

Missing Spectrum-Data Recovery in Cognitive Radio Networks Using Piecewise Constant Nonnegative Matrix Factorization

Alireza Zaeemzadeh, Mohsen Joneidi, Behzad Shahrabi, and Nazanin Rahnavard
 School of Electrical Engineering and Computer Science
 University of Central Florida

Emails: {zaeemzadeh, joneidi, behzad.shahrabi}@knights.ucf.edu, and nazanin@eecs.ucf.edu

Abstract—In this paper, we propose a missing spectrum data recovery technique for cognitive radio (CR) networks using Nonnegative Matrix Factorization (NMF). It is shown that the spectrum measurements collected from secondary users (SUs) can be factorized as product of a channel gain matrix times an activation matrix. Then, an NMF method with piecewise constant activation coefficients is introduced to analyze the measurements and estimate the missing spectrum data. The proposed optimization problem is solved by a Majorization-Minimization technique. The numerical simulation verifies that the proposed technique is able to accurately estimate the missing spectrum data in the presence of noise and fading.

Index Terms—Nonnegative Matrix Factorization, Cognitive Radio Network, Spectrum Sensing, Missing Data Estimation

I. INTRODUCTION

Recent advances in wireless communications and micro-electronic devices are leading the trend of research toward cognitive radios (CRs) [1]. The main feature of CRs is the opportunistic usage of spectrum. CR systems try to improve the spectrum efficiency by using the spectrum holes in frequency, time, and space domains [1, 2]. This means that secondary users (SUs) are allowed to utilize the spectrum, provided that their transmissions do not interfere with the communication of primary users (PUs) [3]. The fundamental components of CR systems that allow them to avoid interference are spectrum sensing and resource allocation.

However, in a practical CR network, spectrum occupancy measurements for all the frequency channels at all times are not available. This is partially because of energy limitations and network failures. Another highly important and very common reason for occurrence of missing entries in the data set is the hardware limitation. Each SU may want to use different frequency channels, but it may not be capable of sensing all the channels simultaneously [4, 5]. On the other hand, a complete and reliable spectrum sensing data set is needed for a reliable resource allocation. Therefore, we need to develop a method to estimate the missing spectrum sensing measurements. This task is especially more challenging in dynamic environments.

There are different approaches toward the problem of data analysis in the CR networks. In [6], a learning approach is introduced based on support vector machine (SVM) for

spectrum sensing in multi-antenna cognitive radios. SVM classification techniques are applied to detect the presence of PUs. Several algorithms have been proposed using dictionary learning framework [7, 8]. These approaches try to find the principal components of data using dictionary learning and exploit the components to extract information.

The goal of this paper is to estimate the missing spectrum sensing data as accurate as possible in the time varying environments. An approach is introduced based on Nonnegative Matrix Factorization (NMF) [9, 10] to represent the spectrum measurements as additive, not subtractive, combination of several components. Each component reflects signature of one PU, therefore the data can be factorized as the product of signatures matrix times an activation matrix.

Dimension reduction is an inevitable pre-processing step for high dimensional data analysis [11]. NMF is a dimension reduction technique that has been employed in diverse fields [12, 13]. The most important feature of NMF, which makes it distinguished from other component analysis methods, is the non-negativity constraint. Thus the original data can be represented as additive combination of its parts.

In our proposed method, a new framework is introduced to decompose the spectrum measurements in CR networks using a piecewise constant NMF algorithm in presence of missing data. Piecewise constant NMF and its application in video structuring is introduced in [14]. In the proposed method, we try to handle the missing entries in the data and also take a different approach to solve the corresponding optimization problem using an iterative reweighted technique.

In the context of CR networks, NMF is utilized in [15] to estimate the power spectra of the sources in a CR network by factorizing the Fourier Transform of the correlation matrix of the received signals. Our proposed method estimates the missing entries in power spectral density measurements by enforcing a temporal structure on the activity of the PUs and can be used in scenarios when the number of the PUs is not known.

The introduced method takes advantage of a prior information about the activity of the PUs and exploits piecewise constant constraint to improve the performance of the factorization. Moreover, a solution for the introduced minimization

problem is suggested using the Majorization-Minimization (MM) framework.

The rest of the paper is organized in the following order. In Section II, the system model and the problem statement are introduced. Section III describes the proposed new NMF problem. In Section IV, a method is presented to solve the piecewise constant NMF problem in MM framework with missing data. Section V presents the simulation results and finally Section VI draws conclusions.

II. SYSTEM MODEL

Due to the nature of wireless environments, trustworthy information cannot be extracted from measurements of a single SU. To find the spectrum holes in frequency, time, and space, there exists a fusion center that collects and combines the measurements from all the SUs [4]. Cooperative spectrum sensing makes the missing data estimation algorithm more robust. Fusion center predicts the missing entries by using the collected measurements. However, since each SU is not able to sense the whole spectrum all the time, the data set collected from the SUs contains missing entries. Network failures, energy limitations, and shadowing can also cause loss of data.

Without loss of generality, we want to reconstruct the power map in a single frequency band. The network consists of N_P primary transmitters and N_R spectrum sensors that are randomly spread over the entire area of interest. Figure 1 illustrates an example of a network with $N_P = 2$ PUs and $N_R = 10$ SUs in a 100×100 area.

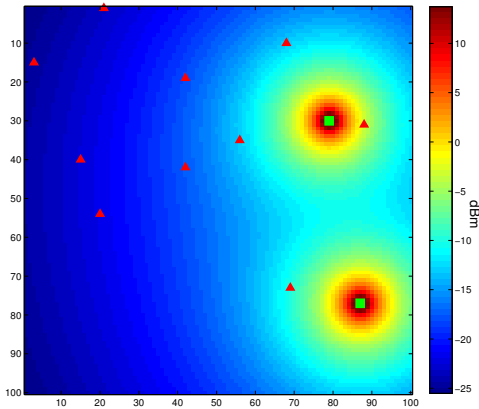


Fig. 1: The power distribution of 2 PUs (squares) and deployment of 10 SUs (triangles) without considering shadowing effect.

The received power of the r^{th} sensor at time t can be written as

$$s_r(t) = \sum_{j=1}^{N_P} p_j(t) \gamma_{rj}(t) + z_r(t), \quad (1)$$

where $p_j(t)$ is the transmit-power of the j^{th} PU at time t , γ_{rj} is the channel gain from the j^{th} PU to the r^{th} SU, and $z_r(t)$ is the zero-mean Gaussian measurement noise at the r^{th} sensor

with variance σ_r^2 . Considering a Rayleigh fading model, the channel gain coefficient can be modeled as:

$$\gamma_{rj} = \frac{C |h_{rj}|^2}{d_{rj}^\alpha}, \quad (2)$$

where the channel constant $C = \frac{G_P G_R c^2}{(4\pi f)^2}$, f is the carrier frequency, c is the speed of light, and G_P and G_R are the transmitter and receiver antenna gains. α is the path loss exponent which determines the rate at which power decays with the separation distance d_{rj} between the r^{th} SU and the j^{th} PU and $|h_{rj}|^2$ models the fading effect.

At time slot t , measurements from SUs can be stacked in a vector $\mathbf{s}(t)$, given as

$$\mathbf{s}(t) = \sum_{j=1}^{N_P} p_j(t) \boldsymbol{\gamma}_j(t) + \mathbf{z}(t), \quad (3)$$

where $\mathbf{s}(t) = [s_1(t) \ s_2(t) \ \dots \ s_{N_R}(t)]^T$, $\boldsymbol{\gamma}_j(t) = [\gamma_{1j}(t) \ \gamma_{2j}(t) \ \dots \ \gamma_{N_R j}(t)]^T$, and $\mathbf{z}(t) = [z_1(t) \ z_2(t) \ \dots \ z_{N_R}(t)]^T$. At each time slot, only a few SUs observe the power levels and report them to the fusion center. Therefore the vector $\mathbf{s}(t)$ contains some *missing* entries. Furthermore, each PU can be active or inactive in each time slot.

Some of the characteristics of the environment can be exploited to simplify the problem. It is assumed that channel gains are slowly time varying such that they can be considered as constant in a time window. Therefore, matrix representation of (3) can be written as:

$$\mathbf{S} = \mathbf{\Gamma} \mathbf{P} + \mathbf{Z}, \quad (4)$$

where \mathbf{S} is an $N_R \times T$ matrix, which includes measurements from sensors in T time slots, $\mathbf{\Gamma} = [\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{N_P}]$ is a $N_R \times N_P$ matrix, which consists of N_P channel gain vectors in the N_R -dimensional space of data, and \mathbf{P} is an $N_P \times T$ matrix that indicates the power levels of PUs in each time slot ($p_{jt} = 0$ if the j^{th} PU is inactive at time t).

Here, the goal is to estimate the missing data using the partial observations. To achieve this goal, the data is decomposed using piecewise constant NMF. Then the components of data and the activation matrix are used to estimate the missing data.

III. PC-NMF: PIECEWISE CONSTANT NONNEGATIVE MATRIX FACTORIZATION

Promoted by (4), it is easy to see that the measurements of each time slot can be represented as an additive, not subtractive, combination of few vectors. This algebraic representation has a geometric interpretation. Figure 2 helps us to visualize the structure of data in a 3-dimensional space of data. In this figure, 3 SUs are measuring power levels in an area with 3 PUs. It is easy to notice that measurement vectors lie within a pyramid in the positive orthant with $N_P = 3$ edges proportional to $\boldsymbol{\gamma}_j$. This is due to fact that all the points in the pyramid can be written as an additive combination of the edges.

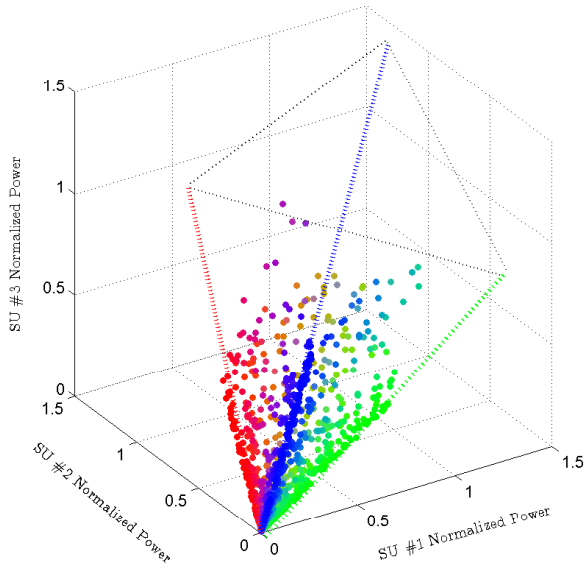


Fig. 2: Structure of data generated by $N_P = 3$ PUs in $N_R = 3$ dimensional space.

Although it is assumed that the channel gains are stationary for a time window of length T , PUs can become activated/deactivated in this time window any number of times and can change their transmission power in each activation. Hence, the power levels of PUs tend to be piecewise constant.

NMF is a widely-used technique to decompose data to its nonnegative components. Here, the structure of the power level matrix \mathbf{P} is exploited while handling the missing entries. As a result, the general objective function is presented as follows:

$$\begin{aligned} & \underset{\mathbf{\Gamma}, \mathbf{P}}{\text{minimize}} && D_W(\mathbf{S}|\mathbf{\Gamma}\mathbf{P}) + \beta F(\mathbf{P}), \\ & \text{subject to} && \mathbf{\Gamma} \geq 0, \mathbf{P} \geq 0, \end{aligned} \quad (5)$$

where $D_W(\mathbf{S}|\mathbf{\Gamma}\mathbf{P})$ is a weighted measure of fit and $F(\mathbf{P})$ is a penalty, which favors piecewise constant solutions. β is a nonnegative scalar weighting the penalty. The constraints denote that all the entries of $\mathbf{\Gamma}$ and \mathbf{P} are nonnegative. W is an $N_R \times T$ weight matrix that is used to estimate the weighted distance between \mathbf{S} and $\mathbf{\Gamma}\mathbf{P}$. The coefficients of the weight matrix denote the presence of data ($w_{rt} = 0/w_{rt} = 1$ if the measurement of the r^{th} SU at time slot t is unavailable/available).

NMF algorithms utilize different measures of fit such as Euclidean distance, generalized Kullback-Leibler (KL) divergence, and the Itakura-Saito divergence. In all the cases, the distance can be calculated as the sum of the distances between different coefficients [16–18].

$$D_W(\mathbf{S}|\mathbf{\Gamma}\mathbf{P}) = \sum_{t=1}^T \sum_{r=1}^{N_R} w_{rt} d(s_{rt} | \sum_{j=1}^{N_P} \gamma_{rj} p_{jt}), \quad (6)$$

In our case, Euclidean Distance is used as the measure of fit. This objective function is commonly used for problems with Gaussian noise model, a common noise model in communication systems, hence:

$$d(s_{rt} | \sum_{j=1}^{N_P} \gamma_{rj} p_{jt}) = \frac{1}{2} (s_{rt} - \sum_{j=1}^{N_P} \gamma_{rj} p_{jt})^2. \quad (7)$$

Since there exist sharp transitions in power level of PUs and power level of each PU is constant in each transmission period, rows of \mathbf{P} tend to be piecewise constant. In order to favor the piecewise constant solutions, the penalty function is defined as:

$$F(\mathbf{P}) = \sum_{t=2}^T \sum_{j=1}^{N_P} \lim_{n \rightarrow 0} |p_{jt} - p_{j(t-1)}|^n. \quad (8)$$

When n tends to 0, this penalty function represents the sum of ℓ_0 norm of the transition vectors, i.e. $\mathbf{p}_t - \mathbf{p}_{(t-1)}$, where \mathbf{p}_t is an $N_P \times 1$ vector containing power levels of PUs in time t . This penalty favors the solutions with a lower number of transitions. However, since it is not differentiable, it can be replaced with a differentiable approximation:

$$\begin{aligned} F_\epsilon(\mathbf{P}) &= \sum_{t=2}^T \sum_{j=1}^{N_P} \rho_\epsilon(p_{jt} - p_{j(t-1)}), \\ & \text{with} \quad \rho_\epsilon(x) = \frac{x^2}{x^2 + \epsilon^2}, \end{aligned} \quad (9)$$

where ϵ^2 is a small positive constant and is much less than all the non-zero elements of $(p_{jt} - p_{j(t-1)})^2 \forall j, t$ to avoid division by zero.

In Section IV, an algorithm is derived to find the minimizer of the following problem:

$$\begin{aligned} & \underset{\mathbf{\Gamma}, \mathbf{P}}{\text{minimize}} && D_W(\mathbf{S}|\mathbf{\Gamma}\mathbf{P}) + \beta F_\epsilon(\mathbf{P}), \\ & \text{subject to} && \mathbf{\Gamma} \geq 0, \mathbf{P} \geq 0. \end{aligned} \quad (10)$$

After estimating \mathbf{P} and $\mathbf{\Gamma}$, the missing entries of \mathbf{S} can be approximated using the equation $\hat{\mathbf{S}} \simeq \mathbf{\Gamma}\mathbf{P}$.

IV. MAJORIZATION-MINIMIZATION FOR PIECEWISE CONSTANT NMF

In this section, an iterative algorithm is described to find the solution of the optimization problem proposed in (10). For that, Majorization-Minimization (MM) framework is employed [16, 17]. MM algorithm and its variants have been used in various applications such as parameter learning and image processing [19, 20]. The update rules are derived to calculate the entries of \mathbf{P} given the entries of $\mathbf{\Gamma}$ and then the entries of $\mathbf{\Gamma}$ given the entries of \mathbf{P} , using an iterative reweighted algorithm.

First, the update rules for \mathbf{P} given $\mathbf{\Gamma}$ are derived. Then, the update rules for $\mathbf{\Gamma}$ will be derived in a similar manner.

As it is clear in (6), we can write the distance measure as a sum of different time slots:

$$D_W(\mathbf{S}|\mathbf{\Gamma}\mathbf{P}) = \sum_{t=1}^T C(\mathbf{p}_t), \quad (11)$$

where $C(\mathbf{p}_t)$ is the weighted Euclidean distance between \mathbf{s}_t and $\Gamma\mathbf{p}_t$, given \mathbf{s}_t and Γ . In MM framework, the update rules are derived by minimizing an auxiliary function [16]. By definition, $G(\mathbf{p}_t, \hat{\mathbf{p}}_t)$ is an auxiliary function of $C(\mathbf{p}_t)$ if and only if $G(\mathbf{p}_t, \hat{\mathbf{p}}_t) \geq C(\mathbf{p}_t)$ and $G(\mathbf{p}_t, \mathbf{p}_t) = C(\mathbf{p}_t)$ for $\forall \mathbf{p}_t$. If $G(\mathbf{p}_t, \hat{\mathbf{p}}_t)$ is chosen such that it is easier to minimize, the optimization of $C(\mathbf{p}_t)$ can be replaced with iterative minimization of $G(\mathbf{p}_t, \hat{\mathbf{p}}_t)$ over \mathbf{p}_t . Thus, in the literature, convex functions are frequently used as the auxiliary functions [12, 16]. It is shown in [16] that $C(\mathbf{p}_t)$ is non-increasing under the update

$$\mathbf{p}_t^{i+1} = \underset{\mathbf{p}_t}{\operatorname{argmin}} G(\mathbf{p}_t, \mathbf{p}_t^i). \quad (12)$$

This is due to the fact that in the i^{th} iteration we have $C(\mathbf{p}_t^{i+1}) \leq G(\mathbf{p}_t^{i+1}, \mathbf{p}_t^i) \leq G(\mathbf{p}_t^i, \mathbf{p}_t^i) = C(\mathbf{p}_t^i)$.

Following a similar approach as [16], the auxiliary function for the weighted Euclidean distance can be formulated as:

$$G(\mathbf{p}_t, \mathbf{p}_t^i) = C(\mathbf{p}_t^i) + (\mathbf{p}_t - \mathbf{p}_t^i) \nabla C(\mathbf{p}_t^i) + \frac{1}{2} (\mathbf{p}_t - \mathbf{p}_t^i)^T K(\mathbf{p}_t^i) (\mathbf{p}_t - \mathbf{p}_t^i), \quad (13)$$

where $K(\mathbf{p}_t^i)$ is an $N_P \times N_P$ diagonal matrix with

$$k_{jj}(\mathbf{p}_t^i) = \frac{q_{jt}^i}{p_{jt}^i}, \quad (14)$$

$$\mathbf{q}_t^i = \Gamma^T (\mathbf{w}_t \odot \Gamma \mathbf{p}_t^i),$$

and k_{jj} is the j^{th} diagonal entry of $K(\mathbf{p}_t^i)$ and \odot is element-wise multiplication.

To solve the problem presented in (10), the contribution of $F_\epsilon(\mathbf{P})$ should be considered in the auxiliary function. For that, a convex version of $F_\epsilon(\mathbf{P})$ is employed:

$$F_\epsilon(\mathbf{P}) = \sum_{t=2}^T \sum_{j=1}^{N_P} y_{jt} (p_{jt} - p_{j(t-1)})^2, \quad (15)$$

with
$$y_{jt} = \frac{1}{(p_{jt}^{(i-1)} - p_{j(t-1)}^{(i-1)})^2 + \epsilon}.$$

Now the update rules can be obtained using the iterative version of (15). This means that y_{jt} is updated in each iteration using the values of \mathbf{P} in the previous iteration.

To form the penalized auxiliary function, $G_\beta(\mathbf{p}_t, \mathbf{p}_t^i)$, we add up $G(\mathbf{p}_t, \mathbf{p}_t^i)$ with the contribution of \mathbf{p}_t to $F(\mathbf{P})$. Thus, $G_\beta(\mathbf{p}_t, \mathbf{p}_t^i)$ can be written as:

$$G_\beta(\mathbf{p}_t, \mathbf{p}_t^i) = G(\mathbf{p}_t, \mathbf{p}_t^i) + \beta \left[\sum_{j=1}^{N_P} y_{jt}^i (p_{jt} - p_{j(t-1)}^i)^2 + \sum_{j=1}^{N_P} y_{j(t+1)}^i (p_{j(t+1)}^i - p_{jt}^i)^2 \right]. \quad (16)$$

It is worthwhile to mention that $y_{j1} = y_{j(T+1)} = 0 \forall j$. Since $G_\beta(\mathbf{p}_t, \mathbf{p}_t^i)$ is convex, it can be easily minimized over \mathbf{p}_t by setting the gradient to zero. Hence the update rule is

attained as:

$$p_{jt}^{i+1} = \frac{-\nabla_j C(\mathbf{p}_t^i) + p_{jt}^i k_{jj}(\mathbf{p}_t^i) + 2\beta y_{jt}^i p_{j(t-1)}^i + 2\beta y_{j(t+1)}^i p_{j(t+1)}^i}{k_{jj}(\mathbf{p}_t^i) + 2\beta y_{jt}^i + 2\beta y_{j(t+1)}^i},$$

$$\nabla C(\mathbf{p}_t^i) = -\Gamma^T (\mathbf{w}_t \odot \mathbf{s}_t - \mathbf{w}_t \odot (\Gamma \mathbf{p}_t^i)), \quad (17)$$

where $\nabla_j C(\mathbf{p}_t^i)$ is the j^{th} element of the gradient $\nabla C(\mathbf{p}_t^i)$.

Finding the update rule for Γ is simple. This is due to the fact that $F(\mathbf{P})$ is not a function of Γ . Hence, the update rule for Γ is similar to the update rule for standard NMF, except the missing entries must be taken into account [21]. The update rules can be written in matrix form as:

$$\Gamma^{i+1} = \Gamma^i \odot \frac{(\mathbf{W} \odot \mathbf{S}) \mathbf{P}^T}{(\mathbf{W} \odot (\Gamma^i \mathbf{P})) \mathbf{P}^T} \quad (18)$$

where \odot is the element-wise multiplication and the division is also performed in an element-wise manner.

The obtained update rules in (17) and (18) are exploited alternatively to estimate Γ and \mathbf{P} . Then the missing entries of \mathbf{S} are predicted by $\hat{\mathbf{S}} = \Gamma \mathbf{P}$.

However, by using the objective function in (10), the optimization problem results in solutions with entries of \mathbf{P} tend toward 0 and $\|\Gamma\|$ tends toward ∞ . We take advantage of the scale ambiguity between Γ and \mathbf{P} to avoid this issue. Let Λ be a diagonal $N_P \times N_P$ matrix with its j^{th} diagonal entry equal to $\|\gamma_j\|_2$. In each iteration, the rescaled matrix pair $(\Gamma \Lambda^{-1}, \Lambda \mathbf{P})$ is used instead of the original matrix pair (Γ, \mathbf{P}) .

As a practical scenario, we should also consider the case when the secondary network has no information about the number of the PUs, i.e. N_P . In this case, the common dimension of matrices Γ and \mathbf{P} is not known. There have been some efforts in model order selection in NMF [18]. In the numerical experiments, $K > N_P$ is used as the common dimension to factorize the data in such conditions. This is only possible if the secondary network has some information about the upper bound of N_P .

V. NUMERICAL RESULTS

For the numerical experiments, one frequency channel is considered with $N_p = 3$ active PUs in the area. Figure 3 illustrates the topology of the network. Incomplete measurements are collected from $N_R = 20$ SUs.

We use the same simulation environment and the same network topology as in [8]. The simulation parameters are set as follows, unless otherwise is stated. The path loss is computed as $(\frac{d}{d_0})^\alpha$, where d is the distance, $d_0 = 0.01$, and $\alpha = 2.5$. γ_{rj} is computed by multiplying the pathloss by the fading coefficient $|h_{rj}|^2$ where

$$h_{rj}(t) = \eta h_{rj}(t-1) + \sqrt{1 - \eta^2} \nu_{rj}(t), \quad (19)$$

$\eta = 0.9995$, and $\nu_{rj}(t)$ is circularly symmetric zero mean complex Gaussian noise with variance 1 [8].

PUs' activity is modeled by a first order Markov model. All the PUs utilize the spectrum $\lambda = 0.3$ of the time slots. Transition matrix of the j^{th} PU is $\begin{pmatrix} 1 - a_j & a_j \\ b_j & 1 - b_j \end{pmatrix}$ and

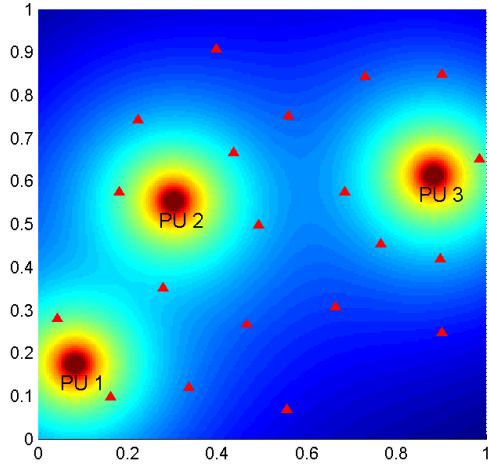


Fig. 3: Network topology consisting of 3 PUs and $N_R = 20$ SUs marked by triangles.

$\lambda = \frac{b_j}{a_j + b_j}$. a_j is the probability that the j^{th} PU stops transmitting from time $t-1$ to t and b_j is the probability that the PU activates transmitting. The parameter a_j is uniformly distributed over $[0.05, 0.15]$.

Each time a PU becomes activated, it chooses the transmission power from a uniform distribution with support $[100, 200]$. Each SU makes a measurement with 70% of chance. The measurements are contaminated by additive white Gaussian noise. The noise variance is 10^{-5} for all the SUs.

Partial measurements are generated for $T = 600$ time slots. To reduce the computational burden, the first 300 time slots are used to estimate Γ . Next, by using the obtained Γ and the update rule (17), \mathbf{P} is estimated for all 600 time slots. The regularization factor β is set to 5×10^{-3} and $K = 5$ factors are used to factorize the data.

Figure 4 shows the true power levels and the reconstructed one at a randomly selected SU versus time for the time window of $T = 300$ samples. It can be seen that the missing entries are accurately recovered through the proposed method, and it is evident that the proposed algorithm can easily track abrupt changes in power level.

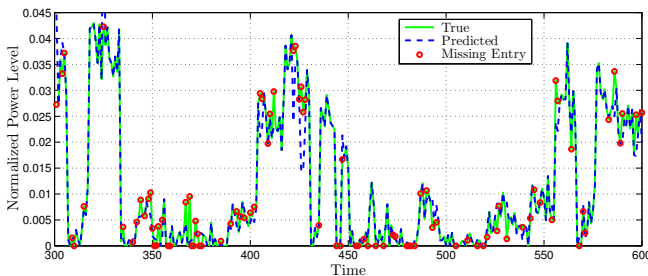
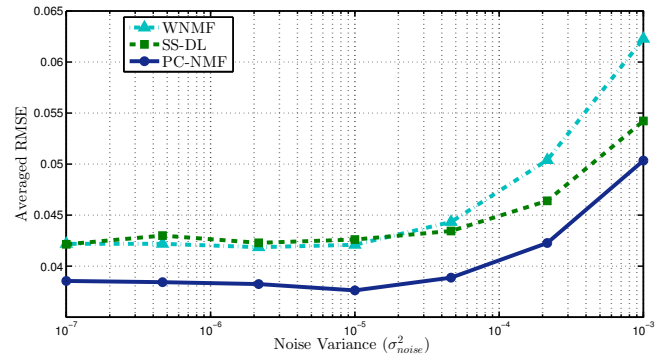


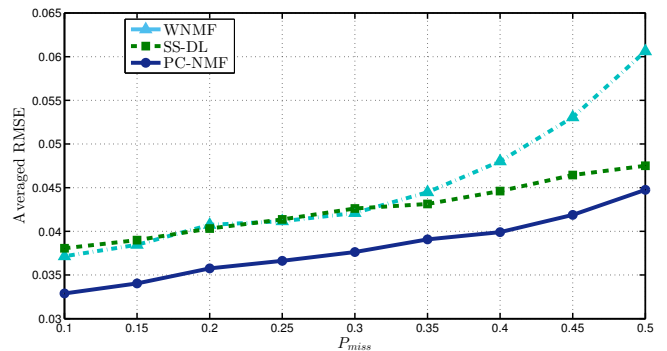
Fig. 4: Power levels of a single SU.

Figure 5 compares the RMSE, averaged over SUs, of the proposed method with two similar methods. The method introduced in [8] exploits the spatial correlation between

adjacent SUs' measurements and semi-supervised dictionary learning (SS-DL) to estimate the missing entries. For the numerical results, the batch version of SS-DL is employed and the parameters are set to their optimal values. Furthermore, to emphasize the effect of the piecewise constant penalty, the results are also compared with the weighted NMF, i.e. WNMF [21, 22]. WNMF employs binary weight matrix to deal with the missing entries. This figure shows that the proposed method outperforms its competitors in different noise levels (Figure 5.(a)) and different probabilities of miss (Figure 5.(b)). P_{miss} denotes the ratio of the missing entries among the spectrum data.



(a) $P_{miss} = 0.3$, $K = 5$



(b) $\sigma_{noise}^2 = 10^{-5}$, $K = 5$

Fig. 5: Performance of the proposed method for different noise levels and probability of miss, averaged over 200 Monte Carlo trials.

This figure shows that WNMF and SS-DL almost perform the same for low noise variance and low P_{miss} . However, for harsh environments with high noise variance or high P_{miss} , SS-DL produces more accurate results. The PC-NMF method outperforms both methods in different noise levels and different probabilities of miss. For instance, PC-NMF has 11.6% less RMSE compared to the SS-DL method for $\sigma_{noise}^2 = 10^{-5}$ and $P_{miss} = 0.3$.

This improvement in the performance does not increase the computational burden of the algorithm. Table I shows the running times¹ for different methods averaged over 100 Monte Carlo trials for $\sigma_{noise}^2 = 10^{-5}$, $P_{miss} = 0.3$, and $K = 5$.

¹All simulations have been performed under MATLAB 2014a environment on a PC equipped with Intel Xeon E5-1650 processor (3.20 GHz) and 8 GB of RAM.

TABLE I: Average Running Time

| Method | Average Running Time (s) |
|--------|--------------------------|
| SS-DL | 10.3039 |
| WNMF | 0.0944 |
| PC-NMF | 0.0952 |

It is known that the NMF methods converge much faster than methods based on gradient descent [21]. However, Table I also illustrates that the proposed method does not require more computational resources compared with WNMF.

To study the effect of the piecewise constant penalty on the output of the algorithm, Figure 6 depicts the power level of two PUs and the estimated activation levels using the introduced method and WNMF. Both methods can estimate the power levels up to a scale factor. The number of factors is set to 3, i.e. $K = N_P$.

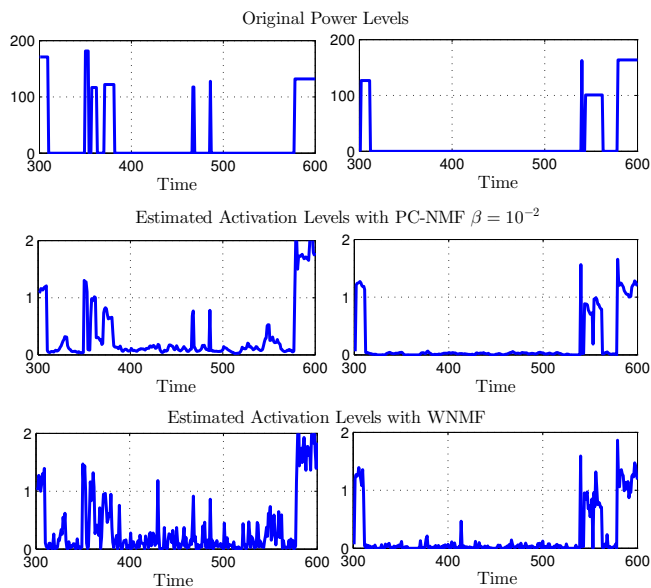


Fig. 6: Original power levels of different PUs and the estimated activation levels with PC-NMF and WNMF.

This figure illustrates the fact that the proposed method produces a more accurate factorization by taking advantage of piecewise constant constraint as a prior information. As it was expected, power levels estimated by PC-NMF are piecewise constant, while the results generated by WNMF are noisy. In fact, the piecewise constant penalty decreases the effect of noise and fading. Moreover, the sharp transitions are preserved in the factors returned by PC-NMF.

VI. CONCLUSIONS

By exploiting inherent structural feature of cognitive radio networks, we proposed a piecewise constant NMF approach that can decompose the data set into its components. Majorization-Minimization framework is utilized to solve the optimization problem of the piecewise constant NMF. Numerical simulations suggest that this method is able to predict the missing entries in the spectrum sensing database accurately.

REFERENCES

- [1] Q. Zhao, "A survey of dynamic spectrum access: signal processing, networking, and regulatory policy," in *IEEE Signal Processing Magazine*, pp. 79–89, 2007.
- [2] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Comput. Netw.*, vol. 50, pp. 2127–2159, Sept. 2006.
- [3] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 23, pp. 201–220, Feb 2005.
- [4] W.-Y. Lee and I. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *Wireless Communications, IEEE Transactions on*, vol. 7, pp. 3845–3857, October 2008.
- [5] B. Shahrasbi and N. Rahnavard, "A clustering-based coordinated spectrum sensing in wideband large-scale cognitive radio networks," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, pp. 1101–1106, Dec 2013.
- [6] O. Awe, Z. Zhu, and S. Lambotharan, "Eigenvalue and support vector machine techniques for spectrum sensing in cognitive radio networks," in *Technologies and Applications of Artificial Intelligence (TAI), 2013 Conference on*, pp. 223–227, Dec 2013.
- [7] S. Amini, M. Sadeghi, M. Joneidi, M. Babaie-Zadeh, and C. Jutten, "Outlier-aware dictionary learning for sparse representation," in *Machine Learning for Signal Processing (MLSP), 2014 IEEE International Workshop on*, pp. 1–6, Sept 2014.
- [8] S.-J. Kim and G. Giannakis, "Cognitive radio spectrum prediction using dictionary learning," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, pp. 3206–3211, Dec 2013.
- [9] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.
- [10] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.
- [11] M. Rahmani and G. Atia, "Randomized robust subspace recovery for high dimensional data matrices," *arXiv preprint arXiv:1505.05901*, 2015.
- [12] S. Essid and C. Févotte, "Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring," *Multimedia, IEEE Transactions on*, vol. 15, pp. 415–425, Feb 2013.
- [13] Z. Hu, R. Ranganathan, C. Zhang, R. Qiu, M. Bryant, M. Wick, and L. Li, "Robust non-negative matrix factorization for joint spectrum sensing and primary user localization in cognitive radio networks," in *IEEE WDD Conference*, 2012.
- [14] N. Seichepine, S. Essid, C. Févotte, and O. Cappe, "Piecewise constant nonnegative matrix factorization," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 6721–6725, May 2014.
- [15] X. Fu, N. Sidiropoulos, and W.-K. Ma, "Tensor-based power spectra separation and emitter localization for cognitive radio," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th*, pp. 421–424, June 2014.
- [16] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, pp. 556–562, 2001.
- [17] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β -divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011.
- [18] V. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization with the /spl beta/-divergence," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, pp. 1592–1605, July 2013.
- [19] J. M. Bioucas-Dias, M. Figueiredo, J. P. Oliveira, et al., "Adaptive total variation image deconvolution: A majorization-minimization approach," in *Signal Processing Conference, 2006 14th European*, pp. 1–4, IEEE, 2006.
- [20] S. Minaee and Y. Wang, "Screen content image segmentation using least absolute deviation fitting," *arXiv preprint arXiv:1501.03755*, 2015.
- [21] Y. Mao and L. K. Saul, "Modeling distances in large-scale networks by matrix factorization," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, IMC '04*, (New York, NY, USA), pp. 278–287, ACM, 2004.
- [22] Y.-D. Kim and S. Choi, "Weighted nonnegative matrix factorization," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 1541–1544, April 2009.