Int. J. Appl. Math. and Comp. Sci., 2000, Vol.10, No.2, 357–379

OPTIMAL SELECTION OF MEASUREMENT LOCATIONS FOR PARAMETER ESTIMATION IN DISTRIBUTED PROCESSES

DARIUSZ UCIŃSKI*

The problem of locating pointwise sensor measurements so as to optimally estimate unknown parameters in a class of distributed systems is studied. Based on a scalar measure of performance defined on the corresponding Fisher information matrix, two approaches are developed for this problem: introduction of continuous designs, which allows for adaptation of well-known sequential algorithms of classical optimum experimental design, and application of standard non-linear programming techniques. In each case, particular algorithms are delineated and analysis of the appropriate sensor placements is made. The relative advantages and shortcomings of both the approaches are discussed and demonstrated by applying them to a two-dimensional diffusion process.

Keywords: sensor location, distributed-parameter systems, parameter estimation, optimum experimental design, spatial statistics

1. Introduction

The design of measurement for estimating unknown parameters of a physical system is related to the optimal choice of measurement conditions so as to obtain the best information about these parameters. As far as a distributed-parameter system (DPS) is concerned, which is understood here as a dynamic system governed by a partial differential equation (PDE), it is generally impossible to observe the system states in overall spatial domain. What is more, cost limitations usually severely restrict the number of sensors which will be available for parameter estimation. Consequently, special care should be exercised when locating a limited number of measurement transducers, since an improper placement of sensor resources may lead to inaccurate estimates or even involve serious problems with identifiability.

Referring to the optimal sensor location problem for parameter estimation, a number of works have appeared in the last two decades. The existing techniques can be split into three main groups (Korbicz and Uciński, 1994):

1. Methods which convert the problem to state estimation,

* Institute of Robotics and Software Engineering, Technical University of Zielona Góra, ul. Podgórna 50, 65–246 Zielona Góra, Poland, e-mail: D.Ucinski@irio.pz.zgora.pl

2. Methods which reduce the task to random fields analysis, and

3. Methods which make use of optimum experimental design theory.

The approach characteristic of Group 1 is to augment the state vector by including the parameters to be identified and then to use well-developed methods of optimal sensor location for state estimation. However, since the state and parameter estimation are to be carried out simultaneously, the whole problem becomes strongly non-linear. To overcome this difficulty, a sequence of linearizations at consecutive state trajectories was performed by Malebranche (1988) and a special suboptimal filtering algorithm was used by Korbicz *et al.* (1988). Nevertheless, the viability of this approach is rather questionable owing to the well-known severe difficulties inherent in non-linear state estimation.

The methods of Group 2 are based on random fields theory. Since DPS's are described by PDE's, direct application of that theory is impossible, and therefore this description should be replaced by characteristics of a random field, e.g. mean and covariance functions. Such a method for a beam vibrating due to the action of a stochastic loading was considered by Kazimierczyk (1989) who made extensive use of optimum experimental design for random fields (Brimkulov *et al.*, 1986). Although the flexibility of this approach seems rather limited, it can be useful in some case studies (see e.g. Sun, 1994)

In turn, the methods of Group 3 originate from the classical theory of optimum experimental design (Fedorov and Hackl, 1997; Pázman, 1986; Pukelsheim, 1993; Walter and Pronzato, 1997) where the adopted optimization criteria are various scalar measures of performance based on the Fisher information matrix (FIM) associated with the parameters to be identified. First investigations in this spirit for dynamic DPS's date back to the work (Quereshi *et al.*, 1980) where the determinant of the FIM was maximized and examples regarding a damped vibrating string and a heat-diffusion process were used to illustrate the advantages and peculiarities of the method. The same optimality criterion was used by Rafajłowicz (1978) in order to optimize both sensor positions and a distributed control for parameter estimation of a static linear DPS. Reduction of the problem to a form, where results of the classical theory of optimum experimental design can be applied, was accomplished after eigenfunction expansion of the solution to the PDE considered and subsequent truncation of the resulting infinite series. Consequently, the FIM was associated with system eigenvalues, rather than with the system parameters. A separation principle was proved which allows the possibility of finding an optimal control and an optimal sensor configuration independently of each other. The delineated approach was generalized in (Rafajłowicz, 1981) to a class of DPS's described by linear hyperbolic equations with known eigenfunctions and unknown eigenvalues. The aim was to find conditions for optimality of measurement design and of optimal spectral density of the stochastic input. It was indicated that common numerical procedures from classical experimental design for linear regression models could be adopted to find optimal sensor location. Moreover, the demonstrated optimality conditions imply that the optimal input comprises a finite number of sinusoidal signals and that optimal sensor positions are not difficult to find in some cases. A similar problem was studied in

(Rafajłowicz, 1983) in a more general framework of DPS's which can be described in terms of Green's functions.

The idea of generalizing methods of optimum experimental design for parameter identification of lumped systems was also applied to solve the optimal measurement problem for moving sensors (Rafajłowicz, 1986b). The approach was based on looking for a time-dependent measure, rather than for the trajectories themselves. Various sufficient optimality conditions were presented, among others the so-called *quasi-maximum principle*. In spite of their somewhat abstract forms, they made it possible to solve relatively easily a number of non-trivial examples. The problem of moving sensors in DPS's was also revisited in (Rafajłowicz, 1988; Rafajłowicz, 1989), but without direct reference to parameter estimation. On the other hand, a slightly different approach to the design of sensor motions was proposed in (Uciński, 1999; Uciński, 2000a; Uciński, 2000b), where the problem was formulated as an optimal-control one with state-variable inequality constraints representing geometric constraints induced by the admissible measurement regions and allowable distances between the sensors. Taking account of the dynamic models of the vehicles carrying the sensors, the problem was reduced to determination of both the control forces of the sensors and initial sensor positions. A method of successive linearizations was then employed to construct a quite efficient numerical scheme of determining optimal sensor trajectories.

The approach based on maximization of the determinant of the appropriate FIM is not restricted to theoretical considerations and there are examples which do confirm its effectiveness in practical applications. Thus, in (Munack, 1984) a given number of stationary sensors were optimally located using non-linear programming techniques for a biotechnological system consisting of a bubble column loop fermenter. On the other hand, Sun (1994) advocates using optimum experimental design techniques to solve inverse problems in groundwater modelling. How to monitor the water quality around a landfill place is an example of such a network design. Sun's monograph constitutes an excellent introductory text to applied experimental design for DPS's, as it covers a broad range of issues motivated by engineering problems. Non-linear programming techniques are also used there to find numerical approximations to the respective exact solutions.

A related optimality criterion was given in (Point *et al.*, 1996) by the maximization of the Gram determinant which is a measure of the independence of the sensitivity functions evaluated at sensor locations. The authors argue that such a procedure guarantees that the parameters are identifiable and the correlation between the sensor outputs is minimized. The form of the criterion itself resembles the D-optimality criterion proposed by Quereshi *et al.* and Rafajłowicz, but the counterpart of the FIM takes on much larger dimensions, which suggests that the approach involves more cumbersome calculations. The delineated technique was successfully applied to a laboratory-scale catalytic fixed-bed reactor (Vande Wouwer *et al.*, 1999).

Our main purpose here is to show how some well-known methods of optimum experimental design for linear regression models can be extended to the setting of the sensor location problem. In our opinion, the main contribution of this paper concerns characterizations of continuous (or approximated) designs for a wide class of

design criteria, which allows an easy testing of any given sensor setting for optimality. This constitutes a significant generalization of Rafajłowicz's results mentioned above, which were centred on D-optimal designs. Moreover, it is indicated how to adapt well-known algorithms of optimum experimental design for finding numerical approximations to the solutions. As an alternative, it is also shown how to exploit non-linear programming techniques to tackle problems with a moderate number of sensors. In the final part of the paper, numerical results are presented to illustrate the application of the delineated sensor placement methodology.

2. Optimal Measurement Problem Based on Pointwise Observations

In what follows, we shall consider a bounded simply-connected open domain $\Omega \subset \mathbb{R}^2$ with sufficiently smooth boundary $\partial\Omega$. The mathematical model of our scalar distributed system is given by

$$\frac{\partial y}{\partial t} = \mathcal{F}\left(x, t, y, \frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \frac{\partial^2 y}{\partial x_1^2}, \frac{\partial^2 y}{\partial x_2^2}, \theta\right), \quad x \in \Omega, \quad t \in T, \tag{1}$$

where $x = (x_1, x_2) \in \bar{\Omega} = \Omega \cup \partial\Omega$ is the corresponding spatial coordinate vector, t stands for time, $T = (0, t_f)$, $y = y(x,t)$ denotes the state variable with values in $\mathbb{R}$ and $\mathcal{F}$ is some known function which may include terms accounting for given *a-priori* forcing inputs.

The boundary and initial conditions for the system are

$$\mathcal{E}\left(x, t, y, \frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \theta\right) = 0, \quad x \in \partial\Omega, \quad t \in T \tag{2}$$

and

$$y(x, 0) = y_0(x), \quad x \in \Omega, \tag{3}$$

respectively, where $\mathcal{E}$ and y_0 denote some known functions.

We assume the existence of a unique solution to (1)–(3), which is sufficiently regular. The system model above contains an unknown constant parameter vector denoted by $\theta \in \mathbb{R}^m$ (note that it may also appear in the boundary conditions), which is assumed to belong to a parameter space Θ. The objective of parameter estimation is to choose a parameter $\theta^\star$ in Θ so that the solution y to (1)–(3) corresponding to $\theta = \theta^\star$ agrees with the 'true' observed state $\tilde{y}$. In general, however, measurements of the state may not be possible, rather only measurements for some observable part of the actual state $\tilde{y}$ may be available. In what follows, we consider the observation process

$$z(t) = y_{\mathrm{m}}(t) + \varepsilon_{\mathrm{m}}(t), \quad t \in T, \tag{4}$$

where

$$y_{\mathrm{m}}(t) = \mathrm{col}\,[y(x^1,t),\ldots,y(x^N,t)],$$

$$\varepsilon_{\mathrm{m}}(t) = \mathrm{col}\,[\varepsilon(x^1,t),\ldots,\varepsilon(x^N,t)],$$

$z(t)$ is the N-dimensional observation vector, $x^j \in X$, $j = 1,\ldots,N$ denote the pointwise and stationary sensor locations, $X \subset \bar{\Omega}$ is the part of spatial domain where the measurements can be made (it is assumed to be a compact set), and $\varepsilon(x^j,t)$ denotes the measurement noise.

It is customary to assume that the measurement noise is zero-mean, Gaussian, spatial uncorrelated and white (Quereshi *et al.*, 1980), i.e.

$$\mathrm{E}\{\varepsilon(x^i,t)\varepsilon(x^j,t')\} = \sigma^2\delta_{ij}\delta(t-t'), \tag{5}$$

where $\sigma > 0$ is the standard deviation of the measurement noise, δ_{ij} and δ standing for the Kronecker and Dirac delta functions, respectively.

Parameter estimation is usually cast as an optimization problem, which leads to the *least-squares* formulation in which we seek to minimize the fit-to-data criterion

$$\mathcal{J}(\theta) = \int_T \|z(t) - \hat{y}_{\mathrm{m}}(t;\theta)\|^2\,\mathrm{d}t, \tag{6}$$

where

$$\hat{y}_{\mathrm{m}}(t;\theta) = \mathrm{col}\,[\hat{y}(x^1,t;\theta),\ldots,\hat{y}(x^N,t;\theta)],$$

$\hat{y}(\cdot,\cdot;\theta)$ stands for the solution to (1)–(3) corresponding to a given parameter θ, and $\|\cdot\|$ signifies the Euclidean norm.

It goes without saying that the parameter estimate $\hat{\theta}$ resulting from minimization of the fit-to-data criterion depends on the sensor positions owing to the presence of the quantity z in the integrand on the right-hand side of (6). This fact suggests that we may attempt to select sensor locations which lead to best estimates of the system parameters. To form a basis for the comparison of different locations, a quantitative measure of the 'goodness' of particular locations is required. A logical approach is to choose a measure related to the expected accuracy of the parameter estimates to be obtained from the data collected. Such a measure is usually based on the concept of the *Fisher Information Matrix* (FIM) (Rafajłowicz, 1986b; Sun, 1994) whose inverse is the Cramér-Rao lower bound on the covariance matrix of any unbiased estimator of θ (Goodwin and Payne, 1977). When the time horizon is large, the nonlinearity of the model with respect to its parameters is mild and the measurement errors are independently distributed and have small magnitudes, it is legitimate to assume that our estimator is *efficient* (minimum-variance) in the sense that the parameter covariance matrix achieves the lower bound (Rafajłowicz, 1986a). This leads to a great simplification since the minimum variance given by the Cramér-Rao lower bound can be easily computed in a number of estimation problems, even though the exact covariance matrix of a particular estimator is very difficult to obtain.

It turns out that the FIM is a deterministic function of the spatial location of the measurement sensors, as it is given by (Quereshi *et al.*, 1980)

$$M = \frac{1}{\sigma^2} \sum_{j=1}^{N} \int_0^{t_f} g(x^j, t) g^{\mathrm{T}}(x^j, t)\, \mathrm{d}t. \tag{7}$$

Here

$$g(x, t) = \left(\frac{\partial y(x, t; \theta)}{\partial \theta} \right)^{\mathrm{T}}_{\theta = \theta^0} \tag{8}$$

stands for the so-called *sensitivity vector*, θ^0 being a prior estimate to the unknown parameter vector θ (Sun, 1994). We assume that both $y(\cdot, \cdot; \theta^0)$ and $\partial y(\cdot, \cdot; \theta^0)/\partial \theta_i$, $i = 1, \ldots, m$ are continuous in $\bar{\Omega} \times T$.

Optimal sensor positions for system identification can be found by choosing x^j, $j = 1, \ldots, N$ so as to minimize some scalar measure of performance Ψ based on the FIM. Various choices exist for such a function (Fedorov and Hackl, 1997; Pázman, 1986; Pukelsheim, 1993; Walter and Pronzato, 1997), including e.g. the following:

- The D-optimality (determinant) criterion

$$\Psi(M) = -\log \det M, \tag{9}$$

- The A-optimality (trace) criterion

$$\Psi(M) = \operatorname{trace} M^{-1}, \tag{10}$$

- The sensitivity criterion

$$\Psi(M) = -\operatorname{trace} M. \tag{11}$$

A D-optimum design minimizes the volume of the uncertainty ellipsoid for the estimates. An A-optimum design suppresses the average variance of the estimates. In turn, the sensitivity criterion is often used due to its simplicity, but it sometimes leads to serious problems with identifiability as it may result in a singular FIM (Zarrop and Goodwin, 1975), so in principle it should be used only to obtain startup locations for other criteria. The introduction of an optimality criterion renders it possible to formulate the sensor location problem as an optimization problem.

Two simplifications come in handy, but they involve no loss of generality. Namely, since in practice all design criteria satisfy the condition

$$\Psi(\beta M) = \gamma(\beta) \Psi(M), \quad \beta > 0, \tag{12}$$

γ being a positive function, we may set $\sigma = 1$. Similarly, operating on the so-called *average* (or *normalized*) FIM

$$\bar{M} = \frac{1}{N t_f} \sum_{j=1}^{N} \int_0^{t_f} g(x^j, t) g^{\mathrm{T}}(x^j, t)\, \mathrm{d}t \tag{13}$$

is slightly more convenient, so in the sequel we will constantly use it in lieu of M. For simplicity of notation, we will also drop the bar over M.

3. Extended Formulation Based on Continuous Designs

The introduction of the design criterion Ψ defined on the FIM permits the optimal experimental design to be cast as a minimization problem

$$\Psi\big[M(x^1,\ldots,x^N)\big] \longrightarrow \min, \tag{14}$$

where x^j, $j = 1,\ldots,N$ belong to the set $X \subset \bar{\Omega}$ in which the measurements are allowed. This leads to the so-called exact designs which can then be calculated with the use of numerous widely accessible non-linear programming solvers if N is not too large (c.f. Section 7). Unfortunately, the problem quickly becomes computationally too demanding and intractable for larger N's. A similar predicament in lumped systems has been addressed in plentiful works on optimum experimental design and the most efficient solution therein is no doubt the introduction of the so-called continuous designs (Ermakov, 1983; Fedorov, 1972; Fedorov and Hackl, 1997; Goodwin and Payne, 1977; Pázman, 1986; Pukelsheim, 1993; Walter and Pronzato, 1997). Such an approach will also be adopted in what follows.

Owing to assumption (5), we admit of replicated measurements, i.e. some values x^j may appear several times in the optimal solution (this is an unavoidable consequence of independent measurements). Consequently, it is sensible to distinguish only the components of the sequence $x^1,\ldots,x^N$ which are different and, if there are ℓ such components, to relabel them as $x^1,\ldots,x^\ell$ while introducing $r_1,\ldots,r_\ell$ as the corresponding numbers of replications. The redefined x^i's are said to be the *design* or *support* points. The collection of variables

$$\xi_N = \begin{Bmatrix} x^1, & x^2, & \ldots, & x^\ell \\ p_1, & p_2, & \ldots, & p_\ell \end{Bmatrix}, \tag{15}$$

where $p_i = r_i/N$, $N = \sum_{i=1}^{\ell} r_i$, is called the *exact design* of the experiment. The proportion p_i of observations performed at x^i can be considered as the percentage of experimental effort spent at that point.

On account of the above remarks, we rewrite the FIM in the form

$$M(\xi_N) = \sum_{i=1}^{\ell} p_i \frac{1}{t_f} \int_0^{t_f} g(x^i,t) g^{\mathrm{T}}(x^i,t)\,\mathrm{d}t. \tag{16}$$

Here the p_i's are rational numbers, since both r_i's and N are integers. Removing this constraint by assuming that they can be any real numbers of the interval $[0,1]$ such that $\sum_{i=1}^{\ell} p_i = 1$, we may think of the designs as probability distributions on X. But if so, we may attempt to take one more step to widen the class of admissible designs a bit further, i.e. to all probability measures ξ over X which are absolutely continuous with respect to the Lebesgue measure and satisfy by definition the condition

$$\int_X \xi(\mathrm{d}x) = 1. \tag{17}$$

Such an extension of the design concept allows us to replace (16) by

$$M(\xi) = \int_X \Upsilon(x)\,\xi(\mathrm{d}x), \tag{18}$$

where

$$\Upsilon(x) = \frac{1}{t_f} \int_0^{t_f} g(x,t) g^{\mathrm{T}}(x,t)\, \mathrm{d}t$$

and the integration in (17) and (18) is to be understood in the Stieltjes-Lebesgue sense. This leads to the so-called *continuous* designs which constitute the basis of the modern theory of optimal experiments (Fedorov and Hackl, 1997; Rafajłowicz, 1996; Pukelsheim, 1993; Walter and Pronzato, 1997). It turns out that such an approach drastically simplifies the design.

From now on, $\Xi(X)$ denotes the set of all probability measures on X. Let us also introduce the notation $\mathfrak{M}(X)$ for the set of all admissible information matrices, i.e.

$$\mathfrak{M}(X) = \{M(\xi) : \xi \in \Xi(X)\}. \tag{19}$$

Then we may redefine an optimal design as a solution to the minimization problem

$$\xi^\star = \arg \min_{\xi \in \Xi(X)} \Psi[M(\xi)]. \tag{20}$$

4. Representation Properties of the Information Matrices

Information matrices possess a number of very attractive features which make it possible to take advantage of some results of convex optimization theory. In order to demonstrate them in detail, we need the following assumptions:

(A1) X is compact, and

(A2) $g \in C(X \times T; \mathbb{R}^m)$.

We begin with certain convexity and representation properties of $M(\xi)$.

Lemma 1. *For any $\xi \in \Xi(X)$ the information matrix $M(\xi)$ is symmetric and non-negative definite.*

Proof. The first part is a direct consequence of the definition (18). The other results from the dependence

$$\begin{aligned} b^{\mathrm{T}} M(\xi) b &= \int_X b^{\mathrm{T}} \Upsilon(x) b\, \xi(x) \\ &= \frac{1}{t_f} \int_X \left\{ \int_0^{t_f} [b^{\mathrm{T}} g(x,t)]^2\, \mathrm{d}t \right\} \xi(\mathrm{d}x) \geq 0 \end{aligned} \tag{21}$$

are valid for any $b \in \mathbb{R}^m$. ∎

Lemma 2. *$\mathfrak{M}(X)$ is compact and convex.*

Proof. Let us notice that by Assumption (A2) the function Υ is continuous in X (Kołodziej, 1979, Th. 22, p. 360). Helley's theorem (Ermakov and Zhigljavsky, 1987, Lem. 1.4, p. 91) then implies that $\Xi(X)$ is weakly compact, i.e. from any sequence $\{\xi_i\}_{i=1}^{\infty}$ of $\Xi(X)$ we can extract a subsequence $\{\xi_{i_j}\}_{j=1}^{\infty}$ which is weakly convergent to a probability measure $\xi_\star \in \Xi(X)$ in the sense that

$$\lim_{j\to\infty} \int_X f(x)\,\xi_{i_j}(\mathrm{d}x) = \int_X f(x)\,\xi_\star(\mathrm{d}x), \quad \forall f \in C(X). \tag{22}$$

Choosing f consecutively as the components of the matrix Υ, we get

$$\lim_{j\to\infty} M(\xi_{i_j}) = M(\xi_\star) \tag{23}$$

which establishes the first part of our claim. The other follows immediately from the implication

$$M\big[(1-\lambda)\xi_1 + \lambda\xi_2\big] = (1-\lambda)M(\xi_1) + \lambda M(\xi_2), \quad \forall \xi_1, \xi_2 \in \Xi(X) \tag{24}$$

valid for any $\lambda \in [0,1]$. ■

Remark 1. Let us observe that Assumption (A2) may be slightly weakened: For the continuity of Υ it suffices to require only $g(\,\cdot\,,t)$ to be continuous and to impose the condition

$$\forall x \in X, \ \|g(x,t)\| \le h(t) \tag{25}$$

almost everywhere in T for some $h \in L^2(T)$.

Let us recall that the support of a function $f : X \to \mathbb{R}$ is defined to be the closure of the set of points in $\mathbb{R}^r$ at which f is non-zero. It turns out that, despite a rather abstract framework for continuous designs, the results obtained through their use are surprisingly closely related to discrete designs whose support consists of a finite number of x values. In other words, the optimal design can be chosen to be of the form

$$\xi^\star = \left\{ \begin{matrix} x^1, & x^2, & \ldots, & x^\ell \\ \pi_1, & \pi_2, & \ldots, & \pi_\ell \end{matrix} \; ; \quad \sum_{i=1}^{\ell} \pi_i = 1 \right\}, \tag{26}$$

where $\ell < \infty$, which concentrates $N\pi_1$ measurements at x_1, $N\pi_2$ at x_2, and so on. In fact, we have the following assertion.

Lemma 3. *For any $M_0 \in \mathfrak{M}(X)$ there always exists a purely discrete design ξ with no more than $m(m+1)/2+1$ support points such that $M(\xi) = M_0$. If M_0 lies on the boundary of $\mathfrak{M}(X)$, then the number of support points is less than or equal to $m(m+1)/2$.*

Proof. We first observe that due to the symmetry of FIM's, $\mathfrak{M}(X)$ can be identified with a closed convex set of $\mathbb{R}^{m(m+1)/2}$ (it suffices to use only the elements which are on and above the diagonals). It is easy to check that the average information matrices $M(\xi_x) = \Upsilon(x)$ which correspond to one-point designs $\xi_x = \{\begin{smallmatrix} x \\ 1 \end{smallmatrix}\}$, i.e. the

designs concentrated at a single point x, are the only extreme points of $\mathfrak{M}(X)$. Hence, from Carethéodory's theorem (Pázman, 1986, Prop. III.8, p. 57), the first part of our lemma follows (any point of a compact convex set A of $\mathbb{R}^{m(m+1)/2}$ can be expressed as a convex combination of $m(m+1)/2+1$ or less extreme points of A).

The second part is established based on the assertion that any boundary point of a compact convex set A of $\mathbb{R}^{m(m+1)/2}$ can be expressed as a convex combination of $m(m+1)/2$ or less extreme points of A (Ermakov and Zhigljavsky, 1987, Th. 1.4, p. 96). ■

The above lemma makes it justified to restrict our attention only to discrete designs with a limited number of supporting points, so the introduction of continuous designs, which may seem at first sight a superfluous complication, leads to very tangible results.

5. Characterization of Optimal Solutions

Our objective in this section is to study some properties of optimal designs. For this purpose, the following additional assumptions about the design criterion $\Psi : \mathbb{R}^{m\times m} \to \mathbb{R}$ will be needed:

(A3) Ψ is convex,

(A4) If $M_1 \le M_2$, then $\Psi(M_1) \ge \Psi(M_2)$ (monotonicity),

(A5) There exists a finite real q such that

$$\{\xi : \Psi[M(\xi)] \le q < \infty\} = \Xi(q) \neq \emptyset,$$

(A6) For any $\xi \in \Xi(q)$ and $\bar{\xi} \in \Xi(X)$, we have

$$\Psi\left[M(\xi) + \lambda\big(M(\bar{\xi}) - M(\xi)\big)\right] = \Psi[M(\xi)] + \lambda \int_X \psi(x,\xi)\,\bar{\xi}(\mathrm{d}x) + o(\lambda;\xi,\bar{\xi}), \quad (27)$$

where ψ stands for a function such that $\psi(\,\cdot\,,\xi)$ is continuous in X and o denotes the usual Landau symbol, i.e.

$$\lim_{\lambda\downarrow 0} \frac{o(\lambda;\xi,\bar{\xi})}{\lambda} = 0.$$

Assumption (A3) is quite natural, since it allows us to stay within the framework of convex analysis, which greatly facilitates subsequent considerations. In turn, Assumption (A4) characterizes Ψ as a linear ordering of Ξ. (As regards the notation in (A4), we adopt that of the Loewner ordering of symmetric matrices, i.e. $M_1 \le M_2$ iff $M_2 - M_1$ is non-negative definite.) As for Assumption (A5), it only states that there exist designs with finite values of Ψ, which constitutes a rather mild and quite

logical requirement. At this juncture, only Assumption (A6) calls for an appropriate comment, as at first sight it may seem a bit odd. In practice, however, (A6) simply amounts to the existence of the directional derivative

$$\delta_+\Psi(M(\xi), M(\bar{\xi}) - M(\xi)) = \left.\frac{\partial\Psi[M(\xi) + \lambda(M(\bar{\xi}) - M(\xi))]}{\partial\lambda}\right|_{\lambda=0^+}, \quad (28)$$

whose form must be on one hand specific, i.e. $\int_X \psi(x,\xi)\,\bar{\xi}(\mathrm{d}x)$, but on the other hand, for most practical criteria such a condition is not particularly restrictive.

In fact, requiring Ψ to be differentiable with respect to individual elements of its matrix argument, we obtain

$$\begin{aligned}
&\delta_+\Psi\big(M(\xi), M(\bar{\xi}) - M(\xi)\big) \\
&\qquad = \operatorname{trace}\Big[\overset{\circ}{\Psi}(\xi)\big(M(\bar{\xi}) - M(\xi)\big)\Big] \\
&\qquad = \int_X \operatorname{trace}\Big[\overset{\circ}{\Psi}(\xi)\Upsilon(x)\Big]\,\bar{\xi}(\mathrm{d}x) - \operatorname{trace}\Big[\overset{\circ}{\Psi}(\xi)M(\xi)\Big] \\
&\qquad = \int_X \left\{\operatorname{trace}\Big[\overset{\circ}{\Psi}(\xi)\Upsilon(x)\Big] - \operatorname{trace}\Big[\overset{\circ}{\Psi}(\xi)M(\xi)\Big]\right\}\bar{\xi}(\mathrm{d}x), \quad (29)
\end{aligned}$$

where

$$\overset{\circ}{\Psi}(\xi) = \left.\frac{\partial\Psi(M)}{\partial M}\right|_{M=M(\xi)}$$

and therefore

$$\psi(x,\xi) = c(\xi) - \phi(x,\xi), \quad (30)$$

the functions c and φ being respectively defined as

$$c(\xi) = -\operatorname{trace}\Big[\overset{\circ}{\Psi}(\xi)M(\xi)\Big] \quad (31)$$

and

$$\begin{aligned}
\phi(x,\xi) &= -\operatorname{trace}\Big[\overset{\circ}{\Psi}(\xi)\Upsilon(x)\Big] \\
&= -\frac{1}{t_f}\int_0^{t_f} g^{\mathrm{T}}(x,t)\overset{\circ}{\Psi}(\xi)g(x,t)\,\mathrm{d}t. \quad (32)
\end{aligned}$$

For most popular critera we have (see e.g. Ermakov and Zhigljavsky, 1987, Th. 3.3, p. 309 and Th. 3.4, p. 310)

- If $\Psi(M) = -\ln\det M$, then

$$\overset{\circ}{\Psi}(\xi) = -M^{-1}(\xi),$$

- If $\Psi(M) = \operatorname{trace} M^{-1}$, then

$$\overset{\circ}{\Psi}(\xi) = -M^{-2}(\xi),$$

- If $\Psi(M) = -\operatorname{trace} M$, then

$$\overset{\circ}{\Psi}(\xi) = -I,$$

where I is the identity matrix.

The next result provides a characterization of the optimal designs.

Theorem 1. *Let Assumptions (A1)–(A6) hold. Then:*

(i) An optimal design exists comprising not more than $m(m+1)/2$ points (i.e. one less than predicted by Lemma 3).

(ii) The set of optimal designs is convex.

(iii) A design $\xi^\star$ is optimal iff

$$\min_{x \in X} \psi(x, \xi^\star) = 0. \tag{33}$$

(iv) For any purely discrete optimal design $\xi^\star$, the function $\psi(\,\cdot\,, \xi^\star)$ has value zero at all support points.

Proof. The theorem can be established in exactly the same way as Theorem 2.3.2 of (Fedorov and Hackl, 1997, p. 31) as the explicit form of the FIM is not essential in the proof. ■

It is now clear that the function ψ is of paramount importance in our considerations, as it determines the location of the support points in the optimal design $\xi^\star$ (they are situated among its points of global minimum). Moreover, given any design ξ, it indicates points at which a new observation contributes to the greatest extent. Indeed, adding a new observation atomized at a single point x^+ amounts to constructing a new design

$$\xi^+ = (1 - \lambda)\xi + \lambda \xi_{x^+} \tag{34}$$

for some $\lambda \in (0, 1)$. If λ is sufficiently small, then from (27) it may be concluded that

$$\Psi\big[M(\xi^+)\big] - \Psi\big[M(\xi)\big] \approx \lambda \psi(x^+, \xi), \tag{35}$$

i.e. the resulting decrease in the criterion value is approximately equal to $-\lambda\psi(x^+, \xi)$. This fact also clarifies why the function $\phi(x, \xi) = -\psi(x, \xi) + c(\xi)$ is usually called the *sensitivity function* (Fedorov and Hackl, 1997) (this terminology is somewhat reminiscent of the sensitivity coefficients introduced in (8), but we hope that it will cause no confusion).

Analytical determination of optimal designs is possible only in simple situations and for general systems it is usually the case that some iterative design procedure will be required. The next theorem is useful in checking the optimality of designs.

Theorem 2. *The following characterizations of an optimal design $\xi^\star$ are equivalent in the sense that each implies the other two:*

(i) the design $\xi^\star$ minimizes $\Psi[M(\xi)]$,

(ii) the design $\xi^\star$ minimizes $\max_{x\in X}\phi(x,\xi) - c(\xi)$, and

(iii) $\max_{x\in X}\phi(x,\xi^\star) = c(\xi^\star)$.

All the designs satisfying (i)–(iii) and their convex combinations have the same information matrix $M(\xi^\star)$.

Proof. With minor modifications, it may be adopted, e.g. from (Ermakov and Zhigljavsky, 1987, Th. 2.3, p. 109) and therefore it is omitted. ■

When formulated for a particular design criterion, Theorem 2 is usually called an *equivalence theorem* and the most famous is the Kiefer-Wolfowitz equivalence theorem corresponding to D-optimum designs. In our framework, this specializes to our next assertion.

Theorem 3. *The following conditions are equivalent:*

(i) the design $\xi^\star$ maximizes $\det M(\xi)$,

(ii) the design $\xi^\star$ minimizes
$\max_{x\in X}\frac{1}{t_f}\int_0^{t_f} g^{\mathrm{T}}(x,t)M^{-1}(\xi)g(x,t)\,\mathrm{d}t$, *and*

(iii) $\max_{x\in X}\frac{1}{t_f}\int_0^{t_f} g^{\mathrm{T}}(x,t)M^{-1}(\xi^\star)g(x,t)\,\mathrm{d}t = m.$

An interesting interpretation of continuous designs in terms of the randomized choice is given in (Rafajłowicz, 1986b). Namely, for ξ_N given by (15), if N sensors are randomly allocated to the points x^i, $i = 1,\ldots,\ell \le N$ according to the distribution p_i, $i = 1,\ldots,\ell$ and such that the measurement process is repeated many times, then (16) is the expected value of the FIM. This justifies our results as theoretically exact from a slightly different point of view.

6. Sequential Numerical Design Algorithms

The results of the previous section provide us with tests for the optimality of designs. In particular,

1. If the sensitivity function $\phi(x,\xi)$ is less than or equal to $c(\xi)$ for all $x \in X$, then ξ is optimal.

2. If the sensitivity function $\phi(x,\xi)$ exceeds $c(\xi)$, then ξ is not optimal.

The interesting thing about these results is that in addition to revealing striking minimax properties of optimal designs, they also provide sequential numerical design algorithms. The underlying idea is quite simple. Suppose that we have an arbitrary (non-optimal) design ξ_k obtained after k iteration steps. Further, let $\phi(\,\cdot\,,\xi_k)$ attain its maximum (necessarily $> c(\xi_k)$) at $x = x_k^0$. Then the design

$$\xi_{k+1} = (1-\lambda_k)\xi_k + \lambda_k \xi_{x_k^0} \tag{36}$$

(recall that $\xi_{x_k^0}$ stands for the unit-weight design concentrated at x_k^0) leads to a decrease in the value of $\Psi[M(\xi_{k+1})]$ for a suitably small λ_k. This follows since the derivative with respect to λ_k is negative, i.e.

$$\frac{\partial}{\partial \lambda_k} \Psi\big[M(\xi_{k+1})\big]\Big|_{\lambda_k = 0^+} = c(\xi_k) - \phi(x_k^0, \xi_k) < 0 \tag{37}$$

The steps in using the outlined gradient method can be briefly summarized as follows (Ermakov, 1983; Fedorov and Hackl, 1997; Rafajłowicz, 1996; Walter and Pronzato, 1997):

Step 1. Guess a discrete non-degenerate starting design measure ξ_0 (we must have $\det M(\xi_0) \neq 0$). Choose some positive tolerance $\eta \ll 1$. Set $k = 0$.

Step 2. Determine $x_k^0 = \arg\max_{x \in X} \phi(x, \xi_k)$. If $\phi(x_k^0, \xi_k) < c(\xi_k) + \eta$, then *STOP*.

Step 3. For an appropriate value of $0 < \lambda_k < 1$, set

$$\xi_{k+1} = (1-\lambda_k)\xi_k + \lambda_k \xi_k^0$$

increment k by one and go to Step 2.

In the same way as for the classical first-order algorithms in common use for many years, it can be shown that the above algorithm converges to an optimal design provided that the sequence $\{\lambda_k\}$ is suitably chosen. For example, the choices which satisfy one of the conditions below will yield the convergence:

(i) $\lim\limits_{k\to\infty} \lambda_k = 0$, $\sum\limits_{k=0}^{\infty} \lambda_k = \infty$ (Wynn's algorithm),

(ii) $\lambda_k = \arg\min\limits_{\lambda} \Psi\big[(1-\lambda)M(\xi_k) + \lambda M(\xi_{x_k^0})\big]$ (Fedorov's algorithm),

Computationally, Step 2 is of crucial significance but at the same time it is the most time-consuming step in the algorithm. Complications arise, among other things, due to the necessity of calculating a global maximum of $\phi(\,\cdot\,,\xi_k)$ which is usually multimodal (getting stuck in one of local maxima leads to precocious termination of the algorithm). Therefore, while implementing this part of the computational procedure an effective global optimizer is essential. Based on numerous computer experiments it was found that the extremely simple adaptive random search (ARS) strategy from (Venot *et al.*, 1986; Walter and Pronzato, 1997, p. 216) is especially suited for that purpose if the design region X is a hypercube.

Notwithstanding the fact that the problem outlined in this section is slightly different from the classical formulation encountered in works on optimum experimental design, the details regarding implementations of the corresponding algorithms remain in principle the same and hence this topic will not be further discussed. Instead, we refer the reader to the excellent specialized literature (Ermakov, 1983; Fedorov and Hackl, 1997; Rafajłowicz, 1996; Rafajłowicz, 1998; Skubalska-Rafajłowicz and Rafajłowicz, 1998; Torsney, 1988; Walter and Pronzato, 1997).

At this very moment, some interpretation of the resulting optimal design of the form (26) would be relevant. Since we manipulate continuous designs, the products $N\pi_i$, $i = 1, \ldots, \ell$ are not necessarily integers. In the spatial setting, however, the number of sensors may be quite large and the set of candidate points is continuous so that we can expect that some rounding procedures (Pukelsheim and Rieder, 1992) of the considered approximate designs calculated by the afore-mentioned algorithms will yield sufficiently good designs. Alternatively, some *exchange algorithms* can be adopted from the classical theory of optimal experiments if N is relatively small, but such a procedure does not change the underlying idea and therefore it will not be pursued.

7. Direct Use of Non-Linear Programming Techniques

When the total number of sensors to be located in a given domain is moderate, the very first idea, which suggests itself, is to exploit numerous well-known numerical techniques of constrained optimization. In principle, such an approach is not difficult to apply (Uciński, 1998) and only computation of the gradient of the design criterion necessitates some comments if gradient methods are to be employed.

For abbreviation, write

$$s = (x^1, \ldots, x^N). \tag{38}$$

Accordingly, the design criterion to be minimized may be rewritten as

$$J(s) = \Psi[M(s)], \tag{39}$$

where

$$M(s) = \frac{1}{Nt_f} \sum_{j=1}^{N} \int_0^{t_f} g(x^j, t) g^{\mathrm{T}}(x^j, t)\, \mathrm{d}t. \tag{40}$$

Using the chain rule, we get

$$\frac{\partial J(s)}{\partial s_r} = \operatorname{trace} \left\{ \overset{\circ}{\Psi}(s) \frac{\partial M(s)}{\partial s_r} \right\}, \tag{41}$$

where s_r stands for the r-th component of s, and

$$\overset{\circ}{\Psi}(s) = \left. \frac{\partial \Psi(M)}{\partial M} \right|_{M=M(s)}. \tag{42}$$

As for computation of $\partial M/\partial s_r$, let us observe first that s_r appears at only one term of the sum in (40), since s_r is just a spatial coordinate of one of the sensors. If we use the symbol j_r to denote the index of the corresponding sensor, then obviously we have

$$\frac{\partial M}{\partial s_r} = \frac{1}{Nt_f}\int_0^{t_f}\left\{\frac{\partial g(x^{j_r},t)}{\partial s_r}g^{\mathrm{T}}(x^{j_r},t) + g(x^{j_r},t)\frac{\partial g^{\mathrm{T}}(x^{j_r},t)}{\partial s_r}\right\}\mathrm{d}t. \tag{43}$$

Hence, on account of the symmetry of $\overset{\circ}{\Psi}(s)$, it follows that

$$\frac{\partial J(s)}{\partial s_r} = \frac{2}{Nt_f}\,\mathrm{trace}\left\{\overset{\circ}{\Psi}(s)\int_0^{t_f}\frac{\partial g(x^{j_r},t)}{\partial s_r}g^{\mathrm{T}}(x^{j_r},t)\,\mathrm{d}t\right\}. \tag{44}$$

We see at once that calculation of $\nabla J(s)$ requires an efficient procedure to determine spatial derivatives of the sensitivity coefficients. Note, however, that this does not present a problem if we take advantage of spline interpolation (Uciński, 1999).

Note that direct application of optimization techniques by no means excludes the phenomenon of sensor clusterization which is encountered when a group of different sensors take measurements at the same point. Unfortunately, in many applications this is indesirable since several sensors placed at a point may influence one another and their measurements could then hardly be considered as independent. One way to attempt to avoid this effect is to include into the non-linear programming formulation appropriate constraints on the admissible distances between the sensors. An alternative approach consists in taking account of mutual correlations between the measurements made by different sensors, i.e. we generalize (5) to

$$\mathrm{E}\{\varepsilon(x^i,t)\varepsilon(x^j,t')\} = c_{ij}\delta(t-t') \tag{45}$$

and assume that the covariance matrix $C = [c_{ij}]$ may not be diagonal. For example, its elements could be of the following isotropic form (Nychka and Saltzman, 1998):

$$c_{ij} = \sigma^2\exp\big(-\|x^i - x^j\|/\beta\big), \tag{46}$$

where β is a positive constant. Occasionally, its extension

$$c_{ij} = \sigma(x^i)\sigma(x^j)\exp\big(-\|x^i - x^j\|/\beta\big) \tag{47}$$

is also used, which allows for different marginal variances.

Let us note that if any two sensors are placed at one point, then the corresponding columns (and rows) of C are identical, which means that C becomes singular.

It can be shown (Uciński, 1999) that the average FIM is then given by

$$M(s) = \frac{1}{Nt_f}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_0^{t_f} d_{ij}(s)g(x^i,t)g^{\mathrm{T}}(x^j,t)\,\mathrm{d}t, \tag{48}$$

where d_{ij}'s are the elements if the inverse of C (i.e. $D = [d_{ij}] = C^{-1}$). A first inconvenience is that the form of $M(s)$ is much more cumbersome than in the case of independent measurements. But a more severe difficulty is that the functional

dependence of M on s is much less regular owing to the occurence that C may be singular or nearly singular, which necessitates the notion of pseudo-inverses and involves serious problems with differentiability and numerical stability. Consequently, in practice it is much easier to simply impose additional constraints on the distances between the sensors which will warrant the assumption of independent measurements.

8. Numerical Simulation

In this section we illustrate, by a simple illustration study, the approach to the sensor placement developed in the previous sections. For this purpose, we considered estimation of the spatially-varying parameter $\kappa = \kappa(x)$ in the heat-conduction process through a thin flat isotropic plate whose flat surfaces were insulated and which occupied the region $\Omega = [0,1]^2$ with boundary $\partial\Omega$ along which heat was lost to the surroundings. The unsteady state temperature $y = y(x,t)$ over the time horizon $T = (0,1)$ was described by a linear parabolic equation of the form

$$\frac{\partial y(x,t)}{\partial t} = \frac{\partial}{\partial x_1}\left(\kappa(x)\frac{\partial y(x,t)}{\partial x_1}\right) + \frac{\partial}{\partial x_2}\left(\kappa(x)\frac{\partial y(x,t)}{\partial x_2}\right) \quad \text{in } \Omega \times T. \tag{49}$$

The initial and boundary conditions of (49) were

$$y(x,0) = 5 \qquad \text{in } \Omega, \tag{50}$$

$$y(x,t) = 5(1-t) \quad \text{on } \partial\Omega \times T. \tag{51}$$

In our simulation study, the following true parameter was assumed:

$$\kappa(x) = \theta_1 + \theta_2 x_1 + \theta_3 x_2, \tag{52}$$

where $\theta_1 = 0.1$, $\theta_2 = \theta_3 = 0.3$. On the basis of simulated data generated with the specified κ, we tried to determine a continuous design over $X = \bar{\Omega}$ such that the D-optimality criterion for $\theta = (\theta_1, \theta_2, \theta_3)$ would be minimized.

In order to numerically solve the measurement location problem, a computer programme was written in Essential Lahey Fortran 90 v.4.0 (Meissner, 1997) using a PC (Pentium II, 300 Mhz, 128 MB RAM) running Windows NT 4.0. The state and sensitivity equations were first solved using the finite-element method on an even grid (with 15 divisions along each space axis and 30 divisions of the time interval). The sensitivity coefficients were then interpolated via tri-cubic spline interpolation and the corresponding spline parameters stored in computer memory. Finally, Fedorov's version of the first-order algorithm was applied to maximize the determinant of the FIM. (The maximum number of evaluations for the performance index in the ARS strategy was 2000.)

Starting from the design

$$\xi_0 = \begin{Bmatrix} (0.6,0.2), & (0.2,0.5), & (0.1,0.2) \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$$

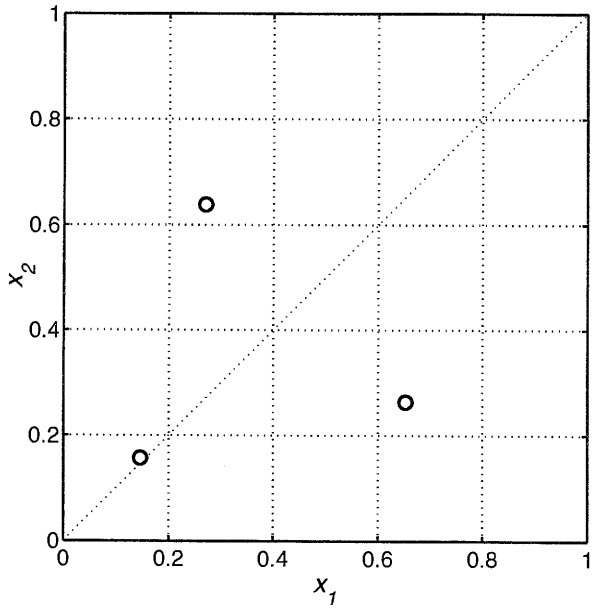


Fig. 1. Support points of the optimal continuous design (the axis of symmetry is represented by a sloping dotted line).

after 13 iterations (which took about two minutes), the following approximation to the optimal design was obtained:

$$\xi^\star = \left\{ \begin{array}{lll} (0.65224, 0.26353), & (0.27083, 0.63834), & \dots \\ 0.33570, & 0.33410, & \dots \\ (0.14647, 0.15668) & & \\ 0.33019 & & \end{array} \right\}$$

for the tolerance $\eta = 10^{-2}$.

The design is concentrated at three support points with approximately equal weights, which means that if we are to locate N sensors, then we should strive to distribute them as evenly as possible between the three calculated potential locations (as outlined before, sensor clusterization is inherent to the approach due to the assumption that the measurements are independent even though some of the sensors take measurements at the same point).

Let us observe that the diffusivity coefficient κ together with the system of boundary and initial conditions assume one axis of symmetry, i.e. the line $x_2 = x_1$. We feel by intuition that this symmetry should also be preserved in a way in the optimal design. In fact, this is confirmed in Fig. 1 where the optimal sensor positions are displayed. They are slightly shifted towards the lower-left part of the system, at

which place the diffusivity coefficient is smaller and the system output is the most sensitive to changes in θ.

In order to assess the applicability of the direct non-linear approach, six sensors were to be placed with the use of the direct non-linear programming approach. At first, the case of independent measurements was tested based on a sequential constrained quadratic programming (SQP) method (cf. Bertsekas, 1999; Miller, 1998; Spellucci, 1998a; Spellucci, 1998b). Starting from an initial solution generated via the ARS procedure, the SQP algorithm found the approximate optimal solution

$$s^\star = \begin{pmatrix} 0.1505197, & 0.1505197, \ldots \\ 0.1505197, & 0.1505197, \ldots \\ 0.2724469, & 0.6376952, \ldots \\ 0.2724469, & 0.6376952, \ldots \\ 0.6376952, & 0.2724469, \ldots \\ 0.6376952, & 0.2724469 \end{pmatrix}$$

shown in Fig. 2(a). This means that we have three pairs of sensors and each of these pairs tends to measure the system state at the same point. In principle, this result should not be surprising, since it was already predicted by the continuous-design approach, where virtually the same support points were obtained. On the other hand, it tallies with some results on replications of D-optimal designs for non-linear models (Haines, 1993).

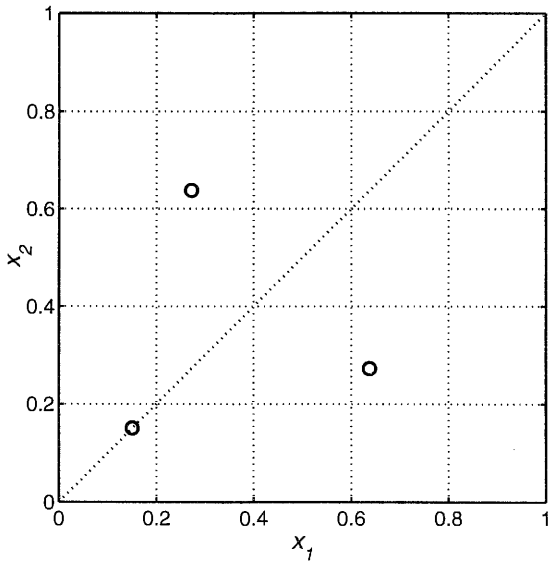
The case of correlated observations was also tested for the model (46) with $\beta = 10^{-2}$. Since the gradient algorithms are not appropriate for this type of performance indices, the ARS technique was employed to assess the optimal solution as

$$s^\star_{\text{corr}} = \begin{pmatrix} 0.128788, & 0.128788, \ldots \\ 0.167856, & 0.167856, \ldots \\ 0.246423, & 0.621963, \ldots \\ 0.301227, & 0.650355, \ldots \\ 0.621963, & 0.246423, \ldots \\ 0.650355, & 0.301226 \end{pmatrix},$$

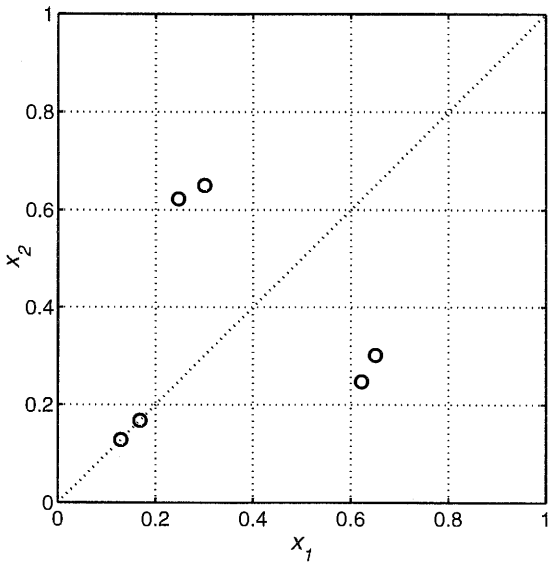
which is illustrated in Fig. 2(b). We see at once that the introduction of interrelations between the sensors results in removing clusterizaton. During experiments, however, some numerical instabilities were observed in addition to a considerably increased computational burden (three minutes versus half a minute for the correlation-free case).

9. Concluding Remarks

An approach to the sensor placement for the purpose of parameter estimation of DPS's has been presented. It consists in the adaptation of some well-known methods



(a)



(b)

Fig. 2. Optimal sensor location calculated via the direct approach: (a) Independent measurements, (b) Correlated measurements.

of modern optimum experimental design for linear regression models. The advantage of introducing continuous designs lies in the fact that the problem dimensionality is dramatically reduced. Moreover, with some minor changes, sequential numerical design algorithms, which have been continually refined since the early 1960s, can be employed here. If the number of sensors is not large, we can always resort to standard optimization routines which ensure that the constraints on the design measure and region are satisfied (as indicated, computation of the gradient does not present a problem). Although the numerical examples presented here are clearly not real-world problems and their purpose is primarily to illustrate our considerations in an easily interpretable manner, they are complex enough to provide evidence for the effectiveness of the proposed approaches.

The approach suggested here has the advantage that it is independent of a particular form of the partial-differential equation describing the distributed system under consideration. The only requirement is the existence of sufficiently regular solutions to the state and sensitivity equations, and consequently non-linear systems can also be treated within the same framework practically without any changes. Moreover, it can easily be generalized to three spatial dimensions, the only limitation being the amount of required computations.

Let us note that, in general, the optimal solutions will depend on the preliminary estimate θ^0 of the unknown parameter vector, cf. (8), as the FIM will depend on it, so logically, the optimal sensor location can never be found at the design stage unless θ^0 is very close to the true parameters or the sensitivity vector g is insensitive to the values of the model parameters (in practice, the latter is unlikely in the considered applications). But this problem is unavoidable in non-linear optimum experimental design (Pázman, 1993). A way out of this predicament is to employ some robust-design concepts (Walter and Pronzato, 1997), but this will constitute the subject of a separate paper.

References

Bertsekas D.P. (1999): *Nonlinear Programming.* — 2-nd Edn., Belmont: Athena Scientific.

Brimkulov U.N., Krug G.K. and Savanov V.L. (1986): *Design of Experiments in Investigating Random Fields and Processes.* — Moscow: Nauka, (in Russian).

Ermakov S.M. (Ed.) (1983): *Mathematical Theory of Experimental Design.* — Moscow: Nauka, (in Russian).

Ermakov S.M. and Zhigljavsky A.A. (1987): *Mathematical Theory of Optimal Experiments.* — Moscow: Nauka, (in Russian).

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

Fedorov V.V. and Hackl P. (1997): *Model-Oriented Design of Experiments.* — New York: Springer.

Goodwin G.C. and Payne R.L. (1977): *Dynamic System Identification. Experiment Design and Data Analysis.* — New York: Academic Press.

Haines L.M. (1993): *Optimal design for nonlinear regression models.* — Communications in Statistics—Theory and Methods, Vol.22, No.6, pp.1613–1627.

Kazimierczyk P. (1989): *Optimal experiment design; vibrating beam under random loading.* — Technical Report 7/1989, Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw.

Kołodziej W. (1979): *Mathematical Analysis.* — Warsaw: Polish Scientific Publishers, (in Polish).

Korbicz J. and Uciński D. (1994): *Sensors allocation for state and parameter estimation of distributed systems.* — Proc. IUTAM Symposium, Zakopane, Poland, 31 August–3 September, 1993, Berlin: Springer, pp.178–189.

Korbicz J., Zgurovsky M.Z. and Novikov A.N. (1988): *Suboptimal sensors location in the state estimation problem for stochastic non-linear distributed parameter systems.* — Int. J. Syst. Sci., Vol.19, No.9, pp.1871–1882.

Malebranche H. (1988): *Simultaneous state and parameter estimation and location of sensors for distributed systems.* — Int. J. Syst. Sci., Vol.19, No.8, pp.1387–1405.

Meissner L.P. (1997): *Essential Fortran 90 & 95. Common Subset Edition.* — Unicomp.

Miller A.J. (1998): *Non-linear programming using DONLP2: Fortran 90 version.* — Code obtainable from `http://www.ozemail.com.au/~milleraj/donlp2.html`.

Munack A. (1984): *Optimal sensor allocation for identification of unknown parameters in a bubble-column loop bioreactor*, In: Analysis and Optimization of Systems, Part 2, (A.V. Balakrishnan and M. Thoma, Eds.). — Berlin: Springer, Vol.63, pp.415–433.

Nychka D. and Saltzman N. (1998): *Design of air-quality monitoring networks*, In: Case Studies in Environmental Statistics (D. Nychka, W.W. Piegorsch and L.H. Cox, Eds.). — New York: Springer, Vol.132, pp.51–76.

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

Pázman A. (1993): *Nonlinear Statistical Models.* — Dordrecht: Kluwer.

Point N., Vande Wouwer A. and Remy M. (1996): *Practical issues in distributed parameter estimation: Gradient computation and optimal experiment design.* — Contr. Eng. Practice, Vol.4, No.11, pp.1553–1562.

Pukelsheim F. (1993): *Optimal Design of Experiments.* — New York: Wiley.

Pukelsheim F. and Rieder S. (1992): *Efficient rounding of approximate designs.* — Biometrica, Vol.79, No.4, pp.763–770.

Quereshi Z.H., Ng T.S. and Goodwin G.C. (1980): *Optimum experimental design for identification of distributed parameter systems.* — Int. J. Contr., Vol.31, No.1, pp.21–29.

Rafajłowicz E. (1978): *Design of experiments for parameter identification of the static distributed systems.* — Syst. Sci., Vol.4, No.4, pp.349–361.

Rafajłowicz E. (1981): *Design of experiments for eigenvalue identification in distributed-parameter systems.* — Int. J. Contr., Vol.34, No.6, pp.1079–1094.

Rafajłowicz E. (1983): *Optimal experiment design for identification of linear distributed-parameter systems: Frequency domain approach.* — IEEE Trans. Automat. Contr., Vol.28, No.7, pp.806–808.

Rafajłowicz E. (1986a): *Choice of Optimum Input Signals in Linear Distributed-Parameter Systems Identification.* — Technical University Press, Wrocław, (in Polish).

Rafajłowicz E. (1986b): *Optimum choice of moving sensor trajectories for distributed parameter system identification.* — Int. J. Contr., Vol.43, No.5, pp.1441–1451.

Rafajłowicz E. (1988): *Distributed parameter systems analysis by tracking stationary points of their states.* — Int. J. Syst. Sci., Vol.19, No.4, pp.613–620.

Rafajłowicz E. (1989): *Reduction of distributed system identification complexity using intelligent sensors.* — Int. J. Contr., Vol.50, No.5, pp.1571–1576.

Rafajłowicz E. (1996): *Algorithms of Experimental Design with Implementations in Mathematica.* — Warsaw: Academic Publishing House, (in Polish).

Rafajłowicz E. (1998): *Selective random search for optimal experiment designs*, In: MODA 5 – Advances in Model-Oriented Data Analysis and Experimental Design (A.C. Atkinson, L. Pronzato and H.P. Wynn, Eds.). — Proc. 5th Int. Workshop in Marseilles, Heidelberg: Physica-Verlag, chapter 9, pp.75–83.

Skubalska-Rafajłowicz E. and Rafajłowicz E. (1998): *Searching for optimal experimental designs using space-filling curves.* — Appl. Math. Comp. Sci., Vol.8, No.3, pp.647–656.

Spellucci P. (1998a): *A new technique for inconsistent QP problems in the SQP method.* — Math. Meth. Ops. Res., Vol.47, pp.355–400.

Spellucci P. (1998b): *An SQP method for general nonlinear programs using only equality constrained subproblems.* — Math. Programm., Vol.82, pp.413–448.

Sun N.-Z. (1994): *Inverse Problems in Groundwater Modeling.* — Dordrecht: Kluwer.

Torsney B. (1988): *Computing optimising distributions with applications in design, estimation and image processing*, In: Optimal Design and Analysis of Experiments (Y. Dodge, V.V. Fedorov and H.P. Wynn, Eds.). — Amsterdam: Elsevier, pp. 316–370.

Uciński D. (1998): *Optimal location of scanning sensors for parameter identification of distributed systems from noisy experimental data.* — Proc. 9-th Int. Symp. *System-Modelling-Control*, Zakopane, Poland, 27 April–1 May, 1998, published on CD-ROM, (P.S. Szczepaniak, Ed.).

Uciński D. (1999): *Measurement Optimization for Parameter Estimation in Distributed Systems.* — Technical University Press, Zielona Góra.

Uciński D. (2000a): *Optimal sensor location for parameter estimation of distributed processes.* — Int. J. Contr., (accepted).

Uciński D. (2000b): *Optimization of sensors' allocation strategies for parameter estimation in distributed systems.* — Systems Analysis–Modelling–Simulation, Vol.37, pp.243–260.

Vande Wouwer A., Point N., Porteman S. and Remy M. (1999): *On a practical criterion for optimal sensor configuration — Application to a fixed-bed reactor.* — Proc. 14-th *IFAC World Congress*, Beijing, China, Vol.I: Modeling, Identification, Signal Processing II, Adaptive Control, pp. 37–42.

Venot A., Pronzato L., Walter É. and Lebruchec J.-F. (1986): *A distribution-free criterion for robust identification, with applications in system modelling and image processing.* — Automatica, Vol.22, No.1, pp.105–109.

Walter É. and Pronzato L. (1997): *Identification of Parametric Models from Experimental Data.* — Berlin: Springer.

Zarrop M.B. and Goodwin G.C. (1975): Comments on *Optimal inputs for system identification.* — IEEE Trans. Automat. Contr., Vol.AC-20, No.2, pp.299–300.

Received: 28 March 2000