Sensor Selection for Estimation, Filtering, and Detection

Sundeep Prabhakar Chepuri and Geert Leus
Invited paper

***Abstract*—Sensor selection is a crucial aspect in sensor network design. Due to the limitations on the hardware costs, availability of storage or physical space, and to minimize the processing and communication burden, the limited number of available sensors has to be smartly deployed. The node deployment should be such that a certain performance is ensured. Optimizing the sensors' spatial constellation or their temporal sampling patterns can be casted as a sensor selection problem. Sensor selection is essentially a combinatorial problem involving a performance evaluation over all possible choices, and it is intractable even for problems of modest scale. Nevertheless, using convex relaxation techniques, the sensor selection problem can be solved efficiently. In this paper, we present a brief overview and recent advances on the sensor selection problem from a statistical signal processing perspective. In particular, we focus on some of the important statistical inference problems like estimation, tracking, and detection.**

***Index Terms*—Sensor placement, sensor selection, sparsity, convex optimization, sensor networks, statistical inference.**

I. Introduction

Sensors are widely used in a variety of applications and services related to field monitoring, safety and security, logistics, and surveillance, to list a few. Every sensor is capable of sensing, processing, and communicating to other nodes or a central unit (often referred to as a fusion center). This enables spatially deployed sensor nodes to function as a network, which has to carry out one or more specified tasks. Each sensor often provides an excessively large dataset from which we seek to extract relevant information by optimally (pre-)processing the data. For example, such processing includes identifying the informative observation and discarding redundant or identical ones. In other words, this is simply a sensor selection problem which can be interpreted as the problem of choosing the best subset of sensors that guarantees some prescribed performance.

The number of sensors used to perform these specific tasks are often limited. This may be due to limitations on the hardware costs, availability of storage or physical space, or to reduce the burden of processing and communication. For location-related services (like source localization and field estimation), the spatial deployment of sensors is important, as certain spatial constellations can significantly deteriorate the performance. Sensor deployment can also be interpreted as a sensor selection problem in which the best subset of the available sensor locations are selected subject to a performance constraint. Sensor selection has been applied to a wide variety of problems: dynamical systems [1]–[5], network monitoring [6], field estimation [7], array optimization [8], source-informative sensor identification [9], anchor placement [10], and outlier detection [11].

The purpose of this partly tutorial paper is to describe the sensor selection problem from a statistical signal processing perspective, and to provide a brief overview with a specific emphasis on some recent advances based on results from [12]–[15].

II. Optimization problem

Sensor selection can be formulated as the problem of designing a selection vector whose entries are chosen from a prescribed finite alphabet set. More specifically, the problem is to design a selection vector

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

such that the number of selected sensors are minimized and a certain performance is achieved. Here, the variable $w_m = (0)1$ indicates that the sensor is (not) selected, and M is the number of sensors available. Mathematically, sensor selection is a constrained cardinality minimization problem where the constraints are based on some appropriate performance measure, $f(\mathbf{w})$. The performance measure can take a number of different forms depending on whether the inference problem is an estimation problem or a detection problem. In essence, the performance measure quantifies the system requirement, i.e., the estimation accuracy or detection probability.

We can write the sensor selection problem as the following generic constrained optimization problem:

$$\arg\min_{\mathbf{w}} \ \|\mathbf{w}\|_0 \tag{1a}$$

$$\text{s.to} \quad f(\mathbf{w}) \leq \lambda, \tag{1b}$$

$$\mathbf{w} \in \{0,1\}^M, \tag{1c}$$

where the threshold λ specifies the accuracy requirement and also induces the sparsity in $\mathbf{w}$. Naturally, the optimization problem in (1) can also be casted as

$$\arg\min_{\mathbf{w}} \ f(\mathbf{w}) \tag{2a}$$

$$\text{s.to} \quad \|\mathbf{w}\|_0 = K, \tag{2b}$$

$$\mathbf{w} \in \{0,1\}^M, \tag{2c}$$

The authors are with the faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, The Netherlands. Email: {g.j.t.leus; s.p.chepuri}@tudelft.nl.

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

where K is the desired number of sensors. The problems in (1) and (2) are equivalent in the sense that with a certain threshold λ^*, K sensors can be selected. The problem of the form (1) might be appropriate for certain designs where the number K is not known in which case λ has to be designed. However, if K is *a priori* known, then the problem of the form (2) is the obvious choice. In this paper, we will focus on sensor selection solvers of the form (1), while the obtained results also hold for solvers of the form (2) with some straightforward adaptation.

The optimization problem in (1) is a non-convex optimization problem with a non-convex cost function given as (1a). The non-convex Boolean constraint (1c) incurs a combinatorial search over all the 2^M (or $\binom{K}{M}$ for the optimization problem of the form (2) where K is known a priori) possible combinations. To simplify this problem, standard convex relaxations are used. The ℓ_0-(quasi) norm in (1a) is relaxed to the ℓ_1-norm, and the Boolean constraint in (1c) is relaxed to the box constraint $[0,1]^M$. As a result, the following relaxed sensor selection problem is obtained

$$\arg\min_{\mathbf{w}} \ \|\mathbf{w}\|_1 \tag{3a}$$

$$\text{s.to} \quad f(\mathbf{w}) \leq \lambda, \tag{3b}$$

$$\mathbf{w} \in [0,1]^M. \tag{3c}$$

An approximate Boolean solution can then be recovered from the solution of the above convex optimization problem either by simple thresholding or randomized rounding [13]. Alternatively, the ℓ_0-(quasi) norm can be approximated using the sum-of-logs, $\sum_{m=1}^{M} \ln(w_m + \delta)$ with $\delta > 0$, which results in an iteratively weighted ℓ_1-norm optimization problem [13]. Typically, log-based heuristics result in a sparser solution, and thus better approximate the ℓ_0-(quasi) norm.

We shall discuss different performance measures $f(\mathbf{w})$ for estimation, filtering, and detection problems in Sections III and IV, respectively.

III. Estimation and filtering

In this section, we discuss sensor selection for estimation and filtering. We will focus on non-linear measurement models (linear models being a special case) which are frequently encountered in problems related to source/target localization and field estimation.

A. Sensor selection for estimation

Consider the following measurement model

$$y_m = h_m(\boldsymbol{\theta}, n_m), m = 1, 2, \ldots, M, \tag{4}$$

where $\boldsymbol{\theta} \in \mathbb{R}^N$ is the unknown parameter vector, $h_m(\cdot)$ is (in general) a non-linear function, and n_m is the measurement noise. The sensors can be either active or passive sensors.

From the viewpoint of estimation, the sensor selection problem can now be described as follows. Out of the M potential sensors choose the best subset of sensors ($\geq N$) that yields the lowest error covariance.

The error covariance matrix is denoted by

$$\mathbf{E} =: \mathbb{E}\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\} \in \mathbb{R}^{M \times M},$$

where $\hat{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$. The sensor selection is determined by evaluating scalar functions of the error covariance $\mathbf{E}$. In literature, the most prominent choices of the functions are related to:

1) *A-optimality*: average-variance criterion, $\mathrm{tr}\{\mathbf{E}\}$.
2) *E-optimality*: largest-eigenvalue criterion, $\lambda_{\max}\{\mathbf{E}\}$.
3) *D-optimality*: determinant criterion, $\ln\det\{\mathbf{E}\}$.

All the above criteria are related to the error concentration ellipsoid defined by $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{E}^{-1} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq 1$. The average-variance criterion is the sum of the semi-axes of this ellipsoid, or the mean-squared-error (MSE). Moreover, a weighted trace criterion can be used when one or more elements of the unknown parameter are relatively more important than the others. The largest-eigenvalue criterion is the semi-major axis of the error ellipsoid, or the worst-case error. The determinant criterion finally is related to the volume or the geometric mean of the semi-axes of the error concentration ellipsoid. All the performance measures are equally credible, although neither of them completely characterizes the error covariance. There is no general answer to the question of how does one performance metric compare with the other. See [16], [17] for a more detailed discussion on this issue.

In many cases, the matrix $\mathbf{E}$ cannot be computed in closed-form or its expression is too complicated for numerical optimization. For example, when the observations are related to a non-linear model as in (4) or to linear additive non-Gaussian noise models. For such cases, we can use the Cramér-Rao bound (CRB), a weaker performance criterion but easier to evaluate and optimize [13]. The error covariance of any unbiased estimate $\hat{\boldsymbol{\theta}} \in \mathbb{R}^N$ of the unknown parameter is related to the CRB through the inequality [18]

$$\mathbb{E}\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\} \geq \mathbf{C}(\boldsymbol{\theta}) = \mathbf{F}^{-1}(\boldsymbol{\theta}),$$

where $\mathbf{C}(\boldsymbol{\theta})$ is the CRB matrix. The Fisher information matrix (FIM), $\mathbf{F}(\boldsymbol{\theta}) \in \mathbb{R}^{N \times N}$ is given by

$$\begin{aligned} \mathbf{F}(\boldsymbol{\theta}) &= \sum_{m=1}^{M} \mathbb{E}\left\{ \left(\frac{\partial \ln p(y_m; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln p(y_m; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T \right\} \\ &= \sum_{m=1}^{M} \mathbf{F}_m(\boldsymbol{\theta}), \end{aligned} \tag{5}$$

with $\ln p(y_m; \boldsymbol{\theta})$ the log-likelihood of the measurements which is the probability density function (pdf) of y_m parameterized by the unknown vector $\boldsymbol{\theta}$. Here, we assume that the pdf $p(y_m; \boldsymbol{\theta})$ satisfies the regularity conditions which is a prerequisite for the CRB to exist. The expression in (5) is valid as long as the observations across the sensors are independent. The expression in (5) indicates that every measurement provides some additional information, and thus the information provided by all the measurements reduces the uncertainty.

We can now write the FIM in (5) using the variable $\mathbf{w}$ as $\mathbf{F}(\mathbf{w}, \boldsymbol{\theta}) = \sum_{m=1}^{M} w_m \mathbf{F}_m(\boldsymbol{\theta})$. In other words, we use the selection variable to choose the most informative sensors. We evaluate the functions of the FIM (or the CRB) matrix $\mathbf{F}(\mathbf{w}, \boldsymbol{\theta})$

based on the A-, E-, and D-optimality measures that were discussed earlier. Specifically, we have for

1) *A-optimality*: $f(\mathbf{w}) := \mathrm{tr}\{(\sum_{m=1}^{M} w_m \mathbf{F}_m(\boldsymbol{\theta}))^{-1}\}$.
2) *E-optimality*: $f(\mathbf{w}) := \lambda_{\max}\{(\sum_{m=1}^{M} w_m \mathbf{F}_m(\boldsymbol{\theta}))^{-1}\}$.
3) *D-optimality*: $f(\mathbf{w}) := \ln\det\{(\sum_{m=1}^{M} w_m \mathbf{F}_m(\boldsymbol{\theta}))^{-1}\}$.

Note that the above performance measures depend on the unknown parameter. In practice, the unknown parameter $\boldsymbol{\theta}$ takes values within a certain known domain denoted by $\mathcal{U}$. Hence, for non-linear models the constraint (3b) should be satisfied for every point within the domain $\mathcal{U}$.

A specialization of the non-linear model in (4) is the linear Gaussian measurement model

$$y_m = \mathbf{h}_m^T \boldsymbol{\theta} + n_m, m = 1, 2, \ldots, M.$$

For such models, the FIM (performance measure) is independent of the unknown parameter vector and is simply given by

$$\mathbf{F} = \sum_{m=1}^{M} \mathbf{h}_m \mathbf{h}_m^T.$$

In that case, there is no need to constrain the performance for every point within $\mathcal{U}$. This is the key difference between sensor selection for linear and non-linear models.

Each one of the performance measures discussed above is a convex function of $\mathbf{w}$. As a result, using one of the above measures as constraints in (3b), the relaxed sensor selection problem will be a convex optimization problem. For example, let us consider the MSE (A-optimality) criterion. Writing the relaxed sensor selection (3) in the epigraph form, we obtain [12]

$$\begin{aligned} &\underset{\mathbf{w}\in\mathbb{R}^M,\, \mathbf{x}\in\mathbb{R}^N}{\arg\min} \quad \|\mathbf{w}\|_1 \\ &\text{s.to } x_n \geq \boldsymbol{\delta}_n^T \left(\left(\sum_{m=1}^{M} w_m \mathbf{F}_m(\boldsymbol{\theta}) \right)^{-1} \right) \boldsymbol{\delta}_n, \, \forall \boldsymbol{\theta} \in \mathcal{U} \\ &\quad x_n \geq 0, n = 1, 2, \ldots, N, \\ &\quad 0 \leq w_m \leq 1, \quad m = 1, 2, \ldots, M, \\ &\quad \mathbf{1}_N^T \mathbf{x} \leq \lambda, \end{aligned}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T \in \mathbb{R}^N$ is a variable, and $\boldsymbol{\delta}_n$ is the nth unit vector in $\mathbb{R}^N$. This can be transformed into a semi-definite programming (SDP) problem, and hence, can be solved efficiently. The SDP problem is given by

$$\begin{aligned} &\underset{\mathbf{w}\in\mathbb{R}^M,\, \mathbf{x}\in\mathbb{R}^N}{\arg\min} \quad \|\mathbf{w}\|_1 \\ &\text{s.to } \begin{bmatrix} \sum_{m=1}^{M} w_m \mathbf{F}_m(\boldsymbol{\theta}) & \boldsymbol{\delta}_n \\ \boldsymbol{\delta}_n^T & x_n \end{bmatrix} \succeq \mathbf{0}_{N+1}, \, \forall \boldsymbol{\theta} \in \mathcal{U}, \\ &\quad x_n \geq 0, n = 1, 2, \ldots, N, \\ &\quad 0 \leq w_m \leq 1, \quad m = 1, 2, \ldots, M, \\ &\quad \mathbf{1}_N^T \mathbf{x} \leq \lambda. \end{aligned} \tag{6}$$

The computational complexity of the relaxed sensor selection problem based on SDP is of the order of M^3. For large-scale problems, i.e., when the number of sensors is very large (≥ 1000), low-complexity algorithms (with cheaper per iteration costs) are often needed. One such computationally less intensive implementation of the relaxed sensor selection problem is based on the projected subgradient algorithm [13]. Alternative approaches to convex optimization based sensor selection exploit the submodularity of the objective function using proxies for the MSE, like the mutual information [19] or the frame potential [7]. The submodularity of the objective function helps in developing low-complexity greedy algorithms, which sequentially adds sensors that maximize the increase in the cost.

In addition to the performance constraints, logical constraints [12] like select sensor i only when sensor j is chosen (via the inequality $w_i \leq w_j$), or select at least one of the sensors i or j (via the inequality $w_i + w_j \geq 1$) can also be imposed.

B. Sensor selection for filtering

In this section, we consider an extension to the static sensor selection problem discussed in the previous section. We develop a framework for dynamic (or adaptive) sensor selection pertinent to tracking time-varying parameters (e.g., target tracking). We assume that the unknown parameter $\boldsymbol{\theta}_t$ follows a dynamical model:

$$\boldsymbol{\theta}_{t+1} = \mathbf{A}_t \boldsymbol{\theta}_t + \mathbf{v}_t, \tag{7}$$

where $\mathbf{A}_t \in \mathbb{R}^{N\times N}$ is an invertible state transition matrix at time t, and $\mathbf{v}_t \in \mathbb{R}^{N\times 1}$ is the process noise at time t that accounts for any unmodeled dynamics. Here, we model $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$, where $\mathbf{Q}_t \in \mathbb{R}^{N\times N}$ represents the covariance matrix. The measurement equation takes the same from as in (4) except for an additional subscript t with which we denote that the state parameter is time-varying. The measurement equations are given as

$$y_{t,m} = h_{t,m}(\boldsymbol{\theta}_t, n_{t,m}), m = 1, 2, \ldots, M, \tag{8}$$

where $y_{t,m}$ is the measurement at time t, $h_{t,m}(\cdot)$ is (in general) a non-linear function, and $n_{t,m}$ is the measurement noise at time t.

We consider the problem of selecting at each time step the best subset of sensors out of M available state measurements that guarantees a prescribed *a posteriori* estimation error. As earlier, we use a selection vector

$$\mathbf{w}_t := [w_{t,1}, w_{t,2}, \ldots, w_{t,M}]^T \in \{0, 1\}^M.$$

The sensor selection is determined by evaluating the posterior CRB (PCRB) $\mathbf{F}_t^{-1}(\mathbf{w}_t, \boldsymbol{\theta}_t)$. When the observations are independent, the PCRB can be recursively expressed as [14]

$$\begin{aligned} \mathbf{F}_t(\mathbf{w}_t, \boldsymbol{\theta}_t) &= (\mathbf{Q}_t + \mathbf{A}_t \mathbf{F}_{t-1}^{-1}(\boldsymbol{\theta}_{t-1}) \mathbf{A}_t^T)^{-1} \\ &\quad + \sum_{m=1}^{M} w_{t,m} \mathbf{F}_{t,m}(\boldsymbol{\theta}_t), \\ &= \mathbf{J}_{t-1}(\boldsymbol{\theta}_{t-1}) + \sum_{m=1}^{M} w_{t,m} \mathbf{F}_{t,m}(\boldsymbol{\theta}_t), \end{aligned} \tag{9}$$

where

$$\mathbf{F}_{t,m}(\boldsymbol{\theta}_t) := \left(\frac{\partial \ln p(y_{t,m};\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}\right)\left(\frac{\partial \ln p(y_{t,m};\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}\right)^T.$$

Due to the non-linearity of the measurement model, the posterior FIM (inverse PCRB) in (9) depends on the unknown state at time t as well as the state at time $t-1$. For additive Gaussian noise models, using the past estimate $\hat{\boldsymbol{\theta}}_{t-1|t-1}$ and the predicted estimate $\hat{\boldsymbol{\theta}}_{t|t-1}$ for $\boldsymbol{\theta}_{t-1}$ and $\boldsymbol{\theta}_t$, respectively, results in the posterior covariance matrix. The past estimate and prediction can be computed from an extended Kalman filter (EKF) for instance [18]. One of the functions related to A-, E-, or D-optimality of the error covariance matrix computed using the past state estimate can then be used as a performance measure to perform selection [20]. However, since the past state estimate (not the true state) is used to compute the covariance matrix, depending on the non-linearity of the model and the noise variance, the selection will be suboptimal. To alleviate this problem, we can constrain the performance for every point within a domain around the past estimate and the predicted estimate [14]. For example, adaptive (relaxed) sensor selection based on the average-variance criterion, can be formulated as

$$\underset{\mathbf{w}_t \in [0,1]^M}{\arg\min} \quad \|\mathbf{w}_t\|_1 \tag{10a}$$

$$\text{s.to } \mathrm{tr}\{(\mathbf{J}_{t-1}(\boldsymbol{\theta}_{t-1}) + \sum_{m=1}^{M} w_{t,m}\mathbf{F}_{t,m}(\boldsymbol{\theta}_t))^{-1}\} \leq \lambda, \quad \forall \boldsymbol{\theta}_{t-1} \in \mathcal{U}_{t-1}, \forall \boldsymbol{\theta}_t \in \mathcal{U}_t, \tag{10b}$$

where the domains $\mathcal{U}_{t-1}$ and $\mathcal{U}_t$ are specified based on the past a posteriori covariance matrix and the predicted covariance matrix, respectively. The above problem can also be transformed to an SDP, as in (6).

When the state-space equations follow a linear model $\mathbf{y}_t := \mathbf{H}_t\boldsymbol{\theta}_t + \mathbf{n}_t$, i.e., for a discrete-time linear dynamical system, the PCRB is independent of the state. Hence, then there is no need to constrain the performance within the domains $\mathcal{U}_{t-1}$ and $\mathcal{U}_t$. An important and obvious specialization is the case when the dynamics matrix, regression matrix, and the related noise covariance matrices are not time-varying (or slowly time-varying). For such cases, the selection has to be performed only when either the dynamics or regression matrix changes. Furthermore, the selection can be simultaneously performed over both space and time over a horizon $t = 1, 2, \ldots, T$ (smoothing, moving horizon estimators, for instance). To choose the best subset of space-time sensors corresponds to simultaneously selecting the best subset of spatial sensors and time instances to obtain the measurements (only spatial sensor selection for each time step is described in this paper).

IV. Detection

Sensor selection for hypothesis testing problems has received less attention compared to sensor selection for estimation and filtering. This section is devoted to sensor selection for event detection, which is fundamental for instance to radar, sonar, communications, and spectrum sensing applications.

The observation at each sensor is related to the state of nature $\mathcal{H}$. Here, we consider a binary hypothesis testing problem, where $\mathcal{H}$ takes one of two possible values, i.e., either $\mathcal{H}_0$ or $\mathcal{H}_1$.

In the classical setting, the optimal detector is the Neyman-Pearson detector which maximizes the probability of detection, $\mathrm{P_d}$ for a fixed probability of false alarm, $\mathrm{P_{fa}}$ [21]. The sensor selection problem for binary hypothesis testing can be interpreted as the problem to choose the best subset of sensors that guarantees a certain prescribed detection probability. The optimal sensor subset is the subset of sensors that maximizes $\mathrm{P_d}$. However, in many cases, optimizing of the error probabilities is very difficult. This may be because these error probabilities do not admit a known closed-form or the expressions are too complicated for numerical optimization. Therefore, a weaker performance criterion which is easier to evaluate and optimize is often used. The relative entropy or the Kullback-Leibler (KL) distance is a frequently used performance criterion for design problems related to hypothesis testing [15], [22]. The sensor selection problem in [15] is formulated as the design of a selection matrix which is a non-convex optimization problem (even after appropriate relaxation). However, similar to the sensor selection framework developed for estimation and filtering in Section III, we can also formulate sensor selection for hypothesis testing problems. We next briefly describe the convex optimization formulation.

Consider the following observation model

$$\mathcal{H}_0: \quad y_m = n_m,\ m = 1, 2, \ldots, M, \tag{11a}$$

$$\mathcal{H}_1: \quad y_m = h_m(\boldsymbol{\theta}, n_m),\ m = 1, 2, \ldots, M, \tag{11b}$$

where n_m is the measurement noise as defined earlier. The observations are collected in $\mathbf{y} = [y_1, y_2, \ldots, y_M]^T \in \mathbb{R}^M$. Let the probability density function (pdf) of $\mathbf{y}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ be denoted by $p(\mathbf{y}|\,\mathcal{H}_0)$ and $p(\mathbf{y}|\,\mathcal{H}_1)$, respectively. Defining the log-likelihood ratio as

$$l(\mathbf{y}) := \ln \frac{p(\mathbf{y}|\,\mathcal{H}_1)}{p(\mathbf{y}|\,\mathcal{H}_0)},$$

the KL distance is given by

$$\begin{aligned} D(\mathcal{H}_1\|\mathcal{H}_0) &= \mathbb{E}_{|\mathcal{H}_1}\{l(\mathbf{y})\} \\ &= \int l(\mathbf{y}) p(\mathbf{y}|\mathcal{H}_1) d\mathbf{y}. \end{aligned} \tag{12}$$

The notation $\mathbb{E}_{|\mathcal{H}_1}\{l(\mathbf{y})\}$ indicates that $l(\mathbf{y})$ is averaged under the pdf $p(\mathbf{y}|\mathcal{H}_1)$. Under the assumption that the observations conditioned on $\mathcal{H}$ are independent across the sensors, the KL distance can be expressed as [23]

$$D(\mathcal{H}_1\|\mathcal{H}_0) = \sum_{m=1}^{M} D_m(\mathcal{H}_1\|\mathcal{H}_0), \tag{13}$$

where

$$D_m(\mathcal{H}_1\|\mathcal{H}_0) = \int l_m(y) p_m(y|\mathcal{H}_1) dy.$$

Here, we define

$$l_m(y) = \ln \frac{p_m(y|\mathcal{H}_1)}{p_m(y|\mathcal{H}_0)}$$

as the local log-likelihood ratio related to the mth sensor with the conditional pdfs of y, $p_m(y|\mathcal{H}_i)$ for $i = 0, 1$.

As the asymptotic (in the number of sensors) rate of convergence of the probability of miss detection ($\mathrm{P_m} := 1 - \mathrm{P_d}$) to zero is parameterized by the KL distance, we want to maximize its value. Hence, using the selection variable $\mathbf{w}$ defined earlier, for detection problems, the cost to be minimized is

$$f(\mathbf{w}) := -\sum_{m=1}^{M} w_m D_m(\mathcal{H}_1 \| \mathcal{H}_0).$$

We can formulate the relaxed sensor selection problem for hypothesis testing as

$$\arg\min_{\mathbf{w}} \ \|\mathbf{w}\|_1 \tag{14a}$$

$$\text{s.to} \quad \sum_{m=1}^{M} w_m D_m(\mathcal{H}_1 \| \mathcal{H}_0) \geq \lambda, \tag{14b}$$

$$0 \leq w_m \leq 1, \quad m = 1, 2, \ldots, M, \tag{14c}$$

where λ specifies the required detection probability. A similar extension can be made to detection problems under the Bayesian setting in which prior probabilities are assigned to the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. The sensor selection for detection in the Bayesian setting then consists of choosing the best subset of sensors that minimizes the probability of error.

V. Concluding remarks

Sensor networks are predominantly used for field estimation, localization, and environment monitoring, to list a few. All these tasks typically solve a statistical inference problem like detection, estimation, or filtering. One of the key aspects in designing an efficient sensor network involves optimizing the spatio-temporal sampling patterns. This includes node deployment and sampling time selection. These problems are casted as a sensor selection problem in which the best subset of sensors that guarantees a prescribed estimation accuracy or detection probability is selected. In this paper, we have presented a brief overview on sensor selection based on convex optimization techniques for problems that are frequently encountered in statistical signal processing.

References

[1] Yinfei Fu, Qing Ling, and Zhi Tian, "Distributed sensor allocation for multi-target tracking in wireless sensor networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 4, pp. 3538–3553, 2012.

[2] M.L. Hernandez, T. Kirubarajan, and Y. Bar-Shalom, "Multisensor resource deployment using Posterior Cramer-Rao bounds," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 2, pp. 399–416, 2004.

[3] F. Jiang, J. Chen, and A. L. Swindlehurst, "Linearly reconfigurable Kalman filtering for a vector process," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[4] Long Zuo, Ruixin Niu, and P.K. Varshney, "Posterior CRLB based sensor selection for target tracking in sensor networks," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2007, vol. 2, pp. II–1041–II–1044.

[5] Pengcheng Zhan, D.W. Casbeer, and A.L. Swindlehurst, "Adaptive mobile sensor positioning for multi-static target tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 120–132, 2010.

[6] V. Kekatos, G.B. Giannakis, and B. Wollenberg, "Optimal placement of phasor measurement units via convex relaxation," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1521–1530, 2012.

[7] J. Ranieri, A. Chebira, and M. Vetterli, "Near-optimal sensor placement for linear inverse problems," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1135–1146, March 2014.

[8] V. Roy, S.P. Chepuri, and G. Leus, "Sparsity-enforcing sensor selection for DOA estimation," in *Proc. of the IEEE Workshop on Comp. Adv. in Multi-Sensor Adaptive Proc. (CAMSAP)*, Dec. 2013.

[9] I.D. Schizas, "Distributed informative-sensor identification via sparsity-aware matrix decomposition," *IEEE Trans. Signal Process.*, vol. PP, no. 99, pp. 1–1, 2013.

[10] S. P. Chepuri, G. Leus, and A.-J. van der Veen, "Sparsity-exploiting anchor placement for localization in sensor networks," in *Proc. of the 21st European Signal Processing Conference (EUSIPCO)*, Sept. 2013.

[11] V. Kekatos and G.B. Giannakis, "From sparse signals to sparse residuals for robust sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3355–3368, 2011.

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

[13] S.P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process. (Submitted)*, Sep. 2013.

[14] S.P. Chepuri and G. Leus, "Sparsity-promoting adaptive sensor selection for non-linear filtering," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 5100–5104.

[15] 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.

[16] F. Pukelsheim, *Optimal design of experiments*, vol. 50, SIAM, 1993.

[17] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, New York, NY, USA, 2004.

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

[19] 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, 2008.

[20] E. Masazade, M. Fardad, and P.K. Varshney, "Sparsity-promoting extended kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 845–848, 2012.

[21] S. M. Kay, *Fundamentals of statistical signal processing: Detection theory*, Englewood Cliffs, NJ: Prentice-Hall, 1998.

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

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