Int. J. Appl. Math. Comput. Sci., 2018, Vol. 28, No. 1, 39–54
DOI: 10.2478/amcs-2018-0003

DISTRIBUTED SCHEDULING OF MEASUREMENTS IN A SENSOR NETWORK FOR PARAMETER ESTIMATION OF SPATIO–TEMPORAL SYSTEMS

MACIEJ PATAN [a,*], DAMIAN KOWALÓW [a]

[a]Institute of Control and Computation Engineering
University of Zielona Góra, ul. Szafrana 2, 65-516 Zielona Góra, Poland
e-mail: {m.patan,d.kowalow}@issi.uz.zgora.pl

The main aim of the paper is to develop a distributed algorithm for optimal node activation in a sensor network whose measurements are used for parameter estimation of the underlying distributed parameter system. Given a fixed partition of the observation horizon into a finite number of consecutive intervals, the problem under consideration is to optimize the percentage of the total number of observations spent at given sensor nodes in such a way as to maximize the accuracy of system parameter estimates. To achieve this, the determinant of the Fisher information matrix related to the covariance matrix of the parameter estimates is used as the qualitative design criterion (the so-called D-optimality). The proposed approach converts the measurement scheduling problem to a convex optimization one, in which the sensor locations are given *a priori* and the aim is to determine the associated weights, which quantify the contributions of individual gaged sites to the total measurement plan. Then, adopting a pairwise communication scheme, a fully distributed procedure for calculating the percentage of observations spent at given sensor locations is developed, which is a major novelty here. Another significant contribution of this work consists in derivation of necessary and sufficient conditions for the optimality of solutions. As a result, a simple and effective computational scheme is obtained which can be implemented without resorting to sophisticated numerical software. The delineated approach is illustrated by simulation examples of a sensor network design for a two-dimensional convective diffusion process.

Keywords: sensor network, distributed parameter system, parameter estimation, experimental design, sensor scheduling.

1. Introduction

Optimized measurement strategies for parameter estimation of spatio-temporal systems, commonly known as distributed parameter systems (DPSs), have become of considerable interest in both theory and practice for a great variety of engineering problems. A key difficulty here is that it is impossible to observe the system states over the entire spatial domain. This leads directly to the issue of determining a sensor placement so as to get the most valuable information about the unknown system parameters.

At the same time, technological advances in communication systems and sensor technology have begun a strong trend of measurement devices to become miniaturized and cheap, while making the networks of sensors increasingly powerful. The performance of distributed sensor networks improved to the point where their wide applicability is unquestionable (Zhao and Guibas, 2004; Ögren *et al.*, 2004; Jain and Agrawal, 2005; Cassandras and Li, 2005; Wu, 2006; Boukerche, 2006; Zhong and Cassandras, 2011). This is especially important in the context of spatio-temporal systems where a large number of sensors can be used for the task of monitoring the dynamics of a system. A cooperated and scalable network of sensors has a potential to substantially enhance the performance of the observation systems. Although in the context of DPSs the number of sensor placement techniques dedicated to manage the configuration for sensor networks of practical scale is limited, some effective approaches have been proposed to cover a number of different experimental settings, including stationary (Nehorai *et al.*, 1995; Uciński, 2000; Point *et al.*, 1996; Joshi and Boyd, 2009; Patan and Kowalów, 2014; Kowalów *et al.*, 2015; Patan and Uciński, 2016a), scanning (Uciński and Patan, 2002; Patan, 2006; 2008; 2012a; Tricaud *et al.*, 2008) or moving observations (Rafajłowicz, 1986; Porat and Nehorai, 1996; Uciński,

*Corresponding author

2000; 2012; Patan *et al.*, 2008; Demetriou and Hussein, 2009; Demetriou, 2010). Their adaptation to the sensor-network setting was reported by Tricaud and Chen (2012) as well as Song *et al*. (2009).

Some 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 minimized. The 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 estimator is employed so that the optimal sensor placement can be determined independently of the estimator used. This leads to a great simplification since the Cramér–Rao lower bound for the aforementioned covariance matrix is merely the inverse of the FIM, which can be computed with relative ease, even though the exact covariance matrix of a particular estimator is very difficult to obtain. A related optimality criterion was given by Point *et al*. (1996) by the maximization of the Gram determinant, which is a measure of the independence of the sensitivity functions evaluated at sensor locations. The form of the criterion itself resembles the D-optimality criterion, being the determinant of the FIM, but the counterpart of the information matrix takes on much larger dimensions, which may lead to more intensive calculations.

However, most approaches communicated by various authors rely on centralized techniques, which assume the existence of some superior entity to maintain the whole network, responsible for optimization of the observation strategy. The distributed nature of the measurement scheduling problem is taken into account very occasionally. Recent advancements in sensor networks necessitate effective and distributed algorithms for computation and information exchange.

In this paper we wish to focus on certain computational aspects of sensor placement problems for parameter estimation in DPSs. Given a finite set of stationary located sensors, the observation horizon is divided into consecutive stages (time subintervals). Then, the non-negative weights are assigned to each sensor at a given stage so as to maximize the determinant of the FIM. The weight assigned to a measurement point can be interpreted as the proportion of observations performed at this point, or the frequency of sensor activation during replications of the experimental trials. Thus, the observation scheduling problem is embedded in the context of weight optimization, which is close in spirit to the classical optimum experimental design theory for lumped systems (Atkinson *et al*., 2007; Fedorov and Hackl, 1997; Pázman, 1986). The potential solutions are of considerable interest while assessing which sensors are more informative than the others and allow complexity reduction of the measurement system.

The approach developed here relies on a very simple fixed-point iteration algorithm for computation of a D-optimum design on a finite set, devised and analysed by Pázman (1986), Silvey *et al*. (1978) or Torsney and Mandal (2004) for the case of rank-one information matrices per single observation, and then extended to the multi-response systems case by Uciński (2004, p. 62). In the paper, we further generalize it to the framework of scanning observations, with possible sensor switchings between the consecutive time subintervals. Although there exist some alternative approaches to the problem, via application of semi-definite programming (Joshi and Boyd, 2009) or branch-and-bound discrete optimization (Uciński and Patan, 2007), unfortunately there is no straightforward and efficient way to adapt them to the framework of decentralized computations.

A key contribution of this work consists in partial decomposition of the optimization problem into subproblems for each time subintervals coupled by the global FIM. To achieve this, the necessary and sufficient conditions for the optimality of solutions are derived, together with appropriate mathematical substantiation based on the concept of the so-called *continuous designs*. What is more, this result also opens the possibility of embedding the optimization task into the context of consensus type algorithms over a distributed network. The general idea is based on the class of so-called randomized *gossip* algorithms in which each node communicates with no more than one neighbor at each time instant (Xiao and Boyd, 2004; Boyd *et al*., 2006), introduced in the domain of communication and distributed data averaging. It was extended to temporal measurements by Braca *et al*. (2008), although still without direct relations to measurement scheduling. Therefore, this work can be considered the first attempt to extend this concept to the area of distributed sensor scheduling for parameter estimation of DPSs.

In the presented algorithmic scheme, the different nodes of the network independently calculate and store the desired quantities, and their final estimates are obtained in a fully decentralized fashion. The advantage of such information exchange is clear as it is robust with respect to individual sensor faults and the global estimate of the information matrix is stored at all sensor nodes so can easily be recovered. Since our intention is to develop an algorithmic approach able to cope with the networks of practical scale comprising hundreds or even thousands of nodes, in order to reduce the communication payload on the network nodes the problem of reduction of measurement space is also discussed. Here, the additional contribution of this work is derivation of a suitable thresholding rule for rejecting non-informative sensor nodes, being an extension of the result derived

originally by Pronzato (2003) to the case of scanning observations of multi-output DPSs.

The paper has the following structure. Section 2 formally states the measurement scheduling problem in the context of parameter estimation of spatio-temporal systems. The continuous approximation of the original discrete resource allocation problem, its decomposition and optimality conditions are discussed in Section 3. Section 4 provides the algorithmic solution for distributed computing of a spatio-temporal measurement schedule over sensor network. In Section 5, the illustration of algorithm performance is given on a simulation example of a two-dimensional convective diffusion process. Conclusion and a discussion on related open issues are delivered in Section 6. The proofs of some essential properties of the approach developed in the paper are contained in Appendix.

Notation. Notation throughout the paper is more or less standard. We use $\mathbb{N}$ and $\mathbb{R}$ to denote the sets of integers and real numbers, respectively. Further, $\mathbb{N}_+$ and $\mathbb{R}_+$ denote the sets of non-negative integer and 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 $\mathbb{R}^{n\times k}$. 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 I_n. In $\mathbb{S}^n$, two sets are of special importance: the cone of non-negative definite matrices and the cone of positive definite matrices, denoted by $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$, which are respectively given by

$$\mathbb{S}^n_+ = \{A \in \mathbb{R}^n :\ x^{\mathsf{T}}Ax \geq 0 \text{ for all } x \in \mathbb{R}^n\},$$
$$\mathbb{S}^n_{++} = \{A \in \mathbb{R}^n :\ x^{\mathsf{T}}Ax > 0 \text{ for all } 0 \neq x \in \mathbb{R}^n\}.$$

The symbols $\mathbf{1}$ and $\mathbf{0}$ denote vectors whose components are all one and zero, respectively. The context makes their lengths clear. Given a set of points A, $\mathrm{conv}(A)$ stands for its convex hull. The probability (or canonical) simplex in $\mathbb{R}^n$ is defined as $\mathbb{P}_n = \mathrm{conv}\big(\{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n\}\big) = \big\{p \in \mathbb{R}^n_+ \mid \mathbf{1}^{\mathsf{T}}p = 1\big\}$, where $\boldsymbol{e}_j$ is the usual unit vector along the j-th coordinate of $\mathbb{R}^n$.

2. Measurement scheduling for identification of DPSs

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

The state y is observed by N scanning sensors, which possibly change their positions at time instants $0 < t_0 < t_1 < \cdots < t_K = t_f$ and will remain stationary for the duration of each subinterval $T_k = [t_{k-1}, t_k]$, $k = 1,\ldots,K$. Thus, the 'scanning' observation strategy considered can be formally represented as (Patan and Uciński, 2010; 2016b)

$$z^j(t) = y(x^j_k,t;\theta) + \varepsilon(x^j_k,t), \quad x^j_k \in X, \quad t \in T_k, \tag{1}$$

for $j = 1,\ldots,N$ and $k = 1,\ldots,$ where x^j_k is the element selected from *a-priori* defined set of sensor locations $X = \{x^1,\ldots,x^N\}$ and $\varepsilon(x^j_k,t)$ is measurement noise which is assumed zero-mean, Gaussian, spatial uncorrelated and white (Patan, 2008; Patan and Uciński, 2016a), i.e., $\mathrm{E}[\varepsilon(x^j_k,t)] = 0$ and $\mathrm{cov}[\varepsilon(x^j_k,t)] = C(x^j_k,t)$, where $C(x,t) \in \mathbb{S}^s_{++}$ is a known covariance matrix representing potential correlations between outputs of the system. For the sake of simplicity of our further discussion, the multiple outputs are assumed to be uncorrelated and noise normalized, i.e., C is an identity matrix. A comprehensive treatment of the setting of correlated outputs is proposed Patan (2012b). Also, although white noise is not possible in real applications, it constitutes here a fair approximation to disturbances whose adjacent samples are uncorrelated at all time instants for which the time increment is small compared with the time constants of the DPS (Kubrusly and Malebranche, 1985). Such an approach is often accepted in the literature as there is a large number of practical application where such an approximation is satisfactory, including examples of environmental monitoring (Jacobson, 1999; Cacuci *et al.*, 2014), thermal systems (Alifanov *et al.*, 1995), computer assisted impedance tomography (Holder, 2004) and many others.

A further assumption is that the estimation of the unknown parameter vector θ is performed via minimization of the least-squares criterion

$$\mathcal{J}(\theta) = \sum_{j=1}^{N}\sum_{k=1}^{K}\int_{T_k} \|z^j(t) - \hat{y}(x^j_k,t;\theta)\|^2\,\mathrm{d}t, \tag{2}$$

with $\theta \in \Theta_{\mathrm{ad}}$ and Θ_{ad} being the set of admissible parameters and $\hat{y}(\cdot,\cdot;\theta)$ denoting the system model response suitable for θ. The estimate of the true value of $\theta^\star$ is a vector $\hat{\theta}$ which minimizes $\mathcal{J}(\theta)$. Obviously, the covariance matrix $\mathrm{cov}(\widehat{\theta})$ of the above least-squares estimator depends on the active sensor locations x^j_k. This gives us the possibility to improve the quality of estimates of the system parameters through proper selection of measurements over the sensor networks. 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 Ψ is customarily based on the concept of the *Fisher information matrix* (FIM), which is widely used in optimum experimental design theory for lumped systems (Fedorov and Hackl, 1997; Atkinson *et al.*, 2007). In our setting,

owing to the character of noise in (1), the so-called *average* (or *normalized* per measurement) FIM is given by (Uciński and Patan, 2002; Patan, 2012b)

$$M = \frac{1}{N}\sum_{k=1}^{K}\sum_{j=1}^{N}\Upsilon_k(x_k^j), \tag{3}$$

where $\Upsilon_k(x)$ is a local information matrix at the k-th subinterval at spatial point x given by

$$\Upsilon_k(x) = \frac{1}{t_f}\int_{T_k} G^{\mathsf{T}}(x,t)G(x,t)\,\mathrm{d}t, \tag{4}$$

$$G(x,t) = \left(\frac{\partial y(x,t;\theta)}{\partial\theta}\right)_{\theta=\theta^0}, \tag{5}$$

with G denoting the Jacobi matrix of the sensitivity coefficients. As for θ^0, which stands for a prior estimate to the unknown parameter vector θ, usually some known nominal values of the parameters θ can be used or we can apply estimates obtained from previous experiments (Sun, 1994; Uciński, 2004; Atkinson *et al*., 2007). In the remainder of this paper we shall assume that elements of $G \in \mathbb{R}^{s\times m}$ are continuous. Up to a constant scalar multiplier, the inverse of the FIM constitutes a good approximation of the covariance matrix of any unbiased estimator (Atkinson *et al*., 2007).

As for a specific form of Ψ, various options exist (Fedorov and Hackl, 1997; Atkinson *et al*., 2007), but the most popular choice is to maximize the D-optimality criterion

$$\Psi(M) = \log\det(M), \tag{6}$$

which yields the minimum volume of the uncertainty ellipsoid for the estimates. In the following, we assume a Ψ to be defined as in (6).

3. Problem decomposition and optimality conditions

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

$$\Psi[M] \to \max, \tag{7}$$

with respect to x_k^j belonging to the admissible set X.

Owing to the assumption on independent measurement noise, we admit replicated measurements, i.e., some values x_k^j may appear several times in the optimal solution (we can use multiple sensors located at the same position or we can repeat the process). Consequently, we introduce r_k^j as the number of replicated measurements corresponding to the sensor nodes locations $x^1,\ldots,x^N$ at consecutive time subinterval T_k. The sequence of pairs

$$\xi_k = \{(x_k^1, p_{k1}),\ (x_k^2, p_{k2}),\ \ldots,\ (x_k^N, p_{kN})\}, \tag{8}$$

where $p_{kj} = r_k^j/N$, $N = \sum_{j=1}^{N} r_k^j$, is called the *exact design* of experiment for subinterval T_k. The proportion p_{kj} of observations performed at x_k^j can be considered the percentage of experimental effort spent at that point.

Introducing

$$\xi = (\xi_1, \xi_2, \ldots, \xi_K) \tag{9}$$

as the total experimental design, we rewrite the FIM in the form

$$M(\xi) = \sum_{k=1}^{K} M_k(\xi_k), \tag{10}$$

where

$$M_k(\xi_k) = \frac{1}{t_f}\sum_{j=1}^{N} p_{kj}\int_{T_k} G^{\mathsf{T}}(x_k^j,t)G(x_k^j,t)\,\mathrm{d}t. \tag{11}$$

Here the p_{kj}'s are rational numbers, since both r_k^j's and N are integers. This discrete nature of N-observation exact designs leads directly to integer programming problems, therefore causing serious difficulties that cannot be solved easily by standard optimization techniques, particularly when N is large. A reasonable approach is to relax the definition of the design. When N is large, the weights p_{kj} can be approximated with real numbers in the interval [0, 1], not necessarily integer multiples of $1/N$. Obviously, we must have $\sum_{j=1}^{N} p_{kj} = 1$, for $k = 1,\ldots,K$, so we may think of the designs as probability distributions on set of admissible locations X. This leads to the so-called *continuous* designs, which constitute the basis of the modern theory of optimal experiments (Atkinson *et al*., 2007; Uciński, 2004; Walter and Pronzato, 1997). It turns out that such an approach drastically simplifies the design. Thus, we shall operate on designs which concentrate approximately Np_{kj} measurements at each location x_k^js.

Relaxing the notion of the design ξ_k even further to all probability measures over X which are absolutely continuous with respect to the Lebesgue measure and satisfy by definition the conditions

$$\int_X \xi_k(\mathrm{d}x) = 1, \quad k = 1,\ldots,K, \tag{12}$$

we obtain the equivalent notation, where $M_k(\xi_k)$ in (10) takes the form

$$M_k(\xi_k) = \sum_{k=1}^{K}\int_X \Upsilon_k(x)\,\xi_k(\mathrm{d}x). \tag{13}$$

The integration in (12) and (13) is to be understood in the Lebesgue–Stieltjes sense. Then, we may redefine an optimal design as a solution to the following optimization problem.

Problem 1. Determine $\xi^\star \in \Xi(X)$ to maximize $\Psi[M(\xi)]$, where $\Xi(X)$ denotes the set of all designs of the form (9).

A number of characterizations of the optimal design $\xi^\star$ can be derived based on some general results given by Uciński (2004) and Patan (2012b).

Lemma 1. *For any $\xi \in \Xi(\mathrm{X})$ the information matrix $M(\xi)$ is symmetric and non-negative definite. Let us introduce the notation $\mathfrak{M}(X)$ for the set of all admissible information matrices, i.e., $\mathfrak{M}(X) = \{M(\xi) : \xi \in \Xi(X)\}$.*

Lemma 2. *$\mathfrak{M}(X)$ is compact and convex.*

The next result constitutes the necessary and sufficient condition for the optimality of designs. It is usually called an *equivalence theorem* (Uciński, 2004; Patan, 2012b).

Theorem 1. *The following conditions are equivalent:*

(i) the design $\xi^\star$ maximizes $\Psi[M(\xi)]$,

(ii) $\max\limits_{x\in X} \phi_k(x,\xi^\star) = \varsigma_k(\xi^\star),\ \ k = 1,\dots,K,$

where

$$\phi_k(x,\xi) = \mathrm{trace}\Big[M^{-1}(\xi)\Upsilon_k(x)\Big], \tag{14}$$

$$\varsigma_k(x,\xi) = \mathrm{trace}\Big[M^{-1}(\xi)M_k(\xi_k)\Big]. \tag{15}$$

Proof. See Appendix. ∎

Remark 1. The sensitivity functions $\phi_k(x,\xi)$ are of vital importance here, providing us with a simple test for the optimality of designs. In particular, if they take values less than or equal to $\varsigma_k(\xi)$ for all $x \in X$, then ξ is optimal, otherwise ξ is not optimal.

Remark 2. Theorem 1 establishes decomposition of the original optimization task for the scanning sensor network into a finite set of 'virtually independent' problems, where continuous-time observations are taken by stationary sensors over subintervals T_k. This opens a great possibility of almost direct application of the classical algorithms from the theory of optimum experimental design valid for stationary sensors (Uciński, 2004; Patan, 2012b). Also, since the only element joining the subproblems for consecutive time subintervals is the common global FIM $M(\xi)$, this also raises attractive possibilities for parallel and/or distributed implementations of the algorithm, which will be discussed in the next section.

4. Distributed algorithm

4.1. Experimental effort optimization. Analytical determination of optimal designs is possible only in simple situations, and for general systems it is usually the case that some iterative design procedure will be required. Our basic assumption is that the set of admissible support points X, where the observations are possible, is finite. Because the number of locations from X is limited, any design $\xi \in \Xi(X)$ is uniquely determined by the collection of the corresponding weights. Such a description is very convenient as the problem is reduced to optimization of weights. As has already been mentioned, the weight assigned to a measurement point can be interpreted as the proportion of observations performed at this point. The potential solutions are of considerable interest while assessing which sensors are more informative than others at a given time instant and allow complexity reduction of the measurement system.

In the case under consideration, i.e., the design for fixed sensor locations, we focus our attention on the FIM defined by (10) and (11). Thus, Problem 1 can be rewritten as follows.

Problem 2. Find $P^\star = [p^\star_{kj}],\quad k = 1,\dots K,\quad i = 1,\dots,N$, to maximize the criterion $\mathcal{P}$ defined as

$$\mathcal{P}(P) = \Psi[M(\xi)], \tag{16}$$

subject to $P \in \mathbb{R}^{k\times N}$ being a stochastic matrix, i.e.,

$$P\mathbf{1} = \mathbf{1},\quad p_{kj} \ge 0,\qquad k = 1,\dots K,\quad i = 1,\dots,N. \tag{17}$$

The constraint (17) simply means that each row $p_k = (p_{k1},\dots,p_{k\ell}) \in \mathbb{P}_N$ (i.e., belongs to the canonical simplex of order N). Therefore, Problem 2 is equivalent to a finite-dimensional optimization one over the product of canonical simplices $\mathbb{P}_N^K$.

In the case considered in the paper, i.e., the design for fixed sensor locations, an efficient decentralized computational algorithm can be derived based on the mapping $\mathcal{T} : \Xi(X) \to \Xi(X)$ defined for $k = 1,\dots,K$ by

$$\mathcal{T}\xi = \Big\{\Big(x_k^1, p_{k1}\frac{\phi(x_k^1,\xi)}{\varsigma_k(\xi)}\Big),\dots,\Big(x_k^N, p_{kN}\frac{\phi_k(x_k^N,\xi)}{\varsigma_k(\xi)}\Big)\Big\}. \tag{18}$$

From Theorem 1 it follows that a design $\xi^\star$ is D-optimal if it is a fixed point of the mapping $\mathcal{T}$, i.e.,

$$\mathcal{T}\xi^\star = \xi^\star. \tag{19}$$

Also, we have the valuable property that the sum of weights p_{kj} is invariant with respect to $\mathcal{T}$ for any $k = 1,\dots,K$ (i.e., it is always equal to 1). As for interpretation of this mapping, let us note that in the case of ξ not being an optimal design, it increases the weights for those sensor locations where we observe high values of the sensitivity function at the cost of decreasing the weights of non-informative locations with small values of ϕ. Using this simple idea, it is possible

to develop a decentralized configuration algorithm for the sensor network being a distributed generalization of the classical optimum experimental design problem consisting in iterative computation of a D-optimum design on a finite set. It was proposed in the context of convex experimental design theory for non-linear regression models and widely exploited in applications (Silvey *et al*., 1978; Silvey, 1980; Pázman, 1986; Torsney and Mandal, 2004; Patan and Bogacka, 2007; Patan and Patan, 2010). As for convergence of the fixed-point iteration (18), it was proved in the case of only one system response by Pázman (1986, p.139) and extended for MIMO systems by Uciński (2004).

4.2. Decentralized communication scheme. In the following we assume the asynchronous time model for the configuration process. Let $r = 0, 1, 2 \dots$ be the discrete time index, which partitions the continuous configuration time axis into time slots $Z_r = [z_{r-1}, z_r)$.

Owing to Theorem 1 we can use the mapping $\mathcal{T}$ to iteratively improve the experimental effort distribution. Unfortunately, $\phi_k(\cdot, \xi)$ and $\varsigma_k(\xi)$ computed according to (14) in a centralized fashion are not very useful in our setting. But the only component which cannot be calculated independently of other network nodes is the global information matrix $M(\xi)$. Moreover, from (10) it is clearly seen the information matrix is a weighted average of the subinterval information matrices $\Upsilon_k(x_k^j)$.

Thus, our task is closely related to the problem of distributed averaging on a network, which appears in many applications and has been a subject of extensive studies (Xiao and Boyd, 2004; Boyd *et al*., 2006; Braca *et al*., 2008).

Distributed averaging can be achieved in many ways. One of straightforward techniques is pairwise communication flooding, also known as a *gossip* scheme, which in its classic version assumes that at the k-th reconfiguration time slot the i-th sensor contacts some neighboring node j with probability Q_{ij}, i.e., a pair $(i-j)$ is randomly and independently selected. At this time, both nodes set their values equal to the average of their current values. Denoting by $M_k^j(\xi^{(r)})$ the estimate of (11) for subinterval T_k maintained by the j-th sensor at configuration time slot Z_r, as we have ($\leftarrow$ is an update operator)

$$M_k^{\ell}(\xi^{(r)}) \leftarrow \frac{p_{ki} M_k^i(\xi^{(r)} + p_{kj} M_k^j(\xi^{(r)})}{p_{ki} + p_{kj}}, \qquad (20)$$

where $\ell \in \{i, j\}, \quad k = 1, \dots, K$. However, in our setting the averaging problem is not typical since the contribution of the nodes with weights tending to zero as r increases should be eliminated from the averaging. This can be achieved by increasing the contribution of the local information matrix via the concept of the so-called running consensus (Braca *et al*., 2008),

$$M_k^i(\xi^{(r)}) \leftarrow \frac{r-1}{r} M_k^i(\xi^{(r)}) + \frac{1}{r} \Upsilon_k(x_k^i). \qquad (21)$$

The first term enforces consensus among the nodes (represents the average information from the rest of the network), while the second accounts for the increase in the total contribution of the local node. Note that there is no need to record the whole design structure of $\xi^{(r)}$ at each node as the information about the current solution is propagated via local estimates of the global FIM, i.e., $M_k^i(\xi^{(r)})$. The whole idea of the communication process is given by Algorithm 1.

Algorithm 1. Distributed optimization of experimental effort. (Indices i and j denote, respectively, data from a local repository and obtained from a neighbor, r is an index of current configuration slot Z_r. The function *EXCHANGE* is responsible for both sending and receiving data to/from connector neighbor (order descending on who initiated communication).)

1: **for** $k \leftarrow 1, K$ **do**
2: EXCHANGE($M_k^j(\xi^{(r)})$, $M_k^i(\xi^{(r)})$) {sends and receives}
3: EXCHANGE(p_{ki} and p_{kj}) {FIM weights}
4: $p \leftarrow p_{ki} + p_{kj}$
5: $M_k^i(\xi^{(r)}) \leftarrow (p_{ki} M_k^i(\xi^{(r)}) + p_{kj} M_k^j(\xi^{(r)}))/p$
6: Calculate $\phi_k(x_k^{\ell}, \xi^{(r)})$ and $\varsigma_k(\xi^{(r)})$, $\ell \in \{i, j\}$
7: Update $p_{ki} \leftarrow p_{ki} \phi_k(x_k^i, \xi^{(r)})/\varsigma_k(\xi^{(r)})$
8: $p_{ki} \leftarrow p \cdot p_i(p_{ki} + p_{kj})$ {normalization p_{ki}'s to sum up to 1}
9: Update $M_k^i(\xi^{(r)}) \leftarrow \frac{r-1}{r} M_k^i(\xi^{(r)}) + \frac{1}{r} \Upsilon_k(x_k^i)$
10: **end for**
11: $r \leftarrow r + 1$

For all subintervals T_k data are calculated separately to reach the most informative configuration for the above-mentioned part of the experiment. At $r = 0$ each network node starts with global FIM estimates $M_k^i(\xi^{(r)})$ initialized with its local information matrix $\Upsilon_k(x_k^i)$ and starting with non-zero weights (this is important since, once the weight attains zero, a multiplicative update forces it to remain zero). Then at each of asynchronous configuration time slots Z_r an appropriate pair of nodes exchanges information according to Algorithm 1. It can be shown that the algorithm converges to an optimal solution, which is a consequence of independent weight updates for each network node and the stochastic convergence of gossip algorithms (Xiao and Boyd, 2004).

Obviously, the sensor network topology and the choice of a proper communication scheme significantly influence the convergence rate. In general, under some assumptions on the network connectivity graph, a suitable gossip algorithm can be provided via semidefinite programming with an exponential convergence rate. Since

a discussion on those very important issues is far beyond the scope of this work, we refer the reader to the seminal papers of Xiao and Boyd (2004), Boyd *et al*. (2006) or Patan and Romanek (2016) for details.

4.3. Reduction of the measurement space. It is clear that the numerical complexity of the communication protocol considered depends linearly on the number of sensor nodes taking part in the averaging. Note that the choice of the initial weights for algorithm is not crucial for the convergence; however, no weights can be equal to zero, because it would then be impossible to change them. This simple observation may lead to a significant improvement, since if during the run of the procedure a weight achieves a value close to zero, it is rather impossible that it will be increased in the next iterations. Deletion of such points from the communication schedule significantly decreases the communication and computational burden.

In order to give more formal justification of this fact, first note that it is easy to check that the sum of weights is invariant with respect to the mapping $\mathcal{T}$. As for interpretation of this function, consider the situation when a design ξ is not optimal. Then the mapping $\mathcal{T}$ increases the weights of those support points of ξ at which the sensitivity function takes high values, i.e., $\phi_k(x,\xi) > \varsigma_k(\xi)$, in such a way decreasing its maximal values. This is attained at the cost of decreasing the weights for support points with small values of $\phi(x,\xi)$ (where $\phi(x,\xi) < \varsigma_k(\xi)$). More precisely, for the case of a multi-output system we can formulate the following extension of the result provided by Prozato (2003) in the context of single response regression model.

Theorem 2. *Let X be some finite design space, $X = \{x^j \in \mathbb{R}^d,\ j = 1,\dots,N\}$, s denote the number of system outputs and $\xi^{(r)}$ be any design measure on X, with*

$$\epsilon_r = \max_{x_k^j \in X} \phi_k(x_k^j, \xi^{(r)}) - \varsigma_k(\xi). \tag{22}$$

Then any point x_k^j such that

$$\phi_k(x_k^j, \xi^{(r)}) < \varsigma_k(\xi) d(\epsilon_k) \tag{23}$$

with

$$d(\epsilon) = 1 + \frac{s\epsilon}{2} - \frac{\sqrt{s\epsilon(4+s\epsilon)}}{2} \tag{24}$$

cannot be a support point of a D-optimum design measure on X.

Proof. See Appendix. ∎

Armed with such a result, when using any algorithm of the D-optimum design, one can remove all the points satisfying (23) from the design space X in each iteration. Clearly, the acceleration that can be expected depends on the employed algorithm and the cardinality of X. Removing support points based on Theorem 2 implies some additional computations. Consequently, the best results can be obtained for a high power of the set X. Nevertheless, this gives us the mathematical justification for removing the non-informative sensor nodes from the communication scheme and directly leads to a significant decrease in pairwise communications between the sensor nodes during the network configuration. In fact, usually in practical implementations, instead of calculating the condition (23), we can resort to a simple heuristic choice and delete from the activation schedule the nodes with negligible weights below some arbitrarily chosen threshold. In most situations, the number of $1/N$, where N is a total number of measurements, is a reasonable choice (Patan, 2012b). This means that, if ϵ_{r} is fixed at $1/N$, no measurements will be taken by a given sensor at a given time instant with a lower weight.

5. Simulation example

As a demonstrative example illustrating the ideas of underlying approach, consider sensor scheduling for parameter estimation of the air pollutant transport process over a given urban area. The spatial domain was normalized to the unit square $\Omega = (0,1)^2$. Inside this area two sources of pollutant are located emitting the pollutant which spreads over the entire area (cf. Fig. 1). Let $y = y(x,t)$ denote the pollutant concentration at a given spatial point at time t belonging to the normalized time interval $T = [0,1)$. Mathematically, the spatio-temporal changes in concentration are described by the following advection–diffusion equation:

$$\frac{\partial y(x,t)}{\partial t} + \nabla \cdot \big(v(x,t)y(x,t)\big) = \nabla \cdot \big(\alpha(x)\nabla y(x,t)\big) + f_1(x) + f_2(x), \tag{25}$$

where $x \in \Omega$ while the boundary and initial conditions are respectively given by

$$\frac{\partial y(x,t)}{\partial n} = 0 \qquad \text{on } \Gamma \times T, \tag{26}$$

$$y(x,0) = y_0 \qquad \text{in } \Omega, \tag{27}$$

Terms $f_\ell(x) = 50\exp(-50||x - c_\ell||^2),\ \ell = 1,2$ represent the pollutant sources located at the positions $c_1(t)$ and $c_2(t)]$ (in general, we allow moving sources, which will be examined in different scenarios) and $\partial y/\partial n$ denote the derivative of y in the direction of the outward normal to the boundary Γ. The average spatio-temporal changes of the wind velocity field over the area were approximated with $v(x,t) = (2(x_1 + x_2 - t), x_2 - x_1 + t)$. The assumed functional form of the spatially-varying diffusion coefficient is

$$\alpha(x) = \theta_1 + \theta_2 x_1 x_2 + \theta_3 x_1^2. \tag{28}$$

Such a class of models using partial differential equations (PDEs) is typically used to describe the dynamics of distributed parameter system, especially in the case of a environmental monitoring, e.g., the smog prediction problem in big agglomerations.

A MATLAB program was written using a PC equipped with an Intel Core i5 processor (1.9 GHz, 12 GB RAM) running Windows 10. The nominal values of the systems parameters were assumed to be $\theta_1^0 = 0.05$, $\theta_2^0 = 0.01$, $\theta_3^0 = 0.005$. Calculations were performed with a finite element method for a spatial mesh composed of 590 triangles, 326 nodes and evenly partitioned time subintervals $T_k = (\frac{k-1}{3}, \frac{k}{3}]$, where $k = 1, 2, 3$. The sensitivity coefficients were then linearly interpolated and stored within the sensor nodes repositories. It was assumed that observation grid consists of only internal points of the finite element mesh, and there are $N = 266$ such internal points in $\Omega \cup \Gamma$ where measurements are supposed to be taken. We aim at finding the weights for sensor nodes at each time subinterval representing the proportions of the total amount of measurements in such a way as to obtain D-optimum least-squares estimates of the parameters θ. It was assumed that the network is fully connected with a uniform probability distribution for the connection between selected pairs of nodes. For a better understanding of the physical phenomena and the sensor configuration process, a set of different scenarios is studied: using two stationary sources (with a fixed position along the time interval), where failures of sensors were implemented, and with two mobile sources.

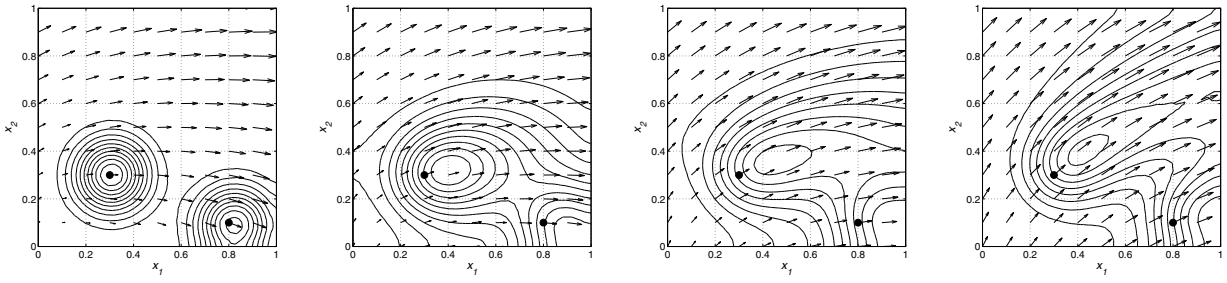
5.1. Scenario 1. The first example take a into account model given by (25)–(27) with two sources with fixed locations, $c_1 = [0.3, 0.3]$ and $c_2 = [0.8, 0.1]$. Temporal changes in wind velocity (arrows) and pollutant concentration with marked sources (black circles) are shown in Fig. 1. Sensor configurations at different stages of the algorithm are presented in Fig. 2. They tend to form a pattern reflecting the areas of the greatest changes in the pollutant concentration, but the observations are averaged over time and it is not trivial to follow the complex process of contaminant proliferation. It can be observed that during each time subinterval the system dynamics enforce the distribution of experimental effort. In the first stage T_1, the pollutant concentration takes significant values only in the close vicinity of the one source located at point c_1, and thus all measurement effort is spent at this location (cf. Figs. 2(a)–(d)). As time elapses, the pollutant spreads over the whole spatial domain. Therefore, the sensors tend to measure the system state in other regions, where we observe the greatest changes in the concentration and gather rich information about the system dynamics (Figs. 2(e)–(h)). At the final stages shown in Figs. 2(i)–(l) two subregions where sources are located are taken as most informative. Additionally, the pattern of the final design was achieved relatively quickly after reaching a reasonably small number of r. Then, the algorithm slows down and achieving a very high accuracy of the solution becomes expensive in the sense of the number of pairwise communications. The convergence of the log-determinant of the FIM for all time intervals is presented in Fig. 3(a)–(c). The reduction of the number of active sensors in each time subinterval along the configuration time is illustrated in Fig 3(d). It is clearly seen that the final number of active nodes at each time subinterval is reduced to approximately 5% of the initial value. This yields a significant reduction of the total computational burden.

Table 1. Active sensor reduction ratio for the first scenario.

Number of sensors	Approx. no. of active sensors T_1	T_2	T_3	Reduction ratio
25	7.7317	8.8022	10.3217	64%
266	11.4235	9.1211	9.3211	96%
1121	98.2131	123.5147	125.8143	90%

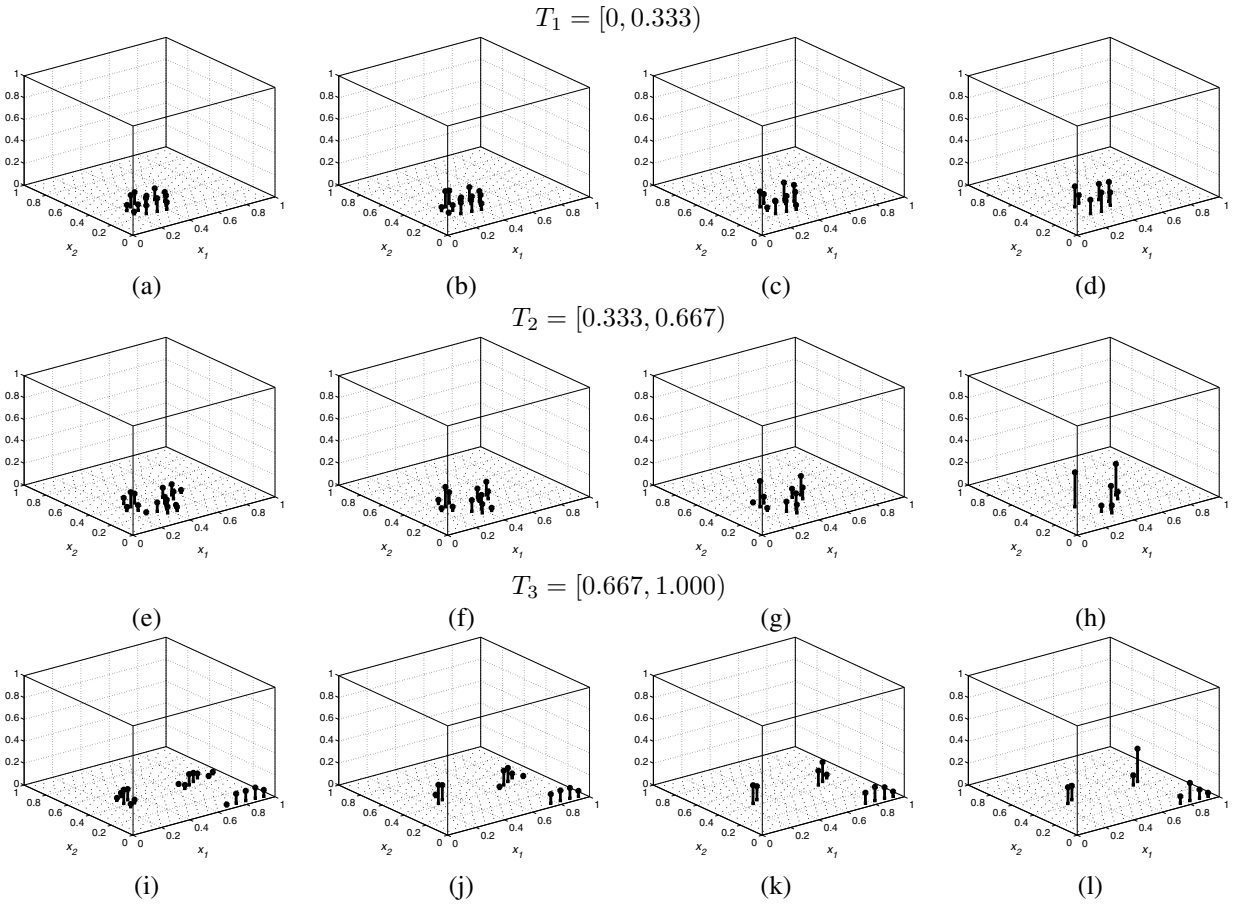
To explore the influence of the spatial discretization on the solutions, additional simulation with a lower number of admissible locations for sensor nodes was done. This time, the number of N potential locations was assumed to be only $N = 25$. During the experiment it was noticed that for this thinner mesh the pattern is somewhat similar for subintervals T_1 and T_2, but for T_3 the most informative sensor locations are distributed along a much larger spatial area than in the case of the denser grid. The number of active sensors was reduced to approximately 36% of all the available ones. The convergence rate of the log-determinant of the FIM is similar, but the quality of the solution is about five times worse in the sense of the criterion. Let us notice that the total minimal number of active measurement points is similar for both networks with 266 and 25 nodes and is equal to about ten active sensors. Intuitively, sensors networks consisting of more elements benefit more during reduction of unnecessary information. A comparison of the efficiency of reduction rate for different numbers sensors nodes in the network is shown in Table 1 (the results are averaged over a few trials of the algorithm). This gives us clear evidence that the reduction rule works better for denser grids.

5.2. Scenario 2. To illustrate the efficacy of the distributed calculation scheme with respect to potential communication failures of individual sensors, the extension of the previous scenario is considered. In separate simulations, two situations were analyzed. First, temporary deactivation of a given random subset of sensors that are not able to communicate and transmit data to other sensors within some given time subinterval (in this case, 40% of sensor nodes was deactivated,



(a) (b) (c) (d)

Fig. 1. Temporal changes in the wind velocity field and pollutant concentration for constant placement of sources for the first scenario: $t = 0$ (a), $t = 0.28$ (b), $t = 0.64$ (c), $t = 1$ (d).

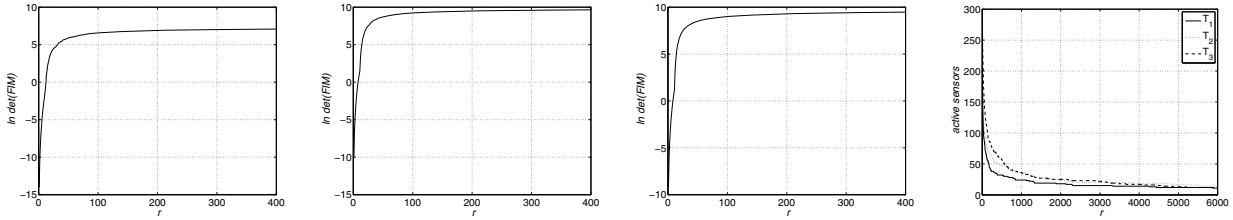


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Fig. 2. Allocation of active sensors with experimental effort (stem plot) in consecutive time subintervals for different numbers of communications r for the first scenario: $r = 500$ (a), $r = 1000$ (b), $r = 2000$ (c), $r = 6000$ (d), $r = 500$ (e), $r = 1000$ (f), $r = 2000$ (g), $r = 6000$ (h), $r = 500$ (i), $r = 1000$ (j), $r = 2000$ (k), $r = 6000$ (l).



(a) (b) (c) (d)

Fig. 3. Convergence (average value along all sensor nodes) for the log-determinant of FIM estimates in time subintervals: $T_1 = [0, 0.333)$ (a), $T_2 = [0.333, 0.667)$ (b), $T_3 = [0.667, 1.000)$ (c), and the number of active sensors during the configuration process (d).

Table 2. Convergence comparison for different scenarios of sensor failures of sensors.

	No. of sensors	Determinant of FIM at T_k		
		T_1	T_2	T_3
Temporary failure	115	4.5440	7.1766	6.8037
	266	7.0751	9.6505	9.4696
	425	8.4501	11.0506	10.8961
Permanent failure	115	4.3953	7.1672	6.7696
	266	7.0548	9.6083	9.3998
	425	8.4442	10.9923	10.8935
No. of failures	115	4.5163	7.1526	6.8243
	266	7.0733	9.6487	9.4586
	425	8.4541	11.0502	10.8948

different subsets in each subinterval). Second, permanent deactivation of a randomly chosen subset of sensors during the whole time of the experiment (again 40%, this time the same subset for any subinterval). Both these situations were compared with Scenario 1 for three sizes of a network: containing 115, 266 and 425 sensors. The determinants of the FIMs calculated for each consecutive subinterval are gathered in Table 2. It is clear that even with such a significant loss of communication resources, despite the sensor failures (permanent or temporal), the distributed calculation scheme converges to virtually the same accuracy (the difference is within no more than 1 % in comparison with Scenario 1). This is strongly consistent with reported results on application of randomized gossip algorithms (Boyd *et al*., 2006).

5.3. Scenario 3. The third simulation takes into account two moving sources. The first pollutant source c_1 is changing the position with a constant speed along linear trajectory starting from $c_1 = [0.3, 0.3]$, finishing at $[0.1, 0.9]$ and $c_2 = [0.8, 0.1]$ moving to position $[0.1, 0.1]$ also at a constant speed. In Figs. 4 and 5 we can observe the simulation results. This time, the sensor allocations correspond to intuition as the pattern follows the movements of sources, at least to some reasonable extent. Convergence and reduction in the number of active sensors is shown in Fig. 6, and results are comparable to previous scenarios where after 100 iterations of configuration time the log-determinant of the FIM is achieves saturation.

6. Concluding remarks

The problem of selection of an optimal spatio-temporal observation schedule in view of accurate parameter estimation in a distributed parameter system has been addressed. The main contribution of this work is a proper mathematical and algorithmic characterization of the underlying problem's relaxation and decomposition. This further allows its adaptation to a distributed scheme of computation and information exchange. The advantage of introducing continuous designs lies in the fact that the problem dimensionality is dramatically reduced. A decided majority of the existing approaches to the sensor location problem are designed only for centralized monitoring systems, and in the context of modern sensor networks often this appears impractical. On the other hand, numerous effective routines for distributed information fusion designed for sensor, peer-to-peer or wireless ad hoc networks are not directly applicable for the purpose of parameter estimation in systems with spatio-temporal dynamics. A decided advantage of the proposed distributed calculation scheme is its great simplicity and the ability to produce solutions reasonably fast.

Naturally, there still remain some open problems which need attention. Potential directions of further research include generalization towards a wider class of design criteria as well as the development of the approaches taking into account the parametric uncertainty via adaptation of the so-called *robust* experimental designs. Additionally, extensions to more sophisticated monitoring systems, i.e., sensor networks with mobile nodes and refinements of the proposed approach taking into account numerous constraints inherent to network design, e.g., energy consumption or cost of deployment, can be considered.

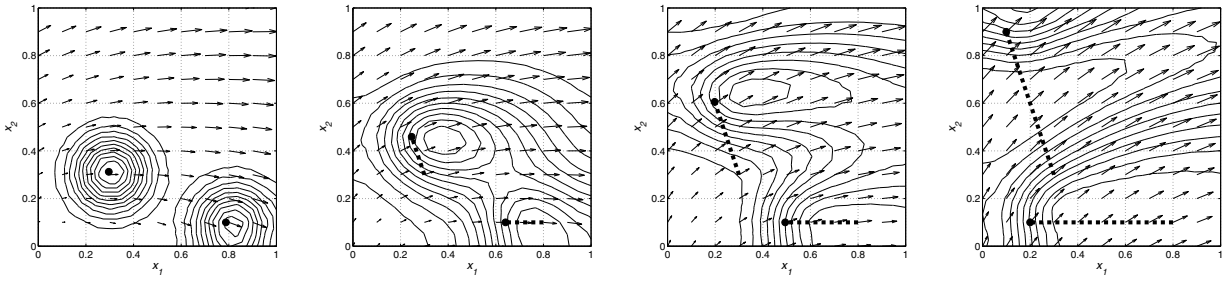


Fig. 4. Temporal changes in the wind velocity field and pollutant concentration for moving sources for the third scenario: $t = 0$ (a), $t = 0.28$ (b), $t = 0.64$ (c), $t = 1$ (d).

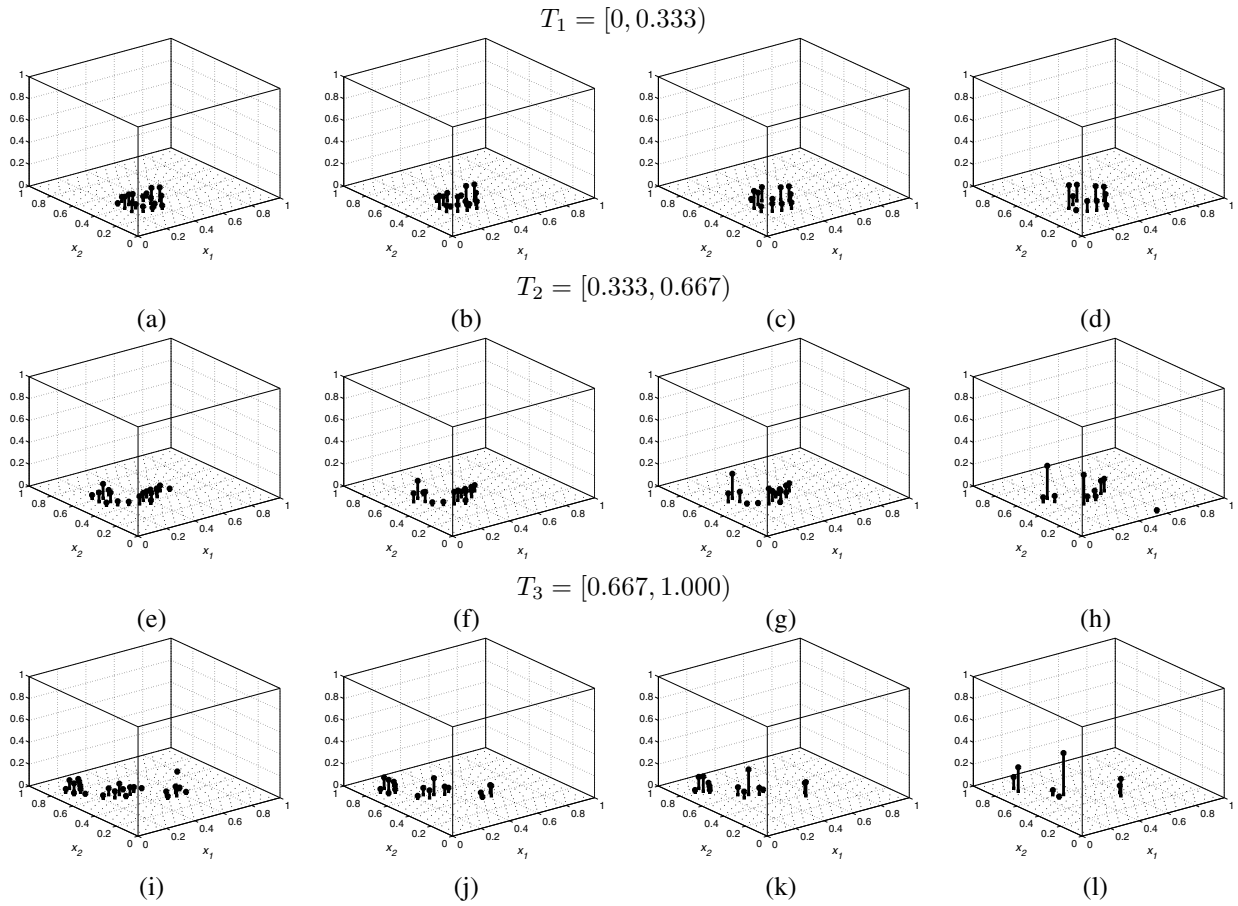


Fig. 5. Allocation of active sensors with experimental effort (stem plot) in consecutive time subintervals for different numbers of communications r for the third scenario: $r = 200$ (a), $r = 500$ (b), $r = 2000$ (c), $r = 6000$ (d), $r = 200$ (e), $r = 500$ (f), $r = 2000$ (g), $r = 6000$ (h), $r = 200$ (i), $r = 500$ (j), $r = 2000$ (k), $r = 6000$ (l).

Acknowledgment

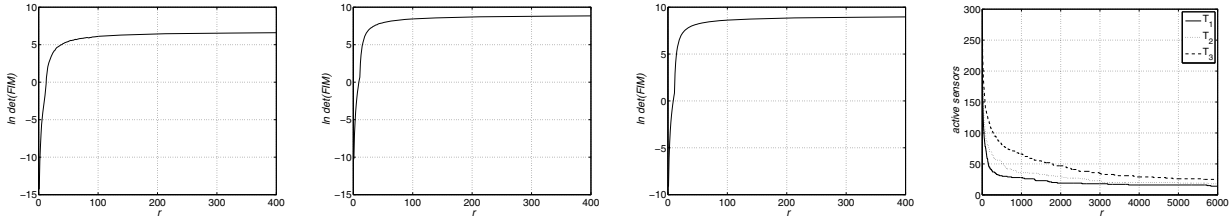
This work is supported by the National Science Center in Poland under the grant no. 2014/15/B/ST7/03208.

References

Alifanov, O.M., Artyukhin, E.A. and Rumyantsev, S.V. (1995). *Extreme Methods for Solving Ill Posed Problems with Applications to Inverse Heat Transfer Problems*, Begell House, New York, NY.

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

Boukerche, A. (2006). *Handbook of Algorithms for Wireless Networking and Mobile Computing*, Chapman & Hall/CRC, Boca Raton, FL.



(a) (b) (c) (d)

Fig. 6. Convergence (average value) for log-determinant of FIM estimates in time subintervals: $T_1 = [0, 0.333)$ (a), $T_2 = [0.334, 0.667)$ (b), $T_3 = [0.667, 1.000)$ (c), and the number of active sensors during exchange of information for the third scenario (d).

Boyd, S., Ghosh, A., Prabhakar, B. and Shah, D. (2006). Randomized gossip algorithms, *IEEE Transactions on Information Theory* **52**(6): 2508–2530.

Braca, P., Marano, S. and Matta, V. (2008). Enforcing consensus while monitoring the environment in wireless sensor network, *IEEE Transactions on Signal Processing* **56**(7): 3375–3380.

Cacuci, D.G., Navon, I.M. and Ionescu-Bujor, M. (2014). *Computational Methods for Data Evaluation and Assimilation*, CRC Press, Boca Raton, FL.

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

Demetriou, M.A. (2010). Guidance of mobile actuator-plus-sensor networks for improved control and estimation of distributed parameter systems, *IEEE Transactions on Automatic Control* **55**(7): 1570–1584.

Demetriou, M.A. and Hussein, I.I. (2009). Estimation of spatially distributed processes using mobile spatially distributed sensor network, *SIAM Journal on Control and Optimization* **48**(1): 266–291.

Fedorov, V.V. and Hackl, P. (1997). *Model-Oriented Design of Experiments*, Lecture Notes in Statistics, Vol. 125, Springer-Verlag, New York, NY.

Holder, D.S. (Ed.) (2004). *Electrical Impedance Tomography: Methods, History and Applications*, Taylor & Francis, Philadelphia, PA.

Jacobson, M.Z. (1999). *Fundamentals of Atmospheric Modeling*, Cambridge University Press, Cambridge.

Jain, N. and Agrawal, D.P. (2005). Current trends in wireless sensor network design, *International Journal of Distributed Sensor Networks* **1**(1): 101–122.

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

Kowalów, D., Patan, M., Paszke, W. and Romanek, A. (2015). Sequential design for model calibration in iterative learning control of DC motor, *20th International Conference on Methods and Models in Automation and Robotics, Międzyzdroje, Poland*, pp. 794–799.

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

Nehorai, A., Porat, B. and Paldi, E. (1995). Detection and localization of vapor-emitting sources, *IEEE Transactions on Signal Processing* **43**(1): 243–253.

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.

Patan, K. and Patan, M. (2010). Selection of training data for locally recurrent neural network, *in* K. Diamantaras *et al.* (Eds.), *Artificial Neural Networks—ICANN 2010*, Lecture Notes in Computer Science, Vol. 6353, Springer, Berlin/Heidelberg, pp. 134–137.

Patan, M. (2006). Optimal activation policies for continuous scanning observations in parameter estimation of distributed systems, *International Journal of Systems Science* **37**(11): 763–775.

Patan, M. (2008). A parallel sensor scheduling technique for fault detection in distributed parameter systems, *in* E. Luque *et al.* (Eds.), *Parallel Processing—Euro Par 2008*, Lecture Notes in Computer Science, Vol. 5168, Springer, Berlin/Heidelberg, pp. 833–843.

Patan, M. (2012a). Distributed scheduling of sensor networks for identification of spatio-temporal processes, *International Journal of Applied Mathematics and Computer Science* **22**(2): 299–311, DOI: 10.2478/v10006-012-0022-9.

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

Patan, M. and Bogacka, B. (2007). Optimum experimental designs for dynamic systems in the presence of correlated errors, *Computational Statistics & Data Analysis* **51**(12): 5644–5661.

Patan, M., Chen, Y. and Tricaud, C. (2008). Resource-constrained sensor routing for parameter estimation of distributed systems, *17th IFAC World Congress, Seoul, South Korea,* pp. 7772–7777.

Patan, M. and Kowalów, D. (2014). Robust sensor scheduling via iterative design for parameter estimation of distributed systems, *19th International Conference on Methods*

and Models in Automation and Robotics, Międzyzdroje, Poland, pp. 618–623.

Patan, M. and Romanek, A. (2016). Communication scheduling for fast distributed averaging in sensor networks, *2016 International Symposium on Wireless Communication Systems, Poznań, Poland*, pp. 109–113.

Patan, M. and Uciński, D. (2010). Time-constrained sensor scheduling for parameter estimation of distributed systems, *49th IEEE Conference on Decision and Control, Atlanta, GA, USA*, pp. 7–12.

Patan, M. and Uciński, D. (2016a). Cost-constrained D-optimum node activation for large-scale monitoring networks, *IEEE 2016 American Control Conference (ACC), Boston, MA, USA*, pp. 1643–1648.

Patan, M. and Uciński, D. (2016b). D-optimal spatio-temporal sampling design for identification of distributed parameter systems, *55th IEEE Conference on Decision and Control, Las Vegas, NV, USA*, pp. 3985–3990.

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

Point, N., Wouwer, A.V. 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 designs algorithms, *Statistics and Probability Letters* **63**(3): 223–228.

Rafajłowicz, E. (1986). Optimum choice of moving sensor trajectories for distributed-parameter system identification, *International Journal of Control* **43**(5): 1441–1451.

Silvey, S.D. (1980). *Optimal Design. An Introduction to the Theory for Parameter Estimation*, Chapman & Hall, London.

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* **7**(14): 1379–1389.

Song, Z., Chen, Y., Sastry, C.R. and Tas, N.C. (2009). *Optimal Observation for Cyber-Physical Systems: A Fisher-Information-Matrix-Based Approach*, Springer-Verlag, London.

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

Torsney, B. and Mandal, S. (2004). Multiplicative algorithms for constructing optimizing distributions: Further developments, *in* A. Di Bucchianico *et al.* (Eds), *mODa 7: 7th International Workshop on Model-Oriented Data Analysis*, Physica-Verlag, Heidelberg, pp. 163–171.

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

Tricaud, C., Patan, M., Uciński, D. and Chen, Y. (2008). D-optimal trajectory design of heterogeneous mobile sensors for parameter estimation of distributed systems, *American Control Conference, Seattle, WA, USA*, pp. 663–668.

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

Uciński, D. (2004). *Optimal Measurement Methods for Distributed Parameter System Identification*, CRC Press, Boca Raton, FL.

Uciński, D. (2012). Sensor network scheduling for identification of spatially distributed processes, *International Journal of Applied Mathematics and Computer Science* **22**(1): 25–40, DOI: 10.2478/v10006-012-0002-0.

Uciński, D. and Patan, M. (2002). Optimal location of discrete scanning sensors for parameter estimation of distributed systems, *15th IFAC World Congress, Barcelona, Spain*, pp. 22–26.

Uciński, D. and Patan, M. (2007). D-optimal design of a monitoring network for parameter estimation of distributed systems, *Journal of Global Optimization* **39**(2): 291–322.

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

Wu, J. (Ed.) (2006). *Handbook on Theoretical and Algorithmic Aspects of Sensor, Ad Hoc Wireless, and Peer-to-Peer Networks*, Auerbach Publications/Taylor & Francis Group, Boca Raton, FL.

Xiao, L. and Boyd, S. (2004). Fast linear iterations for distributed averaging, *Systems & Control Letters* **53**(1): 65–78.

Zhao, F. and Guibas, L.J. (2004). *Wireless Sensor Networks: An Information Processing Approach*, Morgan Kaufmann, Amsterdam.

Zhong, M. and Cassandras, C.G. (2011). Distributed coverage control and data collection with mobile sensor networks, *IEEE Transactions on Automatic Control* **56**(10): 2445–2455.

Maciej Patan was born in 1975. He received his MSc and PhD degrees, both in electrical engineering, from the University of Zielona Góra, Poland, in 2000 and 2004, respectively. From 2005 to 2006 he was a research assistant in statistics at the School of Mathematical Sciences, Queen Mary, University of London. He has been an IEEE member since 2009. Since 2014 he has been an associated professor with the Faculty of Computer, Electrical and Control Engineering, University of Zielona Góra. His research interests include optimal experimental design in distributed parameter systems, distributed computing and optimization in sensor networks, system identification and mobile robotics.

Damian Kowalów was born in 1988. He received his MSc degree in automatic control and robotics in 2013 from the University of Zielona Góra. Currently he is a PhD student there. His research interests include iterative learning control, modeling and control of spatio-temporal systems and sensor selection for identification problems.

Appendix

A1. Proof of Theorem 1

To prove the discussed necessary and sufficient conditions for optimality of designs, we first have to derive some auxiliary assertions.

Lemma A1. *For any design* $\xi = (\xi_1, \ldots, \xi_K) \in \Xi(X)$ *and all* $k = 1, \ldots, K$, *we have*

(i) $\int_X \phi_k(x,\xi)\,\xi_k(\mathrm{d}x) = \varsigma_k(\xi)$, *and*

(ii) $\max_{x\in X} \phi_k(x,\xi) \geq \varsigma_k(\xi)$.

Proof. From (14) and (15), we have

$$\begin{aligned}\int_X \phi_k(x,\xi)\,\xi_k(\mathrm{d}x) &= \int_X \operatorname{trace}\left[M^{-1}(\xi)\Upsilon_k(x)\right]\xi_k(\mathrm{d}x)\\ &= \operatorname{trace}\left[M^{-1}(\xi)\int_X \Upsilon_k(x)\,\xi_k(\mathrm{d}x)\right]\\ &= \operatorname{trace}\left[M^{-1}(\xi)M_k(\xi_k)\right] = \varsigma_k(\xi).\end{aligned} \tag{A1}$$

This asserts *(i)*.

As for the second part, we have

$$\begin{aligned}\varsigma_k(\xi) &= \int_X \phi_k(x,\xi)\,\xi_k(\mathrm{d}x)\\ &\leq \int_X \max_{x\in X}\phi_k(x,\xi)\,\xi_k(\mathrm{d}x) = \max_{x\in X}\phi_k(x,\xi),\end{aligned} \tag{A2}$$

which proves *(ii)*. ∎

Lemma A2. *If* $\xi = (\xi_1, \ldots, \xi_K) \in \tilde{\Xi} = \{\xi : \Psi[M(\xi)] < \infty\} \neq \emptyset$, $\bar{\xi} = (\bar{\xi}_1, \ldots, \bar{\xi}_K) \in \Xi(X)$ *and* $\xi_\alpha = (1-\alpha)\xi + \alpha\bar{\xi}$, *then*

$$\left.\frac{\partial\Psi[M(\xi_\alpha)]}{\partial\alpha}\right|_{\alpha=0^+} = \sum_{k=1}^K \int_X \phi_k(x,\xi)\,\bar{\xi}_k(\mathrm{d}x) - \varsigma(\xi),$$

where

$$\varsigma(\xi) = \sum_{k=1}^K \varsigma_k(\xi).$$

Proof. Taking into account (14) and (15), we have

$$\begin{aligned}\left.\frac{\partial\Psi[M(\xi_\alpha)]}{\partial\alpha}\right|_{\alpha=0^+} &= \left.\frac{\partial\Psi[(1-\alpha)M(\xi)+\alpha M(\bar{\xi})]}{\partial\alpha}\right|_{\alpha=0^+}\\ &= \operatorname{trace}\left\{M^{-1}(\xi)[M(\bar{\xi})-M(\xi)]\right\}\\ &= \sum_{k=1}^K \int_X \left\{\operatorname{trace}\left[M^{-1}(\xi)\Upsilon_k(x)\right]\right\}\\ &\quad - \operatorname{trace}\left\{\left[M^{-1}(\xi)M_k(\xi_k)\right]\right\}\bar{\xi}(\mathrm{d}x)\\ &= \sum_{k=1}^K \int_X \phi_k(x,\xi)\,\bar{\xi}_k(\mathrm{d}x) - \sum_{k=1}^K \varsigma_k(\xi).\end{aligned} \tag{A3}$$

∎

Now, we are in position to prove Theorem 1. First, introduce the design $\xi_\alpha = (1-\alpha)\xi^\star + \alpha\bar{\xi}$, where $\xi^\star = (\xi_1^\star, \ldots, \xi_K^\star) \in \tilde{\Xi}$, and $\bar{\xi} = (\bar{\xi}_1, \ldots, \bar{\xi}_K) \in \Xi(X)$.

(i) $\Rightarrow$ *(ii)* If $\xi^\star = \arg\min_{\xi\in\Xi(X)} \Psi[M(\xi)]$, then

$$\left.\frac{\partial\Psi[M(\xi_\alpha)]}{\partial\alpha}\right|_{\alpha=0^+} \geq 0, \quad \forall\bar{\xi}\in\Xi(X). \tag{A4}$$

From Lemma A2, for any $i = 1, \ldots, K$, setting $\xi = \xi^\star$ and $\bar{\xi}$ such that

$$\bar{\xi}_k = \begin{cases} \xi_k^\star & \text{if } k \neq i,\\ \{(x,1)\} & \text{if } k = i,\end{cases} \tag{A5}$$

we obtain

$$\begin{aligned}\left.\frac{\partial\Psi[M(\xi_\alpha)]}{\partial\alpha}\right|_{\alpha=0^+} &= \sum_{k=1}^K \varsigma_k(\xi^\star)\\ &\quad - \sum_{\substack{k=1\\k\neq i}}^K \int_X \phi_k(x,\xi^\star)\,\xi_k^\star(\mathrm{d}x) - \phi_i(x,\xi^\star)\\ &= \sum_{k=1}^K \varsigma_k(\xi^\star) - \sum_{k=1}^K \int_X \phi_k(x,\xi^\star)\,\xi_k^\star(\mathrm{d}x)\\ &\quad - \phi_i(x,\xi^\star) + \int_X \phi_i(x,\xi^\star)\,\xi_i^\star(\mathrm{d}x)\\ &= \varsigma_i(\xi^\star) - \phi_i(x,\xi^\star) \geq 0.\end{aligned} \tag{A6}$$

The result is

$$\varsigma_i(\xi^\star) \geq \phi_i(x,\xi^\star) \geq \max_{x\in X}\phi_i(x,\xi^\star). \tag{A7}$$

Part *(ii)* of Lemma A1 now yields the desired conclusion.

(ii) $\Rightarrow$ *(i)* Let $\xi^\star \in \Xi(X)$ satisfy (ii). We have

$$\begin{aligned}\frac{\partial \Psi[M(\xi_\alpha)]}{\partial \alpha}\bigg|_{\alpha=0^+} &= \sum_{k=1}^{K}\Big[\varsigma_k(\xi^\star) - \int_X \phi_k(x,\xi^\star)\,\bar{\xi}_k(\mathrm{d}x)\Big] \\ &\geq \sum_{k=1}^{K}\big[\varsigma_k(\xi^\star) - \max_{x\in X}\phi_k(x,\xi^\star)\big] = 0,\end{aligned} \tag{A8}$$

which completes the proof.

A2. Proof of Theorem 2

In order to derive the main result, first we need to claim some useful lemmas.

Lemma A3. *Let A be a symmetric matrix of size s. Then*

$$\big(\operatorname{trace}(A)\big)^2 \leq s\operatorname{trace}(A^2). \tag{A9}$$

Proof. Denote by $\lambda_1,\ldots,\lambda_s$ a sequence of eigenvalues of matrix A. The inequality considered can be rewritten as

$$\Big(\sum_{i=1}^{s}\lambda_i\Big)^2 \leq s\sum_{i=1}^{s}\lambda_i^2. \tag{A10}$$

This is equivalent to

$$\sum_{k=1}^{s}\sum_{i=1}^{s}\lambda_i\lambda_k \leq \sum_{k=1}^{s}\sum_{i=1}^{s}\lambda_i^2, \tag{A11}$$

and further to

$$\begin{aligned}0 &\leq \frac{1}{2}\sum_{k=1}^{s}\sum_{i=1}^{s} -\frac{1}{2}\sum_{k=1}^{s}\sum_{i=1}^{s}2\lambda_i\lambda_k \\ &= \frac{1}{2}\sum_{(i,k)}\Big(\lambda_i^2+\lambda_k^2-2\lambda_i\lambda_k\Big) = \frac{1}{2}\sum_{(i,k)}(\lambda_i-\lambda_k)^2,\end{aligned} \tag{A12}$$

which completes the proof. ∎

Lemma A4. *Let A and B be any matrices of the same size. Then*

$$\operatorname{trace}(AB^{\mathsf{T}}AB^{\mathsf{T}}) \leq \operatorname{trace}(AA^{\mathsf{T}}BB^{\mathsf{T}}). \tag{A13}$$

Proof. Using the cyclic permutation rule, we can rewrite the inequality in the form

$$\operatorname{trace}(B^{\mathsf{T}}AB^{\mathsf{T}}A) \leq \operatorname{trace}(B^{\mathsf{T}}AA^{\mathsf{T}}B). \tag{A14}$$

Hence, introducing matrix $C = B^{\mathsf{T}}A$, we obtain

$$\operatorname{trace}(C^2) \leq \operatorname{trace}(CC^{\mathsf{T}}). \tag{A15}$$

Since for every square matrix $\operatorname{trace}(C) = \operatorname{trace}(C^{\mathsf{T}})$, we have

$$\begin{aligned}&2\Big(\operatorname{trace}(CC^{\mathsf{T}}) - \operatorname{trace}(C^2)\Big) \\ &\quad= \operatorname{trace}(CC^{\mathsf{T}} + C^{\mathsf{T}}C - CC - C^{\mathsf{T}}C^{\mathsf{T}}) \\ &\quad= \operatorname{trace}\big\{(C-C^{\mathsf{T}})(C^{\mathsf{T}}-C)\big\} \\ &\quad= -\operatorname{trace}(C^{\mathsf{T}}-C)^2 \geq 0.\end{aligned} \tag{A16}$$

The last inequality is a result of the fact that $C^{\mathsf{T}} - C$ is a skew-symmetric matrix. Hence its has purely imaginary eigenvalues. Therefore, its square has to be semi-negative definite, which establishes our claim. ∎

Lemma A5. *Let $B \in \mathbb{S}_+^n$. Then, for any matrix Y,*

$$\operatorname{trace}(Y^{\mathsf{T}}BY) \leq \operatorname{trace}(Y^{\mathsf{T}}Y)\operatorname{trace}(B) \tag{A17}$$

Proof. For any non-negative definite matrices A and B, we have

$$\operatorname{trace}(AB) \leq \operatorname{trace}(A)\operatorname{trace}(B).$$

Then

$$\begin{aligned}\operatorname{trace}(Y^{\mathsf{T}}BY) &= \operatorname{trace}(YY^{\mathsf{T}}B) \\ &\leq \operatorname{trace}(YY^{\mathsf{T}})\operatorname{trace}(B) \\ &= \operatorname{trace}(Y^{\mathsf{T}}Y)\operatorname{trace}(B).\end{aligned} \tag{A18}$$

∎

Lemma A6. *Let $\mathfrak{M}$ be some convex set of $m\times m$ positive definite matrices and $R, Q \in \mathfrak{M}$, where*

$$Q = \arg\max_{M\in\mathfrak{M}} \log\det M. \tag{A19}$$

Define $r = \operatorname{trace}(G^{\mathsf{T}}P^{-1}G)$, $q = \operatorname{trace}(G^{\mathsf{T}}Q^{-1}G)$ with G being some matrix of $\mathbb{R}^{m\times s}$, and let ϵ be some positive real value. If

$$\operatorname{trace}\{R^{-1}(Q-R)\} \leq \epsilon, \tag{A20}$$

then

$$qd(\epsilon) \leq r, \tag{A21}$$

where

$$d(\epsilon) = 1 + \frac{s\epsilon}{2} - \frac{1}{2}\sqrt{s\epsilon(s\epsilon+4)}.$$

Proof. We start with

$$\begin{aligned}r - q &= \operatorname{trace}\left\{G^{\mathsf{T}}R^{-1}G\right\} - \operatorname{trace}\left\{G^{\mathsf{T}}Q^{-1}G\right\} \\ &= \operatorname{trace}\left\{G^{\mathsf{T}}R^{-1}(Q-R)Q^{-1}G\right\}.\end{aligned} \tag{A22}$$

Since R and Q are positive definite, it is possible to introduce the following decompositions:

$$R^{-1} = UU^{\mathsf{T}}, \quad Q^{-1} = VV^{\mathsf{T}}. \tag{A23}$$

In addition to this, define Z such that $Z = V^{\mathsf{T}}(Q - R)R^{-1}G$. Then, using Lemmas A3 and A4, we have

$$\begin{aligned}(r-q)^2 &= \left(\operatorname{trace}\left\{G^{\mathsf{T}}R^{-1}(Q-R)Q^{-1}G\right\}\right)^2 \\ &\le s \operatorname{trace}\left\{\left[G^{\mathsf{T}}R^{-1}(Q-R)Q^{-1}G\right]^2\right\} \\ &= s \operatorname{trace}\left\{(Z^{\mathsf{T}}V^{\mathsf{T}}G)^2\right\} \\ &\le s \operatorname{trace}\left\{Z^{\mathsf{T}}ZG^{\mathsf{T}}Q^{-1}G\right\} \\ &\le s \operatorname{trace}\left\{Z^{\mathsf{T}}Z\right\} \operatorname{trace}\left\{G^{\mathsf{T}}Q^{-1}G\right\}.\end{aligned} \tag{A24}$$

Further, according to Lemma A5, we obtain

$$\begin{aligned}&\operatorname{trace}\left(Z^{\mathsf{T}}Z\right) \\ &\quad= \operatorname{trace}\left\{G^{\mathsf{T}}UU^{\mathsf{T}}(Q-R)Q^{-1}(Q-R)UU^{\mathsf{T}}G\right\} \\ &\quad\le \operatorname{trace}\left\{G^{\mathsf{T}}R^{-1}G\right\} \\ &\qquad\times \operatorname{trace}\left\{U^{\mathsf{T}}(Q-R)Q^{-1}(Q-R)U\right\} \\ &\quad= \operatorname{trace}\left\{G^{\mathsf{T}}R^{-1}G\right\} \\ &\qquad\times \operatorname{trace}\left\{(R^{-1}-Q^{-1})(Q-R)\right\}.\end{aligned} \tag{A25}$$

In such a way,

$$\begin{aligned}(r-q)^2 &\le s \operatorname{trace}\left\{G^{\mathsf{T}}Q^{-1}G\right\} \operatorname{trace}\left\{G^{\mathsf{T}}R^{-1}G\right\} \\ &\qquad\times \operatorname{trace}\left\{(R^{-1}-Q^{-1})(Q-R)\right\} \\ &= srq \operatorname{trace}\left\{(R^{-1}-Q^{-1})(Q-R)\right\}.\end{aligned} \tag{A26}$$

At this point, introduce $M_\alpha = (1-\alpha)R + \alpha Q, \ \ \alpha \in [0,1]$ and the function $\Psi(\alpha) = \log\det M_\alpha$. Its Fréchet derivative is given by

$$\overset{\circ}{\Psi}(\alpha) = \operatorname{trace}\{M_\alpha^{-1}(Q-R)\}. \tag{A27}$$

Further,

$$\overset{\circ}{\Psi}(\alpha)\Big|_{\alpha=0^+} = \operatorname{trace}\{R^{-1}(Q-R)\} \le \epsilon$$

and $\Psi(\alpha) \ge \Psi(1) = \log\det Q$. This, taken in conjunction with the concavity of Ψ, implies $0 \le \overset{\circ}{\Psi}(1) \le \overset{\circ}{\Psi}(0) \le \epsilon$. Finally,

$$\epsilon \ge \overset{\circ}{\Psi}(1) - \overset{\circ}{\Psi}(0) = \operatorname{trace}\left\{(R^{-1}-Q^{-1})(Q-R)\right\}. \tag{A28}$$

Combination of (A26) and (A28) gives

$$(p-q)^2 \le \epsilon srq. \tag{A29}$$

Solving this quadratic inequality with respect to r and taking the lower bound, we get

$$q\left(1 + \frac{s\epsilon}{2} - \frac{1}{2}\sqrt{s\epsilon(s\epsilon+4)}\right) \le r, \tag{A30}$$

which completes the proof. ■

Assume that $\xi^\star$ is a D-optimum design measure; then, from Lemma A1 and definition of $\epsilon_k^{(r)}$, we have

$$\begin{aligned}\epsilon_k^{(r)} &\ge \int_X \phi_k(x,\xi^{(r)})\,\xi^\star(\mathrm{d}x) - \varsigma_k(\xi^{(r)}) \\ &= \operatorname{trace}\left\{M^{-1}(\xi^{(r)})\left[M(\xi^\star) - M(\xi^{(r)})\right]\right\}.\end{aligned} \tag{A31}$$

Then, the assertion of Theorem 2 follows immediately from Lemma A6. In fact, taking $R = M(\xi^{(r)})$ and $Q = M(\xi^\star)$ in Lemma A6, in conjunction with the claim *(ii)* of Theorem 1, for any support point x_k^j in $\xi^\star$, we obtain

$$d(\epsilon_k^{(r)})\varsigma_k(\xi^\star) \le \phi_k(x_k^j, \xi^{(k)}),$$

which constitutes the desired result.

Received: 31 January 2017
Revised: 17 July 2017
Accepted: 9 August 2017