A note on optimal designs for estimating the slope of a polynomial regression

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

Viatcheslav B. Melas

St. Petersburg State University

Department of Mathematics

St. Petersburg , Russia

email: vbmelas@yandex.ru

Petr Shpilev

St. Petersburg State University

Department of Mathematics

St. Petersburg , Russia

email: pitshp@hotmail.com

September 15, 2020

Abstract

In this note we consider the optimal design problem for estimating the slope of a polynomial regression with no intercept at a given point, say z. In contrast to previous work, which considers symmetric design spaces we investigate the model on the interval $[0, a]$ and characterize those values of z, where an explicit solution of the optimal design is possible.

AMS subject classification: 62K05

Keywords and phrases: polynomial regression, slope estimation, c-optimal designs

1

1 Introduction

Consider the common polynomial regression model of degree n with no intercept

$$Y_i = f^\top(x_i)\theta + \varepsilon_i = (x_i, x_i^2, \ldots, x_i^n)^\top \theta + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{1.1}$$

where $\varepsilon_1, \ldots, \varepsilon_N$ denote independent random variables with $\mathbb{E}[\varepsilon_i] = 0$; $\mathrm{Var}(\varepsilon_i) = \sigma^2 > 0$, $\theta = (\theta_1, \ldots, \theta_n)^\top \in \mathbb{R}^n$ is a vector of unknown parameters and the explanatory variables $x_1, \ldots, x_N$ vary in the interval $[0, a]$ for some $a > 0$. An approximate optimal design [in the sense of Kiefer (1974)] minimizes an appropriate function of the (asymptotic) covariance matrix of the statistic $\sqrt{N}\hat\theta$, where the $\hat\theta$ denotes the least squares estimate of the parameter θ in the regression model (1.1) [see Silvey (1980) or Pukelsheim (2006)].

In a recent paper Dette et al. (2020) considered model (1.1) on the symmetric interval $[-1, 1]$ and determined explicitly the approximate optimal design for estimating the derivative of the regression function

$$\frac{d}{dx} f^\top(x)\theta \Big|_{x=z} = \sum_{j=1}^n j\theta_j z^{j-1}$$

at the point z, which minimizes the variance of the best linear unbiased estimate of $\sum_{j=1}^n j\theta_j z^{j-1}$. The corresponding optimality criterion is a special case of the well known c-optimality criterion [see, for example, Elfving (1952); Studden (1968) or Pukelsheim (2006), Chapter 2].

In practice, however, polynomial regression models with no intercept are usually used on a positive interval, where x corresponds, for example, to speed, concentration or time, and the response at the initial point $x = 0$ is [see, for example, Huang et al. (1995); Li et al. (2005)]. Therefore the goal of this note is to provide some optimal designs for estimating the slope of polynomial regression model with no intercept in the case where the design space is given by the interval $[0, a]$. In Section 2 we introduce the basic optimal design problem and review a geometric characterization of c-optimal designs. The main result can be found in Section 3 where the optimal designs for estimating the slope at the point in a polynomial regression model with no intercept are determined explicitly and the theory is illustrated by several examples.

2 c-optimal designs

Consider the regression model (1.1) on the interval $[0, a]$. Following Kiefer (1974) we call a discrete probability measure

$$\xi = \begin{pmatrix} x_1 & \cdots & x_m \\ \omega_1 & \cdots & \omega_m \end{pmatrix}$$

with support points $x_1, \ldots, x_m \in [0, a]$ and weights $\omega_1, \ldots, \omega_m$ an approximate design (on the interval $[0, a]$). If N observations can be taken this means that the quantities $N\omega_i$ are rounded to integers, say n_i, with $\sum_{i=1}^m n_i = N$ and n_i observations are taken at each experimental condition x_i $(i = 1, \ldots, m)$. For an approximate design ξ we denote by

$$M(\xi) = \int_{[0,a]} f(x) f^\top(x) \xi(dx)$$

its information matrix in the model (1.1), where $f(x) = (x, \ldots, x^n)^\top$ is the vector regression functions. The covariance matrix of the least squares estimate for the parameter θ, say $\hat{\theta}$, can be approximated (if $N \to \infty$, $n_i/N \to \omega_i$) by $\sigma^2/N M^{-1}(\xi)$ and an optimal design minimizes an appropriate real valued function of the matrix $M^{-1}(\xi)$. In this paper we are interested in designs minimizimng the asymptotic variance of the best linear unbiased estimate $c^\top \hat{\theta}$ of the linear combination $c^\top \theta$ for a given vector $c \in \mathbb{R}^n$. To be precise, we call a design ξ c-optimal in the regression model (1.1), if it minimizes the function

$$\Phi(\xi) = \begin{cases} c^\top M^-(\xi) c, & \text{if there exists a vector } v \in \mathbb{R}^d \text{ such that } c = M(\xi)v; \\ \infty, & \text{otherwise,} \end{cases}$$

where $M^-(\xi)$ is a generalized inverse for the matrix $M(\xi)$. In the first case the design ξ is called *admissible for estimating the linear combination $c^\top \theta$* in the regression model (1.1) and the value of the quadratic form does not depend on the choice of the generalized inverse [see Pukelsheim (2006)]. The choice $c = f'(z) = (1, 2z, \ldots, nz^{n-1})^\top$ for some z corresponds to the minimization of the variance of the best unbiased prediction of the derivative of the regression function $\theta^\top f(x)$ at the point z. The optimal design is called *optimal design for estimating derivative at the point z* in this case.

A useful tool for the determination of c-optimal designs is a geometric characterization of the c-optimal design and which is called Elfving's theorem in the literature [see Elfving (1952)]. We formulate it here in a slightly different form, which can be directly used to check optimality of a given design [see Dette et al. (2004) for details].

3

Theorem 2.1 *An admissible design ξ^* for estimating the linear combination $c^\top \theta$ with support points $x_1, x_2, \ldots, x_m \in \mathcal{X} = [0, a]$ and weights $\omega_1, \omega_2, \ldots, \omega_m$ is c-optimal if and only if there exists a vector $p \in \mathbb{R}^n$ and a constant h such that the following conditions are satisfied:*

(1) $|p^\top f(x)| \leq 1$ for all $x \in \mathcal{X}$;

(2) $|p^\top f(x_i)| = 1$ for all $i = 1, 2, \ldots, m$;

(3) $c = h \sum_{i=1}^m f(x_i) \omega_i p^\top f(x_i)$.

Moreover, in this case we have $c^\top M^-(\xi^*) c = h^2$ and the function $p^\top f(x)$ is called extremal polynomial.

3   Optimal designs for estimating the slope

For the linear model through the origin (that is $n = 1$) it is easy to see using Elfving's theorem that the optimal design for estimating the slope is unique and puts all mass at the point a (independently of the point z). However, in the case $n > 1$ the situation is more complicated. By Theorem 2.1 it follows that the support points of the optimal design are extremal points of a polynomial of the form $p^\top f(x) = \sum_{i=1}^n p_i x^i$. In fact it is possible to identify a candidate for this optimal polynomial explicitly. For this purpose let

$$T_n(x) = \cos(n \arccos(x))$$

denote the nth Chebyshev polynomial of the first kind [see Szegö (1975)] and consider the polynomial

$$S_n(x) = T_n\left(\frac{x}{a}(1 + \cos\frac{\pi}{2n}) - \cos\frac{\pi}{2n}\right). \tag{2.1}$$

It is easy to see that $S_n(x)$ has exactly n extremal points $s_1 < s_2 < \cdots < s_n$ on the interval $[0, a]$, which are given by

$$s_i = a \cdot \frac{\cos\frac{(i-1)\pi}{n} + \cos\frac{\pi}{2n}}{1 + \cos\frac{\pi}{2n}}, \quad i = 1, 2, \ldots, n. \tag{2.2}$$

For the statement of our main result we define $\bar{L}_1, \ldots, \bar{L}_n$ as the Lagrange basis interpolation polynomials without intercept corresponding to the nodes $s_1, \ldots, s_n$, that is

$$\bar{L}_i(z) = \frac{z \prod_{j \neq i}(z - s_j)}{s_i \prod_{j \neq i}(s_i - s_j)}, \tag{2.3}$$

4

and denote by $\bar{L}_i'$ the derivative of $\bar{L}_i$ $(i = 1, \ldots, n)$.

Theorem 3.1 *Consider the polynomial regression model of degree $n > 1$ with no intercept on the interval $[0, a]$. The optimal design $\xi^*(z)$ for estimating the slope of this model at the point z is supported at the points $s_1, \ldots, s_n$ defined in (2.2) if and only if*

$$z \in \bigcup_{i=1}^{n} (\omega_{1,i-1}, \omega_{n,i}),$$

where $-\omega_{1,0} = \omega_{n,n} = \infty$ and $\omega_{i,k}$ is k-th root of the function

$$\omega_i(z) = \frac{|\bar{L}_i'(z)|}{\sum_{j=1}^{n} |\bar{L}_j'(z)|} \quad , \quad i = 1, \ldots, n, \tag{2.4}$$

$k = 1, \ldots, n-1, j = 1, \ldots, n$. Moreover, in this case the weight of the design $\xi^*(z)$ at s_i is given by $\omega_i(z)$ $(i = 1 \ldots, n)$.

To prove this Theorem we use the following Lemma. The proof can be found in Sahm (1998) or in Dette et al. (2020).

Lemma 3.2 *Let $P_1(x)$ and $P_2(x)$ be polynomials of degree n with n distinct roots $t_{(1,1)} < t_{(1,2)} < \ldots < t_{(1,n)}$ and $t_{(2,1)} < t_{(2,2)} < \ldots < t_{(2,n)}$, respectively. Assume that the roots are interlacing in the following sense:*

$$t_{(1,1)} \leq t_{(2,1)} < t_{(1,2)} \leq t_{(2,2)} < \ldots < t_{(1,n)} \leq t_{(2,n)} \; ,$$

where at least one of the inequalities $t_{(\ell,1)} \leq t_{(\ell,2)}$ $(\ell = 1, \ldots n)$ is strict. Then the roots $v_{(1,1)} \leq v_{(1,2)} \leq \ldots \leq v_{(1,n-1)}$ and $v_{(2,1)} \leq v_{(2,2)} \leq \ldots \leq v_{(2,n-1)}$ of the derivatives $P_1'(x)$ and $P_2'(x)$ are strictly interlacing, that is

$$v_{(1,1)} < v_{(2,1)} < \ldots < v_{(1,n-1)} < v_{(2,n-1)}.$$

Proof of Theorem 3.1.
We will check the optimality of the design $\xi^*(z)$ by an application of Theorem 2.1. Note, that the polynomial $S_n(z)$ defined in (2.1) obviously satisfies to conditions (1) and (2) of this theorem.

It now remains to characterize those values of z such that the system of equations defined by condition (3) in Theorem 2.1 admits a solution with nonnegative weights ω_i satisfying $\sum_{i=1}^{m} \omega_i = 1$. Note that condition (3) in Theorem 2.1 can be rewritten in the form

$$c = f'(z) = (1, z, \ldots, nz^{n-1})^\top = hF\beta, \tag{2.5}$$

where

$$F = ((s_j)^i)_{i,j=1}^{n} = (f(s_1), \ldots, f(s_n)) \in \mathbb{R}^{n \times n} \text{ and } \beta_i = \omega_i(p^\top f(s_i))$$

In order to investigate the system of equations defined by (2.5) note that the identity $F^{-1}F = I_n$ (here I_n is the identity matrix) implies

$$e_i^\top F^{-1} f(s_j) = \delta_{ij} \quad (i, j = 1, \ldots, n),$$

where δ_{ij} is the Kroneker symbol and $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top \in \mathbb{R}^n$ the ith unit vector. As these equations characterize the ith Lagrange basis interpolation polynomial $\bar{L}_i(z) = a_i^T f(z)$ with nodes $s_1, \ldots, s_n$ we have

$$e_i^\top F^{-1} f(z) = \bar{L}_i(z), \ i = 1, \ldots, n.$$

Differentiating both sides of the equation with respect to z yields

$$e_i^\top F^{-1} f'(z) = \bar{L}_i'(z), \ i = 1, \ldots, n,$$

or equivalently

$$F^{-1} f'(z) = (\bar{L}_1'(z), \ldots, \bar{L}_n'(z))^\top.$$

Therefore we obtain for the solution of (2.5)

$$h\beta = (\bar{L}_1'(z), \ldots, \bar{L}_n'(z))^\top$$

or equivalently (since $\beta_i = \omega_i(p^\top f(s_i))$)

$$h\beta_i = h\omega_i(-1)^{n-i} = \bar{L}_i'(z) , \quad i = 1, \ldots, n. \tag{2.6}$$

Consequently applying Lemma 3.2 to the pairs of polynomials $\bar{L}_i(z), \bar{L}_{i+1}(z)$ from (2.3) $i = 1, \ldots, n-1$ we obtain that the roots of functions $\bar{L}_i'(z)$ are strictly interlacing, that is

$$\omega_{(n,1)} < \ldots < \omega_{(1,1)} < \omega_{(n,2)} < \ldots < \omega_{(1,2)} < \ldots < \omega_{(1,n-2)} < \omega_{(n,n-1)} < \ldots < \omega_{(1,n-1)}.$$
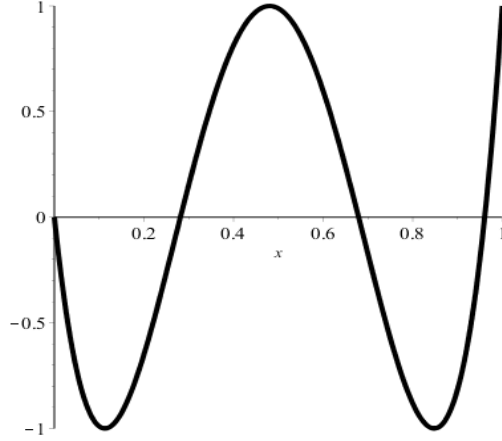
Figure 1: *The extremal polynomial $S_4(x)$ on the interval $[0,1]$ $(a = 1)$.*

This immediately implies that each of the functions $\bar{L}'_i(z)$ has only one root in the intervals $[\omega_{(n,i)}, \omega_{(1,i)}]$ and has no roots in the intervals $A_1, \ldots, A_n$, where the set A_i is defined by $A_i = (\omega_{1,i-1}, \omega_{n,i})$, $i = 1, \ldots, n$. Moreover, for $z \in A_n$ we have

$$\text{sign}((-1)^{n-i}\bar{L}'_i(z)) = 1 \, , \quad i = 1, \ldots, n$$

(since $\text{sign}(\bar{L}'_i(z)) = (-1)^{n-i}$), and for $z \in A_j$

$$\text{sign}((-1)^{n-i}\bar{L}'_i(z)) = (-1)^{n+j} \, , \quad i, j = 1, \ldots, n.$$

This implies that

$$|\bar{L}'_i(z)| = (-1)^{n+j}(-1)^{n-i}\bar{L}'_i(z) \quad (i, j = 1, \ldots, n)$$

for $z \in A_j$.

The proof is now completed observing (2.6), which implies that for $z \in A_j$, the weights are given by

$$\omega_i(z) = \frac{(-1)^{n-i}\bar{L}'_i(z)}{h} = \frac{(-1)^{n+j}|\bar{L}'_i(z)|}{h} \quad (i = 1, \ldots, n)$$

with $h = (-1)^{n+j}\sum_{i=1}^{n}|\bar{L}'_i(z)|$. $\qquad\square$

Example 3.3 In this example we illustrate potential applications of Theorem 3.1 determining optimal designs for estimating the slope of a polynomial regression with no intercept on the interval $[0,1]$.
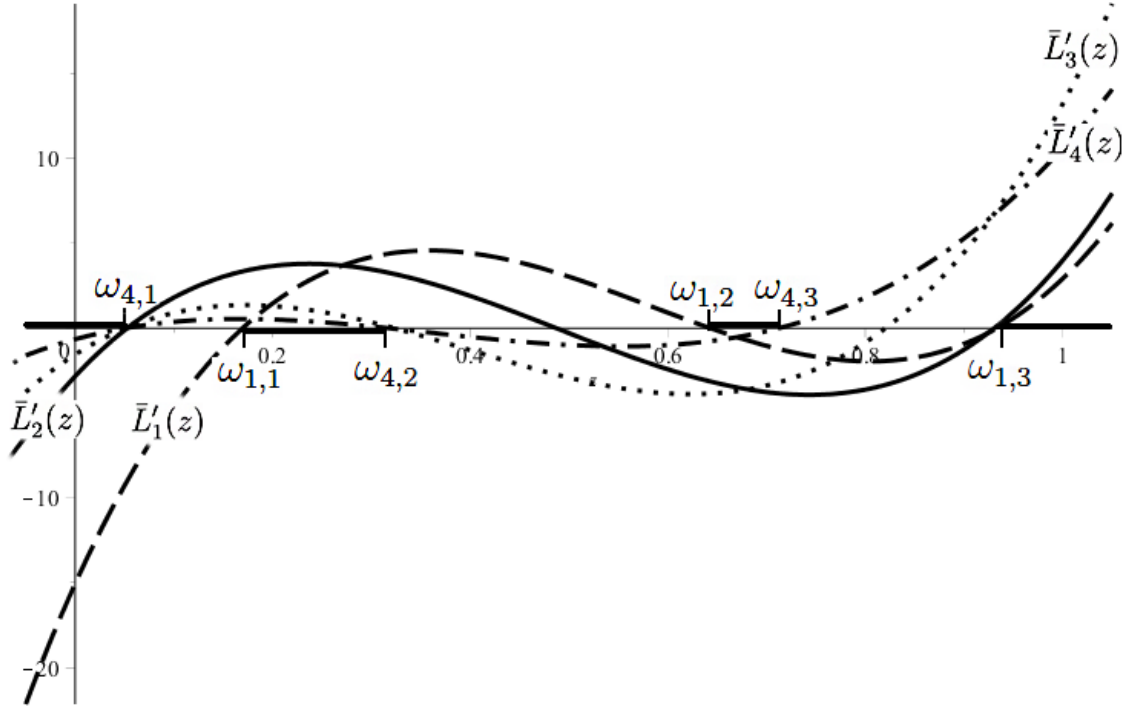
7

Figure 2: *The functions $\bar{L}'_i(z)$, $z \in \mathbb{R}$ for $i = 1, 2, 3, 4$ ($n = 4$, $a = 1$).*

We start with the case of a quadratic regression model, that is $n = 2$. Here the extremal points in (2.2) are given by $s_1 = \sqrt{2} - 1$, $s_2 = 1$ and the derivatives of the polynomials in (2.3) are calculated as

$$\bar{L}'_1(z) = -\frac{4 + 3\sqrt{2}}{2}(2z - 1) \; , \; \bar{L}'_2(z) = -\frac{2 + \sqrt{2}}{2}(-2z + \sqrt{2} - 1).$$

The corresponding roots of the functions (2.4) are obtained as

$$\omega_{1,1} = \frac{1}{2} \; , \; \omega_{2,1} = \frac{1}{2}(1 - \sqrt{2})$$

and the optimal design for estimating the slope of the polynomial regression without intercept is supported at points $\sqrt{2} - 1, 1$ if and only if

$$z \in \left(-\infty, \frac{1}{2}(1 - \sqrt{2})\right) \cup \left(\frac{1}{2}, \infty\right)$$

As a second example we consider the cubic regression model with no intercept, that is $n = 3$. In this case the extremal points of the polynomial $S_3(x)$ are $s_1 = 3\sqrt{3} - 5$, $s_2 = \sqrt{3} - 1$,

8

$s_3 = 1$ and the derivatives of the Lagrange interpolation polynomials in (2.3) are given by

$$\bar{L}'_1(z) = 35.490z^2 - 40.981z + 8.6607 \ , \ \bar{L}'_2(z) = -28.548z^2 + 22.767z - 1.8680$$
$$\bar{L}'_3(z) = 13.933z^2 - 8.6240z + .66745,$$

The roots of the roots of the functions (2.4) are obtained as

$$\omega_{1,1} = 0.2785 \ , \ \omega_{2,1} = 0.0935 \ , \ \omega_{3,1} = 0.090 \ , \ \omega_{1,2} = 0.8758 \ , \ \omega_{2,2} = 0.7045 \ , \ \omega_{3,2} = 0.528$$

By Theorem 3.1 the optimal design for estimating the slope of the polynomial regression without intercept at the point z is supported at the points $\{3\sqrt{3} - 5, \sqrt{3} - 1, 1\}$ if and only if

$$z \in (-\infty, 0.090) \cup (0.2785, 0.528) \cup (0.8758, \infty)$$

Finally we consider model (1.1) with $n = 4$, where the extremal polynomial $S_4(x)$ is displayed in Figure 1. The corresponding extremal points are given by $s_1 = 0.1127$, $s_2 = 0.4802$,, $s_3 = 0.8477$ and $s_1 = 1$. The derivatives of the polynomials in (2.3) are calculated as

$$\bar{L}'_1(z) = -148.08z^3 + 258.55z^2 - 128.47z + 15.072,$$
$$\bar{L}'_2(z) = 118.63z^3 - 174.42z^2 + 62.631z - 2.8327$$
$$\bar{L}'_3(z) = -114.72z^3 + 137.04z^2 - 37.110z + 1.5517,$$
$$\bar{L}'_4(z) = 56.968z^3 - 61.552z^2 + 15.858z - 0.65327$$

and displayed in Figure 2. The roots of the functions (2.4) are given by

$$\omega_{1,1} = 0.1696 \ , \ \omega_{2,1} = 0.05268 \ , \ \omega_{3,1} = 0.05102 \ , \ \omega_{4,1} = 0.05071,$$
$$\omega_{1,2} = 0.6432 \ , \ \omega_{2,2} = 0.4872 \ , \ \omega_{3,2} = 0.3232 \ , \ \omega_{4,2} = 0.3175,$$
$$\omega_{1,3} = 0.9332 \ , \ \omega_{2,3} = 0.9305 \ , \ \omega_{3,3} = 0.8205 \ , \ \omega_{4,3} = 0.7123.$$

Therefore, by Theorem 3.1 the optimal design for estimating the slope of the polynomial regression without intercept at the point z is supported at the points $0.1127, 0.4802, 0.8477, 1$ if and only if

$$z \in (-\infty, 0.05071) \cup (0.1696, 0.3175) \cup (0.6432, 0.7123) \cup (0.9332, \infty)$$

References

Dette, H., Melas, V. B., and Shpilev, P. (2020). Some explicit solutions of c-optimal design problems for polynomial regression with no intercept. *Accepted for publication by Annals of the Institute of Statistical Mathematics.* https://doi.org/10.1007/s10463-019-00736-0.

Dette, H., Melas, V. B., and Pepelyshev, A. (2004). Optimal designs for estimating individual coefficients in polynomial regression - a functional approach. *Journal of Statistical Planning and Inference*, 118(1):201 – 219.

Elfving, G. (1952). Optimal allocation in linear regression theory. *The Annals of Mathematical Statistics*, 23:255–262.

Huang, M.-N. L., Chang, F.-C., and K., W. W. (1995). *D*-optimal designs for polynomial regression without an intercept. *Statistica Sinica*, 5(2):441–458.

Kiefer, J. (1974). General Equivalence Theory for Optimum Designs (Approximate Theory). *The Annals of Statistics*, 2(5):849–879.

Li, K.-H., Lau, T.-S., and Zhang, C. (2005). A note on *D*-optimal designs for models with and without an intercept. *Statistical Papers.*, 46(3):451–458.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. SIAM, Philadelphia.

Sahm, M. (1998). *Optimal designs for estimating individual coefficients in polynomial regression*. PhD thesis, Fakultät für Mathematik, Ruhr-Universität Bochum, Germany.

Silvey, S.(1980). *Optimal Design*. Chapman and Hall, London.

Studden, W. J. (1968). Optimal designs on Tchebycheff points. *Annals of Mathematical Statistics*, 39(5):1435–1447.

Szegö, G.(1975). *Orthogonal Polynomials*. American Mathematical Society, Providence, R.I.