D-optimal designs
for estimating extremum point
of multivariate quadratic regression model

Andrey Pepelyshev

St.Petersburg State University

Contents

1. Response surface methodology

Response surface methodology is a branch of experimental design.

Box, G. E. P. and Wilson, K.B. (1951) On the Experimental Attainment of Optimum Conditions (with discussion). Journal of the Royal Statistical Society Series B 13(1):1-45.

The purpose is to find the conditions $x_1, \ldots, x_k$ for some output variable to be of maximal value.

$$y = f(x_1, x_2, \ldots, x_k) + \varepsilon$$

Box and Wilson suggest using a first-degree polynomial model to do this. They acknowledge that this model is only an approximation, but use it because such a model is easy to estimate and apply, even when little is known about the process.

An easy way to estimate a first-degree polynomial model is to use a factorial experiment or fractional factorial designs.

1. First stage of response surface methodology

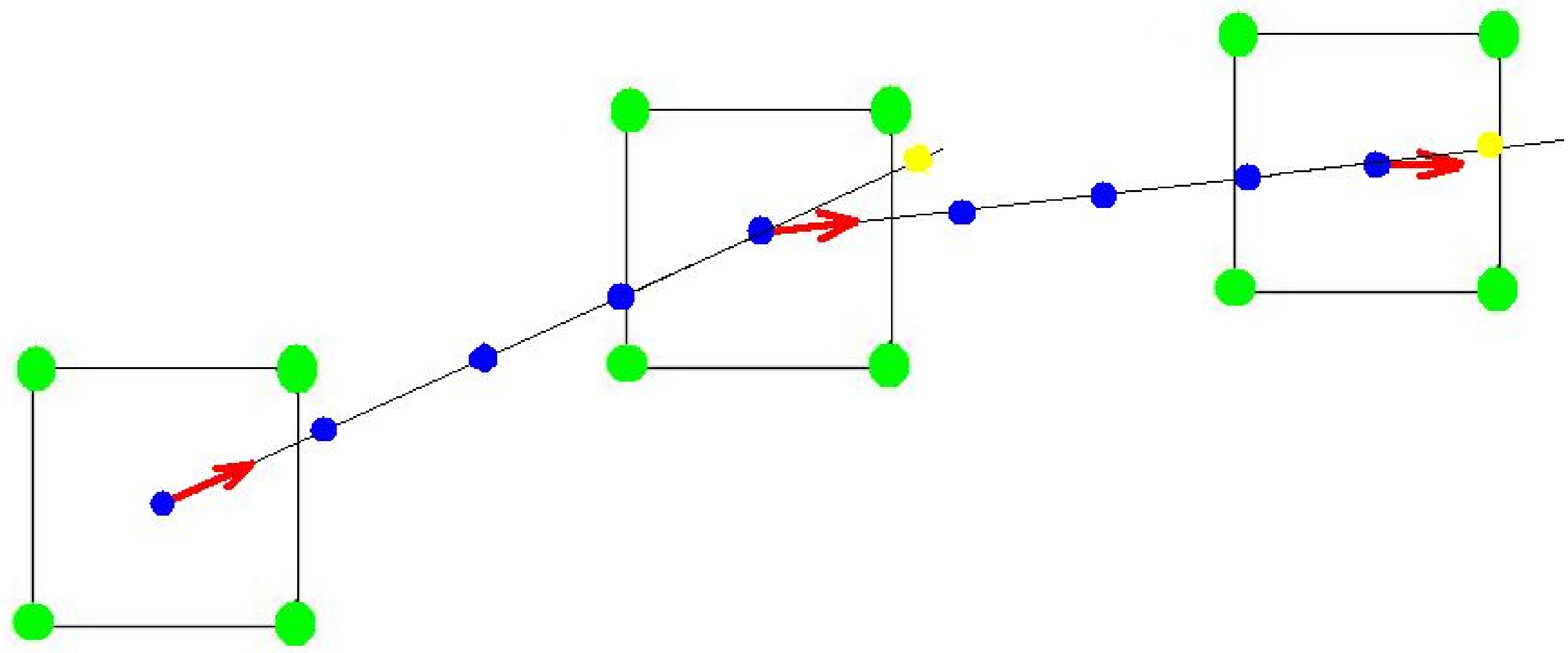


Figure. The points of observation for $k = 2$.

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon$$

This procedure works well when
the current point is far from an extremum point.

1. Second stage of response surface methodology

Let
the current point be close to an extremum point or
an extremum point be outside of design region.

A second-degree polynomial model should be used

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + a_{11} x_1^2 + a_{12} x_1 x_2 + a_{13} x_1 x_3 + \ldots + a_{kk} x_k^2 + \varepsilon$$

- central composite design
- D-optimal design for estimating all parameters
- locally D-optimal design for estimating extremum point

2. Multivariate quadratic regression model

Let the experimental results at the design points $x_{(i)}$, $i = 1, 2, \ldots, n$, $x_{(i)} \in \mathfrak{X}$ be described by the equation

$$y_i = \eta(x_{(i)}, A, \beta, \gamma) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where

$$\eta(x) = \eta(x, A, \beta, \gamma) = x^T A x + \beta^T x + \gamma,$$

A is a positive definite $k \times k$ matrix, β is a k dimensional vector, γ is a real number, $\{\varepsilon_i\}$ are i. i. d. random errors such that $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \sigma^2$, $i = 1, 2, \ldots, n$.

Elements of the matrix A and of the vector β as well as γ are unknown and we should evaluate the extremum point

$$b = \arg\max_x \eta(x, A, \beta, \gamma) = -\frac{1}{2} A^{-1} \beta.$$

2. Multivariate quadratic regression model

Let us rewrite the regression function in the form

$$\bar{\eta}(x, \Theta) = (x - b)^T A (x - b) + c,$$

where

$$\begin{aligned} \Theta &= (b_1, \ldots, b_k, a_{11}, \ldots, a_{kk}, a_{12}, \ldots, a_{1k}, a_{23}, \ldots, a_{k-1k}, c)^T, \\ c &= \gamma - \beta^T A^{-1} \beta / 4. \end{aligned}$$

The asymptotic variance matrix $\mathrm{Cov}\hat{\Theta}$ for the (nonlinear) LSE of Θ is

$$\sigma^2 M^{-1}(\xi)$$

where ξ is a discrete probability measure (experimental design) given by

$$\{x_{(1)}, \ldots, x_{(n)}; \mu_1, \ldots, \mu_n\},$$

$x_{(i)}$ are experimental conditions (design points) and μ_i are proportions of the total number of experiments to be performed at the design points, $\sum \mu_i = 1$. **The goal of the experimental design** is to determine a design ξ which minimizes the determinant of $\mathrm{Cov}\hat{b}$.

$$M(\xi) = \begin{pmatrix} 2A & 0 \\ 0 & I \end{pmatrix} \bar{M}(\xi) \begin{pmatrix} 2A & 0 \\ 0 & I \end{pmatrix},$$

where I is the identify matrix, $f(x) = \partial\bar{\eta}/\partial\Theta$,

$$\begin{aligned} \bar{M}(\xi) &= \bar{M}(\xi, b) = \sum_{l=1}^{n} f(x_{(l)}) f^T(x_{(l)}) \mu_l, \\ f(x) &= f(x, b) = \big((b_1 - x_1), \ldots, (b_k - x_k), (x_1 - b_1)^2, \ldots, (x_k - b_k)^2, \\ &\quad 2(x_1 - b_1)(x_2 - b_2), \ldots, 2(x_{k-1} - b_{k-1})(x_k - b_k), 1\big)^T. \end{aligned}$$

Rewrite the matrix $\bar{M}(\xi)$ in the block form

$$\bar{M}(\xi) = \begin{pmatrix} M_1 & M_2^T \\ M_2 & M_3 \end{pmatrix},$$

where M_1 is a $k \times k$ matrix.

3. Information matrix

Let

$$M_b = M_b(\xi) = M_1 - X^T M_3 X,$$

where $X = M_3^{-1} M_2$ if the matrix M_3 is nonsingular, otherwise X is an arbitrary solution of the equation $M_3 X = M_2$. Then

$$\mathrm{Cov}\hat{b} \approx \frac{\sigma^2}{4} A M_b^{-1}(\xi) A.$$

A design ξ is called locally D-optimal if it maximize the quantity

$$\det M_b(\xi)$$

that corresponds to the known truncated D-criterion.

Our purpose is to find locally optimal designs for any given b for the hypercube $\mathfrak{X} = [-1, 1]^k$ and the hyperball $\mathfrak{X} = \{x; \sum_{i=1}^k x_i^2 \leq 1\}$.

4. The case of line segment, $\mathfrak{X} = [-1, 1]$

Let us consider the following designs

$$(1)\quad \xi = \xi(b) = \begin{cases} \xi_{(1)}, & |b| > \frac{1}{2} \\ \xi_{(2)}, & 0 \le b \le \frac{1}{2} \\ \xi_{(3)}, & -\frac{1}{2} \ge b \ge 0 \end{cases} \quad \textit{where} \quad \begin{array}{rcl} \xi_{(1)} & = & \{-1, 0, 1; \frac{1}{4} - \nu, \frac{1}{2}, \frac{1}{4} + \nu\},\ \nu = \frac{1}{8b}, \\ \xi_{(2)} & = & \{2b - 1, 1; 1/2, 1/2\}, \\ \xi_{(3)} & = & \{-1, 1 + 2b; 1/2, 1/2\}. \end{array}$$

Introduce also the following 1×1 matrices

$$(2)\quad \begin{array}{rcl} M_b^{-1} & = & 16b^2,\ |b| > 1/2, \\ M_b^{-1} & = & 1/(1 - |b|)^2,\ |b| \le 1/2. \end{array} .$$

Theorem. *For the problem at the unit segment there exists a unique locally optimal design. This design has the form (1) and the corresponding values of* M_b^{-1} *are given by the formula (2).*

5. Locally optimal designs on hypercube

Let k be an arbitrary natural number and $b = b_{(0)} \in \mathrm{Int}[0,1]^k$. Consider all hyperparallelepiped with center in the point b and take a maximal one.

Let ξ^* be the experimental design that consists of all vertices of this hyperparallelepiped and equal weights, $m_l = 1/n$, $l = 1, \ldots, n$, $n = 2^k$.

Theorem. *For an arbitrary k the design ξ^* is a locally optimal design for estimation of an extremum point if and only if $|b_i| \le 1/2$, $i = 1, \ldots, k$.*

Let

$$s_k = \left[\det M_b(\xi^*, b)/(\det M_b(\bar{\xi}, b)\right]^{1/k}$$

under a fixed b, where ξ^* is the locally optimal design for estimating the extremum and $\bar{\xi}$ is the usual D-optimal design. Note that the design ξ^* will require s_k times less observations than design $\bar{\xi}$ with the same accuracy.
For $b = (1/2, \ldots, 1/2)^T$ we have

k	1	2	3	4	5
s_k	1.5	1.78	2.08	2.38	2.68

5. Locally optimal designs on hypercube

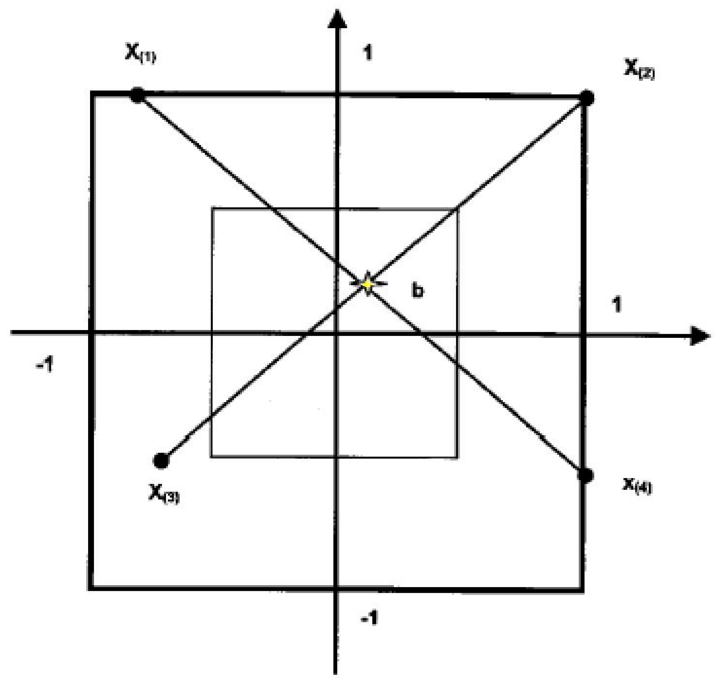


Figure. Points of the design ξ^* for $k = 2$.

The design ξ^* consists of all vertices of maximal inscribed hyperparallelepiped with the center at b and equal weights.

Note that the design ξ^* corresponds to the full factorial design, i.e. its number of points is 2^k. For $k \geq 4$ it is possible to construct a locally optimal design with the number of points less than 2^k.

Let ν be a natural number such that

$$2^{\nu-1} \geq k, \nu \leq k.$$

Theorem. *If $b \in [-1/2, 1/2]^k$ and $k \geq 4$ then there exists a locally optimal design with $n = 2^\nu$ points for the estimation of an extremum point.*

An explicit form of this design is given in (Melas, Pepelyshev, Cheng, 2003).
Let ν be the minimal number satisfying above inequality for a given k, and set $n^*(k) = 2^\nu$. Then
for $k = 5, 6, 7, 8$ we have $n^*(k) = 16$,
for $k = 9, 10, \ldots, 16$ we have $n^*(k) = 32$.

5. Locally optimal designs on hypercube

It is impossible to construct the design explicitly for $b \notin [-1/2, 1/2]^k$.

Теорема. *Suppose that ξ^* be an optimal design. Then there does not exist two points of the design ξ^* which are situated inside some hypercube or some hyperside or some hyperedge.*

Let $k = 2$. Optimal design $\xi_0 = \xi^*(b_{(0)})$ for $b_{(0)} = (\gamma, \gamma)$ equals

$$\xi^*((\gamma,\gamma)) = \begin{pmatrix} (-1,-1) & (-1,0) & (0,-1) & (0,0) \\ \frac{2\gamma-1}{16\gamma} & \frac{2\gamma-1}{16\gamma} & \frac{2\gamma-1}{16\gamma} & 1/4 \\ & (0,1) & (1,0) & (1,1) \\ & \frac{2\gamma+1}{16\gamma} & \frac{2\gamma+1}{16\gamma} & \frac{2\gamma+1}{16\gamma} \end{pmatrix}, \ \gamma \geq 1/2.$$

Optimal design $\xi_0 = \xi^*(b_{(0)})$ for $b_{(0)} = (2, 4)$ equals

$$\xi^*((2,4)) = \begin{pmatrix} (-1,-1) & (-1,-0.0968) & (0.2885,-1) & (-0.0265, 0.0145) \\ 0.1224 & 0.0860 & 0.1242 & 0.2041 \\ & (-0.2990, 1) & (1, 0.1180) & (1,1) \\ & 0.1743 & 0.1140 & 0.1749 \end{pmatrix}.$$

5. Locally optimal designs on square

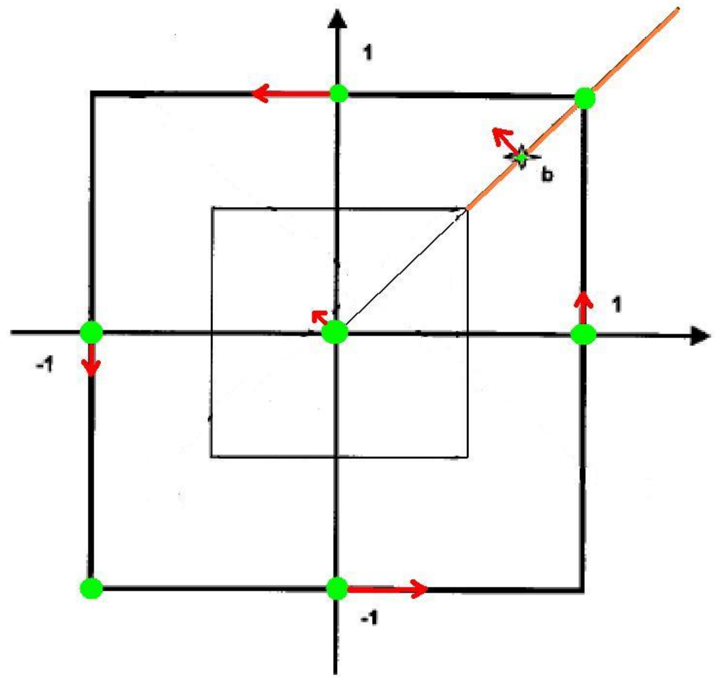


Figure. Points of the design ξ^* for $k = 2$.

Green points are points of design $\xi^*((\gamma, \gamma))$ for $\gamma \geq 1/2$.

6. Locally optimal designs on ball

Let us now consider the problem at the unit circle: $k = 2$,

$$\mathfrak{X} = \{x = (x_1, x_2)^T; x_1^2 + x_2^2 \leq 1\},\ b = (b_1, b_2)^T \in \mathbf{R}^2.$$

Let $\beta = ||b|| = \sqrt{b_1^2 + b_1^2}$, $\tilde{b} = \beta e_1$, $e_1 = (1, 0)^T$, $\beta \geq 0$, $\varphi = \mathrm{arc}\cos(b_1/||b||)$,

$$\tilde{x} = Lx,\ \ L = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}.$$

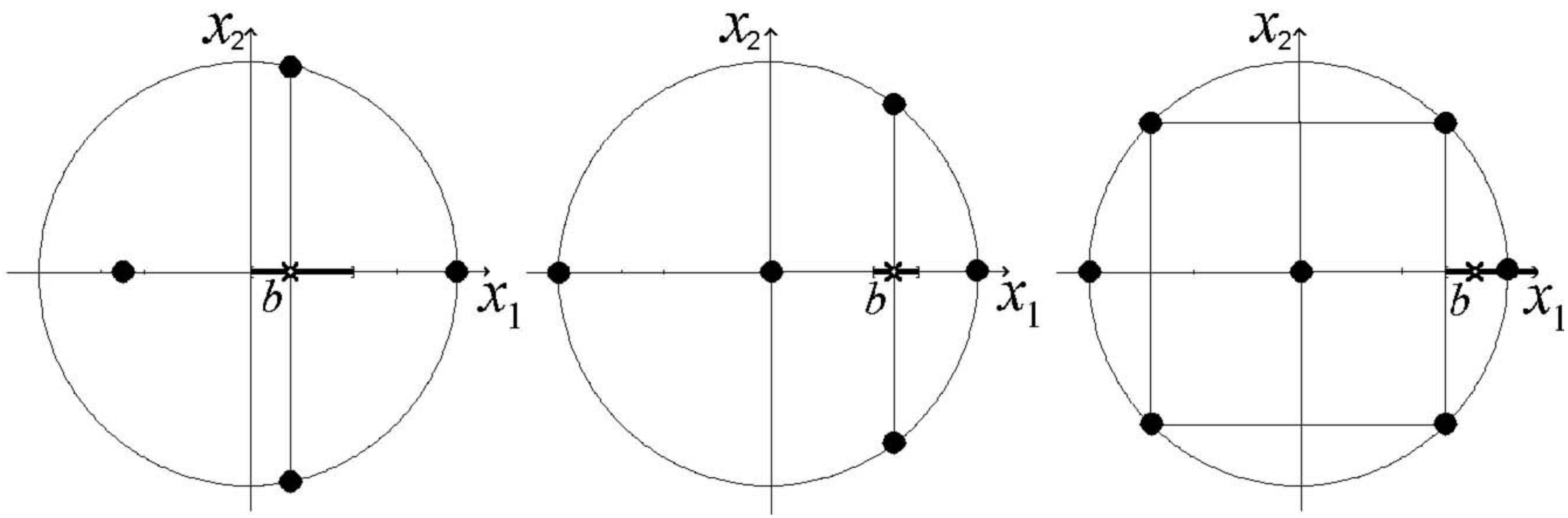
The matrix L is orthogonal and

$$\begin{aligned} \eta(x, \Theta) &= (x - b)^T A (x - b) + c = \\ &= (\tilde{x} - \tilde{b})^T \tilde{A} (\tilde{x} - \tilde{b}) + c, \end{aligned}$$

where $\tilde{A} = LAL^T$, $\det \tilde{A} = \det A$, $\tilde{x} \in \mathfrak{X}$. Thus by rotation of coordinate axis the problem can be reduced to the estimation of the vector of the form $\tilde{b} = \beta e_1$.

6. The theorem and the plot of points of optimal design on ball

Theorem. For the problem at the unit circle with $b = \beta e_1$ locally optimal designs are given by formula



	$\xi_3 = (\xi_{(2)} + \xi_{(4)})/2$	$\xi_2 = (\xi_{(1)} + \xi_{(4)})/2$	$\xi_1 = (\xi_{(1)} + \xi_{(3)})/2$
$(M_b(\xi))^{-1}$	$\mathrm{diag}\{\frac{2}{(1-\beta)^2}, \frac{2}{1-\beta^2}\}$	$\mathrm{diag}\{32\beta^2, \frac{2}{1-\beta^2}\}$	$\mathrm{diag}\{32\beta^2, 8\beta^2\}$
	$0 \le \beta \le \frac{1}{2}$	$\frac{1}{2} < \beta \le \frac{\sqrt{2}}{2}$	$\beta > \frac{\sqrt{2}}{2}$

Besides, for any fixed $\beta \le \sqrt{2}/2$ the design $\xi(b)$ is the unique locally optimal design. For any fixed $\beta > \sqrt{2}/2$ this design is a unique locally optimal design with the minimal number of points.

6. Locally optimal designs on ball

The designs in the theorem are given by

$$\xi_{(1)} = \{(-1,0),(0,0),(1,0); \frac{1}{4}-\nu, \frac{1}{2}, \frac{1}{4}+\nu\},\ \nu = \frac{1}{8\beta},$$

$$\xi_{(2)} = \{(2b-1,0),(1,0); 1/2, 1/2\},$$

$$\xi_{(3)} = \{(-\sqrt{2}/2, \pm\sqrt{2}/2),(\sqrt{2}/2, \pm\sqrt{2}/2); \frac{1}{4}-\mu. \frac{1}{4}-\mu, \frac{1}{4}+\mu, \frac{1}{4}+\mu\},$$

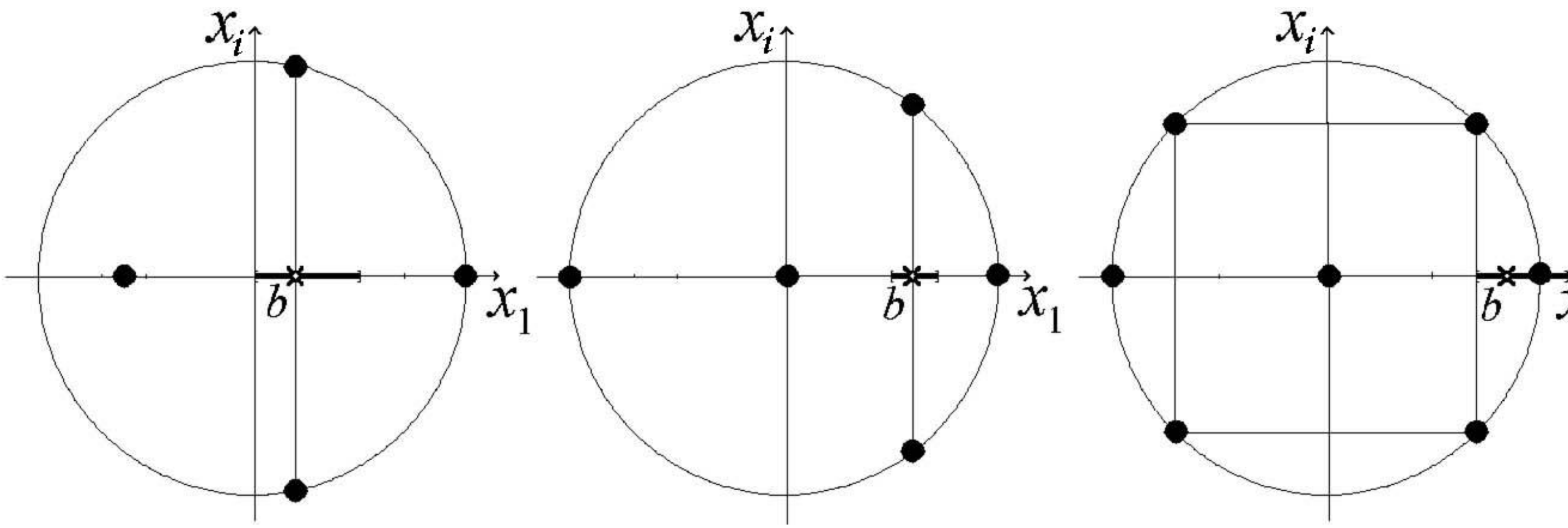
$$\mu = \frac{\sqrt{2}}{8\beta},$$

$$\xi_{(4)} = \{(\beta, -\sqrt{1-\beta^2}),(\beta, \sqrt{1-\beta^2}); \frac{1}{2}, \frac{1}{2}\}.$$

Designs $\xi_{(1)}$ and $\xi_{(2)}$ differ from the optimal designs at the unit segment $[-1,1]$ only by adding the second coordinate x_2 and its value is equal to 0.

6. Locally optimal designs on hyperball

Theorem. For the problem at the unit hyperball with $b = \beta e_1$ locally optimal designs are given by formula



$\xi_3 = \frac{1}{k}\xi_{(2)} + \frac{k-1}{k}\xi_{(4)}$	$\xi_2 = \frac{1}{k}\xi_{(2)} + \frac{k-1}{k}\xi_{(4)}$	$\xi_1 = \frac{1}{k}\xi_{(1)} + \frac{k-1}{k}\xi_{(3)}$
$\mathrm{diag}\{\frac{k}{(1-\beta)^2}, \frac{k}{1-\beta^2}, \ldots, \frac{k}{1-\beta^2}\}$	$\mathrm{diag}\{16k\beta^2, \frac{k}{1-\beta^2}, \ldots, \frac{k}{1-\beta^2}\}$	$\mathrm{diag}\{16k\beta^2, 4k\beta^2, \ldots, 4k\beta^2\}$
$0 \le \beta \le \frac{1}{2}$	$\frac{1}{2} < \beta \le \frac{\sqrt{2}}{2}$	$\beta > \frac{\sqrt{2}}{2}$

For $\beta \le \frac{\sqrt{2}}{2}$ it is the unique optimal design. For $\beta > \frac{\sqrt{2}}{2}$ and $k = 1$ it is the unique optimal design as well. And for $\beta > \frac{\sqrt{2}}{2}$ and $k \ge 2$ this design is the unique optimal design with the minimal number of design points (with accuracy up to an arbitrary rotation of axes $x_2, \ldots, x_k$ if $k > 2$).

The designs in theorem are given by

$$\begin{aligned}
\xi_{(1)} &= \{-e_1, e_0, e_1; \frac{1}{4} - \frac{1}{8\beta}, \frac{1}{2}, \frac{1}{4} + \frac{1}{8\beta}\}, \\
e_1 &= (1, 0, \ldots, 0)^T, \; e_0 = (0, \ldots, 0)^T \in \mathbf{R}^k, \\
\xi_{(2)} &= \{(2\beta - 1)e_1, e_1; 1/2, 1/2\},
\end{aligned}$$

$$\xi_{(3)} = \sum_{i=2}^{k} \xi_{(3)i}/(k-1), \quad \xi_{(4)} = \sum_{i=2}^{k} \xi_{(4)i}/(k-1),$$

$$\begin{aligned}
\xi_{(4)i} &= \{(\beta, \pm\sqrt{1-\beta^2}e_i; 1/2, 1/2\}, \\
i &= 2, 3, \ldots, k-1, \; e_2 = (0, 1, 0, \ldots, 0)^T, \ldots, \; e_k = (0, \ldots, 0, 1)^T, \\
\xi_{(3)2} &= \{(\pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{2}}{2}, 0, \ldots, 0); \frac{1-\mu}{4}, \frac{1-\mu}{4}, \frac{1+\mu}{4}, \frac{1+\mu}{4}\}, \ldots, \\
\xi_{(3)k} &= \{(\pm\frac{\sqrt{2}}{2}, 0, \ldots, 0, \pm\frac{\sqrt{2}}{2}); \frac{1-\mu}{4}, \frac{1-\mu}{4}, \frac{1+\mu}{4}, \frac{1+\mu}{4}\}, \mu = \frac{\sqrt{2}}{2\beta}.
\end{aligned}$$

6. Comparison with a usual D-optimal design

Let

$$s_k = \left[\det M_b(\xi^*, b) / (\det M_b(\bar{\xi}, b)\right]^{1/k}$$

under a fixed b, where ξ^* is the locally optimal design for estimating the extremum and $\bar{\xi}$ is the usual D-optimal design. Note that the design ξ^* will require s_k times less observations than design $\bar{\xi}$ with the same accuracy.

Table. Values s_k for the hyperball, $\beta = ||b||$.

$\beta \backslash k$	2	3	4	5	6
$1/4$	1.38	1.38	1.41	1.45	1.50
$1/2$	1.64	1.80	1.97	2.15	2.33
$1/\sqrt{2}$	1.56	1.80	2.03	2.28	2.52
1	1.38	1.59	1.82	2.06	2.30
2	1.25	1.44	1.66	1.89	2.13

7. Conclusions

- Optimal designs for estimating the extremum for hypercube and hyperball are constructed.
- The advantage of locally optimal designs comparing to usual D-optimal designs consists of the total number of observations with $k \geq 5$ reduced more then two times.
- One more advantage is that locally optimal designs are concentrated in substantially less number of distinct points.

For implementation of locally D-optimal designs a sequential procedure is needed.

The results are published in the papers:

[1] Cheng, R.C.H., Melas, V.B., Pepelyshev, A.N. (2000). Optimal design for evaluation of an extremum point. Optimum Design 2000. Eds. A.Atkinson, B.Bogacka, A.Zhigljavsky. Kluwer, 15-24.

[2] Melas V.B., Pepelyshev A.N., Cheng R.C.H. (2003). Designs for estimating an extremal point of quadratic regression models in a hyperball. Metrika, 58. 193–208.

8. Remarks

Local optimal design is only an ideal.

In practice a design $\alpha\xi_0 + (1-\alpha)\xi^*$ should be used where
ξ_0 is the usual D-optimal design or central composite design,
ξ^* is the local optimal design.

Asymptotically best value of α is $\alpha^*/\sqrt{N}$
where α^* does not depend on N,
N is the total number of observations.