EXPERIMENTAL DESIGN FOR (SEMI-)LOCAL REGRESSION

E. Rafajłowicz* and R. Schwabe†

1 Introduction

The mainstream of the research in the theory of optimal design of experiments for estimating a response function is dedicated to a model based approach, since the existing results strongly rely on the assumption that a parametric structure of the underlying response is a priori exactly known (see [2], [11], [23], [31], [24], besides others, for monographs in this field). Furthermore, in many situations the support of an optimal design is saturated and, hence, there is no possibility to test the correctness of the assumed underlying model structure. On the other hand, even if it is possible to test a model structure and a simpler submodel seems to be adequate, then, as in polynomial regression, it usually occurs that a design, which has been chosen optimally for estimating the parameters in the original model is not optimal for the submodel and may result in unsatisfactory efficiencies (see e. g. [8]). We refer also to [3] for appropriate designs for estimation after a model check.

The above mentioned drawbacks of the model based approach were realized by researchers and practioners already a long time ago, but possible remedies are discussed in a small number of papers only. In [16] and [1] a combination of partial criteria for different models is proposed as a global design optimality criterion (see also [8]). Recently, in [29] and [30], it was established that product designs are model robust for certain classes of regression structure in multi–dimensional models, in the sense that a design which is optimal for a larger model is also optimal for suitable submodels. Note that in the one–dimensional case the

*Institute of Engineering Cybernetics, Technical University of Wrocław, Wybrzeże Wyspiańskiego 27, PL-50 370 Wrocław, Poland. Main results of this paper have been obtained during a visit of E. Rafajłowicz at Freie Universität Berlin. I would like to express my sincere thanks for warm hospitality and stimulating discussions as well as for the financial support by the *Deutsche Forschungsgemeinschaft* to conduct this research.

†Freie Universität Berlin, Institut für Mathematik I, Arnimalle 2–6, D-14 195 Berlin, Germany. Work partly supported by grant Ku719/2-2 of the *Deutsche Forschungsgemeinschaft*.

trigonometric regression is another well–known example where a simultaneously optimal design exists for different degrees of regression.

On the other extreme, one can find, at least, two model free approaches. The first given by [20] is based on the minimization of the integrated mean squared error of a nonparametric regression estimate (see also [25], where it is shown that it is desirable to use points of optimal quadratures in order to obtain the optimal rate of L_2 convergence in nonparametric regression). The second one proposed by [12] aims at space–filling properties, where the prescription for allocating n points $x_1, x_2, \ldots, x_n$ in a design region $\mathcal{X}$ is as follows: find $x_1^*, x_2^*, \ldots, x_n^*$ for which $\max_i \min_j \|x_i - x_j\|$ is attained (see also [18] and [19] for properties and constructions of those maximin designs). The last approach does not even postulate the explicit existence of a functional relationship between the response and the design variables.

Our aim in the present paper is to discuss an approach to experimental design, which is located between model–based and model–free paradigms, somewhat closer, however, to the traditional model–based approach. More precisely, a regression model is assumed but it is suspected a priori that the model is only approximately correct over the whole region of interest $\mathcal{R}$. In particular, it is expected that the assumed model is not equally adequate in the whole region $\mathcal{R}$ and/or different sets of model parameters are suitable in different subregions of $\mathcal{R}$. The starting point to the proposed approach is the local regression model (see [6], [7], [13], [27] and [10]), which is slightly reinterpreted in the next section. It should be added that the present approach to local regression, although based on models for subdomains, differs from those of [22] and [14]. A further approach can be found in [21].

In the next section we recall properties of local regression weighted by a kernel. Section 3 provides some assumptions while in Section 4 the performance of the estimators is investigated. Locally optimal designs are proposed in Section 5 and, finally, those local solutions are combined to globally optimal designs in Sections 6 and 7.

2 Remarks on local weighted least squares regression

Let (x_i, y_i), $i = 1, 2, \ldots, n$ be a sequence of observations, where the $x_i \in R^d$ denote the settings of the independent variables (design points), while the values y_i are the corresponding observations of an unknown regression function $f : R^d \to R$, i. e.,

$$y_i = f(x_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n. \tag{1}$$

Here, the sequence $\{\varepsilon_i\}$ denotes univariate zero mean random errors, which are assumed to be uncorrelated and homoscedatsic, having the same finite variance

σ^2, say, unknown.

We adopt the approach to local least squares regression as exhibited in [27] with the essential difference that the values $x_1, \ldots, x_n$ of the explanatory variables can be chosen by the experimenter. Thus, the local least squares regression is not treated as a purely nonparametric procedure, here, as will be explained later.

Denote by $\mathcal{R} \subset R^d$ the experimental region where the estimation of the response f is of interest and where observations are available. Let $v(x) = (v_1(x), v_2(x), \ldots, v_r(x))^\top$ be a given vector of continuous functions on R^d which are linearly independent for x in a vicinity of zero satisfying $v_1(x) \equiv 1$, $v_2(0) = \ldots = v_r(0) = 0$. The above conditions hold, e. g., if the components of $v(x)$ are monomials in the components of $x \in R^d$.

Denote by the kernel $K(x)$ a nonnegative function satisfying $\int K(x)\,dx = 1$, $\int xK(x)\,dx = 0$ and $\int x^\top x\,K(x)\,dx < \infty$. Additionally, we assume that K is continuous on its support S which is defined as the closure of $\{x : K(x) > 0\}$. Hence, possible discontinuities of K may occur only on the boundary of the support S.

Standard one–dimensional examples of kernels include the Gaussian kernel

$$K_G(x) = (2\pi)^{-1/2} \exp(-x^2/2)$$

with support $S = R^d$, the uniform kernel

$$K_U(x) = \begin{cases} 1/2 & \text{for } x \in [-1,1] \\ 0 & \text{otherwise} \end{cases},$$

the triangular kernel

$$K_T(x) = \begin{cases} 1-|x| & \text{for } x \in [-1,1] \\ 0 & \text{otherwise} \end{cases}$$

and the quadratic (Epanechnikov) kernel

$$K_E(x) = \begin{cases} (3/4)\,(1-x^2) & \text{for } x \in [-1,1] \\ 0 & \text{otherwise} \end{cases}$$

where the latter three are all supported on the compact standard interval $S = [-1,1]$. Multivariate kernels can easily be generated from the above by either taking tensor products $K(x) = K^{(1)}(x^{(1)}) \cdot \ldots \cdot K^{(d)}(x^{(d)})$ of one–dimensional kernels $K^{(1)}, \ldots, K^{(d)}$, where $x^{(m)}$ denotes the mth component of $x = (x^{(1)}, \ldots, x^{(d)})^\top$, or by substituting a norm $\|x\|$ of x into a one–dimensional kernel $K^{(1)}$, i. e. $K(x) = c_d K^{(1)}(\|x\|)$, with a suitable renormalizing constant c_d to achieve $\int K(x)\,dx = 1$.

Note that for the tensor product kernel the support S is the Cartesian product of the supports $S^{(m)}$ of the one–dimensional kernels $K^{(m)}$ and, in pareticular, S is the standard hypercube $[-1,1]^d$ if all one–dimensional kernels are supported on the standard interval, $S^{(1)} = \ldots = S^{(d)} = [-1,1]$. In contrast, the isotropic kernel $K(x) = cK^{(1)}(\|x\|)$ is supported by the unit ball $S = \{x : \|x\| \leq 1\}$ with respect to the norm $\|\cdot\|$ if $K^{(1)}$ is fintely supported on the standard interval $S^{(1)} = [-1, 1]$.

A rescaled kernel K_h is defined by $K_h(x) = h^{-d}K(x/h)$, where $h > 0$ is a scalar parameter, called the bandwidth of the kernel K_h.

According to the methodology of local least squares regression, an estimate of f at a point $\kappa \in R^d$ is obtained by the minimization of the local measure of fit

$$Q(a, \kappa) = \sum_{i=1}^{n} \left(y_i - a^\top v(x_i - \kappa)\right)^2 K_h(x_i - \kappa). \tag{2}$$

with respect to the parameters $a = (a_1, \ldots, a_r)^\top \in R^r$. A minimizer of $Q(a, \kappa)$ depends on the location κ of interest and will, hence, further be denoted by $\hat{a}(\kappa)$. In a purely nonparametric approach to local least squares regression, it is customary to take the first component of $\hat{a}(\kappa)$, i. e. the local fit $\hat{a}^{(1)}(\kappa) = \hat{a}(\kappa)^\top v(0)$ at κ, as the estimator of $f(\kappa)$. Furthermore, $\hat{a}(\kappa)$ is recalculated for every κ in a domain of interest. This approach tacitly assumes that the model $a(\kappa)^\top v(x - \kappa)$ approximates $f(x)$ badly even in a vicinity of κ. In [9] it was shown that $\hat{a}^{(1)}(\kappa)$ is a consistent estimator of $f(\kappa)$, if f is sufficiently smooth and the bandwidth $h = h_n$ satisfies the growth conditions $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$.

On the other hand, if K is the uniform kernel and h is sufficiently large, such that an interval of size h covers the whole region $\mathcal{R}$ of interest, then the minimization of Q is equivalent to the traditional regression problem, incorporating all observations to calculate only a single estimate of the parameters a which can be used simultaneously for all κ.

The approach to local least squares regression considered here is a compromise between these two methods and it is, therefore, called semi–linear regression which will be explained more precisely in the following sections.

3 Assumptions

As usually, in local least squares regression, we assume that the response $f(\kappa)$ is approximated properly by a finite parametrization $a(\kappa)^\top v(x - \kappa)$ in a vicinity of κ. The main difference between local least squares regression and our approach is that at the present stage of investigations the set of parmaeters $a(\kappa)$ is estimated at a finite number of locations $\kappa_1, \kappa_2, \ldots, \kappa_s$ rather than over the whole region $\mathcal{R}$. From a formal point of view these locations are chosen deliberately by the experimenter, but we intend to impose some restrictions on their choice. Roughly speaking, $\kappa_1, \kappa_2, \ldots, \kappa_s$ should form a regular pattern, which covers the domain $\mathcal{R}$ of interest, while the distances between each site κ_j and its neighbours depend on the bandwidth h. For example, for kernels with a finite support having radius 1, it is reasonable to choose $h < ||\kappa_j - \kappa_i|| \leq 2h$, when κ_i and κ_j are neighbours.

Thus, the choice of the bandwidth h reflects our prior knowledge on the size of the domain in which an approximation $a(\kappa_j)^\top v(x-\kappa_j)$ is appropriate. We do not require that all locations $\kappa_1, \kappa_2, \ldots, \kappa_s$ are chosen in advance. However, if $\kappa_1, \kappa_2, \ldots, \kappa_s$ are selected in a data–driven way, then they should be placed at a number of specified sites only forming a regular pattern, since the locations κ_j will be points where f is estimated and will be, simultaneously, the centers of local experiments.

Next, our aim is to formulate a number of assumptions, which clarify the above statements and try to reflect the experience concerning the correctness of a model.

A1) Let κ be a point where the function f is to be estimated. As a measure of model correctness in the vicinity of κ we employ $K_h(x-\kappa)$ as a function of x.

The choice of the kernel K and the bandwidth h is at our disposal, but we require K to be nonincreasing in any direction starting from zero which is satisfied for both the tensor product kernels and the norm based isotropic kernels if the one–dimensional components are unimodal at zero. Choosing larger values of h and K closer to the uniform kernel (in a vicinity of zero), we specify a larger neighbourhood of κ, where $a(\kappa)^\top v(x-\kappa)$ is expected to be adequate. Note however, that in this paper we specify a constant value of the bandwidth h over the whole region of interest which reflects the believe that the approximation is equally appropriate at all sites $\kappa_1, \kappa_2, \ldots, \kappa_s$.

Define the domains

$$\begin{aligned} D_j = \{x \in \mathcal{R}: \quad & K_h\left(x-\kappa_j\right) > 0, \\ & K_h\left(x-\kappa_j\right) \geq K_h\left(x-\kappa_\ell\right), \ \ell = 1, 2, \ldots, s, \\ & K_h\left(x-\kappa_j\right) > K_h\left(x-\kappa_\ell\right), \ \ell < j \ \}, \end{aligned}$$

$j = 1, 2, \ldots, s$, which result essentially in a partition of the region $\mathcal{R}$.

A2) For a certain set of (unknown) parameters $a(\kappa_j) \in R^r$ the model $a(\kappa_j)^\top v(x-\kappa_j)$ is expected to be an adequate approximation of f in the domain D_j, $j = 1, 2, \ldots, s$.

In other words, the function $\tilde{f}(x) = \sum_{j=1}^s \mathbf{1}_{D_j}(x) a(\kappa_j)^\top v(x-\kappa_j)$ is assumed to be a sufficiently close approximation of $f(x)$, where $\mathbf{1}_D$ denotes the indicator function of the set D. Note that $\tilde{f}$ is allowed to be discontinuous although f may be known to be continuous, which means that $\tilde{f}$ may loose some smoothness properties of f. In fact, if f is sufficiently smooth, then testing the equalities $\hat{a}(\kappa_j) = \hat{a}(\kappa_i)$ in adjacent domains D_i and D_j and re–estimating the parameters we can find a smoother estimate of the response f.

A3) For each κ_j the local parameter vector $a(\kappa_j)$ is estimated by $\hat{a}(\kappa_j)$ which minimizes $Q(a, \kappa_j)$ defined in (2).

A4) Within the domain D_j we accept the locally fitted function $\hat{f}_j(x) = \hat{a}(\kappa_j)^\top v(x-\kappa_j)$ as the estimator of the response $f(x)$, i. e.

$$\hat{f}(x) = \sum_{j=1}^{s} \mathbf{1}_{D_j}(x)\hat{a}(\kappa_j)^\top v(x - \kappa_j). \tag{3}$$

Thus, $\hat{f}$ is the estimator of the approximation $\tilde{f}$ rather than of the response f itself. We consider $\hat{f}$ as a pilot estimate of f for experimental design purposes. After collecting data and testing the model adequacy in subregions, one can obtain more smooth estimates of f, but a more detailed descriptions of such a procedure is beyond the scope of this paper.

To illustrate the above assumptions it is convenient to consider the example of a quadratic regression in one variable. Let $d = 1$, $v(x) = (1, x, x^2)^\top$, $\kappa_1 = -1/2$, $\kappa_2 = 1/2$. Suppose that the uniform kernel K_U and the bandwidth $h = 1/2$ reflects our prior knowledge concerning the correctness of quadratic models in vicinities of κ_1 and κ_2. Thus, $D_1 = [-1, 0]$, $D_2 = (0, 1]$ and $\hat{f}$ is of the form

$$\hat{f}(x) = \hat{a}^{(1)}(\kappa_j) + \hat{a}^{(2)}(\kappa_j)(x - \kappa_j) + \hat{a}^{(3)}(\kappa_j)(w - \kappa_j)^2, \tag{4}$$

i. e. a piecewise quadratic model is assumed. Note that (4) can be interpreted as the second order Taylor expansion in the neighbourhoods of κ_j. Note also that $\{x : K_h(x - \kappa_1) > 0\} \cap \{x : K_h(x - \kappa_2) > 0\} = \{0\}$. Hence, zero is the only common point of observations which is used for both the estimators $\hat{a}(\kappa_1)$ and $\hat{a}(\kappa_2)$ of $a(\kappa_1)$ and $a(\kappa_2)$. If h is increased to 1 then $D_1 = [-3/2, 0]$, $D_2 = (0, 3/2]$, and $\hat{a}(\kappa_1)$ is calculated from observations with x values in $[-3/2, 1/2]$, while $\hat{a}(\kappa_2)$ is based on the design region $[-1/2, 3/2]$. This is admitted by the assumptions A1)–A4) and does not lead to misinterpretation, since $D_1 \cap D_2 = \emptyset$.

4 Performance of the estimators

In [27] asymptotic expressions are given for the variance of $\hat{f}(x)$. Unfortunately, they are not applicable here for the following reasons:

1) Our aim is to consider experimental design problems for a finite number of observations.

2) According to A4), our approach is not purely local. It can be called semi–local (or quasi–local), since we do not use only $\hat{a}^{(1)}(\kappa_j)$ as the estimate of $f(\kappa_j)$, but use also higher order components of $\hat{a}(\kappa)$ for estimating the response f in the whole subdomain D_j. This, in turn, implies that the accuracy of the whole estimation vectors $\hat{a}(\kappa_j)$ is important.

3) It is obvious that a nonparametric approach to the minimization of the asymptotic variance of $\hat{f}(x)$ leads to the trivial design which concentrates all observations at the location x of interest.

Next, consider the matrix

$$M_n(\kappa_j) = \sigma^{-2} \sum_{i=1}^{n} K_h(x_i - \kappa_j)\, v(x_i - \kappa_j)\, v(x_i - \kappa_j)^\top = V_{n,j}^\top \mathcal{K}_j V_{n,j}, \qquad (5)$$

where the weight matrix $\mathcal{K}_j$ is a $n \times n$ diagonal matrix with entries $K_h(x_i - \kappa_j)$, $i = 1, 2, \ldots, n$, on the diagonal, while $V_{n,j} = (v(x_1 - \kappa_j), v(x_2 - \kappa_j), \ldots, v(x_n - \kappa_j))^\top$ is the local design matrix.

Furtheron, we will treat $M_n(\kappa_j)$ as if it were the information matrix of $\hat{a}(\kappa_j)$ and $M_n(\kappa_j)^{-1}$ the covariance matrix of this estimator. This is justifiable by the structure under the assumption of local model correctness. Note, however, that $M_n(\kappa_j)^{-1}$ is only an approximation of the covariance $cov(\hat{a}(\kappa_j))$, which is of the form

$$cov(\hat{a}(\kappa_j)) = \sigma^2 \left(V_{n,j}^\top \mathcal{K}_j V_{n,j}\right)^{-1} \left(V_{n,j}^\top \mathcal{K}_j^2 V_{n,j}\right) \cdot \left(V_{n,j}^\top \mathcal{K}_j V_{n,j}\right)^{-1}, \qquad (6)$$

which can be verified directly or deduced from the results in [27]. It is easy to see that (6) coincides with $M_n(\kappa_j)^{-1}$ for the uniform kernel, up to a multiplicative constant depending on the bandwidth h.

Concerning the notion of experimental design, we will deal with the so–called generalized (or approximate) designs, which are (discrete) probability measures defined on the experimental region $\mathcal{R}$ (see, e. g., [11] and [23] for the definition and interpretations). For a generalized design ξ the normalized information matrix is defined as

$$M(\kappa_j, \xi) = \int v(x - \kappa_j)\, v(x - \kappa_j)^\top K_h(x - \kappa_j)\, \xi(dx). \qquad (7)$$

Note that $M(\kappa_j, \xi) = (\sigma^2/n) M_n(\kappa_j)$ if ξ assigns equal mass $1/n$ to each observational point $x_1, \ldots, x_n$, i. e. replications bear weights proportional to the number of replicates. This indicates that the definition (7) is the appropriate generalization of the relation (5).

For a given compact set $X \subset R^d$ the set of all designs will be denoted by $\Xi(X)$. We remark that, unlike in the classical theory, we will deal with designs over varying domains X.

Note that the following technical assumption

A5) For a given compact set X, the sets $\mathcal{M}(\kappa) = \{M(\kappa, \xi) : \xi \in \Xi(X)\}$ are compact in $R^{r \times r}$ for all $\kappa \in R^d$.

is satisfied by the continuity of K and v on S.

5 Optimal designs for estimation in a single domain

In this section we focus on optimal experimental design problems in the case $s = 1$, i. e. when our aim is to estimate f in a vicinity of a single specified point κ_1. Problems of this kind can be treated in the framework of the classical theory of optimal experimental designs. In particular, under appropriate assumptions concerning the kernel K, one can directly apply the equivalence theorems for finding optimal approximate experimental designs. The only difference is caused by the presence of the kernel K and the bandwidth h, which can influence the positions and the weights of the design points.

In the sequel we will focus on D–optimal designs which aim at minimizing the determinant of the approximate covariance matrix $M(\kappa,\xi)^{-1}$ which is proportional to the volume of the confidence ellipsoid of the parameters under normality assumptions. However, the present ideas can be extended to other design criteria and, in particular, to those based on the approximate variance

$$d(x,\xi) = d_j(x,\xi) = v(x-\kappa_j)^\top M(\kappa_j,\xi)^{-1} v(x-\kappa_j), \quad \text{for } x \in D_j, \tag{8}$$

of the prediction $\hat{f}(x)$, like the G–criterion which aims at the minimization of the maximal variance $\sup_{x\in\mathcal{R}} d(x,\xi)$, or the integrated mean squared error criterion which measures the L_2–distance $E\left(\int_{\mathcal{R}}\left(\hat{f}(x)-\tilde{f}(x)\right)^2\,dx\right) = \int_{\mathcal{R}} d(x,\xi)\,dx$.

5.1 Exponential kernels

Consider a polynomial regression model of order $r-1$ in one factor and suppose that we are interested in the accurate estimation of $f(x)$ in a vicinity of $\kappa_1 = 0$.

For the Gaussian kernel K_G Karlin and Studden [15] gave explicit characterizations of the D–optimal experimental design, treating the exponential weights as heterogeneous variances of the observations. The D–optimal design is supported on the roots $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_r$ of the rth degree Hermite polynomials with equally distributed weights $1/r$.

In order to re–interpret the results of Karlin and Studden in terms of a local regression near zero, note that it suffices to substitute $z = \frac{x}{h}$, say, in (7) and recall the equivariance of the D–optimality criterion. This leads to the conclusion that the D–optimal experimental design is supported on the points $h\hat{x}_i$, $i = 1, 2, \ldots, r$, with uniform weights. This simple interpretation indicates the role of h in shrinking the design points toward $\kappa_1 = 0$ as our uncertainty increases concerning the domain of model correctness. Note also that the above interpretation of the results by Karlin and Studden seems to be more adequate than the assumption that the heterogeneous error variance follows an exponential decay in practice.

Next we briefly consider the multi–dimensional case $d > 1$. The multi–dimensional isotropic Gaussian kernel is defined by $K(x) = c_d K_G(\|x\|_2)$, where $\|\cdot\|_2$ denotes the Euclidean norm and the constant c_d depends on the dimension. This kernel can be also written as a tensor product $K(x) = \prod_{m=1}^{d} K_G(x^{(m)})$. If we expect that in a vicinity of $\kappa_1 = 0$ a regression function is spanned by the Kronecker products of univariate polynomials, then the D–optimal design can be obtained as the product of the optimal one–dimensional designs introduced above. We do not give details here, since they can easily be reconstructed from the results in [26] and [28].

5.2 Kernels with finite support

To exhibit to which extent the presence of the kernel influences the locally optimal designs we consider the following simple examples.

Example 1 *Consider a one–dimensional locally linear regression in a vicinity of* $\kappa_1 = 0$, *i. e.* $\tilde{f}(x) = a_1 + a_2 x$. *Let* $h = 1$ *and the kernel be of the form*

$$K_\alpha(x) \quad = \quad \begin{cases} c(\alpha)\,(1 - |x|^\alpha) & \text{for } x \in [-1, 1] \\ 0 & \text{otherwise} \end{cases}\,, \tag{9}$$

where $\alpha \geq 1$ *and* $c(\alpha) = (\alpha + 1)/2\alpha$ *is the normalizing constant. It is easy to verify that the D–optimal experimental design for estimating* $\tilde{f}$ *is concentrated on* $\pm 1/(\alpha+1)^{1/\alpha}$ *with equal weights* $1/2$. *In particular, for the triangular kernel* $(\alpha = 1)$ *the design is concentrated on* $\pm 1/2$, *while for the Epanechnikov kernel* $(\alpha = 2)$ *the design points are* $\pm 1/\sqrt{3} \approx 0.57$. *As* $\alpha \to \infty$ *the limiting support becomes* ± 1 *in accordance with the fact that* K^α *approaches the uniform kernel for* α *large.*

Example 2 *If we let* $0 < \alpha < 1$ *in Example 1, then the D–optimal design is concentrated on three support points* $x_{1,2} = \pm\,(1 - \alpha/2)^{1/\alpha}$ *and, additionally,* $x_3 = 0$ *with corresponding weights* $p_{1,2} = 1/(4 - 2\alpha)$ *and* $p_3 = (1 - \alpha)/(2 - \alpha)$, *respectively. In particular, for* $\alpha = 1/2$ *we have* $x_{1,2} = \pm 9/16$ *with all weights equal to* $1/3$.

Example 3 *Consider the local linear regression* $\tilde{f}(x) = a_1 + a_2 x^{(1)} + a_3 x^{(2)} + a_4 x^{(1)} x^{(2)}$ *in two factors with interactions. Let* $h = 1$ *and* $K(x) = c(\alpha)^2 (1 - |x^{(1)}|^\alpha)(1 - |x^{(2)}|^\alpha)$, *for* $|x^{(1)}| \leq 1$, $|x^{(2)}| \leq 1$ *and zero otherwise, be the tensor product on the one–dimensional kernels* K_α *considered in the previous examples. Then the D–optimal design is the product of the one–dimensional D–optimal designs obtained in Examples 1 and 2. In particular, for* $\alpha \geq 1$ *the D–optimal design assigns equal weights* $1/4$ *to the points* $(\pm 1/(\alpha + 1)^{1/\alpha}, \pm 1/(\alpha + 1)^{1/\alpha})$.

Example 4 *Consider the local linear regression* $\tilde{f}(x) = a_1 + a_2 x^{(1)} + a_3 x^{(2)}$ *in two factors without interactions. Let* $h = 1$ *and* $K(x) = c(1 - \|x\|_2^\alpha$ *for* $\|x\|_2 \leq 1$ *and zero otherwise, be an isotropic kernel based on the one–dimensional kernels* K_α *considered in Example 1,* $\alpha \geq 1$*. Then standard invariance considerations show that the D–optimal designs are concentrated on the sphere* $\{x : \|x\|_2^\alpha = 4/(4+3\alpha)\}$ *with radius* $(4/(4+3\alpha))^{1/\alpha}$*. In particular, any design is D–optimal which assigns equal weights* $1/k$ *to* $k \geq 3$ *equally spaced points on that sphere, i. e. to the corners of a regular polygon inscribed to the corresponding ball.*

Example 5 *Consider the local polynomial regression of* m*th degree in a vicinity of* $\kappa_1 = 0$ *with* $h = 1$ *and a weighted quadratic kernel*

$$K^\alpha(x) = \begin{cases} c_1(\alpha)\,(1-x^2)^{1+\alpha} & \textit{for } x \in [-1,1] \\ 0 & \textit{otherwise} \end{cases}, \tag{10}$$

where $\alpha \geq -1$ *and* $c_1(\alpha)$ *is a normalizing constant. This situation has been considered in [15] (specialized to our case by imposing a symmetry condition on* K*), where it was shown that the D–optimal design is saturated, i. e. it has minimal support, and equal weights* $1/(m+1)$ *are assigned to the roots of the* $(m+1)$*st Jacobi polynomial with both parameters equal to* α*. In particular, for* $m = 1$ *the D–optimal design is concentrated on* $\pm 1/\sqrt{3+2\alpha}$*, while for* $m = 2$ *it is supported on* $\pm\sqrt{3/(5+2\alpha)}$ *and* 0*.*

6 Experimental design for several domains of interest

Here we discuss experimental design problems which occur when a regression function is estimated in several subdomains.

6.1 The extendability of a design

We confine our attention to kernels with bounded support $\{x : K(x) > 0\}$. Let us suppose that a local generalized design $\xi^\star$, say, is given in the vicinity of $\kappa^\star$. In this subsection we admit that $\xi^\star$ may be any suitable design, for example $\xi^\star$ may be rotatable or it may be locally D–optimal. Let us assume that $\xi^\star$ has a finite support, denoted by $\mathrm{supp}(\xi^\star) = (x_1^\star, \ldots, x_m^\star)$, i. e. $\xi^\star$ is a discrete probability measure concentrated on $x_j^\star$ with corresponding weights $p_j^\star$, $j = 1, 2, \ldots, m$.

Suppose that a certain number of observations have been performed according to the design $\xi^\star$ and one decides to explore the response function f in certain regions close to $\mathrm{supp}(\xi^\star)$, where the direction of further investigation may be chosen based on the estimate $\hat{a}(\kappa^\star)$ which contains information on the partial derivatives of the response f at $\kappa^\star$.

Let $G^\star$ be a neighbourhood of $\kappa^\star$ bearing the design $\xi^\star$, i. e. $\mathrm{supp}(\xi^\star) \subset G^\star$. One of the possible strategies for additional explorations can be described as follows: As additional designs take "copies" of $\xi^\star$ translated to domains surrounding $G^\star$. It is desirable to have "copies" of $\xi^\star$ which have many common points with the support of $\xi^\star$ itself in order to make additional use of measurements already obtained in previous domains. To achieve this aim some restrictions have to be imposed on the geometrical shape of the exploration regions.

Suitable bases for an explorative choice of subdomains are given by tilings of R^d, i. e. by covering the whole space R^d by touching copies of a shifted or reflected initial domain G. To form a tiling all subdomains are assumed to be polyhedra of the same shape, i. e. which can be mapped to each other by affine transformations like translation, rotations and reflections. The simplest examples of tilings are provided by adjacent hypercubes or adjacent simplices. In R^2 another useful tiling is provided by regular hexagons in addition to squares and triangles.

Let $G_1, G_2, \ldots$ form a tiling of R^d and let $\mathcal{T}$ be the set of affine transformations T_{ij} mapping G_i onto G_j, $i, j = 1, 2, \ldots$.

Definition 1 *Let ξ_i be a design in G_i with finite support $\{x_1, \ldots, x_m\} \subset G_i$ and corresponding weights $p_1, \ldots, p_m$ and let the affine transformation $T \in \mathcal{T}$ map G_i onto another of the sets forming the tiling. The design ξ' which is the image of ξ_i under the transformation T, i. e. supp$(\xi') = \{T(x_1), \ldots, T(x_m)\}$ with weights $p'_j = p_j$ assigned to $T(x_j)$, $j = 1, 2, \ldots, mm$ will be called the transformed design (under T). We write $\xi' = T(\xi_i)$ for brevity.*

Definition 2 *Let ξ_i be a design with finite support in G_i We say that ξ_i is extendable under a transformation $T \in \mathcal{T}$ to $G_j = T(G_i)$ adjacent to G_i, if either supp$(\xi_i) \cap$ supp$(T(\xi_i)) = \emptyset$ or the weights attached to the design points in the intersection $G_i \cap T(G_i)$ coincide for both the original design ξ_i and its transformed image $\xi_j = T(\xi_i)$.*

Definition 3 *The design $\xi^\star$ is said to be extendable under $\mathcal{T}$ from $G^\star$ to R^d, if $T(\xi^\star)$ is extendable to all adjacent regions for all transformations $T \in \mathcal{T}$.*

Remark 1 *The above definitions can easily be translated to the case of exact designs if the requirement of equal weights is replaced by the requirement of equal numbers of observations at common points in adjacent regions.*

An example illustrating the notion of extendability is shown in Fig. 1 if only translations are admitted. The following simple result reduces extendability to a local property.

Conjecture 1 *If a design $\xi^\star$ is extendable from $G^\star$ to all sets $T(G^\star)$, which are adjacent to $G^\star$ then $\xi^\star$ is extendable to the whole R^d.*

Below we have a sufficient condition for extendability.

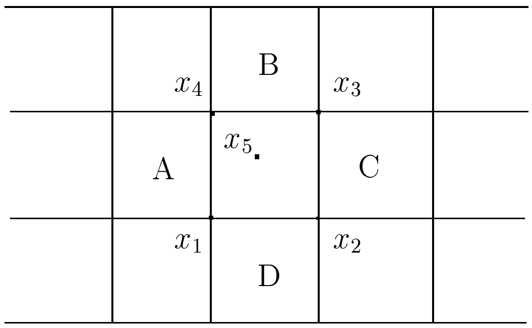

Figure 1: A design which is not extendable under translations to all domains A, B, C or D unless $p_1 = p_2 = p_3 = p_4$. Note that the weight p_5 does not influence the extendability.

Conjecture 2 *Let $\mathcal{T}_\star$ be the set of all affine transformations of $G_\star$ onto itself. If a design $\xi^\star$ is invarinat under each transformation in $\mathcal{T}_\star$, i. e. if $T_\star(\xi^\star) = \xi^\star$ for all $T_\star \in \mathcal{T}_\star$ then this design is extendable to the whole R^d.*

This conjecture has the following interpretation: if $\xi^\star$ has as many symmetries as $G^\star$, then $\xi^\star$ is extendable to R^d. In particular, such a design $\xi^\star$ which is supported by the corners of a regular polyhedron has to assign equal weights to the corners.

The importance of the extendability of a design arises from the fact that having performed observations in one domain $G^\star$ one can use part of the observations when it is decided to extend observations to one of (or more) adjacent domains. The reasons for extending observations can be at least twofold, namely, to test the hypothesis whether $a(\kappa_i) = a(\kappa_j)$ or look for an extremum in the spirit of the evolutionary operation procedure by Box and Draper [4].

It occurs that a number of classical (not necessarily optimal) designs are extendable, the most important example being the full factorial experiments.

6.2 Properties of locally optimal extendable designs

In the previous subsection we started our definition of extendability from a certain design $\xi^\star$, which was not required to be optimal. Now, suppose that our starting design $\xi^\star$ is locally D–optimal for $\kappa^\star = 0$ given the kernel K and the bandwidth h. Then, either

A) the design $\xi^\star$ is extendable in the sense of Definition 3, or

B) extendability of $\xi^\star$ is lacking.

Note that in case B) one may add additional restrictions on the admissibility of designs in order to force extendability. Below we concentrate on case A), since in many problems of theoretical and practical importance D–optimal designs are extendable in a natural way (see the examples in Section 5).

Having a locally D–optimal extendable design $\xi^\star$ and the corresponding $G^\star$ centered at 0, one can choose a collection of affine transformations $T_i \in \mathcal{T}$, $i = 1, 2, \ldots, s$, which extend $\xi^\star$ to neighbouring regions. From a formal point of view the choice of the transformations T_i is arbitrary, but from a practical point of view it is important that the T_i can be selected in a data driven way.

For a given set T_i, $i = 2, 3, \ldots, s$ one can define the centers $\kappa_i = T_i(0)$, the transformed designs $\xi_i^\star = T_i(\xi^\star)$ for the regions $G_i = T_i(G^\star)$, $i = 2, 3, \ldots, s$, which, in turn, specifies the regions $S_h(\kappa_i)$ of interest and the domains D_i. For notational convenience we set $G_1 = G^\star$, $\xi_1^\star = \xi^\star$ and take T_1 as the identity transformation.

In order to compare efficiencies of different designs it is useful to distinguish the following cases.

I. The transformations $\{T_i\}_{i=2}^s$ are chosen in such a way that $S_h(\kappa_i) \cap S_h(\kappa_j) = \emptyset$ for $i \neq j$, what implies:

$$\mathrm{supp}(\xi_i^\star) \cap \mathrm{supp}(\xi_j^\star) = \emptyset, \quad i \neq j. \tag{11}$$

II. The transformations $\{T_i\}_{i=2}^s$ are such that

a) $\mathrm{Int}(G_i) \cap \mathrm{Int}(G_j) = \emptyset$, $i \neq j$, $i, j = 1, 2, \ldots, s$, where $\mathrm{Int}(G)$ denotes the set of interior points of the set G,

b) $\bigcup_{i=1}^s G_i$ forms a tiling of the connected domain $\mathcal{R} \subset R^d$.

III. At least one of the sets $\mathrm{Int}(G_i) \cap \mathrm{supp}(\xi_i^\star) \cap \mathrm{Int}(G_j) \cap \mathrm{supp}(\xi_j^\star)$ is not empty for $i \neq j$.

Case III will be excluded from further considerations, since it would require a different definition of design extendability than adopted here. In fact, situation III means that some local designs have common support points, which are in the interior of the domains G_i. In contrast, situation II allows that adjacent designs have common points of their supports only at the boundary of the domains G_i. This case is schematically shown in Figure 2. On the other extreme, condition I describes the simplest case of completely disjoint design supports and disjoint estimation regions $S_h(\kappa_j)$, $j = 1, 2, \ldots, s$. Case I is considered briefly for completeness and comparisons only, since it does not allow for future use of observations made in a previous region.

Translation invariance of local design optimality. Our aim is to give sufficient conditions for the translation invariance of optimality properties of a starting design $\xi^\star$.

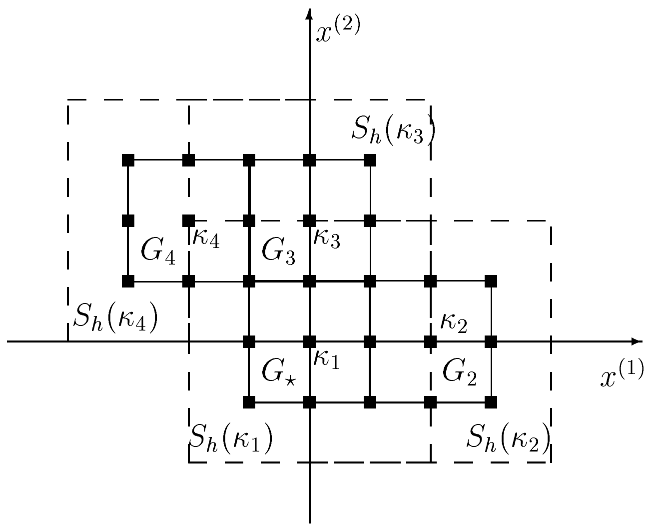

Figure 2: Example of domains and expandable designs, for which condition II holds.

Proposition 1 *Let A1)-A6) hold and for the translations* $\{T_i\}_{i=2}^{s}$ *either condition I or II is fulfilled. If* $\xi^\star$ *is* D*–optimal in* $\Xi(S_h(0))$, $\xi^\star$ *is extendable to* R^d *and the following condition holds*

$$S_h(\kappa_i) \cap supp(\xi_j^\star) = G_i \cap supp(\xi_j^\star), \quad i \neq j, \ i,j = 1,2,\ldots,s, \tag{12}$$

then each transformed design $\xi_i^\star = T_i(\xi^\star)$, $i = 2,3,\ldots,s$, *is also locally* D*–optimal, i. e.* $\xi_i^\star$ *minimizes* $\det M(\kappa_i,\xi)^{-1}$. *Furthermore, all the information matrices* $M(\kappa_i,\xi_i^\star)$, $i = 1,2,\ldots,s$, *coincide.*

Proof. It suffices to notice that the regression function, the kernel and the designs $\xi_i^\star(x) = \xi_1^\star(x-\kappa_i)$ are translation equivariant, resulting in local information matrices satisfying $M(\kappa_i,\xi_i^\star) = M(0,\xi^\star)$. •

Remark 2 *Condition (12) is trivially fulfilled in case I. In case II this condition means that only boundary points of adjacent designs, those at which equality of weights holds, are allowed to enter as data for calculating the estimates and the information matrices in adjacent regions.*

Remark 3 *If* $\mathcal{T}$ *also contains general affine transformations like reflections or rotations rather than merely translations then the situation is more delicate. Note that each affine transformation may be decomposed into a translation and a second component which is a reflection, a rotation or a composition of both. If the design* *problem for the original region $S_h(0)$ is equivariant with respect to all these second components, i. e. the D–optimality of $\xi^\star$ is preserved under these transformations, then, again, the transformed designs $\xi_i^\star$ are locally D–optimal.*

7 Global properties of saturated local designs

In this section we discuss the important special case when the local designs are saturated, i. e. the number of support points equals the number r of locally estimated parameters and equal weights are assigned to each design point. We are concerned with the fundamental question, when locally D–optimal saturated extendable designs can be "glued" together to obtain a globally D–optimal design for the whole region of interest. The answer is confirmative for a local regression spanned by a Tschebyschev system of functions defined on hyperrectangles.

Before formulating particular results it is expedient to state the following remarks.

Let the designs $\xi_i^\star$ be saturated for the corresponding local regressions. Then, the following statements can be directly verified.

S1 Locally weighted least squares estimates of the values of the local response functions at the design points are equal to the arithmetic means of the observations performed at that points. In other words: The local response interpolates the arithmetic means observed at the design points.

S2 If condition II holds, then the values of the local response functions in adjacent regions coincide at common design points.

S2 follows directly from S1 and it can be strengthened, under appropriate assumptions, to contuity of the whole response, if the local response functions are polynomials. Hence, the estimated response is a first degree spline with nodes at the design points. Note, however, that we cannot guarantee continuity of derivatives on the boundary of adjacent subregions.

In stating sufficient conditions for global optimality of extendable designs we concentrate on the one–dimensional case and indicate briefly generalizations to higher dimensions.

Let $f_h(x) = \sqrt{K_h(x)}v(x)$, where $v(x) = (1, x, \ldots, x^m)^\top$. Note that here $r = m + 1$. Let $\xi_1^\star$ be locally D–optimal for estimating the response spanned by $v(x)$ with kernel $K_h(x)$ in $S_h(0)$, i. e. $\kappa_1 = 0$. Assume that $\xi_1^\star$ is saturated and, hence, assigns equal weights $1/r$ to the support points $x_0^\star, x_1^\star, \ldots, x_m^\star$. Sufficient conditions for K, which lead to saturated D–optimal designs for polynomial regressions can be found in [15], where also the following formula, rewritten in our notation, is proved

$$(m+1)^{-1} f_h(x)^\top M(0, \xi_1^\star)^{-1} f_h(x) = K_h(x) \sum_{j=0}^{m} \varrho_j^2(x), \tag{13}$$

where, by definition,

$$\begin{aligned}
\varrho_j(x) &= K_h(x_j^\star)^{-1/2}\ell_j(x), &(14)\\
\ell_j(x) &= \ell(x)\left[\ell'(x_j^\star)(x-x_j^\star)\right]^{-1}, &(15)\\
\ell(x) &= \prod_{j=0}^{m}(x-x_j^\star). &(16)
\end{aligned}$$

Furthermore, ϱ_j, $j=0,\ldots,m$, are the fundamental Lagrange polynomials which have the property

$$\sqrt{K_k(x_i^\star)}\varrho_j(x_i^\star)=\delta_{ij},\quad i,j=1,2,\ldots,m. \tag{17}$$

Note that the polynomials ϱ_j depend on h, but we suppress this in the notation.

Let T_1 be the identity and let $T_2,\ldots,T_s$ denote transformations of $S_h(0)$ forming a tiling of a connected interval for which condition II holds. The support points of the transformed designs $\xi_i^\star=T_i(\xi_1^\star)$, $i=2,3,\ldots,s$, will be denoted by $x_{i,j}^\star$, $j=0,1,\ldots,m$ (with the convention $x_{1,j}^\star=x_j^\star$, $j=0,1,\ldots,m$). The fundamental polynomials, corresponding to the transformed ϱ_j, which are interpolating at $x_{i,j}^\star$, $i=2,3,\ldots,s$, $j=0,1,\ldots,m$, will be denoted by $\varrho_{i,j}(x)$. Analogously, $\varrho_{1,j}(x)$ is identified with $\varrho_j(x)$.

With the above notation, the local regressions can be equivalently expressed as

$$\bar{y}_i(x)=\sqrt{K_h(x-\kappa_i)}\sum_{j=0}^{m}\beta_i^{(j)}\varrho_{i,j}(x),\quad i=1,2,\ldots,s, \tag{18}$$

where the $\beta_i^{(j)}$ denote the unknown local parameters in the vicinity of κ_i, $i=1,2,\ldots,s$.

It is easy to note, but fundamental for further considerations, that $\bar{y}_i(x_{i,m}^\star)=\beta_i^{(m)}$ and $\bar{y}_{i+1}(x_{i+1,0}^\star)=\beta_{i+1}^{(0)}$, what imposes additional constraints on the parameters $\beta_i^{(j)}$.

Lemma 1 *Let the local polynomial regressions and the designs $\xi_i^\star$, $i=1,2,\ldots,s$, be constructed according to the description at the beginning of this subsection. If the designs $\xi_i^\star$ satisfy $x_{i,m}^\star=x_{i+1,0}^\star$, $i=1,2,\ldots,s-1$, then the number of independently estimated parameters among $\left\{\beta_i^{(j)}\right\}_{i=1,2,\ldots,s}^{j=0,1,\ldots,m}$ equals $sm+1$. Furthermore, the following constraints hold*

$$\beta_i^{(m)}=\beta_{i+1}^{(0)},\quad i=1,2,\ldots,s-1. \tag{19}$$

Proof. Note that the designs $\xi_i^\star$, $i=1,2,\ldots,s$ are saturated for estimating the local response. Thus, the estimate $\hat{y}_i^{(j)}$ of $\bar{y}_i(x_{i,j}^\star)$, is the arithmetic mean of all the observations taken at $x_{i,j}^\star$, $i=1,2,\ldots,s$, $j=0,1,\ldots,m$. The assumed equalities

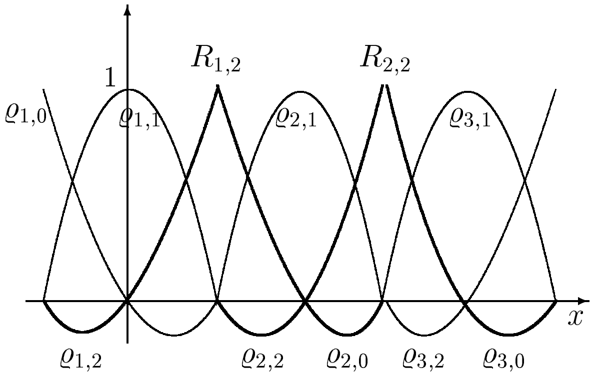

Figure 3: Example to the constructing functions $R_{i,j}$ for quadratic local regressions and the uniform kernel K. Thin curves represent the cases when $R_{i,j}(x) = \varrho_{i,j}(x)$. Thick lines correspond to "glued" curves, defined by (21).

$x^\star_{i,m} = x^\star_{i+1,0}$ imply $\hat{y}_i^{(m)} = \hat{y}_{i+1}^{(0)}$ as the average of the same observations. This, in turn, imposes additional constraints on the local response, namely, $\bar{y}_i(x^\star_{i,m}) = \bar{y}_{i+1}(x^\star_{i+1,0})$, which yields (19) by invoking (17). •

Now, our aim is to obtain a global regression model, which incorporates the constraints (19). To this end, define for $i = 1, 2, \ldots, s-1$

$$R_{i,j}(x) = \sqrt{K_h(x-\kappa_i)}\,\varrho_{i,j}(x), \quad j = 1, 2, \ldots, m-1, \tag{20}$$

$$R_{i,m}(x) = \begin{cases} \sqrt{K_h(x-\kappa_i)}\,\varrho_{i,m}(x) & \text{if } x \le x^\star_{i,m} \\ \sqrt{K_h(x-\kappa_{i+1})}\,\varrho_{i+1,0}(x) & \text{if } x > x^\star_{i,m} = x^\star_{i+1,0} \end{cases} \tag{21}$$

and $R_{1,0}(x) = \sqrt{K_h(x-\kappa_1)}\,\varrho_{1,0}(x)$, $R_{s,j}(x) = \sqrt{K_h(x-\kappa_s)}\,\varrho_{s,j}(x)$ for $j = 1, 2, \ldots, m$. The above construction of the functions $R_{i,j}(x)$ is illustrated in Figure 3 in the case of a quadratic local regression.

Consider a regression function of the form

$$\tilde{y}(x) = \gamma_1^{(0)} R_{1,0}(x) + \sum_{i=1}^{s} \sum_{j=1}^{m} \gamma_i^{(j)} R_{i,j}(x), \tag{22}$$

where the $\gamma_i^{(j)}$ denote the unknown parameters. From the construction of $\tilde{y}(x)$ we immediately obtain

Lemma 2 *The functions of the form (22) and the set of functions $\bar{y}_i$, $i = 1, 2, \ldots, s$, with constraints (19) span the same subspace of continuous functions.*

Define the global design $\xi^\star$ as follows. The set of its support points equals $\bigcup_{i=1}^{s} \mathrm{supp}(\xi_i^\star)$, taking into account that $x_{i,m}^\star = x_{i+1,0}^\star$, $i = 1, 2, \ldots, s-1$. To avoid ambiguity these repeated points are labeled as $x_{i,m}^\star$. Thus, the cardinality of $\mathrm{supp}(\xi^\star)$ equals $q = sm + 1$ and equal weights $1/q$ are assigned to each point of $\mathrm{supp}(\xi^\star)$.

Remark 4 *The above construction is motivated by the expected way of using locally optimal, extendable and saturated designs. Namely, repeating design $\xi_1^\star$, translated to adjacent subregions, it is reasonable to use (approximately) the same number of observations in each of them.*

Now, we are in the position to state the main result of this section.

Proposition 2 *Let $\xi_i^\star$ be defined as in Lemma 1, then $\xi^\star$ is D–optimal for estimating the parameters $\{\gamma_i^{(j)}\}$ in $\tilde{y}(x)$.*

Proof. Let us note that for $R_{i,j}$ the following relationships, analoguous to (17), hold

$$R_{i,j}(x_{ik}^\star) = \delta_{jk}, \quad j, k = 1, 2, \ldots, m. \tag{23}$$

Analoguous relationships hold for $R_{1,0}(x)$. Let $\bar{R}(x)$ be the q–dimensional column vector of the functions $R_{i,j}(x)$. From (23) it follows that for the information matrix $\int_{\mathcal{R}} \bar{R}(x)\bar{R}(x)^\top \xi^\star(dx) = (1/q) I_q$, where I_q is the $q \times q$ identity matrix. Thus, the variance of estimating $\tilde{y}(x)$ equals $var(\tilde{y}(x)) = q\bar{R}(x)^\top \bar{R}(x)$. From the construction of $R_{i,j}$ and (23) it follows that $\sup_x \bar{R}(x)^\top \bar{R}(x) \leq 1$, what finishes the proof by invoking the Kiefer–Wolfowitz equivalence theorem. $\bullet$.

Remark 5 *In fact, proposition 2 establishes the (constrained) D–optimality of $\xi^\star$ for the parameters $\{\beta_i^{(j)}\}$ by the well–known fact that a design remains D–optimal after a reparametrization of the response function.*

Remark 6 *An alternative way of interpreting Proposition 2 is to treat it as a way of finding a D–optimal design for estimating the parameters $\{\gamma_i^{(j)}\}$ in a regression function $\tilde{y}(x)$, which is spanned by finite elements $\{R_{i,j}(x)\}$ (see, e. g., [17] for the definition and basic properties of finite elements).*

Remark 7 *Using results from [26] and [28] one can extend Proposition 2 to the multi–factor case, if the following conditions hold.*
1) The local regressions are of the form $a(\kappa_i)^\top \bigotimes_{j=1}^{d} v^{(k)}(x^{(j)})$, where $\bigotimes_{j=1}^{d}$ denotes the dfold Kronecker product of vectors, while the regression functions $v_j(x^{(j)}) = (1, x^{(j)}, \ldots, (x^{(j)})^{m_j})^\top$ are $(m_j + 1)$–dimensional vectors of monomials in $x^{(j)}$.
2) The kernel K is a tensor product kernel of the form $K(x) = c_d \prod_{j=1}^{d} K^{(j)}(x^{(j)})$ with marginals $K^{(j)}$
3) The local regressions are defined on hypercubes and the union of these hypercubes forms a hyperrectangle.

The above results can serve as a justification of the following simple procedure of explorating a polynomial regression function over a certain region.
Step 1. Start from a locally D–optimal extendable and saturated design. Perform observations and calculate estimates of the regression parameters.
Step 2. Transform the design obtained earlier to one of the adjacent regions and, after collecting observations and calculating estimates of the local regression, test the hypothesis on equality of parameters of local regressions in neighbouring domains.
Step 3. Repeat Step 2 unless all the domains of interest are covered.

Let us note that the above procedure is similar to that of Box and Wilson [5], but the goal of explorations is different, namely, response surface estimation and validation of a model structure in a selected region rather than finding the extremum of the response.

References

[1] Atkinson A. C. and Donev A. N. *Optimum Experimental Designs.* Clarendon, Oxford, 1992.

[2] Bandemer H. et al. *Theorie und Anwendung der optimalen Versuchsplanung I. Handbuch zur Theorie.* Akademie–Verlag, Berlin, 1977.

[3] Benda N. and Schwabe R. Designing experiments for adaptively fitted models. In *MODA 5 - Advances in Model-Oriented Data Analysis and Experimental Design.* (A. C. Atkinson, L. Pronzato and H. P. Wynn, eds.) Physica, Heidelberg, 165–175, 1998.

[4] Box G. E. P. and Draper N. R. *Evolutionary Operation – A Statistical Method for Process Improvement.* Wiley, New York, 1969.

[5] Box G. E. P. and Wilson K. B. On the experimental attainment of optimum conditions. *J. R. Statist. Soc., Ser. B*, 13:1–45, 1951.

[6] Cleveland W. S. Robust locally weigthed regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, 74:998–1004, 1979.

[7] Cleveland W. S. and Devlin S. J. Locally weighted regression: An approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.*, 83:596–610, 1988.

[8] Dette H. and Studden W. J. *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis.* Wiley, New York, 1997.

[9] Fan J. Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.*, 21:196–216, 1993.

[10] Fan J. and Gijbels I. *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London, 1996.

[11] Fedorov V. V. *Theory of Optimal Experiments.* Academic Press, New York, 1972.

[12] Hardin R. H. and Sloane N. J. A. A new approach to the construction of optimal designs. *J. Statist. Plann. Inference*, 37:339–369, 1993.

[13] Hastie T. and Loader C. Local regression: Automatic kernel carpentry. *Statist. Science*, 8:120–143, 1993.

[14] Kaishev V. K. Optimal experimental designs for the B-spline regression. *Comput. Statist. Data Anal.*, 8:39–47, 1989.

[15] Karlin S. and Studden W. J. Optimal experimental designs. *Ann. Math. Statist.*, 37:783–815, 1966.

[16] Läuter E. Experimental design in a class of models. *Math. Operationsforsch. Statist.*, 5:379–398, 1974.

[17] Lankaster P. and Salkauskas K. *Curve and Surface Fitting. An Introduction.* Academic Press, London, 1986.

[18] Mitchell T., Morris M. and Ylvisaker, D. Asymptotically optimum experimental designs for prediction of deterministic functions given derivative information. *J. Statist. Plann. Inference*, 41:377–389, 1994.

[19] Morris M. D. and Mitchell, T. J. Exploratory designs for computational experiments. *J. Statist. Plann. Inference*, 43:381–402, 1995.

[20] Müller H.-G. Optimal designs for nonparametric kernel regression. *Statist. Probab. Letters*, 2:285–290, 1984.

[21] Müller W. G. Optimal design for local fitting. *J. Statist. Plann. Inference*, 55:389–397, 1996.

[22] O'Hagan A. Curve fitting and optimal design for prediction. *J. R. Statist. Soc., Ser. B*, 40:1–42, 1978.

[23] Pázman A. *Foundations of Optimum Experimental Eesign.* Reidel, Dordrecht, 1986.

[24] Pukelsheim F. *Optimal Design of Experiments.* Wiley, New York, 1993.

[25] Rafajłowicz E. Nonparametric orthogonal series estimators of regression: A class attaining the optimal convergence rate in L_2. *Statist. Probab. Letters*, 5:219–224, 1987.

[26] Rafajłowicz E. and Myszka W. Optimum experimental design for a regression on a hypercube – generalization of Hoel's result. *Ann. Inst. Statist. Math.*, 40:821–827, 1988.

[27] Ruppert D. and Wand M. P. Multivariate locally weighted least squares regression. *Ann. Statist.*, 22:1346–1370, 1994.

[28] Schwabe R. *Optimum Designs for Multi–Factor Models. Lecture Notes in Statistics*, Vol. 113. Springer, New York, 1996.

[29] Schwabe R. Model robust experimental design in the presence of interactions: The orthogonal case. In *PROBASTAT'94. Proceedings of the Second International Conference on Mathematical Statistics, Smolenice 1994.* (A. Pázman and V. Witkovský, eds.) *Tatra Mountains Math. Publ.*, 7: 97–104, 1996.

[30] Schwabe R. Optimal designs for hierarchical interaction structures. *J. Statist. Plann. Inference*, 70:181–190, 1998.

[31] Silvey S. D. *Optimal Design.* Chapman and Hall, London, 1980.