Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Sensor Selection with Nonsmooth Design Criteria Based on Semi-Infinite Programming

Dariusz Uciński

Institute of Control and Computation Engineering
University of Zielona Góra, ul. Szafrana 2, 65–516 Zielona Góra,
Poland (e-mail: d.ucinski@issi.uz.zgora.pl)

Abstract: A problem of optimal node activation in large-scale sensor networks is considered. The resulting measurements are supposed to be used to estimate unknown parameters of a spatiotemporal process described by a partial differential equation. In this setting, the sensor subset selection problem may quickly become computationally intractable when an excessively complex sensor location algorithm is employed. The is even more pronounced when the design criterion is nondifferentiable. A vital example of this criterion is the sum of an arbitrary number of smallest eigenvalues of the Fisher information matrix, being a generalization of the well-known E-optimality criterion. A simple branch-and-bound algorithm is exposed here to maximize this criterion. Its key component to produce upper bounds to the maximum of the objective function implements a relaxation procedure for solving semi-infinite programming problems. It alternates between solving a linear programming subproblem and evaluation of the eigenvalues and eigenvectors of the current information matrix, which makes it extremely easy to implement. The paper is complemented with a numerical example of computing actual sensor locations.

Keywords: distributed parameter systems, parameter estimation, optimal experiment design, sensor systems, global optimization.

1. INTRODUCTION

The importance of the sensor location problem for distributed parameter systems (DPSs) has been recognized in many application domains, cf. Kubrusly and Malebranche (1985), Uciński (2005). One of them is parameter estimation whose accuracy conditions the quality of the underlying mathematical model, which is most often a partial differential equation (PDE).

Technical limitations make it desirable to determine best sensor positions prior to the actual data collection. This makes statistical experimental design, see Atkinson et al. (2007) or Pronzato and Pàzman (2013), the most appropriate framework to formulate the problem mathematically. Therefore, various scalar performance measures defined on the Fisher information matrix (FIM) associated with the estimated parameters are traditionally used to quantify the estimation accuracy. The approach has been followed in the bulk of the literature since the seminal works by Rafajłowicz (1981). For overviews of this active research area till the 2010s, see the monographs by Uciński (2005) and Patan (2012).

In the last decade considerable progress has been made via inclusion of the unknown initial state as an additional unknown parameter. This makes the resulting parameter space infinite dimensional and the overall inverse problem ill-posed, which is reflected by extreme unstability of the estimates even for very low noise in the data. A panacea is the Bayesian framework which takes account of prior statistical information about the unknown parameters and/or states, see, e.g., works by Alexanderian et al. (2014), Haber et al. (2010), Gejadze and Shutyaev (2012). At present, however, a considerable inconvenience caused by this approach are still the attendant large-scale computations.

Most methods of sensor location carry out a search of a predefined finite (but possibly very large) set of candidate locations. This can also equivalently be interpreted as the sensor selection problem in which the observation system comprises a large number of candidate sensors whose positions are already specified and only a subset from among them are supposed to be activated, while keeping the other sensors dormant. This setting corresponds to the operation of sensor networks, see Tricaud and Chen (2012). With no loss of generality, this setting will be adopted in what follows.

The main impediment to progress in sensor selection is its combinatorial nature. For moderate-size problems, a branch-and-bound technique was proposed by Uciński and Patan (2007) to reduce the search space. On the other hand, for large-scale sensor networks continuous relaxations were advocated. Specifically, Patan and Uciński (2016) looked for an optimal density of active sensors instead of the individual positions of active sensors. But in most approaches, however, the 0–1 constraints on the decision variables indicating whether or not the sensors are active are relaxed by allowing them to additionally take any fractional values in the interval $[0, 1]$. In this way, a convex optimization problem is obtained, which paves

the way for application of a plethora of computationally efficient algorithms. These include primarily interior point methods as described by Joshi and Boyd (2009), which are usually applied to problems cast as semidefinite programming ones, cf. Chepuri and Leus (2015), or polyhedral approximation methods, see Uciński and Patan (2007) and Herzog et al. (2018). In addition, efficient algorithms of constrained optimum experimental design can be almost directly applied here, see Harman and Benková (2017). When combined with some sparsity-enhancing schemes and rounding, cf. Chepuri and Leus (2015) and Patan and Uciński (2017), they can be used to produce approximations to the solutions of the original NP-hard discrete sensor selection problem.

The above-mentioned methods are primarily focused on smooth optimality criteria. An important criterion which does not fall into this class is the E-optimality criterion (the minimum eigenvalue of the FIM). Maximizing it with respect to the sensor locations, we minimize the length of the largest axis of the uncertainty ellipsoid for the estimates. This criterion turns out to be nondifferentiable when the minimum eigenvalue of the FIM is multiple. It is easy to show, see Joshi and Boyd (2009) or Uciński (2005, p. 61), that the corresponding relaxed problem can be reformulated in terms of an LMI-constrained convex minimization problem, but when the number of candidate sensors grows, solving SDP problems becomes computationally demanding and the interior-point solvers quickly run into time and memory issues on mediocre computers.

Instead, Pronzato and Pàzman (2013, p. 326) made use of the formulation in terms of a semi-infinite programming problem and applied the relaxation algorithm set forth by Shimizu and Aiyoshi (1980), which resulted in an extremely simple computational scheme alternating between computing the eigenvalue of a matrix and solving a linear programming problem. The approach has then extended by Burclová and Pázman (2016) for the case of the E_k-optimality criterion (the sum of k smallest eigenvalues of the FIM). These results remain, however, quite obscure in the engineering literature.

This paper is aimed at adopting the algorithm of Burclová and Pázman (2016) for E_k-optimality to the sensor selection problem. To this end, this algorithm is combined with the branch-and-bound scheme, whose structure was set forth by Uciński and Patan (2007), to compute E_k-optimal locations of active sensors. Thus, applicability of the methodology of Burclová and Pázman (2016) has substantially been extended. The resulting extremely efficient scheme based on a sequence of linear programming problems serves here as tool to produce upper bounds to the maximum of the objective function in traversing the branch-and-bound tree.

2. E_k-OPTIMAL SENSOR ACTIVATION

Consider a bounded spatial domain $\Omega \subset \mathbb{R}^d$ with a sufficiently smooth boundary Γ, and a bounded time interval $T = (0, t_f]$. Over the spatiotemporal domain $\Omega \times T$ a DPS is defined, which is decribed by a PDE. Its scalar state at a spatial point $\boldsymbol{x} \in \bar{\Omega} \subset \mathbb{R}^d$ and a time instant $t \in \bar{T}$ is denoted by $y(\boldsymbol{x}, t; \boldsymbol{\theta})$ (the bar over a set denotes its closure). Here $\boldsymbol{\theta} \in \mathbb{R}^m$ stands for a vector of unknown parameters in the PDE which must be estimated using observations of the state.

Let $\boldsymbol{x}^i$, $i = 1, \ldots, N$ be given spatial positions of candidate sensor network nodes. It is assumed here that, over the time interval T, the state y is observed continuously in time by $n < N$ active sensors selected from among N candidate points ($N - n$ nonselected sensors will remain dormant). Introducing for each location $\boldsymbol{x}^i$ a variable v_i which takes the value 1 or 0 depending on whether or not a sensor located at $\boldsymbol{x}^i$ is active over T, respectively, we can represent the observations obtained from network nodes as follows:

$$z_i(t) = v_i \big[y(\boldsymbol{x}^i, t; \boldsymbol{\theta}) + \varepsilon(\boldsymbol{x}^i, t) \big], \quad i = 1, \ldots, N \tag{1}$$

for $t \in T$, where $\varepsilon(\boldsymbol{x}^i, t)$ denotes the measurement noise. It is customary to assume that the noise is zero-mean, Gaussian and white.

The accuracy of the resulting least-squares estimates of $\boldsymbol{\theta}$ depends on the selection of gauged sites. A quantitative measure Φ of the 'goodness' of particular sensor configurations is customarily based on the concept of the *Fisher Information Matrix* (FIM) which is widely used in optimum experimental design theory for lumped systems, cf., Atkinson et al. (2007) or Pronzato and Pàzman (2013). In our setting, the FIM is given by

$$\boldsymbol{M}(\boldsymbol{v}) = \sum_{i=1}^{N} v_i \boldsymbol{M}_i, \tag{2}$$

where $\boldsymbol{v} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N)$,

$$\boldsymbol{M}_i = \int_T \boldsymbol{g}(\boldsymbol{x}^i, t) \boldsymbol{g}(\boldsymbol{x}^i, t)^\top \, \mathrm{d}t, \quad i = 1, \ldots, N. \tag{3}$$

Here

$$\boldsymbol{g}(\boldsymbol{x}, t) = \left[\frac{\partial y(\boldsymbol{x}, t; \boldsymbol{\vartheta})}{\partial \vartheta_1}, \ \ldots, \ \frac{\partial y(\boldsymbol{x}, t; \boldsymbol{\vartheta})}{\partial \vartheta_m} \right]^\top_{\boldsymbol{\vartheta} = \boldsymbol{\theta}^0} \tag{4}$$

stands for the *sensitivity vector* (see (Uciński, 2005, Ch. 2.6) for its computation for a given PDE), $\boldsymbol{\theta}^0$ being a prior estimate to the unknown parameter vector $\boldsymbol{\theta}$, cf. Uciński (2005). It is easy to check that the $m \times m$ matrices $\boldsymbol{M}_i$ are nonnegative definite and, therefore, so is $\boldsymbol{M}(\boldsymbol{v})$.

As for a specific form of Φ, we are focused on the E_k-optimality criterion of Burclová and Pázman (2016),

$$\Phi(M) = \sum_{\ell=1}^{k} \lambda_\ell(\boldsymbol{M}), \tag{5}$$

where $\lambda_1(\boldsymbol{M}) \leq \cdots \leq \lambda_m(\boldsymbol{M})$ denote the eigenvalues of $\boldsymbol{M}$ in increasing order. For $k = 1$ (5) simplifies to the well-known E-optimality criterion. Note that (5) is isotonic, i.e., it preserves Loewner's ordering (this means that $\boldsymbol{M}_1 \succeq \boldsymbol{M}_2$ implies $\Phi(\boldsymbol{M}_1) \geq \Phi(\boldsymbol{M}_2)$), and concave over the cone of nonnegative definite matrices (Marshall et al., 2011, Fact F.3, p. 688).

Our design problem is thus formulated as follows:

Problem P: Find a sequence $\boldsymbol{v} = (v_1, \ldots, v_N) \in \mathbb{R}^N$ to maximize

$$\mathcal{P}(\boldsymbol{v}) = \Phi\big(\boldsymbol{M}(\boldsymbol{v})\big) \tag{6}$$

subject to the constraints

$$\sum_{i=1}^{N} v_i = n, \quad v_i \in \big\{0, 1\big\}, \quad i = 1, \ldots, N. \tag{7}$$

3. BRANCH-AND-BOUND SCHEME

3.1 General Idea

The branch-and-bound (BB) is a general technique for finding optimal solutions of combinatorial optimization problems. Set $\mathcal{I} = \{1, \ldots, N\}$. We then partition the feasible set

$$V = \left\{ \boldsymbol{v} \in [0,1]^N \;\Big|\; \sum_{i=1}^{N} v_i = n, \quad v_i = 0 \text{ or } 1, \ \forall i \in \mathcal{I} \right\}, \tag{8}$$

into subsets

$$V(\mathcal{I}_0, \mathcal{I}_1) = \{ \boldsymbol{v} \in V \mid v_i = 0, \ \forall i \in \mathcal{I}_0, \ v_i = 1, \ \forall i \in \mathcal{I}_1 \}, \tag{9}$$

where $\mathcal{I}_0$ and $\mathcal{I}_1$ are disjoint subsets of $\mathcal{I}$. Consequently, $V(\mathcal{I}_0, \mathcal{I}_1)$ is the subset of V such that sensors are active at the locations with indices in $\mathcal{I}_1$, sensors are dormant at the locations with indices in $\mathcal{I}_0$, and sensors may be active or dormant at the remaining locations.

Each subset $V(\mathcal{I}_0, \mathcal{I}_1)$ is identified with a node in the BB tree. The key assumption in the BB method is that for every nonterminal node $V(\mathcal{I}_0, \mathcal{I}_1)$, i.e., the node for which $\mathcal{I}_0 \cup \mathcal{I}_1 \neq \mathcal{I}$, there is an algorithm that determines an upper bound $\overline{\mathcal{P}}(\mathcal{I}_0, \mathcal{I}_1)$ to the maximum value of design criterion over $V(\mathcal{I}_0, \mathcal{I}_1)$, i.e.,

$$\overline{\mathcal{P}}(\mathcal{I}_0, \mathcal{I}_1) \geq \max_{\boldsymbol{v} \in V(\mathcal{I}_0, \mathcal{I}_1)} \mathcal{P}(\boldsymbol{v}), \tag{10}$$

and a feasible solution $\underline{\boldsymbol{v}} \in V$ for which $\mathcal{P}(\underline{\boldsymbol{v}})$ can serve as a lower bound to the maximum design criterion over V. We may compute $\overline{\mathcal{P}}(\mathcal{I}_0, \mathcal{I}_1)$ by solving the following relaxed problem:

Problem $R(\mathcal{I}_0, \mathcal{I}_1)$: Find a sequence $\overline{\boldsymbol{v}}$ to maximize (6) subject to the constraints

$$\sum_{i=1}^{N} v_i = n, \tag{11}$$

$$v_i = 0, \quad i \in \mathcal{I}_0, \tag{12}$$

$$v_i = 1, \quad i \in \mathcal{I}_1, \tag{13}$$

$$0 \leq v_i \leq 1, \quad i \in \mathcal{I} \setminus (\mathcal{I}_0 \cup \mathcal{I}_1). \tag{14}$$

In Problem $\mathrm{R}(\mathcal{I}_0, \mathcal{I}_1)$ 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, $i \in \mathcal{I}_0 \cup \mathcal{I}_1$ are fixed at either 0 or 1. A simple and efficient method for its solution is given in Section 4. As a result of its application, we set $\overline{\mathcal{P}}(\mathcal{I}_0, \mathcal{I}_1) = \mathcal{P}(\overline{\boldsymbol{v}})$.

As for $\underline{\boldsymbol{v}}$, 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 upper 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).

3.2 Branching Rule and BB Algorithm

The result of solving Problem $\mathrm{R}(\mathcal{I}_0, \mathcal{I}_1)$ can serve as a basis to construct a branching rule for the binary BB tree. We adopt here the approach in which the tree node/subset $V(\mathcal{I}_0, \mathcal{I}_1)$ 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

$$i_\star = \arg \min_{i \in \mathcal{I} \setminus (\mathcal{I}_0 \cup \mathcal{I}_1)} |v_i - 0.5|. \tag{15}$$

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

(ii) Partition $V(\mathcal{I}_0, \mathcal{I}_1)$ into sets $V(\mathcal{I}_0 \cup \{i_\star\}, \mathcal{I}_1)$ and $V(\mathcal{I}_0, \mathcal{I}_1 \cup \{i_\star\})$ 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, $\mathrm{R}(\mathcal{I}_0, \mathcal{I}_1)$, while each edge corresponds to fixing one relaxed variable at 0 or 1.

The above scheme is complemented with the depth-first search strategy to incrementally explore all the nodes of the BB tree implemented as Algorithm 1. The operators involved in this implementation are as follows:

- SINGULARITY-TEST$(\mathcal{I}_0, \mathcal{I}_1)$ returns true only if expansion of the current node will result in a singular FIM, see Uciński and Patan (2007) for details.
- RELAXED-SOLUTION$(\mathcal{I}_0, \mathcal{I}_1)$ returns a solution to Problem $\mathrm{R}(\mathcal{I}_0, \mathcal{I}_1)$.
- PHI-FIM$(\boldsymbol{v})$ evaluates the design criterion Φ at the FIM corresponding to $\boldsymbol{v}$.
- INTEGRAL-TEST$(\boldsymbol{v})$ returns true only if the current solution $\boldsymbol{v}$ is integral.
- INDEX-BRANCH$(\boldsymbol{v})$ returns the index defined by (15).

4. ALGORITHM FOR THE RELAXED PROBLEM

Let $r = n - |\mathcal{I}_1|$, $q = I - |\mathcal{I}_1 \cup \mathcal{I}_0|$ (for a set A the notation $|A|$ means its cardinality). Consider any bijection π from $\{1, \ldots, q\}$ to $\mathcal{I} \setminus (\mathcal{I}_1 \cup \mathcal{I}_0)$ such that $w_j = v_{\pi(j)}$, $j = 1, \ldots, q$.

Accordingly, we obtain the following formulation:

Problem $R'(\mathcal{I}_0, \mathcal{I}_1)$: Find a sequence $\boldsymbol{w} = (w_1, \ldots, w_q) \in \mathbb{R}^q$ to maximize

$$\mathcal{Q}(\boldsymbol{w}) = \Phi\big(\boldsymbol{G}(\boldsymbol{w})\big) \tag{16}$$

subject to the constraints

$$\sum_{j=1}^{q} w_j = r, \qquad 0 \leq w_j \leq 1, \quad j = 1, \ldots, q, \tag{17}$$

where

$$\boldsymbol{G}(\boldsymbol{w}) = \boldsymbol{A} + \sum_{j=1}^{q} w_j \boldsymbol{S}_j, \qquad \boldsymbol{A} = \sum_{i \in \mathcal{I}_1} \boldsymbol{M}_i, \tag{18}$$

$$\boldsymbol{S}_j = \boldsymbol{M}_{\pi(j)}, \quad j = 1, \ldots, q. \tag{19}$$

(Note that $|\mathcal{I}_1|$ sensors have already been activated at locations $\boldsymbol{x}^i$, $i \in \mathcal{I}_1$, and thus a decision about the activation of r remaining sensors has to be made.)

In the sequel, W will stand for the set of all vectors $\boldsymbol{w}$ satisfying (17). Note that it forms a polyhedral set in $\mathbb{R}^q$.

Using the *minimum principle* (see Appendix A), we can express (16) as

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 design criterion over feasible solutions found so far and the solution at which it is attained.

procedure RECURSIVE-DFBB(E_0, E_1)
  if $\mathcal{I}_0 \cup \mathcal{I}_1 = \mathcal{I}$ **then** ▷ Deepest level of the BB tree has been attained
    phi_v ← PHI-FIM($v(E_0, E_1)$)
    if *phi_v* > *LOWER* **then**
      v_best ← $v(\mathcal{I}_0, \mathcal{I}_1)$
      LOWER ← *phi_v*
    end if
    return
  end if
  if SINGULARITY-TEST($\mathcal{I}_0, \mathcal{I}_1$) **then**
    return ▷ Only singular FIMs can be expected
  end if
  v_relaxed ← RELAXED-SOLUTION($\mathcal{I}_0, \mathcal{I}_1$)
  phi_relaxed ← PHI-FIM(*v_relaxed*)
  if *phi_relaxed* ≤ *LOWER* **then**
    return ▷ Pruning
  else if INTEGRAL-TEST(*v_relaxed*) **then**
    v_best ← *v_relaxed*
    LOWER ← *phi_relaxed*
    return ▷ Relaxed solution is integral
  else
    $i_\star$ ← INDEX-BRANCH(*v_relaxed*) ▷ Partition into two descendants
    RECURSIVE-DFBB($\mathcal{I}_0 \cup \{i_\star\}, \mathcal{I}_1$)
    RECURSIVE-DFBB($\mathcal{I}_0, \mathcal{I}_1 \cup \{i_\star\}$)
  end if
end procedure

$$\begin{aligned}\mathcal{Q}(\boldsymbol{w}) &= \min_{\boldsymbol{S}\in\mathbb{S}} \operatorname{trace}(\boldsymbol{S}^\top \boldsymbol{G}(\boldsymbol{w})\boldsymbol{S}) \\ &= \min_{\boldsymbol{S}\in\mathbb{S}} \sum_{j=1}^{q} w_j \operatorname{trace}(\boldsymbol{S}^\top \boldsymbol{M}_i \boldsymbol{S})\end{aligned} \tag{20}$$

where $\mathbb{S} = \{\boldsymbol{S} \in \mathbb{R}^{m\times k} \mid \boldsymbol{S}^\top\boldsymbol{S} = \boldsymbol{I}_k\}$. In this way, maximization of $\mathcal{Q}(\boldsymbol{w})$ becomes a maximin optimization problem, as we seek $\boldsymbol{w}^\star = \arg\max_{\boldsymbol{w}\in W} \min_{\boldsymbol{S}\in\mathbb{S}} \operatorname{trace}(\boldsymbol{S}^\top\boldsymbol{B}\boldsymbol{S})$. Specifically, it can be formulated as the following semi-infinite programming (SIP) one: Find $\boldsymbol{w} \in W$ and an auxiliary scalar α so as to maximize α subject to the infinite set of constraints

$$\left\{\sum_{j=1}^{q} w_j \operatorname{trace}(\boldsymbol{S}^\top \boldsymbol{M}_i \boldsymbol{S}) \geq \alpha, \quad \boldsymbol{S} \in \mathbb{S}\right\}. \tag{21}$$

Various numerical approaches to solving SIP problems are characterized, e.g., in (Hettich and Kortanek, 1993). In practice, however, the simple relaxation algorithm proposed by Shimizu and Aiyoshi (1980), as was also suggested by Burclová and Pázman (2016), turns out to perform very well. It consists in relaxing the problem by taking account only of a finite number of constraints (21). The tailored version of the relaxation procedure is presented as Algorithm 2. Note its striking simplicity, as it alternates between computing the eigenvalues and eigenvectors of a matrix and solving a linear-programming problem. Thus it can be implemented with great ease. What is more, it is naturally applicable to large-scale problems.

Algorithm 2 Algorithm solving the SIP problem.

Step 0: (Initialization)
Set

$$\boldsymbol{w}^{(0)} = (\underbrace{r/q, \ldots, r/q}_{q \text{ times}}).$$

and $\mathbb{S}^{(0)} = \emptyset$. Select $0 < \epsilon \ll 1$, a parameter used in the stopping rule, and set $\tau = 0$.

Step 1: (Solution of the eigenproblem)
Determine

$$\begin{aligned}\boldsymbol{S}^{(\tau)} &= \arg\min_{\boldsymbol{S}\in\mathbb{S}} \operatorname{trace}(\boldsymbol{S}^\top \boldsymbol{G}(\boldsymbol{w}^{(\tau)})\boldsymbol{S}) \\ &= \left[\boldsymbol{q}_1(\boldsymbol{G}(\boldsymbol{w}^{(\tau)})) \ \ldots \ \boldsymbol{q}_k(\boldsymbol{G}(\boldsymbol{w}^{(\tau)}))\right],\end{aligned} \tag{22}$$

where $\boldsymbol{q}_1(\boldsymbol{G}(\boldsymbol{w}^{(\tau)})), \ldots, \boldsymbol{q}_k(\boldsymbol{G}(\boldsymbol{w}^{(\tau)}))$ are orthonormal eigenvectors of $\boldsymbol{G}(\boldsymbol{w}^{(\tau)})$ corresponding to the eigenvalues $\lambda_1(\boldsymbol{G}(\boldsymbol{w}^{(\tau)})), \ldots, \lambda_k(\boldsymbol{G}(\boldsymbol{w}^{(\tau)}))$.
Set $\mathbb{S}^{(\tau+1)} = \mathbb{S}^{(\tau)} \cup \{\boldsymbol{S}^{(\tau)}\}$.

Step 2: (Solution of the linear programming problem)
Find $(\boldsymbol{w}^{(\tau+1)}, \alpha^{(\tau+1)})$ to maximize $\alpha^{(\tau+1)}$ subject to the constraints

$$\sum_{j=1}^{q} w_j^{(\tau+1)} \operatorname{trace}(\boldsymbol{S}^\top \boldsymbol{M}_i \boldsymbol{S}) \geq \alpha^{(\tau+1)}, \quad \boldsymbol{S} \in \mathbb{S}^{(\tau+1)}. \tag{23}$$

Step 3: (Termination check)
If

$$\Phi(\boldsymbol{G}(\boldsymbol{w}^{(\tau)})) \geq \alpha^{(\tau+1)}(1-\epsilon) \tag{24}$$

then $\boldsymbol{w}^{(\tau)}$ is an approximation to the sought maximin solution. Otherwise, increment τ and go to Step 1.

5. NUMERICAL EXAMPLE

Consider the heat equation modelling the temperature evolution in an anisotropic rectangular square copper plate

$$\begin{aligned}\frac{\partial y(\boldsymbol{x},t)}{\partial t} &= \frac{\partial}{\partial x_1}\left(\mu(\boldsymbol{x};\boldsymbol{\theta})\frac{\partial y(\boldsymbol{x},t)}{\partial x_1}\right) \\ &\quad + \frac{\partial}{\partial x_2}\left(\mu(\boldsymbol{x};\boldsymbol{\theta})\frac{\partial y(\boldsymbol{x},t)}{\partial x_2}\right) + u(\boldsymbol{x},t), \\ &\quad \boldsymbol{x} \in \Omega = (0,1)\times(0,1), \quad t \in (0,1)\end{aligned} \tag{25}$$

subject to the homogeneous initial and boundary conditions. The diffusion coefficient to be identified has the form $\mu(\boldsymbol{x};\boldsymbol{\theta}) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1^2 + \theta_5 x_1 x_2 + \theta_6 x_2^2$. The assumed nominal value $\boldsymbol{\theta}^0$ necessary to determine the sensitivity coefficients is

$$\boldsymbol{\theta}^0 = (0.2, -0.05, 0.2, -0.1, 0.05, -0.2). \tag{26}$$

It defines the quadratic surface displayed in Fig. 1.

As regards the forcing term in our model, it has the form

$$u(\boldsymbol{x},t) = 20\exp\bigl(-50(x_1 - t)^2\bigr). \tag{27}$$

It models the action of a line heat source which is oriented parallely to the x_2-axis and moves with a constant speed from the left to the right boundary of square domain Ω. The corresponding evolution of the state variable $y(\boldsymbol{x},t)$ is shown in Fig. 2.

A uniform mesh of $N = 31 \times 31 = 961$ candidate points was assumed as the set of sites where sensor network nodes are placed (they are marked with dots in Fig. 3). Our purpose is activate best $n = 100$ sensors from among them, as quantified by the E_k-optimality criterion, in order to

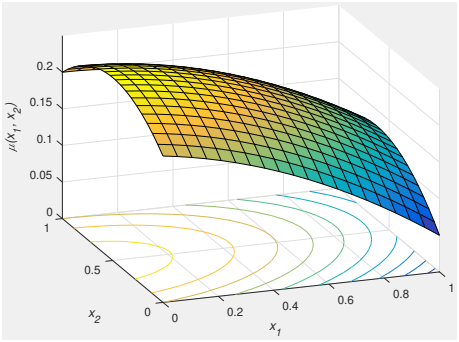


Fig. 1. Diffusion coefficient.

estimate μ (i.e., the parameters θ_1–θ_6) as accurately as possible.

Basically, the number of eigenvalues strongly influences the results. When the number of eigenvalues included in the criterion is low, the active sensors attempt to cluster in regions where the value of $\mu(\boldsymbol{x})$ is low, as the sensitivity to changes in the value of this coefficient is the highest there. An increase in the number k of the eigenvalues included in the criterion results in a decrease in the number of sensor clusters. If all the eigenvalues are included, which corresponds to maximizing the trace of the FIM, only one group of sensors is formed.

The size of the solutions space is $\binom{961}{100} > 9 \times 10^{137}$. The optimal solutions displayed in Fig. 3 were obtained in no more than 30 seconds on a decent laptop (Intel(R) Core(TM) i7-6700HQ 2.60GHz, 24 GB RAM) running Windows 10 and MATLAB 2016b. The state and sensitivity equations were solved using the Partial Differential Equation Toolbox. The parameter required by Algorithm 2 was set as $\epsilon = 10^{-5}$. This fine performance was possible owing to a good initial lower bound to the optimal solution. It was obtained by solving a fully relaxed problem and activating sensors which corresponded to the largest relaxed weights. Otherwise, the solutions obtained by literally following Algorithm 1 consumed about five minutes of the CPU time, which is not bad, either.

6. CONCLUDING REMARKS

An alternative approach to select a best n-element subset of active sensors from among a given N-element set of all candidate sensors could be to employ a greedy exchange algorithm, see, e.g., (Uciński, 2005, p. 105). This strategy is very popular in practice. Such algorithms begin with an n-point starting sensor configuration which then sequentially evolves through addition of new elements selected from among dormant sensors and deletion of sites at which active sensors have provisionally been planned, in an effort to maximally increase the value of the E_k-optimality design criterion. It goes without saying that such an approach would outperform the BB technique proposed here as far as the running time is concerned. Note, however, that greedy exchange algorithms are only capable of finding globally competitive solutions (i.e., nearly optimal ones), with an explicit trade-off between global optimality and speed. The approach presented here is superior in the sense that it always produces global maxima and, what is more, does it within tolerable time.

REFERENCES

Alexanderian, A., Petra, N., Stadler, G., and Ghattas, O. (2014). A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification. *SIAM Journal on Scientific Computing*, 36(5), A2122–A2148.

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

Burclová, K. and Pázman, A. (2016). Optimal design of experiments via linear programming. *Statistical Papers*, 57(4), 893–910.

Chepuri, S.P. and Leus, G. (2015). Sparsity-promoting sensor selection for non-linear measurement models. *IEEE Transactions on Signal Processing*, 63(3), 684–698.

Gejadze, I.Y. and Shutyaev, V. (2012). On computation of the design function gradient for the sensor-location problem in variational data assimilation. *SIAM Journal on Scientific Computing*, 34(2), B127–B147. doi: 10.1137/110825121.

Haber, E., Horesh, L., and Tenorio, L. (2010). Numerical methods for the design of large-scale nonlinear discrete ill-posed inverse problems. *Inverse Problems*, 26(2), 025002.

Harman, R. and Benková, E. (2017). Barycentric algorithm for computing D-optimal size- and cost-constrained designs of experiments. *Metrika*, 80(2), 201–225.

Harville, D.A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, New York.

Herzog, R., Riedel, I., and Uciński, D. (2018). Optimal sensor placement for joint parameter and state estimation problems in large-scale dynamical systems with applications to thermo-mechanics. *Optimization and Engineering*, 19(3), 591–627.

Hettich, R. and Kortanek, K.O. (1993). Semi-infinite programming: Theory, methods and applications. *SIAM Review*, 35(3), 380–429.

Joshi, S. and Boyd, S. (2009). Sensor selection via convex optimization. *IEEE Transactions on Signal Processing*, 57(2), 451–462.

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

Marshall, A.W., Olkin, I., and Arnold, B.C. (2011). *Inequalities: Theory of Majorization and Its Applications*. Springer-Verlag, New York, 2nd edition.

Patan, M. (2012). *Optimal Sensor Networks Scheduling in Identification of Distributed Parameter Systems*. Springer-Verlag, Berlin.

Patan, M. and Uciński, D. (2016). Cost-constrained D-optimum node activation for large-scale monitoring networks. In *2016 American Control Conference (ACC)*, 1643–1648. doi:10.1109/ACC.2016.7525152.

Patan, M. and Uciński, D. (2017). A sparsity-enforcing method for optimal node activation in parameter estimation of spatiotemporal processes. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 3089–3094. doi:10.1109/CDC.2017.8264110.

Pronzato, L. and Pàzman, A. (2013). *Design of Experiments in Nonlinear Models. Asymptotic Normality, Optimality Criteria amd Small-Sample Properties.*

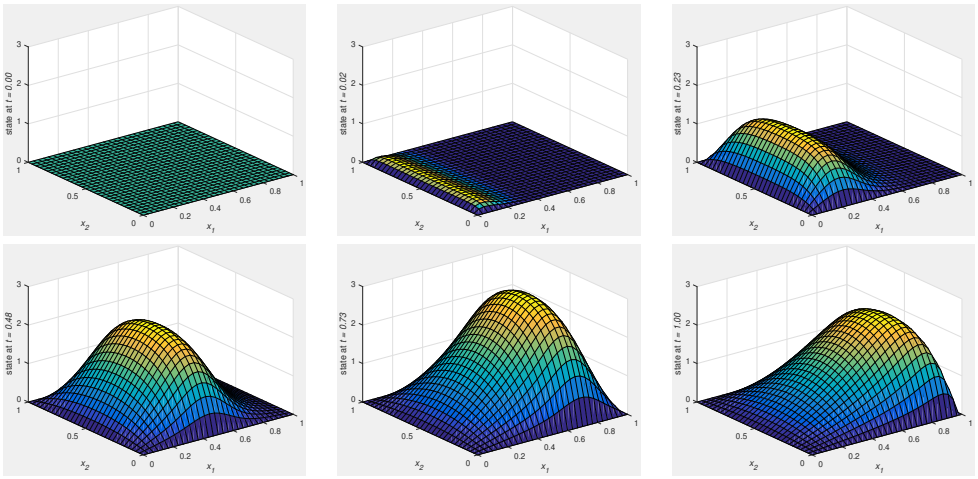
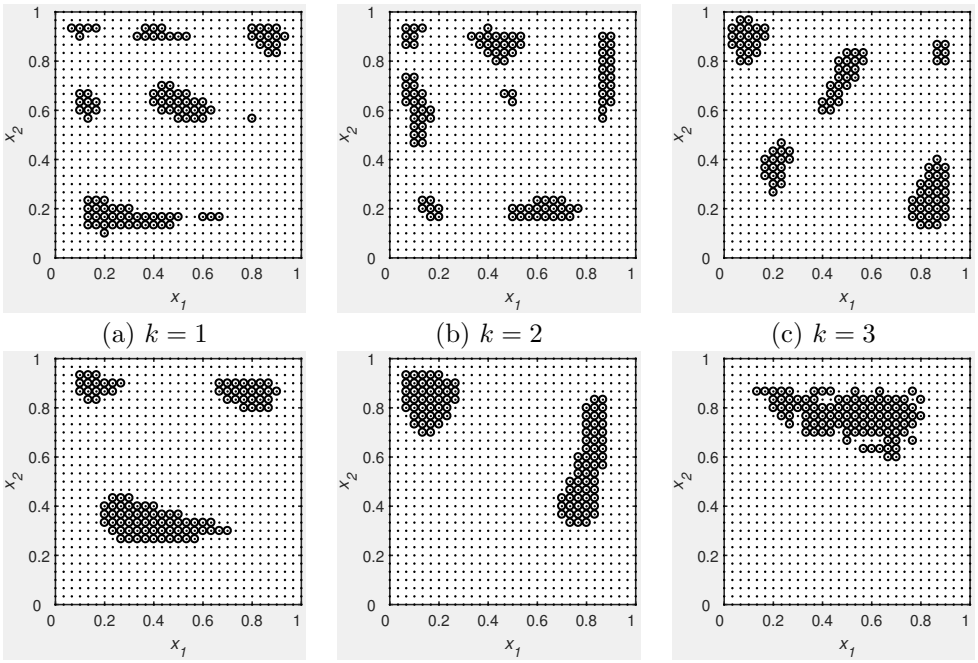


Fig. 2. Temperature at consecutive time moments.



(a) $k = 1$ (b) $k = 2$ (c) $k = 3$

(d) $k = 4$ (e) $k = 5$ (f) $k = 6$

Fig. 3. E_k-optimal locations of activated sensors.

Springer-Verlag, New York.

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

Shimizu, K. and Aiyoshi, E. (1980). Necessary conditions for min-max problems and algorithms by a relaxation procedure. *IEEE Transactions on Automatic Control*, AC-25(1), 62–66.

Tricaud, C. and Chen, Y. (2012). *Optimal Mobile Sensing and Actuation Policies in Cyber-physical Systems*. Springer-Verlag, London.

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.

Appendix A. MINIMUM PRINCIPLE

The following result can be found, e.g., in (Harville, 1997, Theorem 21.12.4,p. 556):

Theorem 1. Let $\boldsymbol{B} \in \mathbb{R}^{m\times m}$ be symmetric with (not necessarily distinct) eigenvalues $\lambda_1 \le \cdots \le \lambda_m$. Then for all $k = 1, \ldots, m$ we have

$$\sum_{\ell=1}^{k} \lambda_\ell = \min\bigl\{\mathrm{trace}(\boldsymbol{S}^\top \boldsymbol{B}\boldsymbol{S}) \mid \boldsymbol{S} \in \mathbb{R}^{m\times k} \text{ and } \boldsymbol{S}^\top\boldsymbol{S} = \boldsymbol{I}_k\bigr\}, \tag{A.1}$$

where $\boldsymbol{I}_k$ is the $k \times k$ identity matrix. The minimum on the right-hand side is attained if the columns of $\boldsymbol{S}$ are orthonormal eigenvectors of $\boldsymbol{B}$ corresponding to $\lambda_1, \ldots, \lambda_k$, respectively.