

Sensor Selection for Hypothesis Testing in Wireless Sensor Networks: a Kullback-Leibler Based Approach

Dragana Bajović, Bruno Sinopoli and João Xavier

Abstract—We consider the problem of selecting a subset of p out of n sensors for the purpose of event detection, in a wireless sensor network (WSN). Occurrence or not of the event of interest is modeled as a binary Gaussian hypothesis test. In this case sensor selection consists of finding, among all $\binom{n}{p}$ combinations, the one maximizing the Kullback-Leibler (KL) distance between the induced p -dimensional distributions under the two hypotheses. An exhaustive search is impractical if n and p are large, as the resulting optimization problem is combinatorial. We propose a suboptimal approach with computational complexity of order $\mathcal{O}(n^3 p)$. This consists of relaxing the 0/1 constraint on the entries of the selection matrices to let the optimization problem search over the set of Stiefel matrices. Although finding the Stiefel matrix is a nonconvex problem, we provide an algorithm that guarantees to produce a global optimum for $p = 1$, through a series of judicious problem reformulations. The case $p > 1$ is tackled by an incremental, greedy approach. The obtained Stiefel matrix is then used to determine the sensor selection matrix which best approximates its range space. Extensive simulations are used to assess near optimality of the proposed approach. They also show how the proposed approach performs better than exhaustive searches once an upper bound on the computation time is set.

I. INTRODUCTION

In wireless sensor networks (WSN), energy is a scarce resource. To guarantee prolonged unattended deployment, sensors need to be frugal with respect to energy usage. Sensing, computing and communication need to be minimized to ensure maximization of network lifetime. In addition, it may be infeasible to collect and process all the possible sensor data due to bandwidth and computational constraint. Therefore, it is of practical interest to activate only a limited number of sensors at any time instant. The power constraint dictates how many sensors should be operating. For any particular application, the following question arises: which is the optimal subset of sensors to choose in order to maximize a pre-specified performance metric ?

Partially supported by grant SFRH/BD/33517/2008 (through the Carnegie Mellon/Portugal Program managed by ICTI) from Fundação para a Ciência e Tecnologia and also by ISR/IST plurianual funding (POSC program, FEDER).

D. Bajović is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal, and with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA dragana@isr.ist.utl.pt, dbajovic@andrew.cmu.edu

B. Sinopoli is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA brunos@ece.cmu.edu

J. Xavier is with the Institute for Systems and Robotics (ISR), Instituto Superior Técnico (IST), Lisbon, Portugal jjxavier@isr.ist.utl.pt

The problem of sensor selection has attracted the interest of many researchers in the past few years. In the context of estimation [1], the authors propose an efficient heuristic for finding a subset of sensors for optimal estimation. In [2], the authors consider estimation of a convection-dispersion field via Kalman filtering. The subset of sensors that will report their measurements in each iteration is chosen to minimize the next step error covariance matrix. In [20], the authors determine the optimal sensor selection for estimation of the state of a dynamical system in the case sensors communicate using a Carrier Sensed Medium Access (CSMA) protocol. The benefit of using more sensors is off-set by the bottleneck they would create in the network. In [3], the authors propose an information-driven dynamic sensor selection for target tracking tasks, using information utility measures such as entropy, Mahalanobis and Kullback-Leibler distances. A survey of sensor selection schemes in WSN can be found in [4].

In this paper, we are interested in selecting sensors for event detection in WSN. Sensor selection for event detection in WSNs arises in many applications, namely in monitoring pipes from breaking in building systems, or monitoring CO₂ sequestration sites for gas leakage. Our goal is to find the most informative subset of sensors whose readings are then processed by a central node. This is different than distributed detection. In the latter approach, each sensor makes a local decision and transmits it to the fusion center, which, after gathering all local decisions, performs the final detection.

Problem statement. Let us consider the problem of event detection for a WSN composed of n sensors. Such problem can be modeled as a binary Gaussian hypothesis test where H_1 (H_0) corresponds to the occurrence (non-occurrence) of the event. The mean and the covariance matrix of the Gaussian distributed sensor data vector $x \in \mathbb{R}^n$ under hypothesis H_i are denoted by (m_i, S_i) respectively and assumed known, e.g., they are estimated from available training data. Let us assume that only $p < n$ sensors are allowed to transmit their readings to a central node. In other words, the fusion node conducts the hypothesis test based on a p -dimensional data vector $y = E^T x$, where E is an $n \times p$ selection matrix having one (resp. at most one) entry equal to 1 in each column (resp. row), the remaining entries being 0. We address the question of which subset of p sensors to choose, i.e., which matrix E to pick, in order to maximize detection performance. We are interested in evaluating the performance of the proposed approach and compare to the optimal one when this is feasible to compute, and to a suboptimal approach based on a feasible exhaustive search over a subset of the all possible

choices.

Contributions. To find the optimal subset of p sensors, we propose to maximize the Kullback-Leibler (KL) distance between the distributions of the selected p measurements under the two hypotheses. Performing this maximization by searching over all $\binom{n}{p}$ possible sensor selections is impractical for large n and p . We propose a suboptimal method with computational cost of order $\mathcal{O}(n^3 p)$ for maximizing the KL-based cost function, showing near-optimal performance in extensive simulations.

Our approach consists of relaxing the choice of 0/1 selection matrices to the set Stiefel ones [19]. This can be interpreted as finding the linear map $\mathbb{R}^n \mapsto \mathbb{R}^p$ which maximizes the KL distance between the “projected” Gaussian distributions in the lower p -dimensional space. This problem also appears in linear dimensionality reduction applications, e.g., see [16], [15] for closely related approaches involving the J -divergence and Chernoff distances. It is a difficult, nonconvex optimization problem. To our best knowledge, existing work on this topic either does not solve the problem in full generality (i.e., unequal means and covariance matrices) or does not guarantee global optimality of their solutions [16], [15].

Here, we solve this nonconvex problem globally for the case $p = 1$, with small computational cost and full generality (i.e., we allow arbitrary means and covariances). This is achieved by reducing the original problem to a one-dimensional one with an interval constraint. Our optimal solution method for $p = 1$ is then used as a building block for the case $p > 1$. We propose an incremental, greedy approach in which the p columns of the Stiefel matrix $E : n \times p$ are obtained one after the other: addition of each column consists of using our $p = 1$ subroutine in successively lower dimensional spaces. Thus, our results are directly applicable to other linear dimensionality reduction problems (here, we focus on the KL distance, but our $p = 1$ method can be extended to the J -divergence and Chernoff distance).

The range space of an optimal Stiefel matrix can be interpreted as the most informative p -dimensional subspace in the observation space \mathbb{R}^n . After the Stiefel matrix is designed, we find the 0/1 selection matrix with the range space closest to the suggested one. This can be done in closed-form and involves almost no computations.

We perform extensive simulations to show how our approach closely approximates the optimal one in the case where the latter can be calculated, and how it performs better than other suboptimal schemes with comparable computational complexity.

Paper organization. The rest of the paper is organized as follows. In section II we formulate the sensor selection problem. In section III we detail our sensor selection algorithm. In section IV simulation results are presented and section V summarizes the work presented in this paper.

II. PROBLEM FORMULATION

Gaussian hypothesis test. Let $x = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n$ be the vector of all sensor measurements, where x_i denotes

the i th sensor measurement. It is assumed that the data vector $x \in \mathbb{R}^n$ was generated by one of two known Gaussian distributions. Accordingly, we face the hypothesis test

$$\begin{aligned} H_0 &: x \sim \mathcal{N}(m_0, S_0) \\ H_1 &: x \sim \mathcal{N}(m_1, S_1) \end{aligned} ,$$

where $\mathcal{N}(m_i, S_i)$ denotes the Gaussian distribution with mean m_i and covariance matrix S_i . Hypothesis H_1 (H_0) corresponds to occurrence (non-occurrence) of the event of interest.

Sensor selection. To meet power consumption restrictions, we assume only p sensors are allowed to transmit their measurements to a fusion node. Mathematically, this translates into selecting p out of the n sensors and conduct the hypothesis test in the lower dimensional space \mathbb{R}^p . Sensor selection corresponds to finding a selection matrix $E \in \mathbb{R}^{n \times p}$ with 0/1 entries such that $y = E^\top x$. Each column of E has exactly a 1 at some coordinate i , meaning that sensor i is selected (there is at most one 1 per row). The remaining entries of E are 0. Thus, each choice of E implements a data projection

$$\mathbb{R}^n \rightarrow \mathbb{R}^p \quad x \mapsto y = E^\top x,$$

and induces a hypothesis test in the lower dimension space \mathbb{R}^p given by

$$\begin{aligned} H_0 &: y \sim \mathcal{N}(E^\top m_0, E^\top S_0 E) \\ H_1 &: y \sim \mathcal{N}(E^\top m_1, E^\top S_1 E) \end{aligned} .$$

Stein’s lemma. Before proceeding, we recall Stein’s Lemma [14], a fundamental result from detection theory. It relates the performance of a Neyman-Pearson (NP) detector to the dissimilarity of the two distributions to be tested. More precisely, it states that if a NP detector is employed to discriminate between $H_0 : x \sim p_0$ and $H_1 : x \sim p_1$, then the probability of false alarm decreases exponentially fast to zero when the number of data samples – assumed independent and identically distributed (i.i.d.) – goes to infinity. Indeed, if $P_{\text{FA}}(k)$ denotes the probability of false alarm when k i.i.d. samples are processed, then

$$\lim_{k \rightarrow \infty} \frac{\log P_{\text{FA}}(k)}{k} = -D_{\text{KL}}(p_1 \| p_0)$$

where

$$D_{\text{KL}}(p_1 \| p_0) = \int p_1(x) \log \left(\frac{p_1(x)}{p_0(x)} \right) dx$$

denotes the Kullback-Leibler distance between the distributions p_0 and p_1 . Thus, more dissimilar distributions lead to lower probabilities of false alarm in the asymptotic regime.

Optimization problem. Inspired by Stein’s lemma, we propose to search for the selection matrix E which yields the largest KL distance between the induced p -dimensional distributions. That is, we formulate the following optimization problem

$$\begin{aligned} &\text{maximize} && f(E) \\ &\text{subject to} && E_{ij} \in \{0, 1\} \\ &&& E^\top E = I_p \end{aligned} \quad (1)$$

where I_p denotes the $p \times p$ identity matrix and

$$f(E) := D_{\text{KL}} \left(\mathcal{N}(E^\top m_1, E^\top S_1 E), \mathcal{N}(E^\top m_0, E^\top S_0 E) \right).$$

It can be shown that

$$f(E) = \frac{1}{2} \left\{ \text{tr} \left((E^\top S_0 E)^{-1} (E^\top S_1 E) \right) + \frac{(m_1 - m_0)^\top E (E^\top S_0 E)^{-1} E^\top (m_1 - m_0)}{\log \frac{\det(E^\top S_1 E)}{\det(E^\top S_0 E)} - p} \right\}.$$

III. SENSOR SELECTION ALGORITHM

The optimization problem (1) is combinatorial and solving it by searching over all $\binom{n}{p}$ combinations of sensors is intractable for sufficiently large n and p . We propose a sub-optimal approach with computational cost of order $\mathcal{O}(n^3 p)$. It consists of two phases:

Phase 1: relaxation. We discard the 0/1 constraints and address the relaxed optimization problem

$$\begin{aligned} & \text{maximize} && f(E) \\ & \text{subject to} && E^\top E = I_p \end{aligned} \quad (2)$$

Thus, we seek to maximize the objective f over the manifold of Stiefel matrices. This phase has computational complexity of order $\mathcal{O}(n^3 p)$.

Phase 2: polishing. After an approximate solution of (2) is found, a local search around a nearby selection matrix is performed. This entails $(n-p)p$ evaluations of f .

A. Phase 1: relaxation

Problem (2) is nonconvex and still difficult to solve.

Case $p = 1$. For $p = 1$, the Stiefel matrix $E : n \times p$ reduces to an unit-norm vector, say $E = e$, and problem (2) consists of

$$\begin{aligned} & \text{maximize} && \frac{1}{2} \left\{ \frac{e^\top S_1 e}{e^\top S_0 e} + \frac{((m_1 - m_0)^\top e)^2}{e^\top S_0 e} - \log \frac{e^\top S_1 e}{e^\top S_0 e} - 1 \right\} \\ & \text{subject to} && e^\top e = 1 \end{aligned} \quad (3)$$

It can be shown that (3) reduces to the following one-dimensional problem:

$$\begin{aligned} & \text{maximize} && \gamma(t) \\ & \text{subject to} && t \in [1/\lambda_{\max}(S), 1/\lambda_{\min}(S)] \end{aligned} \quad (4)$$

where the objective is

$$\gamma(t) = \lambda_{\max}(S - tS + M) + \log(t), \quad (5)$$

$S := S_0^{-1/2} S_1 S_0^{-1/2}$ and $M := mm^\top$ with $m = S_0^{-1/2}(m_1 - m_0)$. Here, $\lambda_{\min}(S)$ and $\lambda_{\max}(S)$ denote the minimum and maximum eigenvalues of the matrix S , respectively. The equivalence of (3) and (4) is a major contribution of this paper. Let t^* solve (4) and let u_{\max} be an unit-norm eigenvector associated with the maximum eigenvalue of $S - t^*S + M$. Then $S_0^{-1/2} u_{\max} / \|S_0^{-1/2} u_{\max}\|$ solves (3).

We can solve (4) by discretizing the constraint interval $[1/\lambda_{\max}(S), 1/\lambda_{\min}(S)]$ and picking the point in the grid with the highest objective γ . However, a more efficient approach is a bound maximization iterative approach which

works as follows. Let t_k be the current iteration. We consider a lower bound function γ_k which touches γ at $t = t_k$, i.e., $\gamma_k(t) \leq \gamma(t)$ for all feasible t and $\gamma_k(t_k) = \gamma(t_k)$. The next iteration is given by maximizing the lower bound

$$\begin{aligned} t_{k+1} = & \arg \max_{t \in [1/\lambda_{\max}(S), 1/\lambda_{\min}(S)]} \gamma_k(t) \end{aligned} \quad (6)$$

Note that $\gamma(t_{k+1}) \geq \gamma_k(t_{k+1}) \geq \gamma_k(t_k) = \gamma(t_k)$. Thus, we're making progress with respect to the true objective γ . This iterative approach is attractive if the bound γ_k is simple enough to make (6) easy to solve. In our case, we can find such a convenient bound. We reason as follows. It is straightforward to show that $t \mapsto \phi(t) := \lambda_{\max}(S - tS + M)$ is a convex function. Thus,

$$\phi(t) \geq \phi(t_k) + g_k(t - t_k) \quad \text{for all } t, \quad (7)$$

where g_k denotes an element of the subdifferential of ϕ at t_k . We can pick $g_k = -u_k^\top S u_k$ where u_k denotes a unit-norm eigenvector associated with the maximum eigenvalue of $S - t_k S + M$. Plugging (7) into (5) yields the concave lower bound

$$\gamma_k(t) = \phi(t_k) + g_k(t - t_k) + \log(t). \quad (8)$$

For this choice of γ_k , the solution of (6) is trivial and given by

$$t_{k+1} = -\frac{1}{g_k}.$$

In our simulations, we observed that usually it takes only 5 to 7 iterations for this iterative method to converge to the global maximizer of (4) with precision 10^{-7} .

Case $p > 1$. We tackle the generic case $p > 1$ through an incremental, greedy approach in which the p columns of $E = [e_1 e_2 \dots e_p] \in \mathbb{R}^{n \times p}$ are obtained one after the other: first e_1 , then e_2 , and so on. Appending one column consists of solving (4) with the problem data (m, S) living in a successively lower-dimensional space, as orthogonality with the previously found columns is enforced at each step. Algorithm 1 details this approach.

Algorithm 1 Greedy algorithm

- 1: **for** $j = 1$ to p **do**
 - 2: Compute $U^{(j)} \in \mathbb{R}^{n \times (n-j+1)}$ ($U^{(1)} := I_n$), an orthonormal basis for the orthogonal complement of the $j-1$ dimensional subspace $\text{span}\{e_1, e_2, \dots, e_{j-1}\}$
 - 3: Compute the projected means and covariances $m_i^{(j)} = U^{(j)\top} m_i$, $S_i^{(j)} = U^{(j)\top} S_i U^{(j)}$ for $i = 0, 1$
 - 4: Compute $S^{(j)} = (S_0^{(j)})^{-1/2} S_1^{(j)} (S_0^{(j)})^{-1/2}$, $m^{(j)} = (S_0^{(j)})^{-1/2} (m_1^{(j)} - m_0^{(j)})$, $M^{(j)} = m^{(j)} m^{(j)\top}$
 - 5: Let $e^{(j)} \in \mathbb{R}^{n-j+1}$ denote a solution of (3) for $(m, S) = (m^{(j)}, S^{(j)})$
 - 6: Compute the j th column of E as $e_j = U^{(j)} e^{(j)}$
 - 7: **end for**
-

B. Phase 2: polishing

Let \widehat{E} be the Stiefel matrix produced by the greedy algorithm. In general, this will not be a selection matrix, i.e., with 0/1 entries. Thus, we need a mechanism to return to the feasible set of (1). A tempting approach would be to round each column of \widehat{E} to the nearest 0/1 vector (with only one entry equal to 1). However, note the intrinsic ambiguity exhibited by the objective function of (1). We have $f(E) = f(EQ)$ for every Stiefel matrix E and orthogonal $p \times p$ matrix Q . This invariance with respect to right multiplication of orthogonal matrices highlights the fact that f depends on E only through the p -dimensional subspace spanned by its columns, and not on its particular entries. Thus, instead of rounding the entries of \widehat{E} , we round its range space. More precisely, we look for the 0/1 selection matrix with range space closest to the range space of \widehat{E} . Since EE^\top is the orthogonal projector onto the range space of a Stiefel matrix E , this translates into solving

$$\begin{aligned} \widetilde{E} = \arg \min & \quad \left\| EE^\top - \widehat{E}\widehat{E}^\top \right\| \\ \text{subject to} & \quad E_{ij} \in \{0, 1\} \\ & \quad E^\top E = I_p, \end{aligned} \quad (9)$$

where $\|\cdot\|$ denotes the Frobenius norm. Note that this rounding approach captures the aforementioned invariance: both \widehat{E} and $\widehat{E}Q$ (where Q is orthogonal) yield the same solution \widetilde{E} (up to a permutation of columns, which is unavoidable), whereas rounding directly \widehat{E} and $\widehat{E}Q$ could lead to fundamentally distinct selection matrices. Solving (9) is straightforward: if (j_1, j_2, \dots, j_p) denote the indices of the largest entries of the diagonal of the projector $\widehat{E}\widehat{E}^\top$, then $\widetilde{E} = [i_{j_1} \ i_{j_2} \ \dots \ i_{j_p}]$ where i_j stands for the j th column of the identity matrix I_n .

Local refinement. We finish with a local maximization of f around \widetilde{E} . We sweep the p columns of \widetilde{E} , trying to improve each one. This consists of p steps. In the i th step ($i = 1, 2, \dots, p$) all columns of the current selection matrix are fixed except the i th one, which is viewed as an optimization variable. Solving (1) only over this column is straightforward and consists of comparing $n - p + 1$ evaluations of f . We denote the result of this local refinement by E^* .

Computational complexity. We now comment on the overall computational complexity of the proposed algorithm, by going through both of its phases. Phase 1 (relaxation phase) consists of p loops, where the complexity decreases at each next loop (see algorithm 1). For the j -th loop ($j = 1, \dots, p$), steps 2) and 4) bear main computational burden; step 2) requires computation of the orthogonal complement of the the collection of vectors $\{e_1, e_2, \dots, e_{j-1}\}$ and has complexity of order $\mathcal{O}((n - j + 1)^3)$. Step 4) requires computation of $-1/2$ power of matrix $S_0^{(j)}$ of the dimension $n - j + 1$ and it also has complexity of order $\mathcal{O}((n - j + 1)^3)$.

In phase 2) (polishing phase), major computation is consumed in the local refinement step. It consists of $p(n - p + 1)$ evaluations of the cost function $f(\cdot)$, where each cost function evaluation requires inversion of a $p \times p$ matrix. A naive implementation of phase 2 entails $\mathcal{O}(np^4)$. However,

TABLE I
KL DISTANCES FOR OPTIMAL AND SUBOPTIMAL SELECTION

KL distance	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 8$
SUBOPT						
n						
10	1.0430	1.8802	3.7991	4.8723	11.3714	59.1499
20	1.0579	2.4793	3.7191	4.4632	4.8182	8.8611
30	1.0225	2.7789	4.4219	5.6727	6.5402	10.0490
40	1.1092	3.1995	7.0466	8.5552	13.0797	19.0771
50	1.0875	2.3586	3.3338	5.2191	6.1373	
OPTIMAL						
n						
10	1.0430	2.3332	3.7991	6.1536	11.3714	59.1499
20	1.0579	2.4793	3.7191	4.4632	5.3216	8.9598
30	1.0225	2.7789	4.4219	5.6727	7.0785	10.0490
40	1.1092	3.1995	7.0466	9.4557	13.0797	21.9207
50	1.0875	2.3586	3.6754	5.3809	6.8906	

structure can be exploited to lower down this estimate, and the details are left for a journal version of this work.

Combining complexities of phase 1 and phase 2 we get that the overall complexity of the proposed algorithm is $\mathcal{O}(n^3 p)$.

IV. SIMULATION RESULTS

In this section we will provide simulation results on the performance of the proposed algorithm. All simulation results are obtained using MATLAB.

To test the performance of our algorithm, we compare the values of the KL distance for the subset of sensors obtained by our algorithm with the KL distance of the optimal subset of sensors, obtained by exhaustive search. We randomly generated parameters for two Gaussian distributions for cases $n = 10, 20, 30, 40, 50$, and for each of them we ran our algorithm for $p = 1, 2, 3, 4, 5, 8$. The results are shown in Table 1. It can be seen that the values of KL distance match in most of the cases the optimal ones, and, in the worst case, stay within a mere 20% distance from the global optimum.

We also compared the detector performances of our sub-optimal solution to the one relative to the optimal choice of sensors. To accomplish this we used a maximum likelihood detector (MLD). For each of the cases stated above, we ran 10^5 instantiations of tests for the samples coming from each of the two distributions. Based on the data from tests, we calculated probabilities of false alarm (PFA). The results are shown in Table 2. Values of PFA for suboptimal choice of sensors are very close to the ones for the optimal choice.

Note that, in Table 2, the probability of false alarm of the suboptimum approach is lower than for the optimal choice for some simulated configurations, e.g., ($n = 20, p = 8$). This is no paradox. Indeed, Stein's lemma only states an asymptotic connection between the probability of false alarm and the KL distance. In our case, we are not in the asymptotic regime as we decide on the basis of only $k = 1$ sample. To illustrate further this point, we also performed simulations closest to the asymptotic regime, namely, with $k = 10$. That

TABLE V
KL DISTANCES FOR SUBOPTIMAL SELECTION MATRIX AND BEST RANDOM SELECTION MATRIX

	$n = 50$			$n = 100$			$n = 200$		
	$p = 5$	$p = 10$	$p = 20$	$p = 10$	$p = 20$	$p = 40$	$p = 20$	$p = 40$	$p = 80$
SUBOPTIMAL KL dist	0.784146	2.046895	5.699933	0.435614	1.633185	7.035000	0.763321	2.596785	11.050338
BEST RANDOM KL dist	0.784513	1.592929	4.827790	0.303585	1.006640	5.127912	0.509236	1.863333	7.983869

TABLE II

PROBABILITY OF FALSE ALARM (PFA) FOR OPTIMAL AND SUBOPTIMAL SELECTION

PFA	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 8$
SUBOPT						
n						
10	0.2324	0.1667	0.1109	0.1097	0.0641	0.0185
20	0.2332	0.1349	0.0880	0.0687	0.0623	0.0341
30	0.2371	0.1250	0.0821	0.0605	0.0502	0.0330
40	0.2293	0.1120	0.0560	0.0417	0.0260	0.0090
50	0.2294	0.1380	0.1005	0.0608	0.0464	
OPTIMAL						
n						
10	0.2324	0.1373	0.1109	0.1102	0.0641	0.0185
20	0.2332	0.1349	0.0880	0.0687	0.0685	0.0412
30	0.2371	0.1250	0.0821	0.0605	0.0503	0.0330
40	0.2293	0.1120	0.0560	0.0361	0.0260	0.0102
50	0.2294	0.1380	0.0923	0.0595	0.0444	

TABLE III

KL DISTANCES FOR OPTIMAL AND SUBOPTIMAL SELECTION

KL distance	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 8$
SUBOPT						
n						
10	0.0011	0.1808	0.3723	0.6157	0.5887	1.9549
20	0.0012	0.1245	0.3190	0.4792	0.9392	1.9089
30	0.0012	0.0807	0.2062	0.3692	0.5840	1.7459
40	0.0011	0.0463	0.0975	0.1365	0.2571	0.5275
50	0.0012	0.0610	0.1972	0.2893	0.4292	0.8135
OPTIMAL						
n						
10	0.0011	0.1808	0.3723	0.6157	0.9044	1.9549
20	0.0012	0.1609	0.4744	0.6936	1.1677	3.8261
30	0.0012	0.1270	0.3221	0.4751	0.6951	1.8829
40	0.0011	0.0739	0.1348	0.1996	0.2846	0.6690
50	0.0012	0.0933	0.1972	0.2952	0.4381	0.9775

is, the detector chooses one of the hypothesis on the basis of a batch of $k = 10$ i.i.d. samples. Tables 3 and 4 are the counterparts of Tables 1 and 2 for this scenario, although different Gaussian distributions were used. We can see that in the configurations where the aforementioned phenomenon occurs the performance gap decreases significantly.

The receiver operating characteristics (ROC) of MLD for cases ($n = 40, p = 2$) and ($n = 40, p = 4$) are shown in Fig.1. For the first case, ($n = 40, p = 2$), the algorithm found the

TABLE IV

PROBABILITY OF FALSE ALARM (PFA) FOR OPTIMAL AND SUBOPTIMAL SELECTION

PFA	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 8$
SUBOPT						
n						
10	0.4731	0.1757	0.1029	0.0508	0.0487	0.0023
20	0.4741	0.2230	0.1176	0.0702	0.0285	0.0043
30	0.4705	0.2679	0.1605	0.1051	0.0609	0.0088
40	0.4707	0.3257	0.2491	0.2016	0.1396	0.0679
50	0.4767	0.2897	0.1679	0.1229	0.0943	0.0269
OPTIMAL						
n						
10	0.4731	0.1757	0.1029	0.0508	0.0290	0.0023
20	0.4741	0.1976	0.0778	0.0523	0.0296	0.0014
30	0.4705	0.2186	0.1175	0.0791	0.0478	0.0093
40	0.4707	0.2804	0.2176	0.1704	0.1255	0.0513
50	0.4767	0.2533	0.1679	0.1188	0.0804	0.0277

optimal selection and thus the ROC curves coincide. For the other case, ($n = 40, p = 4$), it can be seen that the performance of the suboptimal sensor subset is very close to the optimal one.

It is also easy to see that, for fixed n , the detection performance increases with p , as expected. In addition, for practical issues, finding (sub)optimal KL distance for specific choice of p gives a measure of best achievable detection performance for given p . This provides invaluable information for design. By computing (sub)optimal solutions for different values of p we can determine the minimum number of sensors required to satisfy a certain performance specification. Our algorithm provides tight lower bound on best performance with small computational cost.

Finally, we tested the performance of our algorithm at scale. As the number of sensors grows the combinatorial problem becomes infeasible. In this case we compared the values of KL distance given by the proposed approach, with the values of KL distance for the best subset of sensors (the one that gives highest value of KL distance) among 10^5 randomly chosen subsets, for $n = 50, 100, 200$ and $p = 10, 20, 40$ % of n . The results are shown in Table 5. The comparison shows that with the exception of the case ($n = 50, p = 5$) for which the difference between values of KL distance is very small, our proposed approach yields far greater performance together with significant savings in computation time. For example, for the case $n = 100, p = 10$ it takes 2 sec to find

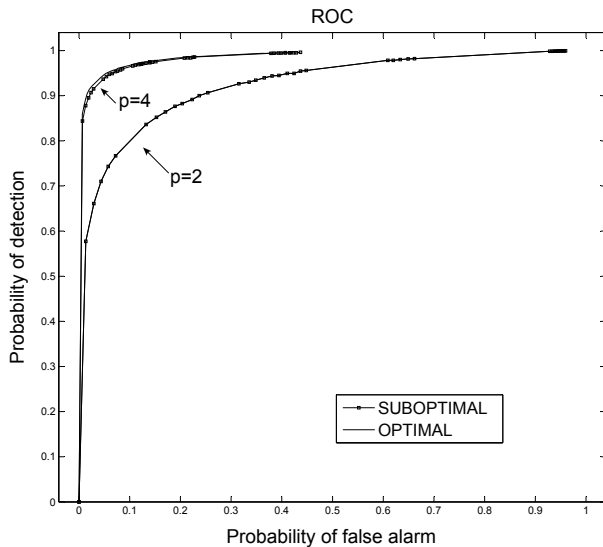


Fig. 1. ROC curves for optimal and suboptimal selection

sensor selection matrix, while to check 10^5 100×10 matrices it takes about 2 min on a 1.7 GHz personal computer.

V. CONCLUSION

In this paper we propose a sensor selection algorithm for event detection in wireless sensor networks. We propose to reduce the dimensionality of the problem by selecting p sensors, out of n , that maximize the KL distance between the selected measurement distributions. Selecting sensors is amenable to sensor networks as it decreases the sensing, computational and communication burden, increasing this way its lifetime. The computational complexity of our selection strategy is of order $\mathcal{O}(n^3 p)$ operations, compared to the non scalable original combinatorial problem. Our solution, although suboptimal in principle, shows near optimal performance in most cases. For large systems, where the original combinatorial solution is not computable, our algorithm shows better performance and lower computational time if compared with Montecarlo-bases suboptimal solutions. Our results are also of immediate interest for linear dimensionality reduction applications since, as an intermediate step, we addressed the problem of finding the linear map $\mathbb{R}^n \mapsto \mathbb{R}^p$ which maximizes the KL distance between distributions in the lower p -dimensional space. We globally solved this nonconvex problem for the case $p = 1$ and capitalized on that tool to tackle the generic case $p > 1$ via an incremental, greedy approach, which provides near optimal result with small computational cost.

REFERENCES

- [1] S. Joshi, S. Boyd Sensor Selection via Convex Optimization, *IEEE Transactions on Signal Processing*, 57(2) pp 451-462, February 2009
- [2] J. E. Weimer, B. Sinopoli, B. H. Krogh A Relaxation Approach to Dynamic Sensor Selection in Large-Scale Wireless Networks, in *Proceedings of the 2008 The 28th International Conference on Distributed Computing Systems Workshops* pp 501-506, 2008
- [3] F. Zhao, J. Shin, J. Reich Information-Driven Dynamic Sensor Collaboration for Tracking Applications, *IEEE Signal Processing Magazine* vol. 19(2), pp 61-72, March

- [4] H. Rowaihy, S. Eswaran, M. Johnson, D. Verma, A. Bar-Noy, T. Brown, and T. L. Porta A survey of sensor selection schemes in wireless sensor networks, Technical report, Pennsylvania State University, March 2007
- [5] John N. Tsitsiklis Decentralized detection by a large number of sensors, *Mathematics of Control, Signals and Systems*, vol. 1, pp 167-182, 1988
- [6] Z. Chair and P. K. Varshney Optimal data fusion in multiple sensor detection systems, *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 1, pp 98-101, Jan. 1986
- [7] P. K. Willett, P. F. Swaszek, and R. S. Blum The good, bad, and ugly: Distributed detection of a known signal in dependent Gaussian noise, *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp 3266-3279, Dec. 2000
- [8] M. Kam, Q. Zhu, and W. S. Gray Optimal data fusion of correlated local decisions in multiple sensor detection systems, *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 7, pp 916-920, July 1992
- [9] C. Rago, P. Willett, Y. Bar-Shalom Censoring sensors: a low communication - rate scheme for distributed detection, *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp 554-568, April 1996
- [10] R. Jiang, B. Chen Fusion of censored decisions in WSN, *IEEE Trans. Wireless Commun.*, vol. 4, pp 2668-2673, Nov. 2005
- [11] S. Kumar, K. K. R. Kambhatla, B. Zan, F. Hu, Y. Xiao An Energy-aware and intelligent cluster-based event detection scheme in WSN, *Internal Journal of Sensor Networks*, vol. 3, Issue 2 pp 123-133, Feb. 2008
- [12] E. B. Fox, J. W. Fisher, A. S. Willsky Detection and localization of material releases with sparse sensor configurations, *IEEE Trans. Signal Processing*, vol. 55, no. 5, May 2007
- [13] A. D'Costa, V. Ramachandran, A. M. Sayeed Distributed classification of Gaussian space-time sources in WSN, *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, August 2004
- [14] T. M. Cover, J. A. Thomas *Elements of information theory*, Wiley, 1991
- [15] L. Rueda, M. Herrera. Linear dimensionality reduction by maximizing the Chernoff distance in the transformed space, *Pattern Recognition*, vol. 41(10), pp 3138-3152, October 2008
- [16] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation and Time series Analysis*, Addison and Wesley, 1991
- [17] D. H. Johnson, S. Sinanovic Symetrizing the Kullback-Leibler Distance, Computer and Information Technology Institute, Department of Electrical Engineering, Rice University, Houston
- [18] J. Yang, H. Ye, Z. David A new LDA-KL combined method for feature extraction and its generalization, *Pattern Analysis and Applications*, vol. 7(1), pp 40 - 50, April 2004
- [19] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002
- [20] R. Ambrosino, B. Sinopoli, K. Poolla, S. S. Sastry, Optimal Sensor Density for Remote Estimation OverWireless Sensor Networks, *Forty-Sixth Annual Allerton Conference*, September 2008