

Dynamic Sensor Selection for Reliable Spectrum Sensing via E-optimal Criterion

Mohsen Joneidi, Alireza Zaeemzadeh, and Nazanin Rahnavard
 Department of Electrical and Computer Engineering, University of Central Florida
 {joneidi, zaeemzadeh, nazanin}@eecs.ucf.edu

Abstract—Reliable and efficient spectrum sensing through dynamic selection of a subset of spectrum sensors is studied. The problem of selecting K sensor measurements from a set of M potential sensors is considered where $K \ll M$. In addition, K may be less than the dimension of the unknown variables of estimation. Through sensor selection, we reduce the problem to an under-determined system of equations with potentially infinite number of solutions. However, the sparsity of the underlying data facilitates limiting the set of solutions to a unique solution. Sparsity enables employing the emerging compressive sensing technique, where the compressed measurements are selected from a large number of potential sensors. This paper suggests selecting sensors in a way that the reduced system of equations constructs a well-conditioned measurement matrix. Our criterion for sensor selection is based on E-optimality, which is highly related to the restricted isometry property that provides some guarantees for sparse solution obtained by ℓ_1 minimization. Moreover, the proposed framework exploits a feedback mechanism to evolve the selected sensors dynamically over time. The evolution aims to maximize the reliability of the sensed spectrum.

Key-words: Sensor Selection, E-optimality, Restricted Isometry Property (RIP), Matrix Subset Selection, Compressive Spectrum Sensing and Sparse Recovery.

I. INTRODUCTION

In the last decade, complex systems containing very large numbers of data-gathering devices were developed. An example is wireless sensor networks. In such systems, the processing unit has to deal with an excessively large number of observations acquired by the various sensors. Often there exist some redundancies within the sensed data and they should be pruned. Sensor selection and sensor scheduling aim to address this problem. In many applications the sensor selection task is non-trivial and possibly consists of addressing an NP-hard problem (i.e., there are $\binom{M}{K}$ possibilities of choosing K distinct sensors out of M available ones). This essentially implies that an optimal solution cannot be efficiently computed, in particular when the number of sensors becomes excessively large. A convex relaxation of the original NP-hard problem has been suggested in [1]. The most prominent advantage of this approach over other methods is its practicality, thanks to many well-established computationally-efficient convex optimization techniques. In addition to convex relaxation, a sub-modular cost function as the criterion of sensor selection allows us to take advantage of greedy optimization methods for selecting sensors [2].

The existing studies on sensor selection mostly consider heuristic approaches. For example, in [1] the volume of the reduced bases is considered. This method is called *D-optimality*.

In addition, *A-optimality* [3] and *E-optimality* [3] are suggested as some other alternative heuristics already introduced in convex optimization. These heuristics are presented without any specific justification for sensor selection application. In this paper we are going to exploit a criteria more judiciously in favor of compressed sensing (CS) theoretical guarantees.

Inspired by the compressed sensing theory, this paper suggests to design and optimize a sensor selection method. The goal is to reduce a measurement matrix to only a small fraction of its rows in order to optimize the proposed E-optimal criterion.

The main contributions of the paper are summarized as follows:

- The link between matrix subset selection, especially volume sampling and sensor selection is investigated.
- The E-optimal criterion for matrix subset selection is proposed, which results in a new sensor selection method,
- The suitability of the E-optimal criterion is discussed which is an upper bound for RIP coefficients for compressive sensing, and
- The reliability concept for power spectrum map is introduced and it is exploited for reliable sensor selection.

Table I presents the employed notations throughout this paper.

The rest of paper is organized as follows. Section II-C states the problem of sensor selection and reviews some existing methods. E-optimal sampling is introduced in Section III and a new sensor selection method is proposed. Section IV propose dynamic sensor selection based on the reliability. The optimization method for solving the proposed problem is explained in Section V. Section VI presents the simulation results and Section VII concludes the paper.

Table I: Employed notations and variables in this paper.

Variable Type	Notation
Constant Scalar	X
Vector	\mathbf{x}
s^{th} entry of Vector	x_s
Matrix	\mathbf{X}
Set	\mathbb{X}
Selected Rows of \mathbf{A} by set \mathbb{X}	$\mathbf{A}_{\mathbb{X}}$
Number of non-zero entries of \mathbf{x}	$\ \mathbf{x}\ _0$
Trace of Matrix \mathbf{X}	$\text{Tr}(\mathbf{X})$
Projection of \mathbf{X} on its rows set \mathbb{T}	$\pi_{\mathbb{T}}(\mathbf{X})$
M	number of potential sensors
K	number of selected sensors

II. BACKGROUND

This paper address a joint framework for spectrum sensing with partial sensing from a big set of sensors. The partial data are selected through a sensor selection procedure. Viewing spectrum sensing and sensor selection together in a joint problem is resulted to inspiring theoretical results in addition to a new application.

The prerequisite background of the proposed framework is reviewed in this section. The first subsection reviews compressed sensing theory and then the system model of spectrum sensing is introduced in the second subsection. The third subsection introduces the sensor selection problem. The last subsection review theoretical results in matrix subset selection literature which is highly related to our proposed selection method.

A. Compressed Sensing

Compressed sensing is a technique by which sparse signals can be measured at a rate less than conventional Nyquist sampling theorem. [4]. There exist vast applications of CS in signal and image processing [5], channel estimation [6] and spectrum sensing [7]. CS aims to recover a sparse vector, \mathbf{x} , using a small number of measurements \mathbf{y} . The CS problem can be formulated as,

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_0 \quad \text{s.t. } \mathbf{y} = \mathbf{\Phi}\mathbf{x}, \quad (1)$$

where, $\|\cdot\|_0$ represents the number of non-zero elements of a vector. $\mathbf{\Phi} \in \mathbb{R}^{K \times N}$ is called measurement matrix that provides us K measurements collected in \mathbf{y} . Exact solution of the above optimization problem is through the combinational search among all possible subsets. Due to its high computational burden, this algorithm is impractical for high dimension scenarios. Many sub-optimal algorithms have been proposed such as OMP [8], smoothed ℓ_0 [9] and basis pursuit [10]. Basis pursuit is based on relaxing ℓ_0 to ℓ_1 norm and is popular due to theoretical guarantees and reasonable computational burden [11]. The theoretical guarantees for ℓ_1 minimization arise from several sufficient conditions based on some suggested metrics. These include the mutual coherence [12], null space property [13], spark [14] and restricted isometry property (RIP) [15]. Except for the mutual coherence, none of these measures can be efficiently calculated for an arbitrary given measurement matrix $\mathbf{\Phi}$. For example, the RIP requires enumerating over an exponential number of index sets. RIP is defined as follows.

Definition 1. [15] *A measurement matrix is said to satisfy symmetric form RIP of order S with constant δ_S if δ_S is the smallest number that*

$$(1 - \delta_S)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1 + \delta_S)\|\mathbf{x}\|_2^2, \quad (2)$$

holds for every S -sparse \mathbf{x} (i.e. \mathbf{x} contains at most S nonzero entries).

Based on this definition several guarantees are proposed in terms of δ_{2S} , δ_{3S} and δ_{4S} in [16] and [17] in order to guarantee recovering S -sparse vectors. By S -sparse we mean a vector that has S non-zero entries. In [18] an asymmetric form of

definition 1 is introduced in order to more precisely quantify the RIP.

Definition 2. [18] *For a measurement matrix the asymmetric RIP constants δ_S^L and δ_S^U are defined as,*

$$\begin{aligned} \delta_S^L(\mathbf{\Phi}) &= \underset{c>0}{\operatorname{argmin}} (1 - c)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2, \quad \forall \mathbf{x} \in \mathcal{X}_S^N, \\ \delta_S^U(\mathbf{\Phi}) &= \underset{c>0}{\operatorname{argmin}} (1 + c)\|\mathbf{x}\|_2^2 \geq \|\mathbf{\Phi}\mathbf{x}\|_2^2, \quad \forall \mathbf{x} \in \mathcal{X}_S^N, \end{aligned} \quad (3)$$

where, \mathcal{X}_S^N refers to the set of S -sparse vectors in \mathbb{R}^N .

Remark 1. [18] *Although both the smallest and largest singular values of $\mathbf{\Phi}_S^T \mathbf{\Phi}_S^{-1}$ affect the stability of the reconstruction algorithms, the smaller eigenvalue is dominant for compressed sensing in that it allows distinguishing between sparse vectors, \mathcal{X}_S^N , given their measurements by $\mathbf{\Phi}$.*

B. Spectrum sensing problem statement

Cognitive radio (CR) is a promising solution to alleviate today's spectrum deficiency caused by an increased demand for the wireless technologies [19, 20]. The CR paradigm allows the unlicensed or secondary users (SUs) to coexist with the PUs. The SUs are allowed to access the spectrum provided that they do not interfere with the licensed users. The under-utilized spectrum bands that can be used by the SUs are called *spectrum holes* [21]. An ideal CR is able to efficiently detect and utilize spectrum holes.

Due to the scarce presence of active PUs and their narrow band transmission, sparse recovery methods are exploited to perform cooperative spectrum sensing [22, 23]. These approaches decompose the power spectrum density (PSD) of CRs, in terms of some appropriate bases which are related to the network topology and parameters.

We assume a network setup the same as that of [22]. Consider N_s points in a grid as the potential locations of transmitters and M receivers in an area. The receivers receive a superposition of transmitters signals. Figure 1 shows a setup consisting $N_s = 25$ potential transmitters in which 2 of them are active and there exist $M = 60$ sensors. The received signals are contaminated by channel gain and additive noise, represented by the following equation,

$$y_m = \mathbf{a}_m^T \mathbf{x} + v_m, \quad \forall m = 1 \dots M, \quad (4)$$

where, \mathbf{a}_m contains the corresponding channel gains and v_m represents noise power at the m^{th} receiver. Coefficients of \mathbf{x} correspond to the transmitted power at different grid points. The following problem aims to estimate \mathbf{x} collaboratively using all the measurements [22].

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \gamma \|\mathbf{x}\|_1, \quad (5)$$

in which, \mathbf{y} and \mathbf{A} are concatenation of y_m and \mathbf{a}_m respectively. Each entry of \mathbf{x} determines the contribution of the s^{th} source on the sensed data. Due to scarce presence of active transmitters

¹ \mathbb{S} represents a set with cardinality of S and $\mathbf{\Phi}_{\mathbb{S}}$ represents the corresponding selected rows of $\mathbf{\Phi}$.

and their narrow band communication, $\|\mathbf{x}\|_1$ is exploited which encourages sparsity. It should be noted that by estimating \mathbf{x} , we will know the location and power of transmitters at different frequency bands. These information enable us to build a radio environment map [24].

C. Sensor Selection Problem Statement

Solving the sensor selection problem by evaluating the performance for each of the possible choices of $\binom{M}{K}$ is impractical unless the sizes are sufficiently small.

Suppose we want to estimate a vector $\mathbf{x} \in \mathbb{R}^N$ from M linear measurements where each one is collected from a sensor, corrupted by additive noise, given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}, \quad (6)$$

where, $\mathbf{y} \in \mathbb{R}^M$ and $\mathbf{A} \in \mathbb{R}^{M \times N}$ and \mathbf{v} is normally distributed with zero mean and σ^2 variance. In other words, we want to only select just K rows of \mathbf{A} to have K measurements out of maximum M measurements. The corresponding rows of \mathbf{A} construct the measurement matrix, Φ , in compressed sensing literature. The maximum likelihood (ML) estimator is given by [1],

$$\hat{\mathbf{x}}_{ML} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}. \quad (7)$$

The estimation error $\mathbf{x} - \hat{\mathbf{x}}$ has zero mean and the covariance matrix is equal to

$$\Sigma_{ML} = \sigma^2 (\mathbf{A}^T \mathbf{A})^{-1}. \quad (8)$$

To involve selection operator in the equations let us first write the ML solution as follows,

$$\hat{\mathbf{x}}_{ML} = \left(\sum_{m=1}^M \mathbf{a}_m \mathbf{a}_m^T \right)^{-1} \sum_{m=1}^M y_m \mathbf{a}_m, \quad (9)$$

where, \mathbf{a}_m^T is the m^{th} row of \mathbf{A} . The estimation error is distributed in a high dimensional ellipsoid that its center is located at origin and its shape is according to the covariance matrix of error [1]. Minimization of volume of this ellipsoid (D-optimality) is the heuristic used in [1] that results in the following problem:

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \log \det \left(\sum_{m=1}^M w_m \mathbf{a}_m \mathbf{a}_m^T \right)^{-1}, \quad (10)$$

subject to $\|\mathbf{w}\|_0 = K$ and $\mathbf{w} \in \mathbb{B}^M$,

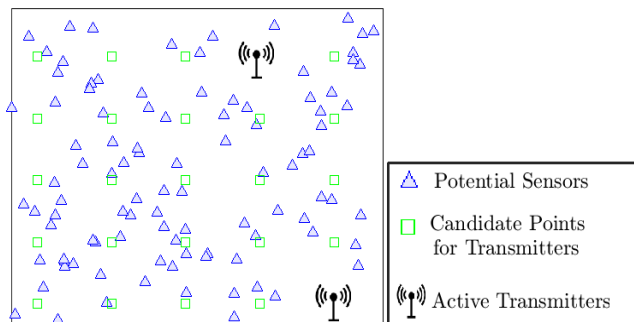


Figure 1: An example setup with 25 candidate points as transmitters.

where \mathbf{w} determines whether or not each column is involved and $\mathbb{B} = \{0, 1\}$.

The practical algorithms alternative to the combinatorial search are divided into two main categories, convex relaxation and greedy selection. The first approach approximates the search space to the nearest convex set and exploits convex optimization methods to solve the problem, while greedy methods gradually select suitable sensors or prune inefficient ones.

D. Matrix subset selection

The sensor selection problem is highly related to column/row sub-matrix selection, a fundamental problem in applied mathematics. There exists many efforts in this area [25–28]. Generally, they aim at devising a computationally efficient algorithm in which the span of the selected columns/rows cover the columns/rows space as close as possible. Mathematically, a general guarantee can be stated as one of the following forms [26, 29],

$$\mathbb{E}\{\|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2\} \leq (K+1)\|\mathbf{A} - \mathbf{A}_K\|_F^2,$$

$$\|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2 \leq p(K, M, N)\|\mathbf{A} - \mathbf{A}_K\|_F^2,$$

in which, $\pi_{\mathbb{T}}(\mathbf{A})$ represents projection of rows of \mathbf{A} on to the span of selected rows indexed by \mathbb{T} set. \mathbb{E} indicates expectation operator with respect to \mathbb{T} , i.e., all the combinatorial selection of K rows of \mathbf{A} out of M are considered. Moreover, $p(K, M, N)$ is a polynomial function of the number of selected elements, the number of columns and the number of rows. \mathbf{A}_K is the best rank- K approximation of \mathbf{A} that can be obtained by singular value decomposition. The first form suggests the distribution of potential sets for selection and it expresses an upper bound for expected value of error. The second form guarantees existence of a deterministic subset that bounds the error by a polynomial function of the parameters.

Volume sampling is the most well-known approach to achieve the desired selection that satisfies one of the aforementioned bounds. The following theorem expresses the probabilistic form volume sampling.

Theorem II.1 ([26]). *Let \mathbb{T} be a random K -subset of rows of a given matrix \mathbf{A} chosen with probability*

$$Pr(\mathbb{T}) = \frac{\det(\mathbf{A}_{\mathbb{T}} \mathbf{A}_{\mathbb{T}}^T)}{\sum_{|\mathbb{U}|=K} \det(\mathbf{A}_{\mathbb{U}} \mathbf{A}_{\mathbb{U}}^T)}$$

Then,

$$\mathbb{E}\{\|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2\} \leq (K+1)\|\mathbf{A} - \mathbf{A}_K\|_F^2.$$

Volume sampling considers more probability of selection for those rows whose volume is greater. The volume of a subset of a matrix, $\mathbf{A}_{\mathbb{T}}$, is proportional to the determinant of $\mathbf{A}_{\mathbb{T}} \mathbf{A}_{\mathbb{T}}^T$. The same heuristic criterion in (10) aims to find the subset which has the greatest volume. It indicates the most probable subset according to volume sampling. It shows the relation of volume sampling and sensor selection in which they are solving the same problem. However this heuristic criterion is not justified for ant specific task.

Volume sampling and D-optimality pursue the same heuristic objective. This heuristic does not promote a well-shaped matrix for compressive sensing purposes based on RIP. However, the analysis of optimization w.r.t the RIP coefficient is not an easy task due to the columns combinatorial behavior in addition to row selection for the basic sensor selection problem. To eliminate the column combinations, we consider all of the columns and consequently we come up with an optimization problem w.r.t the minimum eigenvalue that is known as E-optimality in the optimization literature [3]. Assume a simple selection from rows of $\mathbf{A} \in \mathbb{R}^{100 \times 3}$. Each row of \mathbf{A} , associated with a sensor, corresponds to a point in \mathbb{R}^3 . We are to select 2 sensors out of 100 based on D-optimality and E-optimality. Both solutions are initialized by the same sensor (sensor 1) and the criteria for the next selection varies. The D-optimal solution aims to maximize the surrounded area (gray area in Fig. 2) which is vulnerable to be an ill-shaped area while, E-optimal solution comes up with a well-shaped area due to maximizing the minimum eigenvalue (shaded area in Fig. 2).²

The following simple example illustrates the effect of E-optimality. Consider two matrices, $\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. The determinant of both matrices are equal, thus D-optimality does not favor one over the other, however, the second matrix is optimum based on E-optimality.

As we will see in the next section, for selection of K rows of $\mathbf{A} \in \mathbb{R}^{M \times N}$, the E-optimal criterion is equivalent to optimizing the RIP coefficient of order N , which is an upper bound for any arbitrary order of RIP coefficients. In the next section E-optimality will be exploited to develop a new sampling method for which its performance guarantee is analyzed. E-optimal criterion suggests optimization of an upper bound for a specific order of RIP. Moreover, in this paper we suggest a method to approximate a specific order of RIP. Based on it, a new RIP-based sensor selection algorithm is proposed.

III. E-OPTIMAL SAMPLING

Remark 1 promotes us to develop a new matrix subset selection method that reduces the matrix to have a well-conditioned sub-matrices in the CS sense. The dominant factor of RIP constant comes from the minimum eigenvalue of the reduced matrix. It suggests to exploit the following optimization problem for sensor selection₁

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \left\| \left(\sum_{m=1}^M w_m \mathbf{a}_m \mathbf{a}_m^T \right) \right\|, \quad (11)$$

subject to $\|\mathbf{w}\|_0 = K$ and $\mathbf{w} \in \mathbb{B}^M$.

In which, $\|\cdot\|$ denotes the spectral norm of a matrix that is defined as its maximum eigenvalue. The following lemma shows that the minimum eigenvalue is an upper bound for δ_S^L .

²The presented intuition about D-optimality and E-optimality relates to the condition number of a matrix in linear algebra [30]. Diverged eigenvalues results in a large condition number and an ill-conditioned system of equations; accordingly, we refer to the polygon of an ill-conditioned system of equations as ill-shaped where the vertexes of shape are the rows of the matrix. On the other hand, close eigenvalues correspond to a small condition number and a well-conditioned system of equations. The corresponding polygon is referred as well-shaped in Fig 2. Having well-conditioned matrices, is a central concern in CS as evidenced by the role played by the RIP [31].

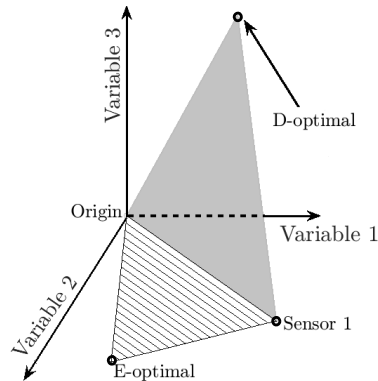


Figure 2: Comparison of D-optimality and E-optimality for selecting 2 sensors in the 3D space. The gray area is the maximum achievable area by selecting the second sensor based on D-optimality. The shaded area is a well-shaped polygon obtained by E-optimality.

Lemma III.1. For any $\mathbf{A} \in \mathbb{R}^{M \times N}$, the following inequality holds.

$$1 - \sigma_{\min}(\mathbf{A}\mathbf{A}^T) = \delta_N^L(\mathbf{A}) \geq \delta_{N-1}^L(\mathbf{A}) \geq \dots \geq \delta_2^L(\mathbf{A}).$$

Proof: It can be concluded directly by the interlacing property of eigenvalues [32].

Lemma III.1 suggests that E-optimality, i.e., minimization of δ_N^L , actually minimizes an upper bound for an arbitrary order of RIP coefficient.

Similar to volume sampling, we design a probability of sampling according to their minimum eigenvalue.

Definition 3. Given a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, E-optimal sampling is defined as picking a subset of \mathbb{T} with the following probability,

$$Pr(\mathbb{T}) = \frac{\sigma_{\min}^2(\mathbf{A}_{\mathbb{T}})}{\sum_{|\mathbb{U}|=K} \sigma_{\min}^2(\mathbf{A}_{\mathbb{U}})}.$$

Definition 4. Given a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\bar{\delta}_K^L$ is defined as one minus the mean of minimum eigenvalues of \mathbf{A} 's sub-matrices with K columns. Mathematically, it can be expressed as follows,

$$\bar{\delta}_K^L(\mathbf{A}) = 1 - \mathbb{E}\{\sigma_{\min}^2(\mathbf{A}_{\mathbb{S}})\},$$

in which \mathbb{S} indicates a subset of K columns of \mathbf{A} .

Definition 5. [14] Given a matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, the spark of \mathbf{A} is defined as the smallest number of columns that are linearly dependent. It can be stated as follows,

$$\operatorname{Spark}(\mathbf{A}) = \min \|\mathbf{x}\|_0 \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{0} \text{ and } \mathbf{x} \neq \mathbf{0}.$$

The upper bound for spark is the rank of matrix plus 1. However any linear dependencies among some columns of the matrix may decrease the spark. Based on the above definitions we present the following theorem that expresses an upper bound for projection error of E-optimal sampling.

Theorem III.2. Assume spark of $\mathbf{A} \in \mathbb{R}^{M \times N}$ is greater than $K + 1$. E-optimal selection of K rows implies

$$\mathbb{E}\{\|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2\} \leq \frac{M - K}{C(K + 1)} \frac{1 - \bar{\delta}_{K+1}^L(\mathbf{A}^T)}{1 - \bar{\delta}_K^L(\mathbf{A}^T)},$$

Table II: Complexity of different selection strategies.

Algorithm	Complexity
Convex Optimization [1]	$O(M^3)$
Volume sampling [25]	$O(KNM^2 \log M)$
Greedy Submodular Selection [2]	$O(MK^3)$
Greedy E-optimal selection	$O(MNK^2)$

where C is a positive number a function of the dependencies of rows of \mathbf{A} .

proof: First, let us write the expansion of expected value operator according to the definition.

$$\begin{aligned} & \mathbb{E}\{\|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2\} \\ &= \frac{1}{\sum_{|\mathbb{T}|=K} \sigma_{\min}^2(\mathbf{A}_{\mathbb{T}})} \sum_{|\mathbb{T}|=K} \sigma_{\min}^2(\mathbf{A}_{\mathbb{T}}) \|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2. \end{aligned} \quad (12)$$

Based on the assumption on the spark of \mathbf{A} , there exist a positive constant, α , that satisfies the following equation for every $|\mathbb{T}| = K$ and $|\mathbb{S}| = K + 1$ in which $\mathbb{T} \subset \mathbb{S}$.

$$\sigma_{\min}^2(\mathbf{A}_{\mathbb{S}}) = \alpha \sigma_{\min}^2(\mathbf{A}_{\mathbb{T}}) d(\mathbf{a}_m, \pi_{\mathbb{T}}(\mathbf{A})).$$

Where \mathbf{a}_m is the innovation of \mathbb{S} w.r.t \mathbb{T} and $d(\cdot, \cdot)$ represents the Ecludian distance of a vector from a subspace. Let us take summation on all of the possible combinations,

$$\begin{aligned} \sum_{|\mathbb{S}|=K+1} \sigma_{\min}^2(\mathbf{A}_{\mathbb{S}}) &= \sum_{|\mathbb{T}|=K} \sigma_{\min}^2(\mathbf{A}_{\mathbb{T}}) \sum_{m=1}^M \alpha_m d(\mathbf{a}_m, \pi_{\mathbb{T}}(\mathbf{A})) \\ &\geq C \sum_{|\mathbb{T}|=K} \sigma_{\min}^2(\mathbf{A}_{\mathbb{T}}) \|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2, \end{aligned} \quad (13)$$

where C is the minimum value of α_m 's for all of the possible combinations. The assumption on the spark guarantees the existence of a positive constant. Note that the summation of the distances for all of the rows \mathbf{A} can be stated as the Frobenius norm. Let us re-write the obtained inequality as follows,

$$\sum_{|\mathbb{T}|=K} \sigma_{\min}^2(\mathbf{A}_{\mathbb{T}}) \|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2 \leq \frac{1}{C} \sum_{|\mathbb{S}|=K+1} \sigma_{\min}^2(\mathbf{A}_{\mathbb{S}})$$

Dividing both sides of the inequality by $\sum_{|\mathbb{T}|=K} \sigma_{\min}^2(\mathbf{A}_{\mathbb{T}})$ results in $\mathbb{E}\{\|\mathbf{A} - \pi_{\mathbb{T}}(\mathbf{A})\|_F^2\}$ in the left side. After a simple simplification in terms of the coefficients introduced in Definition 2 and Definition 4, the right side turns into the desired expression. Please note the defined CS coefficients in (3) work on the columns while we are to select some rows. ■

E-optimal sampling implies an upper bound for the expectation of projection error in a probabilistic manner. However, we need to select some sensors deterministically. To this aim, we propose the following problem.

$$\mathbb{S} = \underset{\mathbb{S}}{\operatorname{argmax}} \lambda_{\min}(\mathbf{A}_{\mathbb{S}} \mathbf{A}_{\mathbb{S}}^T).$$

Algorithm 1 shows an iterative greedy method to solve this problem. Actually, this algorithm is an approximation for the maximum likelihood estimator in which the likelihood comes from the suggested probability in Definition 3.

Table II compares computational burden of three well-known selection methods with the proposed method. Convex relaxation is not able to work effectively for big data sets since

Algorithm 1 Greedy E-Optimal Sensor Selection

Require: \mathbf{A} and K

Output: The selected set \mathbb{S} .

- 1: **Initialization:** \mathbb{S} with a random sensor
 - 2: for $k = 1, \dots, K$
 - 3: for $m = 1, \dots, M$
 - 4: $\mathbb{T} = \mathbb{S} \cup \{m\}$
 - 5: $p(m) = \sigma_{\min}(\mathbf{A}_{\mathbb{T}})$
 - 6: end
 - 7: $s_k = \operatorname{argmax} p(m)$
 - 8: $\mathbb{S} = \mathbb{S} \cup s_k$
 - 9: end
-

the complexity of the algorithm grows with M^3 [1]. Complexity of volume sampling also depends on M^2 . Likewise, complexity of greedy algorithms which process data one-by-one increase linearly w.r.t size of data.

IV. RELIABILITY ESTIMATION AND DYNAMIC SENSOR SELECTION

Collaborative sensor networks may collect redundant information which results in a larger number of sensor nodes than is needed. While, pruning unnecessary data is essential, Algorithm 1 is measurement-independent and it reduces the underlying equations of the network to shrink the equations to a well-conditioned set of sub-equations regardless of dynamic of the network. This measurement-independent approach is optimal in an averaged sense, i.e., for different possible measurements. It is appropriate for a static regime or initialization of a dynamic sensor selection. This section proposes a dynamic sensor selection framework which considers measurements for sensor selection. First of all, let us define the dynamic sensor selection systematically as follows,

Definition 6. (Dynamic Sensor Selection) [33]: For a given model \mathcal{M} on the data, determine set \mathbb{S} such that the estimation error of the rest of sensors, \mathbb{S}^c , is minimized. The estimation is obtained based on the model, \mathcal{M} , and observed sensors, \mathbb{S} .

We assume the compressed sensing model (4) for power spectrum sensing as described in Section II-B. Let us denote the obtained spectrum power vector by the subset \mathbb{S} of sensors at time t as $\mathbf{x}_{\mathbb{S}}^t$. A proper selection of \mathbb{S} enables to predicting the power spectrum throughout the network's area.

In order to keep track of the network's dynamic, we propose to sample most of the nodes in a low rate mode; while some selected nodes should provide us with data sampled at a high rate enabling estimation of a high temporal resolution power spectrum map. In this framework, there is no completely switched off sensors, but we collect data from low-sampling rate sensors to dynamically select the sensors with high sampling rate. Therefore, we have two following types of sensors in our proposed framework,

- 1) *High-sampling-rate selected (active) sensors:* These are a small fraction of sensors selected by an underlying sensor selection mechanism in order to access real-time data

and generate a dynamic power spectrum map. The active sensors report their sensing at rate $f_h = 1$ sample per time block.

- 2) *Low-sampling-rate sensors*: All sensors collect and report their data in a low-rate mode, resulting in less bandwidth and power consumption. The low-rate data enables us to validate the estimated power spectrum map. The low-sampling sensors report their sensing at rate $f_l = \frac{1}{n_l}$ sample/time. I.e., 1 sample per n_l time blocks is collected. It should be mentioned that the measurements from low-sampling rate sensors will not contribute in estimating \mathbf{x} . They will be used to determine the reliability of estimation as we will discuss below.

The dynamic sensor selection aims to select some sensors as the active-mode set. The rest of sensors are marked as power efficient low sampling rate sensors. If the active set is selected properly, the rest of sensors can be predicted accurately by the assumed model and the active selected sensors. The ability of sensing is assumed same for all sensors and only the sensing time is different. However, different bandwidth for sensing can be considered in a more sophisticated framework which is out of scope of this paper. Selected sensors contain sufficient information enabling them to predict the rest of sensors by the assumed model on the spectrum (4). Low sampling rate data may cause obsolete information vulnerable to large deviation from the model. Moreover, changes in the dynamic of network also may cause large deviations between the model's estimation and the low sampling rate data. The following expression defines a new metric called *reliability* for sensor m at time t .

$$r_m^{(t)} = \frac{\exp(-\sigma(t - t_m))}{1 + |y_m - E^{(t)}(m, \mathbb{S})|^2}, \quad \forall m \in \{1, \dots, M\} \quad (14)$$

in which,

$$E^{(t)}(m, \mathbb{S}) = \mathbf{a}_m^T \mathbf{x}^{(t-1)}(\mathbb{S})$$

In (14), $\mathbf{x}^{(t-1)}(\mathbb{S})$ is the estimation of power propagation at time $t - 1$ based on collected data from active sensors indexed by \mathbb{S} . Moreover, $E^{(t)}(m, \mathbb{S})$ is the estimation of the measurement of m^{th} sensor at time t . σ is a temporal forgetting factor. t_m is the last time that sensor m is sampled and the corresponding measurement is y_m . The reliability of each sensor consists of two terms. The numerator indicates how fresh is our observation. Obsolete data results in unreliable observation. The denominator shows the power of model for estimation of unseen regions. Accurate estimation of the observation of sensor m using the active demonstrates that the sensor m has a reliable sensing. The proposed dynamic sensor selection framework is illustrated in Fig. 3. We propose to consider the reliability of sensors in the sensor selection procedure in order to determine a proper subset which is able to compensate large model's error for the low-rate sampled sensors. Mathematically speaking, the static E-optimal sensor selection algorithm is modified as follows,

$$\mathbb{S} = \underset{|\mathbb{S}| \leq K}{\operatorname{argmax}} \lambda_{\min}(A_{\mathbb{S}} A_{\mathbb{S}}^T) + \gamma \|\mathbf{u}_{\mathbb{S}}\|_2^2, \quad (15)$$

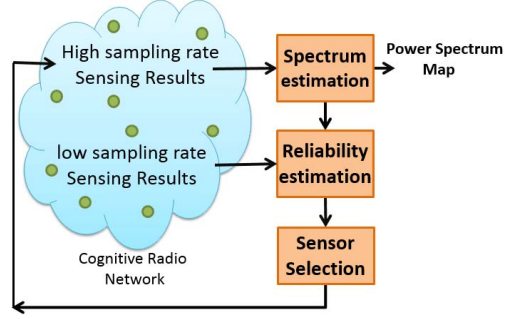


Figure 3: The main framework of the proposed reliability based sensor selection.

in which, γ is the regularization parameter and $u_m = r_m^{-1}$ represents unreliability and $\mathbf{u}_{\mathbb{S}}$ is the sub-vector of \mathbf{u} indexed by set \mathbb{S} . The superscript (t) is removed due to simplicity of notation. It means we are looking for unreliable sensors to select them for the next time slot in order to compensate the model's error.

V. OPTIMIZATION AND COMPLEXITY

In order to cast the dynamic sensor selection (15) in a tractable formulation, first let us rewrite the minimum eigenvalue as the following problem.

$$\lambda_{\min}(\mathbf{A}) = \min \|\mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_2 = 1. \quad (16)$$

Problem (15) can be written in the following form,

$$\mathbf{W}^{(t)} = \underset{\mathbf{W}}{\operatorname{argmax}} \min_x \|\mathbf{W}\mathbf{A}\mathbf{x}\|_2^2 + \gamma \|\mathbf{W}\mathbf{u}^{(t)}\|_2^2 \quad \text{s.t.} \quad (17)$$

$$\|\mathbf{x}\|_2 = 1, W_{ij} \in \{0, 1\}, \|\mathbf{w}_k\|_0 = 1 \text{ and } \|\mathbf{w}^m\|_0 \leq 1.$$

In which $\mathbf{W} \in \mathbb{R}^{K \times M}$ reduces the matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ by some selected rows. \mathbf{w}_k represents the k^{th} row of \mathbf{W} and \mathbf{w}^m indicates the m^{th} column of \mathbf{W} . The last constraint $\|\mathbf{w}^m\|_0 \leq 1$ avoids repetitive selection of the same row (sensor). This problem implies eigenvalue optimization over combination of rows of \mathbf{A} that it is shown to be NP-hard [34]. Accordingly, we propose a greedy algorithm to solve (17).

Algorithm 2 shows the steps of our proposed greedy algorithm to solve the obtained optimization problem. This algorithm optimizes the reduction matrix row-by-row where the reliability of the non-selected sensors are being considered. Assume the algorithm aims to select a new sensor at the k^{th} iteration. Up to current iteration, $k - 1$ sensors already are selected. The algorithm evaluate the non-selected sensors one-by-one in order to find the sensor that maximizes the objective function. The objective function is a weighted summation of the minimum eigenvalue of the restricted set of rows (sensors) and their corresponding unreliability weights. To evaluate each sensor we need to compute the most dominant k eigen components which implies performing singular value decomposition (SVD). However, truncated SVD up to the k^{th} component will be sufficient. An online algorithm is proposed that observes the non-selected sensors with a low sampling rate as depicted in Fig. 3. In each sequence, the observed set of sensors is updated as well as their corresponding reliability

Algorithm 2 Reliable E-optimal Sensor Selection

Require: A, S, K and r **Output:** The selected set \mathbb{S} and reduction matrix W .

Initialization: $W = \mathbf{0} \in \mathbb{R}^{K \times M}$ and $\mathbb{S} = \emptyset$
for $k = 1, \dots, K$ (Optimization of the k^{th} row of W)
 for $\forall m \in \mathbb{S}^c$
 SVD: $A(\mathbb{S} \cup m, :) = V^T \Lambda U$
 $\mathbf{x}^* = U(:, k)$
 $p(m) = \|\mathbf{A}\mathbf{x}^*\|_2^2 + \gamma u(m)$
 end
 $s_k = \text{argmax } p(m)$
 $\mathbb{S} = \mathbb{S} \cup s_k$ and $W_{k, s_k} = 1$
end for

weights. The first step to update the reliability is estimating the propagation using only the current active sensors. To this aim the following problem must be solved.

$$\mathbf{x}^{(t)}(\mathbb{S}) = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{W}^{(t)}(\mathbf{y} - \mathbf{A}\mathbf{x})\|_2^2 + \lambda_{LASSO} \|\mathbf{x}\|_1. \quad (18)$$

Here λ_{LASSO} regularizes sparsity and W indicates the reduction matrix to the selected set \mathbb{S} . Those sensors whose measurements are matched with the estimated power density map are marked as reliable. A consistent definition is proposed in (14) which considers the deviation of actual measurements from the estimation of the model as a metric for reliability. The subscript t is removed in Algorithm 2 for simplification. Algorithm 3 shows the overall process of spectrum sensing using the selected sensors.

The bottleneck of complexity order of Algorithm 2 at the k^{th} iteration is performing a truncated singular value decomposition to obtain the first k eigen components. Thus, the complexity of the algorithm in the k^{th} iteration will be $O(kMN^2)$ [35]. Therefore, selection of K sensors implies complexity order of $O(K^2MN^2)$.

Algorithm 3 Spectrum Sensing using Dynamic Sensor Selection

Require: A, S, K, λ, f_l and λ_{LASSO} .**Output:** Power spectrum for each time $\mathbf{x}^{(t)}$.

Initialization: \mathbb{S} = Output of Algorithm 1 and $\mathbf{x}(\mathbb{S})$ = Result of Problem (18)
2: for a new time block (t)
 sample $M \times f_l$ sensors
4: Update $t_m = t$ for the sensed sensors
 Update reliability using (14)
6: $\mathbb{S}^{(t)}$ = Output of Algorithm 2
 $\mathbf{x}^{(t)}(\mathbb{S}^{(t)})$ = Result of Problem (18)
8: end for

VI. EXPERIMENTAL RESULTS

The simulations are performed for collaborative spectrum sensing. The setup for generating data are employed from [22]. Our goal is to estimate vector \mathbf{x} that indicates transmitted spectrum power at some candidate points.

For the first simulation suppose we have potentially 300 sensors and they are estimating an $\mathbf{x} \in \mathbb{R}^{36}$ that has only 5

active transmitters. The location of sensors are derived from a uniform distribution and the active transmitters are selected randomly and the results are averaged for 200 different realizations. The following linear measurements are sensed by sensors $m = 1 \dots M$,

$$y_m = \mathbf{a}_m^T \mathbf{x} + v_m,$$

where, v_m indicates additive white Gaussian noise. a_{ms} shows the s^{th} entry of \mathbf{a}_m is the channel gain between the m^{th} sensor and the s^{th} potential source. The channel gain between two points is assumed by one over squared distance of two points. Since, the ability of sensors is considered the same over spectrum, thus the simulations are performed for a single spectrum band. The same procedure can be performed for multi-band spectrum regime independently. Figure 4 shows the performance of different static algorithms versus the number of selected sensors. Static refers to measurement-independent methods. In this experiment the SNR is set to +20dB. Successful recovery is defined as true estimation of the support of sparse vector using the measurements. Problem (18) is solved 200 for each algorithm. The Sparse solution is obtained using the iterative re-weighted least square algorithm [36]. As it can be seen in Fig. 4, E-optimal based sensor selection has the best performance.

Fig. 5 exhibits the effect of involving reliability on the static sensor selection. Suppose there are 300 potential sensors and the low-sampling rate is set equal to $\frac{1}{30}$. It means in each time block 10 new measurements contribute to construct the reliability weights (14). Observation of new measurements of one time block makes an improvement in normalized estimation error; similarly, usage of 5 time blocks significantly improves the performance to be close to the estimation after 30 time blocks in which all the sensors are observed. The forgetting factor is set to 0 as the state of network is not changed during observation of 30 time blocks. Thus, aggregating the measurements without the forgetting factor is optimum. The normalized error, $\|\mathbf{x}^* - \mathbf{x}(\mathbb{S})\|_2 / \|\mathbf{x}^*\|_2$, is defined as the criterion for performance, where, \mathbf{x}^* is the ground truth solution.

Fig. 6 visualizes the error of spectrum sensing in the area of network for the setup of Fig. 5. We are to choose 8 sensors.

Fig. 7 shows that the error of estimation is significantly decreased by setting $\gamma = 0.7$ for the setup of Fig. 5. However, an efficient value of γ depends on the problem setup and

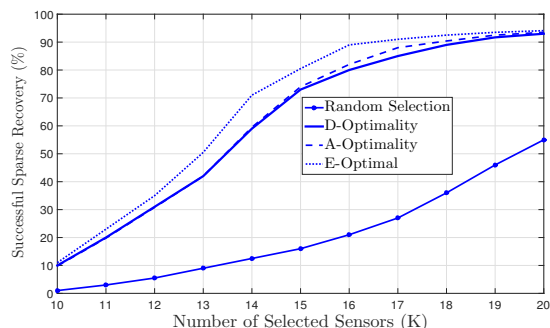


Figure 4: Performance of different static sensor selection algorithms in terms of number of selected sensors.

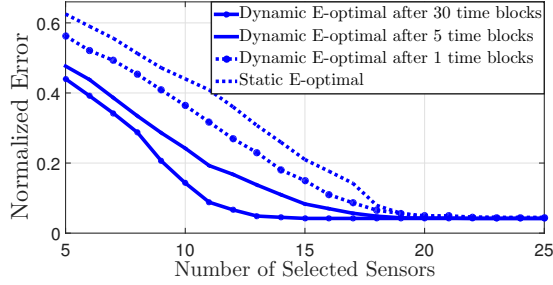


Figure 5: Performance of static and dynamic E-optimal-based sensor selection algorithms vs. the number of selected sensors.

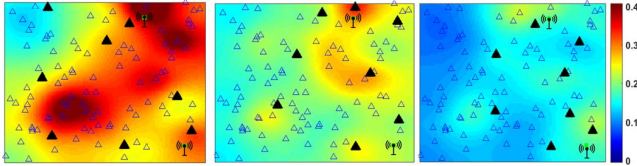


Figure 6: The error of estimated spectrum in the area of interest. (Left) E-optimal, Algorithm 1. (Middle) Reliable E-optimal, Algorithm 2 after sensing in one time block. (Right) Reliable E-optimal, Algorithm 2 while all the sensors are sensed after 30 time blocks. γ is assumed 0.7

should be tuned. Setting $\gamma = 0$ is equivalent to the static E-optimal sensor selection. Simulation shows the proposed reliable sensor selection performs better than the static sensor selection for a relatively wide range of γ , i.e., the problem is not very sensitive to well-tuning of this parameter.

In addition to power spectrum map, the proposed framework is able to generate a new network profile which can provides us trustworthy of the estimated spectrum for each point of the network. We call this side output *reliability map*. Interpolation of the estimated reliability of sensors throughout the network's

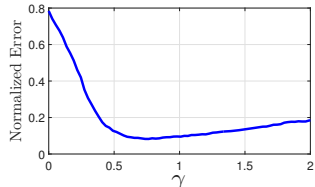


Figure 7: MSE error versus different values of γ .

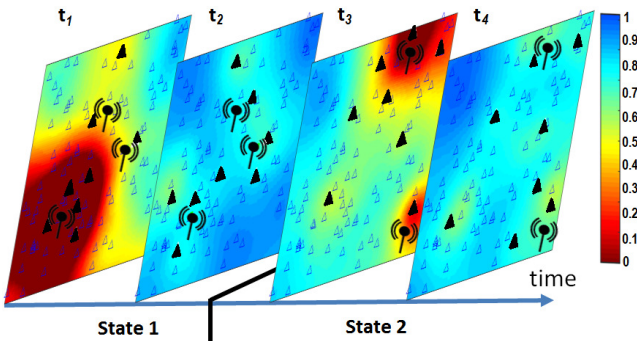


Figure 8: Reliability maps of 4 time blocks illustrate how the proposed framework evolves in time in order to select adapted sensors to the dynamic of network after state transition. Sensors within unreliable (red) areas have more chance of selection.

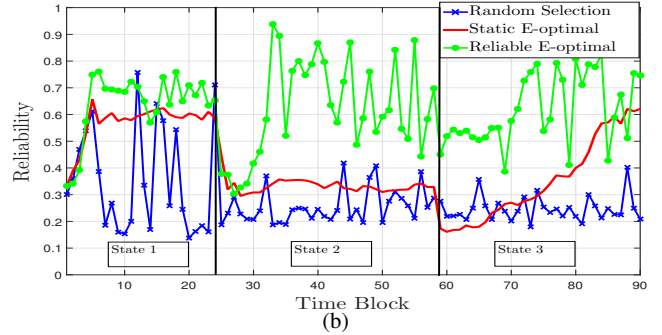
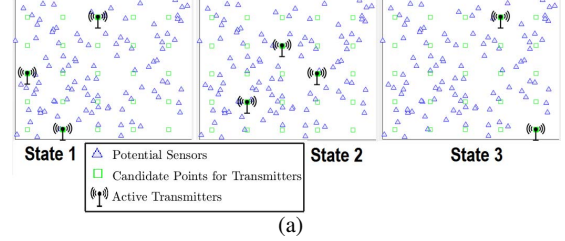


Figure 9: (a) A dynamic network with 3 states for the location of active PUs. The shaded blue squares represent active PUs. (b) The effect of reliable sensor selection for compensation of the model error in the reliable sensor selection procedure.

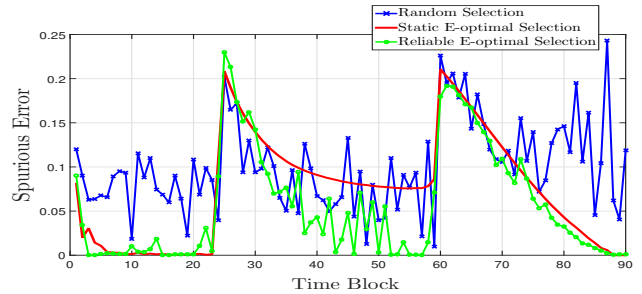


Figure 10: The effect of reliable sensor selection for compensation of the model error in the reliable sensor selection procedure.

area, generates the reliability map. Fig. 8 visualizes the temporal effect of dynamic sensor selection using the reliability map. Unreliable areas are indicated by red and blue areas represent reliable estimation of spectrum. Reliable sensor selection aims to compensate unreliability by considering more chance for red regions. In the next time slot the error for those regions are compensated. In this figure, each state of the network corresponds to a specific set of active PUs.

Fig. 9a shows the location of active PUs for a dynamic network with 3 states. There are 90 time blocks and the state of network is changed in blocks 24 and 59. The forgetting factor is set to $0.1/(\Delta T)$ in which ΔT is the time difference of two consecutive time blocks. Fig. 9b and Fig. 10 show the performance of sensor selection in terms of average network reliability and the spurious error of spectrum sensing which is defined by $\|\hat{\mathbf{x}}_{\text{spurious}}\|_1 = \sum_{i \notin \mathbf{x}^* \text{ support}} |x(\mathbb{S})_i|$. As it can be seen, the reliability is increased and the undesired power propagation is decreased by exploiting the dynamic framework, especially for the second state.

VII. CONCLUSION

The problem of sensor selection is considered and its relation to existing work on matrix subset selection is elaborated. We developed a new subset selection method as an extension of the well-known volume sampling. Our criteria is based on E-optimality, which is in favor of compressive sensing theory. We extended the static E-optimal sensor selection to a dynamic sensor selection method that exploits the measurements in an online manner. The experimental results indicate the efficiency of our suggested sensor selection algorithm in cognitive radio networks' spectrum sensing.

REFERENCES

- [1] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 451–462, Feb 2009.
- [2] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Decision and Control (CDC), 2010 49th IEEE Conf. on*, Dec 2010, pp. 2572–2577.
- [3] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
- [4] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013.
- [5] B. Amizic, L. Spinoulas, R. Molina, and A. Katsaggelos, "Compressive blind image deconvolution," *Image Processing, IEEE Trans. on*, vol. 22, no. 10, pp. 3994–4006, Oct 2013.
- [6] X. Guan, Y. Gao, J. Chang, and Z. Zhang, "Advances in theory of compressive sensing and applications in communication," in *Instrumentation, Measurement, Computer, Communication and Control, 2011 First International Conference on*, Oct 2011.
- [7] Y. Gwon, H. Kung, and D. Vlah, "Compressive sensing with optimal sparsifying basis and applications in spectrum sensing," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, Dec 2012, pp. 5386–5391.
- [8] T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *Information Theory, IEEE Transactions on*, vol. 57, no. 7, pp. 4680–4688, July 2011.
- [9] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ_0 norm," *Signal Processing, IEEE Transactions on*, vol. 57, no. 1, pp. 289–301, Jan 2009.
- [10] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM JOURNAL ON SCIENTIFIC COMPUTING*, vol. 20, pp. 33–61, 1998.
- [11] Q. Geng and J. Wright, "On the local correctness of ℓ_1 minimization for dictionary learning," in *Information Theory (ISIT), 2014 IEEE Symposium on*, June 2014, pp. 3180–3184.
- [12] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Transactions on Information Theory*, 2001.
- [13] A. Cohen, W. Dahmen, and R. Devore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, 2009.
- [14] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.
- [15] E. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.
- [16] E. Candes, "The restricted isometry property and its implications for compressed sensing," *C. R. Academie des Sciences*, no. 356, pp. 689–592, 2008.
- [17] M. Davies and R. Gribonval, "Restricted isometry constants where ℓ_p sparse recovery can fail for $0 < p \leq 1$," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, May 2009.
- [18] J. D. Blanchard, C. Cartis, and J. Tanner, "Compressed sensing: How sharp is the restricted isometry property?" *SIAM Rev.*, vol. 53, no. 1, pp. 105–125, Feb. 2011.
- [19] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127 – 2159, May 2006.
- [20] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *Signal Processing Magazine, IEEE*, vol. 24, no. 3, pp. 79–89, May 2007.
- [21] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.
- [22] J. Bazerque and G. Giannakis, "Distributed spectrum sensing for cognitive radio networks by exploiting sparsity," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1847–1862, March 2010.
- [23] E. Dall'Anese, J. A. Bazerque, and G. B. Giannakis, "Group sparse lasso for cognitive network sensing robust to model uncertainties and outliers," *Physical Communication*, vol. 5, no. 2, pp. 161 – 172, 2012, compressive Sensing in Communications.
- [24] Y. Zhao, L. Morales, J. Gaeddert, K. Bae, J.-S. Um, and J. Reed, "Applying radio environment maps to cognitive wireless regional area networks," in *New Frontiers in Dynamic Spectrum Access Networks, 2007. DySPAN 2007. 2nd IEEE International Symposium on*, April 2007, pp. 115–118.
- [25] A. Deshpande and L. Rademacher, "Efficient volume sampling for row/column subset selection," in *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*. IEEE, 2010, pp. 329–338.
- [26] A. Deshpande, L. Rademacher, S. Vempala, and G. Wang, "Matrix approximation and projective clustering via volume sampling," in *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, 2006, pp. 1117–1126.
- [27] M. Gu and S. C. Eisenstat, "Efficient algorithms for computing a strong rank-revealing qr factorization," *SIAM Journal on Scientific Computing*, vol. 17, no. 4, pp. 848–869, 1996.
- [28] A. Farahat, A. Elgohary, and A. Ghodsi, "Greedy column subset selection for large-scale data sets," *Knowledge and Information Systems*, vol. 45, no. 1, pp. 1–34, 2015.
- [29] C. Boutsidis, M. W. Mahoney, and P. Drineas, "An improved approximation algorithm for the column subset selection problem," in *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '09, 2009.
- [30] P. Van Dooren, "Numerical linear algebra for signal, systems and control," *Draft notes prepared for the Graduate School in Systems and Control*, vol. 250, 2003.
- [31] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [32] W. H. Haemers, "Interlacing eigenvalues and graphs," *Linear Algebra and its applications*, vol. 226, pp. 593–616, 1995.
- [33] C. C. Aggarwal, Y. Xie, and P. S. Yu, "On dynamic data-driven selection of sensor streams," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1226–1234.
- [34] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Trans. Inf. Theor.*, vol. 60, no. 2, pp. 1248–1259, Feb. 2014.
- [35] M. Holmes, A. Gray, and C. Isbell, "Fast svd for large-scale matrices," in *Workshop on Efficient Machine Learning at NIPS*, vol. 58, 2007, pp. 249–252.
- [36] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, March 2008, pp. 3869–3872.