	0.362 (0.192)	0.679 (0.116)	0.909 (0.058)	1.0 (0.020)		3.731495
9	0.165 (0.198)	0.478 (0.149)	0.742 (0.088)	0.929 (0.047)	1.0 (0.018)		3.879493
10	0.0 (0.188)	0.295 (0.167)	0.565 (0.118)	0.788 (0.069)	0.942 (0.038)	1.0 (0.014)	4.011976

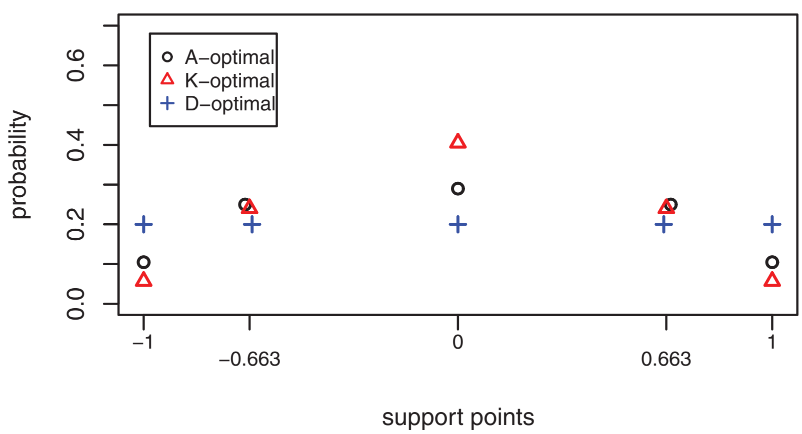

FIG. 1. *K-optimal, D-optimal, and A-optimal designs for $p = 4$.*

(1) They are all symmetric.
(2) They all have $p + 1$ support points.
(3) For each p, the $p + 1$ support points of K-optimal design are almost the same as those of D-optimal design or A-optimal design.
(4) The distribution of K-optimal design is closer to that of A-optimal design than D-optimal design.

Therefore K-optimal designs can enjoy some of the good properties of A-optimal or D-optimal designs and at the same time they are numerically more stable when solving the normal equations to obtain the least squares estimator.

8. Conclusions. The K-optimal criterion based on the condition number of the information matrix is introduced to construct K-optimal designs for regression models. The K-optimal designs are very useful in statistical analysis, since they can reduce the variance inflation factor and the error sensitivity for the least squares estimator. We have focused on the polynomial regression models in this paper; however, the K-

optimal criterion can be applied to any regression models. Theoretical properties of K-optimal designs, such as symmetry and the number of support points, are investigated and obtained.

The theoretical distributions of the K-optimal designs are obtained for $p = 1, 2$, but they are hard to derive for $p \geq 3$. Since we can prove that the condition number of the information matrix is a smooth function of the moments and the optimization problem can be transformed to a semidefinite programming problem, two numerical algorithms are proposed to compute K-optimal designs. The numerical results indicate that the K-optimal designs are unique with respect to the moments. In addition, the K-optimal designs are similar to A-optimal designs for the polynomial regression models.

An interesting research problem in the future is to study K-optimal designs for other regression models, such as polynomial regression models with other bases and multiple regression models. In addition, the condition number can be a useful measure to compare various models and to do model selection by avoiding mulicollinearity. The two numerical algorithms proposed in section 7 are efficient to find K-optimal designs for small p, but it is still challenging to design effective algorithms for large p.

Acknowledgments. We would like to thank Professor Pauline van den Driessche for showing us the interlacing theorem and the connection between the information matrix and the matrix of the direct sums in section 5. We would like to thank anonymous referees for their insightful comments which helped us to improve the presentation of the paper.

REFERENCES

[1] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2003.
[3] S. Boyd, L.E. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM Stud. Appl. Math. 15, SIAM, Philadelphia, 1994.
[4] R. D. Braatz and M. Morari, *Minimizing the Euclidean condition number*, SIAM J. Control Optim., 32 (1994), pp. 1763–1768.
[5] X. Chen, R.S. Womersley, and J.J. Ye, *Minimizing the condition number of a gram matrix*, SIAM J. Optim., 21 (2011), pp. 127–148.
[6] F.H. Clarke, Y.S. Ledyaev, R.J. Stern, and P.R. Wolenski, *Nonsmooth Analysis and Control Theory*, Springer, New York, 1998.
[7] J.M. Danskin, *The Theory of Max-Min and its Applications to Weapons Allocation Problems*, Springer, New York, 1967.
[8] V.V. Fedorov, *Theory of Optimal Experiments*, Academic Press, New York, 1972.
[9] V.V. Fedorov and P. Hackl, *Model-Oriented Design of Experiments*, Springer-Verlag, New York, 1997.
[10] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, MD, 1993.
[11] C. Greif and J.M. Varah, *Minimizing the condition number for small rank modifications*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 82–97.
[12] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1985.
[13] C.R. Kaplan, *Systematic use of condition numbers in experimental design*, M.Sc. paper, University of Maryland, College Park, MD, 1985.
[14] C.R. Kaplan and J.W. Gentry, *Use of condition numbers for shortcut experimental design*, Amer. Inst. Chem. Engrg. J., 33 (1987), pp. 681–685.
[15] S. Karlin and W. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, John Wiley & Sons, New York, 1966.
[16] M. Laurent, Updated version of "Sums of Squares, Moment Matrices and Optimization over Polynomials", http://homepages.cwi.nl/~monique/files/moment-ima-update-new.pdf (February 6, 2010).

[17] Z. Lu and T.K. Pong, *Minimizing condition number via convex programming*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 1193–1211.
[18] P. Maréchal and J.J. Ye, *Optimizing condition numbers*, SIAM J. Optim., 20 (2009), pp. 935–947.
[19] D.C. Montgomery, E.A. Peck, and G.G. Vining, *Introduction to Linear Regression Analysis*, 4th ed., Wiley, New York, 2006.
[20] M.L. Overton, *Large-scale optimization of eigenvalues*, SIAM J. Optim., 2 (1992), pp. 88–120.
[21] A. Pinkus, *Totally Positive Matrices*, Cambridge University Press, Cambridge, 2010.
[22] F. Pukelsheim, *Optimal Design of Experiments*, Wiley, New York, 1993.
[23] J.F. Sturm, *Using SeDuMi* 1.02*, a MATLAB toolbox for optimization over symmetric cones*, Optim. Methods Softw., 11/12 (1999), pp. 625–653.
[24] H. Wolkowicz, R. Saigal, and L. Vandenberghe, *Handbook of Semidefinite Programming*, Kluwer Academic, Boston, 2000.