Efficient Sensing Topology Management for Spatial Monitoring with Sensor Networks

Periklis Liaskovitis · Curt Schurgers

Received: 9 February 2009 / Revised: 3 March 2010 / Accepted: 5 March 2010 / Published online: 13 April 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract We focus on exploiting redundancy for sensor networks in the context of spatial interpolation. The network acts as a distributed sampling system, where sensors periodically sample a physical phenomenon of interest, e.g. temperature. Samples are then used to construct a continuous spatial estimate of the phenomenon over time through interpolation. In this regime, the notion of sensing range typically utilized to characterize redundancy in event detection applications is meaningless and sensor selection schemes based on it become unsuitable. Instead, this paper presents pragmatic approaches for exploiting redundancy in such applications. Their underlying characteristic is that no a-priori assumptions need to be made on the statistical properties of the physical phenomenon. These are instead learned by the network after deployment. Our approaches are evaluated through real as well as synthetic sensor network data showing that significant reductions in the number of active sensors are indeed possible.

Keywords Sensor networks · Spatial monitoring · Sampling of random processes · Sensor selection · Sensing topology management · Energy efficiency · Hilbert space

1 Introduction

Envisioned large sensor deployments will consist of cheap devices, small in size, seamlessly sensing physical phenomena, processing readings and communicating data to user applications. Deployment may target remote regions for environmental and wildlife monitoring or adversarial

regions for military purposes. Exact preplanning and post-deployment alterations of sensor locations in these networks are difficult because of their sheer volume [5].

Lack of prior knowledge about physical phenomena of interest also hinders exact characterization of the sensing behavior of devices before deployment. After deployment, it may well be the case that some of the deployed sensors are redundant from an application point of view. Consider the example of a network monitoring temperature inside a warehouse. For an *event detection* application the objective would be to determine ‘when and where’ an event occurs, i.e., when and where an outlier temperature exists in the sensor readings. Any point in space can potentially be the source of an event, so the application demands that a minimum number of sensors k be sensing each point in space. This is well known as k -coverage [1, 31]. To determine associated redundancy, the intuitive notion of *sensing range* is used, as the maximum distance from a sensor over which an event can be detected reliably [3, 27].

On the other hand, we focus on a different application domain, namely *spatial interpolation*. In the temperature measuring scenario for example, it would be desirable to know how temperature varies as one is moving away from windows or air-conditioning units. The goal is to construct a continuous temperature surface, i.e., even for areas where no sensors exist, by viewing sensor measurements as spatio-temporal samples. Our focus on interpolation is an important one, since existing literature has studied reliability and spatial redundancy notions mainly in the context of aforementioned event detection and sensing range. By contrast, interpolation combines collected samples in a nontrivial way and sensing range can no longer capture associated redundancy. Our ultimate goal is to lay foundations for increasing the operational lifetime of the network: obtained interpolation should be able to consistently meet certain quality criteria for as long as possible.

Spatial redundancy in a sampling-interpolation application essentially means that only a few of the gathered

P. Liaskovitis (✉) · C. Schurgers
ECE Department, University of California,
San Diego, CA, USA
e-mail: pliaskov@ucsd.edu

C. Schurgers
e-mail: curts@ece.ucsd.edu

samples are needed to provide a sufficiently good interpolation. By having only the corresponding subset of sensors communicating data packets, overall energy consumption can be decreased. To obtain lifetime gains, multiple disjoint subsets of sensors should be devised, so that data from each of them individually, achieves the desired interpolation fidelity [15, 18, 19, 27]. At each point in time only one such set is made active. Figure 1 shows interpolation being performed sequentially with three different sets of sensors. In this case, the network can be thought of as three superimposed sub-networks, all equivalent with respect to interpolation. Ideally, this leads to a three-fold increase in network lifetime (for a more detailed discussion of energy savings, see Section 3).

In this paper, we propose strategies for *sensing topology management* in the context of our particular application: how to devise multiple sets of sensors, each capable of interpolation that is accurate enough, according to some well defined criterion. Our specific contributions are:

- Proof that the sensing topology management problem for interpolation applications is exponentially hard with the size of the network, thus necessitating heuristic approaches.
- Two heuristic schemes for devising disjoint sets of sensors. The first one, the *Jittered Grid (JiG)* algorithm is based on classical sampling theory notions. The second one, the *Random Variable Greedy (RaVaG)* algorithm, is based on viewing sensors as a Hilbert space of random variables.
- Evaluation of the proposed schemes on both real and synthetic sensor data, which shows that significant reductions in the number of active sensors can be achieved compared to simpler selection methods.

2 Relation to Previous Work

There exists extensive work on sensing topology management for sensor networks, mostly in an event detection context. Many authors have proposed resilient protocols to maintain k -coverage for any point in the observation field at all times [2, 3]. Set k -cover algorithms [1, 27] aim at obtaining subsets of sensors each capable of k -coverage to increase network lifetime. All these approaches define

Figure 1 Rotating sets of active sensors.

coverage on the basis of a circular sensing range, which bears no physical meaning for a spatial interpolation application. The authors in [30] have used correlation regions of roughly equally informative sensors to partition the network, based on a Voronoi tessellation. Their work still only targets an event driven type of application, namely the estimation of a single point Gaussian source, which is different from constructing an entire spatial surface. Our work can be thought of as analogous to set k -cover, for the distinct regime of sampling-interpolation applications.

Numerous efforts on efficiently building a spatial model for the sensed phenomenon also exist in the literature [12, 24, 34]. They are not applicable in our scenario, since they make some kind of assumption on the underlying phenomenon. Temporal redundancy in sampling type applications has been considered in compressive sensing [9, 11] where all sensors report compressed data. Our approach exploits spatial redundancy instead and reduces the number of packets produced by the network as opposed to packet sizes. Spatial redundancy has previously been considered, but from a significantly different point of view. Koushanfar et al. [15] pair up sensors, so that the readings from one can be used to predict those of the other. However, they do not study the effect on the overall interpolation quality, but only consider predicting the value at a specific sensor location. Our own previous work includes devising sets in a sampling-interpolation setting for one-dimensional wide sense stationary physical processes through jittered grids [18], as well as introducing the Hilbert space framework of sensors as random variables [19]. Here, we additionally present a proof of why optimal redundancy elimination in a sampling-interpolation setting is a hard combinatorial problem, with complexity growing exponentially with the size of the network. Furthermore, proposed algorithms can now handle non-stationary behavior in space, i.e., statistics that change over the network area. Finally, unlike our previous work, this paper includes an evaluation of our approaches for spatially non-stationary monitored processes in the presence of noise.

A relevant theme from geostatistics [35] is sampling design, i.e., finding the best locations to sample, out of a finite set of possible locations. In sensor networks it has been examined as optimum sensor placement [16]. However, it suffers from a ‘learning’ disadvantage: either specific statistics for the physical phenomenon have to be assumed or sensors have to be deployed in a ‘test-placement’ to gather statistical data before being optimally (re)deployed. This type of post-processing cannot be applied in our setting where sampling positions are restricted to initial ones. Statistics are learned online. Specifically compared to optimum sensor placements [16], we use only the covariance matrix at sensor locations instead of estimating the continuous covariance function (see Section 4).

Devising subsets of vectors with our *RaVaG* scheme is also related to the topic of sparse approximation of a single [7, 10, 25, 32] or multiple vectors [29, 33] with redundant dictionaries. Existing methods operate in a different Hilbert space, namely the space of finite length real vectors equipped with the usual Euclidean inner product. In our scenario however, the signal space, i.e., the data matrix B (see Eq. (7)), would coincide with the dictionary space, rendering such methods inefficient. Put another way, existing methods would try to sparsely approximate the data matrix B itself, instead of directly minimizing the generalization error of such an approximation [32]. By contrast, our own *RaVaG* algorithm uses the data matrix B only for estimation of inner products in the Hilbert space of second order random variables.

Subset selection in linear regression is another topic related to *RaVaG*, where a response random variable has to be approximated by a subset of available predictor variables. Heuristics have been studied for the general covariance case [20]. Recently, algorithms with performance guarantees appeared for special cases of covariance [6]. In our scenario however, there are no fixed response variables. Essentially all sensors are response variables and the goal is to approximate the space spanned by these variables with an appropriate subspace. Additionally, specialized algorithms [6] assume full knowledge of the covariance between the response variable and the explanatory variables, which may not be available in practice.

As a final comment, aforementioned algorithms for sampling design, sparse approximation and subset selection in regression, all produce a single set of elements (locations, vectors or variables) instead of multiple disjoint ones as we do here.

3 Network Model

We consider a network of N sensors, indexed 1 through N , scattered in a uniformly random fashion over an observation field F . The network is static in the sense that there can be no addition or modification of sensors after initial deployment. We will refer to the 2-D coordinates of sensor p with the tuple $X_0^p = (x^p, y^p)$, $p = 1 \dots N$. Sensor positions are needed for interpolation (regardless of sensing topology management) and are assumed to be known with sufficiently high accuracy by running a localization service in the network [17]. At discrete time instants t_i , a subset of the sensors measures the value of a physical quantity of interest (e.g. temperature). Sensors are time synchronized at a coarse level, so that they can be considered to be sampling at roughly the same time t_i . The physical phenomenon is modeled as the spatial realization of a random process at t_i , denoted by $S(\mathbf{x}, t_i)$, where vector $\mathbf{x}$ represents 2-D coordinates in the observation field F .

An approximation of this realization is constructed at a data fusion center by interpolating values reported only by the sensors ‘active’ at t_i along with their positions. ‘Active’ sensors are those that actually communicate their data. The collection of active sensors is hereafter referred to as *subset* or *set*. A subset is represented by the Boolean vector $\mathbf{m}_k$, of length N , where each element is (1) for an active sensor and (0) otherwise (by convention, $\mathbf{m}_0$ will refer to the set of all available sensors).

We consider interpolation that is *linear* on values measured by the sensors. This covers a broad range of techniques [9, 12, 28, 30, 34]. Denoting the random process values at sensor locations as $S(X_0^p, \mathbf{m}_k, t_i)$, the linear interpolation at point $\mathbf{x}$ can be generally expressed as:

$$\hat{S}_k(\mathbf{x}, t_i) = \sum_{p=1}^N \lambda_p(\mathbf{x}, \mathbf{m}_k) \cdot S(X_0^p, \mathbf{m}_k, t_i) \quad (1)$$

where $\{\lambda_p(\mathbf{x}, \mathbf{m}_k)\}_{p=1 \dots N}$ are coefficients describing how a specific interpolation scheme depends on the particular set of active sensors $\mathbf{m}_k \forall t_i$. The collection of interpolations $\hat{S}_k(\mathbf{x}, t_i)$ for $\forall \mathbf{x} \in F$ forms a surface, as shown in Fig. 1. A single subset $\mathbf{m}_k$ can generate surfaces for multiple t_i . The idea is that interpolation is performed with one subset for some time and then control is turned over to another subset.

To characterize how good a particular interpolated surface $\hat{S}_k(\mathbf{x}, t_i)$ is, we use as distortion measure the mean square error (MSE) with respect to the true random process $S(\mathbf{x}, t_i)$, averaged over the area of the field F . This is the average distortion incurred when using subset $\mathbf{m}_k$:

$$E[D_{\mathbf{m}_k}] = \frac{1}{|F|} \int_F E \left[\left(\hat{S}_k(\mathbf{x}, t_i) - S(\mathbf{x}, t_i) \right)^2 \right] d\mathbf{x} \quad (2)$$

A set of sensors is acceptable for a particular application, if the average distortion associated with it does not exceed an (application defined) threshold, D_0 . *The overall goal of our sensing topology management scheme is to partition sensors in as many disjoint sets as possible, while still meeting the desired distortion bound for each set.*

Presenting practical ways to devise disjoint sets is the primary goal of this paper. Before outlining our specific contributions, we briefly describe related technical issues. Specifically, notice that Eq. 2 is based on a statistical characterization of the underlying random process which is not available a priori. Instead it has to be learned after network deployment. A two-phase strategy addresses this problem:

- During the *learning phase*, all N sensors report their data, where, in addition to interpolation itself, the goal is to estimate relevant statistical properties of the process.
- During the *monitoring phase*, only sensors from an active set report and sets are rotated over time.

Our novel contribution lies in how to intelligently select the sensors comprising each set. The proposed selection algorithms are executed centrally since interpolation is also performed at the application end-point (data fusion center). The motivation is to effectively reduce the number of data packets which have to be communicated through the network. Although the exact energy savings achieved by our schemes will depend on the actual MAC protocol and routing used, reducing the number of data producing nodes is beneficial in exactly the same manner as it is for the event detection case [1, 27]. Ideally, the increase in lifetime will be proportional to the reduction in the number of data packets. Furthermore, the only energy overhead of our schemes compared to an unscheduled network comes from switching between monitoring sets. This energy cost can be kept to a minimum for example with randomized flooding [14] of a bit-mask packet denoting set membership, whenever such a switch is necessary.

4 Set Selection

4.1 Hilbert Space Framework

To analyze the problem of selecting sets of sensors, we map the network onto an equivalent Hilbert space. A Hilbert space is a collection of elements, indiscriminately referred to as points or vectors, which can be entities of any kind and have appropriate operations defined on them. These operations are addition, scalar multiplication, inner product and norm of an element [4, 21]. We make a choice of Hilbert space structure that we believe is naturally suited to our particular sampling-interpolation setup: the Hilbert space of random variables with finite second order moments. To the best of our knowledge, ours [19] is the first interpretation of a sensor network as an instance of this particular Hilbert space, which could potentially find use beyond the context examined here.

The measured value of the physical phenomenon $S(\mathbf{x}, t_i)$ can be viewed as a random variable. The completed span of these random variables, i.e., all their linear combinations and limits of Cauchy sequences thereof form a Hilbert space [4]. For a fixed time instant and under the assumption of mean square ergodicity in time, which is in fact very common in related literature [13, 16], we can define the inner product and the (induced) norm in this space, as ($*$ denotes the complex conjugate):

$$\langle S(\mathbf{x}_1), S(\mathbf{x}_2) \rangle = R(\mathbf{x}_1, \mathbf{x}_2) = E[(S(\mathbf{x}_1) - \mu(\mathbf{x}_1)) \cdot (S(\mathbf{x}_2) - \mu(\mathbf{x}_2))^*] \quad (3)$$

$$\|S(\mathbf{x})\|^2 = \langle S(\mathbf{x}), S(\mathbf{x}) \rangle = E[|S(\mathbf{x}) - \mu(\mathbf{x})|^2] \quad (4)$$

where $\mu(\cdot)$ is a spatial mean function and $R(\cdot, \cdot)$ is a spatial covariance function. The mean function represents a spatially

varying mean in the data that does not change over time. The covariance function on the other hand, quantifies the similarity between readings gathered at different spatial locations and different times in a statistical sense. The covariance structure described by Eq. 3 is allowed to be non-stationary, i.e., the correlation between readings at two points on the field can depend on their location.

The Hilbert space representation essentially enables treatment of a specific set of sensors as a specific set of vectors. We will hereafter call H_S the Hilbert space of random variables $S(\mathbf{x}, t_i)$ across the whole observation field, i.e., $\forall \mathbf{x} \in F$ for a fixed time instant.

A useful concept for linear approximation in H_S is that of ‘orthogonal projection’ [8]. Consider a Hilbert space H_0 , and a subspace of it H_I defined by a basis $\{\xi_k\}$. In terms of minimizing the squared norm of the error, the best approximation of an element $\eta \in H_0$ by an element $\eta_I \in H_I$ will be the orthogonal projection of η onto H_I . Assuming that η is not exactly orthogonal to H_I , the approximation error induced by this orthogonal projection is given by [8]:

$$\min \|\eta - \eta_I\|^2 = \|\eta\|^2 - \frac{\left(\sum_{p=1}^Q |\langle \eta, \xi_p \rangle|^2 \right)}{\left\| \sum_{p=1}^Q \langle \eta, \xi_p \rangle \cdot \xi_p \right\|^2} \quad (5)$$

In our particular setup, H_0 is H_S , H_I is the subspace defined by any reporting subset of sensors $\{S(\mathbf{X}_0^p), \mathbf{m}_k\}_{p=1..N}$ and η is $S(\mathbf{x})$ for a particular $\mathbf{x}$. Then orthogonal projection is the optimal linear approximation described by Eq. 1. In addition, Eq. 5 gives the value of the MSE, i.e., the integrand of the distortion metric defined by Eq. 2. Thus, through orthogonal projection, the Hilbert space context enables us to abstractly refer to the *best* linear interpolation achieved by a sensor subset and readily characterize its MSE performance, as opposed to referring to specific interpolation schemes.

Obtaining the orthogonal projection for every random variable $S(\mathbf{x})$ requires knowledge of all continuous covariances $R(\mathbf{X}_0^p, \mathbf{x})$. Although it is possible to estimate the continuous covariance function [22], this is ultimately an intricate procedure that does not lend itself to distortion guarantees. Instead, we restrict η to elements of a subspace of H_S , namely that spanned by all deployed sensors, H_{X0} . This is termed the *primary subspace*. Essentially, we assume that the initial number of sensors is large enough for H_{X0} to be a close approximation to H_S . The distortion metric is thus defined in relation to the maximum information we could extract with our initial deployment. Formally, it means that in Eq. 2, $S(\mathbf{x}, t_i)$ is replaced with $\hat{S}_0(\mathbf{x}, t_i)$.

A first practical point is that during the learning phase any sensor selection algorithm needs to actually evaluate Eq. 2 in

order to assess its performance. Since for a real system the expectation operator cannot be known, Eq. 2 is approximated with an average over W time instants, where $W \leq \Theta$. During the learning phase, each sensor in the network collects a time series of values $\{S(\mathbf{X}_0^p, t_i)\}_{i=1 \dots \Theta}$. By virtue of mean square ergodicity, surfaces $\{\hat{S}_0(\mathbf{x}, t_q)\}_{q=1 \dots \Theta}$, obtained from these values can be used as the best available approximation to the ground truth (i.e., they can thought of as *reference surfaces*):

$$\hat{E}[D_{m_k}] = \frac{1}{|F|} \int_F \frac{1}{W} \cdot \sum_{i=1}^W (\hat{S}_k(\mathbf{x}, t_i) - \hat{S}_0(\mathbf{x}, t_i))^2 d\mathbf{x} \quad (6)$$

Note that best interpolations $\{\hat{S}_0(\mathbf{x}, t_q)\}_{q=1 \dots \Theta}$ are defined at all points of the observation field, not just at the locations of unused sensors, and do not necessarily coincide with data obtained from these sensors.

A second practical point is that to evaluate orthogonal projection error, Eq. 5 requires inner products between all sensors. From Eq. 3 these inner products correspond to covariances and can be estimated as follows. Let B be the $\Theta \times N$ matrix having as columns the time series produced by each individual sensor as shown in Eq. 7. We will refer to B as the *data matrix*. The data matrix can be thought of as a finite dimensional approximation to the infinite dimensional random variables which are the actual elements of the primary subspace. Firstly, a vector estimate of the spatial mean function $\hat{\mu} = \{\mu_0^p\}_{p=1 \dots N}$ evaluated at the sensor locations is obtained from B , e.g. with standard least squares methods. Then the empirical covariance matrix provides an estimate of the inner products:

$$B = \begin{bmatrix} S(\mathbf{X}_0^1, t_1) & \dots & S(\mathbf{X}_0^N, t_1) \\ \vdots & \dots & \vdots \\ S(\mathbf{X}_0^1, t_\Theta) & \dots & S(\mathbf{X}_0^N, t_\Theta) \end{bmatrix} \quad (7)$$

$$\hat{R} = \frac{1}{\Theta - 1} \cdot (B - \hat{\mu})^T \cdot (B - \hat{\mu}) \quad (8)$$

with $\hat{\mu}$ being a matrix with all rows equal to $\hat{\mu}$. Equation 8 converges to the true covariances for Θ large enough, again by virtue of mean square ergodicity.

4.2 Complexity of Optimal Solution

In the Hilbert space framework, introduced in the previous section, each individual sensor can be thought of as a vector in the primary subspace. The problem of selecting disjoint sets of sensors thus effectively translates into selecting disjoint sets of vectors. The goal is to maximize the number of sets that can be found (or equivalently, minimize the average number of vectors in each set), while ensuring that each set can provide a sufficiently accurate approximation.

As a first step in tackling this problem, we consider a more basic variant, namely that of finding just a single minimal set: *Given an initial set of sensors, find the minimal subset of vectors which yields an approximation with average distortion at most D_0 .*

The problem can be seen as finding an ‘approximate’ basis for the primary subspace. Finding multiple sets is a generalization of this variant and hence computationally at least as hard. The single-set problem is related to sparse signal approximation with general dictionaries, which has been studied in signal processing literature [7, 10, 25, 29, 32, 33]. The term ‘dictionary’ refers to a set of non-orthogonal vectors used for representation in a Hilbert space, without necessarily forming a basis for that space. A known problem in a V -dimensional Hilbert space is how to select the best vectors out of a redundant dictionary of size $P \geq V$ to approximate a given target vector in the space. This requires enumerating all possible subsets of vectors, an operation with cost exponential in P [7, 10]. In the case of a general dictionary, the resulting computation is provably NP-hard [7].

In our scenario, the dimension of the primary subspace H_{X0} is at most N . The set of N vectors that correspond to the initially deployed sensors $\{S(\mathbf{X}_0^p)\}_{p=1 \dots N}$ is therefore a redundant dictionary for this space. A dictionary of size N effectively means that the computational cost for optimal sensor selection grows exponentially with the size of the network N . For our particular case we have also proved a stronger result (see Appendix):

Lemma *For a deployment where the positions of the sensors form a Poisson point process with constant rate β and the monitored random process $S(\mathbf{x})$ is wide sense stationary, vectors $\{S(\mathbf{X}_0^p)\}_{p=1 \dots N}$ are linearly independent on the average.*

Linear independence means that the dimension of H_{X0} is exactly N , i.e., the dictionary $\{S(\mathbf{X}_0^p)\}_{p=1 \dots N}$ is also a basis for the space, rendering optimization over *any* redundant dictionary for this space exponentially hard with the size of the network.

Since the single set selection problem is hard, the same will hold for the extended problem of finding multiple sets. As a result, we have to resort to heuristic approaches to perform the selection of multiple active sets of sensors.

4.3 Jittered Grid Sampling

The first proposed scheme is motivated by the observation that regular grid sampling designs, specifically rectangular and triangular grids, are frequently used in geostatistical monitoring [9, 35]. Additionally, an abundance of theoretical results exists with respect to the distortion performance of such designs when monitoring spatially stationary Gaussian processes [28].

In our setup, sensors are chosen from a pre-deployed set, which does not necessarily form a grid. The key idea is instead to impose a *virtual square grid* over the observation field and then ‘map’ a subset of sensors onto this grid. The mapping should satisfy some closeness criterion. By offsetting the grid, the mapping procedure can be repeated a number of times so as to obtain many different sensor sets. As an example, Fig. 2 shows a virtual square grid of 16 nodes superimposed on a network of 50 sensors.

A virtual square grid is uniquely defined by its side α , while grids of the same size are distinguished by their offsets relative to a fixed point (x_0, y_0) . To choose the appropriate grid size $M = n^2$, the integer n is iteratively increased until the resulting sensor subsets satisfy the distortion bound as computed through Eq. 6. Furthermore, virtual grids are offset along their common diagonal.

The crucial element is how to obtain disjoint sensor subsets from virtual grids. Essentially, a 1-1 mapping is needed, such that each sensor is mapped onto a grid point only once and optimality is achieved with respect to the closeness criterion. Suppose that a total of $\lfloor N/M \rfloor$ virtual grids have been constructed, consisting of $\lfloor N/M \rfloor \cdot M$ points. We construct a complete bipartite graph (U, V, E) consisting of a set of vertices U corresponding to the grid points, a set of vertices V corresponding to the sensors and a set of edges $E = \{[u_i, v_j]\}$ with weights w_{ij} equal to the Euclidean distance between grid point i and sensor j . If the number of grid points $\lfloor N/M \rfloor \cdot M$ is less than the number of sensors, N (i.e., if N is not an integer multiple of M), the set U of grid points is enhanced by adding to it $N - \lfloor N/M \rfloor \cdot M$ points with incident edges of infinite weight. With this formulation, the problem of mapping sensors onto grid points is equivalent to the minimum weight perfect matching problem for bipartite graphs and can be efficiently solved by the Hungarian Algorithm [23]. We call the overall procedure *Jittered Grid (JiG)* sampling, presented in Fig. 3. Note that after completion of the algorithm, $N - \lfloor N/M \rfloor \cdot M$ sensors corresponding to dummy grid nodes are randomly distributed among devised sets.

A drawback of *JiG* is that it is better suited for spatially stationary processes, since virtual grids are regular. However, spatial non-stationarities are in fact commonly encountered in real life physical phenomena and corresponding measurements. For example, at locations

closer to windows or heat sources, temperature readings may vary more rapidly for proximate sensors than they do at dark or isolated locations, e.g. under desks. Another drawback is that it does not provide fine-grained control over the size of sets obtained. For example, if sets of $13^2 = 169$ sensors are inadequate, the next available choice is sets of $14^2 = 196$ sensors. This effect becomes more pronounced for large n .

4.4 Random Variable Greedy Selection

The purpose of the second scheme we propose is exactly to address these disadvantages of *JiG* sampling. It builds on the expressive power of the Hilbert space H_S of sensors as random variables. The key idea is quantifying ‘colinearity’ or ‘orthogonality’ between a given candidate vector and an existing set of vectors. This can be done by using orthogonal projection error (defined in Eq. (5)) as the score function. If the orthogonal projection error of a candidate vector onto a set of existing vectors is maximal among all such vectors, then the descriptive power of the set will maximally grow if we add the candidate to it. If, on the other hand, the orthogonal projection error of a member of the set onto the rest of the vectors in the set is minimal, then the descriptive power of the set will only marginally be affected if we remove the candidate vector from it. The main strength of this approach is that it provides a characterization of how redundant an individual sensor is with respect to any set of sensors for interpolation purposes.

Our *Random Variable Greedy (RaVaG)* algorithm proceeds as follows. It is not known a priori how many sets can be created. Instead, it starts creating the first set by selecting vectors until the distortion criterion is met. Next, the second set will be selected from the remaining vectors, and so forth. Consider, in general, a situation where we are in the process of creating the j^{th} set. At this point, the primary subspace can be considered as being partitioned in three subspaces: 1) the space H_U of vectors in sets 1 through $j-1$; 2) the space H_A of vectors already selected in set j ; 3) the space H_R of vectors not yet selected for any of the sets. Our algorithm considers all candidate vectors η from those not yet belonging to any set. For each

Figure 2 A virtual grid imposed on a random network.

- 1 Using data from the learning phase, construct W reference surfaces $\{\hat{S}_0(\mathbf{x}, t_q)\}_{q=1 \dots W}$.
- 2 Select n , set $M = n^2$.
- 3 Construct sets indexed $k = 1 \dots \lfloor N/M \rfloor$, with application of the Hungarian algorithm on the matrix $\{w_{ij}\}$ of weights defined by the Euclidean distances separating grid node i , $i=1 \dots N$ from sensor j , $j=1 \dots N$.
- 4 Compute D , from eq. (6) for one of these sets (randomly selected). If $D > D_0$ go to step 2 with $n = n + 1$, otherwise finalize the M computed in step 2.

Figure 3 Jittered Grid (JiG) set construction algorithm.

one of them, it computes the error by orthogonal projection in both of the spaces H_A and H_R (always excluding the vector η):

$$\begin{aligned} E_A(\eta) &= \min \left\| \eta - \sum_p c_p \cdot \xi_p \right\|^2 & H_A &= \text{span}\{\xi_p\} \\ E_R(\eta) &= \min \left\| \eta - \sum_q d_q \cdot \zeta_q \right\|^2 & H_R &= \text{span}\{\zeta_q\} \end{aligned} \quad (9)$$

The crucial element is how to use these orthogonal projection errors to populate the sets with ‘good’ vectors. A simple first choice would be to sequentially select the vector η that maximally expands the existing set, i.e., the vector maximizing $E_A(\eta)$. However, this choice does not always lead to a good set of vectors, especially in cases where the random process is spatially non-stationary. To see this, consider a random process showing rapid variations over a small region of the field, while being smooth over the rest of the field. Then a heuristic based on maximizing $E_A(\eta)$ would first choose vectors in the rapidly varying region, because they are likely to be the most orthogonal to each other. This strategy has an immediate drawback: it is possible that sets subsequently constructed cannot contain any of these vectors describing rapid variations, because they will all have been used up. Eventually, such sets will not be able to achieve the target distortion or will need to employ a much larger number of sensors.

Based on this example, there are two competing effects both of which should be taken into consideration when designing a greedy approach: expanding the expressive capability of the set currently being constructed and not ‘crippling’ the expressive capability of the set of sensors that remain. The first effect can be quantified by requiring a high orthogonal projection error onto the space of already selected sensors, i.e., the candidate vector should be as orthogonal to this set as possible. The second effect can be quantified by requiring a low orthogonal projection error onto the space of remaining vectors, i.e., the candidate vector should be as colinear to this set as possible.

The *RaVaG* selection scheme we propose here ranks each candidate vector according to a single score function. A naive score function would take only one of the two effects into account, e.g. the first one. We indeed examine such a case, specifically setting the score function C_1 to be equal to the orthogonal projection error of the candidate vector onto the space of already selected sensors $E_A(\eta)$. A ‘good’ score function however, should ideally incorporate both of the effects. Since the relationship between them is a competitive one, i.e., $E_A(\eta)$ should be high while $E_R(\eta)$ low, it is a natural choice to consider the mathematical difference and/or ratio between these orthogonal projection errors and attempt to maximize it. In this paper, we have experimented and present

results both with the difference as well as the ratio score functions, hereafter referring to them as C_2 and C_3 respectively:

$$\begin{aligned} C_1(\eta) &= E_A(\eta) \\ C_2(\eta) &= E_A(\eta) - E_R(\eta) \\ C_3(\eta) &= E_A(\eta)/E_R(\eta) \end{aligned} \quad (10)$$

The vector η with the *maximum* score amongst all candidates is added to the j^{th} set. The detailed set selection algorithm is presented in Fig. 4. When the algorithm terminates, there may remain some sensors that were not assigned to any active sets (since they could not form a set by themselves satisfying the distortion target). These are distributed in a round robin fashion among existing sets in such a way that each set is assigned the sensor which maximizes the heuristic expression used to construct it in the first place.

5 Evaluation

5.1 Synthetic Data

A purely simulated evaluation setting has the major advantage that it gives us access to the ground truth, i.e., the spatial process itself, to compare interpolation performance with. The setting for experiments was an observation

```

1  Input: vectors  $\{S(X_0^p)\}_{p=1 \dots N}$ ,  $\hat{R}_{pq}$ ,  $\{\hat{S}_0(x, t_q)\}_{q=1 \dots W}$ ,  $D_t$ 
2  Output: subsets of vectors  $m_k$ 
3  begin
4     $m_0 \leftarrow \text{I}[\bigcup_{p=1}^N S(X_0^p)]$ ,  $k = 1$ 
5  repeat
6     $m_k \leftarrow \emptyset$ 
7    repeat
8      if ( $m_k == \emptyset$ ) select first available vector into  $m_k$ 
9    else
10     foreach candidate vector  $S(X)$ 
11        $p_0 \leftarrow \text{I}[S(X)]$ 
12        $S(X^*) \leftarrow$ 

$$\arg \max_{S(X) \in m_0} C(\hat{R}_{p_0 p_0} - \frac{(\sum_{p \in m_1} (\hat{R}_{p_0 p})^2)^2}{\sum_{p \in m_1} \sum_{q \in m_1} \hat{R}_{p_0 p} \cdot \hat{R}_{p_0 q} \cdot \hat{R}_{pq}})$$


$$, \hat{R}_{p_0 p_0} - \frac{(\sum_{p \in m_0} (\hat{R}_{p_0 p})^2)^2}{\sum_{p \in m_0} \sum_{q \in m_0} \hat{R}_{p_0 p} \cdot \hat{R}_{p_0 q} \cdot \hat{R}_{pq}})$$

13      $m_k \leftarrow m_k \cup \text{I}[S(X^*)]$ 
14      $m_0 \leftarrow m_0 \setminus \text{I}[S(X^*)]$ 
15   until ( $\hat{E}[D_{m_k}] \leq D_t$ )
16    $k \leftarrow k + 1$ 
17 until ( $(m_0 == \emptyset)$  or ( $\hat{E}[D_{m_0}] > D_t$ ))
18 end
Note:  $\text{I}[\cdot]$  denotes the index of the argument in  $\bigcup_{p=1}^N S(X_0^p)$ 

```

Figure 4 Random Variable Greedy (*RaVaG*) subset construction algorithm.

field of square shape and size 10^4 m^2 . We considered uniformly random deployments with $N = 1000$, $N = 1500$ sensors. Our algorithms require a learning phase of Θ time instants. Readings obtained during the learning phase are used for estimation of the covariance matrix in Eq. 8 and estimation of distortion in Eq. 6. The latter is a computationally expensive operation that must be repeated at each algorithm iteration for both *JiG* and *RaVaG* (step 17 in Fig. 4 and step 4 in Fig. 3). Especially for *RaVaG*, $\Theta \geq N$, because the orthogonal projection error of Eq. 5 assumes linearly independent vectors and the covariance matrix estimate of (8) should be full rank [8, 21]. In our simulations we initially ran the learning phase for $\Theta = N$ time instants and used all acquired sensor values for covariance matrix estimation, but only $W = 125$ reference surfaces for distortion estimation. In practice, we found that even the requirement $\Theta \geq N$ is not essential to obtain good sets. Performance of *RaVaG* does not significantly degrade even for $\Theta \approx N/4$.

There is no existing solution to actually compare our schemes against. However we have devised and experimented with a reasonable alternate approach. Specifically:

Random selection Select k sensors at random to comprise a set. Keep increasing k until the distortion criterion (6) is satisfied. Repeat until no sensors remain unselected or are too few to meet the target distortion.

5.1.1 Stationary Data

We first conducted experiments with a spatially stationary physical process. Process realizations were generated according to a *simple kriging* model [28], which is commonly used in geostatistics and atmospheric sciences to describe environmental data sets. Specifically, zero mean white noise was fed into a symmetric 2-D low pass spatial filter. The target distortion D_t (see Fig. 4) was set to 0.5. White Gaussian measurement noise of mean zero and variance 0.1 was added to sensor samples in all cases.

Figure 5(a) shows the number of sets obtained with random selection as well as the *JiG* and *RaVaG* algorithms with all different cost functions (see Eq. 10), for $N = 1000$ and $N = 1500$. Figure 5(b) shows the average size of sets for each scheme, along with standard deviations. Sizes are the ones obtained immediately after running our selection algorithms, i.e., before assignment of remaining sensors among devised sets. It can be seen that our schemes greatly improve on the efficiency of an unscheduled network (consisting of only one set) by 6 and 9 times for $N = 1000$ and $N = 1500$ respectively. Additionally, there is improvement on the number of sets devised with random selection, specifically 50% for both $N = 1000$ and $N = 1500$. This improvement can be explained by observing average set sizes. For example, the average set devised with *RaVaG-C2*

Figure 5 Sets devised for $N = 1000$, 1500 and stationary data.

consists of 151 sensors. This is a 33% reduction over the 225 sensors comprising the average random set.

5.1.2 Non-stationary Data

Next, we experimented with processes characterized by a linearly growing trend as well as a covariance structure changing over the observation field, i.e., spatially varying. To generate appropriate process realizations we implemented the model described in [22] for estimation of non-stationary behavior in environmental data. Figure 6(a) shows a cross section of one process realization taken along the main diagonal of the observation field. The data magnitude shows a clear growing trend. Figure 6(b) shows the same cross section when the trend is subtracted. It can be seen that the resulting process shows more rapid variations for the right half of the cross section. The procedure resulted in process realizations that would require approximately 225 randomly selected sensors to achieve a target distortion of 0.5 (same as in the stationary case).

Figures 7(a) and (b) contain the number of sets obtained and average set sizes respectively, for $N = 1000$ and $N = 1500$. The gains in terms of total number of sets obtained are 6 and 8 fold compared to the unscheduled case and 50%, 33% compared to random selection for $N = 1000$ and $N = 1500$ respectively.

Figure 6 Cross section of a non-stationary realization along main diagonal of observation field, shown **a** with trend in data magnitude and **b** after subtracting trend.

5.1.3 Discussion

Based on Figs. 5 and 7, the variants of *RaVaG* more consistent in providing good performance with respect to number of subsets devised, were those utilizing cost functions C_2 and C_3 . The performance of the variant based on C_1 , even though reasonably good in the stationary case, is observed to deteriorate for non-stationary data, as expected. This is indicated by the large deviation in the sizes of subsets devised by this variant (e.g. for $N=1000$ the second subset is much smaller than the rest). It can be attributed to the cost function C_1 being ‘fooled’ into selecting many sensors in the region of the field characterized by lower correlation, which cripples further construction of subsets, as described earlier, in Section 4.4.

Figure 8 shows spatially averaged instantaneous squared error versus ground truth from actual interpolations with devised subsets. Performance is shown for one run with stationary data, for random selection and *RaVaG*- C_3 when $N=1500$. The ensemble mean in time of the spatially averaged squared error is also shown on the figures as ‘ $\hat{E}[D]$ ’. This serves as an approximation to the true MSE of interpolation. Plots for non-stationary data as well as $N=1000$ and *RaVaG*- C_2 are qualitatively very similar and thus omitted.

Subsets were activated in a sequential manner, with vertical lines indicating points of switching between them. Realizations up to 50 correspond to distortion during the learning phase, where all sensors in the network are reporting. This is done to give an idea of the performance of the initial deployment and how a low initial distortion can be effectively traded off for multiple reporting subsets. Normally, the sensor network would be in the monitoring phase for a much longer duration than the 250 time instants shown here; Fig. 8 merely serves to examine whether distortion achieved by devised subsets actually meets the target.

It can be seen that all subsets succeed in meeting the criterion of Eq. 6, since both ensemble means are lower than the distortion target. Thus, *RaVaG*- C_2 and *RaVaG*- C_3 are the overall best choices for exploiting redundancy in terms of number of disjoint subsets devised, for both stationary and non-stationary settings.

Figure 7 Sets devised for $N=1000$, 1,500 and non-stationary data.

Figure 8 Instantaneous squared error vs. ground truth for **a** *Random*, **b** *RaVaG*- C_3 and stationary data.

5.2 Real Data

We used ambient temperature data from the LUCE Sensorscope deployment at EPFL [26]. Samples were collected several times a day by 97 Sensorscope nodes, dispersed over the entire EPFL campus. For our purposes, we chose the largest subset of sensors with data available for a common, sufficiently long period of observation. This period was from April 24th to May 9th, 2007. Specifically, there were $N=83$ sensors, each reporting 12 times a day for 15 days (April 29th was missing from available datasets). Half of these samples were used for the learning phase and the rest for checking instantaneous interpolation distortion.

One of the problems, intrinsic to testing on real data, is that the underlying physical phenomenon (i.e., the ground truth) is unknown. The only information available is what can be learned from all reporting sensors. Consequently, distortion is evaluated with respect to the primary subspace H_{X0} instead of the physical space H_S . The target distortion for devising subsets with the *RaVaG* algorithm of Fig. 4 was set to four degrees Fahrenheit squared. For the chosen target distortion, *RaVaG*- C_2 and C_3 resulted in two sets of 37 and 46 sensors and 32 and 51 sensors respectively.

Figure 9 shows performance of these sets for the C_3 variant in terms of the root of the instantaneous squared error. The square root in this case, with its units of degrees Fahrenheit, gives a better intuitive feel on much error should be expected on the average when predicting temperature. It is also in accordance with the Root Mean Square Error in [12]. Hence, our target in the evaluation corresponded to two degrees

Figure 9 Root of instantaneous squared error for real temperature data.

Fahrenheit of average prediction error. It can be seen that the target distortion is met on the average, rendering our approach a viable alternative to having all sensors reporting.

6 Conclusion

In this paper, we have addressed the problem of sensing topology management for applications aiming at spatial interpolation of a physical quantity. The sensor network essentially behaves as a distributed sampling system and notions such as sensing range or k -coverage are meaningless. Our objective has been to reduce the amount of packets produced by the network while satisfying an application defined performance criterion.

We have presented two methods to devise disjoint sets of sensors based on well established sampling design notions and a Hilbert space framework. The only overhead compared to an unscheduled network is communicating set membership to the sensors. Simulation results have shown substantial gains in the number of disjoint sets that can support a user specified target distortion compared to simpler sensor selection techniques. Our approaches are generic enough to be able to accommodate both spatially stationary and non-stationary physical processes. Finally, our Hilbert space view of the sensor network enables direct application of algebraic tools (e.g. operator theory) and therefore has the potential of being a useful tool in sensor network processing even beyond the specific scope of the problem examined here.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

Proof of Lemma

We initially consider a one-dimensional sensor network where sensor positions $\{x_p\}_{p=1\dots N}$ constitute an orderly Poisson process of constant rate β on the observation interval $[0, L]$. Similarly, we consider the monitored phenomenon as a one-dimensional wide sense stationary process $S(x)$ with correlation function $R(x)$ and associated power spectral density $\Phi(\omega)$. For this special type of process there exists an isometry between the primary subspace H_{X0} and the space $H_{\Phi_{X0}}$ defined by the vectors $\{e^{j\omega x_p}\}_{p=1\dots N}$ [4]. An element of this space and the correspondence with actual process values is given by:

$$a(\omega) = \sum_{\kappa=1}^n a_{\kappa} \cdot e^{j\omega x_{\kappa}} \leftrightarrow A = \sum_{\kappa=1}^n a_{\kappa} \cdot S(x_{\kappa}, t_i) \quad (11)$$

whereas the isometry (i.e., equivalency of inner products) is described by:

$$\begin{aligned} \langle a(\omega), b(\omega) \rangle_{H_{\Phi_{X0}}} &= \int_{-\infty}^{+\infty} a(\omega) \cdot b^*(\omega) \cdot \Phi(\omega) d\omega \\ &= \sum_{\kappa=1}^n \sum_{\ell=1}^n a_{\kappa} \cdot b_{\ell}^* \cdot R(x_{\kappa} - x_{\ell}) = \langle A, B \rangle_{H_{X0}} \end{aligned} \quad (12)$$

In order to prove linear independence on the average, we first look at linear combinations of vectors in the isometric subspace. We want to prove that no vector of the form $e^{j\omega \cdot x_n}$ can be written as a combination of other vectors, when averaging over all realizations $\{x_p\}_{p=1\dots N}$, i.e., over all Poisson deployments of the given rate. Suppose the opposite is true. Then for a non-trivial choice of the coefficients c_p , the following expression should hold:

$$E[\varepsilon_n] = \sum_{p=1}^n c_p \cdot E[e^{j\omega \cdot x_p}] = 0 \quad (13)$$

The expectation term in Eq. 13 is the characteristic function of x_p , i.e., the length of the interval from the origin to the p -th point of the process. This length is often called recurrence time of order p . For a Poisson point process, its probability density function $f_p(u)$ depends only on p and the rate β through the following closed form:

$$f_p(u) = \frac{\beta \cdot (\beta \cdot u)^{p-1} \cdot e^{-\beta \cdot u}}{(p-1)!}, u \geq 0 \quad (14)$$

Plugging the definition of a characteristic function into Eq. 13 we get equivalently:

$$\begin{aligned} \sum_{p=1}^n c_p \cdot \int_{-\infty}^{\infty} f_p(u) \cdot \text{Ind}_{(0,\infty)}(u) \cdot e^{j\omega \cdot u} \cdot du &= 0 \Leftrightarrow \\ \int_{-\infty}^{\infty} \left(\sum_{p=1}^n c_p \cdot f_p(u) \cdot \text{Ind}_{(0,\infty)}(u) \right) \cdot e^{j\omega \cdot u} \cdot du &= 0 \Leftrightarrow \quad (15) \\ \sum_{p=1}^n c_p \cdot f_p(u) &= 0, \forall u \geq 0 \end{aligned}$$

where $\text{Ind}_{(0,\infty)}(u)$ denotes the step function and the last equation follows from the uniqueness of the Fourier transform. Equation 15 can be further scrutinized as:

$$\begin{aligned} \sum_{p=1}^n c_p \cdot \frac{\beta \cdot (\beta \cdot u)^{p-1} \cdot e^{-\beta \cdot u}}{(p-1)!} &= 0, \forall u \geq 0 \Leftrightarrow \\ e^{-\beta \cdot u} \cdot \sum_{p=1}^n c_p \cdot \frac{\beta \cdot (\beta \cdot u)^{p-1}}{(p-1)!} &= 0, \forall u \geq 0 \Leftrightarrow \quad (16) \\ \sum_{p=1}^n c_p \cdot \frac{\beta^p}{(p-1)!} \cdot u^{p-1} &= 0, \forall u \geq 0 \Leftrightarrow \\ c_p \cdot \frac{\beta^p}{(p-1)!} &= 0, \forall p \Leftrightarrow c_p = 0, \forall p \end{aligned}$$

where the last equation follows from the linear independence of any subset of monomials, over any interval of the real line, i.e.:

$$\sum_{p=0}^n a_p \cdot u^p = 0, \forall u \geq 0 \Leftrightarrow a_p = 0, \forall p \quad (17)$$

Equation 16 contradicts our initial hypothesis and therefore, on the average, no vector in a Poisson deployment can be a linear combination of other vectors. To generalize for the case of multidimensional scenarios, we observe that a multidimensional Poisson sampling process can be degraded into a one dimensional one by keeping only a single (e.g. the first) coordinate of each sensor position; by definition of the Poisson process, for a fixed number of sensors this coordinate will be uniformly distributed in the observation interval and, since involved distributions are continuous, the probability that two of the resulting points will coincide is zero. The corresponding correlation function would then be formulated by taking only differences of the chosen coordinate into account. The resulting proof arguments are the same.

Note that the lemma indicates what we should expect on average for a Poisson deployment of fixed rate. However, any specific deployment may deviate from the average, i.e., there is a non-zero (although very small) probability that some small number of vectors are linearly dependent.

References

- Abrams, Z., Goel, A., & Plotkin, S. (2004). *Set K-Cover algorithms for energy efficient monitoring in wireless sensor networks*. in IPSN 2004.
- Ai, X., Srinivasan, V., & Tham, C.-K. (2007). *DRACo: distributed, robust and asynchronous coverage in wireless sensor networks*. in SECON 2007.
- Carbunar, B., Grama, A., Vitek, J., & Carbunar, O. (2004). *Coverage preserving redundancy elimination in sensor networks*. in SECON 2004.
- Cramer, H., & Leadbetter, M. R. (1967). *Stationary and related stochastic processes: Sample function properties and their applications*. Wiley, 1967.
- Culler, D., Estrin D., & Srivastava, M. B. (2004). *Overview of sensor networks*. in Computer, August 2004.
- Das, A., & Kempe, D. (2008). *Algorithms for subset selection in linear regression*. in STOC 2008.
- Davis, G., Mallat, S., & Avellaneda, M. (1997). Adaptive greedy approximations. *Journal of Constructive Approximation*, 13, 57–98.
- Deutsch, F. (2001). *Best approximation in inner product spaces*. Springer-Verlag, 2001.
- Diggle, P. J., & Ribeiro, P. J. J. (2006). *Model-based geostatistics*. Springer 2006.
- Donoho, D. L., & Elad, M. (2003). *Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization*. in PNAS 2003.
- Duarte, M. F., Wakin, M. B., Baron, D., & Baraniuk, R. G. (2006). *Universal distributed sensing via random projections*. in IPSN 2006.
- Guestrin, C., Bodic, P., Thibaux, R., Paskin, M., & Madden, S. (2004). *Distributed regression: An efficient framework for modeling sensor network data*. in IPSN 2004.
- Jindal, A., & Psounis, K. (2006). Modeling spatially correlated data in sensor networks. in *ACM Transactions on Sensor Networks*, 2(4) 466–499.
- Kini, A. V., Veeraraghavan, V., Singhal, N., Weber, S. (2006). *SmartGossip: An improved randomized broadcast protocol for sensor networks*. in IPSN 2006.
- Koushanfar, F., Taft, N., & Potkonjak, M. (2006). *Sleeping coordination for comprehensive sensing using isotonic regression and domatic partitions*. in INFOCOM 2006.
- Krause, A., Guestrin, C., Gupta, A., & Kleinberg, J. (2006). *Near-optimal sensor placements: Maximizing information while minimizing communication cost*. in IPSN 2006.
- Langendoen, K., & Reijers, N. (2003). *Distributed localization in wireless sensor networks: A quantitative comparison*. in Computer Networks, 2003.
- Liaskovitis, P., & Schurgers, C. (2006). *A distortion aware scheduling approach for wireless sensor networks*. in DCOSS 2006.
- Liaskovitis, P., & Schurgers, C. (2007). *Leveraging redundancy in sampling-interpolation applications for sensor networks*. in DCOSS 2007.
- Miller, A. (2002). *Subset selection in regression*. Chapman & Hall/CRC, 2002.
- Moon, T. K., & Stirling, W. C. (2000). *Mathematical methods and algorithms for signal processing*. Prentice Hall, 2000.
- Nott, D. J., & Dunsmuir, W. T. M. (2002). Estimation of non-stationary spatial covariance structure. *Biometrika*, 89, 819–829.
- Papadimitriou, C. H., & Steiglitz, K. (1998). *Combinatorial optimization—algorithms and complexity*. Dover Publications 1998.
- Perillo, M., Ignjatovic, Z., & Heinzelman, W. (2004). *An energy conservation method for wireless sensor networks employing a blue noise spatial sampling technique*, in IPSN 2004.
- Rao, B. D., Egan, K., Cotter, S. F., Palmer, J., & Kreutz-Delgado, K. (2003). Subset Selection in Noise Based on Diversity Measure Minimization. *IEEE Transactions on Signal Processing*, 51(3).
- Sensorscope Wireless Distributed Sensing System for Environmental Monitoring. (2007). http://sensorscope.epfl.ch/index.php/Main_Page.
- Slijepcevic, S., & Potkonjak, M. (2001). *Power efficient organization of wireless sensor networks*. in ICC 2001.
- Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Springer 1999.
- Tropp, J. A., Gilbert, A. C., & Strauss, M. J. (2006). Algorithms for simultaneous sparse approximation part i: greedy pursuit. *Signal Processing*, 86(3), 572–588.
- Vuran, M. C., & Akyildiz, I. F. (2006). Spatial correlation based collaborative medium access control in wireless sensor networks. *IEEE/ACM Transactions on Networking*, 14(2), 316–329.
- Wang, X., Xing, G., Zhang, Y., Lu, C., Pless, R., & Gill, C. (2003). *Integrated coverage and connectivity configuration in wireless sensor networks*. in SenSys 2003.
- Wipf, D., & Rao, B. D. (2004). Sparse Bayesian learning for basis selection. *IEEE Transactions on Signal Processing*, 52(8), 2153–2164.
- Wipf, D., & Rao, B. D. (2007). An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Transactions on Signal Processing*, 55(7), 3704–3716.
- Zhao, Q., & Tong, L. (2003). *Quality-of-service specific information retrieval for densely deployed sensor networks*. in MILCOM 2003.
- Zhu, Z., & Stein, M. L. (2006). Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological and Environmental Statistics*, 11(2), 24–44.

Periklis Liaskovitis received his Ph.D. and M.S. degrees in Electrical Engineering from the University of California, San Diego, in 2009 and 2006 respectively. Prior to that, he had received a Diploma degree in Electrical and Computer Engineering from the National Technical University of Athens (2002) and worked as a researcher at the University of Athens (2003). His research interests include sparse modelling methods and statistical signal processing, especially as applied to efficient sensing and communication systems.

Curt Schurgers is an assistant professor at the Electrical and Computer Engineering Department of the University of California, San Diego. He received his M.S. degree from the Katholieke Universiteit Leuven, Belgium, in 1997, and his Ph.D. from the University of California, Los Angeles in 2002. He was also a researcher at the Interuniversity Microelectronics Center in Belgium (1997-1999), and a postdoctoral researcher at MIT (2003). His research interests include underwater networking, ad hoc networking and sensor networking.