On sufficient conditions for implementing the functional approach

Viatcheslav B. Melas
St.Petersburg State University

Abstract

The present talk is devoted to an extension of the functional approach elaborated in the book Melas (2006) for studying optimal designs in linear and nonlinear regression models.

Here we consider locally optimal and Bayesian efficient designs for nonlinear models under standard assumptions about the observational errors.

Let also the following assumptions are fulfilled:

i) **the regression function depends on a scalar variable belonging to the design interval,**

ii) derivatives of the function by the parameters generate an extended Chebyshev system on the design interval,

iii) the matrix of second derivatives of the optimality criterion by the different information matrix elements is positive definite.

Under non-restrictive further suppositions it can be proved that the Jacobi matrix of the system of differentional equations that defines implicitly support points and weight coefficients of the optimal design is invertible.

This allows one to implement the Implicit Function Theorem for representing the points and the weights by Taylor series. The results are illustrating by constructing Bayesian optimal designs for a nonlinear model.

1. Introduction

Optimal design theory has developed rapidly in the last fifty years. Theoretical results are well documented in the book Pukelsheim (2006).

For a more practical guide we refer to the recent monograph of Atkinson, Donev and Tobias (2007).

In these and other books and papers optimal designs are determined as discrete probability measures giving an extremal value to a functional of the information matrix.

Two approaches are commonly used for constructing the optimal designs.

The first one consists in finding an explicit solution of the corresponding extremal problem, perhaps by reducing the problem to one of the classical mathematical problems.

The second, most frequent, approach is merely the numerical construction of optimal designs.

Both approaches have serious disadvantages.

In addition to these two approaches, there is also the "functional approach".

The main idea of this approach is to express **the support points of optimal designs as implicit functions of some auxiliary parameters.**

In many cases these functions, being real and analytical, can be represented by means of power series.
This approach was introduced in Melas (1978).

In a number of papers the functional approach has been applied to linear regression models and, particularly, to models nonlinear in the parameters.

The results are collected in the monograph Melas (2006). Locally optimal designs for the D-, E- and c- criteria were investigated for non-linear models as well as maximin efficient designs for these criteria.

The functional approach was applied to Bayesian D-efficient designs in Melas, Staroselskii (2006).

The present paper provides a further development of the functional approach for nonlinear regression models.

Some sufficient conditions for implementing functional approach for a wider class of optimality criteria are elaborated.

2. Outline of the problem

Let the results of an experiment be described by the regression equation

$$y_j = \eta(x_j, \theta) + \varepsilon_j, \quad j = 1, \ldots, N, \tag{1}$$

where $\eta(x, \theta)$ is a given function with a vector of unknown parameters $\theta = (\theta_1, \ldots, \theta_m)^T$; x_j, $j = 1, \ldots, N$ – are the observation points of the experiment and $\varepsilon_1, \ldots, \varepsilon_N$ are random errors, such that $E\varepsilon_j = 0$, $E\varepsilon_i\varepsilon_j = 0$ $(i \neq j)$, $E\varepsilon_j^2 = \sigma^2 > 0$.

The observational points are chosen in a given interval $[a, b]$ and, as in many cases of practical and theoretical interest, the regression functions are given in an explicit form.

For example,

$$\eta(x,\theta) = \frac{\theta_{(1)}^T f_{(1)}(x)}{\theta_{(2)}^T f_{(2)}(x)}.$$

The regression functions are usually continuously differentiable in the parameters.

Moreover we can suppose that they are real and analytical on the design interval $[a, b]$. Later we will describe some further restrictions which hold in most models of practical usefulness.

Following Kiefer (1974) any discrete probability measure ξ given by

$$\xi = \begin{pmatrix} x_1 \dots x_n \\ w_1 \dots w_n \end{pmatrix},$$

where $x_i \neq x_j$ $(i \neq j)$, $x_i \in \mathcal{X}$, $i = 1, \dots, n$, $w_i > 0$, $\sum w_i = 1$ where $\mathcal{X} = [a, b]$ in our case, will be called an (approximate) design.

We say that the experiment is performed using a design ξ if r_i observations are taken at the points x_i $i = 1, \dots, n$, where $r_i = \lfloor w_i \cdot N \rfloor$ or $r_i = \lfloor w_i \cdot N \rfloor + 1$, in such a way that $\sum r_i = N$ (N is the total number of observations, $\lfloor a \rfloor$ is the integer part of a).

Let $\hat{\theta}(N)$ be a least square estimator of the parameter vector θ with θ^* the true value of θ. It is known (see Jennrich, 1969) that if some regularity conditions hold then the vector $(\hat{\theta}(N) - \theta^*)\sqrt{N}$ has an asymptotically normal distribution with zero mean and covariance matrix

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

where

$$M(\xi, \theta) = \left(\sum_{s=1}^{N} f_i(x_s, \theta) f_i(x_j, \theta) w_s \right)_{i,j=1}^{m}$$

is the information matrix and

$$f_i(x, \theta) = \frac{\partial}{\partial \theta_i} \eta(x, \theta), \;\; i = 1, \ldots, m.$$

Bayesian approach to optimal design for nonlinear models was considered in a number of papers [see Chaloner (1993) and Dette and Neugebauer (1996) among many others].

Here we will consider Bayesian efficient designs.The construction of Bayesian D-efficient designs requires minimization of the quantity

$$\int \left(\frac{\det M(\xi^*(\theta), \theta)}{\det M(\xi, \theta)} \right)^{1/m} dP(\theta),$$

where $\xi^*(\theta)$ is a design maximizing $\det M(\xi, \theta)$ and $P(\theta)$ is a given prior measure.

This quantity shows how many times more measurements are required for the experiment with the design ξ relative to that with the design $\xi^*(\theta^*)$ (the locally D-optimal design for the true parameter values).

Note that the design $\xi^*(\theta^*)$ can be constructed only by means of a sequential approach to the design of experiments.
The Bayesian L-efficient design (here L stands for "linear") is defined as the design minimizing

$$\int \frac{\mathrm{tr} BM^{-1}(\xi, \theta)}{\mathrm{tr} BM^{-1}(\xi^*(\theta), \theta)} dP(\theta),$$

where B is a given positive definite matrix.

A free choice of the matrix B provides designs with a specified ratio of precision of estimation for different parameters.

Bayesian D-efficient designs were investigated for some classes of regression models by means of the functional approach in Melas and Staroselskii (2008). In the present talk we extend this approach to a general class of optimality criteria including the L-criterion.

3. Sufficient conditions

Note that the information matrix for the nonlinear regression model (1) coincides with the information matrix of the model

$$\beta^T f(x, \theta), \tag{2}$$

where $\beta = (\beta_1, \ldots, \beta_m)^T$ is the vector of parameters to be estimated and $f(x, \theta) = (f_1(x, \theta), \ldots, f_m(x, \theta))^T$.

Therefore the problem of constructing locally D-optimal designs for model (1) coincides with the problem of constructing D-optimal design for model (2) with given θ.

This observation is true for any criterion. Therefore we will write

$$f(x) = f(x, \theta)$$

for any fixed value of θ and, for given criteria, we will consider optimal designs for the model

$$\beta^T f(x). \qquad (3)$$

The design ξ is called Φ-optimal for model (3) if it minimizes the quantity

$$\Phi(M(\xi)), M(\xi) = \int f(x) f^T(x) \xi(dx)$$

for the function Φ on the set of $m \times m$ matrices.

We consider optimality criteria possessing the following properties:

1) $\Phi(M) = \infty$, if $\det M = 0$;

2) $\Phi(M)$ is a continuously differentiable function;

   2a) $\Phi(M)$ is a real analytical function;

3) $\Phi(M_1) > \Phi(M_2)$, if $M_1 \geq M_2$ and $M_1 \neq M_2$ (where $M_1 \geq M_2$ means, that the matrix $M_1 - M_2$ is nonnegative definite);

4) $\Phi(M)$ is a strictly convex function of M on the set of all positive definite matrices.

In fact we need the property that is slightly more stronger that 4), namely

4a) the matrix of second derivatives of the optimality criterion by the different information matrix elements is positive definite.

This property can be proved for the class of Φ_α criteria, introduced by Jack Kiefer.

Lemma

Lemma 3.1. *The functions*

$$
\begin{aligned}
\Phi(M) &= (\det M)^{-1/m}, \\
\Phi(M) &= \mathrm{tr} B M^{-1}, \\
\Phi(M) &= (\mathrm{tr} M^{p})^{-1/p},\ 0 < p < \infty,
\end{aligned}
$$

possess the properties 1), 2a), 3) and 4a), where B is a given positive definite matrix.

The proof of this Lemma can be obtained by standard methods. See Pukelsheim (2006) and Melas (2006).

The following generalization of the Kiefer-Wolfowitz equivalence theorem holds (see, e.g., Pukelsheim, 2006).

Theorem

Theorem 3.1. *Let the functions $f_1(x), \ldots, f_m(x)$ be continuous and linearly independent on a compact set $\mathcal{X}$ and let $\Phi(M)$ possess properties 2) and 4). Then*

a) there exists at least one Φ-optimal design;

b) the information matrices for all Φ-optimal designs are the same; any convex combination of optimal designs is an optimal design.

Theorem

c) *a design ξ^* is a Φ-optimal design if and only if*

$$f^T(x)\frac{\partial \Phi(M)}{\partial M}f(x) \le \operatorname{tr}\frac{\partial \Phi(M)}{\partial M}M \qquad (4)$$

for any $x \in \mathcal{X}$, where $M = M(\xi^*)$;

d) *there is equality in (4) at each support point of any Φ-optimal design.*

We say that continuous functions $\psi_0(x), \psi_1(x), \ldots, \psi_k(x)$ on the interval $[a, b]$ form a Chebyshev system on this interval if, for any set of points

$$a \leq x_0 < x_1 < \ldots < x_k \leq b,$$

the inequality

$$\det(\psi_i(x_j))_{i,j=0}^{k} > 0$$

holds.

It is known that a necessary and sufficient condition for one of the systems $\{\psi_0, \ldots, \psi_{k-1}, \psi_k\}$, $\{\psi_0, \psi_1, \ldots, \psi_{k-1}, -\psi_k\}$ to be a Chebyshev system is that the function $\sum_{i=0}^{k} a_i \psi_i(x)$ has no more than k roots, where $a_0, \ldots, a_k$ are arbitrary real numbers not all equal to zero.

A system of functions $\{\psi_0, \psi_1 \ldots, \psi_k\}$ is said to be an extended Chebyshev system if it is a Chebyshev system, each of these functions has continuous derivatives of order k and any linear combination of these functions has no more than k roots when accounting for the multiplicity of roots.

This definition is equivalent to that given in Karlin and Studden, 1966, Chapter 1.

Further we will have an interest in systems that contain the function $\psi_0 \equiv 1$. There exists a constructive method for checking that the system of functions $\psi_0 = 1, \psi_1, \ldots, \psi_k$ is an extended Chebyshev system.

Put

$$
\begin{aligned}
\mathcal{F}_{11} &= \psi_1', \ldots, \mathcal{F}_{1k} = \psi_k', \\
\mathcal{F}_{22} &= \left(\frac{\mathcal{F}_{12}}{\mathcal{F}_{11}}\right)', \ldots, \mathcal{F}_{2k} = \left(\frac{\mathcal{F}_{1k}}{\mathcal{F}_{11}}\right)', \\
\ldots \\
\mathcal{F}_{ll} &= \left(\frac{\mathcal{F}_{l-1,l}}{\mathcal{F}_{l-1,l-1}}\right)', \ldots, \mathcal{F}_{lk} = \left(\frac{\mathcal{F}_{l-1,k}}{\mathcal{F}_{l-1,l-1}}\right)', \\
\ldots \\
\mathcal{F}_{kk} &= \left(\frac{\mathcal{F}_{k-1,k}}{\mathcal{F}_{k-1,k-1}}\right)'.
\end{aligned}
$$

Lemma

Lemma 3.2. *If* $\prod_{i=1}^{k} \mathcal{F}_{ii}(x) > 0$ *as* $x \in [a, b]$, *then the functions* $\psi_0 = 1, \psi_1(x), \ldots, \psi_k(x)$ *form an extended Chebyshev system on* $[a, b]$.

Proof. Suppose that there exist numbers $a_0, \ldots, a_k$, not all equal to zero, such that the function

$$h(x) = \sum_{i=0}^{k} a_i \psi_i(x)$$

has at least $k+1$ roots with account of their multiplicity.

The derivative of any continuous differentiable function has at least one root between any two roots of this function.
Therefore the function

$$h^{'}(x)=\sum_{i=1}^{k}a_i\psi_i^{'}(x)=\sum_{i=1}^{k}a_i\mathcal{F}_{1i}(x)$$

has at least k roots. Since $\mathcal{F}_{11}(x) \neq 0$ as $x \in [a, b]$, we see that the functions $\frac{\mathcal{F}_{12}}{\mathcal{F}_{11}}, \ldots, \frac{\mathcal{F}_{1k}}{\mathcal{F}_{11}}$ are k-times continuously differentiable and the function

$$\sum_{i=2}^{k} a_i \mathcal{F}_{2i}(x)$$

has at least $k-1$ roots.

Following these arguments, in a similar way we obtain that $\mathcal{F}_{kk}(x)$ has at least one root. But this contradicts the conditions of the lemma.

Hence it follows that any linear combination of these functions has less then $k+1$ roots taking into account the order of the root. Then, for any $a \le x_0 < x_1 < \ldots < x_k \le b$ we have

$$\det\left(\psi_i(x_j)\right)_{i,j=0}^{k} \neq 0.$$

It is easy to check that

$$\lim_{x_i \to x} \frac{\det \left(\psi_i(x_j)\right)_{i,j=0}^{k}}{\prod(x_i - x_j)} = \prod_{i=1}^{k} \mathcal{F}_{ii}^{k-i}(x) > 0$$

as $x \in [a, b]$. Therefore the functions $\psi_0 = 1, \psi_1(x), \ldots, \psi_k(x)$ form an extended Chebyshev system on $[a, b]$. This completes the proof.

Note that the conditions of the lemma were used in Yang (2010) for deriving some complete theorems on locally optimal designs that were generalized in Dette and Melas (2011).

Here we will use the Chebyshev property in order to derive sufficient conditions for the uniqueness of a (locally) optimal design.

Now we can formulate and prove the next Theorem.

Let the functions $\psi_0 = 1, \psi_1, \ldots, \psi_k$ be a Chebyshev system on $[a, b]$. By $\mathcal{L}\{g_1, g_2, \ldots, g_s\}$ denote a linear subspace generated by the functions $g_1, g_2, \ldots, g_s$.

Theorem

Theorem 3.2. *Let the functions $f_1(x), \ldots, f_m(x)$ be continuous and linearly independent on $[a, b]$, and let the criterion of optimality $\Phi(M)$ possess the properties 1)–4) of Section 3.*

Theorem

1) If for any $i, j \in 1:m$

$$f_i f_j \in \mathcal{L}\{\psi_1, \ldots, \psi_k\}, k = 2m-2 \text{ or } \ 2m-1,$$

then there exists a unique optimal design. This design has m support points and at least one of them coincides with the left or right boundary of the interval $[a, b]$. If $k = 2m-2$ then two support points coincide with boundaries of $[a, b]$.

Theorem

2) If for any $i, j \in 1 : m$, $(i, j) \neq (l, l)$, where l is some number from $1 : m$,

$$\begin{aligned} f_i f_j &\in \mathcal{L}\{\psi_0, \psi_1, \ldots, \psi_{k-1}\}, \\ f_l^2 &\in \mathcal{L}\{\psi_0, \psi_1, \ldots, \psi_k\}, \\ & k = 2m-2 \text{ or } 2m-1, \end{aligned}$$

then there exists a unique Φ-optimal design. This design has m support points and at least one of them coincides with the left or right boundary of the interval $[a, b]$. If $k = 2m - 2$ then two support points coincide with boundaries of $[a, b]$.

The conditions of this theorem are simpler and at least formally weaker than the corresponding conditions in Theorem 3.1 of Dette and Melas (2011). Note that this theorem can be used to derive the uniqueness of a locally optimal design.

Now we consider an example illustrating the application of Theorem.

Example 3.1. *Rational model*

Example 3.1. Rational model
Consider the regression function

$$\eta(x,\theta) = P(x,\theta_{(1)})/Q(x,\theta_{(2)}),$$

where

$$\begin{aligned} P(x,\theta_{(1)}) &= \theta_1 + \theta_2 x + \ldots + \theta_l x^{l-1}, \\ Q(x,\theta_{(2)}) &= \theta_{l+1} + \theta_{l+2} x + \ldots + \theta_{s+l} x^{s-1} + x^s, \end{aligned}$$

$\theta = (\theta_1,\ldots,\theta_m)^T$, $\theta_{(1)} = (\theta_1,\ldots,\theta_l)^T$,
$\theta_{(2)} = (\theta_{l+1},\ldots,\theta_{l+s})^T$ and $m = l + s$.

Suppose that

$$Q(x, \theta_{(2)}) \neq 0 \ for \ x \in [a, b]$$

and that P/Q is an irreducible fraction.

Denote
$\psi_0 = 1, \ \psi_i(x) = x^i / Q^4(x, \theta_{(2)}), \ i = 1, 2, \ldots, k,$

$$k = 2(l + s) - 2 = 2m - 2.$$

Hence

$$f_i f_j \in \mathcal{L}(\psi_0, \psi_1, \ldots, \psi_k), \quad i, j \in 1 : m.$$

Now suppose that $(Q^4(x, \theta_{(2)}))^{(2m-1)} > 0$ for $x \in [a, b]$.

Then by Lemma 3.2 the functions

$$\tilde{\psi}_i(x) = \psi_i(x) Q^4(x, \theta^{(2)}), \quad i = 0, 1, \ldots, k$$

form a Chebyshev system on $[a, b]$.

Therefore the functions $\psi_i(x)$, $i = 0, \ldots, k$ also form a Chebyshev system on $[a, b]$.

Thus, provided

$$\left(Q^4(x,\theta_{(2)})\right)^{(2m-1)} > 0, \quad x \in [a,b], \tag{5}$$

for any optimality criterion satisfying conditions 1) – 4) there exists a unique Φ-optimal design for the rational function model. This design has m support points, two of which coincide with the boundaries of the interval.

This result is a generalization of a similar result for locally D-optimal designs in He et al. (1996).

As is shown in Dette and Melas (2011), for fulfilment of (5) it is sufficient that all the roots of the polynomial $Q(x)$ be real and lie outside the interval $[a, b]$.

4. The functional approach

The idea of this approach consists in studying the support points of optimal designs as implicit functions of some auxiliary parameters.

Let

$$\xi^* = \xi^*(\theta) = \begin{pmatrix} x_1^* \dots x_n^* \\ w_1^* \dots w_n^* \end{pmatrix},$$

$$a \le x_1^* < \dots < x_n^* \le b$$

be a locally Φ-optimal design on the interval $[a, b]$.

Then, due to the necessary conditions for an extremum point, we have

$$\begin{aligned} \frac{\partial}{\partial x_i}\Phi\left(M(\xi,\theta)\right)|_{\xi=\xi^*} &= 0, \\ \frac{\partial}{\partial w_i}\Phi\left(M(\xi,\theta)\right)|_{\xi=\xi^*} &= 0 \end{aligned}$$

for all the support points of the Φ-optimal design from the interval (a, b) and for all except one weight coefficient.

Let us consider a triple (n_1, n_2, n_3), where n_1 is the number of support points that coincide with the left boundary of the design interval, n_3 is the similar number for the right boundary and n_2 is the number of support points inside the interval.

This triple will be called the **'type of design'**.

Consider the case $n_1 = n_3 = 1$. The other cases are similar. Note that there exists a one-to-one correspondence between the vector

$$\tau = (x_2, \ldots, x_{n-1}, w_1, \ldots, w_{n-1})$$

of dimension $2n - 3$ and the design of type $(1, n-2, 1)$

$$\xi_\tau = \begin{pmatrix} a & x_2 & \ldots & x_{n-1} & b \\ w_1 & w_2 & \ldots & w_{n-1} & w_n \end{pmatrix},$$

where $w_n = 1 - \sum_{i=1}^{n-1} w_i$.

Introduce the notation

$$\begin{aligned} \varphi(\tau,\theta) &= \Phi\left(M(\xi_\tau,\theta)\right), \\ g_i(\tau,\theta) &= \frac{\partial}{\partial \tau_i}\varphi(\tau,\theta),\ \ i=1,\dots,2n-3, \\ G(\tau,\theta) &= \left(g_1(\tau,\theta),\dots,g_{2n-3}(\tau,\theta)\right)^T, \end{aligned}$$

and consider the equation

$$G(\tau,\theta)=0.$$

Note that for any fixed θ it follows from the above discussion that the τ corresponding to a locally optimal design satisfy this equation.

Let now $\varphi(\tau, \theta)$ be an arbitrary function of two variables.

The Implicit Function Theorem (see e.g. Gunning, Rossi, 1965) asserts that if this function is real analytic and the Jacobi matrix

$$J(\bar{\theta}) = \left(\frac{\partial^2 \varphi(\tau, \bar{\theta})}{\partial \tau_i \partial \tau_j} \right)_{i,j=1}^{2n-3} \Big|_{\tau=\bar{\tau}},$$

is a non-singular matrix then in a vicinity of the point $\bar{\theta}$ there exists a unique continuous vector function $\tau(\theta)$ such that $\tau(\bar{\theta}) = \bar{\tau}$ and $G(\tau(\theta), \theta) = 0$. This function is real analytic.

Let us apply this theorem to studying locally optimal designs.

Consider the following assumptions:

A1. The function $\eta(x,\theta)$ is a real analytical function for all $\theta \in \Omega$, where Ω is an open simply connected subset of R^m, $x \in [a, b]$;

A2. The function $\Phi(M)$ possess properties 1) – 4) on the set of all positive definite matrices and is real analytic

A3. For any $\theta \in \Omega$ there exist a unique Φ-optimal design of type $(1, n-2, 1)$;

A4. The Jacobi matrix

$$J(\theta) = \left(\frac{\partial^2 \varphi(\tau, \theta)}{\partial \tau_i \partial \tau_j} \right)_{i,j=1}^{2n-3} \Big|_{\tau=\tau^*},$$

is a non-singular matrix for all $\theta \in \Omega$, where τ^* is a vector such that ξ_{τ^*} is a Φ-optimal design.

Let the vector $\tau^*=\tau^*(\theta)$ be such that ξ^*_τ is a locally $\Phi-$optimal design.

Theorem

Theorem 4.1. *The vector τ^* is a real analytical vector-function of θ for $\theta \in \Omega$ if the assumptions A1 – A4 hold.*

The proof of this Theorem is based on use of the Implicit Function Theorem in a way very similar to that in Chapter 2 of Melas (2006).

Note that assumption A1 (the regression function is real analytic) applies to many practically used regression models.

It follows from Lemma 3.1 that assumption A2 (standard properties of the optimality criteria)holds for the D- and L-optimality criteria.

Also, assumption A3 (about the type of optimal design) is fulfilled if the condition of Theorem 3.2 holds.

Let us consider assumption A4 (invertibility of the Jacobi matrix).

Theorem

Theorem 4.2.(the invertibility of the Jacobi matrix) *Let functions $f_i(x,\theta), i = 1,\dots,m$, where θ is a fixed vector from R^m, generate an extended Chebyshev system at the design interval, and the criterion of optimality $\Phi(M)$ is one of the functions*

$$\Phi(M) = (\det M)^{-1/m},$$
$$\Phi(M) = \mathrm{tr} B M^{-1},$$
$$\Phi(M) = (\mathrm{tr} M^p)^{-1/p},\ 0 < p < \infty,$$

then condition A4 holds.

The proof of this theorem is based on the representation

$$J = E + B^T Q B,$$

where

$$E = \left(\mathrm{tr} \left(\frac{\partial^2 M(\xi_\tau, \theta)}{\partial \tau_i \partial \tau_j} \right)^T \left(\frac{\partial \Phi(A)}{\partial A} \right) |_{A=M(\xi_tau^*,\theta)} \right)_{i,j=1}^{s},$$

$$Q = \left(\frac{\partial^2 \Phi(A)}{\partial a_i \partial a_j} \right)_{i,j=1}^{m^2} |_{A=M(\xi_\tau^*,\theta)}, a = vector(A),$$

$$B = \left(\frac{\partial a}{\partial \tau_1}, \ldots, \frac{\partial a}{\partial \tau_s} \right) |a = vector(M(\xi_\tau, \theta)), \tau = \tau^*.$$

Note that E is a diagonal matrix and $E \geq 0$ since

$$\frac{\partial^2}{\partial x_i \partial w_j} \sum_{v=1}^{m} f_p(x_v) f_q(x_v) w_v = 0 \quad p, q = 1, \ldots, m$$

$$\frac{\partial^2}{\partial x_i \partial x_j} \sum_{v=1}^{m} f_p(x_v) f_q(x_v) w_v = 0, \quad i \neq j,\ p, q = q, \ldots, m$$

Besides

$$\frac{\partial^2}{\partial w^2} \sum_{v=1}^{m} f_p(x_v) f_q(x_v) w_v = 0,\ i = 1, \ldots, m-1,\ p, q = q, \ldots, m$$

$$\mathrm{tr} \left(\frac{\partial^2}{\partial \tau_i^2} \right)^T \left(\frac{\partial \Phi(A)}{\partial A} \right) \Bigg|_{A = M(\varepsilon_{\tau *}, \theta)} =$$

$$= \left(\frac{\partial^2}{\partial \tau_i} \mathrm{tr} \left(\frac{\partial}{\partial \tau_i} M(\varepsilon_\tau, \theta) \right)^T \frac{\partial \Phi(A)}{\partial A} \right) \Bigg|_{A = M(\varepsilon_{\tau *}, \theta), \tau = \tau *} \geq 0.$$

Let us prove that $J > 0$ that implies the invertibility of J.

Since $E \geq 0$ it will do to prove that $B^T QB$ is a positive definite matrix.

In Melas (2006) it was proved that B has full rang. And from lemma 3.2 it follows that the matrix Q is positive definite. Thus $B^T QB$ is positive definite as well.

Note that the rational regression function described in Example 3.1 satisfies all assumptions of the theorem.

Another example is

$$\eta(x,\theta) = \theta_0 + \frac{\theta_1}{1 + \exp((\theta_2 - x)/\theta_3)},$$

where $\theta_1 > 0,\ \theta_2 > 0$ and $\theta_3 > 0$.

Other examples are peculiar nonlinear models collected in Jang (2010).

Now consider the problem of finding Bayesian Φ-efficient designs which requires minimization of the quantity

$$\int \frac{\Phi(M(\xi,\theta))}{\Phi(M(\xi^*(\theta),\theta))} dP(\theta),$$

where $\xi^*(\theta)$ is a locally Φ-optimal design and $P(\theta)$ is the prior distribution for the (unknown) parameter vector.

Roughly speaking this quantity is inversely proportional to Bayesian efficiency for the Φ-criterion.

We now define a set of prior distributions that was introduced in Melas and Staroselskii (2008).

Define

$$\begin{aligned}\Omega &= \Omega_z = \{\theta = (\theta_1, \ldots, \theta_m)^T : \\ &(1-z)c_i \le \theta_i \le \frac{c_i}{1-z},\ i = 1, \ldots, n\}, \\ &z \in [0,1) \\ &c = (c_1, \ldots, c_m)^T, c_i > 0, i = 1, \ldots, m,\end{aligned}$$

where c can be considered as an approximation for θ^* and z as a relative error.

Let $P_z(\theta)$ be the uniform distribution on Ω_z.

By $\tau^*(\theta)$ denote a vector such that $\xi_{\tau^*(\theta)}$ is a locally Φ-optimal design.

Consider the problem of constructing Φ-optimal designs among designs of type $(1, m-2, 1)$. This problem reduces to that of minimization of the function

$$\Psi(\tau, z) = \int \frac{\Phi(M(\xi_\tau, \theta))}{\Phi(M(\xi_{\tau^*(\theta)}, \theta)} dP_z(\theta).$$

Write

$$\begin{aligned} \Psi(\tau, 0) &= \frac{\Phi(M(\xi_\tau, c))}{\Phi(M(\xi_{\tau^*(c)}, c))} \\ \Psi(\tau, -z) &= \Psi(\tau, z), \; z \in [0, 1), \end{aligned}$$

For any (scalar, vector or matrix) function $F(z)$ we use the notation

$$F_{(0)} = F(0),\ \ F_{(k)} = \frac{1}{k!}F^{(k)}(0),\ \ k = 1, 2, \ldots$$

$$F_{<k>}(z) = \sum_{s=0}^{k} F_{(s)} z^s,\ \ k = 1, 2, \ldots$$

and denote

$$g(\tau, z) = \left(\frac{\partial}{\partial \tau_i} \Psi(\tau, z) \right)_{i=1}^{2n-3}.$$

Theorem

Theorem 4.3. *If conditions A1-A4 hold then*

I. For some critical value $z^* > 0$ there exist a unique continuous function $\hat{\tau}(z)$ such that $\xi = \xi_{\hat{\tau}(z)}$ is a Bayesian Φ-efficient design for the distribution $P_z(\theta)$. This function is a real analytical function.

Theorem

II. The coefficients of the Taylor expansion of this function can be obtained by the following recurrence formulae

$$\hat{\tau}_{(k+1)} = -\left[J(0)\right]^{-1} g\left(\hat{\tau}_{<k>}(z), z\right)_{(k+1)}, \quad k = 0.1.\ldots,$$

where $\hat{\tau}_0$ *corresponds to the locally* Φ*-optimal design for* $\theta = c$,

$$J(z) = \left(\frac{\partial^2}{\partial\tau_i\partial\tau_j}\Psi(\tau, z)\right)_{i,j=1}^{2k-3}\Big|_{\tau=\hat{\tau}(z)}.$$

Part I of Theorem 4.3 follows from the well known Implicit Function Theorem. Part II follows from Theorem 2.4.3 of Melas (2006).

The next theorem can be used for finding the numerical value of z^*.

Theorem

Theorem 4.4. *Let the regression equation be given by (1), with the function* $\eta(x, \theta)$ *continuously differentiable with respect to the parameters and let* $\Phi(M)$ *be continuously differentiable and convex.*

Theorem

Then the design ξ is the Bayesian Φ-efficient design for the distribution $P_z(\theta)$ if and only if the inequality

$$\int \left(f^T(x,\theta) \frac{\partial \Phi(M(\xi,\theta))}{\partial M} f(x,\theta) - \mathrm{tr} \frac{\partial \Phi(M(\xi,\theta))}{\partial M} M(\xi,\theta) \right) \times$$

$$\times \frac{dP_z(\theta)}{\Phi(M(\xi^*(\theta),\theta))} \leq 0.$$

holds for all $x \in \mathcal{X}$.

The theorem is an analogue of the Equivalence Theorem for the Φ-criterion (see, e.g., Pukelsheim, 2006). Thus it can be proved by standard methods. Note that the Bayesian $\Phi-$efficient criterion is convex that follows from the convexity of the $\Phi-$criterion assumed here.
Theorems 4.3 and 4.4 lead to construction of the Bayesian Φ-efficient design for $z \le z^*$ following the stages:

(i) Find a (locally) Φ-optimal design either numerically or analytically;

(ii) Calculate the coefficients of the Taylor expansion of the function $\hat{\tau}(z)$ by the recurrence formulae from Theorem 4.3;

(iii) Check if the design obtained as a sum of terms of the selected series satisfies the inequality in Theorem 4.4 with the required precision.

5. Numerical example

Consider the regression function

$$\eta(x,\theta) = \frac{\theta_1 x(\theta_2 + x)}{\theta_3 + \theta_4 x + x^2}.$$

This function comes from the assumption of second-order kinetics in a chemical example. (See Crabbe et al., 1986). Let $x \in [a, b]$,

$$\Omega = \Omega_z = \{\theta = (\theta_1, \ldots, \theta_4)^T : (1-z)c_i \leq \theta_i \leq c_i/(1-z), i = 1, 2, 3, 4\}.$$

Following Crabbe et al. (1986) we consider the case $[a, b] = [0.01, 10]$, $c = (2.45, 1.87, 0.08, 5.05)$ and $B = \mathrm{diag}\{0.6, 1, 600, 0.15\}$.

It is easy to check by Theorem 3.1 that any locally L-optimal design has the form

$$\xi = \begin{pmatrix} x_1 & x_2 & x_3 & b \\ w_1 & w_2 & w_3 & w_4 \end{pmatrix}.$$

By a direct numerical calculation we find that the locally L-optimal design for $\theta = c$ is of type $(1, 2, 1)$ and equals

$$\xi = \xi_{\tau_{(0)}},\ \tau_{(0)} = (0.203, 2.937, 0.201, 0.275, 0.350).$$

The first four coefficients of the Taylor expansion of the function $\hat{\tau}(z)$ are shown in the following table.

Table: *Coefficients of the Taylor expansion of the Bayesian L-efficient design at the point* $z_0 = 0$

i	0	1	2	3	4
x_2^*	0.20309	0.0021884	0.07748	0.083202	0.17356
x_3^*	2.9366	0.0044348	0.15691	0.13026	-0.18577
w_1^*	0.20134	0.0015665	0.055427	0.054796	0.052787
w_2^*	0.2746	-0.0022972	-0.08131	-0.078864	-0.0029459
w_3^*	0.3499	0.00038712	0.013491	0.0027333	-0.064029

The corresponding design ξ_{τ^*}, $\tau^* = \hat{\tau}_{<4>}(z)$ satisfies the conditions of Theorem 4.4 with a precision of at least 10^{-4}.

From a practical standpoint there is no difference between this design and the Bayesian L-efficient design.

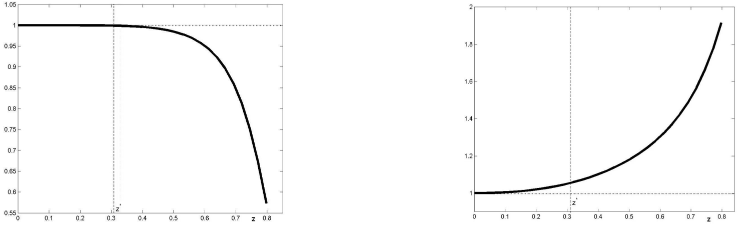

Figure. *The relative efficiency of the locally L-optimal design at the point $\theta = c$ (left panel) with respect to the Bayesian L-efficient design ξ_τ and the numerical value of $\Psi(\tau^*, z)$ (right panel) as $z \in [0, 0.8]$ (Matrix $B = diag\{0.6, 1, 600, 0.15\}$)*

The results given in the figure can be interpreted as follows.

First (see the left panel), with $z \in [0, 0.5]$ the relative efficiency of the locally optimal design for point $\theta = c$ with respect to the Bayesian $L-$efficient design is at least 97 %.

Secondly (see the right panel), for the same values of z the **efficiency of the locally optimal design for the true value of the parameter vector with respect to the Bayesian $L-$efficient design is no greater than 120 %.**

Thus our results show that, for the model considered, the locally optimal designs are rather close to Bayesian $L-$efficient designs in their properties if the relative error in the approximate value for the parameters is less then 50%.

Concluding remarks

It was shown that under some conditions the functional approach allows one to construct locally optimal and Bayesian efficient designs by Taylor series.

The basic assumptions are

i) **the regression function depends on a scalar variable belonging to the design interval,**

ii) derivatives of the function by the parameters generate an extended Chebyshev system on the design interval

These assumptions hold for a lot of practically important regression models.

References

Atkinson, A. C., Donev, A. N., and R. D. Tobias, 2007. *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

Chaloner, K. (1993). A note on optimal Bayesian design for nonlinear prob- lems. *Journal of Statistical Planning and Inference*, **37**, 229-235.

Crabbe, M. J., Jordan, R. M., Ting, H. H., Hoe, S. T. (1986). Bovine lens aldehyde dehydrogenase: activity and non-linear steady-state kinetics. *Experimental eye research* 1986;43(2):177–84.

Dette, H., Melas, V. B. (2011). A note on the de la Garza phenomenon for locally optimal designs. *Annals of Statistics*, **39**, 1266–1281.

Dette, H., Melas, V. B., Pepelyshev, A. (2004). Optimal designs for estimating individual coefficients-a functional approach. *J. Statist. Plann. Inference*, **118**, 201-209.

Dette, H. and Neugebauer, H.-M. (1996). Bayesian optimal one point designs for one parameter nonlinear models. *Journal of Statistical Planning and Inference*, **52**, 17–31.

Fedorov, V.V. (1972). *Theory of Optimal Experiments*. New York: Academic Press.

Gunning, R.C., Rossi, H. (1965). *Analytical Functions of Several Complex Variables*. New York: Prentice-Hall Inc.

He, Z., Studden, W.J., Sun, D. (1996). Optimal designs for rational models. *Ann. Statist.*, **24**, 2128–2147.

Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**, 633–643.

Karlin, S. and Studden, W. (1966). *Tchebysheff systems: with application in analysis and statistics*, Wiley, New York.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849–879.

Melas, V.B. (1978). Optimal designs for exponential regression. *Math. Operationsforsh. Statist.*, **9**, 45–59.

Melas, V.B. (2006). *Functional Approach to Optimal Experimental Design*. Lecture Notes in Statistics, vol. 184. Heidelberg: Springer.

Melas, V.B., Staroselsky, Yu. (2008). *D*-efficient Bayesian designs for a class of nonlinear models. *J. of Statistical Theory and Practice* 2(4), 568–587.

Pukelsheim, F. (2006). *Optimal Design of Experiments*. Classics in Applied Mathematics, vol. 50. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Yang, M. (2010). On the de la Garza Phenomenon.*Annals of Statistics*, **38**,2499-2524.