

SENSOR NETWORK SCHEDULING FOR IDENTIFICATION OF SPATIALLY DISTRIBUTED PROCESSES

DARIUSZ UCINŚKI

Institute of Control and Computation Engineering
University of Zielona Góra, ul. Podgórna 50, 65-246 Zielona Góra, Poland
e-mail: d.ucincki@issi.uz.zgora.pl

The work treats the problem of fault detection for processes described by partial differential equations as that of maximizing the power of a parametric hypothesis test which checks whether or not system parameters have nominal values. A simple node activation strategy is discussed for the design of a sensor network deployed in a spatial domain that is supposed to be used while detecting changes in the underlying parameters which govern the process evolution. The setting considered relates to a 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. The problem is then formulated as the determination of the density of gauged sites so as to maximize the adopted design criterion, subject to inequality constraints incorporating a maximum allowable sensor density in a given spatial domain. The search for the optimal solution is performed using a simplicial decomposition algorithm. The use of the proposed approach is illustrated by a numerical example involving sensor selection for a two-dimensional diffusion process.

Keywords: sensor network, parameter estimation, distributed parameter system, optimum experimental design, Fisher information matrix.

1. Introduction

1.1. Distributed parameter systems and sensor networks. Strong demands made in modern process control are frequently associated with using very accurate models in which spatial dynamics have to be included in addition to the temporal one. The processes in question are often termed Distributed Parameter Systems (DPSs) and they are described by Partial Differential Equations (PDEs). One of the fundamental questions in DPSs is parameter estimation, which refers to model calibration, i.e., the determination from observed data of unknown parameters in the system model such that the predicted response of the model is close, in some well-defined sense, to the process observations made by some suitable collection of sensors forming the observation system. A major difficulty here is related to the impossibility to measure process variables over the entire spatial domain. Moreover, the measurements are inexact by virtue of inherent errors of measurement associated with transducing elements and also because of the measurement environment.

It goes without saying that the inability to take distributed measurements of process states leads to the ques-

tion of where to locate sensors so that the information content of the resulting signals with respect to the distributed state and PDE model is as high as possible. This is an appealing problem since in most applications these locations are not prespecified and therefore provide design parameters. The location of sensors is not necessarily dictated by physical considerations or by intuition and, therefore, some systematic approaches should still be developed in order to reduce the cost of instrumentation and to increase the efficiency of identifiers.

An example which is particularly stimulating in the light of the results reported here constitutes the 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, whose number can be quite large. Then designers must

address the question of how to optimize sensor locations in order to obtain the most accurate model.

In recent years, a revived interest in optimal sensor location has been observed owing to advances in modern observation systems. Sensor and networking technologies now enable large-scale deployment of superior data acquisition systems with adjustable resolutions, called *sensor networks* (see Zhao and Guibas, 2004; Hirsch *et al.*, 2008; Jain and Agrawal, 2005; Sastry and Iyengar, 2005; Chong and Kumar, 2003; Sinopoli *et al.*, 2003; Cassandras and Li, 2005; Bauer, 2008). Each sensor node has a sensing capability, as well as limited energy supply, computing power, memory and communication ability. These inexpensive, low-power communication devices can be placed throughout the physical space, providing dense sensing close to physical phenomena, processing and communicating this information, and coordinating actions with other nodes. Sensor networks have recently come into prominence because they hold the potential to revolutionize a wide spectrum of both civilian and military applications, including environmental monitoring, scene reconstruction, motion tracking, motion detection, battlefield surveillance, remote sensing, global awareness, etc. On the one hand, this offers new possibilities for observation systems. On the other, however, completely new design problems have to be solved. In this paper, we wish to demonstrate how the existing sensor placement techniques can be adapted to solve the node activation problem for a large-scale sensor network which is supposed to detect abnormal changes in the observed DPS.

1.2. Optimal sensor location for parameter identification. The significance of sensor planning has already been recognized in many application domains, e.g., air quality monitoring systems, groundwater-resources management, recovery of valuable minerals and hydrocarbon, model calibration in meteorology and oceanography, chemical engineering, hazardous environments and smart materials (Nehorai *et al.*, 1995; Porat and Nehorai, 1996; Jeremić and Nehorai, 1998; Jeremić and Nehorai, 2000; Navon, 1997; Daescu and Navon, 2004; Christofides, 2001; Banks *et al.*, 1996; Sun, 1994; Uciński, 2005). The operation and control of such systems usually require precise information on the parameters which condition the accuracy of the underlying mathematical model, but that information is only available through a limited number of sensors. Over the past years, this has stimulated laborious research on the development of strategies for efficient sensor placement (for reviews, see the papers by Kubrusly and Malebranche (1985), van de Wal and de Jager (2001) and El Jai and Hamzaoui (2009) or comprehensive monographs by Uciński (2005; 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 candi-

dates 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.

Exceptions to this naive approach constitute the works originating in statistical optimum experimental design (Fedorov and Hackl, 1997; Pázman, 1986; Pukelsheim, 1993; Walter and Pronzato, 1997; Atkinson and Donev, 1992; Uciński and Bogacka, 2005; Uciński and Atkinson, 2004) and its extensions to models for dynamic systems, especially in the context of the optimal choice of sampling instants and input signals (Goodwin and Payne, 1977; Titterton, 1980; Ljung, 1999; Gevers, 2005; Hjalmarsson, 2005). In this vein, various computational schemes have been developed to attack directly the original problem or its convenient approximation. The adopted optimization criteria are essentially the same, i.e., various scalar measures of performance based on the *Fisher Information Matrix* (FIM) associated with the parameters to be identified are maximized. The underlying idea is to express the goodness of parameter estimates in terms of the covariance matrix of the estimates. For sensor-location purposes, one assumes that an unbiased and efficient (or minimum-variance) estimator is employed so that the optimal sensor placement can be determined independently of the estimator used. This leads to a great simplification since the Cramér–Rao lower bound for the aforementioned covariance matrix is merely the inverse of the FIM, which can be computed with relative ease, even though the exact covariance matrix of a particular estimator is very difficult to obtain.

As regards dynamic DPSs, the first treatment of this type for the sensor-location problem was proposed by Uspenskii and Fedorov (1975), who maximized the D-optimality criterion, being the determinant of the FIM associated with the estimated parameters characterizing the source term in a simple one-dimensional linear diffusion equation. The authors observed that the linear dependence of the observed outputs on these parameters makes it possible to directly apply the machinery of optimum experimental design theory. The delineated approach was extended by Rafajłowicz (1981) to cover a class of DPSs described by linear hyperbolic equations with known eigenfunctions and unknown eigenvalues. The aim was to find conditions for the optimality of the measurement design and the spectral density of the stochastic input. It was indicated that common numerical procedures from classical experimental design for linear regression models could be adopted to find optimal sensor location. Moreover, the demonstrated optimality conditions imply that the optimal input comprises a finite number of sinusoidal signals and that optimal sensor positions are not difficult to find in some cases. A similar problem was studied by Rafajłow-

icz (1983) in a more general framework of DPSs which can be described in terms of Green's functions.

Over the past two decades, this methodology has been substantially refined to extend its applicability. A comprehensive treatment of both theoretical and algorithmic aspects of the resulting sensor location strategies is contained in the monograph by Uciński (2005). The potential of the approach for generalizations was exploited, e.g., by Munack (1984), who optimally located a given number of stationary sensors using nonlinear programming techniques for a biotechnological system consisting of a bubble column loop fermenter. On the other hand, Sun (1994) advocates using optimum experimental design techniques to solve inverse problems in groundwater modelling. How to monitor the water quality around a landfill place is an example of such a network design. Nonlinear programming techniques are also used there to find numerical approximations to the respective exact solutions.

A similar approach was used by Kammer (1990; 1992) for on-orbit modal identification of large space structures. Although the respective models are not PDEs but their discretized versions obtained through the finite-element method, the proposed solutions can still be of interest owing to the striking similitude of both the formulations. A fast and efficient approach was delineated for reducing a relatively large initial candidate sensor-location set to a much smaller optimum set which retains the linear independence of the target modes and maintains the determinant of the FIM resulting in more accurate modal-response estimates. Some improvements on this approach by incorporating basic elements of tabu search were proposed by Kincaid and Padula (2002).

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

It should be noted that spatial design methods related to the design of monitoring networks are also of great interest to statisticians and a vast amount of literature on the subject already exists (Müller, 2001; Nychka *et al.*, 1998; Nychka and Saltzman, 1998) contributing to the research field of spatial statistics (Cressie, 1993) motivated by practical problems in agriculture, geology, meteorology, environmental sciences and economics. However, the models considered in the statistical literature are quite different from the dynamic models described by

PDEs discussed here. Spatiotemporal data are not considered in this context and the main purpose is to model the spatial process by a spatial random field, incorporate prior knowledge and select the best subset of points of a desired cardinality to best represent the field in question. The motivation is a need to interpolate the observed behaviour of a process at unobserved spatial locations, as well as to design a network of optimal observation locations which allows an accurate representation of the process. The field itself is modelled by some multivariate distribution, usually Gaussian (Armstrong, 1998). Designs for spatial trend and variogram estimation can be considered. The basic theory of optimal design for spatial random fields is outlined in the excellent monograph by Müller (2001), which bridges the gap between spatial statistics and classical optimum experimental design theory. The optimal design problem can also be formulated in terms of information-based criteria, whose application amounts to maximizing the amount of information (of the Kullback–Leibler type) to be gained from an experiment (Caselton and Zidek, 1984; Caselton *et al.*, 1992). However, the applicability of all those fine statistical results in the engineering context discussed here is not clear for now, and more detailed research into this direction should be pursued in the near future (specifically, generalizations regarding time dynamics are not obvious).

Let us remark that an appealing alternative to stationary sensors is to apply spatially movable ones, which leads to the so-called continuous scanning observations. The complexity of the resulting optimization problem is compensated by a number of benefits. Specifically, sensors are not assigned to fixed positions which are optimal only on the average, but are capable of tracking points which provide at a given time moment the best information about the parameters to be identified. Consequently, by actively reconfiguring a sensor system we can expect the minimal value of an adopted design criterion to be lower than the one for the stationary case. What is more, technological advances in communication systems and the growing ease in making small, low power and inexpensive mobile systems now make it feasible to deploy a group of networked vehicles in a number of environments (Ögren *et al.*, 2004; Chong and Kumar, 2003; Sinopoli *et al.*, 2003; Cassandras and Li, 2005; Martínez and Bullo, 2006). Very prospective approaches in the context of control and state estimation were set forth by Demetriou (cf. Demetriou and Hussein, 2009; Demetriou, 2009; 2010) and Khapalov (2010). But some counterparts exist for parameter estimation, too.

In his seminal article, Rafajłowicz (1986) considers the D-optimality criterion and seeks an optimal time-dependent measure, rather than the trajectories themselves. On the other hand, Uciński (2005; 2000), apart from generalizations of Rafajłowicz's results, develops some computational algorithms based on the FIM. He re-

duces the problem to a state-constrained optimal-control one for which solutions are obtained via the methods of successive linearizations which is capable of handling various constraints imposed on sensor motions. Quite similar techniques were used in the works of Nehorai *et al.* (1995), Porat and Nehorai (1996), Jeremić and Nehorai (1998; 2000) as well as Zhao and Nehorai (2006) to detect and estimate the chemical dispersion of possibly moving sources using model-based integrated sensor array processing. In turn, the work by Uciński and Chen (2005) was intended as an attempt to properly formulate and solve the time-optimal problem for moving sensors which observe the state of a DPS so as to estimate some of its parameters. Some possibilities of exploiting parallel computations on clusters of PCs to reduce the time spent on determining optimal sensor node trajectories were discussed by Zięba and Uciński (2008). More applications in the area of sensor networks can be found in the recent monograph by Song *et al.* (2009).

1.3. Design in the context of fault detection. In spite of the rapid development of fault detection and localization methods for dynamic systems (Korbicz *et al.*, 2004; Isermann, 1997; Patton and Korbicz, 1999; Patton *et al.*, 2000; Chiang *et al.*, 2001), there have been very few attempts to devise techniques tailored to spatiotemporal systems. Patan and Patan (2005) developed a fault detection scheme for DPSs based on the maximization of the power of a parametric hypothesis test regarding the nominal state of a given DPS. They exploited the D_s -optimality criterion which is a standard tool in optimum experimental design theory (Atkinson *et al.*, 2007).

This line of research was further developed by Patan and Uciński (2008), who proposed a practical approach to sensor selection for fault detection which, while being independent of a particular model of the dynamic DPS in question, is versatile enough to cope with practical monitoring networks consisting of many stationary sensors. Specifically, N possible sites at which sensor resided were considered, but only n of them (typically, n is much smaller than N) were supposed to be activated. Consequently, the problem was to divide the N available sites between n gauged sites and the remaining $N - n$ ungauged sites so as to maximize the D_s -optimality criterion defined on the Fisher information matrix 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 was developed, which implicitly enumerates all the feasible sensor configurations, using relaxed optimization problems that involve no integer constraints (cf. also Uciński and Patan, 2007).

The D_s -optimality criterion was also employed by

Uciński and Patan (2010) to determine mobile sensor trajectories to detect and localize moving contamination sources. A receding-horizon scheme was then used to make this technique work on-line.

This overview would be by far incomplete without passing reference to the works which attempt to adapt the common fault detection approach based on observers or Kalman filters, cf., e.g., the work by Demetriou *et al.* (2007). The design problem aiming at optimizing the reliability of the diagnosis has not been considered yet in this setting.

1.4. Objective of this paper. The aim of the present paper is to propose an alternative technique to that set forth by Patan and Uciński (2008). It is going to alleviate problems with the combinatorial nature of the node activation problem in large-scale sensor networks. It consists in operating on the spatial density of sensors (i.e., the number of sensors per unit area), rather than on the sensor locations. It proved reasonable for a sufficiently large number of sensors as far as the reliability of model predictions was the main focus of interest (cf. Uciński, 2010). The underlying idea has its origins in the concept of replication-free designs in spatial statistics, (cf. Fedorov, 1989), and over the past few years successful attempts have been made at adapting it for use in problems ranging from maximization of observability (Uciński, 2005) to optimization of measurement strategies for scanning observations (Uciński, 2005, Chapter 4.1.1). What is more, convenient and efficient mathematical tools of convex programming theory made it possible to derive interesting characterizations of D_s -optimal solutions.

In this paper, the original problem of optimal node activation for fault detection is reduced to maximization of the D_s -optimality criterion over the set of all convex combinations of a finite number of nonnegative definite matrices subject to additional box constraints on the weights of those combinations. Then simplicial decomposition is applied which is a simple and direct method for dealing with large-scale convex optimization problems (von Hohenbalken, 1977; Patriksson, 2001). The decomposition iterates by alternately solving a linear programming subproblem within the set of all feasible points and a nonlinear master problem within the convex hull of a subset of previously generated points. The latter is solved by a gradient projection type method. As a result, an uncomplicated computational scheme is obtained which can be easily implemented without resorting to sophisticated numerical software.

Notation. Throughout the paper, \mathbb{R}_+ and \mathbb{R}_{++} stand for the sets of nonnegative and positive real numbers, respectively. We use \mathbb{S}^m to denote the set of symmetric $m \times m$ 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 $\mathbf{A}, \mathbf{B} \in \mathbb{S}^m$, $\mathbf{A} \succeq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is nonnegative definite (resp. positive definite). The symbols $\mathbf{1}$ and $\mathbf{0}$ denote vectors whose all components are one and zero, respectively. We call a point of the form $\alpha_1 \mathbf{a}_1 + \dots + \alpha_\ell \mathbf{a}_\ell$, where $\alpha_1 + \dots + \alpha_\ell = 1$ and $\alpha_i \geq 0$, $i = 1, \dots, \ell$, a convex combination of the points $\mathbf{a}_1, \dots, \mathbf{a}_\ell$. Given a set of points A , $\text{co}(A)$ stands for its convex hull, i.e., the set of all convex combinations of elements of A . The probability (or canonical) simplex in \mathbb{R}^n is defined as $S_n = \text{co}(\{\mathbf{e}_1, \dots, \mathbf{e}_n\})$, where \mathbf{e}_j is the usual unit vector along the j -th coordinate of \mathbb{R}^n .

2. Sensor selection for fault detection

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

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

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

$$\mathcal{B}(\mathbf{x}, t, y, \boldsymbol{\theta}) = 0 \quad \text{on } \Gamma \times T, \quad (2)$$

$$y = y_0 \quad \text{in } \Omega \times \{t = 0\}, \quad (3)$$

respectively, \mathcal{B} being an operator acting on the boundary Γ and $y_0 = y_0(\mathbf{x})$ a given function. The 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{F} and \mathcal{B} are given explicitly up to an m -dimensional vector of unknown constant parameters $\boldsymbol{\theta}$ which must inevitably be estimated using observations of the system. The implicit dependence of the state y on the parameter vector $\boldsymbol{\theta}$ will be reflected by the notation $y(\mathbf{x}, t; \boldsymbol{\theta})$.

In the sequel, we consider the discrete-continuous observations provided by N stationary pointwise sensors, namely,

$$z_m^\ell(t) = y(\mathbf{x}^\ell, t; \boldsymbol{\theta}) + \varepsilon(\mathbf{x}^\ell, t), \quad t \in T, \quad (4)$$

where $z_m^\ell(t)$ is the scalar output and $\mathbf{x}^\ell \in X$ stands for the location of the ℓ -th sensor ($\ell = 1, \dots, N$), X signifies the part of the spatial domain Ω where the measurements can be made and $\varepsilon(\mathbf{x}^\ell, t)$ 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 *et al.*, 1980; Omatu and Seinfeld, 1989; Amouroux and Babary, 1988), i.e.,

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

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. Uciński, 2005, Appendix C1) 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 $z_m^\ell(\cdot)$, $\ell = 1, \dots, N$, estimate $\boldsymbol{\theta}$ by $\hat{\boldsymbol{\theta}}$, a global minimizer of the output least-squares error criterion

$$\mathcal{J}(\boldsymbol{\theta}) = \sum_{\ell=1}^N \int_T \{z_m^\ell(t) - y(\mathbf{x}^\ell, t; \boldsymbol{\theta})\}^2 dt, \quad (6)$$

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

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; Patan and Uciński, 2008).

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

$$\boldsymbol{\theta} = [\theta_1 \dots \theta_s \vdots \theta_{s+1} \dots \theta_m]^\top = \begin{bmatrix} \boldsymbol{\alpha}^\top \vdots \boldsymbol{\beta}^\top \end{bmatrix}, \quad (7)$$

where $\boldsymbol{\alpha}$ is a vector of s parameters which are essential for a proper fault detection and $\boldsymbol{\beta}$ is the vector of $m - s$ 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

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

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 (for details, see Goodwin and Payne, 1977; Patan and Patan, 2005)

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

where $M \in \mathbb{R}^{m \times m}$ stands for the so-called *Fisher information matrix* which is decomposed as

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

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 *et al.*, 2007). In our setting, the FIM is given by (Quereschi *et al.*, 1980)

$$M(x^1, \dots, x^N) = \sum_{\ell=1}^N \int_T g(x^\ell, t) g^T(x^\ell, t) dt, \quad (11)$$

where

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

stands for the so-called *sensitivity vector*, θ^0 being the nominal value of the parameter vector θ (Uciński, 2005; Sun, 1994; Rafajłowicz, 1981; 1983). Up to a constant scalar multiplier, the inverse of the FIM constitutes a good approximation of $\text{cov}(\hat{\theta})$ 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

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

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)

$$D_{\alpha\alpha} = (M_{\alpha\alpha} - M_{\alpha\beta}M_{\beta\beta}^{-1}M_{\alpha\beta}^T)^{-1} \quad (14)$$

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

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

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 .

It can be demonstrated (Patan and Uciński, 2008) that the D_s -optimality criterion is concave over the cone of symmetric and positive definite $m \times m$ matrices. What is more, its matrix derivative is

$$\overset{\circ}{\Psi}(M) = \frac{\partial \Psi(M)}{\partial M} = \begin{bmatrix} D_{\alpha\alpha} & D_{\alpha\beta} \\ D_{\alpha\beta}^T & D_{\beta\beta} - M_{\beta\beta}^{-1} \end{bmatrix}, \quad (16)$$

which results from the representation

$$\begin{aligned} \Psi[M] &= \log \det(M) - \log \det(M_{\beta\beta}) \\ &= \log \det(M) - \log \det(A^T M A), \end{aligned} \quad (17)$$

where

$$A = \begin{bmatrix} 0 \\ \dots \\ I \end{bmatrix} \in \mathbb{R}^{m \times (m-s)}, \quad (18)$$

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

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

Problem 1 Find locations $x^{\ell*}$, $\ell = 1, \dots, N$, belonging to the admissible set X to maximize

$$\Upsilon(x^1, \dots, x^N) = \Psi[M(x^1, \dots, x^N)]. \quad (19)$$

3. Reformulation in terms of D_s -optimal sensor densities

As far as the number of sensors N is large, which becomes a common situation in applications involving sensor networks, the optimal sensor location problem becomes extremely difficult from a computational point of view due to its combinatorial nature. A way to overcome this predicament is to operate on the spatial density of sensors (i.e., the number of sensors per unit area), rather than on the sensor locations. The density of sensors over X can be approximately described by a probability measure $\xi(d\mathbf{x})$ on the space (X, \mathcal{B}) , where \mathcal{B} is the σ -algebra of all Borel subsets of X . Feasible solutions of this form make it possible to apply convenient and efficient mathematical tools of convex programming theory.

As regards the practical interpretation of the so produced results (provided that we are in a position to calculate at least their approximations), one possibility is to

partition X into non-overlapping subdomains X_i of relatively small areas and then to allocate to each of them the number

$$N_i = \left\lceil N \int_{X_i} \xi(d\mathbf{x}) \right\rceil \quad (20)$$

of sensors (here $\lceil \rho \rceil$ is the smallest integer greater than or equal to ρ).

Accordingly, we define the class of admissible designs as all probability measures ξ over X which are absolutely continuous with respect to the Lebesgue measure and satisfy by definition the condition

$$\int_X \xi(d\mathbf{x}) = 1. \quad (21)$$

Consequently, slightly abusing the notation, we replace (11) by

$$\mathbf{M}(\xi) = \int_X \mathbf{G}(\mathbf{x}) \xi(d\mathbf{x}), \quad (22)$$

where

$$\mathbf{G}(\mathbf{x}) = \int_0^{t_f} \mathbf{g}(\mathbf{x}, t) \mathbf{g}^\top(\mathbf{x}, t) dt.$$

The integration in (21) and (22) is to be understood in the Lebesgue–Stieltjes sense. This leads to the so-called *continuous* designs, which constitute the basis of the modern theory of optimal experiments (Fedorov and Hackl, 1997; Walter and Pronzato, 1997).

We impose the crucial restriction that the density of sensor allocation must not exceed some prescribed level. For a design measure $\xi(d\mathbf{x})$ this amounts to the condition

$$\xi(d\mathbf{x}) \leq \omega(d\mathbf{x}), \quad (23)$$

where $\omega(d\mathbf{x})$ signifies the maximal possible ‘number’ of sensors per $d\mathbf{x}$ (Fedorov and Hackl, 1997; Uciński, 1999; 2005). Logically, this bounding measure must satisfy the condition

$$\int_X \omega(d\mathbf{x}) \geq 1. \quad (24)$$

Consequently we are faced with the following optimization problem:

Problem 2 Find a measure $\xi^* \in \Xi(X)$ to maximize

$$\Lambda(\xi) = \Psi(\mathbf{M}(\xi)) \quad (25)$$

subject to

$$\xi(d\mathbf{x}) \leq \omega(d\mathbf{x}), \quad (26)$$

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

The design ξ^* above is then said to be a (Ψ, ω) -optimal design (Fedorov and Hackl, 1997).

Let us now make the following assumptions:

(A1) X is compact.

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

(A3) There exists a finite real α such that

$$\{\xi : \Psi[\mathbf{M}(\xi)] < \alpha\} = \tilde{\Xi}(X) \neq \emptyset.$$

(A4) $\omega(d\mathbf{x})$ is atomless, i.e., for any $\Delta X \subset X$ there exists a $\Delta X' \subset \Delta X$ such that

$$\int_{\Delta X'} \omega(d\mathbf{x}) < \int_{\Delta X} \omega(d\mathbf{x}). \quad (27)$$

In what follows, we write $\tilde{\Xi}(X)$ for the collection of all the design measures which are specific in that they coincide with ω within their supports, i.e., they satisfy the requirement

$$\xi(\Delta X) = \begin{cases} \omega(\Delta X) & \text{for } \Delta X \subset \text{supp } \xi, \\ 0 & \text{for } \Delta X \subset X \setminus \text{supp } \xi. \end{cases} \quad (28)$$

(Recall that the support of a measure ξ is defined as the closed set $\text{supp } \xi = X \setminus \bigcup \{G : \xi(G) = 0, G \text{ open}\}$ (cf. Rao, 1987, p. 80).)

Given a design ξ , we will say that the function $\psi(\cdot, \xi)$ defined by

$$\psi(\mathbf{x}, \xi) = \frac{1}{t_f} \int_0^{t_f} \mathbf{g}^\top(\mathbf{x}, t) \tilde{\Psi}(\mathbf{M}(\xi)) \mathbf{g}(\mathbf{x}, t) dt \quad (29)$$

separates sets X_1 and X_2 with respect to $\omega(d\mathbf{x})$ 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

$$\int_{\Delta X_1} \psi(\mathbf{x}, \xi) \omega(d\mathbf{x}) \geq \int_{\Delta X_2} \psi(\mathbf{x}, \xi) \omega(d\mathbf{x}). \quad (30)$$

We can now formulate the following characterization of (Ψ, ω) -optimal designs.

Theorem 1. (Uciński, 1999; 2005) Let Assumptions (A1)–(A4) hold. Then

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

From a practical point of view, the above result means that at all the support points of an optimal design ξ^* the values of the mapping $\psi(\cdot, \xi^*)$ should be greater than anywhere else, i.e., preferably $\text{supp } \xi^*$ should coincide with maximum points of $\psi(\cdot, \xi^*)$. In practice, this amounts to allocating observations to the points which provide the most valuable information about the vector α .

Theorem 1 is a valuable tool to check whether or not a given design is (Ψ, ω) -optimal, but the question of how to construct an approximation to this optimal solution is still open. It constitutes the main topic of the next section.

4. Conversion to a weight optimization problem

We are now interested in the question of how to discretize Problem 2 to make it tractable by a computer. In what follows, the basic idea is to make use of the partition of X into a union of small disjoint subdomains X_i , $i = 1, \dots, n$, i.e., $X = \bigcup_{i=1}^n X_i$, as discussed in Section 3. Observe that a measure $\xi \in \Xi(X)$ assigns each subdomain X_i a weight $p_i = \xi(X_i)$. Owing to (20), the knowledge of the p_i 's suffices to determine an optimal distribution of sensor nodes. Assuming that the variation of $G(\cdot)$ over each X_i is negligible (this can be achieved by constructing a sufficiently fine partition of X), we have

$$M(\xi) \approx M(p) := \sum_{i=1}^n p_i M_i, \quad (31)$$

where $p = [p_1, \dots, p_n]^T$, $M_i = G(x^i)$, x^i being an arbitrary point in X_i (e.g., its centre of gravity). Thus, setting $b = [\omega(X_1), \dots, \omega(X_n)]^T$, we arrive at the following approximation to Problem 2.

Problem 3 Find a vector of weights p to maximize

$$\Phi(p) = \Psi[M(p)] \quad (32)$$

subject to

$$0 \preceq p \preceq b, \quad (33)$$

$$\mathbf{1}^T p = 1. \quad (34)$$

In what follows, we let P be the bounded polyhedral set of feasible weights defined by (33) and (34). Moreover, without restriction of generality, we shall further assume that $b \succ 0$.

Note that the performance index Φ is concave over the canonical simplex $S_n = \{p \in \mathbb{R}_+^n \mid \mathbf{1}^T p = 1\}$. What is more, it is differentiable at points in S_n yielding nonsingular FIMs, with $\phi(p) := \nabla \Phi(p)$ given by

$$\phi(p) = \begin{bmatrix} \text{trace}\{\ddot{\Psi}(M(p))M_1\} \\ \vdots \\ \text{trace}\{\ddot{\Psi}(M(p))M_n\} \end{bmatrix}. \quad (35)$$

Accordingly, numerous computational methods can potentially be employed for solving Problem 3, e.g., the conditional gradient method or a gradient projection method. Unfortunately, for large n , these algorithms may lead to unsatisfactory computational times. In what follows, it will be shown how simplicial decomposition can be employed to build a very simple and efficient computational scheme for solving Problem 3.

Algorithm 1 Algorithm model for solving Problem 3 via simplicial decomposition.

Step 0: (Initialization)

Guess an initial solution $p^{(0)} \in P$ such that $M(p^{(0)})$ is nonsingular. Set $I = \{1, \dots, n\}$, $Q^{(0)} = \{p^{(0)}\}$ and $k = 0$.

Step 1: (Termination check)

Set

$$I_{\text{ub}}^{(k)} = \{i \in I \mid p_i^{(k)} = b_i\}, \quad (36)$$

$$I_{\text{im}}^{(k)} = \{i \in I \mid 0 < p_i^{(k)} < b_i\}, \quad (37)$$

$$I_{\text{lb}}^{(k)} = \{i \in I \mid p_i^{(k)} = 0\}. \quad (38)$$

If

$$\phi_i(p^{(k)}) \begin{cases} \geq \lambda & \text{if } i \in I_{\text{ub}}^{(k)}, \\ = \lambda & \text{if } i \in I_{\text{im}}^{(k)}, \\ \leq \lambda & \text{if } i \in I_{\text{lb}}^{(k)} \end{cases} \quad (39)$$

for some $\lambda \in \mathbb{R}_+$, then STOP and $p^{(k)}$ is optimal.

Step 2: (Solution of the column generation problem)

Compute

$$q^{(k+1)} = \arg \max_{p \in P} \phi(p^{(k)})^T p \quad (40)$$

and set

$$Q^{(k+1)} = Q^{(k)} \cup \{q^{(k+1)}\}. \quad (41)$$

Step 3: (Solution of the restricted master problem)

Find

$$p^{(k+1)} = \arg \max_{p \in \text{co}(Q^{(k+1)})} \Phi(p) \quad (42)$$

and purge $Q^{(k+1)}$ of all extreme points with zero weights in the resulting expression of $p^{(k+1)}$ as a convex combination of elements in $Q^{(k+1)}$. Increment k by one and go back to Step 1.

5. Simplicial decomposition for Problem 3

5.1. Algorithm model. Simplicial Decomposition (SD) proved extremely useful for large-scale pseudoconvex programming problems encountered, e.g., in traffic assignment or other network flow problems (Patriksson, 2001). In its basic form, it proceeds by alternately solving linear and nonlinear programming subproblems, called the *Column Generation Problem* (CGP) and the *restricted Master Problem* (RMP), respectively. In the RMP, the original problem is relaxed by replacing the original constraint set P with its inner approximation being the convex hull of a finite set of feasible solutions. In the CGP, this inner approximation is improved by incorporating a point in the original constraint set that lies furthest along the

gradient direction computed at the solution of the RMP. This basic strategy has been discussed in numerous references (von Hohenbalken, 1977; Hearn *et al.*, 1985; Patriksson, 2001), where possible extensions have also been proposed. A marked characteristic of the SD method is that the sequence of solutions to the RMP tends to a solution to the original problem in such a way that the objective function strictly monotonically approaches its optimal value.

The SD algorithm may be viewed as a form of modular nonlinear programming, provided that one has an effective computer code for solving the RMP, as well as access to a code which can take advantage of the linearity of the CGP. A principal aim of this paper is to show that this is the case within the framework of Problem 3. What is more, since we deal with maximization of the concave function Φ over a bounded polyhedral set P , this will automatically imply the convergence of the resulting SD scheme in a finite number of RMP steps (von Hohenbalken, 1977).

Tailoring the SD scheme to our needs, we obtain Algorithm 1. In the sequel, its consecutive steps will be discussed in turn.

5.2. Termination criterion for Algorithm 1. In the original SD setting, the criterion for terminating the iterations is checked only after solving the column generation problem. The computation is then stopped if the current point $\mathbf{p}^{(k)}$ satisfies the “basic” optimality condition of nonincrease, to first order, in the performance measure value in the whole constraint set, i.e.,

$$\max_{\mathbf{p} \in P} \phi(\mathbf{p}^{(k)})^\top (\mathbf{p} - \mathbf{p}^{(k)}) \leq 0. \quad (43)$$

Note, however, that the condition (39) is less costly in terms of the number of floating-point operations. It results from the following characterization of \mathbf{p}^* , which has the property that $\Phi(\mathbf{p}^*) = \max_{\mathbf{p} \in P} \Phi(\mathbf{p})$.

Proposition 1. Suppose that the matrix $\mathbf{M}(\mathbf{p}^*)$ is non-singular for some $\mathbf{p}^* \in P$. The vector \mathbf{p}^* constitutes a global minimum of Φ over P if, and only if, there exists a number λ^* such that

$$\phi_i(\mathbf{p}^*) \begin{cases} \leq \lambda^* & \text{if } p_i^* = b_i, \\ = \lambda^* & \text{if } 0 < p_i^* < b_i, \\ \geq \lambda^* & \text{if } p_i^* = 0 \end{cases} \quad (44)$$

for $i = 1, \dots, n$.

This result follows from direct application of Lemma 1 (see Appendix) after setting $f(\mathbf{p}) = \Phi(\mathbf{M}(\mathbf{p}))$.

5.3. Solution of the column generation subproblem. Setting $\mathbf{c} = \phi(\mathbf{p}^{(k)})$, in Step 2 we deal with the linear

Algorithm 2 Algorithm model for solving the column generation subproblem.

Step 0: (Initialization)

Set $j = 0$ and $v^{(0)} = 0$.

Step 1: (Sorting)

Sort the elements of \mathbf{c} in nonincreasing order, i.e., find a permutation π on the index set $I = \{1, \dots, n\}$ such that

$$c_{\pi(i)} \geq c_{\pi(i+1)}, \quad i = 1, \dots, n-1. \quad (45)$$

Step 2: (Identification of nonzero weights)

Step 2.1: If $v^{(j)} + b_{\pi(j+1)} < 1$, then set

$$v^{(j+1)} = v^{(j)} + b_{\pi(j+1)}. \quad (46)$$

Otherwise, go to Step 3.

Step 2.2: Increment j by one and go to Step 2.1.

Step 3: (Form the ultimate solution)

Set

$$q_{\pi(i)} = \begin{cases} b_{\pi(i)} & \text{for } i = 1, \dots, j, \\ 1 - v^{(j)} & \text{for } i = j+1, \\ 0 & \text{for } i = j+2, \dots, n. \end{cases} \quad (47)$$

programming problem

$$\begin{aligned} & \text{maximize} && \mathbf{c}^\top \mathbf{p} \\ & \text{subject to} && \mathbf{p} \in P. \end{aligned} \quad (48)$$

The following assertion is a direct consequence of Lemma 1 in Appendix.

Proposition 2. A vector $\mathbf{q} \in P$ constitutes a global solution to the problem (48) if, and only if, there exists a scalar ρ such that

$$c_i \begin{cases} \geq \rho & \text{if } q_i = b_i, \\ = \rho & \text{if } 0 < q_i < b_i, \\ \leq \rho & \text{if } q_i = 0 \end{cases} \quad (49)$$

for $i = 1, \dots, n$.

We thus see that, in order to solve (48), it is sufficient to pick the consecutive smallest components c_i of \mathbf{c} and set the corresponding weights q_i as their maximal allowable values b_i . The process is repeated until the sum of the assigned weights exceeds one. Then the value of the last weight which was set in this manner should be corrected so as to satisfy the constraint (34) and the remaining (i.e., unassigned) weights are set as zeros. This

straightforward scheme is implemented as Algorithm 2. Note that its correctness requires satisfaction of the condition $\mathbf{b} \preceq \mathbf{1}$, which is by no means restrictive.

Algorithm 3 Algorithm model for solving the restricted master problem.

Step 0: (Initialization)

Choose an initial point $\mathbf{w}^{(0)} \in S_r$ and set $\kappa = 0$.

Step 1: (Determination of a feasible direction and termination check)

Find

$$\mathbf{v}^{(\kappa)} = \mathcal{D}(\mathbf{w}^{(\kappa)}, \phi(\mathbf{w}^{(\kappa)})) \quad (50)$$

$$\mathbf{a}^{(\kappa)} = \phi^T(\mathbf{w}^{(\kappa)})\mathbf{v}^{(\kappa)}. \quad (51)$$

If $\mathbf{a}^{(\kappa)} = \mathbf{0}$, then STOP and $\mathbf{w}^{(\kappa)}$ is optimal.

Step 2: (Line search)

Determine

$$\begin{aligned} \gamma_{\max} &= \max\{\gamma \in \mathbb{R}_+ \mid \mathbf{w}^{(\kappa)} + \gamma\mathbf{v}^{(\kappa)} \in S_r\} \\ &= \min_{v_i^{(\kappa)} < 0} \left\{ -\frac{w_i^{(\kappa)}}{v_i^{(\kappa)}} \right\} \end{aligned} \quad (52)$$

and find

$$\gamma^{(\kappa)} = \arg \max_{\gamma \in [0, \gamma_{\max}]} \Phi(\mathbf{w}^{(\kappa)} + \gamma\mathbf{v}^{(\kappa)}). \quad (53)$$

Step 3: (Updates for the next iteration)

Set $\mathbf{w}^{(\kappa+1)} = \mathbf{w}^{(\kappa)} + \gamma^{(\kappa)}\mathbf{v}^{(\kappa)}$, increment k by one and go back to Step 1.

5.4. Solution of the restricted master subproblem.

Suppose that in the $(k+1)$ -th iteration of Algorithm 1, we have

$$Q^{(k+1)} = \{\mathbf{q}_1, \dots, \mathbf{q}_r\}, \quad (54)$$

possibly with $r < k+1$ owing to the built-in deletion mechanism of points in $Q^{(j)}$, $1 \leq j \leq k$, which did not contribute to the convex combinations yielding the corresponding iterates $\mathbf{p}^{(j)}$. Step 3 involves maximization of $\Phi(\cdot)$ over

$$\text{co}(Q^{(k+1)}) = \left\{ \sum_{j=1}^r w_j \mathbf{q}_j \mid \mathbf{w} \succeq \mathbf{0}, \mathbf{1}^T \mathbf{w} = 1 \right\}. \quad (55)$$

From the representation of any $\mathbf{p} \in \text{co}(Q^{(k+1)})$ as

$$\mathbf{p} = \sum_{j=1}^r w_j \mathbf{q}_j, \quad (56)$$

or, in component-wise form,

$$p_i = \sum_{j=1}^r w_j q_{j,i}, \quad i = 1, \dots, n, \quad (57)$$

$q_{j,i}$ being the i -th component of \mathbf{q}_j , it follows that

$$\begin{aligned} \mathbf{M}(\mathbf{p}) &= \sum_{i=1}^n p_i \mathbf{M}_i \\ &= \sum_{j=1}^r w_j \left(\sum_{i=1}^n q_{j,i} \mathbf{M}_i \right) = \sum_{j=1}^r w_j \mathbf{M}(\mathbf{q}_j). \end{aligned} \quad (58)$$

From this, we see that the RMP can equivalently be formulated as the following problem:

Problem 4 Find a sequence of weights $\mathbf{w} \in \mathbb{R}^r$ to maximize

$$\Phi(\mathbf{w}) = \Psi[\mathbf{H}(\mathbf{w})] \quad (59)$$

subject to the constraints

$$\mathbf{1}^T \mathbf{w} = 1, \quad (60)$$

$$\mathbf{w} \succeq \mathbf{0}, \quad (61)$$

where

$$\mathbf{H}(\mathbf{w}) = \sum_{j=1}^r w_j \mathbf{H}_j, \quad (62)$$

$$\mathbf{H}_j = \mathbf{M}(\mathbf{q}_j), \quad j = 1, \dots, r. \quad (63)$$

Note that the above RMP is much simpler than Problem 3, since the number of decision variables is usually substantially smaller here and there are no upper bounds to the weights. Instead, we maximize a concave function on the convex hull of a finite set of points. This formulation has captured close attention in optimum experimental design and, basically, the well-known first-order Wynn–Fedorov algorithm could be employed here (cf. Fedorov and Hackl, 1997).

In lieu of this option, a gradient projection type method was applied here due to its slightly faster operation. Specifically, defining for any arbitrary $\mathbf{w} \in S_r = \{\mathbf{w} \in \mathbb{R}_+^r \mid \mathbf{1}^T \mathbf{w} = 1\}$ the set of feasible directions

$$\mathcal{D}(\mathbf{w}, \mathbf{h}) = \arg \max_{\mathbf{v} \in V(\mathbf{w})} \mathbf{h}^T \mathbf{v}, \quad (64)$$

where

$$\begin{aligned} V(\mathbf{w}) &= \{\mathbf{v} \in \mathbb{R}^r \mid \|\mathbf{v}\| \leq 1, \\ &\quad \exists \gamma > 0 : \mathbf{w} + \gamma \mathbf{v} \in S_r\}, \end{aligned} \quad (65)$$

we can represent this scheme in its most basic form as Algorithm 3

In order to increase its efficiency, numerous improvements can be made. That is why, while implementing

this part of the SD scheme, Algorithm 2 of Botkin and Stoer (2005) was used, which accounts for several substantial alterations. Nevertheless, the basic structure of Algorithm 3 is preserved in it. What is more, Botkin and Stoer (2005) proposed an extremely fast procedure for computing $\mathcal{D}(\mathbf{w}, \mathbf{h})$ (it uses only $\mathcal{O}(r \log(r))$ operations), which was also implemented here. As for the line search, a modified Brent method using derivatives was adopted (cf. Press *et al.*, 2007).

6. Simulation example

Consider the two-dimensional diffusion equation

$$\frac{\partial y}{\partial t} = \nabla \cdot (\mu \nabla y) \quad \text{in } \Omega \times T, \quad (66)$$

where $\Omega \subset \mathbb{R}^2$ is the spatial domain with boundary Γ shown in Fig. 1 and $T = (0, 1)$. The assumed form of the diffusion coefficient is

$$\mu(\mathbf{x}) = \theta_1 + \theta_2 x_1 + \theta_3 x_2, \quad (67)$$

where θ_1 , θ_2 and θ_3 are unknown parameters. We assume that only θ_2 and θ_3 are crucial for fault detection, i.e., $\boldsymbol{\alpha} = \text{col}[\theta_2, \theta_3]$ and $\boldsymbol{\beta} = [\theta_1]$. The parameter vector $\boldsymbol{\theta} = \text{col}[\theta_1, \theta_2, \theta_3]$ is supposed to be estimated based on the measurements from a large-scale sensor network. Throughout the design, $\theta_1^0 = 0.1$ and $\theta_2^0 = \theta_3^0 = 0.3$ are to be used as nominal values of θ_1 , θ_2 and θ_3 , respectively.

The PDE (66) is supplemented with the initial and boundary conditions

$$y(\mathbf{x}, 0) = 5 \quad \text{in } \Omega, \quad (68)$$

$$y(\mathbf{x}, t) = 5(1 - t) \quad \text{on } \Gamma \times T. \quad (69)$$

The upper bound in (23) is given by $\omega(A) = c|A|/|X|$ for any Borel subset of \bar{X} , where $c \geq 1$ is fixed, i.e., ω corresponds to a uniform distribution on \bar{X} , and $|A|$ stands for the area (i.e., the Lebesgue measure) of A .

The MATLAB PDE toolbox (COMSOL AB, 1995) was used to generate the triangular mesh \mathcal{T} of 646 nodes and $n = 1196$ triangles shown in Fig. 1. The values of the sensitivities $\mathbf{g} = \text{col}[g_1, g_2, g_3]$ are defined as solutions to the following system of PDEs (Uciński, 2005; 1999):

$$\begin{cases} \frac{\partial y}{\partial t} = \nabla \cdot (\mu \nabla y), \\ \frac{\partial g_1}{\partial t} = \nabla \cdot \nabla y + \nabla \cdot (\mu \nabla g_1), \\ \frac{\partial g_2}{\partial t} = \nabla \cdot (x_1 \nabla y) + \nabla \cdot (\mu \nabla g_2), \\ \frac{\partial g_3}{\partial t} = \nabla \cdot (x_2 \nabla y) + \nabla \cdot (\mu \nabla g_3), \end{cases} \quad (70)$$

in which the first equation constitutes the original state equation and the second and third equations result from

its differentiation with regard θ_1 , θ_2 and θ_3 , respectively. The initial and Dirichlet boundary conditions for the sensitivity equations are homogeneous.

We numerically solved (70) using some routines of the MATLAB PDE toolbox and stored matrices $\mathbf{G}(\mathbf{x}^i)$ computed based on g_1 , g_2 and g_3 interpolated at the gravity centres of individual triangles, cf. Appendix I of Uciński (2005) for details.

The triangular mesh \mathcal{T} was also used when solving Problem 3. All computations were performed within several seconds on a low-cost laptop (Intel Centrino Duo T9300, 3 GB RAM) running Windows Vista Home Premium and Matlab 7 (R2008a).

Simplicial decomposition implemented in accordance with Algorithm 1 lead to solutions in which the weights p_i associated with the respective triangles approximately satisfy the ‘bang-bang’ principle formulated in Theorem 1: they are either zero, or equal to the corresponding upper bound $b_i = c|X_i|/|X|$. Thus, the support of the optimal measure ξ covers approximately $c^{-1} \cdot 100\%$ of the area of $|X|$, cf. Fig. 1. Observe that both the symmetry of the solution imposed by the form of the diffusion coefficient, as well as that of the boundary and initial conditions, are virtually retained (some imprecisions results from a non-symmetric triangular mesh).

7. Conclusions

The paper has been focused on a design problem a monitoring sensor network which is supposed to provide proper diagnostic information about the functioning of a distributed parameter system. Fault detection is accomplished here based on monitoring the values of the parameters which govern the dynamics of the spatiotemporal system. One of the important issues related to this task is the choice of a suitable criterion to quantify the quality of the detection of an abnormal state indicating potential faults. Here the problem is formulated as maximization of the power for a parametric hypothesis test which verifies the nominal values of the system parameters. To this end, the D_s -optimality criterion defined on the Fisher information matrix is proposed. Although this criterion is well known in optimum design theory, there exist only few attempts to exploit it in the context of fault detection for DPSs.

The combinatorial nature of the design problem has been circumvented here by replacing the original problem, which operates directly on the sensor positions, by another one which operates on the spatial density of the sensors. In much the same way as is done in stochastic optimization, an approximated solution is then sought in the form of a constrained probability mass function concentrated at a finite number of given support points. The computational tool to effectively solve the resulting weight optimization problem is simplicial decomposition, which alternates be-

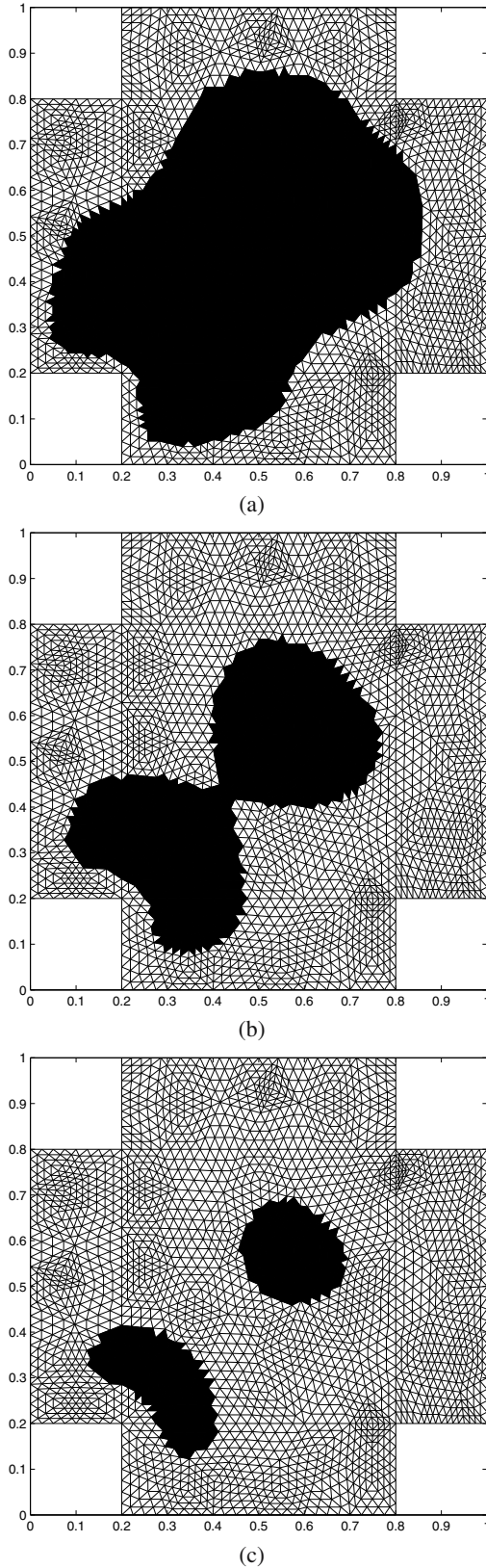


Fig. 1. Support of the D_s -optimal design measure: $c = 2$, 50% coverage of X (a), $c = 4$, 25% coverage of X (b), $c = 10$, 10% coverage of X (c).

tween a simple linear programming problem and a nice convex programming problem being solved by a gradient projection type method.

Observe that some refinements may still improve the algorithm performance. First of all, the basic simplicial decomposition scheme outlined here can be replaced by the so-called restricted simplicial decomposition (Hearn *et al.*, 1985), which is based on the fact that a particular feasible solution, such as the optimal one, can be represented as a convex combination of an often much smaller number of extreme points than that implied by Carathéodory's theorem. Apart from that, some improvements aimed at removing nonoptimal support points can be incorporated in the restricted master problem to speed up its solution. But even the basic scheme proposed here performs well in practice and the aforementioned refinements may be necessary only when the time of computations is a truly critical factor. Clearly, an extremely important issue regarding the technique outlined here is a thorough analysis of its computational efficiency and, specifically, a comparison with alternative approaches (e.g., the speed up, the potential for parallelization, etc.). This topic, however, is beyond the scope of this introductory paper and a detailed discussion of all these issues will appear elsewhere.

Although the approach has been presented in the form which conforms to off-line design, it can be relatively easily adapted to a setting where the computations are dynamic data driven, i.e., the current measurements from the sensor network nodes serve as a basis for the update of parameter estimates and these, in turn, can be used to correct the node activation schedule. To this end, a non-linear model-predictive-control-like approach can be used in much the same way as described by Uciński and Patan (2010).

The application of the D_s -optimality criterion is inherently related to splitting the set of all unknown parameters into the disjoint subsets of the parameters which are and are not relevant to fault detection. Such a clear distinction may be neither obvious nor practical in real world applications. Therefore, an alternative approach consists in arbitrarily weighing the influence of individual parameters based on some *a priori* knowledge gained, e.g., from previous experiments. But the implementation of this simple idea presents a more delicate problem and, therefore, will appear in a forthcoming publication.

Acknowledgment

The author wishes to express his gratitude to the reviewers for several helpful comments which improved the quality of the presentation.

This work was supported by the Polish Ministry of Science and Higher Education under Grant No. N N519 2971 33.

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.
- Armstrong, M. (1998). *Basic Linear Geostatistics*, Springer-Verlag, Berlin.
- Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Designs*, Clarendon Press, Oxford.
- Atkinson, A.C., Donev, A.N. and Tobias, R.D. (2007). *Optimum Experimental Designs, with SAS*, Oxford University Press, Oxford.
- Banks, H.T., Smith, R.C. and Wang, Y. (1996). *Smart Material Structures: Modeling, Estimation and Control*, Research in Applied Mathematics, Masson, Paris.
- Bauer, P.H. (2008). New challenges in dynamical systems: The networked case, *International Journal of Applied Mathematics and Computer Science* **18**(3): 271–277, DOI: 10.2478/v10006-008-0025-8.
- Bernstein, D.S. (2005). *Matrix Mathematics. Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton University Press, Princeton, NJ.
- Botkin, N.D. and Stoer, J. (2005). Minimization of convex functions on the convex hull of a point set, *Mathematical Methods of Operations Research* **62**(2): 167–18.
- Caselton, W.F., Kan, L. and Zidek, J.V. (1992). Quality data networks that minimize entropy, in A. Walden and P. Guttorp (Eds.), *Statistics in the Environmental and Earth Sciences*, Halsted Press, New York, NY, Chapter 2, pp. 10–38.
- Caselton, W.F. and Zidek, J.V. (1984). Optimal monitoring network design, *Statistics & Probability Letters* **2**: 223–227.
- 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.
- Chong, C.-Y. and Kumar, S.P. (2003). Sensor networks: Evolution, opportunities, and challenges, *Proceedings of the IEEE* **91**(8): 1247–1256.
- Christofides, P.D. (2001). *Nonlinear and Robust Control of PDE Systems: Methods and Applications to Transport-Reaction Processes*, Systems & Control: Foundations & Applications, Birkhäuser, Boston, MA.
- COMSOL AB (1995). *Partial Differential Equation Toolbox for Use with Matlab. User's Guide*, The MathWorks, Inc., Natick, MA.
- Cressie, N.A.C. (1993). *Statistics for Spatial Data*, Revised Edn., John Wiley & Sons, New York, NY.
- Daescu, D.N. and Navon, I.M. (2004). Adaptive observations in the context of 4D-Var data assimilation, *Meteorology and Atmospheric Physics* **85**: 205–226.
- Demetriou, M.A. (2009). Natural consensus filters for second order infinite dimensional systems, *Systems & Control Letters* **58**(12): 826–833.
- Demetriou, M.A. (2010). Guidance of mobile actuator-plus-sensor networks for improved control and estimation of distributed parameter systems, *IEEE Transactions on Automatic Control* **55**(7): 1570–1584.
- Demetriou, M.A. and Hussein, I.I. (2009). Estimation of spatially distributed processes using mobile spatially distributed sensor network, *SIAM Journal on Control and Optimization* **48**(1): 266–291.
- Demetriou, M.A., Ito, K. and Smith, R.C. (2007). Adaptive monitoring and accommodation of nonlinear actuator faults in positive real infinite dimensional systems, *IEEE Transactions on Automatic Control* **52**(12): 2332–2338.
- El Jai, A. and Hamzaoui, H. (2009). Regional observation and sensors, *International Journal of Applied Mathematics and Computer Science* **19**(1): 5–14, DOI: 10.2478/v10006-009-0001-y.
- Fedorov, V.V. (1989). Optimal design with bounded density: Optimization algorithms of the exchange type, *Journal of Statistical Planning and Inference* **22**: 1–13.
- Fedorov, V.V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*, Lecture Notes in Statistics, Springer-Verlag, New York, NY.
- Gevers, M. (2005). Identification for control: From the early achievements to the revival of experiment design, *European Journal of Control* **11**(4–5): 335–352.
- Goodwin, G. C. and Payne, R. L. (1977). *Dynamic System Identification. Experiment Design and Data Analysis*, Mathematics in Science and Engineering, Academic Press, New York, NY.
- Hearn, D.W., Lawphongpanich, S. and Ventura, J.A. (1985). Finiteness in restricted simplicial decomposition, *Operations Research Letters* **4**(3): 125–130.
- Hirsch, M.J., Pardalos, P.M., Murphey, R. and Grundel, D. (Eds.) (2008). *Advances in Cooperative Control and Optimization. Proceedings of the 7th International Conference on Cooperative Control and Optimization*, Springer-Verlag, Berlin.
- Hjalmarsson, H. (2005). From experiment design to closed-loop control, *Automatica* **41**: 393–438.
- Isermann, R. (1997). Supervision, Fault Detection and Diagnosis of Technical Systems, *Control Engineering Practice* **5**(5), Special section.
- Jain, N. and Agrawal, D.P. (2005). Current trends in wireless sensor network design, *International Journal of Distributed Sensor Networks* **1**: 101–122.
- Jeremić, A. and Nehorai, A. (1998). Design of chemical sensor arrays for monitoring disposal sites on the ocean floor, *IEEE Transactions on Oceanic Engineering* **23**(4): 334–343.
- Jeremić, A. and Nehorai, A. (2000). Landmine detection and localization using chemical sensor array processing, *IEEE Transactions on Signal Processing* **48**(5): 1295–1305.
- Kammer, D.C. (1990). Sensor placement for on-orbit modal identification and correlation of large space structures, *Proceedings of the American Control Conference, San Diego, CA, USA*, Vol. 3, pp. 2984–2990.

- Kammer, D.C. (1992). Effects of noise on sensor placement for on-orbit modal identification of large space structures, *Transactions of the ASME* **114**: 436–443.
- Khapalov, A.Y. (2010). Source localization and sensor placement in environmental monitoring, *International Journal of Applied Mathematics and Computer Science* **20**(3): 445–458, DOI: 10.2478/v10006-010-0033-3.
- Kincaid, R.K. and Padula, S.L. (2002). D-optimal designs for sensor and actuator locations, *Computers & Operations Research* **29**: 701–713.
- Korbicz, J., Kościelny, J., Kowalczyk, 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.
- Ljung, L. (1999). *System Identification: Theory for the User*, 2nd Edn., Prentice Hall, Upper Saddle River, NJ.
- Martínez, S. and Bullo, F. (2006). Optimal sensor placement and motion coordination for target tracking, *Automatica* **42**: 661–668.
- Müller, W.G. (2001). *Collecting Spatial Data. Optimum Design of Experiments for Random Fields*, 2nd Revised Edn., Contributions to Statistics, Physica-Verlag, Heidelberg.
- Munack, A. (1984). Optimal sensor allocation for identification of unknown parameters in a bubble-column loop bioreactor, in A.V. Balakrishnan and M. Thoma (Eds.), *Analysis and Optimization of Systems, Part 2*, Lecture Notes in Control and Information Sciences, Vol. 63, Springer-Verlag, Berlin, pp. 415–433.
- Navon, I.M. (1997). Practical and theoretical aspects of ad-joint parameter estimation and identifiability in meteorology and oceanography, *Dynamics of Atmospheres and Oceans* **27**: 55–79.
- Nehorai, A., Porat, B. and Paldi, E. (1995). Detection and localization of vapor-emitting sources, *IEEE Transactions on Signal Processing* **43**(1): 243–253.
- Nychka, D., Piegorsch, W.W. and Cox, L.H. (Eds.) (1998). *Case Studies in Environmental Statistics*, Lecture Notes in Statistics, Vol. 132, Springer-Verlag, New York, NY.
- Nychka, D. and Saltzman, N. (1998). Design of air-quality monitoring networks, in D. Nychka, W.W. Piegorsch and L.H. Cox (Eds.), *Case Studies in Environmental Statistics*, Lecture Notes in Statistics, Vol. 132, Springer-Verlag, New York, NY, pp. 51–76.
- Ögren, P., Fiorelli, E. and Leonard, N.E. (2004). Cooperative control of mobile sensor networks: Adaptive gradient climbing in a distributed environment, *IEEE Transactions on Automatic Control* **49**(8): 1292–1302.
- Omatu, S. and Seinfeld, J.H. (1989). *Distributed Parameter Systems: Theory and Applications*, Oxford Mathematical Monographs, Oxford University Press, New York, NY.
- 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. and Uciński, D. (2008). Configuring a sensor network for fault detection in distributed parameter systems, *International Journal of Applied Mathematics and Computer Science* **18**(4): 513–524, DOI: 10.2478/v10006-008-0045-4.
- Patriksson, M. (2001). Simplicial decomposition algorithms, in C.A. Floudas and P.M. Pardalos (Eds.), *Encyclopedia of Optimization*, Vol. 5, Kluwer Academic Publishers, Dordrecht, pp. 205–212.
- 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, *International Journal of Applied Mathematics and Computer Science* **9** (3), Special issue.
- Pázman, A. (1986). *Foundations of Optimum Experimental Design*, Mathematics and Its Applications, D. Reidel Publishing Company, Dordrecht.
- Point, N., Vande Wouwer, A. and Remy, M. (1996). Practical issues in distributed parameter estimation: Gradient computation and optimal experiment design, *Control Engineering Practice* **4**(11): 1553–1562.
- Porat, B. and Nehorai, A. (1996). Localizing vapor-emitting sources by moving sensors, *IEEE Transactions on Signal Processing* **44**(4): 1018–1021.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B. P. (2007). *Numerical Recipes. The Art of Scientific Computing*, 3rd Edn., Cambridge University Press, Cambridge.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*, Probability and Mathematical Statistics, John Wiley & Sons, New York, NY.
- 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.
- Rafajłowicz, E. (1986). Optimum choice of moving sensor trajectories for distributed parameter system identification, *International Journal of Control* **43**(5): 1441–1451.
- Rao, M.M. (1987). *Measure Theory and Integration*, John Wiley & Sons, New York, NY.
- Sastry, S. and Iyengar, S.S. (2005). Real-time sensor-actuator networks, *International Journal of Distributed Sensor Networks* **1**: 17–34.
- Sinopoli, B., Sharp, C., Schenato, L., Schaffert, S. and Sastry, S.S. (2003). Distributed control applications within sensor networks, *Proceedings of the IEEE* **91**(8): 1235–1246.
- Song, Z., Chen, Y., Sastry, C.R. and Tas, N.C. (2009). *Optimal Observation for Cyber-physical Systems: A Fisher-information-matrix-based Approach*, Springer-Verlag, London.

- Sun, N.-Z. (1994). *Inverse Problems in Groundwater Modeling, Theory and Applications of Transport in Porous Media*, Kluwer Academic Publishers, Dordrecht.
- Titterton, D.M. (1980). Aspects of optimal design in dynamic systems, *Technometrics* **22**(3): 287–299.
- Uciński, D. (1999). *Measurement Optimization for Parameter Estimation in Distributed Systems*, Technical University Press, Zielona Góra.
- Uciński, D. (2000). Optimal sensor location for parameter estimation of distributed processes, *International Journal of Control* **73**(13): 1235–1248.
- Uciński, D. (2005). *Optimal Measurement Methods for Distributed-Parameter System Identification*, CRC Press, Boca Raton, FL.
- Uciński, D. (2010). Sensor network design for spatio-temporal prediction of distributed parameter systems, in M. Kuczma and K. Wilmski (Eds.), *Computer Methods in Mechanics: Lectures of the CMM 2009*, Springer-Verlag, Berlin, pp. 193–210.
- Uciński, D. and Atkinson, A. C. (2004). Experimental design for time-dependent models with correlated observations, *Studies in Nonlinear Dynamics & Econometrics* **8**(2), Article No. 13.
- Uciński, D. and Bogacka, B. (2005). T-optimum designs for discrimination between two multivariate dynamic models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**: 3–18.
- Uciński, D. and Chen, Y. (2005). Time-optimal path planning of moving sensors for parameter estimation of distributed systems, *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005, Seville, Spain*, (on CD-ROM).
- 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.
- Uciński, D. and Patan, M. (2010). Sensor network design for the estimation of spatially distributed processes, *International Journal of Applied Mathematics and Computer Science* **20**(3): 459–481, DOI: 10.2478/v10006-010-0034-2.
- Uspenskii, A.B. and Fedorov, V.V. (1975). *Computational Aspects of the Least-Squares Method in the Analysis and Design of Regression Experiments*, Moscow University Press, Moscow, (in Russian).
- van de Wal, M. and de Jager, B. (2001). A review of methods for input/output selection, *Automatica* **37**: 487–510.
- Vande Wouwer, A., Point, N., Porteman, S. and Remy, M. (1999). On a practical criterion for optimal sensor configuration—Application to a fixed-bed reactor, *Proceedings of the 14th IFAC World Congress, Beijing, China*, Vol. I: Modeling, Identification, Signal Processing II, Adaptive Control, pp. 37–42.
- von Hohenbalken, B. (1977). Simplicial decomposition in nonlinear programming algorithms, *Mathematical Programming* **13**: 49–68.
- Walter, É. and Pronzato, L. (1997). *Identification of Parametric Models from Experimental Data*, Communications and Control Engineering, Springer-Verlag, Berlin.
- Zhao, F. and Guibas, L. J. (2004). *Wireless Sensor Networks: An Information Processing Approach*, Morgan Kaufmann Publishers, Amsterdam.
- Zhao, T. and Nehorai, A. (2006). Detecting and estimating biochemical dispersion of a moving source in a semi-infinite medium, *IEEE Transactions on Signal Processing* **54**(6): 2213–2225.
- Zięba, T. and Uciński, D. (2008). Mobile sensor routing for parameter estimation of distributed systems using the parallel tunneling method, *International Journal of Applied Mathematics and Computer Science* **18**(3): 307–318, DOI: 10.2478/v10006-008-0028-5.



Dariusz Uciński was born in 1965. He received the M.Sc. degree in electrical engineering from the Higher College of Engineering in Zielona Góra, Poland, in 1989, and the Ph.D. and D.Sc. degrees in automatic control and robotics from the Technical University of Wrocław, Poland, in 1992 and 2000, respectively. In 2007 he was conferred the title of state professor, the highest scientific degree in Poland. He is currently a professor at the University of Zielona Góra, Poland. His research interests are in the area of measurement optimization for distributed parameter systems. He authored a book entitled *Optimal Measurement Methods for Distributed Parameter System Identification* (CRC Press, 2005). Other areas of his expertise include optimum experimental design, algorithmic optimal control, robotics and parallel computing. Since 1992 he has been serving as the deputy editor of the *International Journal of Applied Mathematics and Computer Science*. Since 2008 he has been the chair of the Control Systems Society Chapter of the IEEE Poland Section.

Appendix

Maximization of a concave function over the intersection of a canonical simplex and a box

Given a vector $\mathbf{b} \in \mathbb{R}_{++}^n$ such that $\mathbf{1}^\top \mathbf{b} \geq 1$, consider the problem

$$\begin{aligned} & \text{maximize} && f(\mathbf{p}) \\ & \text{subject to} && \mathbf{p} \in P, \end{aligned} \quad (71)$$

where we assume throughout that

- $P = S_n \cap B$ for S_n being the probability simplex in \mathbb{R}^n and $B = \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{0} \preceq \mathbf{p} \preceq \mathbf{b}\}$, and
- the function $f : S_n \rightarrow \mathbb{R}$ is concave and continuously differentiable over P .

We have the following.

Lemma 1. *A vector \mathbf{p}^* constitutes a global maximum of the constrained problem (71) if, and only if, there exists a*

number λ^* such that

$$\frac{\partial f(\mathbf{p}^*)}{\partial p_i} \begin{cases} \geq \lambda^* & \text{if } p_i^* = b_i, \\ = \lambda^* & \text{if } 0 < p_i^* < b_i, \\ \leq \lambda^* & \text{if } p_i^* = 0 \end{cases} \quad (72)$$

for $i = 1, \dots, n$.

This characterization can be interpreted as separability of the components of \mathbf{p}^* in terms of the gradient $\nabla f(\mathbf{p}^*)$. Specifically, the values of $\partial f(\mathbf{p}^*)/\partial p_i$ for the components p_i^* of \mathbf{p}^* which are interior points of the closed intervals $[0, b_i]$ must equal some constant λ^* , whereas for the components p_i^* coinciding with endpoints 0 or b_i , the values of $\partial f(\mathbf{p}^*)/\partial p_i$ must be no larger and no smaller than λ^* , respectively.

Proof. The problem (71) can be reformulated as that of finding $\mathbf{p}^* \in \mathbb{R}^n$ to minimize

$$\tilde{f}(\mathbf{p}) = -f(\mathbf{p}) \quad (73)$$

subject to the constraints

$$\mathbf{1}^\top \mathbf{p} - 1 = 0, \quad (74)$$

$$-\mathbf{p} \preceq \mathbf{0}, \quad (75)$$

$$\mathbf{p} - \mathbf{b} \preceq \mathbf{0}. \quad (76)$$

Introducing the dual variables $\lambda \in \mathbb{R}$, $\boldsymbol{\mu} \in \mathbb{R}_+^n$ and $\boldsymbol{\nu} \in \mathbb{R}_+^n$, we define the Lagrangian of (73)–(76) as

$$\mathcal{L}(\mathbf{p}, \lambda, \boldsymbol{\mu}, \boldsymbol{\nu}) = -f(\mathbf{p}) + \lambda(\mathbf{1}^\top \mathbf{p} - 1) - \boldsymbol{\mu}^\top \mathbf{p} + \boldsymbol{\nu}^\top (\mathbf{p} - \mathbf{b}). \quad (77)$$

The first-order Karush–Kuhn–Tucker (KKT) conditions for our problem are

$$-\nabla f(\mathbf{p}) + \lambda \mathbf{1} - \boldsymbol{\mu} + \boldsymbol{\nu} = \mathbf{0}, \quad (78)$$

$$\boldsymbol{\mu} \cdot \mathbf{p} = 0, \quad (79)$$

$$\boldsymbol{\nu} \cdot (\mathbf{p} - \mathbf{b}) = 0, \quad (80)$$

$$\boldsymbol{\mu} \succeq \mathbf{0}, \quad (81)$$

$$\boldsymbol{\nu} \succeq \mathbf{0}, \quad (82)$$

$$\mathbf{0} \preceq \mathbf{p} \preceq \mathbf{b}, \quad (83)$$

$$\mathbf{1}^\top \mathbf{p} = 1. \quad (84)$$

The independence of $2n+1$ linear constraints defined by (74)–(76) implies that (78)–(84) are necessary for optimality. But they are also sufficient since the performance measure (73) is convex. Consequently, the optimality of \mathbf{p}^* amounts to the existence of some values of λ , $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, denoted by λ^* , $\boldsymbol{\mu}^*$ and $\boldsymbol{\nu}^*$, respectively, such that (78)–(84) are satisfied.

Suppose that $p_i^* = b_i$ for some index i . Then from (79) it follows that $\mu_i^* = 0$ and, therefore, (78) reduces to

$$\frac{\partial f(\mathbf{p}^*)}{\partial p_i} = \lambda^* + \nu_i^* \geq \lambda^*, \quad (85)$$

the last inequality owing to (82). In turn, on account of (80), the assumption $p_i^* = 0$ yields $\nu_i^* = 0$, and then (78) simplifies to

$$\frac{\partial f(\mathbf{p}^*)}{\partial p_i} = \lambda^* - \mu_i^* \leq \lambda^*, \quad (86)$$

which is owing to (81). Finally, by (79) and (80), the assumption $0 < p_i^* < b_i$ clearly forces $\mu_i^* = \nu_i^* = 0$, for which (78) gives

$$\frac{\partial f(\mathbf{p}^*)}{\partial p_i} = \lambda^*, \quad (87)$$

Conversely, having found $\mathbf{p}^* \in P$ and $\lambda^* \in \mathbb{R}$ for which (72) is fulfilled, we can define

$$\mu_i^* = \max \left(\lambda^* - \frac{\partial f(\mathbf{p}^*)}{\partial p_i}, 0 \right), \quad (88)$$

$$\nu_i^* = \max \left(\frac{\partial f(\mathbf{p}^*)}{\partial p_i} - \lambda^*, 0 \right), \quad (89)$$

for $i = 1, \dots, n$, which guarantees the satisfaction of (78)–(84). This means that \mathbf{p}^* is a KKT point and this is equivalent to its global optimality. ■

Received: 27 January 2011

Revised: 12 September 2011