amcs

CONFIGURING A SENSOR NETWORK FOR FAULT DETECTION IN DISTRIBUTED PARAMETER SYSTEMS

MACIEJ PATAN, DARIUSZ UCIŃSKI

Institute of Control and Computation Engineering
University of Zielona Góra
ul. Podgórna 50, 65–246 Zielona Góra, Poland
`{M.Patan,D.Ucinski}@issi.uz.zgora.pl`

The problem of fault detection in distributed parameter systems is formulated as that of maximizing the power of a parametric hypothesis test which checks whether or not system parameters have nominal values. A computational scheme is provided for the design of a network of observation locations in a spatial domain that are supposed to be used while detecting changes in the underlying parameters of a distributed parameter system. A considered setting relates to situation where from among a finite set of potential sensor locations only a subset of them can be selected because of the cost constraints. As a suitable performance measure the D_s-optimality criterion defined on the Fisher information matrix for the estimated parameters is applied. Then, the solution of a resulting combinatorial problem is determined based on the branch-and-bound method. As its essential part, a relaxed problem is discussed in which the sensor locations are given *a priori* and the aim is to determine the associated weights, which quantify the contributions of individual gaged sites. The concavity and differentiability properties of the criterion are established and a gradient projection algorithm is proposed to perform the search for the optimal solution. The delineated approach is illustrated by numerical example on a sensor network design for a two-dimensional convective diffusion process.

Keywords: branch-and-bound, constrained experimental design, distributed parameter systems, fault detection, parameter estimation, sensor location,

1. Introduction

Experimental design for spatio-temporal physical systems (especially, in environmental protection, nuclear energetics, oil industry, etc.) is often related to an optimal choice of measurement conditions in order to obtain the best information for estimating unknown parameters which can then be used in optimal control. The impossibility to observe the system states over the entire spatial domain implies the question of where to locate discrete sensors so as to accurately estimate the unknown system parameters. An example which is particularly stimulating in the light of the results reported in this note constitutes optimization of air quality monitoring networks. The question of how to optimize sensor locations acquires especially vital importance in the context of recent advances in distributed sensor networks (Cassandras and Li, 2005).

Over the past years, laborious research on the development of strategies for efficient sensor placement has been conducted (for reviews, see papers (Kubrusly and Malebranche, 1985; Uciński, 1992; Uciński, 2000; van de Wal and de Jager, 2001) and comprehensive monographs (Uciński, 2005; Uciński, 1999)). Nevertheless, although the need for systematic methods was widely recognized, most techniques communicated by various authors usually rely on exhaustive search over a predefined set of candidates and the combinatorial nature of the design problem is taken into account very occasionally (van de Wal and de Jager, 2001). Needless to say that this approach, which is feasible for a relatively small number of possible locations, soon becomes useless as the number of possible location candidates increases.

Additionally, in spite of the rapid development of fault detection and localization methods for dynamic systems (Korbicz, Kościelny, Kowalczuk and Cholewa, 2004; Isermann, 1997; Patton and Korbicz, 1999; Patton, Frank and Clark, 2000; Chiang, Russell and Braatz, 2001), there are no effective methods tailored to spatiotemporal systems. Some successful attempts at exploiting the D_s-optimality criterion were reported by Patan and Patan (2005). The aim of the research reported here was to

develop a practical approach to sensor selection for fault detection which, while being independent of a particular model of the dynamic DPS in question, would be versatile enough to cope with practical monitoring networks consisting of many stationary sensors. Specifically, we consider N possible sites at which to locate a sensor, but limitations on the number of sensors at our disposal allow only n of them (typically, n is much smaller than N) to be selected. Consequently, the problem is to divide the N available sites between n gauged sites and the remaining $N-n$ ungauged sites so as to maximize the determinant of the Fisher Information Matrix (FIM) associated with the parameters to be estimated. Since selecting the best subset of sites to locate the sensors constitutes an inherently discrete large-scale resource allocation problem whose solution may be prohibitively time-consuming, an efficient guided search algorithm based on the branch-and-bound method is developed, which implicitly enumerates all the feasible sensor configurations, using relaxed optimization problems that involve no integer constraints.

The idea of using branch-and-bound (BB) for sensor selection in DPS was already applied by Uciński and Patan (2007) in the context of D-optimality. Obviously, BB constitutes one of the most frequent approaches to solve discrete optimization problems and it has indeed been used in the context of network design (Boer, Hendrix and Rasch, 2001). Nevertheless, the main contribution of this paper consists in the creative usage of the simple and powerful computational scheme to obtain upper bounds to the optimal values of the D_s-optimality criterion which proves especially attractive in the light of applications in fault detection and diagnosis. Moreover, various properties of the D_s-optimality criterion are given and a relaxed problem is introduced whose solution is obtained through a gradient projection algorithm.

The paper has structure as follows. Section 2 states formally the sensor network design problem as a discrete resource allocation problem. The general BB scheme for its solution is discussed in Sect. 3. Section 4 describes the algorithmic solution for computing upper bounds required by the branching rule. In Sect. 5, we illustrate the use of our algorithm on the example of a sensor network design problem regarding a two-dimensional convective diffusion process. We conclude in Sect. 6 with some comments on related open problems. The proofs of some essential properties of the D_s-optimality criterion are contained in the Appendix.

Notation. Throughout the paper, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ stand for the sets of nonnegative and positive real numbers, respectively. We adopt the convention that all vectors have column form. The set of real $m \times n$ matrices is denoted by $\mathbb{R}^{m \times n}$. We use $\mathbb{S}^m$ to denote the set of symmetric $m \times m$ matrices, $\mathbb{S}^m_+$ to denote the set of symmetric nonnegative definite $m \times m$ matrices, and $\mathbb{S}^m_{++}$ to denote the set of

symmetric positive definite $m \times m$ matrices. The curled inequality symbol $\succeq$ (resp. $\succ$) is used to denote generalized inequalities. More precisely, between vectors, it represents a componentwise inequality, and between symmetric matrices, it represents the Löwner ordering: given $A, B \in \mathbb{S}^m$, $A \succeq B$ means that $A - B$ is nonnegative definite (resp. positive definite).

2. Sensor selection for fault detection

Consider a bounded spatial domain $\Omega \subset \mathbb{R}^d$ with sufficiently smooth boundary Γ, a bounded time interval $T = (0, t_f]$, and a distributed parameter system (DPS) whose scalar state at a spatial point $x \in \bar{\Omega} \subset \mathbb{R}^d$ and time instant $t \in \bar{T}$ is denoted by $y(x, t)$. Mathematically, the system state is governed by the partial differential equation (PDE)

$$\frac{\partial y}{\partial t} = \mathcal{F}(x, t, y, \theta) \quad \text{in } \Omega \times T, \qquad (1)$$

where $\mathcal{F}$ is a well-posed, possibly nonlinear, differential operator which involves first- and second-order spatial derivatives and may include terms accounting for forcing inputs specified *a priori*. The PDE (1) is accompanied by the appropriate boundary and initial conditions

$$\mathcal{B}(x, t, y, \theta) = 0 \qquad \text{on } \Gamma \times T, \qquad (2)$$
$$y = y_0 \qquad \text{in } \Omega \times \{t = 0\}, \qquad (3)$$

respectively, $\mathcal{B}$ being an operator acting on the boundary Γ and $y_0 = y_0(x)$ a given function. Conditions (2) and (3) complement (1) such that the existence of a sufficiently smooth and unique solution is guaranteed. We assume that the forms of $\mathcal{L}$ and $\mathcal{B}$ are given explicitly up to an m-dimensional vector of unknown constant parameters θ which must be estimated using observations of the system. The implicit dependence of the state y on the parameter vector θ will be be reflected by the notation $y(x, t; \theta)$.

In what follows, we consider the discrete-continuous observations provided by n stationary pointwise sensors, namely

$$z_{\mathrm{m}}^\ell(t) = y(x^\ell, t; \theta) + \varepsilon(x^\ell, t), \quad t \in T, \qquad (4)$$

where $z_{\mathrm{m}}^\ell(t)$ is the scalar output and $x^\ell \in X$ stands for the location of the ℓ-th sensor ($\ell = 1, \ldots, n$), X signifies the part of the spatial domain Ω where the measurements can be made and $\varepsilon(x^\ell, t)$ denotes the measurement noise. This relatively simple conceptual framework involves no loss of generality since it can be easily generalized to incorporate, e.g., multiresponse systems or inaccessibility of state measurements, cf. (Uciński, 2005, p. 95).

It is customary to assume that the measurement noise is zero-mean, Gaussian, spatial uncorrelated and white (Quereshi, Ng and Goodwin, 1980; Omatu and Seinfeld, 1989; Amouroux and Babary, 1988), i.e.,

$$\mathrm{E}\{\varepsilon(x^\ell, t)\varepsilon(x^{\ell'}, t')\} = \sigma^2 \delta_{\ell\ell'}\delta(t - t'), \qquad (5)$$

where σ^2 defines the intensity of the noise, δ_{ij} and $\delta(\cdot)$ standing for the Kronecker and Dirac delta functions, respectively. Although white noise is a physically impossible process, it constitutes a reasonable approximation to a disturbance whose adjacent samples are uncorrelated at all time instants for which the time increment exceeds some value which is small compared with the time constants of the DPS. A rigorous formulation for a time-correlated setting (cf. Appendix C1 of (Uciński, 2005)) is well beyond the mathematical framework of this paper, but the attendant difficulties are mainly technical and do not substantially affect the basic results to be obtained. What is more, the white-noise assumption is consistent with most of the literature on the subject.

The most widely used formulation of the parameter estimation problem is as follows: Given the model (1)–(3) and the outcomes of the measurements $z_\mathrm{m}^\ell(\cdot)$, $\ell = 1, \ldots, n$, estimate θ by $\widehat{\theta}$, a global minimizer of the output least-squares error criterion

$$\mathcal{J}(\vartheta) = \sum_{\ell=1}^{n} \int_T \left\{ z_\mathrm{m}^\ell(t) - y(x^\ell, t; \vartheta) \right\}^2 \, \mathrm{d}t \qquad (6)$$

where $y(\cdot, \cdot; \vartheta)$ denotes the solution to (1)–(3) for a given value of the parameter vector ϑ.

The basic idea of fault detection is to compare the resulting parameter estimates with the corresponding known nominal values, treating possible differences as residuals which contain information about potential faults. Based on some thresholding techniques, the appropriate decision making system could be constructed to detect abnormal situations in system functioning (Patan and Patan, 2005).

Basically, only some parameters can be useful for the diagnosis. This accounts for partitioning the parameter vector into two two subsets. With no loss of generality, we may write

$$\theta = \begin{bmatrix} \theta_1 & \ldots & \theta_s & \theta_{s+1} & \ldots & \theta_m \end{bmatrix}^\mathsf{T} = \begin{bmatrix} \alpha^\mathsf{T} & \beta^\mathsf{T} \end{bmatrix}, \quad (7)$$

where α is a vector of s parameters which are essential for a proper fault detection and β is the vector of some unknown parameters which are a part of the model but are useless for fault detection. Based on the observations, it is possible to test the simple null hypothesis

$$H_0 : \alpha = \alpha^0, \qquad (8)$$

where α^0 is the nominal value for the vector α corresponding to the normal system performance.

For a fixed significance level (i.e. fixed probability of rejecting H_0 when it is true), the power of the likelihood ratio test for the alternative hypothesis of the form $H_A : \alpha \neq \alpha^*$ (i.e. $1 -$ the probability of accepting H_0 when H_A is true) can be

made large by maximizing the D_s-optimality criterion (see (Patan and Patan, 2005) for details)

$$\Psi_s[M] = \log \det[M_{\alpha\alpha} - M_{\alpha\beta} M_{\beta\beta}^{-1} M_{\alpha\beta}^\mathsf{T}], \qquad (9)$$

where $M \in \mathbb{R}^{m \times m}$ stands for the so-called *Fisher Information Matrix* (FIM) which is decomposed as

$$M = \begin{bmatrix} M_{\alpha\alpha} & M_{\alpha\beta} \\ M_{\alpha\beta}^\mathsf{T} & M_{\beta\beta} \end{bmatrix}, \qquad (10)$$

such that $M_{\alpha\alpha} \in \mathbb{R}^{s \times s}$, $M_{\alpha\beta} \in \mathbb{R}^{s \times (m-s)}$, $M_{\beta\beta} \in \mathbb{R}^{(m-s) \times (m-s)}$. The FIM is widely used in optimum experimental design theory for lumped systems (Fedorov and Hackl, 1997; Pázman, 1986; Pukelsheim, 1993; Walter and Pronzato, 1997; Atkinson and Donev, 1992). In our setting, the FIM is given by (Quereshi *et al*., 1980)

$$M(x^1, \ldots, x^n) = \sum_{\ell=1}^{n} \int_T g(x^\ell, t) g^\mathsf{T}(x^\ell, t) \, \mathrm{d}t, \qquad (11)$$

where

$$g(x, t) = \left[\frac{\partial y(x, t; \vartheta)}{\partial \vartheta_1}, \quad \ldots, \quad \frac{\partial y(x, t; \vartheta)}{\partial \vartheta_m} \right]^\mathsf{T}_{\vartheta = \theta^0} \qquad (12)$$

stands for the so-called *sensitivity vector*, θ^0 being the nominal value of the parameter vector θ (Uciński, 2005; Sun, 1994; Rafajłowicz, 1981; Rafajłowicz, 1983). Up to a constant scalar multiplier, the inverse of the FIM constitutes a good approximation of $\mathrm{cov}(\widehat{\theta})$ provided that 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 (Walter and Pronzato, 1997; Fedorov and Hackl, 1997).

Observe that for the partition

$$M^{-1} = \begin{bmatrix} D_{\alpha\alpha} & D_{\alpha\beta} \\ D_{\alpha\beta}^\mathsf{T} & D_{\beta\beta} \end{bmatrix}, \qquad (13)$$

where $D_{\alpha\alpha} \in \mathbb{R}^{s \times s}$, $D_{\alpha\beta} \in \mathbb{R}^{s \times (m-s)}$, $D_{\beta\beta} \in \mathbb{R}^{(m-s) \times (m-s)}$, we have (Bernstein, 2005, Fact 2.8.7, p.44)

$$D_{\alpha\alpha} = \left(M_{\alpha\alpha} - M_{\alpha\beta} M_{\beta\beta}^{-1} M_{\alpha\beta}^\mathsf{T} \right)^{-1} \qquad (14)$$

and further (Bernstein, 2005, Fact 2.15.8, p.73)

$$\det(D_{\alpha\alpha}) = \frac{\det(M_{\beta\beta})}{\det(M)}. \qquad (15)$$

Consequently, maximization of the D_s-optimality criterion amounts to minimization of $\det(D_{\alpha\alpha})$, which is proportional to the determinant of the covariance matrix for α.

The optimal sensor location problem considered in what follows is as follows: Given a set of N possible candidate sensor locations, where $N > n$, we wish to seek the best subset of n locations from among the N possible, so that the problem is then reduced to a combinatorial one. In other words, the problem is to divide the N available sites between n gauged sites and the remaining $N-n$ ungauged sites so as to maximize the determinant of the FIM associated with the parameters to be estimated. This formulation will be also adopted here.

Specifically, let x^i, $i = 1, \ldots, N$ denote the positions of sites where sensors can potentially be placed. Now that our design criterion has been established, the problem is to find an optimal allocation of n available sensors to x^i, $i = 1, \ldots, N$ so as to maximize the value of the design criterion incurred by the allocation. In order to formulate this mathematically, introduce for each possible location x^i a variable v_i which takes the value 1 or 0 depending on whether a sensor is or is not located at x^i, respectively. The FIM in (11) can then be rewritten as

$$M(v_1, \ldots, v_N) = \sum_{i=1}^{N} v_i M_i, \qquad (16)$$

where

$$M_i = \int_T g(x^i, t) g^{\mathsf{T}}(x^i, t) \, \mathrm{d}t. \qquad (17)$$

It is straightforward to verify that the $m \times m$ matrices M_i are nonnegative definite and, therefore, so is $M(v_1, \ldots, v_N)$.

Then our design problem takes the form:

Problem P:. Find the sequence $v = (v_1, \ldots, v_N)$ to maximize

$$\mathcal{P}(v) = \Psi\big(M(v)\big) \qquad (18)$$

subject to the constraints

$$\sum_{i=1}^{N} v_i = n, \qquad (19)$$

$$v_i = 0 \text{ or } 1, \quad i = 1, \ldots, N. \qquad (20)$$

This constitutes a 0–1 integer programming problem which necessitates an ingenious solution. In (Uciński and Patan, 2007) a general computational scheme is proposed to solve this problem based on the branch-and-bound method which is a standard technique for such class of tasks. Its presentation constitutes the next section of the paper.

3. Branch-and-bound solution

3.1. Outline. The branch-and-bound (BB) constitutes a general algorithmic technique for finding optimal solutions of various optimization problems, especially discrete or combinatorial (Floudas, 2001; Bertsekas, 1999). If applied carefully, it can lead to algorithms that run reasonably fast on average.

Principally, the BB method is a tree-search algorithm combined with a rule for pruning subtrees. Suppose we wish to maximize an objective function $\mathcal{P}(v)$ over a finite set V of admissible values of the argument v called the feasible region. The BB then progresses by iteratively applying two procedures: branching and bounding. *Branching* starts with smartly covering the feasible region by two or more smaller feasible subregions (ideally, partitioning into disjoint subregions). It is then repeated recursively to each of the subregions until no more division is possible, which leads to a progressively finer partition of V. The consecutively produced subregions naturally generate a tree structure called the BB tree. Its nodes correspond to the constructed subregions, with the feasible set V as the root node and the singleton solutions $\{v\}$, $v \in V$ as terminal nodes. In turn, the core of *bounding* is a fast method of finding upper and lower bounds to the maximum value of the objective function over a feasible subdomain. The idea is to use these bounds to economize computation by eliminating nodes of the BB tree that have no chance of containing an optimal solution. If the upper bound for a subregion V_A from the search tree is lower than the lower bound for any other (previously examined) subregion V_B, then V_A and all its descendant nodes may be safely discarded from the search. This step, termed *pruning*, is usually implemented by maintaining a global variable that records the maximum lower bound encountered among all subregions examined so far. Any node whose upper bound is lower than this value need not be considered further and thereby can be eliminated. It may happen that the lower bound for a node matches its upper bound. That value is then the maximum of the function within the corresponding subregion and the node is said to be solved. The search proceeds until all nodes have been solved or pruned, or until some specified threshold is met between the best solution found and the upper bounds on all unsolved problems.

3.2. Relaxed sensor selection problem. Let I denote the index set $\{1, \ldots, N\}$ of possible sensor locations. Consider a slight modification of Problem P, which starts by replacing the feasible set

$$V = \left\{ (v_1, \ldots, v_N) \,\Big|\, \sum_{i=1}^{N} v_i = n, \; v_i = 0 \text{ or } 1, \; \forall i \in I \right\}, \qquad (21)$$

by

$$V(I_0, I_1) = \big\{ v \in V \mid v_i = 0, \; \forall i \in I_0, \; v_i = 1, \; \forall i \in I_1 \big\}, \qquad (22)$$

where I_0 and I_1 are disjoint subsets of I. Consequently, $V(I_0, I_1)$ is the subset of V such that a sensor is placed at

the locations with indices in I_1, no sensor is placed at the locations with indices in I_0, and a sensor may or may not be placed at the remaining locations.

Then we introduce the following relaxed problem:

Problem $R(I_0, I_1)$:. Find a sequence $\bar{v}$ to maximize (18) subject to the constraints

$$\sum_{i=1}^{N} v_i = n, \tag{23}$$

$$v_i = 0, \quad i \in I_0, \tag{24}$$

$$v_i = 1, \quad i \in I_1, \tag{25}$$

$$0 \le v_i \le 1, \quad i \in I \setminus (I_0 \cup I_1). \tag{26}$$

In Problem $R(I_0, I_1)$ all 0–1 constraints on the variables v_i are relaxed by allowing them to take any value in the interval $[0, 1]$, except that the variables v_i, $i \in I_0 \cup I_1$ are fixed at either 0 or 1.

Each subset $V(I_0, I_1)$ can be identified with a node in the BB tree. The key assumption in the BB method is that for every nonterminal node $V(I_0, I_1)$, i.e., the node for which $I_0 \cup I_1 \ne I$, there is an algorithm that determines an upper bound $\bar{\mathcal{P}}(I_0, I_1)$ to the maximum design criterion over $V(I_0, I_1)$, i.e.,

$$\bar{\mathcal{P}}(I_0, I_1) \ge \max_{v \in V(I_0, I_1)} \mathcal{P}(v), \tag{27}$$

and a feasible solution $\underline{v} \in V$ for which $\mathcal{P}(\underline{v})$ can serve as a lower bound to the maximum design criterion over V. Furthermore, we may compute $\bar{\mathcal{P}}(I_0, I_1)$ by solving the relaxed Problem $R(I_0, I_1)$. A simple and efficient method for its solution based on the gradient projection technique constitutes the Section 4 of the paper. As a result of its application, we set $\bar{\mathcal{P}}(I_0, I_1) = \mathcal{P}(\bar{v})$.

As for $\underline{v}$, we can specify it as the best feasible solution (i.e., an element of V) found so far. If no solution has been found yet, we can either set the lower bound to $-\infty$, or use an initial guess about the optimal solution (experience provides evidence that the latter choice leads to much more rapid convergence).

3.3. Branching rule. The result of solving Problem $R(I_0, I_1)$ can serve as a basis to construct a branching rule for the binary BB tree. We adopt here the approach in which the node/subset $V(I_0, I_1)$ is expanded (i.e., partitioned) by first picking out all fractional values from among the values of the relaxed variables, and then rounding to 0 and 1 a value which is the most distant from both 0 and 1. Specifically, we apply the following steps:

(i) Determine

$$i_\star = \arg \min_{i \in I \setminus (I_0 \cup I_1)} |v_i - 0.5|. \tag{28}$$

(In case there are several minimizers, randomly pick one of them.)

(ii) Partition $V(I_0, I_1)$ into $V(I_0 \cup \{i_\star\}, I_1)$ and $V(I_0, I_1 \cup \{i_\star\})$ whereby two descendants of the node in question are defined.

A recursive application of the branching rule starts from the root of the BB tree, which corresponds to the trivial subset $V(\emptyset, \emptyset) = V$ and the fully relaxed problem. Each node of the BB tree corresponds to a continuous relaxed problem, $R(I_0, I_1)$, while each edge corresponds to fixing one relaxed variable at 0 or 1.

The above scheme has to be complemented with a search strategy to incrementally explore all the nodes of the BB tree. Here we use a common depth-first technique (Reinefeld, 2001; Russell and Norvig, 2003) which always expands the deepest node in the current fringe of the search tree. The reason behind this decision is that the search proceeds immediately to the deepest level of the search tree, where the nodes have no successors (Gerdts, 2005). In this way, lower bounds on the optimal solution can be found or improved as fast as possible.

It is simple matter to show that the matrix $M_{\beta\beta}$ is singular only if the whole FIM is singular. A very simple method to test whether or not the current relaxed problem will lead to a nonsingular FIM has been proposed by Uciński and Patan (2007, Prop. 2). Consequently, such a test of the singularity of the FIM can be built into the BB procedure in order to drop the corresponding node from further considerations and forego the examination of its descendants.

A recursive version of the resulting depth-first branch-and-bound is implemented in Algorithm 1. The operators involved in this implementation are as follows:

- SINGULARITY-TEST(I_0, I_1) returns true only if expansion of the current node will result in a singular FIM.

- RELAXED-SOLUTION(I_0, I_1) returns a solution to Problem $R(I_0, I_1)$.

- Ds-FIM(v) returns the D_s criterion value for the FIM corresponding to v.

- INTEGRAL-TEST(v) returns true only if the current solution v is integral.

- INDEX-BRANCH(v) returns the index defined by (28).

4. Algorithmic solution for the relaxed problem

4.1. Problem reformulation. For notational convenience, we replace the variables v_i, $i \in I \setminus (I_0 \cup I_1)$ by w_j, $j = 1, \ldots, q$, where $q = |I \setminus (I_0 \cup I_1)|$, since there exists a bijection π from $\{1, \ldots, q\}$ to $I \setminus (I_0 \cup I_1)$ such that $w_j = v_{\pi(j)}$, $j = 1, \ldots, q$. Consequently, we obtain the ultimate formulation:

Algorithm 1 A recursive version of the depth-first branch-and-bound method. It uses two global variables, *LOWER* and *v_best*, which are respectively the maximal value of the $\mathcal{P}(v)$ over feasible solutions found so far and the solution at which it is attained.

```
 1: procedure RECURSIVE-DFBB(I_0, I_1)
 2:    if |I_0| > N − n or |I_1| > n then
 3:        return                                  ▷ Constraint on the number of sensors would be violated
 4:    end if
 5:    if SINGULARITY-TEST(I_0, I_1) then
 6:        return                                              ▷ Only zero determinants can be expected
 7:    end if
 8:    v_relaxed ← RELAXED-SOLUTION(I_0, I_1)
 9:    Ds_relaxed ← DS-FIM(v_relaxed)                                                 ▷ Bounding
10:    if Ds_relaxed ≤ LOWER then
11:        return                                                                    ▷ Pruning
12:    else if INTEGRAL-TEST(v_relaxed) then
13:        v_best ← v_relaxed
14:        LOWER ← Ds_relaxed
15:        return                                                  ▷ Relaxed solution is integral
16:    else
17:        i_⋆ ← INDEX-BRANCH(v_relaxed)                       ▷ Partition into two descendants (Branching)
18:        RECURSIVE-DFBB(I_0 ∪ {i_⋆}, I_1)
19:        RECURSIVE-DFBB(I_0, I_1 ∪ {i_⋆})
20:    end if
21: end procedure
```

Problem R′(I_0, I_1):. Find $w \in \mathbb{R}^q$ to maximize

$$\mathcal{Q}(w) = \Psi\big(G(w)\big) \tag{29}$$

subject to the constraints

$$\sum_{j=1}^{q} w_j = r, \tag{30}$$

$$0 \leq w_j \leq 1, \quad j = 1, \dots, q, \tag{31}$$

where

$$r = n - |I_1|, \qquad G(w) = A + \sum_{j=1}^{q} w_j S_j, \tag{32}$$

$$A = \sum_{i \in I_1} M_i, \qquad S_j = M_{\pi(j)} \tag{33}$$

for $j = 1, \dots, q$. (Note that $|I_1|$ sensors have already been assigned to locations x^i, $i \in I_1$, and thus a decision about the placement of r remaining sensors has to be made.)

4.2. Gradient projection scheme. It can be demonstrated (the proof is technical and thus it can be found in the Appendix) that the D_s-optimality criterion is concave over the cone $\mathbb{S}_{++}^m$. What is more, its matrix derivative is

$$\mathring{\Psi}(M) = \frac{\partial \Psi(M)}{\partial M} = \left[\begin{array}{c|c} D_{\alpha\alpha} & D_{\alpha\beta} \\ \hline D_{\alpha\beta}^{\mathsf{T}} & D_{\beta\beta} - M_{\beta\beta}^{-1} \end{array} \right], \tag{34}$$

which results from the representation

$$\begin{aligned} \Psi[M] &= \log\det(M) - \log\det(M_{\beta\beta}) \\ &= \log\det(M) - \log\det(A^{\mathsf{T}} M A), \end{aligned} \tag{35}$$

where

$$A = \left[\begin{array}{c} 0 \\ \hline I \end{array} \right] \in \mathbb{R}^{m \times (m-s)}, \tag{36}$$

and Proposition 10.6.2 of (Bernstein, 2005, p.410).

In the sequel, W will stand for the set of all vectors $w = (w_1, \dots, w_q)$ satisfying (30) and (31). Note that it forms a polygon in $\mathbb{R}^q$. The objective function (29) is concave as the composition of the log-determinant with an affine mapping, see (Boyd and Vandenberghe, 2004, p. 79). We wish to maximize it over the polyhedral set W.

For that purpose, a number of possibilities exist, cf. (Bertsekas, 1999, Chapter 2), but gradient projection methods prove very efficient. A simplest version is a feasible direction method of the form

$$w^{k+1} = w^k + \lambda^k (\bar{w}^k - w^k), \tag{37}$$

where

$$\bar{w}^k = \left[w^k + \varrho^k \nabla \mathcal{Q}(w^k) \right]^+. \tag{38}$$

Here, $[\,\cdot\,]^+$ denotes projection on the set W, $\lambda^k \in (0, 1]$ is a stepsize, and ϱ^k is a positive scalar. Thus, to obtain the vector $\bar{w}^k$, we take a step $\varrho^k \nabla \mathcal{Q}(w^k)$ along the gradient, as in steepest ascent. We then project the result

$w^k + \varrho^k \nabla \mathcal{Q}(w^k)$ on W, thereby defining the feasible vector $\bar{w}^k$. Finally, we take a step along the feasible direction $\bar{w}^k - w^k$ using the stepsize λ^k.

An $\mathcal{O}(r)$ algorithm of projecting on W, which constitutes an intersection of a hyperplane and a box, was set forth by Maculan, Santiago, Macambira and Jardim (2003). Let also note that

$$\nabla \mathcal{Q}(w) = \Big[\operatorname{trace}\big(\mathring{\Psi}(G(w))S_1\big) \ \ldots \ \operatorname{trace}\big(\mathring{\Psi}(G(w))S_q\big)\Big]^{\mathsf{T}} \tag{39}$$

5. Computational results

As an illustration of the presented approach to the sensor network design, we consider the problem of sensor placement for fault detection in the transport process of an air pollutant over a given urban area. Within this domain, which has been normalized to the unit square $\Omega = (0,1)^2$, an active source of pollution is present, which influence the pollutant spatial concentration $y = y(x,t)$. The evolution of y over the normalized observation interval $T = (0,1]$ is described by the following advection-diffusion equation:

$$\frac{\partial y(x,t)}{\partial t} + \nabla \cdot \big(v(x)y(x,t)\big) = \\ \nabla \cdot \big(a(x)\nabla y(x,t)\big) + f(x), \ x \in \Omega \tag{40}$$

subject to the boundary and initial conditions:

$$\frac{\partial y(x,t)}{\partial n} = 0, \quad \text{on } \Gamma \times T, \tag{41}$$

$$y(x,0) = y_0, \quad \text{in } \Omega, \tag{42}$$

where term $f(x) = 50 \exp\big(-50\|x - c\|^2\big)$ represents a source of pollutant located at point $c = (0.3, 0.3)$, and $\partial y/\partial n$ stands for the partial derivative of y with respect to the outward normal to the boundary Γ. The mean spatio-temporal changes of the wind velocity field over the area were approximated by $v = (v_1, v_2)$, where

$$v_1 = 2(x_1 + x_2 - t), \quad v_2 = x_2 - x_1 + t. \tag{43}$$

what is also illustrated in Fig. 1. The assumed functional form of the spatial-varying diffusion coefficient $a(x)$ is

$$a(x) = \theta_1 + \theta_2 x_1 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2. \tag{44}$$

The subject of interest here is a detection of significant increase in the intensity of the pollutant emission from the source. As the symptom of this abrupt fault an excessive deviation of the parameters θ_1 and θ_2 from their nominal values was assumed. Therefore, these parameters need estimation based on measurement data from monitoring stations.

In our simulation studies, the described bound-and-branch technique was applied to determine the locations of stationary sensors. Given N prospective sites in $\Omega \cup \Gamma$, we aim at selecting their subset consisting of the locations at which the measurements made by n available sensors would lead to D_s-optimum least-squares estimates of the parameters θ.

In order to determine the elements of sensitivity vector required to calculate FIM the direct-differentiation method (Uciński, 2005) was applied assuming the nominal values of the parameters $\theta_1^0 = 0.02$, $\theta_2^0 = 0.01$ and $\theta_3^0 = \theta_4^0 = 0.005$. We solved the resulting system of PDEs using routines of the MATLAB PDE toolbox for a spatial mesh composed of 682 triangles and 378 nodes. As for the numerical integration required to evaluate information matrices for admissible observation sites the trapezoidal rule was applied with the time step equal to 0.04, based on the sensitivity vector interpolated at the nodes representing admissible locations x^i, cf. Appendix I in (Uciński, 2005) for details.

A complex dynamics of the pollution process is shown in Fig. 1. The pollutant spreads out over the entire domain reflecting the sophisticated combination of diffusion and advection and follows the temporary direction of the wind being the dominant transport factor. In considered scenario, the observation grid was assumed to be created at locations selected from among those elements of mentioned above 378-point triangulation mesh which do not lie on the outer boundary (there were 312 such nodes, which are indicated with dots in Fig. 2).

A Matlab program was written to implement the recursive version of the DFBB procedure embodied by Algorithm 1. In order to solve the relaxed problem, which constitutes the principal part of DFBB, the gradient projection scheme was implemented (Sec. 4) with tolerance and maximum number of iterations set to $\epsilon = 10^{-8}$ and $\kappa_{\max} = 1000$, respectively. Finally, to take full advantage of the efficient gradient projection scheme performed at each node of the BB tree, Algorithm 1 was extended to incorporate a rounding procedure after solving the relaxed problem. Thus, this solution provides not only an upper bound to the currently processed branch, but may also lead to a great improvement in the lower bound LOWER. Intuitively, a proper way of rounding a relaxed solution is to choose target sites for locating spare sensors so that they correspond to the largest weights. In the case that this choice is complicated by the presence of sites with identical weights, the target sites are chosen randomly. In effect, it is not necessary to descend to the bottom level of the BB tree in order to update LOWER. This fact is of crucial importance for large-scale problems, where such an action costs many recursive calls of Algorithm 1, thus the rounding of relaxed solutions significantly speedup the algorithm, as evidenced by numerous simulation experiments.

D_s-optimal sensor configurations for different numbers of allocated sensors are shown in Fig. 2. It is clear

Table 1. Comparison of algorithm performance for different settings.

No. of sensors	No. of recursive calls	Time [s]	Criterion value
10	17	4.697	1.19422e+006
20	41	9.043	4.28668e+006
40	5	3.856	1.59385e+007
60	1	1.092	3.33651e+007
80	1	0.511	5.8198 e+007
100	3	1.698	8.24685e+007
120	1	0.261	1.14538e+008
150	7	3.395	1.59447e+008

that the complexity of the system dynamics makes the proper prediction of the observation locations rather difficult and nonintuitive. The sensors tend to form the pattern reflecting the areas of greatest changes in the pollutant concentration but the observations are averaged over time and it is not trivial to follow the dynamics of the observation strategy. Surprisingly, the measurements in the closest vicinity of the pollution source turned out to be not very attractive for the considered fault detection. The results concerning the algorithm performance are summarized in Tab. 1. The number of RDFBB calls equals to one means that the optimal solution is obtained just by rounding the fully relaxed problem. Examination of the data from Tab. 1 leads to interesting conclusions. Unexpectedly, with an increased number of sensors (and the size of the corresponding search space), the pruning process becomes more efficient. This effect can be explained by observing that a higher density of sensors leads to a better estimate of the lower bound to the optimal value of the design criterion, which results in an increased efficiency of pruning and whereupon the search is speeded up.

6. Conclusions

We consider the problem of monitoring network design which provides the proper diagnostic information about the state of considered distributed parameter system. One of the most important issues related to this task is the choice of a suitable criterion which quantify the quality of the detection of abnormal system state indicating the potential faults. We state the problem in the form of maximization of the power for a parametric hypothesis test which verify the nominal values for the system parameters. Then, as an appropriate performance index D_s-optimality criterion defined on the Fisher information matrix is proposed. Although this criterion is well known in the optimum design theory, there exist only few attempts to exploit it in the context of fault detection for DPSs, cf. (Uciński, 2003; Patan, 2004; Patan and Patan, 2005; Patan, Uciński and Baranowski, 2005). This work contributes to this issue providing the proper mathematical justification of some valuable properties of D_s-optimality such as its concavity and differentiability.

Another crucial difficulty here is the large scale of the resulting global optimization problem, since the monitoring networks encountered in process industry or environmental engineering may often consist of several hundreds of stations. Obviously, this makes the exhaustive search on a candidate-by-candidate basis practically intractable and creates a need for techniques which would implement a guided search and have acceptable performance. With our sensor network design problem we started from the most common formulation, in which the measurement system has a finite number of sensor candidate positions and the aim is to select the best subset of points (of desired cardinality) in the sense of maximizing the D_s-optimality criterion. This fits into the framework of nonlinear 0–1 integer programming. The solution of this combinatorial design problem using the branch-and-bound method constitutes a quite natural option, but the main problem when trying to implement it has been the lack of a low-cost procedure to obtain upper bounds to the optimal values of the D_s-optimality criterion. Our main contribution consists in adapting a specialized gradient projection procedure for criterion maximization to produce such bounds. Consequently, the proposed method can be implemented with great ease and our experience provides evidence that, with this tool, even large-scale design problems can be solved using an off-the-shelf PC.

Naturally, there still remain some open problems which need a close attention. The following points can be raised as the main directions of further research:

- Extension of the considered class of DPSs. The great advantage of the delineated approach is that it is independent of a specific form of PDEs used as a mathematical model of considered DPS. In such a manner the presented approach can be rather easily adopted for the class of multi-output DPSs or/and systems with delays, since only the formula for calculating the elements of FIM will be properly changed.

- The structure of the proposed BB algorithm is well suited for parallel implementation dedicated to the clusters of PCs. Such parallel version of the approach would be very attractive for solving large-scale problems.

- The proposed simple branching rule for the binary BB tree can be refined by incorporating a mechanism driving the search proces towards the most promising branches in terms of the objective function or heuristics exploiting specific properties of the considered DPS.

A. Appendix

Proof of the concavity of the D_s-optimality criterion
Reall that the Euclidean matrix scalar product

$$\langle G, H \rangle = \text{trace}(G^\mathsf{T} H) \tag{A1}$$

turns $\mathbb{R}^{m \times m}$ into a Euclidean space of dimension m^2. We denote by $\mathbb{S}^m$ its subspace of symmetric matrices. It is well known that this subspace contains the convex, pointed and closed cone of nonnegative definite matrices, $\mathbb{S}_+^m$, which has the set of positive definite matrices, $\mathbb{S}_{++}^m$, as its interior relative to the space $\mathbb{S}^m$, cf. (Pukelsheim, 1993, p. 10).

The Loewner partial ordering $\succeq$ of $\mathbb{S}^m$ is defined as follows:

$$G \succeq H \Leftrightarrow G - H \succeq 0 \Leftrightarrow G - H \in \mathbb{S}_+^m. \tag{A2}$$

Recall that

$$G \in \mathbb{S}_+^m \Leftrightarrow \text{trace}(GH) \succeq 0, \quad \forall H \in \mathbb{S}_+^m. \tag{A3}$$

Lemma 1. *If $G \succeq H \succeq 0$, then for all $V \in \mathbb{S}^m$ we have*

$$\text{trace}(GVGV) \geq \text{trace}(HVHV). \tag{A4}$$

Proof. From (A3) it follows that

$$\text{trace}(GVGV) \geq \text{trace}(GVHV) \geq \text{trace}(HVHV). \tag{A5}$$

The two inequalities result from $VGV \succeq VHV$ and $G \succeq H$, respectively. ∎

Lemma 2. *For the partition (10), if $M \in \mathbb{S}_{++}^m$, then*

$$M^{-1} \succeq A(A^\mathsf{T} M A)^{-1} A^\mathsf{T}, \tag{A6}$$

where A is defined in (36).

Proof. First observe that, given $M \in \mathbb{S}_{++}^m$, we have $M_{\beta\beta} = A^\mathsf{T} M A \in \mathbb{S}_{++}^m$ whenever $s < m$.

From Proposition 2.8.7 of (Bernstein, 2005, p.44) we deduce that

$$
M^{-1} - A(A^\mathsf{T} M A)^{-1} A^\mathsf{T}
$$
$$
= \begin{bmatrix} D_{\alpha\alpha} & -D_{\alpha\alpha} M_{\alpha\beta} M_{\beta\beta}^{-1} \\ -M_{\beta\beta}^{-1} M_{\alpha\beta}^\mathsf{T} D_{\alpha\alpha} & M_{\beta\beta}^{-1} M_{\alpha\beta}^\mathsf{T} D_{\alpha\alpha} M_{\alpha\beta} M_{\beta\beta}^{-1} \end{bmatrix}. \tag{A7}
$$

Then for any $y \in \mathbb{R}^m$ partitioned as

$$y = \begin{bmatrix} y_\alpha \\ y_\beta \end{bmatrix} \tag{A8}$$

we obtain

$$
y^\mathsf{T} \big(M^{-1} - A(A^\mathsf{T} M A)^{-1} A^\mathsf{T} \big) y
$$
$$
= \big(y_\alpha - M_{\alpha\beta} M_{\beta\beta}^{-1} y_\beta \big)^\mathsf{T} D_{\alpha\alpha} \big(y_\alpha - M_{\alpha\beta} M_{\beta\beta}^{-1} y_\beta \big)
$$
$$
\geq 0, \tag{A9}
$$

which yields the desired result. ∎

We claim that the criterion (9) is concave over $\mathbb{S}_{++}^m$. Indeed, denoting by $\Psi_s'(M)$ the first Fréchet derivative of Ψ_s at $M \in \mathbb{S}_{++}^m$, from (35) and Proposition 10.6.2(vi) in (Bernstein, 2005, p.410) we obtain

$$
\Psi_s'(M)V
$$
$$
= \text{trace} \left\{ \frac{\partial}{\partial M} \big[\Psi_s(M) \big] V \right\} \tag{A10}
$$
$$
= \text{trace} \left\{ \big[M^{-1} - A(A^\mathsf{T} M A)^{-1} A^\mathsf{T} \big] V \right\}
$$

for any $V \in \mathbb{S}^m$.

As for the second Fréchet derivative of Ψ_s at $M \in \mathbb{S}_{++}^m$, denoted by $\Psi_s''(M)$, its value $\Psi_s''(M)(U, V)$ at given $U, V \in \mathbb{S}^m$ equals the derivative of the mapping $M \mapsto \Psi_s'(M)V$ applied to U, i.e.,

$$
\Psi_s''(M)(U, V)
$$
$$
= \text{trace} \left\{ \frac{\partial}{\partial M} \big[\Psi_s'(M)V \big] U \right\}
$$
$$
= -\text{trace} \left\{ \big[M^{-1} V M^{-1} \right.
$$
$$
\left. - A(A^\mathsf{T} M A)^{-1} A^\mathsf{T} V A (A^\mathsf{T} M A)^{-1} A^\mathsf{T} \big] U \right\}, \tag{A11}
$$

the last equality resulting from Proposition 10.6.4(iii) in (Bernstein, 2005, p.411).

From Lemmas 1 and 2 we then get

$$
\Psi_s''(M)(V, V)
$$
$$
= -\text{trace} \big\{ M^{-1} V M^{-1} V \big\}
$$
$$
+ \text{trace} \big\{ A(A^\mathsf{T} M A)^{-1} A^\mathsf{T} V A (A^\mathsf{T} M A)^{-1} A^\mathsf{T} V \big\}
$$
$$
\leq 0, \quad \forall V \in \mathbb{S}^m, \tag{A12}
$$

which establishes the concavity of Ψ_s.

Acknowledgment

References

Amouroux, M. and Babary, J. P. (1988). Sensor and control location problems, *in* M. G. Singh (Ed.), *Systems & Control Encyclopedia. Theory, Technology, Applications*, Vol. 6, Pergamon Press, Oxford, pp. 4238–4245.

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

Bernstein, D. S. (2005). *Matrix Mathematics. Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton University Press, Princeton, NJ.

Bertsekas, D. P. (1999). *Nonlinear Programming*, Optimization and Computation Series, 2nd edn, Athena Scientific, Belmont, MA.

Boer, E. P. J., Hendrix, E. M. T. and Rasch, D. A. M. K. (2001). Optimization of monitoring networks for estimation of the semivariance function, *in* A. C. Atkinson, P. Hackl and W. Müller (Eds), mODa 6*, Proc. 6th Int. Workshop on Model-Oriented Data Analysis,* Puchberg/Schneeberg, Austria, 2001, Physica-Verlag, Heidelberg, pp. 21–28.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*, Cambridge University Press, Cambridge.

Cassandras, C. G. and Li, W. (2005). Sensor networks and cooperative control, *European Journal of Control* **11**(4–5): 436–463.

Chiang, L. H., Russell, E. L. and Braatz, R. D. (2001). *Fault Detection and Diagnosis in Industrial Systems*, Springer-Verlag, London.

Fedorov, V. V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*, Lecture Notes in Statistics, Springer-Verlag, New York.

Floudas, C. A. (2001). Mixed integer nonlinear programming, MINLP, *in* C. A. Floudas and P. M. Pardalos (Eds), *Encyclopedia of Optimization*, Vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 401–414.

Gerdts, M. (2005). Solving mixed-integer optimal control problems by branch&bound: A case study from automobile test-driving with gear shift, *Journal of Optimization Theory and Applications* **26**: 1–18.

Isermann, R. (1997). *Supervision, Fault Detection and Diagnosis of Technical Systems*, Special Section Control Engineering Practice, Vol. 5, No. 5.

Korbicz, J., Kościelny, J., Kowalczuk, Z. and Cholewa, W. (2004). *Fault Diagnosis. Models, Artificial Intelligence, Applications*, Springer-Verlag, Berlin Heidelberg.

Kubrusly, C. S. and Malebranche, H. (1985). Sensors and controllers location in distributed systems — A survey, *Automatica* **21**(2): 117–128.

Maculan, N., Santiago, C. P., Macambira, E. M. and Jardim, M. H. C. (2003). An $O(n)$ algorithm for projecting a vector on the intersection of a hyperplane and a box in $\mathbb{R}^n$, *Journal of Optimization Theory and Applications* **117**(3): 553–574.

Omatu, S. and Seinfeld, J. H. (1989). *Distributed Parameter Systems: Theory and Applications*, Oxford Mathematical Monographs, Oxford University Press, New York.

Patan, M. (2004). *Optimal Observation Strategies for Parameter Estimation of Distributed Systems*, Zielona Góra: University Press. May be viewed at `http://www.zbc.zgora.pl`.

Patan, M. and Patan, K. (2005). Optimal observation strategies for model-based fault detection in distributed systems, *International Journal of Control* **78**(18): 1497–1510.

Patan, M., Uciński, D. and Baranowski, P. (2005). Optimal observation strategies for fault detection in distributed-parameter systems, *Pomiary Automatyka Kontrola* (9): 71–73.

Patton, R. J., Frank, P. M. and Clark, R. (2000). *Issues of Fault Diagnosis for Dynamic Systems*, Springer-Verlag, Berlin.

Patton, R. J. and Korbicz, J. (Eds) (1999). *Advances in Computational Intelligence*, Special Issue International Journal of Applied Mathematics and Computer Science, Vol. 9, No. 3.

Pázman, A. (1986). *Foundations of Optimum Experimental Design*, Mathematics and Its Applications, D. Reidel Publishing Company, Dordrecht, The Netherlands.

Pukelsheim, F. (1993). *Optimal Design of Experiments*, Probability and Mathematical Statistics, John Wiley & Sons, New York.

Quereshi, Z. H., Ng, T. S. and Goodwin, G. C. (1980). Optimum experimental design for identification of distributed parameter systems, *International Journal of Control* **31**(1): 21–29.

Rafajłowicz, E. (1981). Design of experiments for eigenvalue identification in distributed-parameter systems, *International Journal of Control* **34**(6): 1079–1094.

Rafajłowicz, E. (1983). Optimal experiment design for identification of linear distributed-parameter systems: Frequency domain approach, *IEEE Transactions on Automatic Control* **28**(7): 806–808.

Reinefeld, A. (2001). Heuristic search, *in* C. A. Floudas and P. M. Pardalos (Eds), *Encyclopedia of Optimization*, Vol. 2, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 409–411.

Russell, S. J. and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*, 2nd edn, Pearson Education International, Upper Saddle River, NJ.

Sun, N.-Z. (1994). *Inverse Problems in Groundwater Modeling*, Theory and Applications of Transport in Porous Media, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Uciński, D. (1992). Optimal sensor location for parameter identification of distributed systems, *Applied Mathematics and Computer Science* **2**(1): 119–134.

Uciński, D. (1999). *Measurement Optimization for Parameter Estimation in Distributed Systems*, Technical University Press, Zielona Góra. Available in electronic form at `http://www.issi.uz.zgora.pl/~ucinski/`.

Uciński, D. (2000). Optimal selection of measurement locations for parameter estimation in distributed processes, *International Journal of Applied Mathematics and Computer Science* **10**(2): 357–379.

Uciński, D. (2003). On optimum experimental design for technical diagnostics of processes, *in* Z. Kowalczuk (Ed.), *6-th National Conference on Diagnostics of Industrial Processes, 15-17 September, WŞadysŞawowo, Poland, 2003*, Technical University Press, Gdańsk, pp. 207–212. (In Polish).

Uciński, D. (2005). *Optimal Measurement Methods for Distributed-Parameter System Identification*, CRC Press, Boca Raton, FL.

Uciński, D. and Patan, M. (2007). D-optimal design of a monitoring network for parameter estimation of distributed systems, *Journal of Global Optimization* **39**: 291–322.

van de Wal, M. and de Jager, B. (2001). A review of methods for input/output selection, *Automatica* **37**: 487–510.

Walter, É. and Pronzato, L. (1997). *Identification of Parametric Models from Experimental Data*, Communications and Control Engineering, Springer-Verlag, Berlin.
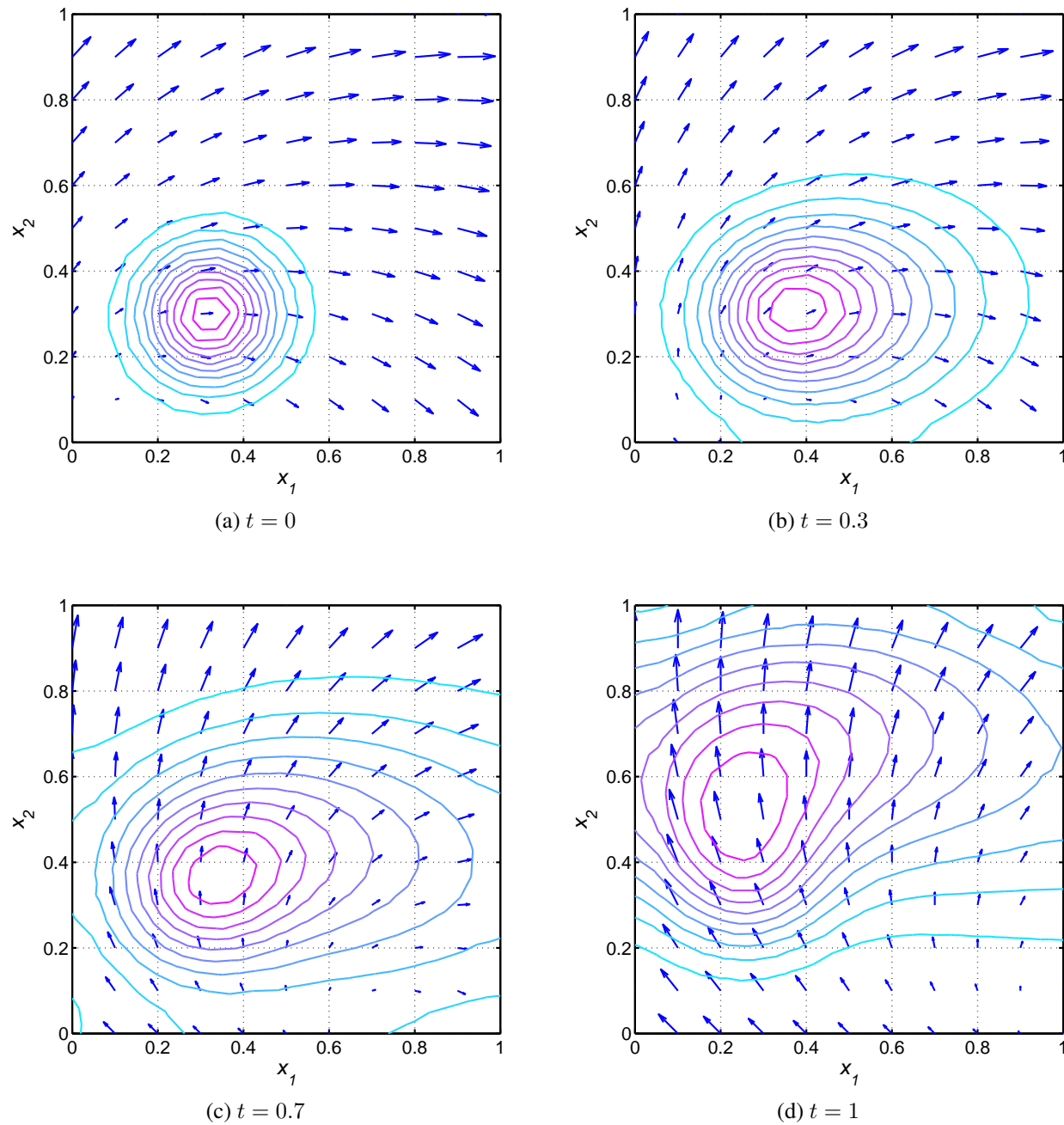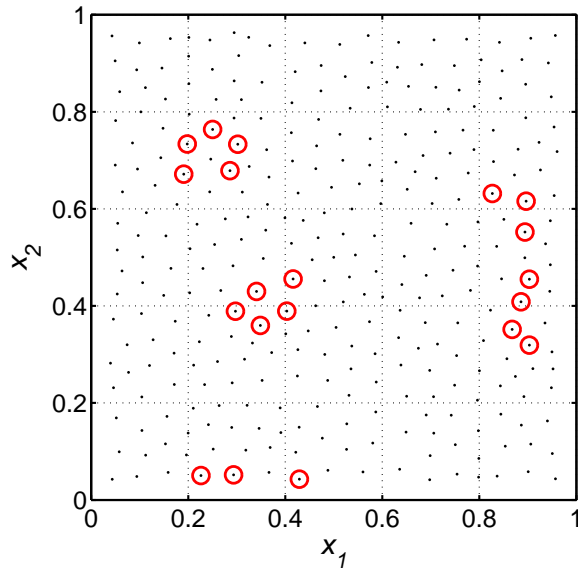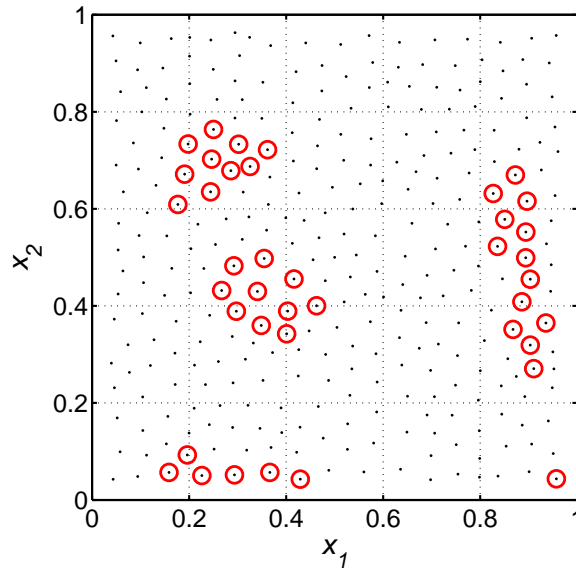
(a) $t = 0$

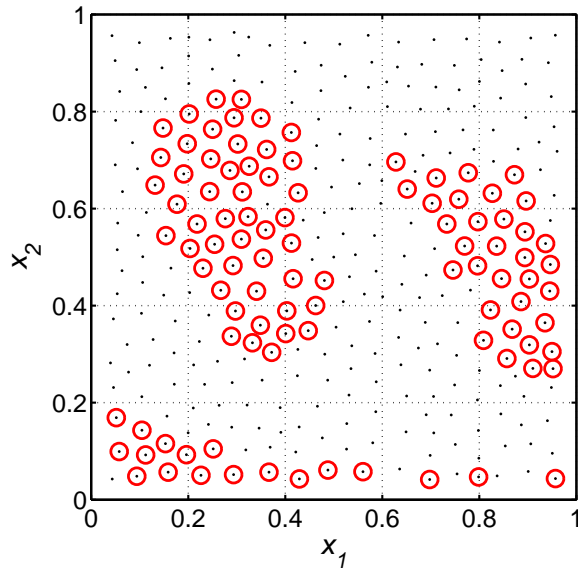(b) $t = 0.3$

(c) $t = 0.7$

(d) $t = 1$

Fig. 1. Temporal changes in the wind velocity field and pollutant concentration.
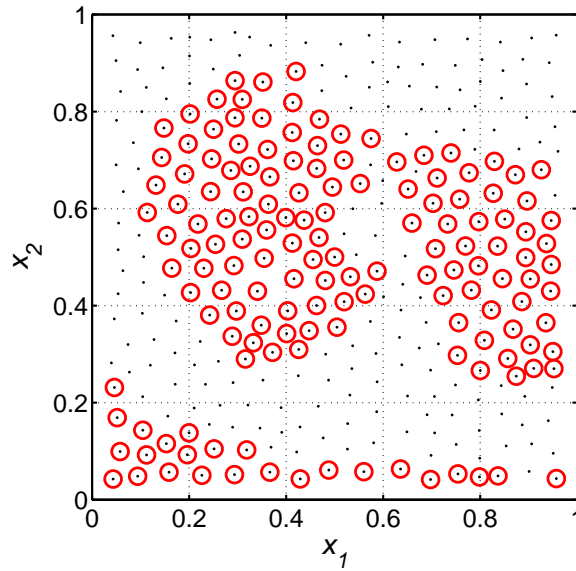
(a) $n = 20$

(b) $n = 40$

(c) $n = 80$

(d) $n = 150$

Fig. 2. The D_s-optimal allocation of different numbers of sensors.