SPARSE SENSING FOR COMPOSITE MATCHED SUBSPACE DETECTION

Mario Coutino, Sundeep Prabhakar Chepuri, and Geert Leus

Delft University of Technology, Delft, The Netherlands.

ABSTRACT

In this paper, we propose sensor selection strategies, based on convex and greedy approaches, for designing sparse samplers for composite detection. Particularly, we focus our attention on sparse samplers for matched subspace detectors. Differently from previous works, that mostly rely on random matrices to perform compression of the subspaces, we show how deterministic samplers can be designed under a Neyman-Pearson-like setting when the generalized likelihood ratio test is used. For a less stringent case than the worst case design, we introduce a submodular cost that obtains comparable results with its convex counterpart, while having a linear time heuristic for its near optimal maximization.

Index Terms— composite hypothesis testing, convex optimization, matched subspace detector, sensor selection, submodular optimization

1. INTRODUCTION

Composite hypothesis testing [1], [2] is one of the problems in signal processing for which until this date no appropriate general solution has been found. Similar to simple hypothesis testing, we desire to select the distribution that best describes the observed data from a set of possible candidates. However, there is a difference: the parameters that describe these distributions are unknown or have uncertainties, e.g., there is prior knowledge that the parameter is greater than zero, or that it lies within a certain range.

A particular interesting instance of such a hypothesis test is the *matched subspace detector* [4]. In this case, the only available information about the signal of interest is the subspace in which it lives. This detector can be seen as the generalization of the celebrated *matched filter* [3] which can be expressed as a rank-1 subspace detector. Matched subspace detectors arise naturally in a myriad of applications such as radar [5], communications [6], and classification [7], to name a few.

In recent years, major attention has been paid to performing detection with compressed or subsampled observations [8]-[12]. However, most of the research is mainly concerned with simple binary hypothesis testing (where simple means non-composite). In these works, it has been shown that it is possible to perfom sensor selection satisfactory by employing numerically amenable performance metrics such as the Kullback-Leibler divergence and the Bhattacharyya distance instead of the probability of error, and probability of detection [12],[14]. In addition, even though robust sensor selection for binary hypothesis testing under parameters with uncertainties has been considered [9], sensor selection for composite detection has still no appropriate solution.

On the other hand, most of the works that focus their efforts on composite hypothesis testing, particularly for subspace detection, rely on random matrices to provide performance guarantees for reduced-size detectors [10]. Hence, they only provide statistical guarantees for an ensemble of random matrices. In addition, due to the probabilistic nature of the guarantees, the samplers need to change constantly, which in most cases is not practical in real systems.

Differently to these works, we are interested in deterministic sparse samplers that produce deterministic performance guarantees by levering the data model. That is why, in this work, we lever the probability distribution of the test statistic for the optimal subspace detector, based on the generalized likelihood ratio test (GLRT), to perform offline sensor selection (a.k.a. sparse sampler design).

2. PROBLEM STATEMENT

Consider the problem of detecting a signal *subspace* in noise and with interference. Particularly, we focus on the case where the interference can be modeled with a low rank subspace. That is, given the acquired signal

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v} + \boldsymbol{n} \in \mathbb{R}^N, \tag{1}$$

the received signal $\boldsymbol{x}$ and interference $\boldsymbol{v}$ can be decomposed using the subspaces in which they exist as

$$\boldsymbol{x} = \boldsymbol{H\theta}, \text{ with } \boldsymbol{H} \in \mathbb{R}^{N\times P}, \tag{2}$$

$$\boldsymbol{v} = \boldsymbol{S\phi}, \text{ with } \boldsymbol{S} \in \mathbb{R}^{N\times Q}, \tag{3}$$

for some $\boldsymbol{\theta} \in \mathbb{R}^P$ and $\boldsymbol{\phi} \in \mathbb{R}^Q$, respectively. Here, $\boldsymbol{n}$ is the zero-mean Gaussian noise vector, i.e., $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. This formulation is still valid for instances in which the interference might not be present, i.e., $\boldsymbol{S} = \boldsymbol{0}$ or $\boldsymbol{\phi} = \boldsymbol{0}$. The matrices $\boldsymbol{H}$ and $\boldsymbol{S}$ are assumed to be full column-rank matrices, and it is also assumed that the matrix $[\boldsymbol{H}\ \boldsymbol{S}]$ has full column-rank with $P + Q \leq N$.

To detect the presence of a signal $\boldsymbol{x} \in \text{span}(\boldsymbol{H})$, we are required to test the following composite hypotheses

$$\begin{aligned} &\mathcal{H}_0 : \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{S\phi}, \sigma^2\boldsymbol{I}) \\ &\mathcal{H}_1 : \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{S\phi} + \boldsymbol{H\theta}, \sigma^2\boldsymbol{I}) \end{aligned}, \tag{4}$$

where for the hypothesis $\mathcal{H}_1$ it is assumed that $\|\boldsymbol{\theta}\|_2^2 > 0$ and for both $\mathcal{H}_0$ and $\mathcal{H}_1$, the noise power, σ^2, is considered unknown. The difficulty in (4) lies in the fact that both $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are unknown vectors, as only information about the subspaces $\boldsymbol{H}$ and $\boldsymbol{S}$ is known.

Furthermore, consider that only a subset of data can be acquired (observed) by means of a linear sensing operation, i.e.,

$$\boldsymbol{y}_{\mathcal{A}} = \boldsymbol{\Phi}_{\mathcal{A}}\boldsymbol{y} \in \mathbb{R}^K, \ \ K \leq N, \tag{5}$$

This research is supported in part by the ASPIRE project (project 14926 within the STW OTP programme), financed by the Netherlands Organization for Scientific Research (NWO), and the KAUST-MIT-TUD consortium under grant OSR-2015-Sensors-2700. Mario Coutino is partially supported by CONACYT.

where $\mathcal{A} \subseteq \mathcal{V} = \{1, 2, \ldots, N\}$ is the considered subset of entries for $\boldsymbol{y}$, and $\boldsymbol{\Phi}_{\mathcal{A}} \in \{0, 1\}^{K \times N}$ denotes the related selection matrix whose rows are the rows of an identity matrix of size $N \times N$ indexed by the set $\mathcal{A}$.

Such a constrained acquisition scheme typically appears in distributed sensor networks where due to economical, space or physical constraints, only a limited set of sensors can be deployed/used for inference [14]. Therefore, in this work, we are interested in the following problem:

How can we optimally design the set $\mathcal{A}$ (of given cardinality) in order to obtain the best detection performance for (4)*?*

For the problem of detecting a signal subspace in noise and interference it is well-known that the GLRT is uniformly most powerful (UMP) invariant test [4]. Therefore, in this work, we propose to lever the parameters of the probability distributions for the GLRT of the hypotheses under test in (4) to design deterministic sparse samplers for optimal detection in a Neyman-Pearson-like setting. In the following, the matched subspace detector, which is the GLRT for (4), and its probability distributions are introduced.

3. MATCHED SUBSPACE DETECTOR

The GLRT for the composite hypothesis test in (4) is given by [4]

$$L(\boldsymbol{y}) \sim \frac{\boldsymbol{y}^T \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{E}_{\boldsymbol{HS}} \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{P}_{\boldsymbol{S}}^{\perp} (\boldsymbol{I} - \boldsymbol{E}_{\boldsymbol{HS}}) \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{y}}. \tag{6}$$

Here, the following orthogonal and *oblique* projections have been defined

$$\boldsymbol{P}_{\boldsymbol{S}}^{\perp} = \boldsymbol{I} - \boldsymbol{S}(\boldsymbol{S}^T \boldsymbol{S})^{-1} \boldsymbol{S}^T, \tag{7}$$

$$\boldsymbol{E}_{\boldsymbol{HS}} = \boldsymbol{H}(\boldsymbol{H}^T \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{P}_{\boldsymbol{S}}^{\perp}. \tag{8}$$

It can be shown that the normalized generalized likelihood ratio (GLR) [cf. (6)] is distributed as

$$\frac{\bar{Q} - \bar{P}}{\bar{P}} L(\boldsymbol{y}) : \begin{cases} F_{\bar{P}, \bar{Q} - \bar{P}}(0) & \text{under} \quad \mathcal{H}_0 \\ F_{\bar{P}, \bar{Q} - \bar{P}}(\lambda^2(\boldsymbol{\theta})) & \text{under} \quad \mathcal{H}_1 \end{cases}, \tag{9}$$

where $\bar{Q} \coloneqq \dim\{\text{span}(\boldsymbol{S}^{\perp})\} = N - Q$ and $\bar{P} \coloneqq \dim\{\text{span}(\boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{H})\}$ are the dimensions of the subspaces spanned by $\boldsymbol{S}$ and $\boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{H}$, respectively. Here, the noncentrality parameter $\lambda^2(\boldsymbol{\theta})$ given by [4]

$$\lambda^2(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \boldsymbol{\theta}^T \boldsymbol{H}^T \boldsymbol{P}_{\boldsymbol{S}}^{\perp} \boldsymbol{H} \boldsymbol{\theta}, \tag{10}$$

completely define the F-distribution given by F. By fixing a threshold η for the GLRT, we obtain the uniform most powerful invariant for testing the hypothesis in (4).

For (6), the false alarm and detection probabilities, with respect to a detection threshold η, are given by

$$P_{\text{fa}} = 1 - P[F_{\bar{P}, \bar{Q} - \bar{P}}(0) \leq \eta]; \tag{11}$$

$$P_{\text{d}} = 1 - P[F_{\bar{P}, \bar{Q} - \bar{P}}(\lambda^2(\boldsymbol{\theta})) \leq \eta]. \tag{12}$$

Based on them, we can design a sparse sampler which maximizes P_{d} for a fixed P_{fa}, i.e., η is fixed. This is achieved by noticing that P_{d} is a monotone function of $\lambda^2(\boldsymbol{\theta})$. Therefore, by maximizing the noncentrality parameter (10), the power of the test can be maximized. In other words, by maximizing $\lambda^2(\boldsymbol{\theta})$ we are effectively maximizing the output power of a signal $\boldsymbol{x} \in \text{span}(\boldsymbol{H})$ after being processed by an *interference filter*, $\boldsymbol{P}_{\boldsymbol{S}}^{\perp}$. Intuitively, under this approach, the optimal sparse sampler aims to maximize the orthogonality between the subspaces defined by $\boldsymbol{S}$ and $\boldsymbol{H}$.

Note that, although we focus on a case where the noise variance is unknown, if the noise level is known, the distribution of the GLRT for the alternative hypothesis $\mathcal{H}_1$ has a noncentralility parameter given by a scaled version of (10) [4]. Hence, the proposed approach extends also to particular instances of the problem in (4) with known σ^2.

4. SPARSE SAMPLER DESIGN

Consider the model for the measurements $\boldsymbol{y}_{\mathcal{A}}$ when the hypothesis $\mathcal{H}_1$ is true and a subset $\mathcal{A} \subseteq \mathcal{V}$ has been chosen

$$\begin{aligned} \boldsymbol{y}_{\mathcal{A}} &= \boldsymbol{\Phi}_{\mathcal{A}} [\boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{S}\boldsymbol{\phi} + \boldsymbol{n}] \\ &= \boldsymbol{H}_{\mathcal{A}} \boldsymbol{\theta} + \boldsymbol{S}_{\mathcal{A}} \boldsymbol{\phi} + \boldsymbol{n}_{\mathcal{A}} \in \mathbb{R}^K. \end{aligned} \tag{13}$$

An optimal sparse sampler aims to obtain the largest noncentrality parameter $\lambda^2(\boldsymbol{\theta})$ by an appropriate selection of the rows of $\boldsymbol{H}$ and $\boldsymbol{S}$. That is, it is desired that the subspaces spanned by $\boldsymbol{H}_{\mathcal{A}}$ and $\boldsymbol{S}_{\mathcal{A}}$ become as orthogonal as possible, i.e., $\boldsymbol{H}_{\mathcal{A}}^T \boldsymbol{S}_{\mathcal{A}} \approx \boldsymbol{0}$. However, the maximization of (10) depends on the unknown parameter $\boldsymbol{\theta}$. Therefore, its maximization in the general case is not possible when we are dealing with composite hypothesis testing problems. Despite this issue, in the following, we propose different metrics that can be considered to design sparse samplers that achieve the best performance possible for a given situation.

4.1. Average Design

The most straightforward design can be obtained by considering the *average* value that λ^2 can obtain, i.e.,

$$\underset{\mathcal{A} \subseteq \mathcal{V}, |\mathcal{A}| = K}{\text{maximize}} \int_{\boldsymbol{\theta} \in \boldsymbol{\Omega}} \boldsymbol{\theta}^T \boldsymbol{H}_{\mathcal{A}}^T \boldsymbol{P}_{\boldsymbol{S}_{\mathcal{A}}}^{\perp} \boldsymbol{H}_{\mathcal{A}} \boldsymbol{\theta} \, d\boldsymbol{\theta}. \tag{14}$$

This design is equivalent to the one obtained through a Bayesian approach where a non-informative prior, e.g., a uniform distribution, is given to all the possible vectors $\boldsymbol{\theta} \in \boldsymbol{\Omega}$. As a result, solving (14) requires to perform integration over the domain $\boldsymbol{\Omega}$ in which $\boldsymbol{\theta}$ is defined. In practice, this can become a cumbersome task as it involves the discretization or sampling of $\boldsymbol{\theta}$. Therefore, in this work we do not put any further attention to this case and opt for two other designs.

4.2. Worst Case (Max-Min) Design

Consider that $\boldsymbol{\theta} \in \boldsymbol{\Omega} \coloneqq \mathbb{R}^P \setminus \{\boldsymbol{0}\}$. The worst case design, in terms of the Neyman-Pearson setting for a fixed η [cf. (12)], can be obtained by maximizing $\lambda^2(\boldsymbol{\theta})$ for the worst case parameter $\boldsymbol{\theta} \in \boldsymbol{\Omega}$. Therefore, the max-min problem for the wost-case design can be posed as

$$\underset{\mathcal{A} \subseteq \mathcal{V}, |\mathcal{A}| = K}{\text{maximize}} \ \underset{\boldsymbol{\theta} \neq \boldsymbol{0}}{\text{minimize}} \ \boldsymbol{\theta}^T \boldsymbol{G}_{\mathcal{A}} \boldsymbol{\theta}, \tag{15}$$

where, by using the definition (7), the matrix $\boldsymbol{G}_{\mathcal{A}}$ is given by

$$\boldsymbol{G}_{\mathcal{A}} \coloneqq \boldsymbol{H}_{\mathcal{A}}^T \big[\boldsymbol{I}_K - \boldsymbol{S}_{\mathcal{A}} (\boldsymbol{S}_{\mathcal{A}}^T \boldsymbol{S}_{\mathcal{A}})^{-1} \boldsymbol{S}_{\mathcal{A}}^T\big] \boldsymbol{H}_{\mathcal{A}}. \tag{16}$$

Without loss of generality, for $\|\boldsymbol{\theta}\|_2^2 = 1$, the worst case $\boldsymbol{\theta}^T \boldsymbol{G}_{\mathcal{A}} \boldsymbol{\theta}$ is obviously given by the minimum eigenvalue of $\boldsymbol{G}_{\mathcal{A}}$. As a result, the problem in (15) can be simplified to

$$\underset{\mathcal{A} \subseteq \mathcal{V}, |\mathcal{A}| = K}{\text{maximize}} \ \lambda_{\min}(\boldsymbol{G}_{\mathcal{A}}), \tag{17}$$

which can be approximated by a relaxed concave problem (see Appendix A) in terms of a selection vector $\boldsymbol{w} \in \{0,1\}^N$ whose nonzero entries are defined by the set $\mathcal{A}$.

4.3. Log-det Design

It is well-known that the max-min criterion for a composite hypothesis test has an inherently pessimistic nature [15]. Therefore, a less stringent design, based on the maximization of the determinant of the matrix $\boldsymbol{G}_{\mathcal{A}}$, can be considered. That is, we could solve

$$\underset{\mathcal{A}\subseteq\mathcal{V},|\mathcal{A}|=K}{\text{maximize}} \ \ln\det(\boldsymbol{G}_{\mathcal{A}}). \tag{18}$$

In (18), instead of only maximizing the minimum eigenvalue, the product of the eigenvalues is maximized. This approach indeed tries to increase the energy in the worst direction, but also tries to distribute the power in other directions as well.

Similar to the optimization problem in (17), the solution of this problem can be approximated by means of a concave problem using the Schur complement. However, instead of only relying on the convex machinery, which requires a series of relaxations (see Appendix B), we propose a submodular relaxation [16] that is constructed in a similar fashion as the convex relaxation but that can be optimized near-optimally through a greedy heuristic [17].

To construct a submodular surrogate for (18) we leverage the fact that $\boldsymbol{G}_{\mathcal{A}} \succeq \mathbf{0},\ \forall\, \mathcal{A} \subseteq \mathcal{V}$. This condition can be expressed using the Schur complement through the following linear matrix inequality

$$\boldsymbol{M}_{\mathcal{A}} := \begin{bmatrix} \boldsymbol{S}^T\boldsymbol{I}_{\mathcal{A}}\boldsymbol{S} & \boldsymbol{S}^T\boldsymbol{I}_{\mathcal{A}}\boldsymbol{H} \\ \boldsymbol{H}^T\boldsymbol{I}_{\mathcal{A}}\boldsymbol{S} & \boldsymbol{H}^T\boldsymbol{I}_{\mathcal{A}}\boldsymbol{H} \end{bmatrix} \succeq \mathbf{0}. \tag{19}$$

As the determinant of the matrices $\boldsymbol{G}_{\mathcal{A}}$ and $\boldsymbol{M}_{\mathcal{A}}$ are related, i.e.,

$$\det(\boldsymbol{M}_{\mathcal{A}}) = \det(\boldsymbol{S}^T\boldsymbol{I}_{\mathcal{A}}\boldsymbol{S})\det(\boldsymbol{G}_{\mathcal{A}}), \tag{20}$$

instead of directly solving problem (18) the following problem can be considered

$$\underset{\mathcal{A}\subset\mathcal{V},|\mathcal{A}|=K}{\text{maximize}} \ \ln\det(\boldsymbol{M}_{\mathcal{A}}). \tag{21}$$

This problem can be solved near-optimally using a greedy heuristic. The greedy algorithm at the kth step adds to the current solution set $\mathcal{A}_k$ the element $i \in \mathcal{V} \setminus \mathcal{A}_k$ which maximizes the gain $f(\mathcal{A}_k \cup \{i\}) - f(\mathcal{A}_k)$. The $(1-1/e)$ optimality [17] of the greedy method when solving (21) is due to the submodularity (see Appendix D) and non-decreasing (see Appendix C) nature of the cost set function $f(\mathcal{A}) := \ln\det(\boldsymbol{M}_{\mathcal{A}})$.

5. NUMERICAL EXPERIMENTS

In this section, we illustrate the proposed methods through numerical examples[1]. We consider an example from array signal processing in which a uniform linear array (ULA) with $N = 15$ elements is used to perform detection. It is assumed that the matrix $\boldsymbol{H}$ is composed by the steering vectors for angle of arrivals (AoAs) $\{-30^o,\ 0^o,\ 50^o\}$. Furthermore, the interference matrix $\boldsymbol{S}$ is considered to be described by the steering vectors corresponding to the AoAs $\{-70^o,\ 30^o\}$. The noise is assumed to be white Gaussian noise with unknown power σ^2. As the noncentrality parameter is monotone in $1/\sigma^2$ and $\|\boldsymbol{\theta}\|_2^2$, without loss of generality, we consider both quantities as unity.

[1]The code to reproduce the figures presented in this paper can be found in https://gitlab.com/fruzti/SparseMatchedDetector

In Fig. 1a, we show a comparison between different methods for approximating the solution (17). In this figure 'Fwd', 'Bck' and 'Init Fwd' stand for the *forward*, *backward*, and *initialized forward* greedy solution, respectively. The distinction between forward and backward lies in the fact that the former *adds* sensors, starting from the empty set, and the later *removes* sensors, starting from the full set. As when $K < P + Q$ the matrix $[\boldsymbol{H}_{\mathcal{A}}\ \boldsymbol{S}_{\mathcal{A}}]$ is rank deficient, the *Fwd* Greedy method considers the minimum nonzero eigenvalue. For the *Init Fwd* method, first the optimal solution for $P + Q = 5$ sensors is found through exhaustive search and it is then used as initialization set for the greedy heuristic. For this reason, only the performance for $K \geq 5$ is reported. From Fig. 1a, it can be seen that the convex relaxation (*CVX minEig*) and *Bck Greedy* perform close to the optimal solution. Even though there is no guarantee for the greedy heuristic to perform near optimally, as $\lambda_{\min}(\boldsymbol{G}_{\mathcal{A}})$ is not a submodular set function, it could be useful to compare the solutions of the convex approach with the greedy alternative to select the best among them. In the figure, the shaded region shows the performance of any random sampler.

In Fig. 1b, the performance of the log determinant case is shown. In this plot we show that the submodular surrogate proposed in this work achieves a comparable performance with respect to its convex counterpart, where both of them perform close to the exhaustive search. However, while the convex relaxation is cast as a semidefinite program (SDP) that has cubic complexity, the submodular surrogate is linear in the number of sensors to select.

Finally, for illustration in Fig. 1c we show an example for $K = 9$, of the obtained beam patterns after the subspace matched filter (SMF) is applied [cf. (6)]. It can be observed that the beam pattern obtained by the *Bck Greedy* method has nulls exactly at the position of the interferer angles and peaks at the desired AoAs. This result shows that due to the nature of the GLRT, the sparse sampler design done through these approaches is not only meaningful for detection but also for estimation.

6. CONCLUSION

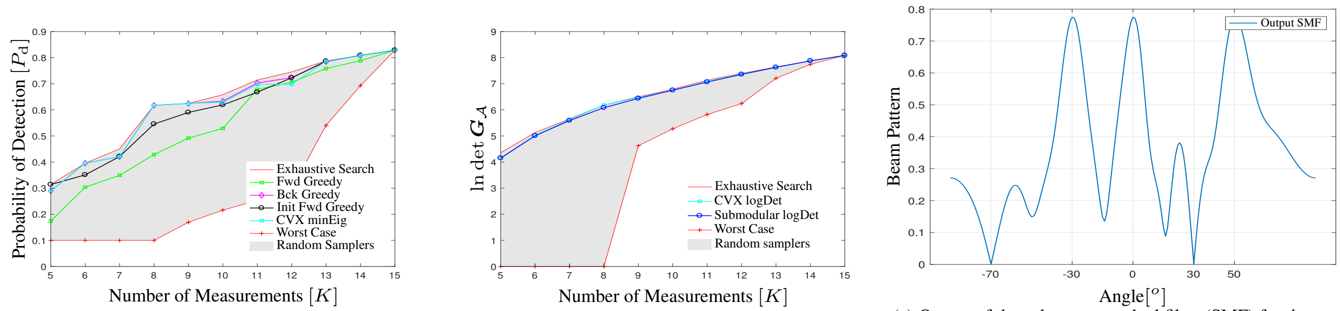
In this paper, we have explored sparse sampler design for composite testing with particular focus on matched subspace detectors. The main idea behind the design strategies is to lever the optimality of the GRLT for this kind of problems. We have shown that a max-min sampler design can be obtained by means of a convex program by applying appropriate convex relaxations. However, due to the inherent pessimistic nature of the max-min criterion, i.e., maximizing the minimum eigenvalue of $\boldsymbol{G}_{\mathcal{A}}$, we proposed to maximize its determinant by means of a submodular surrogate which can be maximized through a greedy heuristic. Even though this work is limited to binary hypothesis tests, future directions aim towards sampler designs for multiple hypotheses testing problems.

A. CONCAVE FORMULATION FOR MAXIMIZING $\lambda_{\min}(\boldsymbol{G}_{\mathcal{A}})$

First, let us consider a selection vector $\boldsymbol{w} \in \{0,1\}^N$ whose nonzero entries are defined by the set $\mathcal{A} \subset \{1, 2, \ldots, N\}$. That is, $w_m = 1$ if $m \in \mathcal{A}$ and $w_m = 0$ for $m \notin \mathcal{A}$.

Recalling the expression for $\boldsymbol{G}_{\mathcal{A}}$ and interchanging the dependency on the set $\mathcal{A}$ with the selection variable $\boldsymbol{w}$, we can express (16) equivalently as

$$\boldsymbol{G}_{\boldsymbol{w}} = \boldsymbol{H}^T\boldsymbol{I}_{\boldsymbol{w}}\boldsymbol{H} - \boldsymbol{H}^T\boldsymbol{I}_{\boldsymbol{w}}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{I}_{\boldsymbol{w}}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{I}_{\boldsymbol{w}}\boldsymbol{H}, \tag{22}$$



(a) Comparison of the different methods with respect to P_d when optimizing $\lambda_{\min}(\boldsymbol{G}_\mathcal{A})$.

(b) Comparison of the different methods when optimizing $\ln\det(\boldsymbol{G}_\mathcal{A})$.

(c) Output of the subspace matched filter (SMF) for the solution of the *Bck Greedy* method for $K = 9$.

Fig. 1

where $\boldsymbol{I_w} = \mathrm{diag}(\boldsymbol{w})$.

By introducing the linear matrix inequality (LMI)

$$\boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{H} - \boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{H} \succeq \lambda_{\min}\boldsymbol{I}, \tag{23}$$

we can lever the Schur complement to express such condition linearly in $\boldsymbol{w}$ and $\lambda_{\min}$ as

$$\begin{bmatrix} \boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{S} & \boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{H} \\ \boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{S} & \boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{H} - \lambda_{\min}\boldsymbol{I} \end{bmatrix} \succeq \boldsymbol{0}. \tag{24}$$

As a result, we can maximize the minimum eigenvalue of $\boldsymbol{G_w}$ through the following relaxed concave program

$$\begin{aligned} &\underset{\boldsymbol{w},\lambda_{\min}}{\text{maximize}} \quad \lambda_{\min} \\ &\text{subject to} \\ &\qquad \begin{bmatrix} \boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{S} & \boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{H} \\ \boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{S} & \boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{H} - \lambda_{\min}\boldsymbol{I} \end{bmatrix} \succeq \boldsymbol{0} \\ &\qquad \boldsymbol{w} \in [0,1]^N, \|\boldsymbol{w}\|_1 \leq K \end{aligned} \tag{25}$$

where the Boolean nature of the selection vector has been relaxed and replaced by a box constraint, and the nonconvex ℓ_0-norm that represents the cardinality constraint has been relaxed and substituted by its convex surrogate, the ℓ_1-norm.

B. CONCAVE FORMULATION FOR MAXIMIZING $\ln\det(\boldsymbol{G}_\mathcal{A})$

Recalling the definition in (22) and introducing a new variable $\boldsymbol{Z} \in \mathbb{R}^{P\times P}$, we can rewrite (22) as

$$\boldsymbol{G_w} = \boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{H} - \boldsymbol{Z}, \tag{26}$$

where $\boldsymbol{Z} = \boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{S}(\boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{S})^{-1}\boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{H}$.

By relaxing the definition of $\boldsymbol{Z}$ from strict equality to an inequality in the cone of the positive semidefinite matrices (PSD), we can use the Schur complement to write the following concave program

$$\begin{aligned} &\underset{\boldsymbol{w},\,\boldsymbol{Z}}{\text{maximize}} \quad \ln\det\left(\boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{H} - \boldsymbol{Z}\right) \\ &\text{subject to} \quad \begin{bmatrix} \boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{S} & \boldsymbol{S}^T\boldsymbol{I_w}\boldsymbol{H} \\ \boldsymbol{H}^T\boldsymbol{I_w}\boldsymbol{S} & \boldsymbol{Z} \end{bmatrix} \succeq \boldsymbol{0} \\ &\qquad\qquad\quad \boldsymbol{w} \in [0,1]^N, \quad \|\boldsymbol{w}\|_1 \leq K \end{aligned} \tag{27}$$

where the nonconvex Boolean nature of the selection vector $\boldsymbol{w}$ has been relaxed to the box $[0,1]^N$, the nonconvex ℓ_0-norm has been substituted by the convex ℓ_1-norm and a linear matrix inequality (LMI) has been introduced to deal with the inverse in (22). The problem (27) can then be used to approximate the solution of (18).

C. MONOTONICITY OF $\ln\det(\boldsymbol{M}_\mathcal{A})$

First, recall the definition of $\boldsymbol{M}_\mathcal{A}$ given by

$$\boldsymbol{M}_\mathcal{A} = \begin{bmatrix} \boldsymbol{S}^T\boldsymbol{I}_\mathcal{A}\boldsymbol{S} & \boldsymbol{S}^T\boldsymbol{I}_\mathcal{A}\boldsymbol{H} \\ \boldsymbol{H}^T\boldsymbol{I}_\mathcal{A}\boldsymbol{S} & \boldsymbol{H}^T\boldsymbol{I}_\mathcal{A}\boldsymbol{H} \end{bmatrix}, \text{ for some } \mathcal{A} \subset \mathcal{V}. \tag{28}$$

In order to show that $f(\mathcal{A}) := \ln\det(\boldsymbol{M}_\mathcal{A})$ is a non-decreasing set function, i.e., $f(\mathcal{A}) \geq f(\mathcal{B})$, $\forall\mathcal{A} \subset \mathcal{B} \subset \mathcal{V}$, we require to prove that

$$\ln\frac{\det(\boldsymbol{M}_\mathcal{A} + \boldsymbol{M}_{\{i\}})}{\det(\boldsymbol{M}_\mathcal{A})} \geq 0, \text{ for some } i \in \mathcal{V}\setminus\mathcal{A}. \tag{29}$$

This condition implies that $\boldsymbol{M}_\mathcal{A} + \boldsymbol{M}_{\{i\}} \succeq \boldsymbol{M}_\mathcal{A}$, which is always true as $\boldsymbol{M}_{\{i\}} \succeq 0\ \forall\{i\} \subset \mathcal{V}$. Therefore, the claim is proven.

D. SUBMODULARITY OF $\ln\det(\boldsymbol{M}_\mathcal{A})$

To show that the cost set function $f(\mathcal{A}) := \ln\det(\boldsymbol{M}_\mathcal{A})$ is a submodular set function, the following inequality has to be proven

$$\Delta(\mathcal{A},\mathcal{B}) := f(\mathcal{A}\cup\{i\}) - f(\mathcal{A}) - f(\mathcal{B}\cup\{i\}) - f(\mathcal{B}) \geq 0, \tag{30}$$

where, without loss of generality, we consider $\mathcal{B} = \mathcal{A}\cup\{j\}$. Expression (30) can be rewritten as

$$\Delta(\mathcal{A},\mathcal{B}) = \ln\frac{\det(\boldsymbol{M}_\mathcal{A} + \boldsymbol{M}_{\{i\}})\det(\boldsymbol{M}_\mathcal{B})}{\det(\boldsymbol{M}_\mathcal{A})\det(\boldsymbol{M}_{\mathcal{B}\cup\{i\}})}, \tag{31}$$

which by using the determinant lemma can be expressed as

$$\ln\frac{\det(\boldsymbol{M}_\mathcal{A})\det(\boldsymbol{I} + \boldsymbol{L}_{\{i\}}^T\boldsymbol{M}_\mathcal{A}^{-1}\boldsymbol{L}_{\{i\}})\det(\boldsymbol{M}_\mathcal{B})}{\det(\boldsymbol{M}_\mathcal{A})\det(\boldsymbol{M}_\mathcal{B})\det(\boldsymbol{I} + \boldsymbol{L}_{\{i\}}^T\boldsymbol{M}_\mathcal{B}^{-1}\boldsymbol{L}_{\{i\}})}, \tag{32}$$

where $\boldsymbol{L}_{\{i\}}$ is defined as any appropriate square-root of $\boldsymbol{M}_{\{i\}}$. The previous expression is reduced to

$$\Delta(\mathcal{A},\mathcal{B}) = \ln\frac{\det(\boldsymbol{I} + \boldsymbol{L}_{\{i\}}^T\boldsymbol{M}_\mathcal{A}^{-1}\boldsymbol{L}_{\{i\}})}{\det(\boldsymbol{I} + \boldsymbol{L}_{\{i\}}^T\boldsymbol{M}_\mathcal{B}^{-1}\boldsymbol{L}_{\{i\}})}. \tag{33}$$

As $\boldsymbol{M}_\mathcal{B} \succeq \boldsymbol{M}_\mathcal{A}$, $\forall\mathcal{A} \subset \mathcal{B}$, we have that

$$\Delta(\mathcal{A},\mathcal{B}) \geq 0,\ \forall\mathcal{A} \subset \mathcal{B}. \tag{34}$$

E. REFERENCES

[1] A. Papoulis, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[2] H. Van Trees, *Detection, estimation, and modulation theory*. John Wiley & Sons, 2004.

[3] S. Kay, *Fundamentals of statistical signal processing: Detection theory, vol. 2.* (1998).

[4] L. Scharf, and B. Friedlander. *Matched subspace detectors.* IEEE Transactions on Signal Processing 42.8 (1994): 2146-2157.

[5] R. Richards, J. A. Scheer, and W. A. Holm. *Principles of modern radar*. SciTech Pub., 2010.

[6] B. Sklar. *Digital communications*. Vol. 2. Upper Saddle River: Prentice Hall, 2001.

[7] G. Giannakis, and M. Tsatsanis. *Signal detection and classification using matched filtering and higher order statistics.* IEEE Transactions on Acoustics, Speech, and Signal Processing, 38.7 (1990): 1284-1296.

[8] Q. Ding, and E. Kolaczyk. *A compressed PCA subspace method for anomaly detection in high-dimensional data*. IEEE Transactions on Information Theory, 59.11 (2013): 7419-7433.

[9] D. Bajovic, B. Sinopoli, and J. Xavier. *Sensor selection for event detection in wireless sensor networks*. IEEE Transactions on Signal Processing 59.10 (2011): 4938-4953.

[10] Z. Wang, R. Gonzalo, and B. Sadler. *Subspace compressive detection for sparse signals.* Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. IEEE, 2008.

[11] M. Davenport, P.T. Boufounos, M.B. Walkin, and R.G. Baraniuk, *Signal processing with compressive measurements*. IEEE Journal of Selected Topics in Signal Processing 4.2 (2010): 445-460.

[12] S.P. Chepuri, and G. Leus. *Sparse sensing for distributed detection*. IEEE Transactions on Signal Processing, 64.6 (2016): 1446-1460.

[13] B.C. Levy, *Principles of signal detection and parameter estimation.* Springer Science & Business Media, 2008.

[14] S.P. Chepuri, and G. Leus. *Sparse Sensing for Statistical Inference.* Foundations and Trends in Signal Processing, 9.34 (2016): 233-368.

[15] M. Feder, and N. Merhav. *Universal composite hypothesis testing: A competitive minimax approach*. IEEE Transactions on information theory, 48.6 (2002): 1504-1517.

[16] S. Fujishige. *Submodular functions and optimization*. Vol. 58. Elsevier, 2005.

[17] G. Nemhauser, et al. *An analysis of approximations for maximizing submodular set functions*. Mathematical Programming, 14.1 (1978): 265-294.