Информационные процессы, Том 16, № 3, 2016, стр. 247–259

INFORMATION THEORY AND INFORMATION PROCESSING

On Testing Degree of a Multivariate Polynomial Regression

M. Malyutov

Mathematics Dept., Northeastern University, 360 Huntington Ave., Boston, MA 02115, USA
E-mail: m.malioutov@neu.edu

Received 10.06.2016

Abstract—This paper continues Separate Testing Inputs vs. Linear Programming relaxation. We study Response Surface Methodology in its decision stage on whether linear model should be replaced with a higher order polynomial model both under sparsity and its absence. Optimal properties of adding repeated measurements in the central point of the Complete Factorial Design under preliminary information on *identical signs* of quadratic coefficients is established under sparsity. Next, we study maximin designs which maximize the minimal power of discrimination under normality and a fixed norm of higher order coefficients.

KEYWORDS: Random design, active inputs, separate testing of inputs, sparsity, multivariate polynomial models, discrimination.

1. INTRODUCTION

Our preceding publication [10] outlines the dramatic history of the sparsity assumption application to the Response Surface Methodology [12], proves the asymptotic optimality of certain estimates of factorial models obtained from the random sample of the Complete Factorial Experiment under sparsity, small noise and rising dimension of the factorial model. The two-step estimates mentioned above use the Separate Testing of Effects for finding active ones followed by Least Squares estimation (LSE) in the reduced model consisting only of active effects. Finally, the Separate Testing of Effects chooses as active those effects that have the maximal Empirical Shannon Information with the output (see [10]). Our Section 2 examines similar and different properties of the estimation, testing and design procedures for the extended second order model including all quadratic terms. Sections 3–6 describe various generalizations of our setting to higher order regression and alternative operability regions.

A performance simulation of our procedures will be carried out later due to temporary unavailability of our [10] coauthors.

2. DISCRIMINATION BETWEEN THE FULL FIRST AND SECOND ORDER MODELS

Let us introduce design ε at support points $x^{(1)}, \dots, x^{(n)}$ with nonnegative weights $p_1, \dots, p_n$ — frequencies of total number N of independent measurements in support points, and a row-vector of outputs-responses $\eta(\theta) = (\eta(x^{(1)}, \theta), \dots, \eta(x^{(n)}, \theta))^T$, where η should be multivariate polinomial in $x \in R^d$ with coefficients formiin parameter $\theta \in R^k$, T is a transposition sign. The measurements' aim is finding a maximum of $\eta(\theta)$. Approximating an unknown response function $\eta(\theta)$ in the whole region by a high-order polynomial is considered inappropriate since the groundbreaking Box and Wilson paper [1]. This is due to an excessive amount of experiments required for getting usually highly correlated and poorly interpretable parameter estimates of complex models. Instead, estimation in a part of the whole region of interest — operability region by a low-order polynomial is recommended.

A sequential experimental strategy of several stages (which we refine a bit in Section 2.2) is introduced in [12]. During the first stage, the response surface is locally fitted as multilinear one yielding an estimates of its gradient and consequent shifting operability regions for carrying out a steepest improvement of a local response model until a stationary region is reached.

Then the second stage of experimentation is initiated to fit the full second order model for the response inside the stationary region for estimating the sensitivity of the response to small parameter variation. Inadequacy of the linear approximation signals about the necessity to the above change of the fitting strategy. For testing adequacy, [1] recommends the complementing the Complete Factorial Design (CFD(d)) used for fitting first order model with several *central points* to enable a reliable estimate for the sum of parameters describing quadratic terms of the response approximation.

This convenient and intuitive recommendation has its drawback: when the above sum of parameters vanishes or is small, the experiments fail to detect inadequacy even if some of the quadratic parameters are large.

Thus, more reliable designs are appropriate for this aim having a guaranteed power of detecting arbitrary quadratic deviation from the multilinear model, if a saddle-type local behavior of the response cannot be excluded beforehand, see Section 3.

Now, we formulate our problem more accurately.

Let the row-vector of measurements $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$ have the Normal distribution $N(\eta(\theta), \sigma^2 diag(p_1^{-1}, \ldots, p_n^{-1}))$. Let η be well approximated by an algebraic or trigonometric polynomial of degree d, but probably, a polynomial of degree $d-k$ will suffice. It seems reasonable at first to place measurements so as to test the hypothesis about the polynomial degree obtained for subsequent optimal estimation of the polynomial coefficients.

We shall test the hypothesis with the help of the F-test in view of its optimal properties [5]. The F-test is the ratio:

p-weighted sum of squares of deviations of the predicted responses from the total mean m $\int(\hat{\eta}(x,\ \theta) - m))^2 d\varepsilon$

over the residual sum of squares (which estimates the identical experimental variance σ^2). Its simplified version χ^2 is used, if σ^2 is known beforehand. The power of the F- and χ^2-tests is an increasing function of the non-centrality parameter Δ which depends on design ε and the vector $\theta_{(1)}$ of coefficients of monomials of degrees from $d-k+1$ to d:

$$\sigma^2 \Delta(\varepsilon,\ \theta_{(1)}) = \theta_{(1)}^T\ D_1^{-1}(\varepsilon)\ \theta_{(1)}, \tag{1}$$

where $D_1(\varepsilon)$ denotes the diagonal block of a covariance matrix D of the LSE for θ restricted to parameters $\theta_{(1)}$ (see [6, 7]).

We use also an alternative expression:

$$\Delta(\varepsilon,\ \theta_{(1)}) = \sigma^{-2} \int \delta^2(x, \varepsilon^\star, \theta_{(1)}),\ (*)$$

where

$$\delta^2(x, \varepsilon^\star, \theta_{(1)}) = \min_{\tilde{\theta}_{(1)}=0} \int (\eta(x,\ \theta) - \eta(x,\ \tilde{\theta}))^2 d\varepsilon = \int (\eta(x,\ \theta) - \eta_0(x, \varepsilon, \theta))^2 d\varepsilon,$$

and the minimum is attained on the function $\eta_0(x, \varepsilon, \theta)$ in the previous equality.

The normalized matrix D_1 rather than conventional expression ND_1 enables getting rid of parameter N and of the integer effects. Thus, we can consider the convex closed set of arbitrary probability measures upon X as the closure of discrete designs facilitating optimization over designs and apply classical theory of convex sets. Discussion of these notions including representation of arbitrary information matrix D^{-1} as that for discrete design with a support in $\leq A(A+1)/2$ points, where A is the total number of parameters, see e.g. in [6].

When the adequacy of a multilinear model is tested, we choose as alternative the significance of at least one component of $\theta_{(1)}$. Choosing a certain norm $||\theta_{(1)}||$ of vector $\theta_{(1)}$, a design is called maximin, if it maximizes $\inf\limits_{\theta_{(1)}\in\Theta} \Delta(\varepsilon,\ \theta_{(1)})$, where $\Theta = \{\theta_{(1)} :\ ||\theta_{(1)}|| = 1\}$.

2.1. Maximin designs on multi-cube for testing second order model

In this section, maximin designs are studied, when X is an n-dimensional cube: $|x_i| \leq 1,\ \ i = 1, 2, \dots, n$, $\eta(x,\ \theta)$ is a multivariate polynomial of degree d, $d = 1$ or 2, $||\theta_{(1)}||^2$ is the sum of squares of the components of $\theta_{(1)}$.

Now, we obtain some bounds for maximin designs.

The Frobenius formula ([6]) gives

$$D_1^{-1} = M_1 - M_{10}M_0^{-1}M_{01}, \tag{2}$$

where blocks of D^{-1} corresponding to the model of smaller order, and cross-parts of D^{-1} have subscripts respectively $0, 01$ or 10.

Lemma 2.1.1. *If ε is a maximin design, then $\bar{\varepsilon} = \int u\varepsilon d\lambda(u)$ is also maximin, where $\lambda(u)$ is the uniform measure on the group $\{u\} = U$ of all coordinate reshuffles and reflections.*

Proof. As is well known (see, e.g. [6]) $D_1^{-1}(\bar{\varepsilon}) - \int D_1^{-1}(u\varepsilon)d\lambda(u)$ is non-negatively definite, but we obviously have $D_1^{-1}(u\varepsilon) = D_1^{-1}(\varepsilon)$ for all U. Thus, $\theta_{(1)}^T D_1^{-1}(\bar{\varepsilon})\theta_{(1)} \geq \theta_{(1)}^T D_1^{-1}(\varepsilon)\theta_{(1)}$ for all $||\theta_{(1)}||$ and, consequently, $\inf\ \theta_{(1)}^T D_1^{-1}(\bar{\varepsilon})\theta_{(1)} \geq \inf\ \theta_{(1)}^T D_1^{-1}(\varepsilon)\theta_{(1)}$. Thus the statement of the lemma follows.

For any symmetric design ε, its matrix D_1^{-1} contains only three entries

$$a = \int x^2 d\varepsilon, b = \int x^4 d\varepsilon, c = \int x^2y^2 d\varepsilon,$$

where x, y are arbitrary different coordinates of multivariate argument. We deduce from (1) for any symmetric design:

Lemma 2.1.2.

$$D_1^{-1} = diag(B, cI_{d(d-1)/2}).$$

Here $B = (c - a^2)l_d l_d^T, l_d$ is the column vector consisting of d ones and I_r is the identity matrix of order r.

Lemma 2.1.3. *l_d is an eigenvector of B and D_1^{-1} with eigenvalue $\rho_1 = b + (d-1)c - da^2$, $rank(B - dI_d) = 1$ implies that $\rho_2 = b - c$ is an eigenvalue of multiplicity $d - 1$. Finally, $\rho_3 = c$ is an eigenvalue of multiplicity $d(d-1)/2$.*

Thus,

$$\Delta = \min\{c, b - c, b + (d-1)c - da^2\}.$$

Obviously, $a \geq b$ for our case implying inequality

$$\Delta \leq \rho_1 \leq b + (d-1)c - db^2.$$

Putting $b = 1/2 + \mu, c = 1/4 + \nu$, we infer:

$$\Delta \leq \min\{\nu, \mu - \nu, (d-1)(\nu - \mu) - d\mu^2\} \leq 1/4.$$

2.2. The Box's family of designs

The mixture $\varepsilon^* = 1/2(CPD(d) + \delta(0))$, where $\delta(0)$ is atom at 0, has $a = b = c = 1/2$ and does not allow unbiased estimates of θ_{ii}. Nevertheless, the following result holds:

Theorem 2.2. *The minimal non-centrality parameter* $\Delta(\varepsilon^*, \theta_{(1)})$ *over all* $\theta_{(1)} : ||\theta_{(1)}|| = 1$, *is maximal over all designs* ε, *when all coefficients of squares* θ_{ii} *have the same sign.*

Here we use normalization $||\theta||^2 = \sum \theta_{ii}^2 + 2\sum_{i<j} \theta_{ij}^2 = 1$.

Proof. For the mixture ε_p of CFD and the central point with frequency p, we have $a = b = c = p$ and $\lambda_1 = d(p - p^2) \leq d/4, \lambda_2 = 0, \lambda_3 = p$. The optimal p is $1/2$. If all θ_{ii} have the same sign and $\sum \theta_{ii}^2 = 1$, then under $p = 1/2$ we have

$$\sigma^2\delta^2(\varepsilon^*) = (d^{-1}\sum \theta_{ii})^2 d/4 + 2^{-1}\sum_{i<j} \theta ij^2 \geq 4^{-1}\sum_i \theta_{ii}^2 = 1/4,$$

$$\Delta(\varepsilon^*) = (1/4)(\sum \theta_{ii})^2 + \sum_{i<j} \theta_{ij}^2 \geq 1/4.$$

Thus, $\Delta(\varepsilon^*) \geq 1/4$ and we will say: ε^* is a quasi-maximin design.

If the quadratic response $\eta(\cdot)$ has a maximum, then D_1^{-1} is non-negative definite and quadratic form $1_i^T D_1^{-1} 1_i \geq 0, i = 1, \ldots, d$, where 1_i is the column-vector of i-indicator. Thus, all coefficients of squares θ_{ii} have the same sign, and $\Delta(\varepsilon^*)$ is not worse than the maximin noncentrality parameter for any design.

2.3. Intuitive sequential procedure

Omitting discussion of tricky multi-decision problems as in [6, 7], we formulate an intuitive strategy of testing adequacy of a linear model.

(i) After an initial sample from CFD(d) of size sufficient for finding all active main (linear effects) and initial estimation of the variance of observations, we increase size of this sampling to enable estimation of active interactions. If any active interaction is found, the null hypothesis of linear relationship is rejected. (Notice, that the increment in sample size depends only of sparsity parameter, since the term involving $\log t$ of upper bound has the unity coefficient due to incommensurability assumption on the active parameters).

(ii) If the null hypothesis has not been rejected in the first series [i], we start adding the central point to the design and after reaching $p = 1/2$, sample from ε_0. When the frequency of central points reaches value $1/2 \geq p > 0$, all entries a, b, c in D_1^{-1} are p and we can evaluate the non-centrality Δ and corresponding power for sequential decision on stopping measurements and deciding about adequacy of the linear model.

(iii) If the null hypothesis is rejected, our design is to be enlarged to form the so-called central composite design or another D-optimal design for fitting the non-linear model efficiently.

Remark. An additional fundamental advantage of using ε_p is *conditional independence* of all main effects, interactions and the sum of squares S, when measured at CFD(d) (S is identically equal to one). The main effects, interactions and S are *dependent* because equality to 0 of one of them implies the same for all the rest.

We find in Sections 3–6 (following preprint [9]) maximin designs for the following cases:

(1) X is an n-dimensional cube: $|x_i| \leq 1, \quad i = 1, 2, \ldots, n$, $\eta(x, \ \theta)$ is an algebraic polynomial of arbitrary degree d, $||\theta_{(1)}||^2$ is the sum of squares of the components of $\theta_{(1)}$ which are coefficients of a polynomial of degree $d + 1$.

(2) X is an n-dimensional ball: $\sum_{i=1}^{n} x_i^2 \leq 1$, $\eta(x,\ \theta)$ is an algebraic polynomial of the degree d, $k = 1$ or 2, $||\theta_{(1)}||$ is invariant under rotations.

(3) X is an n-dimensional torus: $|t_i| \leq \pi$, the points $(t_1, t_2, \ldots, t_{m-1}, \pm\pi, t_{m+1}, \ldots, t_n)$ are considered as identical for all $m, x_1, \ldots, x_{m-1}, x_{m+1}, x_n$.

$$\eta(x,\ \theta) = \sum_{0<|\nu|\leq d} [a_\nu \cos\ (\nu^T t) + b_\nu \sin\ (\nu^T t)] + \frac{a_0}{\sqrt{2}} \tag{3}$$

where $\nu = (\nu_1, \ldots, \nu_n)^T$, $0 \leq \nu_i \leq d$, $|\nu| = \sum \nu_i$, $(t_1, \ldots, t_n)^T$, $||\theta_{(1)}||$ is the sum of squares of the components of $\theta_{(1)}$ and the vector $\theta_{(1)}$ includes simultaneously a_ν, b_ν for a certain set A of indices ν, $0 < \nu \leq d$.

3. ONE-DIMENSIONAL REGRESSION

The results of Section 3, some of which are of independent interest, will be applied in the multidimensional case. We begin with trigonometric regression to which the algebraic case will be reduced. The principal role is played by the following simple result on the trigonometric function orthogonality which is easily obtained by summing up the geometric series $\sum_{r=1}^{N} e^{irx}$.

Lemma 3.1.

(1) $\sum_{r=1}^{N} \sin(m2\pi r/N) \equiv 0$, *all* m, N *being integers;*
(2)
$$\sum_{r=1}^{N} \cos(m2\pi r/N) = \begin{cases} 0, & m/N \text{ is not an integer;} \\ N, & m/N \text{ is an integer.} \end{cases}$$

e_N is the uniform design on the circle $\{[-\pi,\ \pi],\ \pm\pi$ are identified$\}$, giving equal weights $\frac{1}{N}$ to all equidistant points t_i, and in particular, e_N^o is such design e_N, that one of the points of e_N^o is 0.

With the help of known trigonometric formulas, lemma 3.2. follows from lemma 3.1.

Lemma 3.2.

(1) *System of functions:* $\sqrt{2}\sin\ rt, \sqrt{2}\cos rt,\ r \leq d$ *is orthonormal on* e_N *when* $N > 2d$.
(2) *System of functions:* $1, \sqrt{2}\sin\ rt, \sqrt{2}\, r \leq d,\ \sqrt{2}\cos\ rt,\ r < d,\ \cos\ dt$, *is orthonormal on* e_{2d}^0.

Maximin designs for multidimensional trigonometric regression will be constructed in section 6. Here we study the even trigonometric regression:

$$\eta(t,\ \theta) = \sum_{r=0}^{d} \theta_r\ \cos\ rt, \qquad 0 \leq t \leq \pi.$$

Introduce design ε_d with equal weight $p_r = \frac{1}{d}$ at points $r\frac{\pi}{d}$, $r = 1, 2, \ldots, d-1$, and weights $p_0 = p_d = \frac{1}{2d}$ at points 0 and π.

If design ε_d is reflected across the origin of coordinates and the points $\pm\pi$ are identified, we obtain design e_{2d}^o and, by applying evenness of the $\cos x$ and point (2) of lemma 3.2. we get the proof of:

Lemma 3.3. *System of functions* $1,\ \sqrt{2}\cos\ rt,\ \ r < d,\ \ \cos\ dt$, *is orthonormal on* ε_d.

The following theorem is valid:

Theorem 3.1. *Design* ε_d *minimizes* $D\hat{\theta}_d$ *and the non-centrality parameter* $\Delta(\varepsilon,\ \theta_d)$, *when testing the hypothesis:* $\theta_d = 0$.

Proof. Since $\Delta = \frac{\theta_d^2}{D\hat{\theta}_d}$, both statements of the theorem are equivalent. Let us prove the second one. We have

$$\Delta(\varepsilon_d,\ \theta_d) = \min_{\tilde{\theta}_0,\dots,\tilde{\theta}_{d-1}} \int (\eta(t,\ \theta) - \sum_{r=0}^{d-1} \tilde{\theta}_r\ \cos\ rt)^2 d\varepsilon_d$$
$$= \min_{\tilde{\theta}_0,\dots,\tilde{\theta}_{d-1}} \int (\sum_{r=0}^{d-1} (\theta_r - \tilde{\theta}_r)^2\ \cos^2 rt + \theta_d^2\ \cos^2 dt)\ d\varepsilon_d$$

due to orthogonality of the system cos rt, $r \le d$ on ε_d. The minimum is attained, when $\tilde{\theta}_r = \theta_r$, it equals $\tilde{\theta}_d^2$; besides,

$$\delta^2(t,\ \varepsilon_d) = \tilde{\theta}_d^2 \cos^2 dt$$

is maximal in each point of the ε_d support enabling application of theorem A.2.1 establishing the optimality of the design.

Let us go over to the algebraic regression

$$\eta(x,\ \theta) = \sum_{r=0}^{d} \theta_r x^r, \qquad |x| \le 1. \tag{4}$$

Using transformation $x = \cos\ t$ of the independent variable and plugging in the expression

$$\cos^r t = 2^{1-r} \cos\ rt + \sum_{u=1}^{\lfloor \frac{r}{2} \rfloor} a_u\ \cos\ (r-2u)t, \tag{5}$$

where a_u are certain constants, yields

$$\eta(x,\ \theta) = \theta_d\ 2^{1-d} \cos\ dt + \theta_{d-1}\ 2^{2-d} \cos\ (d-1)t + \sum_{r=0}^{d-2} b_r \cos\ rt, \tag{6}$$

where b_r are certain linear combinations of the parameters $\theta_r,\ r \le d$. We remind that polynomial $T_r(x)$ which is uniquely defined by the condition

$$T_r(x) = \cos(r\ \arccos\ x), \qquad |x| \le 1, \tag{7}$$

is called the Chebyshev's polynomial. Thus,

$$\eta(x,\ \theta) = \theta_d\ 2^{1-d} T_d(x) + \theta_{d-1}\ 2^{2-d} T_{d-1}(x) + \sum_{r=0}^{d-2} b_r T_r(x). \tag{8}$$

Introducing design ζ_d with equal weights $p_r = \frac{1}{d}$ at points cos $\frac{r\pi}{d}$, $r = 1,\dots,d-1$ and $p_0 = p_d = \frac{1}{2d}$ at points ± 1 obtained from ε_d by transformation $x = \cos t$, we paraphrase the result of lemma 3.2 as follows:

Lemma 3.3. *The system of polynomials* $\sqrt{2}T_r(x),\ r < d,\ T_d(x)$ *is orthonormal on* ζ_d.

Next is the main result of Section 3

Theorem 3.2.

(i) ζ_d *is a maximin design for testing the hypothesis* $\theta_{(1)} = (\theta_d,\ \theta_{d-1})^T = 0$, *when* $||\theta_{(1)}|| = \theta_d^2 + a\theta_{d-1}^2$, $0 < a < 2$.

(ii) ζ_d *minimizes* $D\,\hat{\theta}_d$.

Proof of (ii) immediately follows from equation (7) and theorem 3.1.

(This fact was proved in [6] in different way). Going over to the proof of the first statement and taking into account equation (7) we get:

$$\Delta(\theta_{(1)},\ \zeta_d) \geq \min_{b_0,\ldots,b_{d-2}} \int (2^{1-d}\theta_d T_d(x) + 2^{2-d}\theta_{d-1}T_{d-1}(x) + \sum (b_r - \tilde{b}_r)T_r(x))^2 d\zeta_d.$$

Using lemma 3.3 yields $\Delta \geq 2^{2-2d}\theta_d^2 + 2^{3-2d}\theta_{d-1}^2 \geq 2^{2-2d}(\theta_r^2 + a\theta_{d-1}^2) = 2^{2-2d}$, the equality is attained, iff $\theta_{d-1} = 0$.

In this case, the function $\delta^2(x,\zeta,d) = 2^{2-2d}T_d^2(x)$ attains its maximum $2^{2-2d}T_d^2(x)$ at all points of the supp ζ_d according to theorem A2.1. For $\tilde{\theta}_{(1)} = (1,\ 0)$, we have

$$\Delta(\tilde{\theta}_{(1)},\ \varepsilon) \leq \Delta(\tilde{\theta}_{(1)},\ \zeta_d) \leq \Delta(\theta_{(1)},\ \zeta_d).$$

Consequently, according to equation (4), the pair $\tilde{\theta}_{(1)},\ \zeta_d$ is a saddle pair, and ζ_d is a maximin design.

Remark. For hypothesis $\theta_d = 0$, the normalization condition is equivalent to the following:

$$\max_{|x|\leq 1} |\eta(x,\ \theta) - E\ \hat{\eta}_{d-1}(x,\ \varepsilon)| \geq 2^{1-d}. \tag{9}$$

An interesting problem is the corresponding reformulation of our normalization $\theta_d^2 + a\theta_{d-1}^2 = 1$ in terms of systematic deviation of $\hat{\eta}_{d-2}(x)$ from $\eta(x,\ \theta)$.

4. MULTIDIMENSIONAL ALGEBRAIC REGRESSION ON A CUBE

Let X be an n-dimensional cube: $|x_i| \leq 1,\ i = 1,\ldots,n,\ \theta_{(1)}$ be the column-vector of the coefficients of monomials of degrees from $d-k+1$ to d of algebraic polynomial $\eta(x,\ \theta)$ and $||\theta_{(1)}||$ is an Euclidean norm of $\theta_{(1)}$.

The direct product of designs ε_1 and ε_2 on the sets X_1 and X_2 is the direct product of the corresponding probability measures on set $X_1 \times X_2$.

Denote as $x_\nu,\ \theta_\nu,\ |\nu|$ respectively the monomial $\prod_{i=1}^{n} x_i^{\nu_i}$, corresponding coefficient and $\sum_{i=1}^{n}\nu_i$. Denote the vector $\theta_{(1)}$ with a single non-zero component $\theta_\nu = 1$ as δ_ν; specifically, $\delta_{d,r}$ corresponds to monomial x_i^d.

Theorem 4.1.

(1) *The direct product* ζ_d^n *of designs* ζ_d *of Section 3 is a maximin design for testing hypothesis* $\theta_{(1)} = 0,\ k$ *equals either* 1 *or* 2;

(2) *matrix* $D_1(\zeta_d^n)$ *is diagonal, the variance* $d_\nu = D\hat{\theta}_\nu$ *of the LSE* $\hat{\theta}_\nu$ *for* $\theta_\nu,\ |\nu| > d-k$, *is*

$$d_\nu = \delta^2 \begin{cases} 2^{2-2d}, & \text{if } \nu = \delta_{d,r} \text{ for certain } r \\ \prod_{\nu_\alpha} 2^{1-2\nu_\alpha}, & \text{otherwise;} \end{cases}$$

(3) *Specifically, when* $k = 1,\ d = 2,\ D_1(\zeta_2^n) = 4\delta^2 I,\ \Delta(\zeta_2^n,\ \theta_{(1)}) = (2\delta)^{-2},\ D\hat{\theta}_0 = n+1.$
The determinant of the sub-matrix $\mathcal{D}(\zeta_2^n)$ *corresponding to* θ_0 *and* $\theta_{\alpha\alpha},\ \alpha = 1,\ldots,n,$ *is minimal.*

A generalization of the theorem for normalization $\sum\limits_{|\nu|=d} \theta^2(\nu) + a \sum\limits_{|\nu|=d-1} \theta^2(\nu) = 1, \quad 0 < a < 2,$ is straightforward.

Proof of Theorem 4.1 is a natural generalization of the proof of theorem 3.2.2. Using orthogonality of the products of Chebyshev's polynomials of various variables on ζ_d^n, we find that $(\zeta_d^n,\ \delta_{d,r})$ is a saddle pair; item (3) is directly verified.

Denote $\pi_\nu(x) = \prod\limits_{i=1}^{n} T_{\nu_i}(x_i),\ \nu_+$ is the number of $\nu_i > 0$ in vector ν.

Lemma 4.1. *When* $|\mu| \le d,\ |\nu| \le d$:

$$\int \pi_\mu(x)\pi_\nu(x) d\zeta_d^n = \begin{cases} 0, & \text{if } \mu \neq \nu \\ 1, & \text{if } \mu = \nu = \delta_{d,r},\ r = 1, \dots, n \\ 2^{-\nu_+}, & \text{if } \mu = \nu \neq \delta_{d,r}. \end{cases}$$

Proof. The left hand side is equal to $\prod\limits_{i=1}^{n} \int T_{\mu_i(x_i)} T_{\nu_i(x_i)} d\zeta_d^n(x_i), \zeta_d^n$ is the direct product of ζ_d. Hence, this lemma immediately follows from lemma 3.2.3.

For definiteness, we shall restrict ourselves to a more complicated case $k = 2$.

Lemma 4.2. *When* $d - 1 \le |\nu| \le d,\quad \delta(x, \zeta_d^n, \delta_\nu) = a_\nu \pi_\nu(x),\ a_\nu = \prod\limits_{\nu_i > 0} 2^{1-\nu_i}$.

Proof. According to equation (5), it is necessary to prove that

$$\min_{\eta_{d-2}} \int (x^\nu - \eta_{d-2}(x))^2 d\zeta_d^n = a_\nu^2 \int \pi_\nu^2 d\zeta_d^n. \tag{10}$$

Transforming multipliers $x_i^{\mu_i}$ in accordance with equation (9) and using the fact that degrees of the monomials in expansion of $T_n(x)$ have the equal evenness, we reduce the first sub-integral expression to the following form :

$$x_\nu - \eta_{d-2}(x) = \prod_{\nu_i > 0} (2^{1-\nu_i} T_{\nu_i}(x_i)) + \sum_{|\mu| \le d-2} b_\mu \pi_\mu(x)$$

with certain constants b_μ. Hence, according to lemma 4.1, the statement follows from equation (11) and lemma 4.1 with help of equation (2).

Let us prove the diagonality of matrix $D_1(\zeta_d^n)$. For this purpose, we prove

Lemma 4.3. $\Delta(\zeta_d^n,\ \theta_{(1)}) = \sum \theta_\nu^2 a_\nu^2 \int \pi_\nu^2 d\zeta_d^n$.

Proof. Using linear dependence $\tilde\eta(x, \zeta_d^n, \theta)$ and lemma 4.2, we get: $\delta(x, \zeta_d^n, \theta_{(1)}) = \sum \theta_\nu a_\nu \pi_\nu(x)$. Further,

$$\begin{aligned} \Delta(\zeta_d^n,\ \theta_{(1)}) &= \int \delta^2(x, \zeta_d^n, \theta_{(1)}) d\zeta_d^n \\ &= \sum \theta_\nu^2 a_\nu^2 \int \pi_\nu^2 d\zeta_d^n + \sum_{\mu \neq \nu} a_\mu a_\nu \int \pi_\mu \pi_\nu d\zeta_d^n. \end{aligned}$$

The latter sum equals 0 according to lemma 4.1.

From lemmas 4.3 and 4.1, point 2 of theorem 4.1 follows, besides, with any $1 \le r \le n$,

$$\Delta(\zeta_d^n,\ \delta_{d,r}) = \min_{||\theta_{(1)}||=1} \Delta(\zeta_d^n,\ \theta_{(1)}). \tag{11}$$

The supp ζ_d^n belongs to the set

$$A_r = \{x : \delta^2(x, \zeta_d^n, \delta_{d,r}) = T_r^2(x) = \Delta(\zeta_d^n,\ \delta_{d,r})\} \tag{12}$$

for any r.

According to lemma 2.2.1, ζ_d^n maximizes $\Delta(\varepsilon,\ \delta_{d,r})$ i.e.

$$\Delta(\varepsilon,\ \delta_{d,r}) \leq \Delta(\zeta_d^n,\ \delta_{d,r}) \leq \Delta(\zeta_d^n,\ \theta_{(1)}).$$

Consequently, $(\zeta_d^n,\ \delta_{d,r})$ is a saddle pair, and ζ_d^n a maximin design. The support of ζ_d^n coincides with $\cap_{r=1}^n A_r$. It follows that the support of any maximin design belongs to supp ζ_d^n.

5. MULTIDIMENSIONAL ALGEBRAIC REGRESSION ON A BALL.

For the case of $d = 2,\ k = 1$, a family of discrete maximin designs was found in [11] on a ball $X = \{x : \sum x_i^2 \leq 1\}$ for the same norm as in section 4. Its support consists of the vertices of the inscribed cube, of the central point and of the 'star' points at the crossing of the spherical boundary with the coordinates. Instead of generalizing this result for $d > 2$, we investigate maximin designs for a natural case of *invariant under all rotations* (rotatable) norms $||\theta_{(1)}||$. More precisely, we demand that $||\theta_{(1)}||$ coincides for a polynomial η_d and for $\eta_d(ux)$, where u is an arbitrary rotation of X. The examples of such norms are $\max\limits_{x\in X} |\eta_d(x,\ \theta) - \eta_{d-1}(x,\ \mu)|$, or $\sqrt{\int(\eta_d(x,\ \theta) - \tilde{\eta}_{d-1}(x,\ \mu))^2\ d\lambda}$, where λ and μ are certain rotatable measures on X.

The following theorem is true in this general situation.

Theorem 5.1.

(1) *The support of any maximin design for testing hypothesis $\theta_{(1)} = 0$ is contained in the spheres $\sum x_i^2 = r_i^2 \leq 1$, the unit sphere included (in this formulation, for even $d, 1/2$ of a sphere is regarded as the center of the ball).*
(2) *The mixture of uniform distributions upon these spheres with some weights is a maximin design.*

Proof. For rotatable norm $||\theta_{(1)}||$, $u\varepsilon$ obtained by the arbitrary rotation of maximin ε, is also maximin.

Lemma 5.1. *If ε is a maximin design, then $\bar{\varepsilon} = \int u\varepsilon d\lambda(u)$ is also maximin, where $\lambda(u)$ is the Haar probability measure on the orthogonal group.*

Proof. As is well known (see, e.g. [3]) $D_1^{-1}(\bar{\varepsilon}) - \int D_1^{-1}(u\varepsilon)d\lambda(u)$ is non- negatively definite, but we have $D_1^{-1}(u\varepsilon) = D_1^{-1}(\varepsilon)$ for all U. Thus $\theta_{(1)}^T D_1^{-1}(\bar{\varepsilon})\theta_{(1)} \geq \theta_{(1)}^T D_1^{-1}(\varepsilon)\theta_{(1)}$ for all $||\theta_{(1)}||$ and, consequently inf $\theta_{(1)}^T D_1^{-1}(\bar{\varepsilon})\theta_{(1)} \geq$ inf $\theta_{(1)}^T D_1^{-1}(\varepsilon)\theta_{(1)}$. Thus, the statement of the lemma follows.

Maximin design $\bar{\varepsilon}$ is, of course, rotatable. Thus, it follows that the corresponding mixed minimax strategy in θ is rotatable and, consequently, is defined by certain rotatable measure ν:

$$\bar{\theta} = \int (u\theta)d\nu(\theta)).$$

Function $\int \delta^2(x, \bar{\varepsilon}, \nu(\theta))\ d\nu(\theta)$ of lemma 2.2 is, evidently, a function of $r^2 = ||x||^2$ only, besides this function being a polynomial of x of degree $2d$, must be a polynomial P_d of r^2 of degree d. The polynomial P_d does not equal a constant, for its coefficient at the degree d cannot be (according to the definition of δ^2) equal to zero. Consequently, the maximum of P_d upon X can be accepted only upon $\frac{d}{2}$ spheres and at the origin, if d is even, and on $\frac{d+1}{2}$ spheres, if d is odd.

Remark 5.1.The orthogonality of basis functions does not imply their independence on the product design even for the simplest case of basic functions on sphere S^1. Here, the Euclidean coordinates become functions $\sin x$ and $\cos x$ on the circumference. Their joint moment generating function (MGF) $E\exp(s\cos x + \sin x) = E\exp(u(\sin(x+\phi))$ is easily reduced to $E\exp(u\sin x)$, where $u = \tan^{-1}(s/t)$. The derivative of $E\exp(usinx)$ can be expressed as a special function $Struvel[1,u]/u$. This exercise boils down to a conclusion that the joint MGF is not a product of a function of s and a function of t. Random samples from the uniform distributions on spheres S^{d-1} is a sample of independent d-dimensional standard normals divided by the root of sum of their squares. Thus, statistical simulation of above designs seems feasible.

6. TRIGONOMETRIC REGRESSION ON A TORE.

Testing hypothesis is simpler here than in previous sections.

Theorem 6.1.

(1) *For the item* (3)*-model advertised in introduction to Section 3, the direct product* ρ^n *of* n *uniform designs* e_{N_i} *depending on the coordinates* $t_i.i = 1,\dots,n$, *is a maximin design, if* $N_i \geq 2d+1,\ \ i = 1,\dots,n$. *In this case*

(2) $D(\rho^n) = 2I$ *and*

(3) $D_1(\rho^n) = \min\limits_{\varepsilon} |D_1(\varepsilon)|$.

Proof. Let us prove the second statement.

Lemma 6.1. *The system of functions* $1,\ \sqrt{2}\ \sin(\nu^T t),\ \sqrt{2}\ \cos(\nu^T t)$ *is orthonormal on* ρ^n *if* $N_i \geq 2d+1,\ |\nu| \leq d,\ i = 1,\dots,n$.
For definiteness, let us find

$$2\int \sin(\mu^T t)\ \sin(\nu^T t)\ d\rho^n = 2\int \sin(\sum \mu_i t_i)\ \sin(\sum \nu_i t_i)\ \Pi\ d\varepsilon_N(t_i).$$

According to lemma 3.2, point (1), *the integral along* t_i *is* 0, *if* $\mu_i \neq \nu_i$, *if* $\mu = \nu$ *coordinates* $t_j,\ j \neq i$ *being fixed. Thus,* (2) *follows. Further on, applying the lemma just proved, we get:*

$$\sum [D\hat{a}_\nu \cos^2(\nu^T t) + D\hat{b}_\nu \sin^2(\nu^T t)] = 2\sum_A [\cos^2(\nu^T t) + \sin^2(\nu^T t)]$$

is the number of parameters $a_\nu,\ b_\nu,\ \nu \in A$, *i.e. the condition of the truncated D-optimality of the design* ρ^n (A2.1.4) *is fulfilled.*

The proof is over.

Remark 6.1. The orthogonality of basis functions does not imply their independence as random variables on the product design even for simplest one-dimensional case of $\sin x$ and $\cos x$.

REFERENCES

1. Box, G. E. P. and Wilson, K.B. On the Experimental Attainment of Optimum Conditions (with discussion). *Journal of the Royal Statistical Society Series B*, **13**, No.1, 1951, 1-45.

2. Brodsky, S. *Introduction to Factorial Design of Experiment (mathematical foundations).* N.Y.:Manhattan Academia, 2013.

3. Budne T.A. Application of Random Balance Designs. *Technometrics*, **1**, No 2, 1959, 139–155.

4. Demyanov V. F., Malozemov V.N. *Introduction to Minimax* (in Russian) Moscow, Nauka, 1972.

5. Dette H., Haines L.M. and Imhof L.A. Maximin and bayesian designs for regression models, *Statistica Sinica*, 17(2007), 463-480.

6. Fedorov V.V *Theory of Optimal Experiments.* New York. Academic Press 1972.

7. Federov V.V. and Malyutov M.B. Optimal Designs in Regression Problems. Operationsforsch Statist 3, Heft 4 1972, pp 281-308.

8. Malyutov M.B. Search for active inputs of a sparse system: a review, *Springer Lecture Notes in Computer Science*, **7777**, 2012, 478–507.

9. Malyutov M. B. Maximin Designs for Testing Degree of a Polynomial, *Design of Optimal Experiments, ed. by Malyutov M. B.*, Moscow University Press, 1975 (In Russian).

10. Malyutov M. B., Grosu P. and Sadaka H. Separate Testing Inputs vs. Linear Programming relaxation, *Information Processes*, vol. 15, No. 3, 2015, 351-376.

11. Malyutov, M. B. and Myatlev, V.D. Minimax Designs for Testing the adequacy of a linear model upon a cube and a ball. (in Russian), *Moscow Univ. Press, Preprint N 23*, Moscow, 1971

12. Myers R.M., Montgomery D.C. and Anderson-Cook C.M. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 4th Edition, Wiley, 2016.

13. Nalimov V. V. and Chernova N. A. *Statisticheskie metody planirovaniya ekstremal'nyh eksperimentov*, M.: Nauka, 1965 (In Russian); English translation: Vassili Nalimov, *Statistical Methods for Design of Extremal Experiments*, Washington, D.C.: Foreign Technology Div., U.S. Air Force Systems Command, Wright-Patterson AFB, Ohio, USA, 1968.

14. Osborne M.J. *Introduction to Game Theory*, Oxford University Press, New York, Oxford, 2004.

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

A1. UNIQUENESS OF THE INFORMATION MATRIX

As usual in [12], we assume that our design is orthogonal for multilinear model and simplify the expression of the non-centrality parameter under this assumption. Formula (2) implies that

$$\delta^2 = -a^2(\sum \theta_{ii}^2) - a^{-1}\sum_{k=1}^{d}(\sum i < j\theta_{ijk})^2 + \sum_{k\geq m}\sum_{i\geq j}[ijkm],$$

where $[ijk]$ and $[ijkm]$ are respectively third and fourth moments for such designs.

It follows that

$$m(\varepsilon) \leq \min_{i,j} \delta^2(\varepsilon, \theta_{i,j}) = [iijj] - a^{-1\sum_k [ijk]^2},$$

where θ_{ij} corresponds to the function $\eta(x) = x_i x_j, i < j$. The averaged design $\bar{\varepsilon}$ is symmetric with $m(\bar{\varepsilon}, \theta_{ij}) = m(\varepsilon)$ and same a. Taking into account the equality $[iijj]_{\bar{\varepsilon}} = [i\bar{ij}j]_{\varepsilon}$, we conclude that equality $m(\varepsilon) = m(\bar{\varepsilon})$ implies equalities for all k and $i < j$: we have $[iijj] = const$ and $[ijk] = 0$.

Now,

$$m(\varepsilon) \leq \delta^2(\varepsilon, \bar{\theta}) = -da^2 - a^{-1}d^{-1}\sum[iii]_{\varepsilon}^2 + d^{-1}\sum[iiii]_{\varepsilon} + 2d^{-1}\sum_{i<j}[iijj]_{\varepsilon},$$

where $\bar{\theta}$ corresponds to the function $d^{-1}\sum_{i=1}^{d} x_i^2$.

Thus, $[iii]_{\varepsilon} = 0$ and $[iiii]_{\varepsilon} = b$ for all $i = 1, \ldots, d$. As a result, all moments of ε and $[\bar{\varepsilon}$ coincide up to the fourth order meaning that their information matrices coincide.

A2. EQUIVALENCE THEOREMS AND ITERATIVE SEARCH FOR AN OPTIMAL DESIGN

The method used in the main body of our paper was a direct finding of a saddle point. In more general cases only an iterative construction of maximin designs is feasible. The equivalence theorems [5, 7, 15] for a number of optimality criteria might be useful. Let us give the corresponding formulations:

Theorem A2.1. *Design $\varepsilon^\star$ with non-degenerate information matrix is ϕ-optimal for*

1. $\phi(D) = \theta_{(1)}^T D_1^{-1}\theta_{(1)} = \Delta(\varepsilon,\ \theta_{(1)})$ *(maximization of the non-centrality parameter, for a specified subvector of parameters* $\theta_{(1)}$;
2. $\phi(D) = \min\limits_{\Theta} \Delta(\varepsilon,\ \theta_{(1)})$ *(maximin design);*
3. $\phi(D) = D_{pp} = D\hat{\theta}_p$ *(the best estimate for the parameter* θ_p*, used in Section* 2*)*;
4. $\phi(D) = |D_1|$*, where* D_1 *is a submatrix of* D *represented by a subvector* $\theta_{(1)}$ *of the parameters and* $||D_1|$ *its determinant (truncated D-optimality, used in section* 6*)*;
5. $\phi(D) = \int\ \Delta(\varepsilon,\ \theta_{(1)})\ d\mu(\theta_{(1)})$ *(Bayesian maximization of the weighted non-centrality parameter),* $d\mu$ *is a probability measure on a certain compact* Θ*;*

if respectively:

1. $\max\limits_{x\in X} \delta^2(x, \varepsilon^\star, \theta_{(1)}) = \Delta(\varepsilon^\star,\ \theta_{(1)})$,
   $\delta^2(x, \varepsilon^\star, \theta_{(1)}) = \min\limits_{\tilde{\theta}_{(1)}=0} \int (\eta(x,\ \theta) - \eta(x,\ \tilde{\theta}))^2 d\varepsilon = \int (\eta(x,\ \theta) - \eta_0(x, \varepsilon, \theta))^2 d\varepsilon$, *where on the function* $\eta_0(x, \varepsilon, \theta)$ *the minimum is reached in the previous equality;*
2. $\max\limits_{x\in X} (\max\{\delta^2(x, \varepsilon^\star, \theta_{(1)}) : \Delta(\varepsilon^\star,\ \theta_{(1)}) = \min\limits_{\tilde{\theta}_{(1)}\in\Theta} \Delta(\varepsilon^\star,\ \tilde{\theta}_{(1)}),\ \ \theta_{(1)} \in \Theta\}) = \min\limits_{\Theta} \Delta(\varepsilon^\star,\ \tilde{\theta}_{(1)})$
3. $\max\limits_{x\in X} \left(\sum\limits_{\alpha=1}^{p} D_{\alpha_p}(\varepsilon^\star) f_\alpha(x) \right) = D_{pp}(\varepsilon^\star)$,
4. $\max\limits_{x\in X} f^T(x)\tilde{D}(\varepsilon^\star) f(x) = dim\ \theta_{(1)}$,
5. $\max\limits_{x\in X} \int \delta^2(x, \varepsilon^\star, \theta_{(1)}) d\mu(\theta_{(1)}) = \Delta(\varepsilon^\star,\ \theta_{(1)}) d\mu(\theta_{(1)})$

The support supp $\varepsilon^\star$ *of the measure is contained in the set, where maximum on* X *is reached in the previous equalities.*

The proof of items (1, 3, 4) is described in [6, 7]. Here, we shall outline the proof of items (2,5). The convexity of all functions ϕ as functions of ε follows from the convexity of the matrix D_1^{-1} (e.g. see [6, 7]). The functional ϕ of item [5] is a differentiable function of elements of D in the vicinity of non-degenerate information matrix of $\varepsilon^\star$, there exists a continuous derivative in any direction:

$$\frac{\partial\phi(\varepsilon_\alpha)}{\partial\alpha}|_{\alpha=0} = \min_{\theta_{(1)}\in R(\varepsilon_0)} \frac{\partial\Delta(\varepsilon_\alpha,\ \theta_{(1)})}{\partial\alpha}$$

where $\varepsilon_\alpha = \alpha\varepsilon_0 + (1-\alpha)\varepsilon_1$, $\quad R(\varepsilon_0) = \{\theta_{(1)} : ||\theta_{(1)}|| = 1, \quad \min\limits_{||\tilde{\theta}_{(1)}||=1} \Delta(\varepsilon_0,\ \tilde{\theta}_{(1)}) = \Delta(\varepsilon_0,\ \theta_{(1)})\}$ (see [4], p.233). Thus, we obtain:

$\frac{\partial\phi(\varepsilon_\alpha)}{\partial\alpha}|_{\alpha=0} = \min\limits_{||\theta_{(1)}||=1} \Delta(\varepsilon_0,\ \theta_{(1)}) - \max\{\delta^2(x, \varepsilon, \theta_{(1)}) : ||\theta_{(1)}|| = 1,\ \Delta(\varepsilon_0,\ \theta_{(1)}) = \min\limits_{||\tilde{\theta}_{(1)}||=1} \Delta(\varepsilon_0,\ \tilde{\theta}_{(1)})\}$.

This concludes the proof. Similar arguments for another maximin model see in [5].

The Bayesian criterion function ϕ in (5) is differentiable and

$$\frac{\partial\phi(\varepsilon_\alpha)}{\partial\alpha}|_{\alpha=0} = \int (\Delta(\varepsilon_0,\ \theta_{(1)}) - \int \delta^2(x, \varepsilon, \theta) d\varepsilon_1) d\mu(\theta_{(1)}),$$

since because of Θ_1 compactness, we can differentiate under the integral sign (the integral of a derivative converges uniformly). This concludes the proof. Note that item (2) can be used for constructing numerical iterative search procedure of maximin designs.

A3. MAXIMIN DESIGNS AND GAME THEORY

Our problem can naturally be interpreted in terms of the game theory, considering that one player chooses a design maximizing $\Delta(\varepsilon_0,\ \theta_{(1)})$ and the other one chooses $\theta_{(1)},\ ||\theta_{(1)}|| = 1$, minimizing $\Delta(\varepsilon_0,\ \theta_{(1)})$.
The discussion of basic notions of the game theory can be found for example in [14]. We are especially interested in the notion of a saddle pair of strategies which in our case are maximin design and the distribution ν upon the set Θ_1 for which:

$$\begin{aligned}\int \Delta(\varepsilon^{\star},\ \theta_{(1)})d\nu(\theta_{(1)}) &= \max_{\varepsilon} \min_{\mu(\Theta)=1} \int \Delta(\varepsilon,\ \theta_{(1)})d\mu \\ &= \min_{\mu(\Theta)=1} \max_{\varepsilon} \int \Delta(\varepsilon,\ \theta_{(1)})d\mu. \qquad (13)\end{aligned}$$

In our case, the optimal strategy in ε can be chosen "pure"and not mixed (the distribution $\nu(\theta_{(1)})$ as in the case of strategy in $\theta_{(1)}$) for the following reasons: It is more convenient to assume the first player choosing information matrix $M(\varepsilon)$ of design [3]. In this case the pay-off Δ becomes strictly convex function of $M(\varepsilon)$ which is continuous in both arguments running over compact subsets of finite dimensional spaces, while the set $\mathcal{M} = \{M(\varepsilon)\}$ of the strategies of the first player is convex by applying (1). It follows that the optimal strategy for the first player is the pure strategy $M^{\star}$ presenting $M^{\star} = M(\varepsilon^{\star})$, we obtain the maximin design. Besides, note that to test whether $\varepsilon^{\star},\ \nu(\theta_{(1)})$ is a saddle pair, is in our case sufficient to test that:

$$\int \Delta(\varepsilon,\ \theta_{(1)})d\nu(\theta_{(1)}) \le \int \Delta(\varepsilon^{\star},\ \theta_{(1)})d\nu(\theta_{(1)}) \le \Delta(\varepsilon^{\star},\ \theta_{(1)}). \qquad (14)$$