SPARSE SENSING FOR DISTRIBUTED GAUSSIAN DETECTION

Sundeep Prabhakar Chepuri and Geert Leus

Faculty of Electrical Engineering, Mathematics, and Computer Science (EEMCS)
Delft University of Technology (TU Delft), The Netherlands
Email: {s.p.chepuri; g.j.t.leus}@tudelft.nl.

ABSTRACT

An offline sampling design problem for Gaussian detection is considered in this paper. The sensing operation is modeled by a selection vector, whose sparsity order is determined by the prescribed global error probability. Since the numerical optimization of the error probability is difficult, equivalent simpler costs, viz., the Kullback-Liebler distance and Bhattacharyya distance are optimized. The sensing problem is formulated and solved sub-optimally using convex optimization techniques. It is shown that the sensing problem can be solved optimally for conditionally independent Gaussian observations. Further, we show that for non-identical sensor observations, the number of sensors required to achieve a certain detection performance decreases as the sensors become more correlated.

Index Terms— Sensor networks, sparse sensing, sensor selection, sensor placement, detection, convex optimization, sparsity.

1. INTRODUCTION

In this work, we focus on distributed detection, where a field (e.g., heat, target signal) is sampled by a set of spatially distributed sensors, and these samples are made available at the fusion center. Subsequently, the fusion center makes a single global decision as to the true state of nature using binary hypothesis testing. Some pertinent examples are: heat detection — temperature is above or below a certain threshold, radar — target is present or absent, and spectrum sensing — primary user is active or inactive.

In such applications, the number of sensors available is limited due to economical or energy constraints, or there might not be sufficient processing capabilities and/or communication bandwidth available. Thus, it is crucial to smartly design the sensing task. Naturally, limiting the number of sensors also restricts the achievable detection performance. In this paper, we are interested in designing the sensing operator $\mathbf{w} \in \{0,1\}^M$ that jointly minimizes the number of sensors and the probability of error. Here, M is the number of candidate sensors. In other words, we assume a certain candidate set of (temporal and/or spatial) locations and the best sampling locations are chosen through $\mathbf{w}$. The problem is that the expressions of the error probabilities are not favorable for numerical optimization. Hence, we adopt simpler substitutes for the error probabilities, viz., the Kullback-Leiber distance and Bhattacharyya distance, which belong to the general class of Ali-Silvey distances [1]. Interestingly, for Gaussian observations optimizing these proxies are optimal in terms of the error probabilities.

The central question of interest, i.e., sensing design for Gaussian detection problems, has been studied in the past [2,3]. In [2], this problem has been solved through techniques which are more likely to lead to a local optimum. Similarly, the formulation in [3] results in a complex non-convex (even after appropriate relaxations) solver on the Stiefel manifold. Different from [2,3], the proposed formulation leads to an elegant convex optimization solver. More interestingly, it enables us to explore cases where we can solve the problem optimally and to further understand the effect of correlation among sensors on sensing design. Some more variants of sensing design, but in the context of estimation and filtering can be found in [4–8] (and references therein).

This work was supported in part by STW under FASTCOM project (10551) and in part by NWO-STW under the VICI program (10382).

2. PROBLEM MODELING

Let M be the number of available sensors. For example, this might be a set of candidate locations where we can place the sensors. The observations at each sensor are related to the state of nature $\mathcal{H}$. In a binary hypothesis testing problem, the random variable $\mathcal{H}$ is drawn from a binary set $\{\mathcal{H}_0, \mathcal{H}_1\}$. Furthermore, in the Bayesian setting, we assume that the prior probabilities $\pi_0 = \Pr(\mathcal{H}_0)$ and $\pi_1 = \Pr(\mathcal{H}_1)$ are known, whereas in the classical setting, the prior probabilities are not known. Consider the case of binary signal detection in Gaussian noise, where the related conditional distributions are given by

$$\begin{aligned} \mathcal{H}_0 :&\quad \mathbf{x} \sim p(\mathbf{x}|\mathcal{H}_0) = \mathcal{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}) \\ \mathcal{H}_1 :&\quad \mathbf{x} \sim p(\mathbf{x}|\mathcal{H}_1) = \mathcal{N}(\boldsymbol{\theta}_1, \boldsymbol{\Sigma}). \end{aligned} \tag{1}$$

Here, $\mathbf{x} = [x_1, x_2, \ldots, x_M]^T$ is an $M \times 1$ observation vector, and the mean vectors $\boldsymbol{\theta}_0 \in \mathbb{R}^M$ and $\boldsymbol{\theta}_1 \in \mathbb{R}^M$, as well as the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times M}$ are assumed to be perfectly known

Let $\widehat{\mathcal{H}}$ denote an estimate of the state of nature $\mathcal{H}$, based on a certain decision rule. Classic approaches to solve the binary hypothesis testing problem (1) include the Neyman-Pearson and Bayes test. In the classical setting, the optimal detector is the well-known Neyman-Pearson detector that minimizes the probability of miss detection (type II error),

$$P_M = 1 - \Pr(\widehat{\mathcal{H}} = \mathcal{H}_1|\mathcal{H}_1)$$

for a fixed probability of false alarm (type I error),

$$P_F = \Pr(\widehat{\mathcal{H}} = \mathcal{H}_1|\mathcal{H}_0).$$

In the Bayesian setting, the optimal detector minimizes the Bayesian error probability,

$$P_E = \Pr(\widehat{\mathcal{H}} \neq \mathcal{H}) = \pi_0 P_F + \pi_1 P_M.$$

Let us model the sensing operator through a Boolean vector

$$\mathbf{w} = [w_1, w_2, \ldots, w_M] \in \{0,1\}^M,$$

where the mth sensor is chosen if $w_m = 1$, otherwise it is not chosen. That is, we gather the data using a compression matrix $\mathbf{\Phi}(\mathbf{w}) = \mathrm{diag_r}(\mathbf{w}) \in \{0,1\}^{K \times M}$ as

$$\mathbf{y} = \mathbf{\Phi}(\mathbf{w})\mathbf{x},$$

where K ($\ll M$) is the number of selected sensors and $\mathrm{diag_r}(\cdot)$ represents a diagonal matrix with the argument on its diagonal, but with the all-zero rows removed. We want to design $\mathbf{w}$ (and, hence $\mathbf{\Phi}$) with as many zeros as possible, where the number of non-zero entries of $\mathbf{w}$ is determined by the prescribed detection performance. We underline the fact that the sparsity order of $\mathbf{w}$ determines the possible reduction in the sensing costs, and other overheads. In this paper, the problem that we address is stated as follows.

Problem statement. *Given the conditional distribution of the observations, design the sensing operator* $\mathbf{w}$ *that chooses* K *sensors out of* M *available ones and also: (i) minimizes the Bayesian probability of error when the prior probabilities are known in the Bayesian setting; (ii) minimizes the probability of miss detection for a fixed probability of false alarm in the classical setting. Such a discrete sensing task is also referred to as sensor selection.*

Parameterizing the error probabilities with the sensing vector $\mathbf{w}$, the above stated problem can be mathematically expressed as the following optimization problem

$$\text{Classical}: \underset{\mathbf{w}\in\{0,1\}^M}{\arg\min}\ P_M(\mathbf{w}) \text{ s.to } P_F(\mathbf{w}) \leq \alpha, \|\mathbf{w}\|_0 = K; \quad (2a)$$

$$\text{Bayesian}: \underset{\mathbf{w}\in\{0,1\}^M}{\arg\min}\ P_E(\mathbf{w}) \text{ s.to } \|\mathbf{w}\|_0 = K, \quad (2b)$$

where α is the prescribed false-alarm rate and the notation $\|\mathbf{w}\|_0$ counts the number of non-zero entries in $\mathbf{w}$. Note that when K is not known, the problem can equivalently be posed as a cardinality minimization problem (i.e., minimize $\|\mathbf{w}\|_0$) subject to a constraint on the (Bayesian/classical) error probabilities. In what follows, we will discuss alternative performance measures for the error probabilities as the error probabilities are in general difficult to optimize.

3. OPTIMALITY CRITERION

In the classical setting or the Neyman-Pearson problem, the decision is based upon the log-likelihood ratio test

$$\log l(\mathbf{y}) = \log \frac{p(\mathbf{y}|\,\mathcal{H}_1)}{p(\mathbf{y}|\,\mathcal{H}_0)} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \gamma,$$

where $\log l(\mathbf{y})$ is the log-likelihood ratio and γ is the threshold. The error probabilities admit the following expressions [9, pg. 475]

$$P_F = \mathcal{Q}\left(\frac{\gamma + s/2}{\sqrt{s}}\right) \quad \text{and} \quad P_M = 1 - \mathcal{Q}\left(\frac{\gamma - s/2}{\sqrt{s}}\right), \quad (3)$$

where $\mathcal{Q}$ is the complementary Gaussian cumulative distribution function

$$\mathcal{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy,$$

and

$$s = \mathbf{m}^T \mathbf{\Phi}^T \mathbf{\Sigma}^{-1}(\mathbf{w}) \mathbf{\Phi} \mathbf{m}, \quad (4)$$

is the *signal-to-noise ratio* [9]. Here, $\mathbf{m} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_0$ and $\mathbf{\Sigma}^{-1}(\mathbf{w})$ is the inverse matrix of $\mathbf{\Sigma}(\mathbf{w}) = \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}^T \in \mathbb{R}^{K\times K}$, which includes only the entries corresponding to the selected sensors. Similarly, the Bayesian error probability is given by [9, pg. 494]

$$P_E = \pi_0 \mathcal{Q}\left(\frac{\gamma' + s/2}{\sqrt{s}}\right) + \pi_1 \left[1 - \mathcal{Q}\left(\frac{\gamma' - s/2}{\sqrt{s}}\right)\right], \quad (5)$$

where the log-likelihood ratio is compared to the threshold $\gamma' = \log(\pi_0/\pi_1)$ in the Bayesian setting.

The optimization problem (2) is difficult to solve because of the involved integral in the expression of the error probabilities (3) and (5). Instead of optimizing the error probabilities, we seek for simpler substitutes, and these are based on a distance measure between the conditional probabilities (1). The idea of distance measure between the probabilities has been extensively used in statistical experimental design (e.g., signal selection and waveform design) [1,10–12]. Some of the prominent distance measures are:

1. *Kullback-Leibler distance* [13] :

$$\mathcal{D}(\mathcal{H}_1 \| \mathcal{H}_0) = \mathbb{E}_{|\mathcal{H}_1}\{\log l(\mathbf{y})\}$$

or

$$\mathcal{D}(\mathcal{H}_0 \| \mathcal{H}_1) = -\mathbb{E}_{|\mathcal{H}_0}\{\log l(\mathbf{y})\}.$$

2. *Bhattacharyya distance* [11]:

$$\mathcal{B}(\mathcal{H}_1 \| \mathcal{H}_0) = -\log \mathbb{E}_{|\mathcal{H}_0}\{\sqrt{l(\mathbf{y})}\}.$$

The notation $\mathbb{E}_{|\mathcal{H}_i}\{\cdot\}$ indicates that the average is computed under the pdf $p(\cdot|\,\mathcal{H}_i)$. The Kullback-Leibler distance measure is the average log-likelihood and it is the best error exponent in the classical setting. The Bhattacharyya distance is the negative logarithm of the average root-likelihood. It is a special case of Chernoff information, which is the best error exponent in the Bayesian setting. For the detection problem (1), the above distance measures are given by [11]

$$\mathcal{D}(\mathcal{H}_1 \| \mathcal{H}_0) = \mathcal{D}(\mathcal{H}_0 \| \mathcal{H}_1) = \frac{s}{2}, \quad (6)$$

$$\mathcal{B}(\mathcal{H}_0 \| \mathcal{H}_1) = \frac{s}{8}. \quad (7)$$

Interestingly, both these distance measures are the same (more precisely, they are simply the signal-to-noise ratio) up to a constant. However, these relations are not universal (e.g., they do not hold for non-Gaussian observations).

Having introduced these distance measures, we now make the following observations, which assert that optimizing one of these distance measures is optimal for (2). For a fixed P_F, say α ($0 \leq \alpha \leq 1$), we have

$$P_M(\mathbf{w}) = 1 - \mathcal{Q}(\mathcal{Q}^{-1}(\alpha) - \sqrt{s(\mathbf{w})}).$$

The $\mathcal{Q}(x)$ function is monotonic in nature, i.e., $\mathcal{Q}(x_0) < \mathcal{Q}(x)$ for $x_0 > x > 0$. Hence, maximizing the signal-to-noise ratio over $\mathbf{w}$ minimizes P_M. In other words, for a certain $s(\mathbf{w}) > s(\mathbf{w}_0)$ it is easy to verify that $P_M(s(\mathbf{w})) < P_M(s(\mathbf{w}_0))$. Furthermore, this also holds for the Bayes test, i.e., for a given $\{\pi_0, \pi_1\}$ pair, if $s(\mathbf{w}) > s(\mathbf{w}_0)$, then $P_E(s(\mathbf{w})) < P_E(s(\mathbf{w}_0))$. In essence, for the considered problem, we can safely replace the error probabilities in (2) with the signal-to-noise ratio s, without any loss of optimality.

4. OPTIMIZATION PROBLEM

In this section, we reformulate (2) as a problem of signal-to-noise ratio (or Kullback-Leibler distance or Bhattacharyya distance) maximization over a Boolean $\mathbf{w}$.

4.1. Equivalent problem

Let $\boldsymbol{\Sigma} = a\mathbf{I} + \mathbf{S}$ with a non-zero $a \in \mathbb{R}$ chosen such that $\mathbf{S} \in \mathbb{R}^{M \times M}$ is invertible and well-conditioned. Using the matrix inversion lemma [14] on $s(\mathbf{w}) = \mathbf{m}^T \boldsymbol{\Phi}^T \left[a\mathbf{I} + \boldsymbol{\Phi}\mathbf{S}\boldsymbol{\Phi}^T\right]^{-1} \boldsymbol{\Phi}\mathbf{m}$ in (4), we can express (4) as

$$s(\mathbf{w}) = \mathbf{m}^T\mathbf{S}^{-1}\mathbf{m} - \mathbf{m}^T\mathbf{S}^{-1}\left[\mathbf{S}^{-1} + a^{-1}\mathrm{diag}(\mathbf{w})\right]^{-1}\mathbf{S}^{-1}\mathbf{m}, \tag{8}$$

where by definition $\boldsymbol{\Phi}^T\boldsymbol{\Phi} = \mathrm{diag}(\mathbf{w})$. Note that in contrast to (4), the design parameter $\mathbf{w}$ only shows up at one place in (8), which makes the optimization problem much easier.

Consequently, the sparse sensing problem can be expressed as:

$$\begin{aligned} \underset{\mathbf{w}\in\{0,1\}^M}{\arg\min} \quad & \mathbf{m}^T\mathbf{S}^{-1}\left[\mathbf{S}^{-1} + a^{-1}\mathrm{diag}(\mathbf{w})\right]^{-1}\mathbf{S}^{-1}\mathbf{m} \\ \text{s.to} \quad & \|\mathbf{w}\|_0 = K, \end{aligned} \tag{9}$$

where only the second term of (8), which depends on $\mathbf{w}$ is optimized (minimization is due to its negative sign). Writing (9) in the epigraph form and using the Schur complement, (9) will equivalently be

$$\begin{aligned} \underset{\mathbf{w}\in\{0,1\}^M, t}{\arg\min} \quad & t \\ \text{s.to} \quad & \|\mathbf{w}\|_0 = K, \\ & \begin{bmatrix} \mathbf{S}^{-1} + a^{-1}\mathrm{diag}(\mathbf{w}) & \mathbf{S}^{-1}\mathbf{m} \\ \mathbf{m}^T\mathbf{S}^{-1} & t \end{bmatrix} \succeq \mathbf{0}, \end{aligned} \tag{10}$$

which is a combinatorial non-convex problem due to the Boolean and cardinality constraints. Here, $t \in \mathbb{R}$ is an auxiliary variable. We remark here that the actual measurements are not needed to solve this problem, i.e., it can be solved offline.

4.2. Relaxed problem

First of all, the Boolean constraint is relaxed with its convex hull, i.e., $0 \leq w_m \leq 1$, $m = 1, 2, \ldots, M$. We also relax the $\|\mathbf{w}\|_0$ constraint in (10) to its best convex approximate $\mathbf{1}^T\mathbf{w}$. Thus, the relaxed convex problem is given by

$$\begin{aligned} \underset{\mathbf{w}, t}{\arg\min} \quad & t \\ \text{s.to} \quad & \mathbf{1}^T\mathbf{w} = K, \\ & \begin{bmatrix} \mathbf{S}^{-1} + a^{-1}\mathrm{diag}(\mathbf{w}) & \mathbf{S}^{-1}\mathbf{m} \\ \mathbf{m}^T\mathbf{S}^{-1} & t \end{bmatrix} \succeq \mathbf{0}, \\ & 0 \leq w_m \leq 1, m = 1, 2, \ldots, M. \end{aligned} \tag{11}$$

This is a semidefinite programming problem that can be solved with off-the-shelf solvers, for example, SeDuMi [15]. The selected sensors (i.e., an approximate Boolean solution) are given by the non-zero entries of $\mathbf{w}$, or they can also be computed using the randomization techniques described in [6].

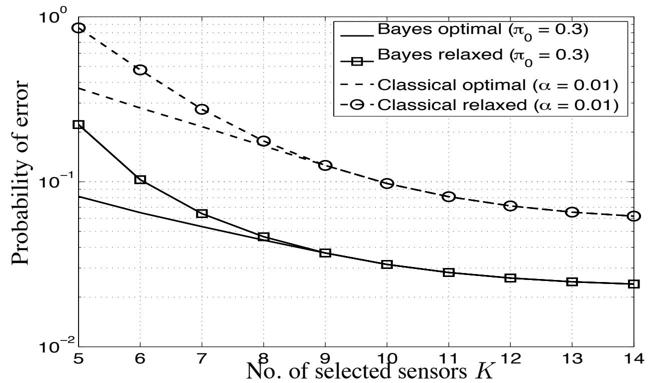
4.3. Numerical example

We illustrate the proposed framework with the following example. Consider the hypothesis testing problem (1) with $M = 15$, $\boldsymbol{\theta}_0 = \mathbf{0}$, and $[\boldsymbol{\theta}_1]_m = \cos 2\pi f m$ with $f := 0.33$ for $m = 1, 2, \ldots, M$. We use a smaller dimension for M to compare the results with the optimal solution of (2) obtained by exhaustive search. Nevertheless, the proposed framework is valid for higher dimensional problems. Let us assume that the covariance matrix $\boldsymbol{\Sigma}$ is of the form

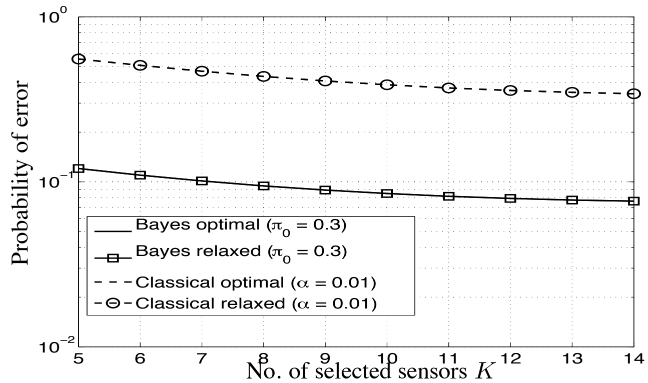
$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} = \sigma^2 \left[(1-\rho)\mathbf{I}_M + \rho\mathbf{1}_M\mathbf{1}_M^T\right],$$

with a known correlation coefficient $\rho = \{0, 0.5\}$ and variance $\sigma^2 = 1$. Finally, we use $\pi_0 = 0.3$, $\pi_1 = 0.7$, $\alpha = 0.01$, and $a = 0.11$. Note that any $a \neq 1 - \rho$ leads to an invertible $\mathbf{S}$ that can be used in the solver (11).

We solve (2) using a brute force evaluation of all the $\binom{K}{M}$ combinations. We remark here that the exhaustive search is computationally intractable even for the modest values of K and M. The convex relaxed problem (11) is solved using SeDuMi [15]. The probability of error, i.e., P_M in the classical setting and P_E in the Bayesian setting for different numbers of selected sensors is shown in Fig. 1a. The P_M and P_E obtained with the selected sensors are computed using (3) and (5), respectively. For this particular example, the performance of the relaxed problem is reasonable for smaller values of K, but the sensor selection is clearly not optimal. The performance for smaller values of K can be further improved with randomized rounding. Nevertheless, for larger values of K, the sensor selection is near-optimal in terms of the error probability.



(a)



(b)

Fig. 1: The (Bayesian/classical) probability of error for (1) with different numbers of selected sensors K out of $M = 15$ sensors. (a) Dependent observations ($\rho = 0.5$). (b) Independent observations ($\rho = 0$).

5. CAN WE SOLVE THE EQUIVALENT PROBLEM OPTIMALLY?

In the previous section, we have discussed a convex optimization approach to solve (2) sub-optimally. However, it is interesting to explore the question —when can we solve (9) optimally in polynomial time? The answer is, for conditionally independent Gaussian observations, i.e., when $\mathbf{\Sigma}$ is a diagonal matrix, we can solve (9) optimally in polynomial time.

For conditionally independent observations, the signal-to-noise ratio s can be expressed as the sum of the local signal-to-noise ratios evaluated at each sensor. For a diagonal matrix $\mathbf{\Sigma}$, defining $\sigma_m^2 = [\mathbf{\Sigma}]_{m,m}$, we can simplify (4) to $s(\mathbf{w}) = \sum_{m=1}^{M} w_m \left([\mathbf{m}]_m^2/\sigma_m^2\right)$. Subsequently, the sensor selection problem simplifies to a Boolean linear program

$$\arg\max_{\mathbf{w}\in\{0,1\}^M} \sum_{m=1}^{M} w_m \left(\frac{[\mathbf{m}]_m^2}{\sigma_m^2}\right) \quad \text{s.to} \quad \|\mathbf{w}\|_0 = K. \tag{12}$$

The above problem admits an explicit solution and computing the optimal solution is straightforward. It is solved simply by ordering the local signal-to-noise ratios and choosing the K largest ones.

More generally, for other distributions, the best subset of K independent sensors are those with the largest local Kullback-Leibler distance or Bhattacharyya distance, and they are optimal in terms of the error exponents. Therefore, for conditionally independent observations (not necessarily Gaussian), convex relaxations are not needed.

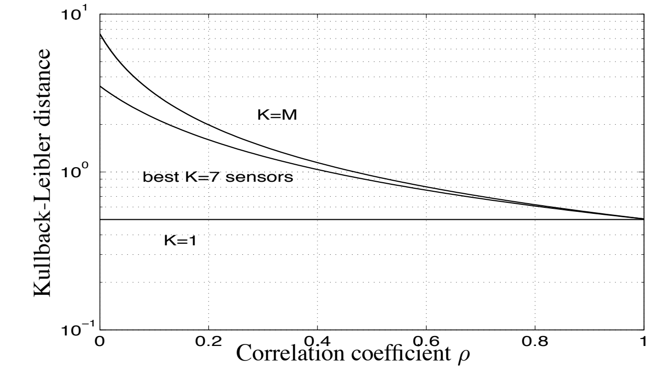
To illustrate the optimality of the proposed sensor selection for independent observations, we refer back to the numerical example introduced in Section 4.3, but we now use $\rho = 0$. We can see in Fig. 1b that the sensor selection (12) based on ordering is optimal in terms of the error probabilities.

6. IS CORRELATION GOOD OR BAD?

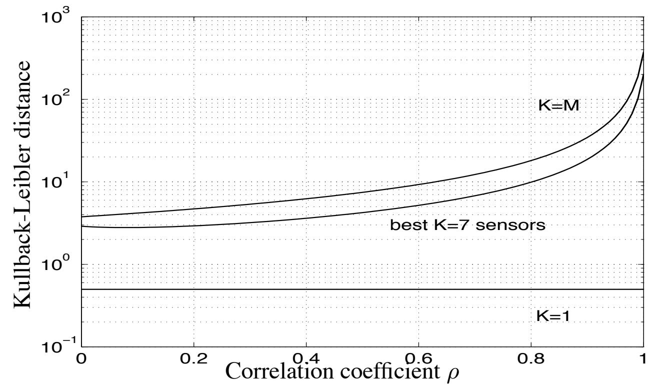
In this section, we extend some of the well-known results from distributed detection to sensor selection. In particular, we are interested in the number of sensors required to achieve a certain detection probability as the correlation coefficient ρ approaches 1. To illustrate this, let us consider the numerical example introduced in Section 4.3 with $f \in \{0, 0.33\}$.

We first consider the case when $f = 0$, where all the M sensors have identical observations ($\boldsymbol{\theta}_1 = \mathbf{1}$). Hence, any subset of sensors is also the best subset of sensors. In other words, for identical observations, random sensing is optimal. As the correlation coefficient ρ approaches 1, the amount of information (Kullback-Liebler distance/Bhattacharyya distance/signal-to-noise ratio) contributed by the best (i.e., a random) subset of $K > 1$ sensors is the same as that of the contribution from $K = 1$ sensor. This can be seen in Fig 2a. Similar results can be found in [16], but in the context of distributed detection over a wireless channel.

A more interesting case, in particular for sensing design problems, is when the observations are not identical ($f = 0.33$). When the observations are non-identical, as the correlation coefficient ρ approaches 1, the amount of information contained in the best subset of $K > 1$ sensors increases significantly; see Fig. 2b. More specifically, to achieve a certain detection performance, the number of required sensors decreases as the correlation coefficient ρ increases.



(a)



(b)

Fig. 2: The Kullback-Leibler distance/Bhattacharyya distance/signal-to-noise ratio for different values of the correlation coefficient ρ. (a) Identical observations ($f = 0$). (b) Non-identical observations ($f = 0.33$).

7. CONCLUSIONS

We have considered the sensing design problem for Gaussian detection. We assume a certain set of candidate sampling locations (temporal and/or spatial) with cardinality M. We choose the best subset out of them through a sensing operator $\mathbf{w} \in \{0,1\}^M$, such that the error probabilities and the cardinality of $\mathbf{w}$ are jointly minimized. In essence, $\mathbf{w}$ should be as sparse as possible such that the resulting prescribed fidelity is reached.

We have considered the detection problem both in the classical setting and Bayesian setting. Since the expressions for the error probabilities are not favorable for numerical optimization, we optimize weaker measures such as the Kullback-Leibler distance and Bhattacharyya distance, both of which are coincidently the same up to a constant for the considered problem. Moreover, they are related to the signal-to-noise ratio, which is an optimal performance criterion for Gaussian detection problems. The sensing design problem has been transformed to a semidefinite program in its most general form. Finally, we conclude this paper with the following remarks:

- For conditionally independent observations, the sensing design problem can be solved optimally by simply ordering the local signal-to-noise ratios. For conditionally dependent observations, the sensing design problem can be solved sub-optimally in polynomial time.
- When the sensor observations are non-identical, the number of sensors required to achieve a certain detection performance decreases as the correlation coefficient increases.

8. REFERENCES

[1] S. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. of the Royal Stat. Society. Series B (Methodological)*, pp. 131–142, 1966.

[2] C.-T. Yu and P. K. Varshney, "Sampling design for gaussian detection problems," *IEEE Trans. Signal Process.*, vol. 45, no. 9, pp. 2328–2337, 1997.

[3] D. Bajovic, B. Sinopoli, and J. Xavier, "Sensor selection for event detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4938–4953, Oct 2011.

[4] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *The Journal of Machine Learning Research*, vol. 9, pp. 235–284, Feb. 2008.

[5] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[6] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.

[7] ——, "Sparsity-promoting adaptive sensor selection for non-linear filtering," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 5100–5104.

[8] S. Liu, E. Masazade, M. Fardad, and P. Varshney, "Sparsity-aware field estimation via ordinary kriging," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3948–3952.

[9] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*. Prentice hall New York, 2000, vol. 1.

[10] T. Grettenberg, "Signal selection in communication and radar systems," *IEEE Trans. Inf. Theory*, vol. 9, no. 4, pp. 265–275, 1963.

[11] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Commun. Technol.*, vol. 15, no. 1, pp. 52–60, February 1967.

[12] T. Kadota and L. A. Shepp, "On the best finite set of linear observables for discriminating two gaussian signals," *IEEE Trans. Inf. Theory*, vol. 13, no. 2, pp. 278–284, 1967.

[13] S. Kullback, *Information theory and statistics*. Courier Dover Publications, 2012.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[15] J. F. Sturm, "Using SeDuMi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[16] J.-F. Chamberland and V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb 2003.