A Matrix-Factorization-Error-Ratio Approach to Cooperative Sensing in Non-Ideal Communication Environment

Rui Zhou, Wenqiang Pu, Licheng Zhao, Ming-Yi You, Qingjiang Shi, and Sergios Theodoridis, *Life Fellow, IEEE*

***Abstract*—A fundamental challenge in cognitive radio is the detection of primary users in a licensed spectrum. Cooperative sensing, which utilizes multiple receivers distributed across different locations, offers the advantage of utilizing multiple antennas and achieving spatial diversity gain. However, successful implementation of cooperative sensing relies on the ideal exchange of information among cooperating receivers, which may not always be feasible in real-world scenarios. In this paper, we consider the cooperative sensing problem in a non-ideal communication scenario, where the raw data broadcasted from a receiving node can be received by only a subset of the nearby nodes. Existing multiantenna detectors can not deal with such a scenario. To tackle this issue, we propose a novel cooperative sensing scheme, where each node sends only its local correlation coefficients to the fusion center. A detection mechanism based on factorizing the partially received sample covariance matrix is developed. To achieve fast convergence and avoid exhaustive step size tuning, a Bregman proximal method, based on an alternating minimization algorithm (with convergence guarantees), is also developed. The advantages of our proposed cooperative scheme is demonstrated through numerical simulations.**

***Index Terms*—Cooperative sensing, matrix factorization, sample covariance matrix, missing entries.**

Manuscript received 25 September 2023; revised 28 April 2024 and 2 July 2024; accepted 6 August 2024. Date of publication 13 August 2024; date of current version 6 September 2024. This work was supported in part by the National Nature Science Foundation of China (NSFC) under Grant 62201362, Grant 62101350, and Grant 62206182; in part by Shenzhen Science and Technology Program under Grant RCBS20221008093126071; and in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515010154. The associate editor coordinating the review of this article and approving it for publication was Dr. Hamdi Joudeh. *(Corresponding author: Wenqiang Pu.)*

Rui Zhou, Wenqiang Pu, and Licheng Zhao are with Shenzhen Research Institute of Big Data, Shenzhen 518172, China (e-mail: rui.zhou@sribd.cn; wpu@sribd.cn; zhaolicheng@sribd.cn).

Ming-Yi You is with the National Key Laboratory of Electromagnetic Space Security, Jiaxing 314033, China (e-mail: youmingyi@126.com).

Qingjiang Shi is with the School of Software Engineering, Tongji University, Shanghai 200092, China, and also with Shenzhen Research Institute of Big Data, Shenzhen 518172, China (e-mail: shiqj@tongji.edu.cn).

Sergios Theodoridis is with the National and Kapodistrian University of Athens, 157 72 Athens, Greece and also with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark (e-mail: s.theodoridis@ieee.org).

Digital Object Identifier 10.1109/TSP.2024.3443291

I. Introduction

MOTIVATED by the spectrum scarcity and congestion caused by the exponential growth of wireless devices, cognitive radio (CR) has emerged as a key technology to enable more efficient and flexible spectrum utilization [1], [2], [3]. This technology is endowed with high spectrum efficiency and data transmission rates by means of exploiting the opportunistic spectrum access to other available networks. In a CR network, CR users are allowed to make use of the frequency band allocated to other primary users, when the latter are inactive [4]. The spectrum sensing technology is necessary for CR users to monitor the occupation status of the frequency band of interest. The cooperative sensing technique is able to make full use of the antennas that provide service to the distributed CR users, and it offers higher spatial diversity gain [5], [6], [7]. For example, a group of unmanned aerial vehicles (UAVs), each of which is equipped with an omnidirectional antenna, may apply the cooperative sensing technology to detect the primary signals.

Several data fusion approaches accomplish spectrum sensing by aggregating raw data at the fusion center. Let $\mathbf{x}_t$ (a p-dimensional complex column vector) represent the received signals of p antennas at time t. The simplest energy detector (ED) uses the energy of the received signals, i.e., $\sum_t \|\mathbf{x}_t\|^2$, to decide the presence of primary signals [8]. Note that, in this case, the noise power is assumed to be known a-priori at the CR receivers so that to keep a satisfactory detection performance [9]. Of course, there also exist numerous detection methods being able to detect signals without the need of prior knowledge of the noise variance. They are realized by taking advantage of the correlation structure in the received data. Given the eigenvalues $\{\lambda_i\}_{i=1}^p$ of the sample covariance matrix, $\mathbf{S} = \frac{1}{T}\sum_t \mathbf{x}_t\mathbf{x}_t^H$, where T is the number of samples, the eigenvalue arithmetic-to-geometric mean (AGM) detector calculates the statistic as the ratio of the eigenvalues' arithmetic mean to the geometric mean [10]. It is based on the fact that the eigenspectrum will spread out if the primary signal exists. Similar methods include the eigenvalue-moment-ratio (EMR) detector [9], the maximum-minimum eigenvalue (MME) detector [11], the scaled largest eigenvalue (SLE) detector [12], and the generalized likelihood ratio test (GLRT) detector [13]. In the cooperative sensing, all aforementioned methods are applicable under the assumption

that one has full access to the raw data that are collected from the distributed receiving nodes (CR users) and sent to the fusion center.

The successful transmission, however, of data from the distributed receiving nodes to a shared fusion center might be challenging for real-world applications [4], [9], [14]. First, the limited transmitting power, due to hardware constraints or security issues, can result in a restricted communication range, as reported in [15] and [16]. Second, high levels of electromagnetic interference in the environment can distort or disrupt the wireless signals [17]. Third, physical obstructions, such as buildings, trees, or other obstacles, can block the wireless signals and prevent successful data transmission [18]. Consequently, direct transmission of raw data from CR nodes to a fusion center is widely acknowledged as problematic [4].

To reduce data transmission, an alternative decision fusion method consists of two steps: First, each CR node detects the presence of the primary signal; second, it transmits a one-bit decision to the fusion center [19], [20]. The fusion center then applies a decision fusion rule to obtain the final decision. Commonly adopted decision fusion rules include the Logical-OR, Logical-AND, and Majority rules [19]. The Logical-OR rule posits that if a CR node detects the primary signal, the fusion center acknowledges its existence. Conversely, the Logical-AND rule requires all CR nodes to confirm the signal before the fusion center recognizes it. The Majority rule demands over half of the CR nodes to report the signal for it to be considered present [21]. More flexibly, a rule may assert the signal's presence when more than K ($1 \leq K \leq p$) CR nodes report it, which incorporates the aforementioned decision rules as special cases. Robust extensions of the decision fusion approach have been explored to accommodate fading environments and dynamic channel characteristics [22], [23]. While decision fusion approaches are efficient in terms of data transmission, their detection performance is notably inferior compared to data fusion methods.

In this paper, we propose a spectrum detecting scheme that is applicable in the aforementioned communication restricted scenario. The major contributions of this paper are as follows:

- We address the cooperative spectrum sensing challenge by introducing a scheme that enables each receiving node to broadcast its raw data to a subset of (possible to receive) nodes. These nodes then compute and send sample covariances of received and own data to the fusion center.
- We propose a novel detection method, which is applicable when the fusion center can only receive the sample covariance matrix with *missing entries*. The employed test statistic is the ratio of matrix factorization error under specific covariance matrix structures, i.e., with or without the primary signals.
- We also propose an efficient and provably convergent algorithm to solve the resulting optimization problem. This consists in approximating the target sample covariance matrix with missing elements via the sum of a low-rank matrix and a diagonal one.
- In addition, we present a practical solution based on the grid search method to address the asynchronism issue that is encountered when collecting raw data.

This paper is organized as follows. We first provide the signal model and our considered sensing scheme in Section II. In Section III, we introduce the proposed detection method and formulate the associated optimization problem. Section IV presents an efficient parameter-free algorithm that addresses the main difficulty of the optimization problem. In Section V, we discuss a practical approach to cope with the asynchronism of the raw data due to the propagation delay. Numerical experiments are given in Section VI. Finally, conclusions are summarized in Section VII.

II. Problem Formulation

A. Signal Model

Consider p distributed CR users, each of which is equipped with a single antenna, whose purpose is to cooperatively sense a frequency band of interest that is occupied by a primary user. The channels between the primary user and each one of the CR users are frequency non-selective (flat) fading and the rank of the respective signal subspace is assumed to be known and equal to r. Note that the spatial rank, r, that is considered here, may result from two cases: *i)* a single primary user with r antennas [13]; *ii)* r primary users with a single antenna [9], which, however, are assumed to be closely located, so that the respective propagation delays to a CR user are regarded almost the same. The time and frequency are assumed to be nearly synchronized to enable the sampling at distributed nodes [24], [25], [26]. There are two hypotheses, i.e., signal absent and signal present events that are denoted as $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Denote as $x_{i,t} \in \mathbb{C}$ the received signal of the i-th ($i = 1, \ldots, p$) receiver at time t ($t = 1, \ldots, T$), and $\mathbf{x}_t = [x_{1,t}, \ldots, x_{p,t}]^T$. The classical hypothesis testing problem is cast as [13]

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{x}_t = \mathbf{n}_t, \\ \mathcal{H}_1 &: \mathbf{x}_t = \mathbf{H}\mathbf{s}_t + \mathbf{n}_t, \end{aligned} \tag{1}$$

where $\mathbf{s}_t \in \mathbb{C}^r$ is the primary signal following the i.i.d. zero-mean circular complex Gaussian (CCG) distribution, $\mathbf{H} \in \mathbb{C}^{p \times r}$ is the unknown channel among the primary user and the receivers, and $\mathbf{n}_t \in \mathbb{C}^p$ is the i.i.d. zero-mean CCG and uncorrelated noise. The channel $\mathbf{H}$ is assumed to remain stable throughout the designated sampling period, which is often termed the channel coherence time [27]. But it may vary among different sampling periods. The hidden node issue [28], which pertains to certain nodes being unable to receive the primary signal due to blockage, is also accounted for in this model. This is accomplished by allowing certain rows in matrix $\mathbf{H}$ to contain zeros. Note that we currently ignore the differences of propagation delays from the primary user to the distributed CR users. Hypothesis testing model in (1) is still valid when one considers different propagation delays, but an extra delay compensation term is needed. Details will be further discussed in the Section V.

Since any spatial correlation and scaling of the primary signal can be absorbed in $\mathbf{H}$, we assume the covariance matrix of $\mathbf{s}$ to

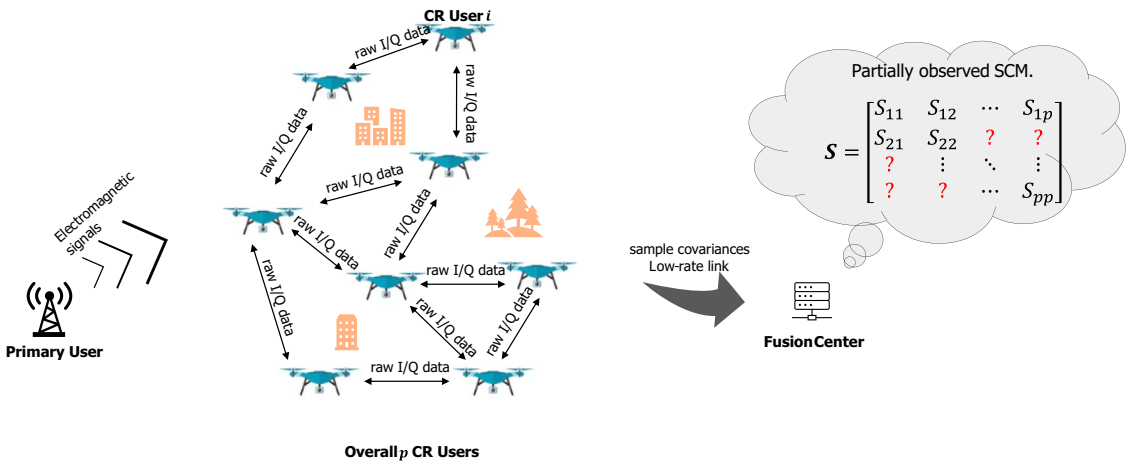


Fig. 1. Distributed detection scheme under a restricted communication distance scenario.

be identity, i.e., $\mathbb{E}\left(\mathbf{s}\mathbf{s}^H\right) = \mathbf{I}_p$. Denote the covariance matrix of the received signal $\mathbf{x}_t$ as $\boldsymbol{\Sigma}$, then the hypothesis testing Problem (1) can equivalently be written as

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{\Sigma} = \boldsymbol{\Psi}, \ \boldsymbol{\Psi} \in \mathcal{D}, \\ \mathcal{H}_1 &: \boldsymbol{\Sigma} = \mathbf{H}\mathbf{H}^H + \boldsymbol{\Psi}, \ \boldsymbol{\Psi} \in \mathcal{D}, \end{aligned} \tag{2}$$

where

$$\mathcal{D} = \{\boldsymbol{\Psi} = \mathrm{Diag}\,(\boldsymbol{\psi}) \mid \boldsymbol{\psi} \in \mathcal{B} := [\boldsymbol{\psi}_{\mathrm{L}}, \boldsymbol{\psi}_{\mathrm{U}}]\}\,, \tag{3}$$

with $\boldsymbol{\psi}_{\mathrm{L}}$ and $\boldsymbol{\psi}_{\mathrm{U}}$ being the predetermined lower and upper bounds of the noise power, respectively ($\mathbf{0} < \boldsymbol{\psi}_{\mathrm{L}} \leq \boldsymbol{\psi}_{\mathrm{U}}$). Various methods have been developed for Problem (2), e.g., EMR [9], MME [11], GLR [13]. It should be noted that all these methods require the raw data to be available, i.e., the matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]^T$ can be correctly formed at the fusion center. If this is not the case, these methods can not be employed since the detection mechanism is developed based on the knowledge of the sample covariance matrix, $\mathbf{S} = \frac{1}{T}\mathbf{X}\mathbf{X}^H$.

B. Cooperative Sensing Scheme

As shown in Fig. 1, the data transmission may fail for various practical reasons, i.e., data missing, decoding errors, etc. The raw data from the involved nodes can not always be successfully transmitted to the fusion center. The missing values of the raw data at the fusion center renders Problem (2) a difficult one. In our alternative scenario, we consider that the raw data are broadcasted and can be correctly received by a subset of the nodes. In this case, for two nearby nodes, say i and j, the respective correlation coefficient, S_{ij}, can be locally computed (c.f. (4)) and, subsequently, sent to the fusion center. Note that this is the only information received by the fusion center from nodes i and j. This significantly reduces the communication burden from the nodes to the fusion center. More precisely, in this proposed sensing scheme, the communication overhead for the fusion center to receive data is limited to a maximum of p^2 complex numbers. Leveraging the symmetry of the matrix $\mathbf{S}$, this overhead can be further reduced through strategic coordination of the CR nodes. For instance, node i need only transmit entries S_{ij} where $j \geq i$. Consequently, this tailored approach reduces the total communication overhead to a maximum of $p(1+p)/2$ complex numbers. On the other hand, for complete data transmission, the communication overhead amounts to $T \times p$ (where T is usually significantly greater than p) complex numbers. However, such a scheme faces a challenge. The fusion center may not "know" the correlation coefficients between mutually faraway nodes. In other words, matrix $\mathbf{S}$ in Fig. 1 involves missing entries. To our knowledge, this task has not been treated before in the literature.

In this paper, we consider the previous cooperative sensing scheme. Each node computes sample covariances with neighboring nodes only. Thus, the fusion center receives partial sample covariances from those nodes. More specifically, the computed covariances at the i-th node, are

$$S_{ij} = \frac{1}{T}\sum_{t=1}^{T} x_{i,t}x_{j,t}^*, \quad \forall j \in \mathcal{I}_i, i = 1, 2, \ldots, \tag{4}$$

where $\mathcal{I}_i$ is the index set of the receiving nodes whose data is available at the i-th node. These locally computed sample covariances are transmitted to the fusion center via the low-rate but reliable communication links, e.g., via satellite communications [29]. If each node can receive reliable data from all the rest, the sample covariance matrix is fully formed in the fusion center. However, since the nodes can only access nearby nodes, the fusion center can only obtain *a partially observed sample covariance matrix*. In the sequel, we first present a matrix factorization formulation for estimating $\boldsymbol{\Sigma}$. Then, a factorization-error-based detector is proposed.

III. FERT Detector

Formally, let $\Omega \subset [p] \times [p]$ denote the symmetric index set, where the pair $(i, j) \in \Omega$ represents that nodes i and j can receive each other's data. Then, the fusion center only receives sample covariance matrix entries that lie in Ω[1]. Our first goal in this paper is to estimate $\boldsymbol{\Sigma}$ from the partially observed sample

[1]Diagonal elements of $\mathbf{S}$ are always collected, i.e., $(i, i) \in \Omega, i = 1, \ldots, p$.

covariance matrix. In line with two hypotheses in (2), we propose the following matrix factorization formulation to estimate $\boldsymbol{\Sigma}$:

$$\hat{\boldsymbol{\Sigma}}_0 = \arg \min_{\boldsymbol{\Sigma}_0 = \boldsymbol{\Psi}, \boldsymbol{\Psi} \in \mathcal{D}} \left\| \mathcal{P}_\Omega \left(\mathbf{S} - \boldsymbol{\Sigma}_0 \right) \right\|_F^2, \tag{5a}$$

$$\hat{\boldsymbol{\Sigma}}_1 = \arg \min_{\boldsymbol{\Sigma}_1 = \mathbf{H}\mathbf{H}^H + \boldsymbol{\Psi}, \boldsymbol{\Psi} \in \mathcal{D}} \left\| \mathcal{P}_\Omega \left(\mathbf{S} - \boldsymbol{\Sigma}_1 \right) \right\|_F^2, \tag{5b}$$

where

$$\mathcal{P}_\Omega \left(\mathbf{Z} \right)_{ij} = \begin{cases} Z_{ij} & (i,j) \in \Omega, \\ 0 & (i,j) \notin \Omega. \end{cases} \tag{6}$$

The algorithm for solving the problems in (5) will be introduced later in Section III-A and Section IV. Next, we derive the matrix factorization error ratio test (FERT) detector based on $\hat{\boldsymbol{\Sigma}}_0$ and $\hat{\boldsymbol{\Sigma}}_1$. The idea is to directly construct the test statistic as the ratio of the matrix factorization error, i.e.,

$$\xi_{\text{FER}} = \frac{f\left(\mathbf{S} \mid \hat{\boldsymbol{\Sigma}}_0 \right)}{f\left(\mathbf{S} \mid \hat{\boldsymbol{\Sigma}}_1 \right)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma, \tag{7}$$

where the threshold γ is set to comply with the desired probability of false alarm and $f\left(\mathbf{S} \mid \boldsymbol{\Sigma} \right) := \left\| \mathcal{P}_\Omega \left(\mathbf{S} - \boldsymbol{\Sigma} \right) \right\|_F^2$ denotes the matrix approximation error. The FERT decides $\mathcal{H}_1$ if $\xi_{\text{FER}} > \gamma$ and vice-versa for $\mathcal{H}_0$. It should be noted that if only a limited number of sample covariance entries are observed, the FERT detector may become infeasible because the denominator in (7) could approach zero. In such scenarios, decision fusion-based cooperative sensing could alternatively be used. Theoretically, the threshold γ could be derived from the cumulative distribution function of ξ_{FER} under pure noise conditions for a given false alarm probability. Nevertheless, due to the intricate nonlinear transformation from data to ξ_{FER}, the threshold γ can be empirically assessed in cases where the absence of the signal is confirmed [9].

A. A Naive Alternating Scheme

Notice that problem (5a) is a projection problem, i.e., project onto a box constraint, which has the following closed-form solution

$$\hat{\boldsymbol{\Sigma}}_0 = \text{Diag}(\boldsymbol{\psi}^*), \ \boldsymbol{\psi}^* = \mathbb{P}_{\mathcal{B}} \left(\text{diag}(\mathbf{S}) \right), \tag{8}$$

where $\mathbb{P}_{\mathcal{B}}(\cdot)$ is the projection operator defined as

$$[\mathbb{P}_{\mathcal{B}}(\boldsymbol{s})]_i = \begin{cases} \boldsymbol{s}_i, [\boldsymbol{\psi}_{\text{L}}]_i \le \boldsymbol{s}_i \le [\boldsymbol{\psi}_{\text{U}}]_i, \\ [\boldsymbol{\psi}_{\text{U}}]_i, \boldsymbol{s}_i > [\boldsymbol{\psi}_{\text{U}}]_i, \\ \boldsymbol{s}_i < [\boldsymbol{\psi}_{\text{L}}]_i. \end{cases} [\boldsymbol{\psi}_{\text{L}}]_i, \tag{9}$$

However, $\hat{\boldsymbol{\Sigma}}_1$ in problem (5b) does not have a closed-form solution since the corresponding objective is a fourth-order polynomial (non-convex) function w.r.t. $\mathbf{H}$. In the following part, we introduce a vanilla gradient descent-type of alternating minimization scheme to obtain the estimate $\hat{\boldsymbol{\Sigma}}_1$. Later, we will show that this alternating minimization scheme faces a step size tuning issue that may restrict its practical usage. To further improve computational efficiency and avoid step size tuning (see discussion in Section IV), the Bregman proximal method is developed to update $\mathbf{H}$, and the details are presented in Section IV.

Algorithm 1 Gradient descent method for Problem (11).

1: Initialize $\mathbf{H}^0$ and η.
2: **for** $l = 0, 1, 2, \ldots$ **do**
3: Compute $\nabla f_H(\mathbf{H}^l)$ as in (13);
4: Update $\mathbf{H}^{l+1} = \mathbf{H}^l - \eta \nabla f(\mathbf{H}^l)$;
5: Terminate when converges;
6: **end for**
7: Return $\mathbf{H}$.

Let us, now, recall problem (5) for $\hat{\boldsymbol{\Sigma}}_1$ and reformulate it as

$$\begin{aligned} \min_{\mathbf{H}, \boldsymbol{\psi}} \quad & f(\mathbf{H}, \boldsymbol{\Psi}) := \left\| \mathcal{P}_\Omega \left(\mathbf{S} - \mathbf{H}\mathbf{H}^H - \boldsymbol{\Psi} \right) \right\|_F^2, \\ \text{s.t.} \quad & \mathbf{H} \in \mathbb{C}^{p \times r}, \ \ \boldsymbol{\Psi} = \text{Diag}\left(\boldsymbol{\psi} \right), \ \ \boldsymbol{\psi} \in \mathcal{B}. \end{aligned} \tag{10}$$

By simply partitioning the variables into two blocks, i.e., $\mathbf{H}$ and $\boldsymbol{\Psi}$, a simple alternating minimization scheme is as follows.

1) Update $\mathbf{H}$: Given a fixed $\boldsymbol{\Psi}$, sub-problem w.r.t. $\mathbf{H}$ is

$$\begin{aligned} \min_{\mathbf{H} \in \mathbb{C}^{p \times r}} f_H(\mathbf{H}) &:= \left\| \mathcal{P}_\Omega \left(\hat{\mathbf{S}} - \mathbf{H}\mathbf{H}^H \right) \right\|_F^2 \\ &:= \sum_{(i,j) \in \Omega} \left| \mathbf{h}_i^H \mathbf{h}_j - \hat{S}_{ij} \right|^2, \end{aligned} \tag{11}$$

where $\hat{\mathbf{S}} = \mathbf{S} - \boldsymbol{\Psi}$ and $\mathbf{h}_i$ is the i-th row of $\mathbf{H}$. This problem is typically a low-rank matrix factorization problem with missing data [30], [31], [32]. It can be solved via the vanilla gradient descent method, whose updating rule at the l-th iteration is

$$\mathbf{H}^{l+1} = \mathbf{H}^l - \eta \nabla f_H(\mathbf{H}^l), \tag{12}$$

where $\eta > 0$ is the step size and $\nabla f_H(\mathbf{H}) \in \mathbb{C}^{p \times r}$ is the gradient. The i-th row of $\nabla f_H(\mathbf{H})$ corresponds to the partial gradient w.r.t. $\mathbf{h}_i$ and is calculated as

$$\nabla_{\mathbf{h}_i} f_H(\mathbf{H}) = 4 \times \sum_{j:(i,j) \in \Omega} \left(\mathbf{h}_j^H \mathbf{h}_i - \hat{S}_{ij} \right) \mathbf{h}_j. \tag{13}$$

The vanilla gradient descent for solving Problem (11) is summarized in Algorithm 1.

2) Update $\boldsymbol{\Psi}$: Given a fixed $\mathbf{H}$, sub-problem w.r.t. $\boldsymbol{\psi}$ is

$$\min_{\boldsymbol{\psi}} \ \left\| \text{diag}\left(\tilde{\mathbf{S}} \right) - \boldsymbol{\psi} \right\|_2^2 \quad \text{s.t. } \boldsymbol{\psi} \in \mathcal{B}, \tag{14}$$

where $\tilde{\mathbf{S}} = \mathbf{S} - \mathbf{H}\mathbf{H}^H$. The same as (5a), the optimal $\boldsymbol{\psi}^\star$ is

$$\boldsymbol{\psi}^\star = \mathbb{P}_{\mathcal{B}} \left(\text{diag}(\tilde{\mathbf{S}}) \right). \tag{15}$$

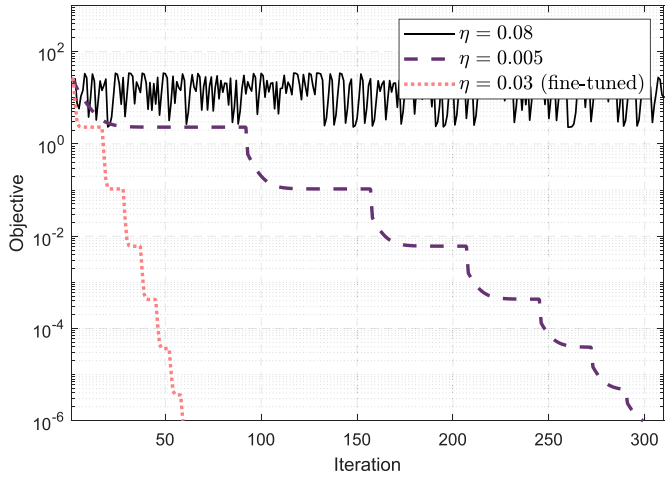
The proposed alternating minimization scheme is summarized in Algorithm 2. Note that Algorithm 2 is a double-loop algorithm, which has shown to be empirically and theoretically inefficient in numerous fields, e.g., [33], [34], [35]. To optimally capitalize on the prospects afforded by spectrum sensing, it is imperative to ensure that the process is executed with the utmost expediency. A practical implementation way is to execute Algorithm 1 for only a few iterations (or even one iteration), which

Algorithm 2 Alternating minimization for Problem (10).

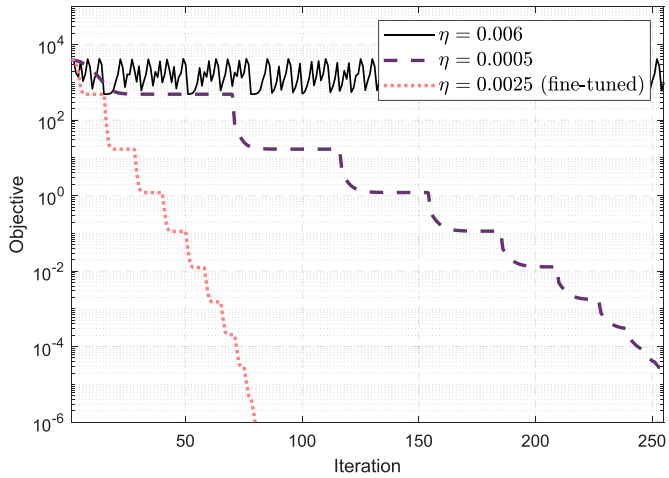
1: Initialize $\mathbf{H}^0$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Update $\mathbf{H}^{k+1}$ using Algorithm 1;
4: Update $\mathbf{\Psi}^{k+1}$ as in (15);
5: Terminate when converges;
6: **end for**
7: Return $\mathbf{H}$ and $\mathbf{\Psi}$.

can numerically achieve fast and stable convergence behavior if the step size η is properly tuned. However, for different sensing scenarios, e.g., different SNRs, configurations of nodes, and etc, Algorithm 1 requires case by case step size tuning, i.e., a bold choice of η may lead to fast convergence but can also cause the objective to diverge. On the other hand, the conservative choice of η may ensure stable convergence but at the cost of slower convergence. The step size tuning introduces issues of inefficiency and is contingent upon the fine-tuned step size. Furthermore, extensions of the classical gradient descent method, such as the line search strategy, are undesired due to their need for repeated evaluation of the objective functions.

To gain some insight on the step size tuning, let us first consider a toy realization of Problem (11), where $\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T$ with $\mathbf{x}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{h}\mathbf{h}^T + \mathrm{Diag}(\boldsymbol{\psi}))$, and $h_j \sim \mathcal{CN}(0,1)$, $\psi_j \sim \mathcal{U}(0,1), j = 1, \ldots, p$. Then, we generate another realization of Problem (11) using the similar setting except $h_j \sim \mathcal{CN}(0,10)$ and $\psi_j \sim \mathcal{U}(0,10)$. For both cases, three different choices of step size η in Algorithm 1 are used, i.e., $\eta = 0.005, 0.03, 0.08$ and $\eta = 0.0005, 0.0025, 0.006$, respectively. The changes of objective values versus iteration are shown in Fig. 2, where one iteration stands for a single computation of the gradient $\nabla f(\mathbf{H}^k)$. The curves show that the proper selection of η varies with different problem realizations.



(a) Case I: $h_j \sim \mathcal{CN}(0,1)$, $\psi_j \sim \mathcal{U}(0,1)$.



(b) Case II: $h_j \sim \mathcal{CN}(0,10)$, $\psi_j \sim \mathcal{U}(0,10)$.

Fig. 2. Example of convergence across distinct numerical scales, showcasing the potential variation in fine-tuned η.

IV. An Efficient Parameter-Free Algorithm

To address step size tuning issue, we propose a parameter-free and efficient algorithm based on the Bregman proximal gradient (BPG) framework. The detailed algorithm development and its convergence analysis are presented as below.

A. Algorithm Design

In this part, we will first present a brief introduction on the BPG method. Thereafter, based on the BPG method, a specific algorithm is developed to solve Problem (11).

Consider the general optimization problem

$$\min_{\mathbf{x}} \ f(\mathbf{x}) \quad \text{s.t. } \mathbf{x} \in \mathcal{X}, \tag{16}$$

where $f(\cdot)$ is a non-convex, continuously differentiable, and globally lower bounder objective function, and $\mathcal{X}$ is a convex set. The BPG-type methods solve this problem by iteratively conducing the following updating rule [36]

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}\in\mathcal{X}} \ f(\mathbf{x}^k) + \left\langle \nabla f(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^k \right\rangle + L \cdot D_\phi(\mathbf{x}, \mathbf{x}^k),$$

where $L > 0$ is a positive number, $D_\phi(\mathbf{x}, \mathbf{x}^k) = \phi(\mathbf{x}) - \phi(\mathbf{x}^k) - \left\langle \nabla\phi(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^k \right\rangle$ is the Bregman distance associated with the kernel generating distance function $\phi(\cdot)$, which is convex over $\mathcal{X}$. Several choices of $\phi(\cdot)$ have been investigated in the literature, e.g., [37], [38], [39].

The essential issue for utilizing the BPG-type method is that the function pair (f, ϕ) should enjoy the following property (also referred as relative smoothness), which can lead to a straightforward descent behavior of the BPG method, e.g., [40].

Definition 1 (Relative Smoothness): Function f is L-smooth relative to ϕ if there exists $0 < L < \infty$ such that

$$f(\mathbf{x}) \le f(\mathbf{y}) + \langle \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + L \cdot D_\phi(\mathbf{x}, \mathbf{y}), \ \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}.$$

Inspired by [39] and realizing that function $f_H(\mathbf{H})$ in Problem (11) is a fourth-order polynominal on the variables $\mathbf{H}$, we consider to take the following kernel generating distance function, which renders $f_H(\mathbf{H})$ to be relatively smooth, i.e.,

$$\phi(\mathbf{H}) = \frac{1}{4}\|\mathbf{H}\|_F^4 + \frac{1}{2}\|\mathbf{H}\|_F^2. \tag{17}$$

Algorithm 3 Bisection method for finding solution to (20).

Initialize $l = 0$, $u = 1/L$.
while not convergent **do**
  if $\rho(\frac{l+u}{2}) > 0$ **then**
    $u = (l+u)/2$;
  else
    $l = (l+u)/2$;
  end if
end while
Return $(l+u)/2$.

Intuitively, preserving the same polynomial order in $\phi(\mathbf{H})$ can leverage the curvature information inherent in $f_H(\mathbf{H})$. By doing so, the developed BPG iteration adapts to the curvature of $f_H(\mathbf{H})$. The following lemma shows that such a choice allows us to construct a strongly convex upper bound function of $f(\mathbf{H}, \mathbf{\Psi})$ w.r.t. $\mathbf{H}$.

Proposition 2: For any $\mathbf{\Psi} \in \mathcal{D}$, $f(\mathbf{H}, \mathbf{\Psi})$ w.r.t. $\mathbf{H}$ is L_H-smooth relative to $\phi(\mathbf{H})$, i.e., for any $\mathbf{H}$ and $\bar{\mathbf{H}}$, the following holds:

$$f(\mathbf{H}, \mathbf{\Psi}) \leq f(\bar{\mathbf{H}}, \mathbf{\Psi}) + \langle \nabla_H f(\bar{\mathbf{H}}, \mathbf{\Psi}), \mathbf{H} - \bar{\mathbf{H}} \rangle + L_H D_\phi(\mathbf{H}, \bar{\mathbf{H}}),$$

where $L_H = 3|\Omega| + p\left\|\boldsymbol{\psi}_U\right\|_\infty + \sum_{(i,j)\in\Omega} |S_{ij}|$.

Proof: See Appendix A. ■

Then, let us choose $L > L_H$ and ignore some constants, the BPG method for updating $\mathbf{H}^{k+1}$ becomes

$$\begin{aligned}\mathbf{H}^{k+1} &= \arg\min_{\mathbf{H}} \left\langle \nabla f_H(\mathbf{H}^k), \mathbf{H} - \mathbf{H}^k \right\rangle + L \cdot \phi(\mathbf{H}) \\ &\quad - L \cdot \left\langle \nabla \phi(\mathbf{H}^k), \mathbf{H} - \mathbf{H}^k \right\rangle \\ &= \arg\min_{\mathbf{H}} \left\langle \nabla f_H(\mathbf{H}^k) - L \cdot \nabla \phi(\mathbf{H}^k), \mathbf{H} \right\rangle \\ &\quad + \frac{L}{4}\left\|\mathbf{H}\right\|_F^4 + \frac{L}{2}\left\|\mathbf{H}\right\|_F^2. \end{aligned} \tag{18}$$

The following proposition shows that the optimal solution of Problem (18) enjoys a nice property, which leads to an equivalent form of Problem (18) that only contains a single optimization variable (see Problem (19)).

Proposition 3: The optimal solution $\mathbf{H}^\star$ to Problem (18) must be of the form: $-t\mathbf{A}^k$, where $t > 0$ is a positive number and $\mathbf{A}^k = \nabla f(\mathbf{H}^k) - L \cdot \nabla \phi(\mathbf{H}^k)$.

Proof: See Appendix B. ■

Proposition 3 suggests that solving Problem (18) is equivalent to solving the following one-dimensional task,

$$\min_{t\geq 0} \quad \frac{L}{4}\left\|\mathbf{A}^k\right\|_F^4 t^4 + \frac{L}{2}\left\|\mathbf{A}^k\right\|_F^2 t^2 - \left\|\mathbf{A}^k\right\|_F^2 t. \tag{19}$$

Taking the gradient to zero, we just need to find the solution to the following equation

$$\rho(t) = L\left\|\mathbf{A}^k\right\|_F^4 t^3 + L\left\|\mathbf{A}^k\right\|_F^2 t - \left\|\mathbf{A}^k\right\|_F^2 = 0. \tag{20}$$

Notice that $\rho(0) < 0$, $\rho(1/L) > 0$, and $\rho(t)$ is strictly increasing when $t > 0$. This implies that the unique solution to (20) can be easily found by the bisection method. The bisection method for finding the root $\rho(t) = 0$ is summarized in Algorithm 3.

Algorithm 4 BPG based algorithm for Problem (10).

Initialize $\mathbf{H}^0$.
for $k = 0, 1, 2, \ldots$ **do**
  Find t^k using Algorithm 3;
  Update $\mathbf{H}^{k+1} = -t^k(\nabla f(\mathbf{H}^k) - L \cdot \nabla \phi(\mathbf{H}^k))$;
  Update $\mathbf{\Psi}^{k+1}$ as in (15);
  Terminate when converges;
end for
Return $\mathbf{H}$ and $\mathbf{\Psi}$.

To reduce the computational burden per iteration, we propose to update $\mathbf{H}$ by solving Problem (18) once per iteration instead of trying to solve exactly Problem (11). The overall algorithm for solving Problem (10) is given in Algorithm 4.

Note that L_H in Proposition 2 is problem-dependent and can be used throughout the entire procedure for each problem-solving instance. In fact, L_H represents a relaxed lower bound for L that ensures the convergence of the algorithm. However, in practice, the convergence behavior of Algorithm 4 can often be improved by carefully selecting a smaller value for L, i.e., $L_H/2$ when a tuning-free point L_H has been identified.

B. Convergence Analysis

Note that Algorithm 4 is essentially a block coordinate descent (BCD) type algorithm consisting of two blocks, i.e., $\mathbf{H}$ and $\mathbf{\Psi}$. Block $\mathbf{H}$ is updated through a single iteration of the BPG method, while the other block, $\mathbf{\Psi}$, is minimized using a closed-form solution. Note that existing works on BPG [37], [38], [39], [40] primarily focus on the BPG itself and do not address such a mixed setting that we have explored in this work. Below, we provide the convergence analysis for Algorithm 4. We first define the stationary point associated with Problem (10), which is used to quantify the obtained solution.

Definition 4 (Stationary Point): Let $(\mathbf{H}^+, \mathbf{\Psi}^+)$ be the iterate generated by Algorithm 4 at $(\mathbf{H}, \mathbf{\Psi})$, if $(\mathbf{H}^+, \mathbf{\Psi}^+) = (\mathbf{H}, \mathbf{\Psi})$, then $(\mathbf{H}^+, \mathbf{\Psi}^+)$ is a stationary point of problem (10). Further, an ϵ-stationary solution ($\epsilon \geq 0$) is defined as $\|(\mathbf{H}^+, \mathbf{\Psi}^+) - (\mathbf{H}, \mathbf{\Psi})\|_F \leq \epsilon$.

Recall constant L_H in Proposition 2, which serves as a reference value for choosing L in Algorithm 4. Then, we provide a basic estimate of the descent behavior of Algorithm 4.

Lemma 5 (Descent Property): Suppose $L - L_H > 0$, then at iteration k, the following descent property holds:

$$\begin{aligned} &f(\mathbf{H}^{k+1}, \mathbf{\Psi}^{k+1}) \\ &\leq f(\mathbf{H}^k, \mathbf{\Psi}^k) - \alpha\|(\mathbf{H}^{k+1}, \mathbf{\Psi}^{k+1}) - (\mathbf{H}^k, \mathbf{\Psi}^k)\|_F^2, \end{aligned} \tag{21}$$

where $\alpha = \min\{L - L_H, 2\}$.

Proof: See Appendix C. ■

Based on Lemma 5, convergence of Algorithm 4 can be established as below.

Theorem 6: Let $\{(\mathbf{H}^k, \mathbf{\Psi}^k)\}$ be a sequence generated by Algorithm 4 and denote $F^k = f(\mathbf{H}^k, \mathbf{\Psi}^k)$. Then the following statements hold.

1) the function value sequence $\{F^k\}$ is nonincreasing;

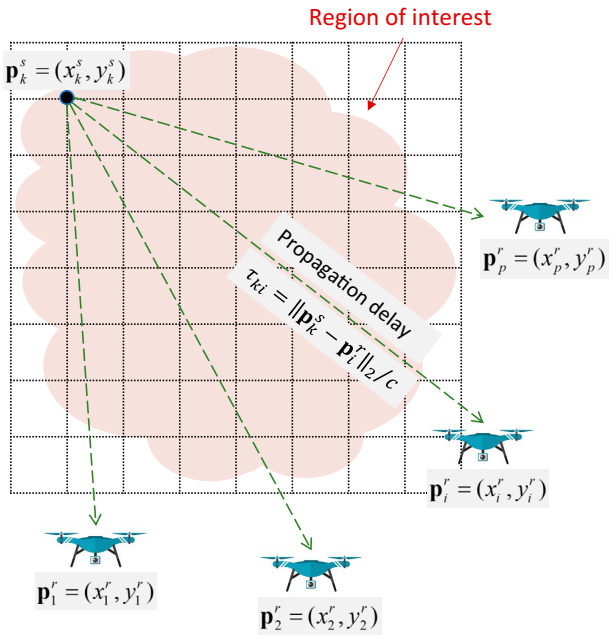


Fig. 3. An example of performing grid-search cooperative sensing over a 2-D region of interest.

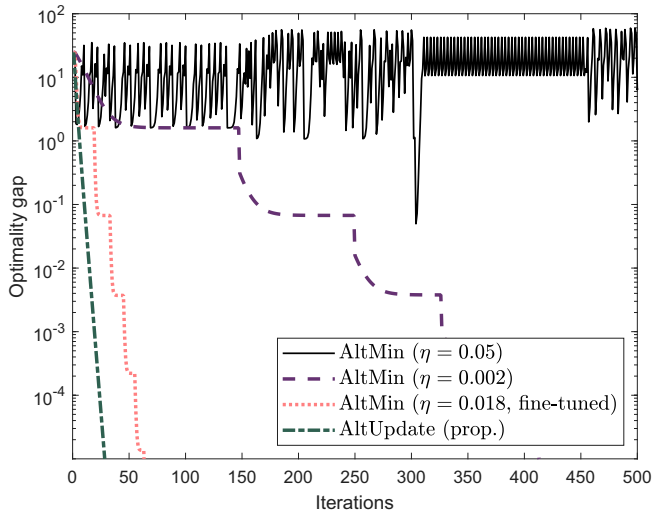


Fig. 4. Comparison on convergence of Algorithm 2 (AltMin) and Algorithm 4 (AltUpdate).

2) let $k \to \infty$, then $\|(\mathbf{H}^{k+1}, \mathbf{\Psi}^{k+1}) - (\mathbf{H}^k, \mathbf{\Psi}^k)\|_F^2 \to 0$;
3) within total K iterations, $\sqrt{C/K}$ stationary solution can be attained ($C = f(\mathbf{H}^0, \mathbf{\Psi}^0)/\alpha$)).

Proof: See Appendix E. ■

The first statement in Theorem 6 unveils a basic behavior of Algorithm 4. That is, the loss function is nonincreasing. The second statement indicates that Algorithm 4 converges to a stationary point. The last statement implies that the iteration complexity of Algorithm 4 is $\mathcal{O}(\epsilon^{-2})$.

V. Asynchronous Issue for Nodes Spreading Over a Large-Scale Region

In practical cooperative sensing scenarios, it is common for sensing nodes to be distributed over a significant large-scale region, with distances between nearby nodes spanning a few kilometers. Besides, the limited communication abilities among the distributed nodes, the propagation delays from the signal sources to the nodes can not be ignored, and this may lead to a model mismatch in Problem (2).

For instance, consider the case where only one signal source exists, i.e., $r = 1$, and the propagation delays from the signal source to nodes i and j are denoted as τ_i and τ_j, respectively. The naively obtained sample covariance S_{ij} actually admits the form

$$\mathbb{E}(S_{ij}) = h_i h_j^* \mathbb{E}(s_{t-\tau_1} s_{t-\tau_2}). \tag{22}$$

In theory, due to our initial assumption on the statistical properties of $\mathbf{s}$, $\mathbb{E}(s_{t-\tau_1} s_{t-\tau_2})$ equals to zero if $\tau_1 \neq \tau_2$, as stated in [41]. Therefore, it is necessary to align the received data before computing sample covariances.

To address the issue of asynchronicity, one intuitive approach is to align each pair of receiving data samples, prior to computing the associated sample covariances. There are three typical techniques for aligning time series data: those based on the proper time scale [42], the pulse per second (PPS) signals [43], and the correlation function [44] methods. The time scale based method is designed to guard against the problem of accidental omission of few data points. It is not applicable to the case considered in this paper. The pulse per second-based method requires the signal source to generate regular pulse signals, which is infeasible for non-cooperative primary users. The correlation-based method looks for the peak of the time shift correlation of two time series sequences, but it fails to produce reliable alignment results within the relatively low SNR scenes that are desirable for cooperative sensing. Therefore, the two-stage approach considered in this study may not be a suitable choice for addressing the asynchronous issue.

Motivated by the use of search grids in radar systems e.g., [45], [46], we propose a grid-search-based cooperative sensing scheme to address the unknown time delays among the involved distributed nodes. This approach is based on the fact that propagation delays can be accurately estimated once the physical locations of signal sources and receiving nodes are known. While determining the locations of non-cooperative sources may seem impossible, we can search the region of interest in a grid-like manner. As shown in Fig. 3, we assume that the primary user is located at the k-th grid point with position coordinates $\mathbf{p}_k^s = (x_k^s, y_k^s)$, and the i-th receiving node is located at coordinates $\mathbf{p}_i^r = (x_i^r, y_i^r)$ for $i = 1, \ldots, p$. We can estimate the propagation delays from the potential signal source at the k-th grid point to the i-th receiving node as $\tau_{ki} = \|\mathbf{p}_k^s - \mathbf{p}_i^r\| / c$, where c is the speed of the electromagnetic wave. We can then adjust the time sequences to align them, and calculate the test statistic ξ_{FER} in (7) using the aligned time sequence at each grid point. The highest ξ_{FER} value among all grid points may be taken as the test statistic for the region of interest.

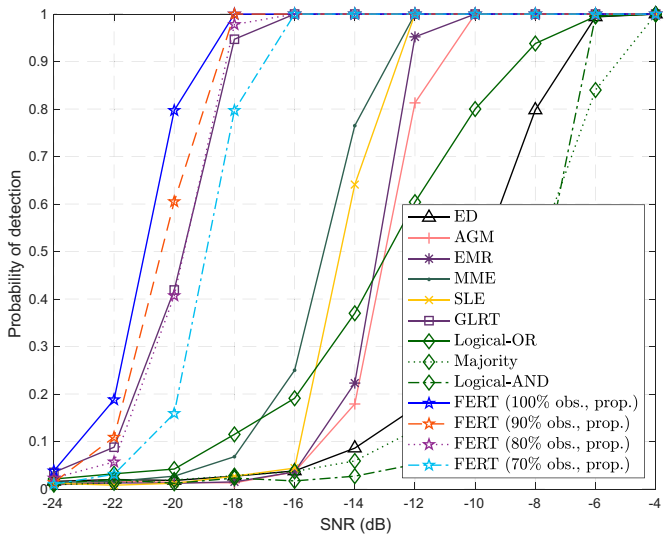


Fig. 5. Probability of detection with synchronized raw data by several detection methods.

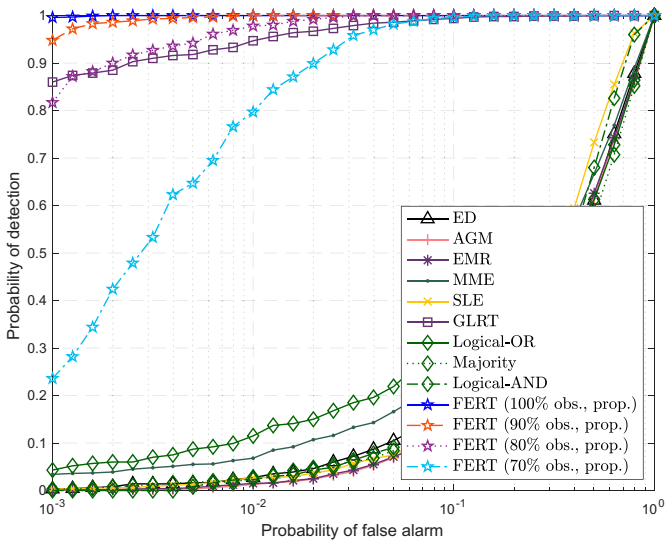


Fig. 6. ROC curve with synchronized raw data by several detection methods at a SNR of -18dB.

To perform the proposed FERT detection for the potential primal user at the k-th grid point, for the i-th receiving node, the calculation of the sample covariance associated with the k-th grid point is calculated using a time shift factor, given by

$$S_{ij}(\tilde{\tau}_{ij,k}) = \begin{cases} \frac{1}{T-\tilde{\tau}_{ij,k}} \sum_{t=1}^{T-\tilde{\tau}_{ij,k}} x_{i,t+\tilde{\tau}_{ij,k}} x^*_{j,t}, & \tilde{\tau}_{ij,k} \geq 0, \\ \frac{1}{T+\tilde{\tau}_{ij,k}} \sum_{t=1}^{T+\tilde{\tau}_{ij,k}} x_{i,t} x^*_{j,t-\tilde{\tau}_{ij,k}}, & \tilde{\tau}_{ij,k} < 0, \end{cases} \tag{23}$$

where $\tilde{\tau}_{ij,k}$ is the difference of time delays, i.e., $\tau_{ki} - \tau_{kj}$. Since $\tilde{\tau}_{ij,k}$ is discrete and finite, we simply require the i-th node to calculate sample covariances for a few time shifts, i.e., the possible sample covariances between node i and nodes j are $S_{ij}(\tilde{\tau}), \tilde{\tau} = 0, \pm 1, \pm 2, \ldots, \delta_{ij}$, where $\delta_{ij} = \max_k |\tilde{\tau}_{ij,k}|$.

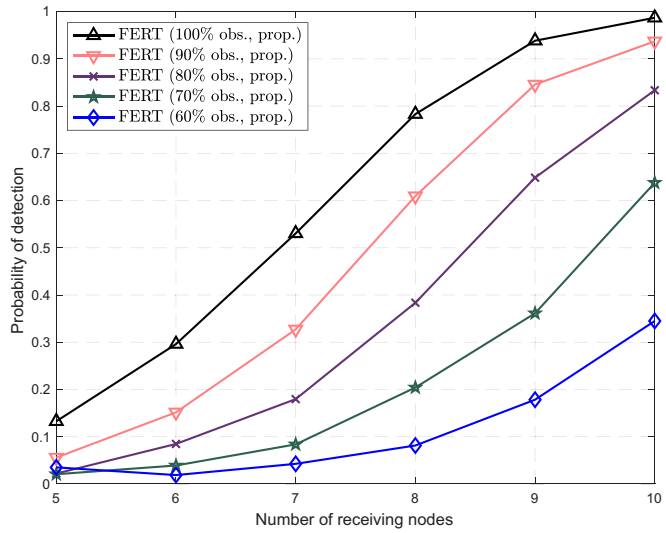


Fig. 7. Probability of detection of our proposed FERT detector versus number of receiving nodes at a SNR of -20dB.

VI. Numerical Experiments

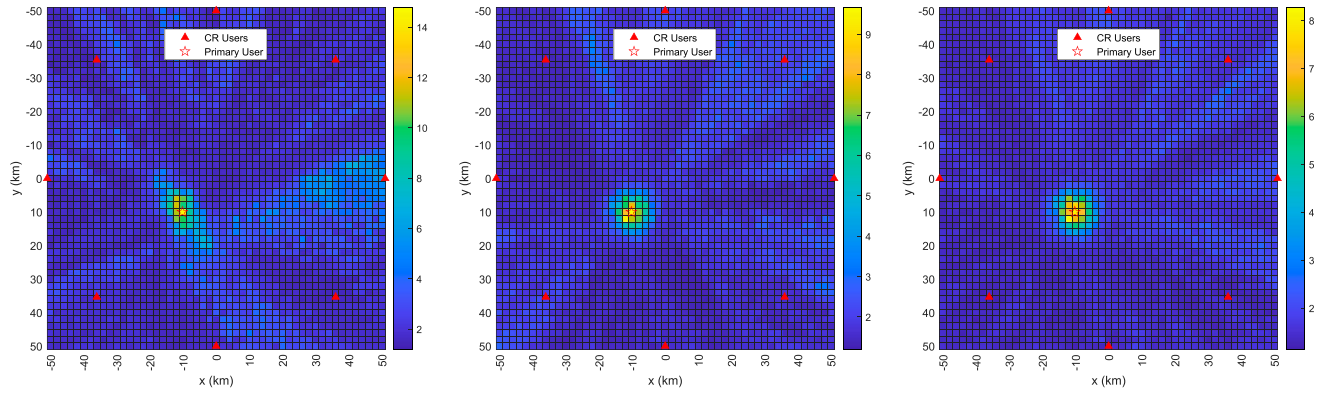
In this section, we validate the performance of our proposed cooperative detector through numerical experiments.

We consider a scenario with $p = 8$ distributed CR users, each of them is equipped with a single omnidirectional antenna. The task is to detect a single-antenna primary signal source, i.e., $r = 1$. The primary signal is carrying a QPSK modulated signals with the baud rate being equal to 20kHz. The sampling rate at each receiver is 100kHz. The channel between each receiver and the source is assumed to be an independent Rician fading channel[2]. The noise power of the CR user i is $\hat{\sigma}_i = \alpha$, $i = 1, \ldots, p$, where α is given in dB and it is uniformally distributed in an interval $[-1, 1]$ dB [10]. Since Algorithm 4 can adapt to different scenarios without tuning L, L is fixed to be L_H for all scenarios. The received signal-to-noise ratio (SNR) is set to be the same for each antenna. Each simulation experiment is conducted using 10000 consecutive samples, which corresponds to 100 milliseconds duration time. For the purpose of assessing the probability of false alarm, we configure the primary source to be inactive, corresponding to $\mathcal{H}_0$ status. Conversely, we activate the primary source, corresponding to $\mathcal{H}_1$ status, to evaluate the probability of detection.

A. Performance of Algorithms

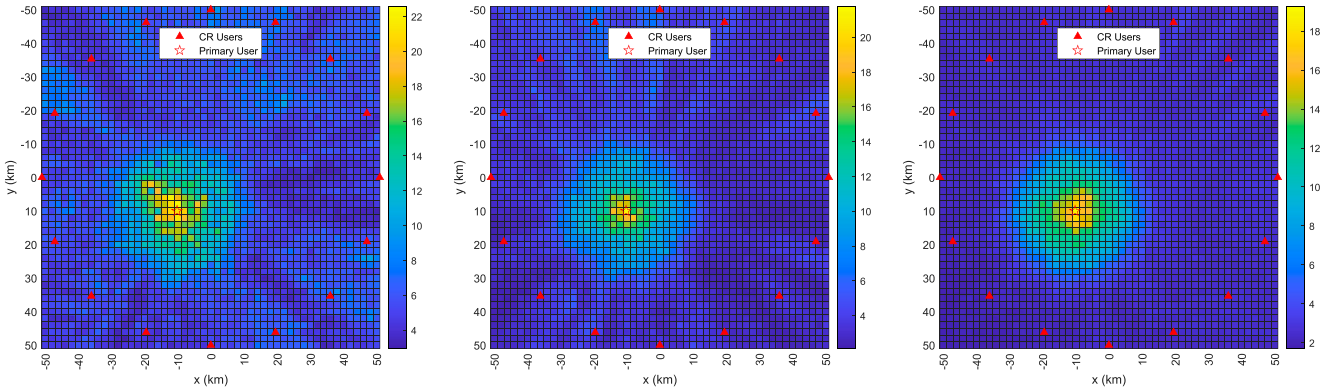
We first illustrate the convergence behavior of our proposed Algorithm 4 for solving problem (10). To evaluate the generalization of our proposed algorithm, we randomly set 50% of the elements of $\mathbf{S}$ in the fusion center to be missing. The initial point of our algorithm is $\mathbf{H}^0$ with $H^0_{ij} \sim \mathcal{CN}(0, \theta)$, where θ is the median of the off-diagonal elements of $|\mathbf{S}|$. The algorithm terminates when the relative changes of the objective and the variables of problem (10) are less than 10^{-5}.

[2]The modulated signals are generated using the MATLAB Communications Toolbox [47].



(a) 50% off-diagonal elements in $\mathbf{S}$ is collected. (b) 70% off-diagonal elements in $\mathbf{S}$ is collected. (c) 100% off-diagonal elements in $\mathbf{S}$ is collected.

Fig. 8. Examples of obtained test statistics map using our proposed FERT detector ($p = 8$, $r = 1$, average SNR is -17dB).



(a) 50% off-diagonal elements in $\mathbf{S}$ is collected. (b) 70% off-diagonal elements in $\mathbf{S}$ is collected. (c) 100% off-diagonal elements in $\mathbf{S}$ is collected.

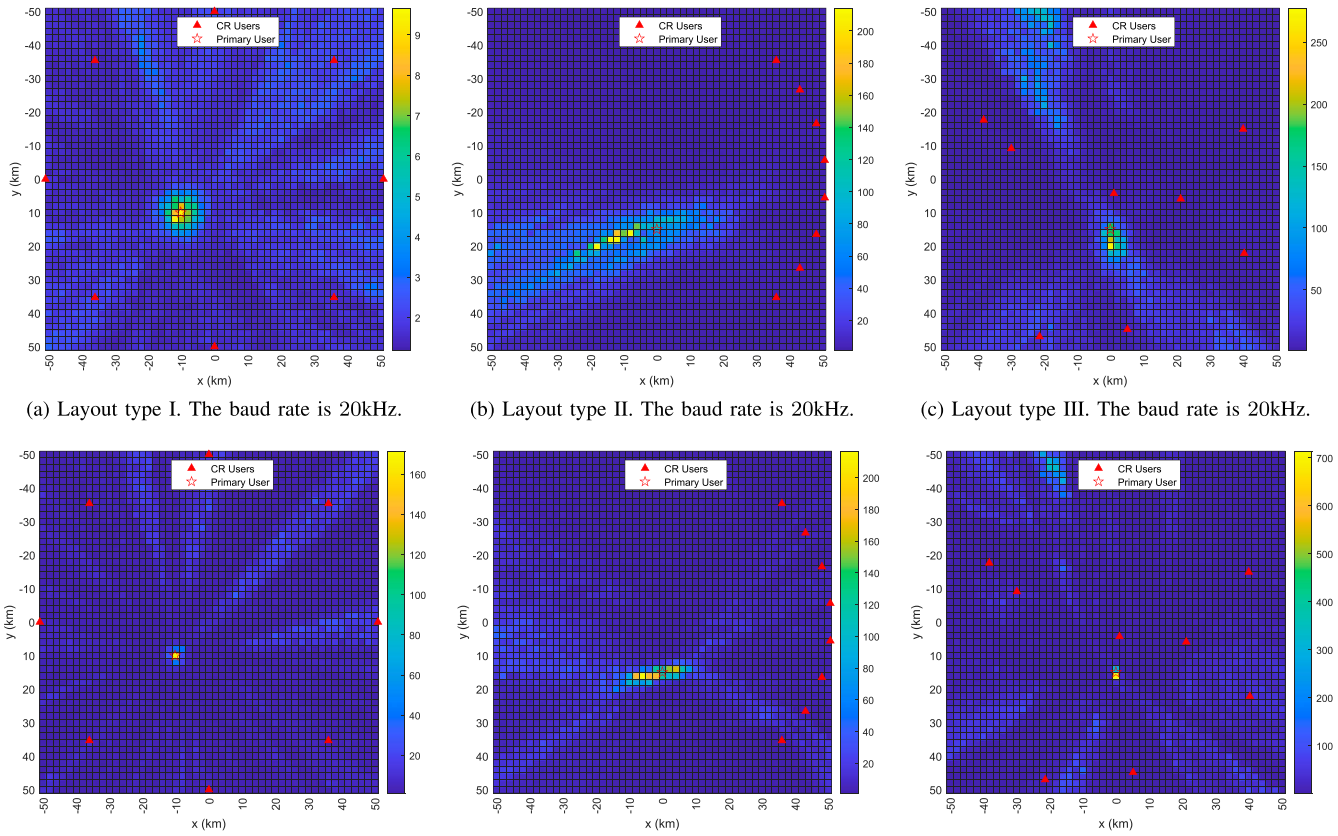
Fig. 9. Examples of obtained test statistics map using our proposed FERT detector ($p = 16$, $r = 3$, average SNR is -17dB).

Fig. 4 shows an example of convergence curves of Algorithm 4 and our proposed Algorithm 4 for solving Problem (10), where one iteration stands for a single computation of the gradient, i.e., $\nabla f(\mathbf{H}^k)$. Similar to the results illustrated in Fig. 2, the choice of step size η is crucial for the convergence speed of the classical GD method. By selecting the appropriate η, the classical GD method can achieve fast convergence. Notably, our proposed Algorithm 4 achieves even faster convergence speed compared to the best case of the GD method without requiring any tuning of extra parameters. Besides, we notice that our proposed Algorithm 4 exhibits superior performance in almost all problem settings, mainly due to the benefits of single-loop iteration and adaptive update of matrix $\mathbf{H}$.

B. Detection Performance

In this section, we demonstrate the effectiveness of our proposed FERT detector using a simplified scenario where the propagation delays from the primary user to the CR users are identical, and the SNR is kept the same for all nodes. Various communication-restricted scenarios are analyzed by randomly removing partial off-diagonal elements of $\mathbf{S}$ at the fusion center. For comparison, we also compare it with several data fusion benchmark methods fed with fully observed data and decision fusion benchmark methods. In the decision fusion benchmarks, all CR nodes employ the ED method with a uniform threshold to obtain local detection results. The test threshold, γ, is chosen to achieve probability of false alarm $P_{FA} = 1\%$.

The simulation results comparing the probability of detection with varying SNR levels are presented in Fig. 5. It is clear that, among all the data fusion benchmark methods, the energy detection method performs the worst and the GLRT method performs the best. This can be explained since the GLRT method is able to cope with the unequal noise variances in different receiving nodes. The decision fusion benchmark methods generally underperform relative to data fusion benchmarks due to their limited cooperation, which occurs solely at the decision level. Our proposed FERT method can achieve similar performance as that of GLRT method if fed with the fully observed $\mathbf{S}$. The small improvement in the performance of the proposed FERT method compared to the GLRT method might be due to the non-convex properties of the related optimization problems. Notably, the non-convex low-rank matrix factorization problem, i.e., Problem (10), is generally more amenable to finding global optima through conventional optimization algorithms [48]. This characteristic, however, is absent in the low-rank matrix maximum likelihood estimation problem central to the



(a) Layout type I. The baud rate is 20kHz. (b) Layout type II. The baud rate is 20kHz. (c) Layout type III. The baud rate is 20kHz.

(d) Layout type I. The baud rate is 60kHz. (e) Layout type II. The baud rate is 60kHz. (f) Layout type III. The baud rate is 60kHz.

Fig. 10. Comparison of obtained test statistics map using our proposed FERT detector under different layout and baud rate settings ($p = 8$, $r = 1$, 70% off-diagonal elements in $\mathbf{S}$ is collected). The three types of layouts can be distinguished as follows: *i)* a circular distribution of CR users surrounding the primary user; *ii)* a sectorial distribution of CR users adjacent to one side of the primary user; *iii)* a random distribution of CR users in the vicinity of the primary user.

GLRT detector [13], [49]. Moreover, note that even when matrix $\mathbf{S}$ is fed with 80% of the entries, the FERT detector is capable of achieving comparable performance to that of the GLRT detector. However, as the number of missing off-diagonal elements in $\mathbf{S}$ increases, the detection performance of the FERT method deteriorates. Note that all the existing benchmark methods require full knowledge of the data. Therefore, it is significant that, even when fed with a partially observed sample covariance matrix, our proposed FERT method can also achieve performance better than all benchmarks except the GLRT method (which, however, requires the full covariance matrix).

Fig. 6 depicts the receiver operating characteristic (ROC) curve, comparing the probability of detection to the probability of false alarm at a SNR of -18dB. Remarkably, our proposed FERT method demonstrates detection performance comparable to the GLRT method, even when utilizing only a 80% observed sample covariance matrix. When provided with a 70% observed sample covariance matrix, the FERT method surpasses other benchmarks except the GLRT method. Fig. 7 illustrates the detection performance of the proposed FERT detector as a function of the number of receiving nodes at a SNR of -20dB and a false alarm probability of 1%. Notably, detection efficacy improves with an increasing number of receiving nodes, while it diminishes when the number of missing data entries increases.

C. Robustness to Asynchronicity

In the sequel, we evaluate the performance of our proposed detection method in a more practical scenario, where time delays and signal attenuation are left to vary. Specifically, the time delay of the raw data received by the i-th node is equivalent to the time taken by the electromagnetic wave to travel from the primary user to the i-th cognitive radio (CR) user. Moreover, the strength of the signal received by the i-th node is inversely proportional to the square of its distance from the primary user [50].

To address the time asynchronous issue, we employ the approach discussed in Section V. The resulting test statistics ξ_{FER} over the grid are presented in heat maps, as shown in Figs. 8 and 9. From both figures, it is evident that the test statistics exhibit significant values around the physical location of the primary user, with a decrease as the distance from the primary user increases. Furthermore, we observe that the test statistic map appears more noisy as a greater number of off-diagonal elements of matrix $\mathbf{S}$ are missing. Nevertheless, even when half of the off-diagonal elements are missing, it remains possible to identify the presence of the signal source. As an extra bonus, the obtained test statistic map appears to be capable of identifying the location of the primary user.

We further investigate the test statistic map by altering the layout of the CR and primary users, as well as the baud rate, and comparing the resultant test statistic maps as depicted in Fig. 10. All other parameters remain consistent with those in Fig. 8. Upon examining Fig. 10(a)–10(c), it becomes apparent that the relative positioning of CR users and the primary user significantly influences the clarity of the test statistic maps. Additionally, a larger baud rate can markedly enhance this clarity as shown in Fig. 10(d)–10(f). However, as this observation is beyond the primary scope of this paper, we will defer a more detailed analysis to future research.

VII. Conclusion

This paper has addressed the cooperative sensing problem where the complete raw data collection at the fusion center is not possible. We have proposed a cooperative sensing scheme, in which each node transmits to the fusion center only the sample covariances between its own accessible received data and its neighboring nodes. A matrix factorization error ratio test method has been proposed to detect the presence of a signal. A provably convergent alternating update based algorithm has also been proposed to solve the matrix factorization problem with missing entries in the target matrix. Additionally, we have proposed a practical grid search approach to handle the asynchronism issue in the received raw data. Numerical experiments demonstrate the efficiency of our algorithm and the effectiveness of our cooperative sensing scheme.

Appendix

A. Proof of Proposition 2

Recall $f(\mathbf{H},\mathbf{\Psi})$ in problem (11), it straightforward to reformulated it as

$$f(\mathbf{H},\mathbf{\Psi}) := g(\mathbf{h}) = \sum_{(i,j)\in\Omega} |\mathbf{h}^H\mathbf{A}_{ij}\mathbf{h} - \hat{S}_{ij}|^2,$$

where $\mathbf{h} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_p]^H$, $\mathbf{A}_{ij} = \bar{\mathbf{A}}_{ij} \otimes \mathbf{I}_r$ ($\bar{\mathbf{A}}_{ij}$ is a $p \times p$ matrix with zero entries except the (i,j) entry equal to 1). By a straightforward linear algebra calculation, we have

$$\nabla^2 g(\mathbf{h}) = \sum_{(i,j)\in\Omega} (2\mathbf{A}_{ij}\mathbf{h}\mathbf{h}^H\mathbf{A}_{ij} + (\mathbf{h}^H\mathbf{A}_{ij}\mathbf{h} - \hat{S}_{ij})\mathbf{A}_{ij}),$$
$$\nabla^2\phi(\mathbf{h}) = (\|\mathbf{h}\|^2 + 1)\mathbf{I}_{pr} + 2\mathbf{h}\mathbf{h}^H.$$

Since $\lambda_{\max}(\mathbf{A}_{ij}) = 1, \forall (i,j) \in \Omega$, we have

$$\lambda_{\max}(\nabla^2 g(\mathbf{h})) \le \sum_{(i,j)\in\Omega} (3\|\mathbf{h}\|^2 + |\hat{S}_{ij}|) \le 3|\Omega|\|\mathbf{h}\|^2 + p\left\|\boldsymbol{\sigma}_u^2\right\|_\infty + \sum_{(i,j)\in\Omega} |S_{ij}|$$

and $\lambda_{\min}(\nabla^2\phi(\mathbf{h})) \ge \|\mathbf{h}\|^2 + 1$. Setting $L_H = 3|\Omega| + p\left\|\boldsymbol{\psi}_{\mathrm{U}}\right\|_\infty + \sum_{(i,j)\in\Omega}|S_{ij}|$ yields $L_H\nabla^2\phi(\mathbf{h}) \ge \nabla^2 g(\mathbf{h})$. Using the definition of relative smoothness completes the proof.

B. Proof of Proposition 3

Assuming there exist a optimal solution $\tilde{\mathbf{H}}$ minimizing Problem (18), we can always set $\mathbf{H}^\star = -(\left\|\tilde{\mathbf{H}}\right\|_F / \left\|\mathbf{A}^k\right\|_F)\mathbf{A}^k$. Since $\langle \mathbf{A}^k, \mathbf{H}^\star\rangle \le \left\langle \mathbf{A}^k, \tilde{\mathbf{H}}\right\rangle$ and $\frac{L}{4}\|\mathbf{H}^\star\|_F^4 + \frac{L}{2}\|\mathbf{H}^\star\|_F^2 = \frac{L}{4}\left\|\tilde{\mathbf{H}}\right\|_F^4 + \frac{L}{2}\left\|\tilde{\mathbf{H}}\right\|_F^2$, the objective is non-increasing. Since Problem (18) is bounded from below, there exist an optimal solution $\tilde{\mathbf{H}}$. Let f^* be the minimal objective value and set $\mathbf{H}^\star = -(\left\|\tilde{\mathbf{H}}\right\|_F / \left\|\mathbf{A}^k\right\|_F)\mathbf{A}^k$. Since $\langle \mathbf{A}^k, \mathbf{H}^\star\rangle \le \left\langle \mathbf{A}^k, \tilde{\mathbf{H}}\right\rangle$ and $\frac{L}{4}\|\mathbf{H}^\star\|_F^4 + \frac{L}{2}\|\mathbf{H}^\star\|_F^2 = \frac{L}{4}\left\|\tilde{\mathbf{H}}\right\|_F^4 + \frac{L}{2}\left\|\tilde{\mathbf{H}}\right\|_F^2$, f^* is no less than the objective w.r.t. $\mathbf{H}^\star$. This implies $\mathbf{H}^\star = \tilde{\mathbf{H}}$ which completes the proof.

C. Proof of Lemma 4

Denote $\iota(\mathbf{\Psi})$ as the indicator function of the constraint set of $\mathbf{\Psi}$ and suppose $(\mathbf{H}^*, \mathbf{\Psi}^*)$ is a stationary point of Problem 4, then $(\mathbf{H}^*, \mathbf{\Psi}^*)$ satisfies

$$\nabla_H f(\mathbf{H}^*, \mathbf{\Psi}^*) = \mathbf{0}, \quad \nabla_\Psi f(\mathbf{H}^*, \mathbf{\Psi}^*) + \mathbf{v}^* = \mathbf{0}$$

where $\mathbf{v}^* \in \partial\iota(\mathbf{\Psi}^*)$ is the subgradient. Recall Algorithm 4, the optimality condition of $(\mathbf{H}^+, \mathbf{\Psi}^+)$ is

$$\nabla_H f(\mathbf{H},\mathbf{\Psi}) + L(\nabla\phi(\mathbf{H}^+) - \nabla\phi(\mathbf{H})) = \mathbf{0},$$
$$\nabla_\Psi f(\mathbf{H}^+, \mathbf{\Psi}^+) + \mathbf{v}^+ = \mathbf{0}$$

where $\mathbf{v}^+ \in \partial\iota(\mathbf{\Psi}^+)$. Since $(\mathbf{H}^+, \mathbf{\Psi}^+) = (\mathbf{H}, \mathbf{\Psi})$, we immediately have $(\mathbf{H}, \mathbf{\Psi})$ satisfying the definition of stationarity.

D. Proof of Lemma 5

By the definition of relative smoothness, we have

$$f(\mathbf{H}^{t+1}, \mathbf{\Psi}^t) \le f(\mathbf{H}^t, \mathbf{\Psi}^t) + \langle \nabla_H f(\mathbf{H}^t, \mathbf{\Psi}), \mathbf{H}^{t+1} - \mathbf{H}^t\rangle + L_H D_\phi(\mathbf{H}^{t+1}, \bar{\mathbf{H}}^t)$$

Recall the first-order optimality condition of $\mathbf{H}^{t+1}$ that

$$\nabla_H f(\mathbf{H}^{t+1}, \mathbf{\Psi}^t) + L(\nabla\phi(\mathbf{H}^{t+1}) - \nabla\phi(\mathbf{H}^t)) = \mathbf{0}.$$

Combine the above two (in)equalities, we have

$$\begin{aligned} &f(\mathbf{H}^{t+1}, \mathbf{\Psi}^t) \\ &\le f(\mathbf{H}^t, \mathbf{\Psi}^t) + \langle \nabla_H f(\mathbf{H}^t, \mathbf{\Psi}), \mathbf{H}^{t+1} - \mathbf{H}^t\rangle \\ &\quad + L_H D_\phi(\mathbf{H}^{t+1}, \bar{\mathbf{H}}^t) \\ &= f(\mathbf{H}^t, \mathbf{\Psi}^t) + L\langle \nabla\phi(\mathbf{H}^t) - \nabla\phi(\mathbf{H}^{t+1}), \mathbf{H}^{t+1} - \mathbf{H}^t\rangle \\ &\quad + L_H D_\phi(\mathbf{H}^{t+1}, \bar{\mathbf{H}}^t) \\ &= f(\mathbf{H}^t, \mathbf{\Psi}^t) - L(D_\phi(\mathbf{H}^{t+1}, \mathbf{H}^t) + D_\phi(\mathbf{H}^t, \mathbf{H}^{t+1})) \\ &\quad + L_H D_\phi(\mathbf{H}^{t+1}, \bar{\mathbf{H}}^t) \\ &= f(\mathbf{H}^t, \mathbf{\Psi}^t) - (2L - L_H) D_\phi(\mathbf{H}^{t+1}, \bar{\mathbf{H}}^t) \end{aligned}$$

where $D_\phi(\mathbf{H}^{t+1}, \mathbf{H}^t) = D_\phi(\mathbf{H}^t, \mathbf{H}^{t+1})$ is used. Since $f(\mathbf{H}, \mathbf{\Psi})$ is strongly convex w.r.t. $\mathbf{\Psi}$ with modulus $\alpha_\Psi = 2$, it is straightforward to obtain the following inequality by making use of the optimality of $\mathbf{\Psi}^{t+1}$,

$$f(\mathbf{H}^{t+1}, \mathbf{\Psi}^{t+1}) \le f(\mathbf{H}^{t+1}, \mathbf{\Psi}^t) - \alpha_\Psi\|\mathbf{\Psi}^{t+1} - \mathbf{\Psi}^t\|_F^2.$$

Setting $\alpha_H = L - L_H \le 2L - L_H$ and summing up the two inequality completes the proof.

E. Proof of Theorem 6

Sum up (21) from $t=0$ to $t=T$, we have

$$\sum_{t=0}^{T} \alpha_H D_\phi(\mathbf{H}^{t+1},\mathbf{H}^t) + \alpha_\Psi \|\mathbf{\Psi}^{t+1}-\mathbf{\Psi}^t\|_F^2 \le f(\mathbf{H}^0,\mathbf{\Psi}^0) - f^*$$

Let $\alpha = \min\{\alpha_H, 2\}$, we have

$$\min_{0\le t\le T} D_\phi(\mathbf{H}^{t+1},\mathbf{H}^t) + \|\mathbf{\Psi}^{t+1}-\mathbf{\Psi}^t\|_F^2 \le \frac{f(\mathbf{H}^0,\mathbf{\Psi}^0)-f^*}{\alpha(T+1)}$$

Since $\|\mathbf{H}^{t+1}-\mathbf{H}^t\|_F^2 \le D_\phi(\mathbf{H}^{t+1},\mathbf{H}^t)$ and $f^* \ge 0$, we have

$$\|(\mathbf{H}^{t+1},\mathbf{\Psi}^{t+1}) - (\mathbf{H}^t,\mathbf{\Psi}^t)\|_F \le \sqrt{f(\mathbf{H}^0,\mathbf{\Psi}^0)/(\alpha K)}.$$

Let $t\to\infty$, the above implies

$$\|\mathbf{H}^{t+1}-\mathbf{H}^t\|_F^2 \to 0, \quad \|\mathbf{\Psi}^{t+1}-\mathbf{\Psi}^t\|_F^2 \to 0.$$

Denote $(\mathbf{H}^+,\mathbf{\Psi}^+)$ as the iterate generated by Algorithm 4 at $(\mathbf{H},\mathbf{\Psi})$ and let $\mathcal{F}(\mathbf{H},\mathbf{\Psi}) = (\mathbf{H}^+,\mathbf{\Psi}^+)$ as the corresponding mapping. It is easy to verify that $\mathcal{F}(\cdot)$ is a continuous function. Note that if $(\mathbf{H}^*,\mathbf{\Psi}^*) = \mathcal{F}(\mathbf{H}^*,\mathbf{\Psi}^*)$, then $(\mathbf{H}^*,\mathbf{\Psi}^*)$ is a stationary point. Let $(\mathbf{H}^*,\mathbf{\Psi}^*)$ be a limit point that sequence $\{\mathbf{H}^t,\mathbf{\Psi}^t\}$ converges to along some subsequence $\mathcal{T}$. Then we have

$$\begin{aligned}
&\|(\mathbf{H}^*,\mathbf{\Psi}^*) - \mathcal{F}(\mathbf{H}^*,\mathbf{\Psi}^*)\|_F \\
&= \lim_{t\to\infty, t\in\mathcal{T}} \|(\mathbf{H}^t,\mathbf{\Psi}^t) - \mathcal{F}(\mathbf{H}^t,\mathbf{\Psi}^t)\|_F \\
&= \lim_{t\to\infty, t\in\mathcal{T}} \|(\mathbf{H}^{t+1},\mathbf{\Psi}^{t+1}) - \mathcal{F}(\mathbf{H}^t,\mathbf{\Psi}^t)\|_F = 0
\end{aligned}$$

This implies $(\mathbf{H}^*,\mathbf{\Psi}^*)$ is a stationary point.

REFERENCES

[1] B. Wang and K. R. Liu, "Advances in cognitive radio networks: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 5–23, Feb. 2010.

[2] Y. Zou, Y.-D. Yao, and B. Zheng, "Cooperative relay techniques for cognitive radio systems: Spectrum sensing and secondary user transmissions," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 98–103, Apr. 2012.

[3] M. Wasilewska, H. Bogucka, and H. V. Poor, "Secure federated learning for cognitive radio sensing," *IEEE Commun. Mag.*, vol. 61, no. 3, pp. 68–73, Mar. 2023.

[4] Y. Zeng, Y.-C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: Challenges and solutions," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–15, Jan. 2010.

[5] S. M. Mishra, A. Sahai, and R. W. Brodersen, "Cooperative sensing among cognitive radios," in *Proc. IEEE Int. Conf. Commun.*, vol. 4. Piscataway, NJ, USA: IEEE Press, 2006, pp. 1658–1663.

[6] J. Oksanen, J. Lundén, and V. Koivunen, "Characterization of spatial diversity in cooperative spectrum sensing," in *Proc. 4th Int. Symp. Commun., Control Signal Process. (ISCCSP)*, 2010, pp. 1–5.

[7] J. Lunden, V. Koivunen, and H. V. Poor, "Spectrum exploration and exploitation for cognitive radio: Recent advances," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 123–140, May 2015.

[8] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, Apr. 1967.

[9] L. Huang, J. Fang, K. Liu, H. C. So, and H. Li, "An eigenvalue-moment-ratio approach to blind spectrum sensing for cognitive radio under sample-starving environment," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3465–3480, Aug. 2015.

[10] R. Zhang, T. J. Lim, Y.-C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A GLRT approach," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 84–88, Jan. 2010.

[11] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1784–1793, Jun. 2009.

[12] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814–823, Feb. 2010.

[13] D. Ramirez, G. Vazquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria, "Detection of rank-p signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764–3774, Aug. 2011.

[14] M. Park and H. Oh, "Cooperative information-driven source search and estimation for multiple agents," *Information Fusion*, vol. 54, pp. 72–84, Feb. 2020.

[15] Y. Huang, W. Mei, J. Xu, L. Qiu, and R. Zhang, "Cognitive UAV communication via joint maneuver and power control," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7872–7888, Nov. 2019.

[16] B. Duo, Q. Wu, X. Yuan, and R. Zhang, "Energy efficiency maximization for full-duplex UAV secrecy communication," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 4590–4595, Apr. 2020.

[17] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[18] R. Meireles, M. Boban, P. Steenkiste, O. Tonguz, and J. Barros, "Experimental study on the impact of vehicular obstructions in VANETs," in *Proc. IEEE Veh. Netw. Conf.*, 2010, pp. 338–345.

[19] P. K. Varshney, *Distributed Detection and Data Fusion*. New York, NY, USA: Springer-Verlag, 2012.

[20] S. Atapattu, C. Tellambura, and H. Jiang, "Energy detection based cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1232–1241, Apr. 2011.

[21] E. C. Y. Peh, Y.-C. Liang, Y. L. Guan, and Y. Zeng, "Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 5294–5299, Nov. 2009.

[22] A. Ghasemi and E. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *1st IEEE Int. Symp. New Frontiers Dynamic Spectr. Access Netw. (DySPAN)*, 2005, pp. 131–136.

[23] V. Balaji, "Reinforcement learning based decision fusion scheme for cooperative spectrum sensing in cognitive radios," in *Proc. Int. Conf. Wireless Commun., Signal Process. Netw. (WiSPNET)*, 2018, pp. 1–5.

[24] D. W. Allan, N. Ashby, and C. C. Hodge, *The Science of Timekeeping*. Palo Alto, CA, USA: Hewlett-Packard, 1997.

[25] S. Mghabghab, H. Ouassal, and J. A. Nanzer, "Wireless frequency synchronization for coherent distributed antenna arrays," in *Proc. IEEE Int. Symp. Antennas Propag. USNC-URSI Radio Sci. Meeting*, New York, NY, USA: Springer-Verlag, 2019, pp. 1575–1576.

[26] M. Rashid and J. A. Nanzer, "High accuracy distributed Kalman filtering for synchronizing frequency and phase in distributed phased arrays," *IEEE Signal Process. Lett.*, vol. 30, pp. 688–692, 2023.

[27] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[28] D. Janu, S. Kumar, and K. Singh, "A graph convolution network based adaptive cooperative spectrum sensing in cognitive radio network," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2269–2279, Feb. 2023.

[29] O. Kodheli et al., "Satellite communications in the new space era: A survey and future challenges," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 1, pp. 70–109, 1st Quart. 2021.

[30] S. Burer and R. D. Monteiro, "A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization," *Math. Program.*, vol. 95, no. 2, pp. 329–357, Feb. 2003.

[31] R. Ge, C. Jin, and Y. Zheng, "No spurious local minima in nonconvex low rank problems: A unified geometric analysis," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2017, pp. 1233–1242.

[32] J. Chen and X. Li, "Model-free nonconvex matrix completion: Local minima analysis and applications in memory-efficient kernel pca." *J. Mach. Learn. Res.*, vol. 20, no. 142, pp. 1–39, 2019.

[33] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*. New York, NY, USA: Springer-Verlag, 2012.

[34] R. G. Downey and M. R. Fellows, *Fundamentals of Parameterized Complexity*. New York, NY, USA: Springer-Verlag, 2013, vol. 4.

[35] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*. Cambridge, MA, USA: MIT Press, 2022.

[36] J. Bolte, S. Sabach, M. Teboulle, and Y. Vaisbourd, "First order methods beyond convexity and Lipschitz gradient continuity with applications to quadratic inverse problems," *SIAM J. Optim.*, vol. 28, no. 3, pp. 2131–2151, 2018.

[37] H. Lu, R. M. Freund, and Y. Nesterov, "Relatively smooth convex optimization by first-order methods, and applications," *SIAM J. Optim.*, vol. 28, no. 1, pp. 333–354, 2018.

[38] T. Gao, S. Lu, J. Liu, and C. Chu, "Leveraging two reference functions in block Bregman proximal gradient descent for non-convex and non-Lipschitz problems," 2019, *arXiv:1912.07527*.
[39] Q. Li, Z. Zhu, G. Tang, and M. B. Wakin, "Provable Bregman-divergence based methods for nonconvex and non-Lipschitz problems," 2019, *arXiv:1904.097129*.
[40] H. H. Bauschke, J. Bolte, and M. Teboulle, "A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications," *Math. Oper. Res.*, vol. 42, no. 2, pp. 330–348, 2017.
[41] A. F. Molisch, *Wireless Communications.* Hoboken, NJ, USA: Wiley, 2012.
[42] M. Rhudy, "Time alignment techniques for experimental sensor data," *Int. J. Comput. Sci. Eng. Surv.*, vol. 5, no. 2, 2014, Art. no. 1.
[43] M. Vollmer, D. Bläsing, and L. Kaderali, "Alignment of multi-sensored data: Adjustment of sampling frequencies and time shifts," in *Proc. Comput. Cardiol. (CinC)*, Piscataway, NJ, USA: IEEE Press, 2019, pp. 1–4.
[44] Y. Yang et al., "Time synchronization algorithm for the skiing monitoring system," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.
[45] Y. Briheche and F. Barbaresco, "Impact of grid quantification in radar search pattern optimization," in *Proc. Int. Radar Conf. (RADAR)*, 2019, pp. 1–6.
[46] M. I. Skolnik, *Radar Handbook*. New York, NY, USA: McGraw-Hill, 2008.
[47] *Communications Toolbox* (Version 7.8). The MathWorks, Inc., Natick, MA, USA, Sep. 2022. [Online]. Available: https://www.mathworks.com/help/comm/
[48] R. Sun and Z.-Q. Luo, "Guaranteed matrix completion via non-convex factorization," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6535–6579, Nov. 2016.
[49] K. Khamaru and R. Mazumder, "Computation of the maximum likelihood estimator in low-rank factor analysis," *Math. Program.*, vol. 176, no. 1, pp. 279–310, 2019.
[50] A. Ryer, *Light Measurement Handbook*, Peabody, MA, USA: International Light Technologies, 1997, pp. 25–28.

Rui Zhou received the B.Eng. degree in information engineering from the Southeast University, Nanjing, China, in 2017, and the Ph.D. degree from The Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2021. Currently, he is a Research Scientist with Shenzhen Research Institute of Big Data, Shenzhen, China. His research interests include optimization algorithms, statistical signal processing, machine learning, and financial engineering.

Wenqiang Pu received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2013 and 2018, respectively. From 2019 to 2020, he was a Postdoctoral Associate with the School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen). Currently, he is a Research Scientist with Shenzhen Research Institute of Big Data. His research interests include signal processing and optimization algorithms. His co-authored paper received the Best Student Paper Award from IEEE SAM 2024. He serves as an Associate Editor of IEEE SIGNAL PROCESSING LETTERS.

Licheng Zhao received the B.S. degree in information engineering from the Southeast University (SEU), Nanjing, China, in 2014, and the Ph.D. degree with the Department of Electronic and Computer Engineering with The Hong Kong University of Science and Technology (HKUST), in 2018. Since 2018, he has been an Algorithm Engineer in recommendation system with the JD.COM, China. Since 2021, he has served as a Research Scientist with Shenzhen Research Institute of Big Data (SRIBD). His research interests include optimization theory and efficient algorithms, with applications in signal processing, machine learning, and deep learning in recommendation system.

Ming-Yi You received the B.S. and Ph.D. degrees in mechanical engineering from Shanghai Jiao Tong University, Shanghai, in 2006 and 2012, respectively. He was invited to visit M. S. Wu Manufacturing Research Center, University of Michigan, from 2007 to 2008. Currently, he is a Senior Expert with the No. 36 Research Institute of CETC and leading a research group in radio direction finding and localization. He has published over 50 papers and co-authored the book titled *Radio Direction Finding: Theory and Practice*. His research interests include condition-based maintenance, radio direction finding, wireless localization, and multi-object tracking.

Qingjiang Shi received the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2011. From 2009 to 2010, he visited Prof. Z.-Q. (Tom) Luo's Research Group, University of Minnesota, Twin Cities. In 2011, he worked as a Research Scientist with Bell Labs, China. In 2012, he was with the School of Information and Science Technology, Zhejiang Sci-Tech University. From 2016 to 2017, he worked as a Research Fellow with Iowa State University, USA. Since 2018, he has been with the School of Software Engineering, Tongji University, where currently he is a Full Professor. He is also with Shenzhen Research Institute of Big Data. His research interests include algorithm design and analysis with applications in machine learning, signal processing, and wireless networks. So far, he has published more than 80 IEEE Journals and filed about 40 National Patents. He was an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the recipient of the IEEE Signal Processing Society Best Paper Award in 2022, the Huawei Technical Cooperation Achievement Transformation Award (2nd Prize) in 2022, the Huawei Outstanding Technical Achievement Award in 2021, the Golden Medal at the 46th International Exhibition of Inventions of Geneva in 2018, the First Prize of Science and Technology Award from China Institute of Communications in 2017, the National Excellent Doctoral Dissertation Nomination Award in 2013, the Shanghai Excellent Doctoral Dissertation Award in 2012, and the Best Paper Award from the IEEE PIMRC'09 conference.

Sergios Theodoridis (Life Fellow, IEEE) received the D.Sc. (Hons.) doctoral degree from the University of Edinburgh, Scotland, U.K., in 2023. He is a Professor Emeritus of signal processing and machine learning with the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece. He is also an Adjunct Professor with Aalborg University, Denmark. He has served as a Visiting Professor with Shenzhen Research Institute of Big Data (SRIBD), Chinese University of Hong Kong, Shenzhen, China (2018–2020). He is the author of the book titled *Machine Learning: From the Classics to Deep Networks, Transformers and Diffusion Models* (Academic Press, 3rd ed., 2024), the co-author of the best-selling book titled *Pattern Recognition* (Academic Press, 4th ed., 2009), the co-author of the book titled *Introduction to Pattern Recognition: A MATLAB Approach* (Academic Press, 2010), the co-editor of the book titled *Efficient Algorithms for Signal Processing and System Identification* (Prentice Hall, 1993), and the co-author of three books in Greek, two of them for the Greek Open University. He is the co-author of seven papers that have received the *Best Paper Awards* including the 2014 IEEE Signal Processing Magazine Best Paper Award and the 2009 IEEE Computational Intelligence Society Transactions on Neural Networks Outstanding Paper Award. His research interests include intersection of machine learning and signal processing. He is the recipient of the 2021 IEEE Signal Processing Society (SPS) *Norbert Wiener Award*, which is the IEEE SP Society's highest honor, the 2017 EURASIP *Athanasios Papoulis Award*, the 2014 IEEE SPS *Carl Friedrich Gauss Education Award* and the 2014 EURASIP *Meritorious Service Award*. He has served as a *Distinguished Lecturer* for the IEEE SP as well as the Circuits and Systems Societies. He was *Otto Monstead Guest* Professor, Technical University of Denmark, 2012, and holder of the *Excellence Chair*, Department of Signal Processing and Communications, University Carlos III, Madrid, Spain, 2011. Currently, he serves as a Chairman of the IEEE SP Society Awards Committee. He has served as a Vice President IEEE Signal Processing Society, as a President of the European Association for Signal Processing (EURASIP), as a Member of the Board of Governors for the IEEE Circuits and Systems (CAS) Society, as a Member of the Board of Governors (Member-at-Large) of the IEEE SP Society and as a Chair of the Signal Processing Theory and Methods (SPTM) technical committee of IEEE SPS. He has served as an Editor-in-Chief for IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is a Fellow of IET, a Corresponding Fellow of the Royal Society of Edinburgh (RSE), and a Fellow of EURASIP.