Appl. Math. and Comp. Sci., 1998, Vol.8, No.3, 647–656

SEARCHING FOR OPTIMAL EXPERIMENTAL DESIGNS USING SPACE-FILLING CURVES

EWA SKUBALSKA-RAFAJŁOWICZ*, EWARYST RAFAJŁOWICZ*

A new approach to the standard problem of searching for optimal experimental designs is considered. It consists in replacing a multidimensional search for global maxima by a one-dimensional global search. The points found in this way are then transformed to the multivariate design domain by using a space-filling curve. It is shown that this approach leads to the optimal design, provided that the one-dimensional global search is reliable. An additional advantage of the proposed approach is the possibility of visualization of the model variance surface. The results are presented for the D-optimality criterion, but their extension to other criteria is not difficult.

1. Introduction

Since the early 1970's the algorithms of Wynn and Fedorov (Fedorov, 1972; Wynn, 1970) and their modifications have been the main tools in searching for optimal experimental designs for estimating regression functions. The most difficult and time-consuming step of these algorithms is to find a point to be entered into a present design. The difficulty is that the point to be entered should be a global maximum of the prediction variance surface corresponding to the present design. In the implementations used so far, this difficulty has been circumvented by reducing the set of candidate points from a subset of the Euclidean space to a finite set of candidate points, usually forming a grid covering the design domain. Efficient second-order algorithms have been recently proposed by Gaffke and Heiligers (1995). In their algorithms it is possible not to address a global search directly by using *a priori* information on the optimal design support points. We refer the reader to this paper also for a detailed survey of other design searching algorithms.

An alternative to the brute grid-search way of finding candidates to be entered is given in (Rafajłowicz, 1998), where a selective random search is proposed. Here, we consider a potentially more accurate approach which is based on searching for candidate points in the unit interval and then transforming them to a multidimensional design domain by using a space-filling transformation, such as the Peano or Hilbert curve (see (Hilbert, 1891; Milne, 1980; Sagan, 1994) for basic properties of space-filling curves).

The reminder of the paper is organized as follows. After stating the problem, in the next section we formulate a theoretical basis for the proposed algorithm which

* Technical University of Wrocław, Institute of Technical Cybernetics, ul. Janiszewskiego 11/12, 50–372 Wrocław, Poland, e-mail: raf@ldhpux.immt.pwr.wroc.pl.

is then described. Finally, we discuss some implementation aspects and preliminary results of numerical experiments with the algorithm.

2. Problem Statement

Let $EY(x) = a^T v(x)$ be a regression function spanned by a given column vector $v(x) \in \mathbb{R}^r$ of functions which are linearly independent in a given domain $X \subset \mathbb{R}^s$. Here $a \in \mathbb{R}^r$ denotes the column vector of unknown constant parameters which are estimated by the least-squares method from observations of $Y(x_i)$, $i = 1, 2, \ldots, n$. We assume that $\mathrm{cov}\,(Y(x_i), Y(x_j)) = 0$ for $i \neq j$ and $\mathrm{var}\,(Y(x)) = \sigma^2 = \mathrm{const} < \infty$. As regards the domain of the experiment X and $v(x)$, we adopt the standard assumptions (Fedorov, 1972; Pazman, 1986). Namely, X is assumed to be closed and bounded, while the components of $v : X \to \mathbb{R}^r$ are continuous functions in X.

We treat the design of experiments as discrete probability measures on X, i.e. ξ is a design of experiment, if it is of the form:

$$\xi = \begin{bmatrix} x_1, x_2, \ldots, x_m \\ p_1, p_2, \ldots, p_m \end{bmatrix} \tag{1}$$

where $x_i \in X$, $p_i \geq 0$, $i = 1, 2, \ldots, m$ and $\sum_{i=1}^m p_i = 1$. We refer the reader to (Atkinson and Donev, 1992; Fedorov, 1972; Pazman, 1986) for the motivations and interpretation of such a design. The class of all designs (1) will be denoted by $\Xi(X)$, admitting a different number of support points x_i, $i = 1, 2, \ldots, m$ in different designs. Referring to the above-mentioned origins, we consider the problem of finding a D-optimal design ξ^*, for which $\det M(\xi^*) = \sup_{\xi \in \Xi(X)} \det M(\xi)$, the information matrix $M(\xi)$ being defined as follows:

$$M_s(\xi) = \sum_{i=1}^{m} p_i v(x_i) v^T(x_i)$$

where the subscript s underlines the dependence on the dimension of x.

The Fedorov-Wynn algorithm of searching for ξ^* is based on the maximization of the function

$$\varphi(x, \xi) = v^T(x) M^{-1}(\xi) v(x) \tag{2}$$

with respect to $x \in X$, provided that for a given design ξ, $M(\xi)$ is nonsingular. If in the n-th iteration a design ξ_n is found, then $\varphi(x, \xi_n)$ is maximized in order to find a point x_n to be entered into ξ_n. The search for x_n is the crucial step of the algorithm sketched above for the following reasons:

1. $\varphi(x_n, \xi_n)$ is expected to be the global maximum of $\varphi(\cdot, \xi_n)$ over X.
2. $\varphi(\cdot, \xi_n)$ is known to be multimodal.
3. Maximization of $\varphi(\cdot, \xi_n)$ is performed at each iteration.

Our aim in this paper is to propose an algorithm for finding global maxima of $\varphi(\cdot, \xi_n)$, which is such that the Fedorov-Wynn algorithm retains its convergence properties.

2.1. Space-Filling Curves and Complete Class Theorem

Denote by I_1 the unit interval $[0,1]$ and let $I_s = [0,1] \times \cdots \times [0,1]$ (s times) be the unit hypercube. For simplicity, we assume that the experiment domain $X = I_s$ (note that D-optimum designs can be transformed by one-to-one transformations to other domains, while retaining their optimality).

We need a space-filling curve $\Phi : I_1 \to I_s$ which has the following properties:

$\Phi 1$) Φ transforms I_1 onto I_s,

$\Phi 2$) Φ is continuous.

Let us note that the Peano-Hilbert curves possess properties $\Phi 1$ and $\Phi 2$ (see (Butz, 1971; Milne, 1980; Sagan, 1994) and the references given therein). Also the Sierpiński curve, originally constructed on the unit square (Sierpiński, 1912) and generalized by the first author (Skubalska-Rafajłowicz, 1994), is such that $\Phi 1$ and $\Phi 2$ hold.

Additionally, the generalized Sierpiński curve is closed, which may be advantageous in our application.

Let us note that Φ cannot have an inverse. The above-mentioned curves possess many other interesting features such as Lipschitz continuity and preservation of the Lebesgue measure (Milne, 1980), which are not used in this paper, but may be of interest in other applications.

Consider the class of all experiment designs on I_1 and denote it by $\Xi(I_1)$. Transform each design $\tau \in \Xi(I_1)$ by Φ as follows: convert each support point $t_i \in I_1$ of τ to $x_i = \Phi(t_i) \in I_s$ and associate with x_i the same weight p_i, as t_i had in design τ. Denote by $\Phi(\Xi(I_1))$ the class of all the designs which can be obtained in the way described above. Clearly, $\Phi(\Xi(I_1)) \subset \Xi(I_s)$, but the nonexistence of the inverse of Φ implies that for a certain $\xi \in \Xi(I_s)$ it may have many counterparts in $\Xi(I_1)$. We shall show below that $\Phi(\Xi(I_1))$ is sufficiently rich to contain at least one solution to the D-optimal design problem.

Theorem 1. *Let $\tau^* \in \Xi(I_1)$ be a D-optimal design for estimating a regression function $a^T v(\Phi(t))$ over I_1. Let t_i^*, $i = 1,2,\ldots,m$ denote the support points of τ^*, p_i^* being the corresponding weights. Then the design ξ^* with the support points $x_i^* = \Phi(t_i^*)$ and weights p_i^*, $i = 1,2,\ldots,m$ is D-optimal in the class $\Xi(I_s)$ for estimating $a^T v(x)$ over I_s.*

Outline of the Proof. The optimality of τ^* implies (by the Kiefer-Wolfowitz equivalence theorem)

$$\sup_{t \in I_1} v^T\big(\Phi(t)\big) M_1^{-1}(\tau^*) v\big(\Phi(t)\big) = r \tag{3}$$

where

$$M_1(\tau^*) = \sum_{i=1}^{m} p_i^* v\big(\Phi(t_i)\big) v^T\big(\Phi(t_i)\big) \tag{4}$$

Note that by construction of ξ^* we have: $M_1(\tau^*) = M_s(\xi^*)$. Furthermore, $\Phi 1$ and $\Phi 2$ imply that for any continuous function $q : I_s \to \mathbb{R}$ we have $\sup_{t\in I_1} q(\Phi(t)) = \sup_{x\in I_s} q(x)$. These facts and (3) imply

$$\sup_{x\in I_s} v^T(x)M^{-1}s(\xi^*)v(x) = r$$

which completes the proof, by invoking again the Kiefer-Wolfowitz equivalence theorem (Fedorov, 1972; Pazman, 1986). ■

3. Fedorov-Wynn Algorithm in One Dimension

The above simple theorem implies that it suffices to apply the Wynn-Fedorov algorithm to $a^T v(\Phi(t))$, $t \in I_1$ in order to obtain a sequence of designs which is convergent to a D-optimal design.

Modified Fedorov-Wynn Algorithm

Step 0. Select $\tau_0 \in \Xi(I_1)$ such that $M_1(\tau_0)$ is nonsingular and set $k = 0$ (the iteration counter).

Step 1. Select (by a one-dimensional global search algorithm) a point $t_k \in I_1$, for which a (global) maximum in I_1 of the function

$$\varphi_k(t) \triangleq v^T\big(\Phi(t)\big)M_1^{-1}(\tau_k)v\big(\Phi(t)\big) \tag{5}$$

is attained.

Step 2. Update τ_k according to

$$\tau_{k+1} = (1-\alpha_k)\tau_k + \alpha_k\delta(t_k) \tag{6}$$

where $0 < \alpha_k < 1$ is chosen as $\arg\max_{0\le\alpha\le 1} \ln\det M_1((1-\alpha)\tau_k + \alpha\delta(t_k))$, $\delta(t_k)$ being a one-point design at t_k.

Step 3. If $\varphi_k(t_k) - r$ is less than a prescribed accuracy $\varepsilon > 0$, then go to Step 4, otherwise set $k := k+1$ and go to Step 1.

Step 4. Transform the support of τ_k by Φ to I_s and attach the corresponding weights of τ_k. The resulting design ξ_k provides an approximate solution to the problem and fulfils the inequality

$$\sup_{x\in I_s} \varphi(x,\xi_k) - r < \varepsilon \tag{7}$$

If we omit the stopping condition, then the above algorithm generates a sequence of designs which is convergent to a D-optimal design. This follows from the convergence of the Fedorov-Wynn algorithm in one-dimensional case and from Theorem 1. Clearly, the crucial point of the above algorithm is to have a reliable method of one-dimensional global search. In our computational experiments, we have used a

simplified version of the global search algorithm from (Sergeyev, 1995). When implementing the above algorithm, it is also necessary to have a method of calculating an approximation to a chosen space-filling curve (Butz, 1971; Milne, 1980).

It seems that the most interesting question concerning the proposed algorithm is where the "curse of dimensionality" is hidden. Trying to give a partial answer, let us note that known space-filling curves Φ are Hölder-continuous with exponent $1/s$. Thus, if the components of $v(x)$ are Hölder continuous with exponent $0 < \mu \leq 1$, say, then $v(\Phi(t))$ and $\varphi_k(t)$ are Hölder continuous with exponent μ/s which is not larger than $1/s$ if $v(x)$ is differentiable. In other words, a smooth global optimization problem over I_s is replaced by a one-dimensional problem with a highly irregular goal function. The gain, however, is in a possibility of visual verification whether a vicinity of a global maximum is found. The above considerations are illustrated in the next section.

4. Computer Experiments

Four examples from a series of experiments of the algorithm performance are reported here. We choose the cases in which optimum designs are well-known, in order to be able to evaluate the algorithm accuracy.

Example 1. Our aim in this example is twofold. Namely, we would like to illustrate a fractal behaviour of the variance surfaces scanned by a space-filling curve and to give some indications concerning the choice of the r parameter in the algorithm of global optimization from (Sergeyev, 1995), which was incorporated into the algorithm considered here.

Consider the quadratic regression function without interactions

$$y = a_0 + a_1 x^{(1)} + a^2 x^{(2)} + a_3 (x^{(1)})^2 + a_4 (x^{(2)})^2 \tag{8}$$

and the experiment design

```
{{{1., 1.}, 1/6}, {{0., 0.}, 1/6}
 {{1., 0.}, 1/6}, {{0., 1.}, 1/6}
{{0.5, 0.}, 1/6}, {{1., 0.5},1/6}}
```

which is treated as an initial design. The prediction variance corresponding to this design is shown in Fig. 1, as a function of $x^{(1)}$ and $x^{(2)}$ (the upper panel). The same function, but scanned by the Sierpiński curve is shown in the middle panels of this figure (in two different scales), while in the lower left and in the lower right panels $-\phi(\Phi(t), \xi)$, $t \in [0, 1]$ for the Hilbert and the Peano curves is plotted. Grey dots on these plots indicate points visited by the global search algorithm. A comparison of the last two rows of Fig. 1 indicates that scanning the same surfaces by the three different curves leads to quite different one-dimensional representations of the surface, but all of them share the same Hölder exponent.

To give some indications concerning the choice of the parameter r in the global search algorithm applied to find a minimum of $-\phi(\Phi(t), \xi)$, $t \in [0, 1]$ the following

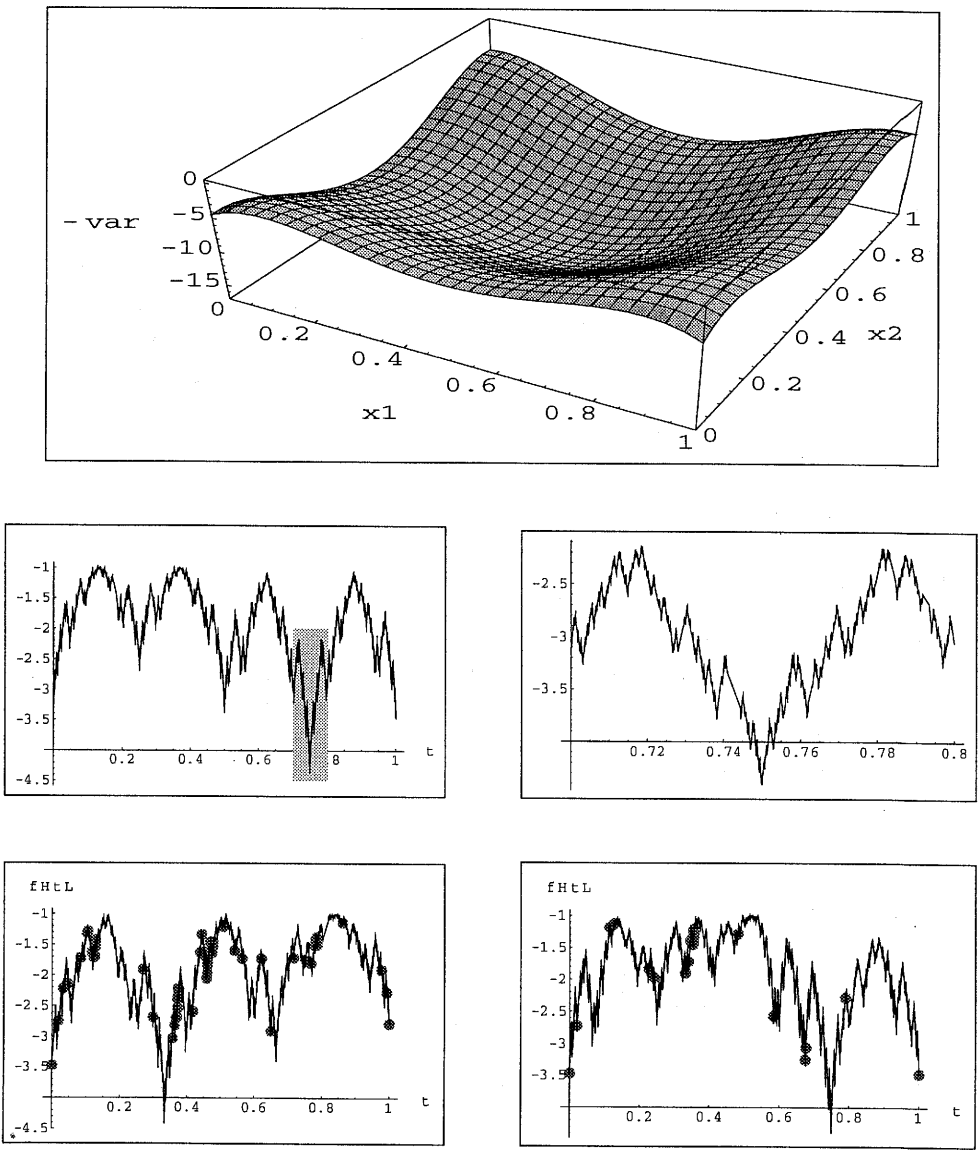

Fig. 1. Upper plot: the surface $-\phi(x,\xi)$ for the design described in Example 1. Middle left plot: the same surface scanned by the Sierpiński curve, i.e. the plot of the function $-\phi(\Phi(t),\xi)$, $t \in [0,1]$. The middle right plot reveals the fractal nature of $-\phi(\Phi(t),\xi)$. The shaded region from the left plot is enlarged on the right plot. In the lower left and the lower right panel the function $-\phi(\Phi(t),\xi)$, $t \in [0,1]$ is shown when Φ is the Hilbert and the Peano curve, respectively.

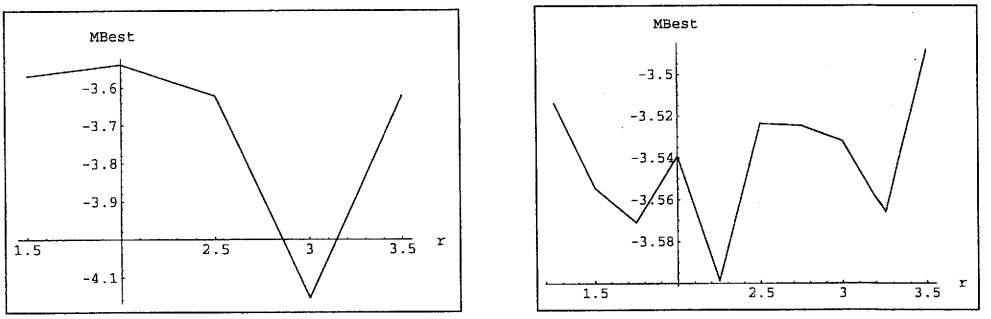

Fig. 2. The best values found by the global search algorithm as a function of the parameter r, obtained for the Hilbert (left panel) and Sierpiński (right panel) curve. See Example 1 for details.

simulations were performed. The global search algorithm was run for each value of the r parameter ranging from 1.5 to 3.5 with the step size 0.5. For each run the smallest value found by the algorithm was stored. These values for each r are plotted in Fig. 2 for the Hilbert (left panel) and the Sierpiński (right panel) curves. As one can notice, the values of r close to $r = 3$ assure good behaviour of the global search method. A similar pattern was obtained for the Peano curve. As a common value for all the simulations reported below $r = 3$ is chosen. ♦

In all the examples reported below the same accuracy 10^{-6} was used in the stopping condition. The 1-D global search algorithm was stopped when the length of the interval between succesive arguments was less than 10^{-10}. In practice, also a space-filling curve has to be approximated. The degree of the approximations which were used in the examples below can be expressed in terms of the edge lengths of subcubes which are treated as one point on an approximate curve. In the examples presented below these egde lenghts were $1/2^{12}$ for the Hilbert and the Sierpiński curve, while for the Peano curve the value $1/3^{12}$ was used.

Example 2. Consider the regression function

$$y = a_0 + a_1 x^{(1)} + a_2 x^{(2)}, \quad (x^{(1)}, x^{(2)}) \in [0,1] \times [0,1] \tag{9}$$

It is well-known that a D-optimal experiment design is concentrated in the corners of $[0,1] \times [0,1]$ with the weights $1/4$. This example turned out to be easy for the proposed algorithm. The optimum design was found in 2 or 3 iterations, depending whether the Sierpiński, Peano or Hilbert curve was used. The starting design was the same as in Example 1. ♦

Example 3. As an example with an intermediate level of difficulty, consider the quadratic regression in two variables $(x^{(1)}, x^{(2)}) \in [0,1] \times [0,1]$:

$$y = a_0 + a_1 x^{(1)} + a_2 x^{(2)} + a_3 x^{(1)} x^{(2)} + a_4 (x^{(1)})^2 + a_5 (x^{(2)})^2 \tag{10}$$

The design search algorithm was run three times from the same starting design for the Peano, Hilbert and Sierpiński curves. The plots in the right column of Fig. 3 illustrate

the history of increasing $\log\det M_1(\tau_k)$ versus iteration number k. As one can notice, after about 10–20 iterations the algorithm was able to find a good approximation of the optimal design, independently of which the space-filling curve was used. Below, the approximation to the optimal design, found by using the Hilbert curve, is shown (rounded to three decimal digits):

```
{{{0., 0.}, 0.146},      {{0.999, 0.499}, 0.080}
{{0, 0.999}, 0.146},     {{0.999, 0.999}, 0.146}
{{0.5, 0.5}, 0.096},     {{0.499, 0.999}, 0.080}
{{0., 0.5}, 0.080},      {{0.499, 0}, 0.080}}
{{0.999, 0.}, 0.146}
```

♦

Example 4. For $x \in [0,1] \times [0,1] \times [0,1]$ a quadratic regression without interaction terms (seven unknown parameters) was considered. The example turned out to be relatively difficult, since the accuracy 10^{-6} was not attained by the design search algorithm after 150 iterations, independently of which the space-filling curves were used. After additional 40 iterations, the algorithm with the Sierpiński curve found the following approximation to the optimum design (three decimel digits are retained):

```
{{0.998, 0.053, 0.007}, 0.111}, {{0.984, 0.982, 0.595}, 0.105}
{{0.002, 0.986, 0.053}, 0.114}, {{0.583, 0.000, 0.416}, 0.118}
{{0.008, 0.486, 0.557}, 0.108}, {{0.500, 0.500, 0.000}, 0.115}
{{0.065, 0.031, 0.992}, 0.112}, {{0.972, 0.446, 0.984}, 0.101}
{{0.532, 0.998, 1.000}, 0.111}
```

We should add, however, that the algorithm was able to improve essentially the starting design relatively quickly, as is documented in Fig. 4. ♦

Due to computer memory limitations of the brute grid-covering technique, we cannot give a full comparison between the new technique and the grid-covering one. We can, however, compare the timing of search for a new point to be entered to the initial design. Consider the initial design from Example 1 having the prediction variance function shown in Fig. 1 (see the upper plot). Using an IBM PC Pentium 200 MHz and Mathematica 3.01, we found the maximum after 0.88 s (23 functions evaluations) by the new technique with the Hilbert curve. Similar timings were obtained when the Peano and Sierpiński curves were used. Assuming the grid step 0.01 with respect to each variable (10000 grid points) the timing was 6.54 sec. Analogously, for the regression function with three factors described in Example 4, the timing of the grid search was 11.25 s for the grid step 0.05 with respect to each variable (9261 grid points). The corresponding time for the new technique was 1.21 s. One can expect that advantages of the new technique would be more visible when more factors enter to a regression function.

Let us also note that the proposed technique is a general-purpose method in the sense that it does not use any additional *a-priori* information on possible positions of the optimal design support points. If such *a-priori* knowledge is available, one should apply more specialized algorithms such as those proposed in (Gaffke and Heiligers, 1995).

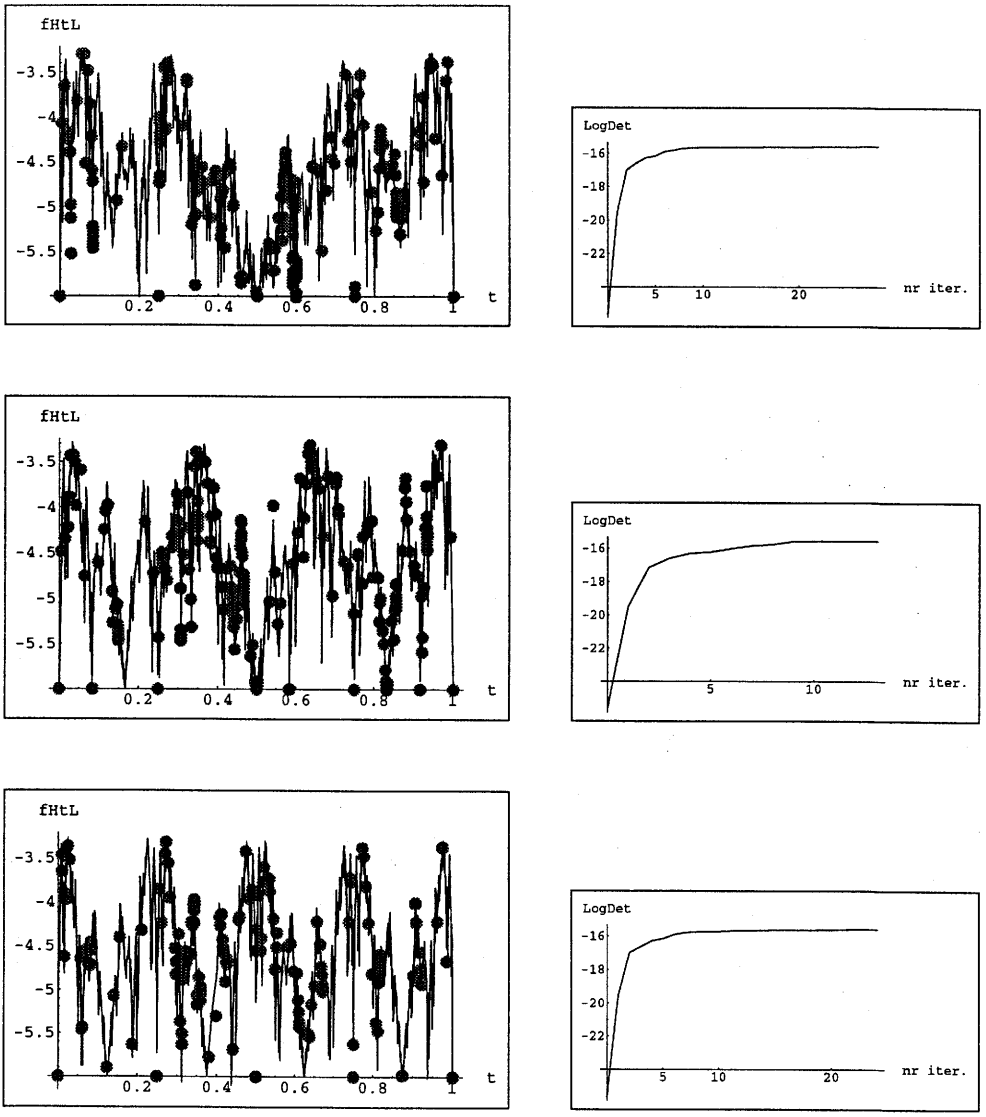

Fig. 3. Left column: functions $\phi_k(t)$ obtained in Example 3 in the last iteration of the design search algorithm when the Peano, Hilbert and Sierpiński space-filling curves were used. Dots indicate the points where $\phi_k(t)$ were evaluated by the 1-D global search algorithm. Right column: the logarithm of the determinant of the information matrix versus the iteration number of the design search algorithm. The plots coorespond to the same curves as the plots in the left column.

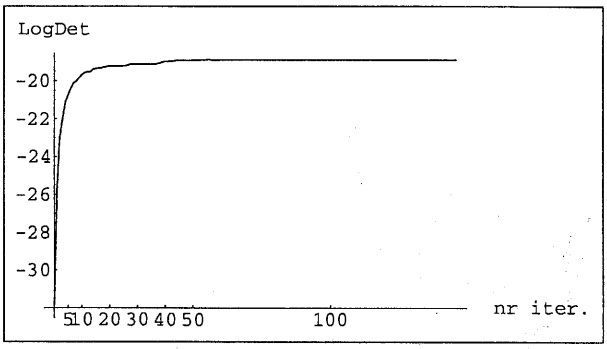

Fig. 4. Growth of $\log\det(M_1(\tau_k))$ obtained in Example 4.

References

Atkinson A.C. and Donev A.N. (1992): *Optimum Experimental Designs.* — Oxford: Clarendon Press.

Butz A.R. (1971): *Alternative algorithm for Hilbert's space-filling curve.* — IEEE Trans. Comput., Vol.C-20, pp.424–426.

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

Gaffke N. and Heiligers B. (1995): *Computing optimal approximate invariant designs for cubic regression on multidimensional balls and cubes.* — J. Statist. Planning and Inference, Vol.47, pp.347–376.

Hilbert D. (1891): *Ueber die stetige abbildung einer linie auf ein flaechenschtueck.* — Mathematische Annalen, Vol.38, pp.459–469.

Milne S.C. (1980): *Peano curves and smoothness of functions.* — Adv. Math., Vol.35, pp.129–157.

Pazman A. (1986): *Foundations of Optimum Experimental Design.* — Dordrecht: D. Reidel.

Rafajłowicz E. (1998): *Selective random search for optimum designs*, In: MODA 5 – Advances in Model-Oriented Data Analysis and Experimental Design (Atkinson A.C., Pronzato L. and Wynn H., Eds.). — Heidelberg, New York: Physica Verlag, pp.75–83.

Sagan H. (1994): *Space-Filling Curves.* — Berlin: Springer.

Sergeyev Y.D. (1995): *An information global optimization algorithm with local tuning.* — SIAM J. Optim., Vol.5, No.2, pp.858–870.

Sierpiński W. (1912): *O pewnej krzywej wypełniającej kwadrat. (Sur une nouvelle courbe continue qui remplit toute une aire plane).* — Bulletin de l'Acad. des Sciences de Cracovie A., pp.463–478.

Skubalska-Rafajłowicz E. (1994): *The closed curve filling multidimensional cube.* — Tech. Rep., No.46/94, Inst. Eng. Cybern., Tech. Univ. Wrocław.

Wynn H.P. (1970): *The sequential generation of D-optimum experimental designs.* — Ann. Math. Stat., Vol.1, No.41, pp.1655–1664.

Received: 26 October 1997
Revised: 30 April 1998