Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Sensor Network Design for Parameter Estimation of Distributed Systems Using Nonsmooth Optimality Criteria

Dariusz Uciński

Abstract— **A computational procedure is presented for designing configurations of a measurement network consisting of a large number of stationary sensors collecting data for parameter estimation of a distributed system. Two widely used minimax criteria defined on the Fisher information matrix, namely those of MV- and E-optimality, are considered here as the measures of the estimation accuracy. The approach applied here is to impose constraints on the sensor density in a given spatial domain and to replace the worst-case criteria by their convex smooth approximations. As a result, a fast iterative procedure is obtained whose each step reduces to replacing less informative sensor locations with points which furnish more informaton about the parameters. This planning algorithms is verified by a numerical example on a two-dimensional heat equation.**

I. INTRODUCTION

This work is focused on certain computational aspects of sensor placement problems for distributed-parameter systems (DPSs), i.e., systems with dynamics described by partial differential equations (PDEs). We are interested in locating a large number of discrete pointwise sensors so as to estimate unknown parameters in the underlying mathematical models as accurately as possible. This is an appealing problem since in most applications sensor locations are not pre-specified and therefore provide design parameters.

The importance of sensor planning has already been recognized in many application domains. An example which is particularly stimulating in the light of the results reported in this note constitutes optimization of air quality monitoring networks. One of the tasks of environmental protection systems is to provide expected levels of pollutant concentrations. But to produce such a forecast, a smog prediction model is necessary, which is usually chosen in the form of an advection-diffusion partial differential equation. Its calibration requires parameter estimation, e.g., the unknown spatially-varying turbulent diffusivity tensor should be identified based on the measurements from monitoring stations. Since measurement transducers are usually rather costly and their number is limited, we are faced with the problem of how to optimize their locations in order to obtain the most precise model.

The sensor location problem was attacked from various angles, cf. [1]–[5] for reviews. One of the most serious problems, which complicate the selection of measurement points, is sensor clusterization being a consequence of the assumption that the measurement noise is spatially uncorrelated. This means that in an optimal solution different sensors often tend to take measurements at the same point, which is most often unacceptable from the technical point of view. Indeed, as pointed out in [6], [7], two special features distinguish the spatial data collection schemes from classical regression designs. First of all, spatial observations are often affected by local correlations which are unaccounted for by standard techniques of optimum experimental design. What is more, there is usually no possibility of replicated measurements, i.e., different sensors cannot take measurements at one point without influencing one another. Anyway, several sensors situated in the close vicinity of one another usually do not give more information than a single sensor. The assumption of independent observations is advantageous from a theoretical point of view, since it allows for direct use of sublime results of convex optimization, but it can hardly be justified when in the optimal solution some sensors are to take measurements near one another. This generates interest in the so-called *clusterization-free* designs where the distances between the sensors are long enough in order to guarantee the independence of their measurements. This is reminiscent of the idea of replication-free designs which have emerged relatively late in the context of spatial statistics (see the monograph [6], the survey [7], and the seminal work [8]). It turns out that this idea can be adapted to the sensor location problems for parameter estimation in DPSs with relative ease [1]. However, some generalizations are still expected in this aspect of the sensor location problem.

This paper extends the method of optimizing clusterization-free sensor configurations presented in [1] to include two widely used minimax criteria, namely those of MV- and E-optimality. As is well-known, they are nondifferentiable (in general, only their directional derivatives exist) which highly complicates their use and hinders direct application of the algorithms given in [1], [2]. We propose an approach to overcome these difficulties, which is based on a convex smooth approximation of the original criteria. In spite of its somewhat abstract assumptions, the resulting algorithm of exchange type is very easy to implement.

II. OPTIMAL MEASUREMENT PROBLEM

Let $y = y(x, t; \theta)$ denote the scalar state of a given DPS at a spatial point $x \in \Omega \subset \mathbb{R}^d$ and time instant $t \in T = [0, t_f]$, $t_f < \infty$. Here θ represents an unknown constant parameter vector which must be estimated using observations of the system.

The work was supported by the State Committee for Scientific Research under Grant 3 T11C 035 27.

D. Uciński is with the Institute of Control and Computation Engineering, University of Zielona Góra, 65-246 Zielona Góra, Poland d.ucinski@issi.uz.zgora.pl

0-7803-9568-9/05/$20.00 ©2005 IEEE

In what follows, we consider the observations provided by N stationary pointwise sensors, namely

$$z_{\mathrm{m}}^{j}(t) = y(x^j, t; \theta) + \varepsilon(x^j, t), \quad t \in T, \tag{1}$$

where $z_{\mathrm{m}}^{j}(t)$ is the scalar output and $x^j \in X$ stands for the location of the j-th sensor ($j = 1, \ldots, N$), X signifies the part of the spatial domain Ω where the measurements can be made 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 [4], [5], [9], i.e.,

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

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

Sensor positions which guarantee the best accuracy of the least-squares estimates of θ are then found by choosing x^j, $j = 1, \ldots, N$ so as to minimize some scalar measure of performance Ψ defined on the *average Fisher Information Matrix* (FIM) given by [9]

$$M = \frac{1}{N t_f} \sum_{j=1}^{N} \int_0^{t_f} g(x^j, t) g^{\mathsf{T}}(x^j, t) \, \mathrm{d}t, \tag{3}$$

where $g(x, t) = \nabla_\theta y(x, t; \theta)\big|_{\theta=\theta^0}$ stands for the so-called *sensitivity vector*, θ^0 being a prior estimate to the unknown parameter vector θ [3], [5], [10], [11]. Such a formulation is generally accepted in optimum experimental design for DPSs, since the inverse of the FIM constitutes, up to a constant multiplier, the Cramér-Rao lower bound on the covariance matrix of any unbiased estimator of θ [12]–[14].

As for Ψ, various choices exist for such a function [12]–[14], including e.g., the following:

- the D-optimality (determinant) criterion:

$$\Psi_D(M) = -\log\det(M), \tag{4}$$

- the A-optimality (trace) criterion:

$$\Psi_A(M) = \mathrm{trace}(M^{-1}), \tag{5}$$

- The MV-optimality criterion

$$\Psi_{MV}(M) = \max_{1 \le i \le m} d_{ii}(M), \tag{6}$$

- The E-optimality criterion

$$\Psi_E(M) = \lambda_{\max}(M^{-1}), \tag{7}$$

where $d_{ii}(M)$ stands for the i-th element on the diagonal of M^{-1}, and $\lambda_{\max}(M^{-1})$ denotes the maximum eigenvalue of M^{-1}.

A D-optimum design minimizes the volume of the uncertainty ellipsoid for the estimates. In turn, an A-optimum design suppresses the average variance of the estimates. In an MV-optimum design, the maximal variance of the estimates $\hat{\theta}_1, \ldots, \hat{\theta}_m$ is minimized. On the other hand, while minimizing the E-optimality criterion, the length of the largest principal axis of the uncertainty ellipsoid of the estimates is suppressed.

The introduction of an optimality criterion renders it possible to formulate the sensor location problem as an optimization problem

$$\Psi\big[M(x^1, \ldots, x^N)\big] \longrightarrow \min \tag{8}$$

with respect to x^j, $j = 1, \ldots, N$ belonging to the admissible set X.

Owing to assumption (2), 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

$$\xi_N = \begin{Bmatrix} x^1, & x^2, & \ldots, & x^\ell \\ p_1, & p_2, & \ldots, & p_\ell \end{Bmatrix}, \tag{9}$$

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

$$M(\xi_N) = \sum_{i=1}^{\ell} p_i \frac{1}{t_f} \int_0^{t_f} g(x^i, t) g^{\mathsf{T}}(x^i, t) \, \mathrm{d}t. \tag{10}$$

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. Such an extension of the design concept allows us to replace (10) by

$$M(\xi) = \int_X \Upsilon(x) \, \xi(\mathrm{d}x), \tag{11}$$

where

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

and the integration in (11) 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 [12], [14]. It turns out that such an approach drastically simplifies the design.

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

$$\xi^\star = \arg \min_{\xi \in \Xi(X)} \Psi[M(\xi)], \tag{12}$$

where $\Xi(X)$ is the set of all probability measures on X.

III. SENSOR NETWORK DESIGN FOR SMOOTH CRITERIA

Let us make the following assumptions:

(A1) X is compact,
(A2) $g \in C(X \times T; \mathbb{R}^m)$,
(A3) Ψ is convex,
(A4) If $M_1 \preceq M_2$, then $\Psi(M_1) \geq \Psi(M_2)$,
(A5) There exists a finite real q such that

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

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

$$\begin{aligned}\Psi[M(\xi) + \lambda(M(\bar{\xi}) - M(\xi))] \\ = \Psi[M(\xi)] + \lambda \int_X \psi(x,\xi)\,\bar{\xi}(\mathrm{d}x) \\ + o(\lambda;\xi,\bar{\xi}),\end{aligned} \tag{13}$$

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

As regards the notation in (A4), we adopt that of the Loewner ordering of symmetric matrices, i.e., $M_1 \preceq M_2$ iff $M_2 - M_1$ is non-negative definite. Note that (A6) simply amounts to the existence of the directional derivative, but 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 get [1], [2]

$$\psi(x,\xi) = c(\xi) - \phi(x,\xi), \tag{14}$$

the functions c and ϕ being respectively defined as

$$c(\xi) = -\operatorname{trace}\Big[\overset{\circ}{\Psi}(\xi) M(\xi)\Big], \tag{15}$$

$$\phi(x,\xi) = -\frac{1}{t_f}\int_0^{t_f} g^{\mathsf{T}}(x,t)\overset{\circ}{\Psi}(\xi) g(x,t)\,\mathrm{d}t, \tag{16}$$

where

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

For particular criteria we have, e.g.,

- the D-optimality criterion:

$$\overset{\circ}{\Psi}(\xi) = -M^{-1}(\xi), \tag{17}$$

- the A-optimality criterion:

$$\overset{\circ}{\Psi}(\xi) = -M^{-2}(\xi). \tag{18}$$

In order to avoid clustered sensor configurations, we implement the idea of operating on the density of sensors (i.e., the number of sensors per unit area), rather than on the sensors' locations, which is justified for a sufficiently large total number of sensors N. In contrast to the classical designs discussed in Section II, however, we impose the crucial restriction that the density of sensor allocation must not exceed some prescribed level. For a design measure $\xi(\mathrm{d}x)$ this amounts to the condition

$$\xi(\mathrm{d}x) \leq \omega(\mathrm{d}x), \tag{19}$$

where $\omega(\mathrm{d}x)$ signifies the maximal possible 'number' of sensors per $\mathrm{d}x$ [1], [2], [14] such that

$$\int_X \omega(\mathrm{d}x) \geq 1. \tag{20}$$

Consequently, we are faced with the following optimization problem: Find

$$\xi^\star = \arg\min_{\xi \in \Xi(X)} \Psi(\xi) \tag{21}$$

subject to

$$\xi(\mathrm{d}x) \leq \omega(\mathrm{d}x). \tag{22}$$

The design $\xi^\star$ above is then said to be a (Ψ,ω)-*optimal design* [14].

Apart from Assumptions (A1)–(A6), a proper mathematical formulation calls for the following proviso:

(A7) $\omega(\mathrm{d}x)$ is atomless, i.e., for any $\Delta X \subset X$ there exists a $\Delta X' \subset \Delta X$ such that

$$\int_{\Delta X'} \omega(\mathrm{d}x) < \int_{\Delta X} \omega(\mathrm{d}x). \tag{23}$$

In what follows, we write $\bar{\Xi}(X)$ for the collection of all the design measures which satisfy the requirement

$$\xi(\Delta X) = \begin{cases} \omega(\Delta X) & \text{for } \Delta X \subset \operatorname{supp}\xi, \\ 0 & \text{for } \Delta X \subset X \setminus \operatorname{supp}\xi. \end{cases} \tag{24}$$

Given a design ξ, we will say that the function $\psi(\,\cdot\,,\xi)$ defined by (14) separates sets X_1 and X_2 with respect to $\omega(\mathrm{d}x)$ if for any two sets $\Delta X_1 \subset X_1$ and $\Delta X_2 \subset X_2$ with equal non-zero measures we have

$$\int_{\Delta X_1} \psi(x,\xi)\,\omega(\mathrm{d}x) \leq \int_{\Delta X_2} \psi(x,\xi)\,\omega(\mathrm{d}x). \tag{25}$$

We can now formulate the main result which provides a characterization of (Ψ,ω)-optimal designs.

Theorem 1 ([1], [2]): Let Assumptions (A1)–(A7) hold. Then:

(i) *There exists an optimal design $\xi^\star \in \bar{\Xi}(X)$, and*
(ii) *A necessary and sufficient condition for $\xi^\star \in \bar{\Xi}(X)$ to be (Ψ,ω)-optimal is that $\psi(\,\cdot\,,\xi^\star)$ separates $X^\star = \operatorname{supp}\xi^\star$ and its complement $X \setminus X^\star$ with respect to the measure $\omega(\mathrm{d}x)$.*

From a practical point of view, the above result means that at all the support points of an optimal design $\xi^\star$ the mapping $\psi(\,\cdot\,,\xi^\star)$ should be less than anywhere else, i.e., preferably $\operatorname{supp}\xi^\star$ should coincide with minimum points of $\psi(\,\cdot\,,\xi^\star)$ (let us note that for the D-optimality criterion this can be expressed as the situation when $\phi(\,\cdot\,,\xi^\star)$ is greater in $\operatorname{supp}\xi^\star$ than in the complement of $\operatorname{supp}\xi^\star$), which amounts to allocating observations to the points at which we know least of all about the system response.

If we were able to construct a design with this property, then it would be qualified as an optimal design. This conclusion forms a basis for numerical algorithms of constructing solutions to the problem under consideration.

As regards the interpretation of the resultant optimal designs (provided that we are in a position to calculate at least their approximations), one possibility is to partition X into subdomains ΔX_i of relatively small areas and then to allocate to each of them the number

$$N^\star(\Delta X_i) = \left\lceil N \int_{\Delta X_i} \xi^\star(\mathrm{d}x) \right\rceil \tag{26}$$

of sensors whose positions may coincide with nodes of some uniform grid [14] (here $\lceil \zeta \rceil$ is the smallest integer greater than or equal to ζ). Additionally, bear in mind that we must also have $\xi^\star(\mathrm{d}x) = \omega(\mathrm{d}x)$ in $X^\star$.

Clearly, unless the considered design problem is quite simple, we must employ a numerical algorithm to make the outlined idea useful. Since $\xi^\star(\mathrm{d}x)$ should be non-zero in the areas where $\psi(\,\cdot\,, \xi^\star)$ takes on a smaller value, the central idea is to move some measure from areas with higher values of $\psi(\,\cdot\,, \xi_k)$ to those with smaller values, as we expect that such a procedure will improve ξ_k. This is embodied by the iterative algorithm presented below:

Step 1. Guess an initial design $\xi_0 \in \bar{\Xi}(X)$. Set $k = 0$.

Step 2. Set $X_{1k} = \operatorname{supp} \xi_k$ and $X_{2k} = X \setminus X_{1k}$. Find

$$x_{1k} = \arg \max_{x \in X_{1k}} \psi(x, \xi_k),$$
$$x_{2k} = \arg \min_{x \in X_{2k}} \psi(x, \xi_k).$$

If $\psi(x_{1k}, \xi_k) < \psi(x_{2k}, \xi_k) + \eta$, where $\eta \ll 1$, then STOP. Else, find two sets $S_{1k} \subset X_{1k}$ and $S_{2k} \subset X_{2k}$ such that $x_{1k} \in S_{1k}$, $x_{2k} \in S_{2k}$ and

$$\int_{S_{1k}} \omega(\mathrm{d}x) = \int_{S_{2k}} \omega(\mathrm{d}x) = \alpha_k$$

(i.e., the measures of S_{1k} and S_{2k} must be identical) for some $\alpha_k > 0$.

Step 3. Construct ξ_{k+1} such that

$$\operatorname{supp} \xi_{k+1} = X_{1,k+1} = (X_{1k} \setminus S_{1k}) \cup S_{2k}.$$

Increment k and to go Step 2.

Convergence is guaranteed if the sequence $\{\alpha\}_{k=0}^{\infty}$ satisfies the conditions [15]

$$\lim_{k \to \infty} \alpha_k = 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty. \tag{27}$$

Within the framework of sensor placement, we usually have $\omega(\mathrm{d}x) = \varrho(x)\mathrm{d}x$, where ϱ is a density function. But in this situation we may restrict our attention to constant ϱ's (indeed, in any case we can perform an appropriate change of coordinates). Moreover, while implementing the algorithm on a computer, all integrals are replaced by sums over some regular grid elements. Analogously, the sets X, X_{1k}, X_{2k}, S_{1k} and S_{2k} then simply consist of grid elements. Consequently, the above iterative procedure may be considered as an exchange-type algorithm with the additional constraint that every grid element must not contain more than one supporting point and the weights of all supporting points are equal to $1/N$. In practice, α_k is usually fixed and, what is more, one-point exchanges are most often adopted, i.e., $S_{1k} = \{x_{1k}\}$ and $S_{2k} = \{x_{2k}\}$, which substantially simplifies implementation. Let us note, however, that convergence to an optimal design is assured only for decreasing α_k's and hence some oscillations in $\Psi[M(\xi_k)]$ may sometimes be observed. A denser spatial grid usually constitutes a remedy for this predicament [6].

IV. MINIMIZATION OF THE MV-OPTIMALITY CRITERION

The MV-optimality criterion

$$\Psi_{MV}(M) = \max\{d_{11}(M), \ldots, d_{mm}(M)\}. \tag{28}$$

is not differentiable (this is because the max function is nondifferentiable), which essentially complicates its minimization.

The approach suggested here is to approximate Ψ_{MV} by a smooth symmetric exponential penalty function [16, p. 248]

$$\Psi_{MV}^{\varepsilon}(M) = \varepsilon \ln\Big(\sum_{i=1}^{m} \mathrm{e}^{d_{ii}(M)/\varepsilon}\Big), \tag{29}$$

where $\varepsilon > 0$ is a parameter.

It is a C^∞ convex function and it is easy to check that it possesses the following uniform approximation property to Ψ_{MV}:

$$0 \le \Psi_{MV}^{\varepsilon}(M) - \Psi_{MV}(M) \le \varepsilon \ln(m). \tag{30}$$

Indeed, we have

$$\begin{aligned} \Psi_{MV}^{\varepsilon}(M) &= \varepsilon \ln\Big(\sum_{i=1}^{m} \mathrm{e}^{d_{ii}(M)/\varepsilon}\Big) \\ &= \Psi_{MV}(M) + \varepsilon \ln\Big(\sum_{i=1}^{m} \mathrm{e}^{(d_{ii}(M) - \Psi_{MV}(M))/\varepsilon}\Big). \end{aligned} \tag{31}$$

But

$$1 \le \sum_{i=1}^{m} \mathrm{e}^{(d_{ii}(M) - \Psi_{MV}(M))/\varepsilon} \le m, \tag{32}$$

which proves (30).

In this manner, we get

$$\lim_{\varepsilon \downarrow 0} \Psi_{MV}^{\varepsilon}(M) = \Psi_{MV}(M), \quad \forall M \succ 0, \tag{33}$$

i.e., we have a family of smooth functions which are uniform approximations to the largest eigenvalue function with the accuracy controlled by the smoothing parameter ε. Consequently, instead of operating on the criterion $\Psi_{MV}(M)$, we may work with its smoothed version $\Psi_{MV}^{\varepsilon}(M)$ for a fixed, sufficiently small ε.

It follows that

$$\frac{\partial \Psi_{MV}^{\varepsilon}(M)}{\partial M} = -M^{-1} A(M) M^{-1}, \tag{34}$$

where

$$\begin{aligned} A(M) = &\left(\sum_{i=1}^{m} \mathrm{e}^{d_{ii}(M)/\varepsilon}\right)^{-1} \\ &\times \operatorname{diag}\big[\mathrm{e}^{d_{11}(M)/\varepsilon}, \ldots, \mathrm{e}^{d_{mm}(M)/\varepsilon}\big]. \end{aligned} \tag{35}$$

As indicated in [16, p. 249], it turns out that when ε is set in the range of 10^{-2}–10^{-6}, Ψ_{MV}^{ε} yields and excellent approximation to Ψ_{MV}. However, care must be taken when trying to implement a computer code, since the exponentials may sometimes lead to numerical complications such as, e.g., problems with a highly ill-conditioned Hessian.

V. CASE OF THE E-OPTIMALITY CRITERION

The criterion (7) is nondifferentiable, either. In fact, if an eigenvalue is repeated, then only its directional derivative exists which is strongly nonlinear in the increment in M (see, e.g., [17]). There appears to be no known specialized algorithm for constructing E-optimum designs such as those discussed in Section III for differentiable criteria. Some attempts at adapting a bundle trust method were reported in [18] and efficient algorithms of semidefinite programming can be applied to find optimal designs for a fixed set of support points, cf. [1, p. 61], but nevertheless the problem is far from being satisfactorily solved.

From a practical point of view, a way out of this predicament can be the reduction of the original minimax problem to the minimization of its smooth convex approximation. Based on the results of the previous section, after replacing the minimization of $\lambda_{\max}[M^{-1}(\xi)]$ by the equivalent problem of minimizing $J(\xi) = \lambda_{\max}[-M(\xi)]$ (this is to avoid matrix inversion), we can consider the smoothed convex approximated design criterion

$$\Psi_E^{\varepsilon}[M(\xi)] = \varepsilon \ln\Big(\sum_{i=1}^{m} \mathrm{e}^{-\lambda_i[M(\xi)]/\varepsilon}\Big), \tag{36}$$

where $\lambda_i[\,\cdot\,]$ is the i-th eigenvalue of its matrix argument and $0 < \varepsilon \ll 1$ stands for a fixed parameter steering the accuracy of such an approximation.

Then the exchange algorithm of Section III can be directly employed to find approximated E-optimum designs. While implementing this idea, it is necessary to use the information provided by the gradient which has the following form:

$$\begin{aligned}\frac{\partial \Psi_E^{\varepsilon}(M)}{\partial M} = &-\left(\sum_{i=1}^{m} \mathrm{e}^{-\lambda_i(M)/\varepsilon}\right)^{-1} \\ &\times \sum_{i=1}^{m} \mathrm{e}^{-\lambda_i(M)/\varepsilon} v_i(M) v_i^{\mathsf{T}}(M),\end{aligned} \tag{37}$$

where $v_i(M)$ is the normalized eigenvector corresponding to the eigenvalue $\lambda_i(M)$.

VI. SIMULATION EXAMPLE

Having developed the approach for calculation of clusterization-free designs in the case of nonsmooth MV- and E-optimality criteria, we go straight to a demonstrative example. For this purpose, consider estimation of the spatially-varying parameter $\kappa = \kappa(x)$ in the heat-conduction process through a thin flat isotropic plate whose flat surfaces are insulated and which occupies the region $\Omega = [0,1]^2 \setminus D$, where D is the disc of radius 0.2 centred at point $(0.5, 0.5)$, with boundary $\partial\Omega$ along which heat was lost to the surroundings. The unsteady state temperature $y = y(x,t)$ over the time horizon $T = (0,1)$ is described by a linear parabolic equation of the form

$$\begin{aligned}\frac{\partial y(x,t)}{\partial t} = &\frac{\partial}{\partial x_1}\left(\kappa(x)\frac{\partial y(x,t)}{\partial x_1}\right) \\ &+ \frac{\partial}{\partial x_2}\left(\kappa(x)\frac{\partial y(x,t)}{\partial x_2}\right) \quad \text{in } \Omega \times T.\end{aligned} \tag{38}$$

The initial and boundary conditions of (38) are

$$y(x,0) = 5 \qquad \text{in } \Omega, \tag{39}$$

$$y(x,t) = 5(1-t) \quad \text{on } \partial\Omega \times T. \tag{40}$$

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

$$\kappa(x) = \theta_1 + \theta_2 x_1 + \theta_3 x_2, \tag{41}$$

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 designs over $X = \bar{\Omega}$ such that the approximated MV- and E-optimality criteria for $\theta = (\theta_1, \theta_2, \theta_3)$ would be minimized.

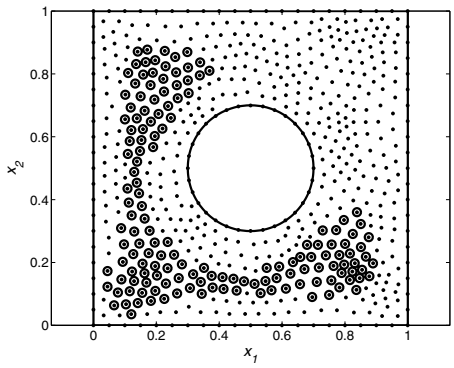
In order to numerically solve the measurement location problem, a computer programme was written in Matlab 7 (R14) using a low-cost PC (Pentium 4, 2.40 GHz, 512 MB RAM) running Windows 2000. The state and sensitivity equations (cf. [1], [2]) were first solved using the MATLAB PDE toolbox [19] on a spatial grid consisting of 624 nodes. The sensitivity coefficients were then stored in computer memory, cf. [1, App. I]

The problem of locating $N = 150$ sensors was considered. For that purpose, the above-mentioned finite-element grid was used to approximate the design space and an initial design was generated by randomly selecting its support points. In order to calculate approximated MV- and E-optimal designs, a simple one-point correction algorithm was employed (after setting $\eta = 10^{-2}$ and $\varepsilon = 10^{-2}$) which quickly produced (practically, in several seconds) the solutions displayed in Figs. 1(b) and (c). For comparison, the D-optimal design is shown in Fig. 1(a). In order to prevent getting stuck in a local minimum, several restarts were performed from different starting solutions.

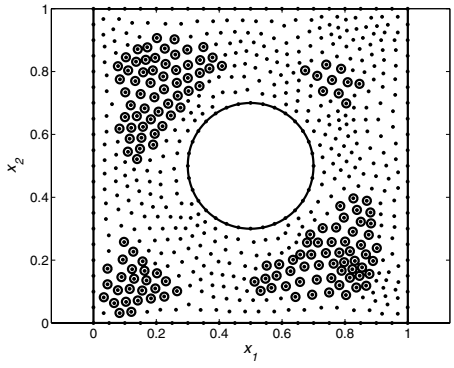
In principle, the MV- and E-optimum configurations look similar, and they involve a deployment of sensors over a wider area (in the D-optimum configuration there are no sensors in the top-right part of Ω). The computed values of the optimality criteria are summarized in Table 1. The different designs are additionally compared there with respect to all the employed design criteria (each row corresponds to a configuration which is optimal in the sense of the criterion given in the first column, and each box indicates the best value of the criterion considered in the respective column). Also note that, as expected, the symmetry imposed by (39)–(41) is retained (cf. the axis of symmetry expressed by the sloping dotted line).

VII. CONCLUDING REMARKS

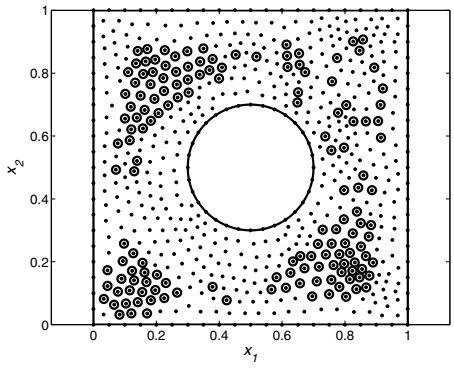
The results contained in this paper show that the difficulties associated with the nondifferentiability of worst-case design criteria used for determination of optimal sensor



(a)



(b)



(c)

Fig. 1. Optimum sensor configurations: D-optimum design (a), approximated MV-optimum design (b) and approximated E-optimum design (c). Dots represent the grid points (these were potential sites where the sensors could be placed, but at most one sensor at one point) and open circles indicate the actual sensor positions.

TABLE I

COMPUTED VALUES OF THE DESIGN CRITERIA.

Criterion	$-\ln\det(M)$	$\lambda_{\max}(M^{-1})$	$\max d_{ii}(M)$
D-optimality	$\boxed{-7.234}$	0.3730	0.2966
E-optimality	-6.771	$\boxed{0.3097}$	0.2916
MV-optimality	-6.947	0.3136	$\boxed{0.2705}$

locations can be overcome to a great extent by appropriately introducing their convex smooth approximations. The approximation accuracy can be steered by selecting a proper value of a scalar parameter. In order to avoid sensor clusterization, we sought to find an optimal design, not within the class of all designs, but rather in a restricted subset of competing clusterization-free designs. As a consequence, this led to a very efficient and particularly simple exchange-type algorithm.

Although the numerical example presented here is clearly not a real-world problem and its purpose is primarily to illustrate our considerations in an easily interpretable manner, it is complex enough to provide evidence for the effectiveness of the proposed approache.

REFERENCES

[1] D. Uciński, *Optimal Measurement Methods for Distributed-Parameter System Identification*. Boca Raton, FL: CRC Press, 2005.

[2] ——, *Measurement Optimization for Parameter Estimation in Distributed Systems*. Zielona Góra: Technical University Press, 1999.

[3] ——, "Optimal sensor location for parameter estimation of distributed processes," *International Journal of Control*, vol. 73, no. 13, pp. 1235–1248, 2000.

[4] C. S. Kubrusly and H. Malebranche, "Sensors and controllers location in distributed systems — A survey," *Automatica*, vol. 21, no. 2, pp. 117–128, 1985.

[5] N.-Z. Sun, *Inverse Problems in Groundwater Modeling*, ser. Theory and Applications of Transport in Porous Media. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1994.

[6] W. G. Müller, *Collecting Spatial Data. Optimum Design of Experiments for Random Fields*, 2nd ed., ser. Contributions to Statistics. Heidelberg: Physica-Verlag, 2001.

[7] V. V. Fedorov, "Design of spatial experiments: Model fitting and prediction," Oak Ridge National Laboratory, Oak Ridge, TN, Tech. Rep. TM-13152, 1996.

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

[9] Z. H. Quereshi, T. S. Ng, and G. C. Goodwin, "Optimum experimental design for identification of distributed parameter systems," *International Journal of Control*, vol. 31, no. 1, pp. 21–29, 1980.

[10] E. Rafajłowicz, "Design of experiments for eigenvalue identification in distributed-parameter systems," *International Journal of Control*, vol. 34, no. 6, pp. 1079–1094, 1981.

[11] ——, "Optimal experiment design for identification of linear distributed-parameter systems: Frequency domain approach," *IEEE Transactions on Automatic Control*, vol. 28, no. 7, pp. 806–808, July 1983.

[12] É. Walter and L. Pronzato, *Identification of Parametric Models from Experimental Data*, ser. Communications and Control Engineering. Berlin: Springer-Verlag, 1997.

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

[14] V. V. Fedorov and P. Hackl, *Model-Oriented Design of Experiments*, ser. Lecture Notes in Statistics. New York: Springer-Verlag, 1997.

[15] V. V. Fedorov, "Optimal design with bounded density: Optimization algorithms of the exchange type," *Journal of Statistical Planning and Inference*, vol. 22, pp. 1–13, 1989.

[16] E. Polak, *Optimization. Algorithms and Consistent Approximations*, ser. Applied Mathematical Sciences. New York: Springer-Verlag, 1997.

[17] E. J. Haug, K. K. Choi, and V. Komkov, *Design Sensitivity Analysis of Structural Systems*, ser. Mathematics in Science and Engineering. Orlando, FL: Academic Press, 1986.

[18] F. Pukelsheim and A. Wilhelm, "A contribution to the discussion of the paper "Constrained optimization of experimental design" by D. Cook and V. Fedorov," *Statistics*, vol. 26, pp. 168–172, 1995.

[19] COMSOL AB, *Partial Differential Equation Toolbox for Use with Matlab. User's Guide*. Natick, MA: The MathWorks, Inc., 1995.