J Glob Optim (2007) 39:291–322
DOI 10.1007/s10898-007-9139-z

ORIGINAL ARTICLE

D-optimal design of a monitoring network for parameter estimation of distributed systems

Dariusz Uciński · Maciej Patan

Received: 25 May 2006 / Accepted: 12 January 2007 / Published online: 21 February 2007

Abstract This paper addresses the design of a network of observation locations in a spatial domain that will be used to estimate unknown parameters of a distributed parameter system. We consider a setting where we are given a finite number of possible sites at which to locate a sensor, but cost constraints allow only some proper subset of them to be selected. We formulate this problem as the selection of the gauged sites so as to maximize the log-determinant of the Fisher information matrix associated with the estimated parameters. The search for the optimal solution is performed using the branch-and-bound method in which an extremely simple and efficient technique is employed to produce an upper bound to the maximum objective function. Its idea consists in solving a relaxed problem through the application of a simplicial decomposition algorithm in which the restricted master problem is solved using a multiplicative algorithm for optimal design. The use of the proposed approach is illustrated by a numerical example involving sensor selection for a two-dimensional convective diffusion process.

Keywords Sensor location · Distributed parameter systems · Optimum experimental design · Branch-and-bound · Simplicial decomposition

1 Introduction and notation

1.1 Sensor location for parameter estimation

One of the crucial design issues in parameter estimation of systems governed by partial differential equations (PDEs), commonly termed 'Distributed Parameter Systems'

D. Uciński (✉) · M. Patan
Institute of Control and Computation Engineering, University of Zielona Góra, ul. Podgórna 50, 65–246 Zielona Góra, Poland
e-mail: d.ucinski@issi.uz.zgora.pl

M. Patan
m.patan@issi.uz.zgora.pl

(DPSs), is the problem of where to locate the measurement sensors. The importance 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; 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 requires 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 possibly expensive sensors. Over the past years, this limitation has stimulated laborious research on the development of strategies for efficient sensor placement [for reviews, see papers (Kubrusly and Malebranche 1985; van de Wal and de Jager 2001) and comprehensive monographs (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 candidates and the combinatorial nature of the design problem is taken into account very occasionally (van de Wal and de Jager 2001). Needless to say that this approach, which is feasible for a relatively small number of possible locations, soon becomes useless as the number of possible location candidates increases.

1.2 Previous work

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; Titterington 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. 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 in (Rafajłowicz 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 (Uciński 2005). The potential of the approach for generalizations was exploited, e.g., by Patan and Patan (2005) who 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. The approach based on maximization of the determinant of the appropriate FIM is by no means restricted to theoretical considerations and there are examples which do confirm its effectiveness in practical applications. Thus, in (Munack 1984) a given number of stationary sensors were optimally located using 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 modeling. 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 in (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 does not lead to a substantial deterioration in the accuracy of modal-response estimates, which is quantified by the determinant of the FIM. 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 (Vande Wouwer et al. 1999).

At this juncture, 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 behavior 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 instant 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). In the seminal article (Rafajłowicz 1986), the D-optimality criterion is considered and an optimal time-dependent measure is sought, rather than the trajectories themselves. On the other hand, Uciński (2000; 2005; Uciński and Korbicz 2001), apart from generalizations of Rafajłowicz's results, develops some computational algorithms based on the FIM. He reduces 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. In turn, the work (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.

1.3 Our results

The aim of the research reported here was to develop a practical approach to sensor selection which, while being independent of a particular model of the dynamic DPS in question, would be versatile enough to cope with practical monitoring networks consisting of many stationary sensors. Specifically, we consider N possible sites at which

to locate a sensor, but limitations on the number of sensors at our disposal allow only n of them (typically, n is much smaller than N) to be selected. Consequently, the problem is to divide the N available sites between n gauged sites and the remaining N–n ungauged sites so as to maximize the determinant of the FIM associated with the parameters to be estimated. Since selecting the best subset of sites to locate the sensors constitutes an inherently discrete large-scale resource allocation problem whose solution may be prohibitively time-consuming, an efficient guided search algorithm based on the branch-and-bound (BB) method is developed, which implicitly enumerates all the feasible sensor configurations, using relaxed optimization problems that involve no integer constraints.

Obviously, this idea is not novel, since BB constitutes one of the most frequent approaches to solve discrete optimization problems and it has indeed been used in the context of network design, cf., e.g., (Boer et al. 2001). The main contribution of this paper consists in the development of a simple, yet powerful, computational scheme to obtain upper bounds to the optimal values of the D-optimality criterion for the restricted problems. These bounds are obtained by relaxing the 0–1 constraints on the design variables, thereby allowing them to take any value in the interval $[0, 1]$ and resulting in a concave problem of determinant maximization over the set of all linear combinations of a finite number of nonnegative definite matrices, subject to additional linear constraints on the coefficients of those combinations. In order to solve it numerically, optimality conditions are first derived and discussed, because they take a surprisingly simple form involving an interesting separability principle for the set of locations at which the weights achieve their upper bounds and the ones at which the weights are zero. Then an original algorithm is proposed which can be interpreted as a simplicial decomposition one with the restricted master problem solved by an uncomplicated multiplicative weight optimization algorithm which is adapted here to take account of constituent matrices possibly having ranks greater than one, as opposed to the standard assumption made in the optimum experimental design literature. The resulting procedure is guaranteed to produce iterates converging to the solution of the relaxed restricted problem. To the best of our knowledge, the proposed combination of the simplicial decomposition algorithm with the known multiplicative algorithm for constructing optimizing probability distributions has not been investigated yet. To illustrate the use of our algorithm, we include implementation details and some numerical experience on a sensor network design problem regarding a two-dimensional convective diffusion process.

The paper is organized as follows. Section 2 states formally the sensor network design problem as a discrete resource allocation problem. The BB algorithm for its solution is discussed in Sect. 3. Section 4 develops the simplicial decomposition algorithm for computing upper bounds required by the branching rule. In Sect. 5, we report the numerical results obtained by applying the algorithm described in Sect. 3 on the optimal design application. We conclude in Sect. 6 with some comments on related open problems. The proofs of the presented theoretical results are contained in the Appendices.

1.4 Conventions and notation

Our notation is more or less standard. Given a set H, $|H|$, and $\bar{H}$, signify its cardinality and closure, respectively. We use $\mathbb{R}$ to denote the set of real numbers and $\mathbb{R}_+$ to denote the set of nonnegative real numbers. The n-dimensional Euclidean vector

space is denoted by $\mathbb{R}^n$, and the Euclidean matrix space of real matrices with n rows and k columns is denoted by $\mathbb{R}^{m\times k}$. We will write $\mathbb{S}^n$ for the subspace of $\mathbb{R}^{n\times n}$ consisting of all symmetric matrices. The identity matrix of order n is denoted by E_n. In $\mathbb{S}^n$ two sets are of special importance: the cone of nonnegative definite matrices and the cone of positive definite matrices, denoted by $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$, respectively. The curly inequality symbol $\succeq$ and its strict form $\succ$ are used to denote the Loewner partial ordering of symmetric matrices: For $A, B \in \mathbb{S}^n$, we have

$$A \succeq B \Longleftrightarrow A - B \in \mathbb{S}^n_+,$$
$$A \succ B \Longleftrightarrow A - B \in \mathbb{S}^n_{++}.$$

We call a point of the form $\alpha_1 u_1 + \cdots + \alpha_\ell u_\ell$, where $\alpha_1 + \cdots + \alpha_\ell = 1$ and $\alpha_i \geq 0$, $i = 1, \ldots, \ell$, a convex combination of the points $u_1, \ldots, u_\ell$ (it can be thought of as a mixture or a weighted average of the points, with α_i the fraction of u_i in the mixture). Given a set of points U, $\mathrm{co}(U)$ stands for its convex hull, i.e., the set of all convex combinations of elements of U,

$$\mathrm{co}(U) = \left\{ \sum_{i=1}^{\ell} \alpha_i u_i \,\Big|\, u_i \in U,\ \alpha_i \geq 0,\ i = 1, \ldots, \ell;\ \sum_{i=1}^{\ell} \alpha_i = 1,\ \ell = 1, 2, 3, \ldots \right\}.$$

The probability (or canonical) simplex in $\mathbb{R}^n$ is defined as

$$P_n = \mathrm{co}\big(\{e_1, \ldots, e_n\}\big) = \left\{ p \in \mathbb{R}^n_+ \,\Big|\, \sum_{i=1}^{n} p_i = 1 \right\},$$

where e_j is the usual unit vector along the j-th coordinate of $\mathbb{R}^n$.

Finally, recall that for any $A \in \mathbb{R}^{n\times n}$ which may depend on a parameter β, there holds

$$\frac{\partial}{\partial \beta} \ln \det(A) = \mathrm{trace}\left(A^{-1} \frac{\partial A}{\partial \beta} \right)$$

whenever A is nonsingular.

2 Sensor location problem

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

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

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

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

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

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

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

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

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

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

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

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

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

where $y(\cdot\,,\cdot\,;\vartheta)$ denotes the solution to (1)–(3) for a given value of the parameter vector ϑ. In practice, a regularized version of the above problem is often considered by adding to $\mathcal{J}(\vartheta)$ a term imposing stability or a-priori information or both (Banks and Kunisch 1989; Vogel 2002).

Inevitably, the covariance matrix $\mathrm{cov}(\widehat{\theta})$ of the above least-squares estimator depends on the sensor locations x^{ℓ}. This fact suggests that we may attempt to select them so as to yield best estimates of the system parameters. To form a basis for the comparison of different locations, a quantitative measure of the 'goodness' of particular sensor configurations is required. Such a measure is customarily based on the concept of the FIM which is widely used in optimum experimental design theory for

lumped systems (Fedorov and Hackl 1997; Pázman 1986; Pukelsheim 1993; Walter and Pronzato 1997; Atkinson and Donev 1992). In our setting, the FIM is given by (Quereshi et al. 1980)

$$M(x^1,\ldots,x^n)=\sum_{\ell=1}^{n}\int_T g(x^\ell,t)g^{\mathsf{T}}(x^\ell,t)\,\mathrm{d}t, \tag{7}$$

where

$$g(x,t)=\left[\frac{\partial y(x,t;\vartheta)}{\partial\vartheta_1},\ldots,\frac{\partial y(x,t;\vartheta)}{\partial\vartheta_m}\right]^{\mathsf{T}}_{\vartheta=\theta^0} \tag{8}$$

stands for the so-called *sensitivity vector*, θ^0 being a prior estimate to the unknown parameter vector θ (Uciński 2005; Sun 1994; Rafajłowicz 1981; 1983). The rationale behind this choice is the fact that, up to a constant scalar multiplier, the inverse of the FIM constitutes a good approximation of $\mathrm{cov}(\widehat{\theta})$ provided that the time horizon is large, the nonlinearity of the model with respect to its parameters is mild, and the measurement errors are independently distributed and have small magnitudes (Walter and Pronzato 1997; Fedorov and Hackl 1997).

As for a specific form of Ψ, various options exist (Walter and Pronzato 1997; Fedorov and Hackl 1997; Atkinson and Donev 1992), but the most popular criterion, called the D-optimality criterion, is the log-determinant of the FIM:

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

The resulting D-optimum sensor configuration leads to the minimum volume of the uncertainty ellipsoid for the estimates.

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

$$\mathcal{R}(x^1,\ldots,x^n):=\Psi\big[M(x^1,\ldots,x^n)\big] \tag{10}$$

with respect to x^ℓ, $\ell=1,\ldots,n$ belonging to the admissible set X. This apparently simple formulation may lead to the conclusion that the only question remaining is that of selecting an appropriate solver from a library of numerical optimization routines. Unfortunately, an in-depth analysis reveals complications which accompany this way of thinking.

A key difficulty in developing successful numerical techniques for sensor location is that the number of sensors to be placed in a given region may be quite large. For example, in the research carried out to find spatial predictions for ozone in the Great Lakes of US, measurements made by approximately 160 monitoring stations were used (Nychka and Saltzman 1998). When trying to treat the task as a constrained nonlinear programming problem, the actual number of variables is even doubled, since the position of each sensor is determined by its two spatial coordinates, so that the resulting problem is rather of large scale. What is more, a desired global extremum is usually hidden among many poorer local extrema. Consequently, to directly find a numerical solution may be extremely difficult. Additionally, a technical complication might also be the sensor clusterization which constitutes a price to pay for the simplifying assumption that the measurement noise is spatially uncorrelated. This means that in an optimal solution different sensors often tend to take measurements at one point, and this is acceptable in applications rather occasionally.

In the literature, a common remedy for the last predicament is to guess a priori a set of N possible candidate locations, where $N > n$, and then to seek the best subset of n locations from among the N possible, so that the problem is then reduced to a combinatorial one. In other words, the problem is to divide the N available sites between n gauged sites and the remaining N–n ungauged sites so as to maximize the determinant of the FIM associated with the parameters to be estimated. This formulation will be also adopted here.

Specifically, let x^i, $i = 1, \ldots, N$ denote the positions of sites where sensors can potentially be placed. Now that our design criterion has been established, the problem is to find an optimal allocation of n available sensors to x^i, $i = 1, \ldots, N$ so as to maximize the value of the design criterion incurred by the allocation. In order to formulate this mathematically, introduce for each possible location x^i a variable v_i which takes the value 1 or 0 depending on whether a sensor is or is not located at x^i, respectively. The FIM in (7) can then be rewritten as

$$M(v_1, \ldots, v_N) = \sum_{i=1}^{N} v_i M_i, \tag{11}$$

where

$$M_i = \int_T g(x^i, t) g^{\mathsf{T}}(x^i, t)\, \mathrm{d}t. \tag{12}$$

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

Then our design problem takes the form:

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

$$\mathcal{P}(v) = \log \det\big(M(v)\big) \tag{13}$$

subject to the constraints

$$\sum_{i=1}^{N} v_i = n, \tag{14}$$

$$v_i = 0 \text{ or } 1, \quad i = 1, \ldots, N. \tag{15}$$

This constitutes a 0–1 integer programming problem which necessitates an ingenious solution. In what follows, we propose to solve it using the BB method which is a standard technique to solve integer-programming problems.

3 Solution by branch-and-bound

3.1 General outline

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

Principally, the BB method is a tree-search algorithm combined with a rule for pruning subtrees. Suppose we wish to maximize an objective function $\mathcal{P}(v)$ over a

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

In what follows, we will use the symbol I to denote the index set $\{1, \ldots, N\}$ of possible sensor locations. Our implementation of BB for Problem P involves the partition of the feasible set

$$V = \left\{ (v_1, \ldots, v_N) \;\middle|\; \sum_{i=1}^{N} v_i = n,\ v_i = 0 \text{ or } 1,\ \forall i \in I \right\}, \tag{16}$$

into subsets. It is customary to select subsets of the form (Bertsekas 1999):

$$V(I_0, I_1) = \{ v \in V \mid v_i = 0,\ \forall i \in I_0,\ v_i = 1,\ \forall i \in I_1 \}, \tag{17}$$

where I_0 and I_1 are disjoint subsets of I. Consequently, $V(I_0, I_1)$ is the subset of V such that a sensor is placed at the locations with indices in I_1, no sensor is placed at the locations with indices in I_0, and a sensor may or may not be placed at the remaining locations.

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

$$\bar{\mathcal{P}}(I_0, I_1) \geq \max_{v \in V(I_0, I_1)} \mathcal{P}(v), \tag{18}$$

and a feasible solution $\underline{v} \in V$ for which $\mathcal{P}(\underline{v})$ can serve as a lower bound to the maximum design criterion over V. We may compute $\bar{\mathcal{P}}(I_0, I_1)$ by solving the following relaxed problem:

Problem R(I_0, I_1) Find the sequence $\bar{v}$ to maximize (13) subject to the constraints

$$\sum_{i=1}^{N} v_i = n, \tag{19}$$

$$v_i = 0, \quad i \in I_0, \tag{20}$$

$$v_i = 1, \quad i \in I_1, \tag{21}$$

$$0 \le v_i \le 1, \quad i \in I \setminus (I_0 \cup I_1). \tag{22}$$

In Problem R(I_0, I_1) all 0–1 constraints on the variables v_i are relaxed by allowing them to take any value in the interval $[0, 1]$, except that the variables $v_i, i \in I_0 \cup I_1$ are fixed at either 0 or 1. A simple and efficient method for its solution is given in Sect. 4. As a result of its application, we set $\bar{\mathcal{P}}(I_0, I_1) = \mathcal{P}(\bar{v})$.

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

3.2 Branching rule and ultimate algorithm

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

(i) Determine

$$i_\star = \arg \min_{i \in I \setminus (I_0 \cup I_1)} |v_i - 0.5|. \tag{23}$$

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

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

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

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

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

- SINGULARITY-TEST(I_0, I_1) returns true only if expansion of the current node will result in a singular FIM, see Sect. 4.2 for details.
- RELAXED-SOLUTION(I_0, I_1) returns a solution to Problem R(I_0, I_1).
- DET-FIM(v) returns the log-determinant of the FIM corresponding to v.
- INTEGRAL-TEST(v) returns true only if the current solution v is integral.
- INDEX-BRANCH(v) returns the index defined by (23).

Algorithm 1 A recursive version of the depth-first BB method. It uses two global variables, *LOWER* and *v_best*, which are respectively the maximal value of the FIM determinant over feasible solutions found so far and the solution at which it is attained.

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

4 Optimality conditions and a simplicial decomposition algorithm for solving the relaxed problem

4.1 Separability form of optimality conditions

The nonleaf nodes of the BB tree are processed by relaxing the original combinatorial problem, which directly leads to Problem R(I_0, I_1). This section provides a detailed exposition of a simplicial decomposition method which is particularly suited for its solution.

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

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

$$\mathcal{Q}(w) = \log\det\big(G(w)\big) \tag{24}$$

subject to the constraints

$$\sum_{j=1}^{q} w_j = r, \tag{25}$$

$$0 \le w_j \le 1, \quad j = 1, \ldots, q, \tag{26}$$

where

$$r = n - |I_1|, \quad G(w) = A + \sum_{j=1}^{q} w_j S_j, \quad A = \sum_{i \in I_1} M_i, \quad S_j = M_{\pi(j)}, \quad j = 1, \ldots, q. \tag{27}$$

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

In the sequel, W will stand for the set of all vectors $w = (w_1, \ldots, w_q)$ satisfying (25) and (26). Note that it forms a polygon in $\mathbb{R}^q$. Recall that the log-determinant is concave and strictly concave over the cones $\mathbb{S}^m_+$ and $\mathbb{S}^m_{++}$, respectively, cf. (Pukelsheim 1993; Boyd and Vandenberghe 2004). Thus, the objective function (24) is concave as the composition of the log-determinant with an affine mapping, see (Boyd and Vandenberghe 2004, p. 79). We wish to maximize it over the polyhedral set W. If the FIM corresponding to an optimal solution $w^\star$ is nonsingular, then an intriguing form of the optimality conditions can be derived.

Proposition 1 *Suppose that the matrix $G(w^\star)$ is nonsingular for some $w^\star \in W$. The vector $w^\star$ constitutes a global solution to Problem R′(I_0, I_1) if, and only if, there exists a number $\lambda^\star$ such that*

$$\varphi(j, w^\star) \begin{cases} \ge \lambda^\star & \text{if } w^\star_j = 1, \\ = \lambda^\star & \text{if } 0 < w^\star_j < 1, \\ \le \lambda^\star & \text{if } w^\star_j = 0, \end{cases} \tag{28}$$

where

$$\varphi(j, w) = \operatorname{trace}\big[G^{-1}(w) S_j\big], \quad j = 1, \ldots, q. \tag{29}$$

Remark 1 The above result can be basically inferred by fitting Problem R′(I_0, I_1) into the abstract setting of Theorem 1 of (Pronzato 2004) dealing with a very general case of directly constrained probability measures. Here, however, we have obtained it in a quite elementary manner, taking account of the specific form of our problem.

Proposition 1 reveals one characteristic feature of the optimal solutions, namely that, when identifying them, the function φ turns out to be crucial and optimality means separability of the components of $w^\star$ in terms of the values of this function. Specifically, the values of $\varphi(\cdot, w^\star)$ for the indices corresponding to the fractional components of $w^\star$ must be equal to some constant $\lambda^\star$, whereas for the components taking the value 0 or the value 1 the values of $\varphi(\cdot, w^\star)$ must be no larger and no smaller than $\lambda^\star$, respectively.

4.2 Discarding singular information matrices

Note that the assumption that $G(w)$ is nonsingular can be dropped, since there is a very simple method to check whether or not the current relaxed problem will lead to a FIM which is nonsingular.

Proposition 2 *The FIM corresponding to the solution to Problem* $R'(I_0, I_1)$ *is singular if and only if so is* $G(\bar{w})$*, where*

$$\bar{w} = \underbrace{(r/q, \ldots, r/q)}_{q \text{ times}}. \tag{30}$$

Consequently, a test of the singularity of the matrix $G(\bar{w}) = A + \frac{r}{q}\sum_{j=1}^{q} S_j$ can be built into the BB procedure in order to drop the corresponding node from further considerations and forego the examination of its descendants. Otherwise, the vector $(r/q, \ldots, r/q)$ may serve as a good starting point for the simplicial decomposition algorithm outlined in what follows.

Remark 2 A solution to Problem $R'(I_0, I_1)$ is not necessarily unique. Note, however, that for nonsingular cases (after all, pruning discards such cases from further consideration), the resulting FIM is unique. Indeed, Problem $R'(I_0, I_1)$ can equivalently be viewed as maximization of the log-determinant over the convex and compact set of matrices $\mathfrak{M} = \{G(w) \mid \sum_{j=1}^{q} w_j = r,\ 0 \le w_j \le 1,\ i = 1, \ldots, q\}$. But the log-determinant is strictly concave over the cone of positive-definite matrices, $\mathbb{S}^m_{++}$, which constitutes the interior of $\mathbb{S}^m_+$ relative to $\mathbb{S}^m$, and this fact implies the unicity of the optimal FIM.

4.3 Simplicial decomposition scheme

Simplicial Decomposition (SD) constitutes an important class of methods for solving large-scale continuous problems in mathematical programming with convex feasible sets (von Hohenbalken 1977; Patriksson 2001; Bertsekas 1999). In the original framework, where a concave objective function is to be maximized over a bounded polyhedron, it iterates by alternately solving a linear programming subproblem (the so-called *column generation problem*) which generates an extreme point of the polyhedron, and a nonlinear *restricted master problem* (RMP) which finds the maximum of the objective function over the convex hull (a simplex) of previously defined extreme points. This basic strategy of simplicial decomposition has appeared in numerous references (Ventura and Hearn 1993; Hearn et al. 1985; 1987), where possible improvements and extensions have also been discussed. A principal characteristic of an SD method is that the sequence of successive solutions to the master problem tends to a solution to the original problem in such a way that the objective function strictly monotonically approaches its optimal value.

Problem $R'(I_0, I_1)$ is perfectly suited for the application of the SD scheme. In this case, it boils down to Algorithm 2. Here $\nabla\mathcal{Q}(w)$ signifies the gradient of $\mathcal{Q}$ at w, and it is easy to check that

$$\nabla\mathcal{Q}(w) = \left[\operatorname{trace}\left(G^{-1}(w)S_1\right), \ldots, \operatorname{trace}\left(G^{-1}(w)S_q\right)\right]^{\mathsf{T}}. \tag{31}$$

Algorithm 2 Algorithm model for simplicial decomposition.

Step 0 (Initialization)
Set $w^{(0)} = (r/q, \ldots, r/q)$ and $Z^{(0)} = \{w^{(0)}\}$. Select $0 < \epsilon \ll 1$, a parameter used in the stopping rule, and set $k = 0$.

Step 1 (Solution of the column generation subproblem)
Determine

$$z = \arg \max_{w \in W} \nabla \mathcal{Q}(w^{(k)})^{\mathsf{T}}(w - w^{(k)}). \tag{32}$$

Step 2 (Termination check)
If $\nabla \mathcal{Q}(w^{(k)})^{\mathsf{T}}(z - w^{(k)}) \leq \epsilon$, then STOP and $w^{(k)}$ is optimal. Otherwise, set $Z^{(k+1)} = Z^{(k)} \cup \{z\}$.

Step 3 (Solution of the restricted master problem)
Find

$$w^{(k+1)} = \arg \max_{w \in \mathrm{co}(Z^{(k+1)})} \mathcal{Q}(w) \tag{33}$$

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

Since we deal with maximization of a concave function $\mathcal{Q}$ over a bounded polyhedral set W, the convergence of Algorithm 2 in a finite number of RMP steps is automatically guaranteed (von Hohenbalken 1977; Bertsekas 1999, p. 221). Observe that Step 3 implements the *column dropping rule* (Patriksson 2001), according to which any extreme point with zero weight in the expression of $w^{(k)}$ as a convex combination of elements in $Z^{(k)}$ is removed. This rule makes the number of elements in successive sets $Z^{(k)}$ reasonably low.

The SD algorithm may be viewed as a form of modular nonlinear programming, provided that one has an effective computer code for solving the restricted master problem, as well as access to a code which can take advantage of the linearity of the column generation subproblem (Hearn et al. 1987). The former issue will be addressed in the next subsection, where an extremely simple and efficient multiplicative algorithm for weight optimization will be discussed. In turn, the latter issue can be easily settled, as in the linear programming problem of Step 1 the feasible region W is defined by one equality contraint (25) and q bound constraints (26). At first sight, its form might suggest that selecting a vertex of W which corresponds to the greatest component of $\nabla \mathcal{Q}(w^{(k)})$ would yield the desired solution z. However, it is impossible to apply such a simple vertex-direction steepest-ascent technique since the nonnegative weight values w_j must sum up to r and, at the same time, they must not exceed one. Consequently, the solution cannot be focused only on one direction and z must be expressed as a linear combination of many vertices. Yet the simple form of the constraint set W can still be exploited directly, since numerous techniques have been proposed to achieve considerable speedup, ranging from improvements on the simplex method (cf. its upper-bounding version described by Pierre (1969, p. 224) to large-scale interior-point methods which are accessible in popular numerical packages (cf. the primal-dual interior-point variant of Mehrotra's predictor-corrector algorithm implemented in MATLAB's Optimization Toolbox (MathWorks 2000), cf. (Nocedal and Wright 1999).

4.4 Multiplicative algorithm for the restricted master problem

Suppose that in the $(k+1)$-th iteration of Algorithm 2, we have

$$Z^{(k+1)} = \{z_1, \ldots, z_s\}, \tag{34}$$

possibly with $s < k+1$ owing to the built-in deletion mechanism of points in $Z^{(i)}$, $1 \le i \le k$, which did not contribute to the convex combinations yielding the corresponding iterates $w^{(\ell)}$. Step 3 of Algorithm 2 involves maximization of the design criterion (24) over

$$\mathrm{co}(Z^{(k+1)}) = \left\{ \sum_{\ell=1}^{s} \alpha_\ell z_\ell \,\middle|\, \alpha_\ell \ge 0,\ \ell = 1, \ldots, s,\ \sum_{\ell=1}^{s} \alpha_\ell = 1 \right\}. \tag{35}$$

From the representation of any $w \in \mathrm{co}(Z^{(k+1)})$ as

$$w = \sum_{\ell=1}^{s} \alpha_\ell z_\ell, \tag{36}$$

or, in component-wise form,

$$w_j = \sum_{\ell=1}^{s} \alpha_\ell z_{\ell j}, \quad j = 1, \ldots, q, \tag{37}$$

$z_{\ell j}$ being the jth component of z_ℓ, it follows that

$$G(w) = A + \sum_{j=1}^{q} w_j S_j = \sum_{\ell=1}^{s} \alpha_\ell \Big(A + \sum_{j=1}^{q} z_{\ell j} S_j \Big) = \sum_{\ell=1}^{s} \alpha_\ell G(z_\ell). \tag{38}$$

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

Problem $\mathbf{P_{RMP}}$ Find the sequence of weights $\alpha = (\alpha_1, \ldots, \alpha_s)$ to maximize

$$\mathcal{T}(\alpha) = \log\det\big(H(\alpha)\big) \tag{39}$$

subject to the constraints

$$\sum_{\ell=1}^{s} \alpha_\ell = 1, \tag{40}$$

$$\alpha_\ell \ge 0, \quad \ell = 1, \ldots, s, \tag{41}$$

where

$$H(\alpha) = \sum_{\ell=1}^{s} \alpha_\ell Q_\ell, \qquad Q_\ell = G(z_\ell). \tag{42}$$

Basically, since the constraints (40) and (41) define the probability simplex P_s in $\mathbb{R}^s$, i.e., a very nice convex feasible domain, it is intuitively appealing to determine optimal weights using a numerical algorithm specialized for solving convex optimization problems. But another, much simpler technique can be employed to suitably guide weight calculation. It fully exploits the specific form of the objective function (39) by giving Problem $\mathrm{P_{RMP}}$ an equivalent probabilistic formulation. Specifically, the

nonnegativeness of the weights $z_{\ell j}$, $j = 1,\ldots,q$ and the nonnegative definiteness of the matrices A and S_j, $j = 1,\ldots,q$ imply that $Q_\ell \succeq 0$, $\ell = 1,\ldots,q$. Defining $\mathcal{X}$ as a discrete random variable which may take values in the set $\{1,\ldots,s\}$ and treating the weights α_ℓ, $\ell = 1,\ldots,s$ as the probabilities attached to its possible numerical values, i.e.,

$$p_{\mathcal{X}}(\ell) = \mathbb{P}(\mathcal{X} = \ell) = \alpha_\ell, \quad \ell = 1,\ldots,s, \tag{43}$$

we can interpret $p_{\mathcal{X}}$ as the probability mass function (pmf) of $\mathcal{X}$ and $H(\alpha) = \sum_{\ell=1}^{s} \alpha_\ell Q_\ell$ in (39) as the $\mathbb{P}$-weighted mean of the function $\mathcal{Q} : \ell \mapsto Q_\ell$. Therefore, Problem $\mathrm{P_{RMP}}$ can be thought of as that of finding a probability mass function maximizing the log-determinant of the mean of $\mathcal{Q}$. This formulation has captured close attention in optimum experimental design theory, where various characterizations of optimal solutions and efficient computational schemes have been proposed (Atkinson and Donev 1992; Fedorov and Hackl 1997; Walter and Pronzato 1997). In particular, in the case of the D-optimality criterion studied here, we can prove the following conditions for global optimality:

Proposition 3 *Suppose that the matrix $H(\alpha^\star)$ is nonsingular for some $\alpha^\star \in P_s$. The vector $\alpha^\star$ constitutes a global solution to Problem P_{RMP} if and only if*

$$\psi(\ell,\alpha^\star) \begin{cases} = m & \text{if } \alpha_\ell^\star > 0, \\ \le m & \text{if } \alpha_\ell^\star = 0 \end{cases} \tag{44}$$

for each $\ell = 1,\ldots,s$, where

$$\psi(\ell,\alpha) = \operatorname{trace}\left[H^{-1}(\alpha)Q_\ell\right], \quad \ell = 1,\ldots,s. \tag{45}$$

A very simple and numerically effective sequential procedure was devised and analyzed in (Pázman 1986; Torsney 1988; Silvey et al. 1978; Torsney and Mandal 2001; 2004) for the case of rank-one matrices Q_ℓ, which was then extended to the general case by Uciński (2005, p. 62). Its version adapted to the RMP proceeds as summarized in Algorithm 3. Clear advantages here are ease of implementation and negligible additional memory requirements.

The idea is reminiscent of the EM algorithm used for maximum likelihood estimation (Lange 1999). The properties of this computational scheme are considered in some detail in (Uciński 2005). Suffice it to say here that Algorithm 3 is globally convergent regardless of the choice of initial weights (they must only be all nonzero and correspond to a nonsingular FIM). Indeed, we have the following result (Uciński 2005 p. 65):

Proposition 4 *Assume that $\{\alpha^{(\kappa)}\}$ is a sequence of iterates constructed by Algorithm 3. Then the sequence $\{\mathcal{T}(\alpha^{(\kappa)})\}$ is monotone increasing and*

$$\lim_{\kappa\to\infty} \mathcal{T}(\alpha^{(\kappa)}) = \max_{\alpha\in P_s} \mathcal{T}(\alpha). \tag{48}$$

The basic scheme of Algorithm 3 can be refined to incorporate various improvements which make convergence much faster. For example, produced solutions often happen to contain many insignificant weights α_ℓ, which results from a limited accuracy of computations and the interruption of Algorithm 3 after a finite number of steps. In practice, these weights may well be disregarded since setting them as zeros and

Algorithm 3 Algorithm model for the restricted master problem.

Step 0 (Initialization)
Select weights $\alpha_\ell^{(0)} > 0$, $\ell = 1,\ldots,s$ which determine the initial pmf $p_{\mathcal{X}^{(0)}}$ for which we must have $\mathcal{T}(\alpha^{(0)}) > -\infty$, e.g., set $\alpha_\ell^{(0)} = 1/s$, $\ell = 1,\ldots,s$. Choose $0 < \eta \ll 1$, a parameter used in the stopping rule. Set $\kappa = 0$.

Step 1 (Termination check)
If

$$\frac{\psi(\ell,\alpha^{(\kappa)})}{m} < 1+\eta, \quad \ell = 1,\ldots,s, \tag{46}$$

then STOP.

Step 2 (Multiplicative update)
Evaluate

$$\alpha_\ell^{(\kappa+1)} = \alpha_\ell^{(\kappa)} \frac{\psi(\ell,\alpha^{(\kappa)})}{m}, \quad \ell = 1,\ldots,s. \tag{47}$$

Increment κ by one and go to Step 1.

distributing the sum of their values among the remaining weights [so as not to violate (40)] involves a negligible change in the value of the performance measure $\mathcal{T}(\alpha^{(\kappa)})$. The sum of the weights removed can be distributed among the other weights for which $\psi(\ell,\alpha^{(\kappa)}) > m$, and additionally, in a manner proportional to $\psi(\ell,\alpha^{(\kappa)}) - m$.

Another improvement is due to Pronzato (2003) who proposed a simple method to identify elements of $Z^{(k+1)}$ which do not contribute to the sought optimal convex combination in $\mathrm{co}(Z^{(k+1)})$. It can be generalized to the general case considered here and used during the search to discard such useless points 'on the fly,' thereby substantially reducing the problem dimensionality.

5 Computational results

The following example serves as a vehicle for the display of some salient features of the proposed approach. Consider simultaneous advection and diffusion of an air pollutant over an urban area normalized to the unit square $\Omega = (0,1)^2$. It is assumed that the mean wind velocity over Ω can be approximated by $\upsilon = (\upsilon_1, \upsilon_2)$, where

$$\upsilon_1 = \upsilon_2 = \frac{1}{2}(x_1 - x_2)^2 - 1, \tag{49}$$

i.e., the air moves with a constant speed along each straight line $x_2 = x_1 + b$, where $b \in [-1,1]$, cf. Fig. 1.

In addition, we take into account an active source of pollution and reaction, which leads to changes in the pollutant concentration $y = y(x,t)$. The evolution of y over the normalized observation interval $T = (0,1]$ is described by the following advection-diffusion equation:

$$\frac{\partial y(x,t)}{\partial t} + \nabla\cdot\big(\upsilon(x)y(x,t)\big) = \nabla\cdot\big(a(x)\nabla y(x,t)\big) + f(x) \qquad \text{in } \Omega\times T \tag{50}$$

subject to the boundary and initial conditions:

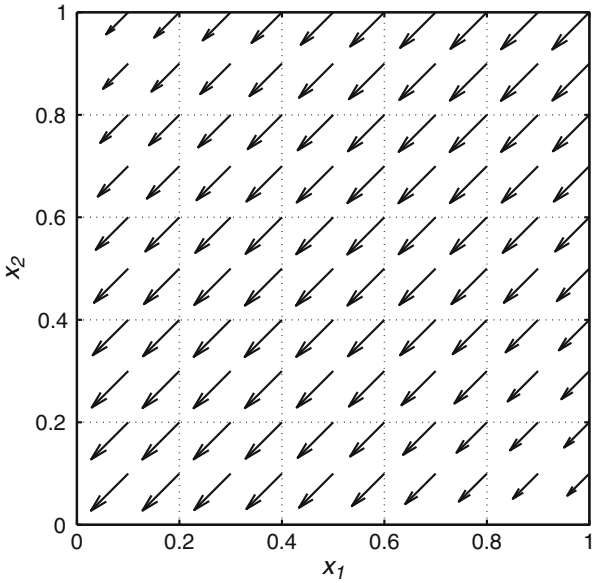

Fig. 1 Domain Ω and the wind velocity field $\upsilon = (\upsilon_1, \upsilon_2)$

$$\frac{\partial y(x,t)}{\partial n} = 0 \qquad \text{on } \Gamma \times T, \tag{51}$$

$$y(x,0) = 0 \qquad \text{in } \Omega, \tag{52}$$

where the term $f(x) = \exp\left(-50\|x - c\|^2\right)$ represents an active source of the pollutant located at point $c = (0.7, 0.7)$, and $\partial y/\partial n$ stands for the partial derivative of y with respect to the outward normal to the boundary Γ. The assumed functional form of the spatial-varying diffusion coefficient is

$$a(x) = \theta_1 + \theta_2 x_1^2 + \theta_3 x_2^2, \tag{53}$$

so that the constant parameters θ_1, θ_2, and θ_3 need estimation based on measurement data from monitoring stations.

Given N prospective sites in $\Omega \cup \Gamma$, we aim at selecting their subset consisting of the locations at which the measurements made by n available sensors would lead to least-squares estimates of θ_1, θ_2, and θ_3 characterized by a minimal volume of the confidence ellipsoid. This task reduces to solving Problem P of Sect. 2.

Note that in the design, the elements of the sensitivity vector (8) at admissible sites are indispensable in order to determine the matrices M_i in (12) which potentially contribute to the FIM (11). Assuming the nominal values $\theta_1^0 = 0.1$, $\theta_2^0 = \theta_3^0 = -0.02$, they can be obtained using the direct-differentiation approach (Uciński 2005; Sun 1994). Roughly speaking, this boils down to solution of the system of four PDEs in which one equation constitutes the original state equation (50), and the other three equations result from its differentiation with respect to θ_1, θ_2, and θ_3, respectively. The initial and Neumann boundary conditions for all the four equations are homogeneous here. We solved this system of PDEs using some routines of the MATLAB PDE toolbox (COMSOL-AB 1995) for a spatial mesh composed of 3,200 triangles and 1,681 nodes. Numerical integration required to evaluate matrices M_i was performed employing the

trapezoidal rule for the time step equal to 1/80, based on the sensitivity vector g interpolated at the nodes representing admissible locations x^i, cf. Appendix I in (Uciński 2005) for details. Despite the impossibility of employing the graphical user interface of the toolbox (it is tailored to single PDEs, and not to systems of PDEs), we could still solve the system of PDEs in question using command-line functions. The GUI was applied here only to conveniently define the spatial domain and then to export the resulting decomposed geometry matrix to MATLAB's workspace.

To provide insight into the performance of the presented approach, two scenarios were considered in our simulation studies, namely the observations were assumed to be taken at locations selected from among the elements of rectangular grids of sizes 21×21 nodes ($N = 441$ admissible sites) and 41×41 nodes ($N = 1,681$ admissible sites) with different numbers of allocated sensors. All procedures were implemented entirely in MATLAB7.1 and tested on a PC equipped with an Intel Centrino 1.73 GHz processor and 1 GB RAM, running Windows XP.

The solution to (50)–(52) is presented in Fig. 2 where the process dynamics can be easily observed. It can be seen that the cloud of pollutant spreads over the entire domain, thereby reflecting the complex combination of diffusion and advection processes, and follows the direction of the wind being the dominant transport factor.

A MATLAB program was written to implement the recursive version of the DFBB procedure embodied by Algorithm 1. For the implementation of its principal part, which is the solution of the relaxed problem with the use of the SD scheme given in Algorithm 2, the following setting was assumed:

- In order to solve the column generation problem in Step 1, the function `linprog` from MATLAB's Optimization Toolbox (Math Works 2000) was used with the default tolerance being set as 10^{-8}.
- The accuracy of the solution of the relaxed problem (the termination check in Step 2) was set on the level of $\epsilon = 5 \cdot 10^{-5}$.
- The tolerance and the maximum number of iterations for the multiplicative algorithm for solving the restricted master problem in Step 3 were set as $\eta = 10^{-5}$ and $\kappa_{\max} = 1,000$, respectively.

Finally, to take full advantage of the efficient SD scheme performed at each node of the BB tree, Algorithm 1 was extended to incorporate a rounding procedure after solving the relaxed problem. Thus, this solution provides not only an upper bound to the currently processed branch, but may also lead to a great improvement in the lower bound *LOWER*. Intuitively, a proper way of rounding a relaxed solution is to choose target sites for locating spare sensors so that they correspond to largest weights. In the event that this choice is complicated by the presence of sites with identical weights, the target sites are chosen randomly. The cost of such a procedure is comparable to the effort of sorting the vector representing the relaxed solution. In such a manner, it is possible to avoid descending to the bottom level of the BB tree in order to update *LOWER*. Since in the case of medium and large-scale problems such a process costs many recursive calls of Algorithm 1, the effort put to round relaxed solutions seems to be insignificant and fully justified, as evidenced by numerous simulation experiments.

The D-optimal sensor configurations for different sizes of the grid and numbers of allocated sensors are shown in Figs. 3 and 4. Each time, the computations were started by randomly generating a guess at the sought solution in order to assign an initial value to the global variable *LOWER*. The obtained locations of sensors perfectly retain the symmetry of the problem with respect to the line $x_2 = x_1$ and tend to

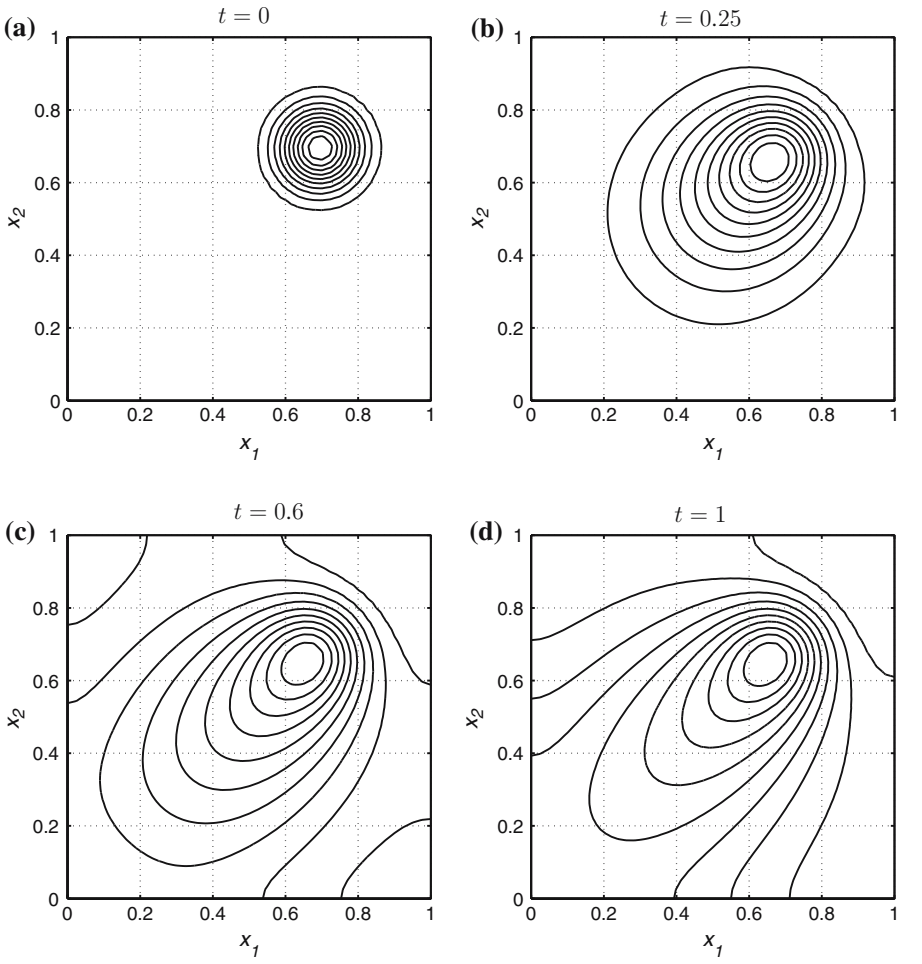

Fig. 2 Concentration of the pollutant at consecutive time instants

form a pattern reflecting the areas of greatest changes in the pollutant concentration. Surprisingly, the measurements in the closest vicinity of the pollution source are not very attractive for parameter estimation. The intuition fails in this case and it is very difficult to predict the solution when armed only with the experimenter's experience.

The results concerning the algorithm performance are summarized in Table 1. To make an informative comparison, the versions with and without rounding of the relaxed solution are given. The number of RDFBB calls equals to one means that the optimal solution is obtained just by rounding the fully relaxed problem. Although the presented examples are rather medium-scale, we have to remember that in the worst case (i.e., when the number of sensors is closest to half the number of available sites) the cardinalities of the search spaces for the 441-point and 1,681-point grids reach even $2.15 \cdot 10^{131}$ and $2.09 \cdot 10^{504}$, respectively. Examination of the data from Table 1 leads to interesting conclusions. Unexpectedly, with an increased number of sensors

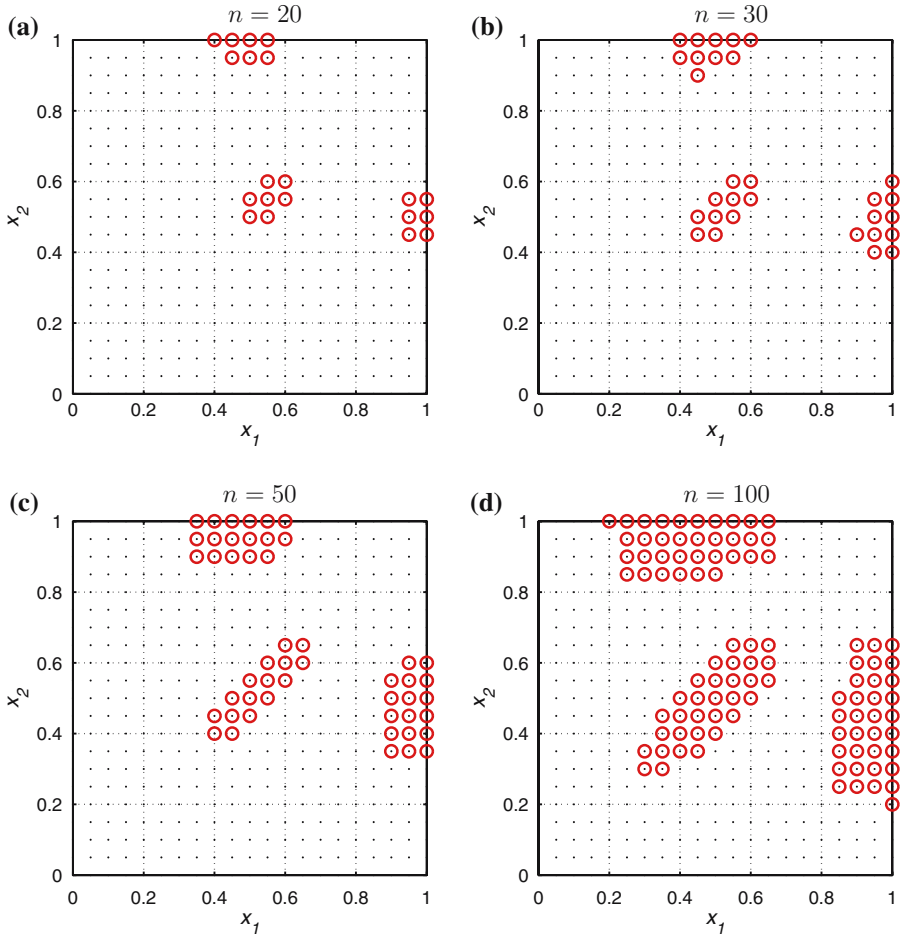

Fig. 3 D-optimal sensor positions (marked with *open circles*) for the 441-point uniform grid of admissible sites (marked with *dots*) and various numbers of sensors

(and the size of the corresponding search space), the pruning process becomes more efficient (only in the case of a finer grid without rounding we observed the influence of the 'curse of dimensionality'). This effect can be explained by observing that a higher density of sensors leads to a better estimate of the lower bound to the optimal value of the design criterion, which results in an increased efficiency of pruning and whereupon the search is speeded up. Moreover, the increase in the sensor density seems to make the rounded solution of the relaxed problem closer to the optimal one in terms of the criterion value. Therefore, the influence of rounding cannot be overestimated since in our example it significantly decreases the number of recursive function calls and, surprisingly, a low number of sensors leads to a harder situation from the computational point of view.

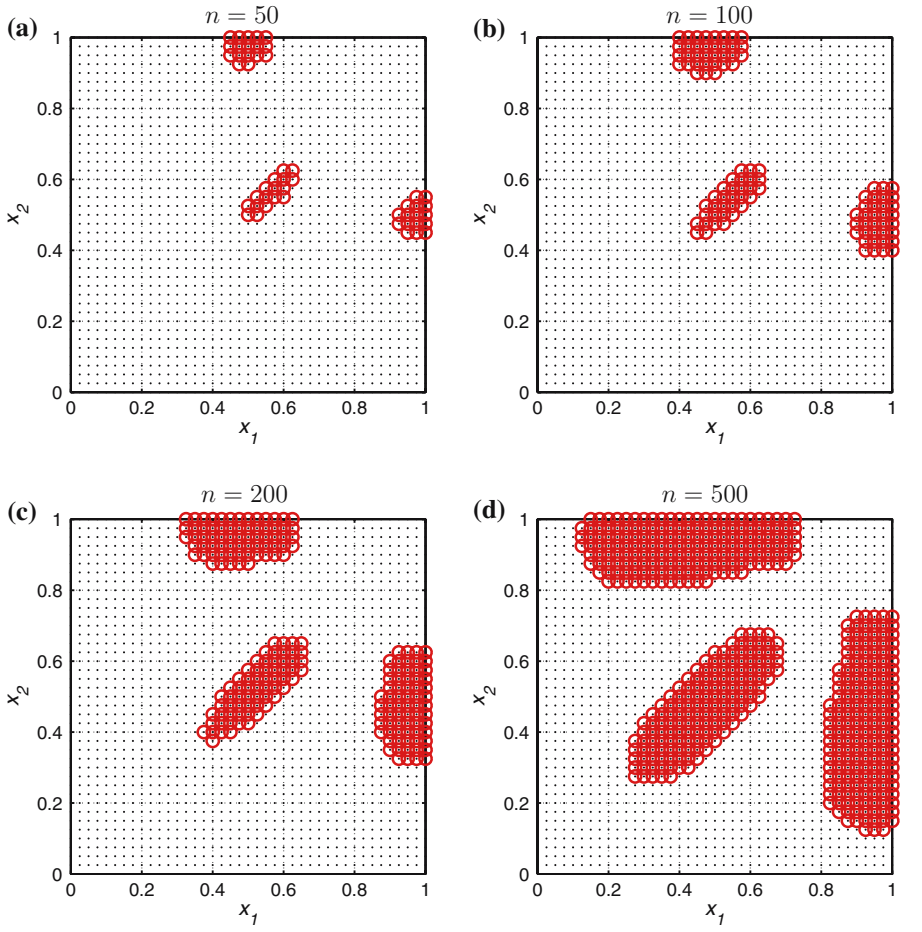

Fig. 4 D-optimal sensor positions (marked with *open circles*) for the 1,681-point uniform grid of admissible sites (marked with *dots*) and various numbers of sensors

6 Discussion and conclusions

We have addressed the problem of selecting optimal observation points in view of accurate parameter estimation for parameter distributed systems, which stand here for dynamical systems governed by partial differential equations. Although it has been approached from various angles since the mid-1970s, there are still few systematic and versatile methods for its solution. In the existing formulations, an optimal sensor placement is thus computed as that which globally maximizes a criterion directly connected with the expected quality of the parameter estimates. But then the key difficulty becomes the large scale of the resulting global optimization problem, since the monitoring networks encountered in process industry or environmental engineering may often consist of several hundreds of stations. Obviously, this makes the exhaustive search on a candidate-by-candidate basis practically intractable and creates a

Table 1 Comparison of the algorithm performance for different scenarios (*RP* relaxed problem, *RDFBB* recursive depth-first BB procedure)

Number of sensors	observational grid: 21 × 21 (441 admissible sites)			
	Without rounding of RP		With rounding of RP	
	Number of RDFBB calls	CPU time [h : min : s]	Number of RDFBB calls	CPU time [h : min : s]
20	413	0 : 08 : 20.93	53	0 : 01 : 02.89
30	131	0 : 02 : 59.32	25	0 : 00 : 27.17
40	115	0 : 02 : 11.59	3	0 : 00 : 02.90
50	115	0 : 01 : 51.76	1	0 : 00 : 00.98
60	17	0 : 00 : 16.50	1	0 : 00 : 01.01
100	13	0 : 00 : 12.05	1	0 : 00 : 00.99
	observational grid: 41 × 41 (1,681 admissible sites)			
50	1,267	1 : 36 : 36.70	37	0 : 02 : 22.38
80	183	0 : 12 : 58.09	1	0 : 00 : 02.74
100	163	0 : 10 : 37.07	1	0 : 00 : 02.86
200	261	0 : 09 : 41.80	1	0 : 00 : 02.26
500	339	0 : 10 : 25.20	1	0 : 00 : 02.37
800	439	0 : 14 : 48.14	1	0 : 00 : 02.36

need for techniques which would implement a guided search and have acceptable performance.

We started from the most common formulation, in which the measurement system has a finite number of sensor candidate positions and the aim is to select the best subset of points of desired cardinality. Choosing the best subset translates to maximizing the determinant of the Fisher information matrix associated with the estimated parameters and fits into the framework of nonlinear 0–1 integer programming. The solution of this combinatorial design problem using the BB method constitutes a quite natural option, but the main problem when trying to implement it has been the lack of a low-cost procedure to obtain upper bounds to the optimal values of the D-optimality criterion. The main contribution of this paper consists in the development of an original and efficient computational scheme to produce such bounds. This was possible by adapting a specialized multiplicative algorithm for determinant maximization, which is in common use by statisticians concerned with optimum experimental design. The link to plug this algorithm into the proposed scheme was a simplicial decomposition being perfectly suited for large-scale problems which can be encountered here. The idea of its application in the context of optimal sensor location for parameter estimation is new. Consequently, the proposed method can be implemented with great ease and our experience provides evidence that, with this tool, even large-scale design problems can be solved using an off-the-shelf PC. A further significant speedup can be achieved by employing a simple rounding heuristic, as described in Sect. 5.

As by-products of the process of constructing the ultimate algorithm, characterizations of the optimal solutions to specific subproblems discussed in Propositions 1 and 3 were obtained. In principle, they can be derived as particular cases of some general results in optimum experimental design, but this would require from the reader fluent knowledge of measure theory. Because of this, the corresponding proofs proceed here independently on a quite elementary level, based on the Karush–Kuhn–Tucker conditions. In this respect, the presented versions of both the proofs are new.

Let us remark that an alternative approach to select a best n-element subset from among a given N-element set of candidate sites could be to employ an exchange algorithm. Typically, algorithms of this type begin with an n-point starting sensor configuration which then sequentially evolves through addition of new elements selected from among vacant sites and deletion of sites at which sensors have provisionally been planned to reside, in an effort to improve the value of the adopted design criterion (Meyer and Nachtsheim 1995). Accordingly, a one-point exchange procedure was used in (Uciński 1999) and further developed in (Uciński and Patan 2002; Uciński 2005) in a sensor-network setting, based on the concept of replication-free designs set forth by Fedorov (1989). A much more efficient extension of this idea could be to adapt the fast algorithm based on multiple simultaneous exchanges, which was developed by Lam et al. (2002). A step in this direction was made by Liu et al. (2005) who refined it and applied the resulting 'sort-and-cut' technique to solve an E-optimum sensor selection problem. It is beyond doubt that this approach outperforms the BB technique proposed here as far as the running time is concerned. On should note, however, that exchange algorithms are heuristics and thus they are only capable of finding globally competitive solutions (i.e., nearly optimal ones), with an explicit trade of global optimality for 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.

Certainly, there is room for refinements and further theoretical developments. We can mention the following points which are the matter of our current research:

- Optimization problems consisting in finding probability mass functions maximizing various experimental design criteria were discussed by Boyd and Vandenberghe (2004, p. 384), cf. also (Uciński 2005 p. 59). In particular, they cast the determinant maximization as a quite specific convex extension of the semidefinite programming problem and developed an interior-point algorithm (Vandenberghe et al. 1998) which can basically be used here in lieu of the multiplicative algorithm solving the restricted master problem. But due to its broad generality, it performs slightly worse than Algorithm 3 which fully exploits the problem specificity. This difference becomes rather striking if problems involving several hundreds of admissible sites are to be solved. Nevertheless, semidefinite programming formulations of A- and E-optimum design problems presented in (Boyd and Vandenberghe 2004; Vandenberghe and Boyd 1999) can be used as a fine starting point to extend the applicability of the BB technique presented here to those criteria, too.
- The proposed simple branching rule for the binary BB tree can be refined by incorporating a mechanism driving the search proces towards the most promising branches in terms of the objective function or heuristics exploiting specific properties of the DPS in question.
- The column generation subproblem reduces to maximization of a linear function over the intersection of a hyperplane and a box. With some additional efforts, this specific form of the feasible set can be fully exploited to eliminate the need for a sophisticated linear-programming solver. The algorithm developed in (D. Uciński 2006) for a constrained D-optimal design problem implements this idea and is almost as simple as a closed-form solution. Our computational experience indicates, however, that incorporation of this improvement in the BB scheme does not yield a significant reduction in the time of execution of the resulting code.
- Since computing on clusters of PCs has become as common as computing on scientific workstations had been a decade ago, we are currently implementing a parallel

version of the BB application aimed at solving large-scale problems. This task is facilitated by the specific structure of the proposed BB algorithm whose components are particularly suited for parallelization.

Acknowledgements The authors would like to thank the anonymous reviewers for their numerous constructive suggestions that greatly helped to improve the quality of the manuscript.

The work was supported by the Polish State Committee for Scientific Research under grant No. 3 T11C 035 27.

Appendix

A: Proof of Proposition 1

Problem $\mathrm{R}'(I_0, I_1)$ can be rewritten as follows: Find $w^\star \in \mathbb{R}^q$ to minimize

$$\widetilde{\mathcal{Q}}(w) = -\log\det\big(G(w)\big) \tag{1}$$

subject to the constraints

$$\sum_{j=1}^{q} w_j - r = 0, \tag{2}$$

$$-w_j \le 0, \quad j = 1, \ldots, q, \tag{3}$$

$$w_j - 1 \le 0, \quad j = 1, \ldots, q. \tag{4}$$

Associating the dual variables

$$\lambda \in \mathbb{R}, \quad \mu_j \in \mathbb{R}_+, \quad \nu_j \in \mathbb{R}_+, \quad j = 1, \ldots, q, \tag{5}$$

with constraints (2), (3), and (4), respectively, we define the Lagrangian of (1)–(4) as

$$\mathcal{L}(w, \lambda, \mu, \nu) = -\log\det\big(G(w)\big) + \lambda\Big(\sum_{j=1}^{q} w_j - r\Big) - \sum_{j=1}^{q} \mu_j w_j + \sum_{j=1}^{q} \nu_j\big(w_j - 1\big). \tag{6}$$

An easy computation shows that

$$\frac{\partial \mathcal{L}}{\partial w_j} = -\varphi(j, w) + \lambda - \mu_j + \nu_j. \tag{7}$$

Let us examine the first-order Karush–Kuhn–Tucker (KKT) conditions for our problem (Bertsekas 1999):

$$-\varphi(j, w) + \lambda - \mu_j + \nu_j = 0, \qquad j = 1, \ldots, q, \tag{8}$$

$$\mu_j w_j = 0, \qquad j = 1, \ldots, q, \tag{9}$$

$$\nu_j\big(w_j - 1\big) = 0, \qquad j = 1, \ldots, q, \tag{10}$$

$$\mu_j \ge 0, \qquad j = 1, \ldots, q, \tag{11}$$

$$\nu_j \ge 0, \qquad j = 1, \ldots, q, \tag{12}$$

$$0 \le w_j \le 1, \qquad j = 1, \ldots, q, \tag{13}$$

$$\sum_{j=1}^{q} w_j = r. \tag{14}$$

For problems with constraints which are both linear and linearly independent, as is the case here, the KKT conditions are necessary for optimality. Additionally, the convexity of the objective function (1) implies that they also become sufficient. Consequently, the optimality of $w^\star$ amounts to the existence of some values of λ, μ_j and ν_j, $j = 1, \ldots, q$, denoted by $\lambda^\star$, $\mu_j^\star$ and $\nu_j^\star$, $j = 1, \ldots, q$, respectively, such that (8)–(14) are satisfied.

Suppose that $w_j^\star = 1$ for some index j. Then from (9) it follows that $\mu_j^\star = 0$ and, therefore, (8) reduces to

$$\varphi(j, w^\star) = \lambda^\star + \nu_j^\star \geq \lambda^\star, \tag{15}$$

the last inequality owing to (12). In turn, on account of (10), the assumption $w_j^\star = 0$ yields $\nu_j^\star = 0$, and then (8) simplifies to

$$\varphi(j, w^\star) = \lambda^\star - \mu_j^\star \leq \lambda^\star, \tag{16}$$

which is owing to (11). Finally, by (9) and (10), the assumption $0 < w_j^\star < 1$ clearly forces $\mu_j^\star = \nu_j^\star = 0$, for which (8) gives

$$\varphi(j, w^\star) = \lambda^\star, \tag{17}$$

Conversely, having found $w^\star \in \mathbb{R}^q$ and $\lambda^\star \in \mathbb{R}$ for which (28) is fulfilled, we can define

$$\mu_j^\star = \max(\lambda^\star - \varphi(j, w^\star), 0), \quad \nu_j^\star = \max(\varphi(j, w^\star) - \lambda^\star, 0), \quad i = 1, \ldots, q, \tag{18}$$

which guarantees the satisfaction of (8)–(14). This means that $w^\star$ is a KKT point and this is equivalent to its global optimality.

B: Proof of Proposition 2

Observe that the following Löwner ordering holds true for any $w = (w_1, \ldots, w_q)$:

$$0 \preceq G(w) = A + \sum_{j=1}^{q} w_j S_j \preceq \frac{q}{r}\Big(A + \frac{r}{q}\sum_{j=1}^{q} S_j\Big) = \frac{q}{r}\, G(\bar{w}). \tag{19}$$

A fundamental property of the determinant is that it preserves this monotonicity (Horn and Johnson 1986 Corr. 7.7.4), which gives

$$0 \leq \det\big(G(w)\big) \leq \Big(\frac{q}{r}\Big)^m \det\big(G(\bar{w})\big). \tag{20}$$

This makes our claim obvious.

C: Proof of Proposition 3

Problem $\mathrm{P_{RMP}}$ can equivalently be formulated as follows: Find $\alpha \in \mathbb{R}^s$ to minimize

$$\widetilde{\mathcal{T}}(\alpha) = -\log\det\big(H(\alpha)\big) \tag{21}$$

subject to the constraints

$$\sum_{\ell=1}^{s} \alpha_\ell - 1 = 0, \tag{22}$$

$$-\alpha_\ell \le 0, \quad \ell = 1, \ldots, s. \tag{23}$$

Characterization of its solution proceeds as follows. First, the Lagrangian is formed,

$$\mathcal{L}(\alpha, \lambda, \mu) = -\log\det\big(H(\alpha)\big) + \lambda\Big(\sum_{\ell=1}^{s} \alpha_\ell - 1\Big) - \sum_{\ell=1}^{s} \mu_\ell \alpha_\ell, \tag{24}$$

where $\lambda \in \mathbb{R}$ and $\mu_\ell \in \mathbb{R}_+$, $\ell = 1, \ldots, s$ are dual variables. It follows easily that

$$\frac{\partial \mathcal{L}}{\partial \alpha_\ell} = -\psi(\ell, \alpha) + \lambda - \mu_\ell. \tag{25}$$

Next, it is observed that the first-order KKT conditions for (21)–(23) are

$$-\psi(\ell, \alpha) + \lambda - \mu_\ell = 0, \qquad \ell = 1, \ldots, s, \tag{26}$$

$$\mu_\ell \alpha_\ell = 0, \qquad \ell = 1, \ldots, s, \tag{27}$$

$$\mu_\ell \ge 0, \qquad \ell = 1, \ldots, s, \tag{28}$$

$$\alpha_\ell \ge 0, \qquad \ell = 1, \ldots, s, \tag{29}$$

$$\sum_{\ell=1}^{s} \alpha_\ell = 1. \tag{30}$$

The independence of $s + 1$ linear constraints defined by (22) and (23) implies that (26)–(30) are necessary for optimality. But they are also sufficient since the performance measure (21) is convex. Accordingly, the optimality of $\alpha^\star$ is equivalent to the existence of some values of λ and μ_ℓ, $\ell = 1, \ldots, s$, denoted by $\lambda^\star$ and $\mu_\ell^\star$, $\ell = 1, \ldots, s$, respectively, such that (26)–(30) are satisfied.

The constant $\lambda^\star$ may be easily inferred from the KKT conditions. Indeed, multiplying each equation in (26) by $\alpha_\ell^\star$, $\ell = 1, \ldots, s$, respectively, and summing the results, we get

$$-\sum_{\ell=1}^{s} \alpha_\ell^\star \operatorname{trace}\big[H^{-1}(\alpha^\star) Q_\ell\big] + \lambda^\star \sum_{\ell=1}^{s} \alpha_\ell^\star - \sum_{\ell=1}^{s} \alpha_\ell^\star \mu_\ell^\star = 0. \tag{31}$$

But

$$\sum_{\ell=1}^{s} \alpha_\ell^\star \operatorname{trace}\big[H^{-1}(\alpha^\star) Q_\ell\big] = \operatorname{trace}\big[H^{-1}(\alpha^\star) H(\alpha^\star)\big] = \operatorname{trace}(E_m) = m, \tag{32}$$

which, taken in conjunction with (27) and (30), simplifies (31) to yield

$$\lambda^\star = m. \tag{33}$$

From this and (26), we conclude that

$$\psi(\ell, \alpha^\star) = m - \mu_\ell^\star \le m, \tag{34}$$

the last inequality being a consequence of (28). Moreover, if $\alpha_\ell^\star > 0$, then (27) implies $\mu_\ell^\star = 0$, i.e.,

$$\psi(\ell, \alpha^\star) = m. \tag{35}$$

Conversely, if (44) is satisfied for some $\alpha^\star \in P_s$, then setting

$$\lambda^\star = m, \quad \mu_\ell^\star = \max(m - \psi(\ell, \alpha^\star), 0), \quad \ell = 1, \ldots, s \tag{36}$$

gives a solution to the system (26)–(30). This shows that $\alpha^\star$ is a KKT point.

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.

Banks, H. T. and Kunisch, K.: 1989, *Estimation Techniques for Distributed Parameter Systems*, Systems & Control: Foundations & Applications, Birkhäuser, Boston.

Banks, H. T., Smith, R. C. and Wang, Y.: 1996, *Smart Material Structures: Modeling, Estimation and Control*, Research in Applied Mathematics, Masson, Paris.

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

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

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

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, 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.

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.

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.

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.

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.

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

Gerdts, M.: 2005, Solving mixed-integer optimal control problems by branch&bound: A case study from automobile test-driving with gear shift, *Optimal Control Applications and Methods* **26**, 1–18.

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.

Hearn, D. W., Lawphongpanich, S. and Ventura, J. A.: 1985, Finiteness in restricted simplicial decomposition, *Operations Research Letters* **4**(3), 125–130.
Hearn, D. W., Lawphongpanich, S. and Ventura, J. A.: 1987, Restricted simplicial decomposition: Computation and extensions, *Mathematical Programming Study* **31**, 99–118.
Hjalmarsson, H.: 2005, From experiment design to closed-loop control, *Automatica* **41**, 393–438.
Horn, R. A. and Johnson, C. R.: 1986, *Matrix Analysis*, Cambridge University Press, Cambridge, UK.
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, *Proc. American Control Conf.,* San Diego, California, 23–25 May 1990, 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.
Kincaid, R. K. and Padula, S. L.: 2002, D-optimal designs for sensor and actuator locations, *Computers & Operations Research* **29**, 701–713.
Kubrusly, C. S. and Malebranche, H.: 1985, Sensors and controllers location in distributed systems — A survey, *Automatica* **21**(2), 117–128.
Lam, R. L. H., Welch, W. J. and Young, S. S.: 2002, Uniform coverage designs for molecule selection, *Technometrics* **44**(2), 99–109.
Lange, K.: 1999, *Numerical Analysis for Statisticians*, Springer-Verlag, New York.
Liu, C. Q., Ding, Y. and Chen, Y.: 2005, Optimal coordinate sensor placements for estimating mean and variance components of variation sources, *IEE Transactions* **37**, 877–889.
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.
MathWorks: 2000, *Optimization Toolbox for Use with Matlab. User's Guide, Version 2*, The MathWorks, Inc., Natick, MA.
Meyer, R. K. and Nachtsheim, C. J.: 1995, The coordinate-exchange algorithm for constructing exact optimal experimental designs, *Technometrics* **37**(1), 60–69.
Müller, W. G.: 2001, *Collecting Spatial Data. Optimum Design of Experiments for Random Fields*, Contributions to Statistics, 2nd revised edn, 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, volume 63, Springer-Verlag, Berlin, pp. 415–433.
Navon, I. M.: 1997, Practical and theoretical aspects of adjoint 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.
Nocedal, J. and Wright, S. J.: 1999, *Numerical Opimization*, Springer-Verlag, New York.
Nychka, D., Piegorsch, W. W. and Cox, L. H. (eds): 1998, *Case Studies in Environmental Statistics*, Lecture Notes in Statistics, volume 132, Springer-Verlag, New York.
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, volume 132, Springer-Verlag, New York, 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.
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.
Patriksson, M.: 2001, Simplicial decomposition algorithms, *in* C. A. Floudas and P. M. Pardalos (eds), *Encyclopedia of Optimization*, Vol. 5, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 205–212.
Pázman, A.: 1986, *Foundations of Optimum Experimental Design*, Mathematics and Its Applications, D. Reidel Publishing Company, Dordrecht, The Netherlands.

Pierre, D. A.: 1969, *Optimization Theory with Applications*, Series in Decision and Control, John Wiley & Sons, New York.
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.
Pronzato, L.: 2003, Removing non-optimal support points in D-optimum design algorithms, *Statistics & Probability Letters* **63**, 223–228.
Pronzato, L.: 2004, A minimax equivalence theorem for optimum bounded design measures, *Statistics & Probability Letters* **68**, 325–331.
Pukelsheim, F.: 1993, *Optimal Design of Experiments*, Probability and Mathematical Statistics, John Wiley & Sons, New York.
Quereshi, Z. H., Ng, T. S. and Goodwin, G. C.: 1980, Optimum experimental design for identification of distributed parameter systems, *International Journal of Control* **31**(1), 21–29.
Rafajłowicz, E.: 1981, Design of experiments for eigenvalue identification in distributed-parameter systems, *International Journal of Control* **34**(6), 1079–1094.
Rafajłowicz, E.: 1983, Optimal experiment design for identification of linear distributed-parameter systems: Frequency domain approach, *IEEE Transactions on Automatic Control* **28**(7), 806–808.
Rafajłowicz, E.: 1986, Optimum choice of moving sensor trajectories for distributed parameter system identification, *International Journal of Control* **43**(5), 1441–1451.
Reinefeld, A.: 2001, Heuristic search, *in* C. A. Floudas and P. M. Pardalos (eds), *Encyclopedia of Optimization*, Vol. 2, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 409–411.
Russell, S. J. and Norvig, P.: 2003, *Artificial Intelligence: A Modern Approach*, 2nd edn, Pearson Education International, Upper Saddle River, NJ.
Silvey, S. D., Titterington, D. M. and Torsney, B.: 1978, An algorithm for optimal designs on a finite design space, *Communications in Statistics — Theory and Methods* **14**, 1379–1389.
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.
Sun, N.-Z.: 1994, *Inverse Problems in Groundwater Modeling*, Theory and Applications of Transport in Porous Media, Kluwer Academic Publishers, Dordrecht, The Netherlands.
Titterington, D. M.: 1980, Aspects of optimal design in dynamic systems, *Technometrics* **22**(3), 287–299.
Torsney, B.: 1988, Computing optimising distributions with applications in design, estimation and image processing, *in* Y. Dodge, V. V. Fedorov and H. P. Wynn (eds), *Optimal Design and Analysis of Experiments*, Elsevier, Amsterdam, pp. 316–370.
Torsney, B. and Mandal, S.: 2001, Construction of constrained optimal designs, *in* A. Atkinson, B. Bogacka and A. Zhigljavsky (eds), *Optimum Design 2000*, Kluwer Academic Publishers, Dordrecht, The Netherlands, chapter 14, pp. 141–152.
Torsney, B. and Mandal, S.: 2004, Multiplicative algorithms for constructing optimizing distributions: Further developments, *in* A. Di Bucchianico, H. Läuter and H. P. Wynn (eds), mODa 7*, Proc. 7th Int. Workshop on Model-Oriented Data Analysis,* Heeze, The Netherlands, 2004, Physica-Verlag, Heidelberg, pp. 163–171.
Uciński, D.: 1999, *Measurement Optimization for Parameter Estimation in Distributed Systems*, Technical University Press, Zielona Góra. Available in electronic form at `http://www.issi.uz.zgora.pl/~ucinski/`.
Uciński, D.: 2000, Optimal 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.: 2006, Construction of constrained D-optimum designs using simplicial decomposition, *Computational Statistics & Data Analysis* . (submitted).
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, *Proc. 44th IEEE Conference on Decision and Control, and the European Control Conference 2005,* Seville, Spain. Published on CD-ROM.
Uciński, D. and Korbicz, J.: 2001, Optimal sensor allocation for parameter estimation in distributed systems, *Journal of Inverse and Ill-Posed Problems* **9**(3), 301–317.

Uciński, D. and Patan, M.: 2002, Optimal location of discrete scanning sensors for parameter estimation of distributed systems, *Proc. 15th IFAC World Congress,* Barcelona, Spain, 22–26 July 2002. Published on CD-ROM.

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, *Proc. 14th IFAC World Congress,* Beijing, China, 5–9 July, 1999, Vol. I: Modeling, Identification, Signal Processing II, Adaptive Control, pp. 37–42.

Vandenberghe, L. and Boyd, S.: 1999, Applications of semidefinite programming, *Applied Numerical Mathematics* **29**, 283–299.

Vandenberghe, L., Boyd, S. and Wu, S.-P.: 1998, Determinant maximization with linear matrix inequality constraints, *SIAM Journal on Matrix Analysis and Applications* **19**(2), 499–533.

Ventura, J. A. and Hearn, D. W.: 1993, Restricted simplicial decomposition for convex constrained problems, *Mathematical Programming* **59**, 71–85.

Vogel, C. R.: 2002, *Computational Methods for Inverse Problems*, Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia.

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.