Recovering Missing Values From Corrupted Historical Observations: Approaching the Limit of Predictability in Spectrum Prediction Tasks

XI LI\textsuperscript{1,2,3}, (Student Member, IEEE), GUOJUN CHEN\textsuperscript{1,2}, (Student Member, IEEE), YINFEI XU\textsuperscript{2}, (Member, IEEE), XIN WANG\textsuperscript{3}, (Senior Member, IEEE), AND TIECHENG SONG\textsuperscript{1,2}, (Member, IEEE)

\textsuperscript{1}National Mobile Communication Research Laboratory, Southeast University, Nanjing 210096, China
\textsuperscript{2}School of Information Science and Engineering, Southeast University, Nanjing 210096, China
\textsuperscript{3}Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794, USA

Corresponding authors: Xi Li (lixixuzhou@seu.edu.cn) and Yinfei Xu (yinfeixu@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771126 and Grant 61901105, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20190343, in part by the Key Research and Development Plan of Jiangsu Province under Grant BE2018108, in part by the Youth Fund of Ministry of Education (MOE), Project of Humanities and Social Sciences, China, under Project 20YJC630245, in part by the Fundamental Research Funds for the Central Universities under Grant 30919013229, and in part by the Electrical, Communications and Cyber Systems, National Science Foundation, USA, under Grant 1731238 and 2030063.

\textbf{ABSTRACT} Spectrum prediction is a key enabler of a range of emerging applications, from adaptive spectrum sensing to agile and proactive decision making, for dynamic spectrum access. Due to the fact that spectrum sensors easily experience an issue of missing readings and anomaly pollution, spectrum prediction with incomplete and corrupted historical observations has caused extensive concern. In this paper, we aim to tackle the challenging problem on how to accurately and efficiently recover the missing values from corrupted historical spectrum observations for approaching the limit of predictability in the spectrum prediction tasks. Specially, we first formulate a Hankelized time-structured spectrum tensor that can naturally preserve both spectral and temporal dependencies among the historical spectrum observations. We then model the spectrum data recovery problem as tensor completion with its latent low-rank structure and sparse anomaly property. To efficiently solve the optimization problem, we design a robust online spectrum data recovery algorithm based on the alternating direction method. In addition, we also introduce the concept of maximum predictability to reveal the harmful effects of missing data and anomalies, as well as to evaluate the effectiveness of our proposed algorithm from an information theory perspective. Extensive experimental evaluations on the real-world spectrum observation dataset show that the proposed algorithm achieves significantly better spectrum data recovery performance in terms of both recovery accuracy and computation efficiency. It also indicates that not only the predictability of the frequency bands but also the prediction accuracy of any predictive algorithm can be improved by the proposed algorithm.

\textbf{INDEX TERMS} Cognitive radio, dynamic spectrum access, spectrum prediction, tensor completion.

I. INTRODUCTION

A. BACKGROUND

Over the past decade, the contradiction between spectrum shortage and spectrum under-utilization has motivated the active research on dynamic spectrum access (DSA). As a key enabling technique for DSA, cognitive radio (CR) provides the capability of harnessing the under-utilized spectrum resources in an opportunistic manner. The fundamental step of implementing CR technology is to capture the relevant information about the spectrum state evolution. It is based on spectrum data analytics, such as spectrum sensing and spectrum prediction. Briefly speaking, spectrum sensing detects the current spectrum state through signal detection, whereas
spectrum prediction infers future spectrum state evolution trajectory by fully exploiting the inherent statistical dependencies and regularities among historical observations. With the ability to foresee the state evolution of radio spectrum, spectrum prediction has attracted much attention during the past few years [1], [2], and has been widely developed for, such as, adaptive sensing scheduling [3], agile decision making [4], and proactive channel switching [5], in CR networks. So far, a diverse group of spectrum prediction methods have been proposed, e.g. Markov model-based spectrum prediction [6], machine learning-based spectrum prediction [7], Bayesian inference-based spectrum prediction [8], and deep learning-based spectrum prediction [9].

B. CHALLENGES

Missing values and anomalies are common in real-world spectrum data. In a typical CR network, large numbers of smart sensors are deployed and served as secondary users (SUs) to monitor the frequency bands in a real-time manner. However, in practical environment, the sensors easily experience an issue of missing readings due to unexpected hardware failures or communication interruptions [10]. Besides the data loss, the observed sensory data are also easily polluted by anomalies occur due to activity from malicious operations, or misconfigurations and failures of network equipments [11]. Previous studies [12]–[14] have demonstrated that the existence of missing values and anomalies in historical spectrum observations will compromise the prediction accuracy of several spectrum prediction algorithms. However, the algorithm-specific evaluations cannot reveal the inherent deficiency caused by missing values and anomalies for spectrum prediction, because the compromise could be incurred by either the structure of algorithms or the missing values and anomalies. Therefore, it is indispensable to introduce an algorithm-free metric to explicitly illustrate the effect of missing values and anomalies on spectrum prediction. The maximum predictability, which refers to the degree to which a correct prediction of the system’s future state can be made ignorant of specific predictive algorithm, is a good option. It is derived from information theory by considering both the randomness and regularity of the system. Maximum predictability was previously introduced by Song et al. for analyzing the mobility of mobile phone user [15]. Recently, it has also been introduced into CR networks for analyzing TV band and GSM band dynamics underlying real-world spectrum measurements [16], [17]. When it comes to the scope of spectrum prediction in this paper, we define the maximum predictability as the highest potential accuracy that could be achieved by any spectrum prediction algorithm. In Section II, we analyze the effect of missing values and anomalies reflected on the maximum predictability. Our results indicate that the accuracy of any spectrum prediction algorithm is limited by missing values and anomalies due to their destructive effects on the quality of historical spectrum observations in terms of predictability. To this end, accurately yet efficiently recovering missing values from corrupted historical spectrum observations is non-trivial and challenging in spectrum prediction tasks.

The key point to tackle this challenge lies on how to accurately model the quantitative dependencies of the missing values on the known ones. As the historical spectrum observations are typically expressed in the form of one-dimensional time series, majority of the existing studies have essentially focused on spectrum data recovery from the vector view, by exploiting the dependencies in time domain. Filtering algorithms of various kinds, such as MAF (Moving Average Filter) [18], SES (Single Exponential Smoothing) [19] and Kriging [20] have been developed for spectrum data recovery. In [21], a minimum Bayesian risk-based scheme is also developed for robust spectrum prediction in the presence of sensing errors. However, such approaches can only learn temporal dependencies. It inevitably ignores the spectral dependencies among different frequency bands, leading to an inaccurate estimation in many circumstances. To solve this issue, some researchers have taken consideration of multiple frequency bands at the same time. They treated the historical spectrum observations as two-dimensional spectral-temporal matrices, whose rows correspond to frequency bands and columns correspond to time slots, and proposed various matrix-based methods to estimate the missing values by capturing the inherent low-rank structure of the spectrum matrices. To the author’s best knowledge, [22] is among the first that introduces nonnegative matrix factorization to estimate the missing spectrum data in the presence of noise and fading. However, [22] assumes that anomalies are free and it only presents a batch algorithm therein. Moreover, in [23], the authors present a robust online scheme based on matrix completion and matrix recovery for spectrum data reconstruction and spectrum prediction. Although the matrix-based methods can well take advantage of the spectral and temporal dependencies, still they are limited by capturing only local dependencies, ignoring global dependencies such as periodicity and seasonality. Recently, tensor models are also introduced to jointly capture dependencies in time and frequency domains as well as the daily periodicity of the spectrum data. Then, based on it, tensor completion algorithms [24], [25] are developed for long-term spectrum prediction in the presence of missing data and anomalies. Despite their good performance on recovering missing values and separating anomalies, the computational complexity of the batch algorithms can easily become large. This will inevitably compromise the efficiency of spectrum prediction algorithms, since accurate but outdated prediction is useless.

C. CONTRIBUTIONS

In this paper, we propose to formulate a hankelized time-structured spectrum tensor to simultaneously capture the spectral-temporal dependencies as well as uncover latent time-directional structure of the spectrum data. Nevertheless, applying this high-level idea still requires addressing several challenges. The first challenge is that, the historical spectrum observations are easily corrupted by anomalies. It needs some
careful design to formulate an optimization problem considering the effects of anomalies for low-rank tensor-based data recovery. Secondly, as the spectrum tensor grows infinitely as time goes by, it deserves an elaborative consideration to efficiently solve the convex optimization problem with a convergence guarantee. To this end, the main contributions of this paper are summarized as follows:

1) We introduce the concept of maximum predictability to represent the upper bound of the potential accuracy that any spectrum prediction algorithm can reach and capture the effects of missing values and anomalies on spectrum prediction from an information theory perspective. We show that the maximum predictability of real-world frequency bands can reach up to 97% on average and degrades significantly with missing data and anomalies. Our findings indicate that the real-world frequency bands are highly predictable but limited by missing data and anomalies.

2) We propose to transfer the historical spectrum observations into a hankelized time-structured spectrum tensor, and introduce CANDECOMP/PARAFAC (CP) decomposition [26] to decouple the interdependency of different tensor modes. We design a robust online tensor completion algorithm by drawing upon recent advances of alternating direction method of multipliers (ADMM) [27] to accurately and efficiently recover the missing values under a circumstance of anomaly pollution.

3) We conduct extensive experiments with a real-world spectrum observation dataset to evaluate our proposed spectrum data recovery algorithm based on hankelized tensor completion (SDR-HTC). Our results show that the SDR-HTC algorithm outperforms the state-of-the-art spectrum data recovery algorithms in terms of both data recovery accuracy and computation efficiency. It is also shown that not only the maximum predictability of frequency bands but also the prediction accuracy of various spectrum prediction algorithms are improved by our SDR-HTC algorithm. This indicates the capability of our SDR-HTC algorithm to approach the limit of predictability in spectrum prediction tasks.

### D. PRELIMINARIES

This subsection presents some basic notations, operations and necessary definitions used throughout this paper. Tensors are denoted by uppercase calligraphic letters (A, B, ...), matrices are written as uppercase boldfaced letters (A, B, ...), vectors are represented by lowercase calligraphic letters (v, w, ...), and scalars are denoted by normal lowercase letters (a, b, ...). An Nth-order tensor can be denoted by \( \mathbf{X} \in \mathbb{R}^{i_1 \times i_2 \times \cdots \times i_N} \), whose elements can be represented by its symbolic name with indexed as subscripts as \( x_{i_1,i_2,\ldots,i_N} \), \( i_n \in \{1,2,\ldots,I_n\}, n \in \{1,2,\ldots,N\} \). Accordingly, when \( N = 2 \), a matrix can be obtained as \( \mathbf{X} \in \mathbb{R}^{i_1 \times i_2} \), where each element can be denoted by \( x_{i_1,i_2} \). Similarly, when \( N = 1 \), a vector can be obtained as \( \mathbf{x} \in \mathbb{R}^{i_1} \) and the elements of the vector can be denoted by \( x_{i_1} \). The key variables used in this paper are summarized in Table 1.

![Table 1. Key variables used in this paper.](image_url)

Some operations are quite essential and will be applied in this paper. Symbol "\( \odot \)" represents the vector outer product, while symbol "\( \otimes \)" represents the matrix Hadamard product. (.)' represents the transposition operation. \( \| . \|_0 \) stands for the \( l_0 \)-norm, which counts the number of non-zero elements in the tensor. \( \| . \|_1 \) represents the \( l_1 \)-norm, which calculates the sum of the magnitudes of the tensor. \( \| . \|_2 \) stands for the \( l_2 \)-norm, which calculates the square root of the sum of the square of all the elements of the vector. \( \| . \|_F \) denotes the frobenius norm, which calculates the square root of the sum of the square of the absolute value of elements of the tensor. \( \| . \|_* \) represents the nuclear norm, which calculates the sum of the singular values (SVs) of the tensor.

Tensors, matrices and vectors with the square bracket, e.g. \( \mathbf{A}[i], \mathbf{A}[i] \) and \( \mathbf{a}[i] \), represent the computed values after performing \( t \)-times updates in the online-based subspace learning algorithm described in Section III.

**Definition 1 (Slices of a tensor [28]):** A Tensor can be represented by a set of two-dimensional slices, which are defined by fixing all but two indexes of the tensor. For a
3rd-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, three kinds of slices, i.e., the horizontal, lateral and frontal slices can be obtained and denoted by $(\mathcal{X})_{i_1, \cdot, \cdot}$, $(\mathcal{X})_{\cdot, i_2, \cdot}$ and $(\mathcal{X})_{\cdot, \cdot, i_3}$, respectively. In this paper, we typically use $\mathcal{X}$ and $\mathcal{X}_r$ to represent the $f$th horizontal slice and the $r$th frontal slice of a tensor $\mathcal{X}$ respectively.

**Definition 2 (Rank-1 tensor):** A rank-1 tensor is also called simple tensor or decomposable tensor. The $N$th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is a rank-1 tensor as long as it can be written as the outer product of $N$ vectors $\mathcal{X} \doteq \mathbf{v}^{(1)} \odot \mathbf{v}^{(2)} \odot \cdots \odot \mathbf{v}^{(N)}$.

**Definition 3 (Candecomp/Parafac (CP) Decomposition):** The idea of CP decomposition is to express a tensor as the sum of a finite number of rank-1 tensors. By this way, the $N$th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ can be expressed as $\mathcal{X} = \sum_{r=1}^{R} \mathcal{V}^{(r)} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, $n \in \{1, 2, \ldots, N\}$. By introducing the factor matrices $\mathcal{V}^{(n)} = [v^{(n)}_1, v^{(n)}_2, \ldots, v^{(n)}_r] \in \mathbb{R}^{I_n \times R}$, the CP decomposition of tensor $\mathcal{X}$ can also be expressed as $\mathcal{X} = \prod_{n=1}^{N} \mathcal{V}^{(n)}$.

**E. ORGANIZATION**
The rest of the paper is organized as follows. In Section II, we present predictability analysis of real-world spectrum observations and evaluate the effect of missing values and anomalies on spectrum prediction. In Section III, we present our model and optimization algorithm. Section IV presents experimental results and analysis, and Section V offers some concluding remarks.

**II. PREDICTABILITY ANALYSIS**
In this section, we detail the methodology of using statistical entropy measures and Fano inequality to quantify the degree of predictability and explore the impact of missing data and anomalies on predicting the state evolution of the frequency bands.

Entropy is probably one of the most fundamental measure to characterize the degree of predictability of a stochastic process. In general, lower entropy corresponds to the higher predictability, and vice versa. Let $\mathbf{y} = \{y_{f, 1}, y_{f, 2}, \cdots, y_{f, T}\}$ denote the spectrum observation sequence of a given frequency band $f$ occurred at $T$ consecutive time slots, where $y_{f, t}, t \in 1, 2, \cdots, T$ is a random variable representing the spectrum observation of frequency band $f$ at time slot $t$. Considering both the probabilities of the spectrum observations and their temporal order, the entropy of $\mathbf{y}$ is given by

$$E_f = -\sum_{s' \subset \mathbf{y}} P(s') \log_2 \left[ P(s') \right].$$

where $P(s')$ represents the probability of finding a particular time-ordered subsequence $s'$ in the spectrum observation sequence $\mathbf{y}$. However, the problem of finding all the subsets of a given set has exponential computational complexity. In practice, we use the estimator proposed by Lempel and Ziv [29] that rapidly converges to the value of the entropy. For the

$T$-length spectrum observation sequence $\mathbf{y}$, the Lempel-Ziv entropy estimator is defined as

$$E_f \approx \left( \frac{1}{T} \sum_t \Lambda_t \right)^{-1} \ln T,$$

where $\Lambda_t$ represents the length of the shortest subsequence that starts at the $t$th time slot and does not appear from time slot 1 to time slot $t - 1$.

We define the predictability $\Phi$ as the probability that an appropriate predictive algorithm can correctly predict the future spectrum state of a frequency band. By invoking Fano’s inequality as in [15], this quantity is upper bounded by the maximum predictability $\Phi_{\max}$, i.e., $\Phi \leq \Phi_{\max}$. Moreover, we can obtain the maximum predictability $\Phi_{\max}$ of frequency band $f$ through numerical calculation as

$$E_f = -\Phi_{\max} \log_2 (\Phi_{\max}) - (1 - \Phi_{\max}) \log_2 (1 - \Phi_{\max}) + (1 - \Phi_{\max}) \log_2 (Q_f - 1),$$

where $Q_f$ represents the number of distinct observation values.

As an illustrative example, Fig. 1 shows the distribution of maximum predictability obtained from a real-world spectrum dataset. The dataset comes from the RWTH Aachen University spectrum measurement campaign [30]. It contains one-week-length power spectral density (PSD) data sequences of total 14746 sampling frequency points in the 300MHz~3000MHz frequency region. The size of the dataset is (339380, 14746). In the following, we first perform data preprocessing. We then determine the entropy and maximum predictability of each frequency point, and finally obtain the distribution of the maximum predictability over all frequency points. Some data preprocessing operations, including resampling and quantization, are quite essential to make it convenient for analysis. In order to speed up the calculation process, we resample the PSD data sequences in the original dataset with a 3 minutes inter-sample time. That is, we obtain a new dataset $D_1$ of size (3393, 14746) by resampling the original dataset. Due to
the high variability of the measured PSD values, we quantize the measured PSD values of each individual frequency point and replace them with corresponding symbols to facilitate further processing. Specifically, we take the maximum value and minimum value of the original dataset as the upper and lower bound respectively, and quantize the PSD values in $D_1$ into 8 levels equally. The measured PSD values inside each quantization interval will be represented by a corresponding symbol, i.e., 1, 2, 3, 4, 5, 6, 7, 8. For the comparison purpose, we also obtain the maximum predictability of independent identical distributed (i.i.d.) Gaussian noise data sequences with the same size of $D_1$ and plot their distribution. From Fig. 1, we observe that the maximum predictability of real-world spectrum observations are much higher than that of the i.i.d. Gaussian noise data. This demonstrates that evolution of the frequency bands are highly predictable owing to their distinct temporal correlations or regularities. On the other hand, we also see that the maximum predictability $\Phi_{\max}$ of each frequency point is lower than 1, which means that regardless of how good our predictive algorithms are, the evolution of the frequency bands cannot be predicted perfectly with no error. Besides, we observe that the average value of $\Phi_{\max}$ over all sampling frequency points is 0.97. It indicates that the evolution of the frequency bands can potentially be correctly predicted from $D_1$ with an accuracy of 0.97.

To characterize the impact of missing data and anomalies on predicting the evolution of the frequency bands, we set up another new dataset $D_2$ by artificially injecting missing data and anomalies into the original real-world spectrum dataset from the RWTH Aachen University spectrum measurement campaign. We first inject anomalies to a portion of randomly selected measured PSD values following the standard anomaly injection method used in existing works [23], [31]. To be specific, we use the exponential weighted moving average method to predict the future entries based on their past values (i.e., $y_t = \alpha y_{t-1} + (1 - \alpha) y_{t-1}$, where $\alpha = 0.8$ in our experiments) and use the maximum difference between the actual and predicted values as the anomaly size to be injected. We vary the fraction of entries to injected anomalies from 5% to 30%. Then we randomly drop a portion of PSD values to simulate the missing data. We vary the fraction of entries to be deleted from 10% to 90%. A noteworthy point is that data preprocessing is also performed with the new dataset $D_2$ in determining the maximum predictability $\Phi_{\max}$. Fig. 2 presents the average maximum predictability over all sampling frequency points for different percentages of injected anomalies and missing data. From the figure we observe that the average maximum predictability decreases significantly with the increase of the percentage of missing data and the percentage of injected anomalies. It indicates that the existence of missing data and anomalies degrades the predictability of the frequency bands due to their destructive effects on the temporal correlation and temporal regularities of the frequency bands. Considering the fact that high prediction accuracy can only be achieved by predictive algorithms when the evolution of a frequency band is highly predictable (i.e., with large maximum predictability), we can also draw a conclusion that the performance of spectrum prediction algorithms are limited by the missing data and anomalies.

So far, we have shown how highly predictable the real-world frequency bands are and the negative effects of missing data and anomalies on spectrum prediction. These findings motivate us in the next section to develop a robust online spectrum data recovery framework to tackle the challenges introduced by missing data and anomalies for spectrum prediction.

III. ROBUST ONLINE ALGORITHM DESIGN

In this section, we first mathematically formulate the optimization problem of our proposed robust online spectrum data recovery framework. Then, we provide detailed solutions and an algorithm to efficiently solve the optimization problem in an online manner.

A. PROBLEM FORMULATION

The objective of spectrum data recovery is to infer the missing or anomalous spectrum data from the obtained incomplete and corrupted spectrum observations. To achieve this objective, we should robustly model the underlying latent structure of the spectrum data. Even if we can keep a matrix form of the spectrum observations similar like [23] to capture the spectral and temporal dependencies among multiple frequency bands, we cannot deal with the similarity among partial temporal variations inside each frequency band. Considering the fact that evolution of frequency bands usually have periodic and seasonality characteristics accompanied with relatively large noise and fluctuation, a novel analysis method uncovering latent time-directional structure inside each frequency band is required. Thus, a hankel matrix representation with the tensor structure plays a crucial role by exploiting a three-dimensional model against noisy and fluctuated signals. To this end, we propose to transfer the spectrum observations into a three-dimensional hankelized time-structured spectrum tensor and analyze it.
E represents the signal of interest, \( Y \) model the spectrum tensor as a mixture of all these effects. In the following, we further can efficiently describe the spectral-temporal dependencies on observations in the spectrum tensor \( Y \), the Hankel matrix \( Y \), the data collection processes. To characterize the missing spectrum sensing capacities, and the transmission losses in such as the mobility of the spectrum sensors, the limited storage from the incomplete and corrupted spectrum tensor \( \mathcal{X} \), we use the set \( \mathcal{P}_\Theta (\cdot) \) to indicate the corresponding entries. We set the entries of \( \mathcal{Y} \) that are not in \( \Theta \) to zero and keep the rest unchanged. Based on the notation, the spectrum tensor can be further expressed as \( \mathcal{P}_\Theta (\mathcal{Y}) \).

To efficiently model the signal subspace \( \mathcal{X} \) from the incomplete and corrupted spectrum tensor \( \mathcal{P}_\Theta (\mathcal{Y}) \), an optimization problem can be formulated by leveraging the inherent low-rank structure of \( \mathcal{X} \) and the sparsity property of \( \mathcal{Y} \) as

\[
\min_{\mathcal{X}, \mathcal{Y}} \text{rank} (\mathcal{X}) + \rho \| \mathcal{Y} \|_0, \\
\text{s.t.} \quad \| \mathcal{P}_\Theta (\mathcal{Y} - \mathcal{X}) \|_F^2 \leq \varepsilon, \quad \text{and} \quad (\mathcal{X})_{y_1,...,y_T} \in \mathcal{S}_H, f \in \{1, \ldots, F\}, \tag{5}
\]

where \( \text{rank} (\cdot) \) stands for the rank of a tensor. \( \rho \) is a positive rank-sparsity controlling parameter, and \( \varepsilon \) is a noise-related parameter. \( (\mathcal{X})_{y_1,...,y_T} \in \mathcal{S}_H \) represents the constraint of each horizontal slice matrix of \( \mathcal{X} \) with the Hankel structure. It should be noted that both rank and 0-norm criteria are in general NP-hard to optimize. To make the optimization problem in (5) tractable, one efficient way is to replace \( \text{rank} (\cdot) \) with its tightest convex surrogate, the nuclear norm \( \| \cdot \|_* \), and replace \( \| \cdot \|_0 \) with its closest convex surrogate, the \( l_1 \)-norm ||.|.1. However, it can be easily observed that the replacements result in a non-smooth optimization problem since both \( ||.|_* \) and \( ||.|_1 \) are not differentiable at the origin, and the size of the optimization problem can become quite large with time. To tackle the above challenges, CP decomposition is introduced in this paper to decompose the signal tensor \( \mathcal{X} \) by leveraging its low-rank property as \([33]\)

\[
\mathcal{X} = \sum_{r=1}^{R} a_r \circ b_r \circ c_r = \| A, B, C \|, \tag{6}
\]

where \( R \) is the rank of \( \mathcal{X} \), \( a_r \in \mathbb{R}^F, b_r \in \mathbb{R}^K \) and \( c_r \in \mathbb{R}^W \) denote the rank-one vectors. \( A = [a_1, a_2, \ldots, a_R] \in \mathbb{R}^{F \times R}, B = [b_1, b_2, \ldots, b_R] \in \mathbb{R}^{K \times R} \) and \( C = [c_1, c_2, \ldots, c_R] \in \mathbb{R}^{W \times R} \) represent three subspace factor matrices, and they can also be denoted by row vectors as \( A = [a_1^\top, a_2^\top, \ldots, a_R^\top]^\top, B = [b_1^\top, b_2^\top, \ldots, b_R^\top]^\top \) and \( C = [c_1^\top, c_2^\top, \ldots, c_R^\top]^\top \), where \( [a_1^\top, b_1^\top, c_1^\top] \in \mathbb{R}^{1 \times R} \). Then, we can obtain the transformed problem formula of (5) as

\[
\min_{A, B, C, \mathcal{Y}} \frac{1}{2} \left( \mathcal{P}_\Theta \right)_{\mathcal{Y}} \left( \mathcal{Y} - \mathcal{X} - \mathcal{Y} \right)_{\mathcal{F}}^2 + \frac{\mu_F}{2} \left( \| A \|_F^2 + \| B \|_F^2 + \| C \|_F^2 \right) + \mu_s \| \mathcal{Y} \|_1, \tag{7}
\]

\( s.t. \quad (\mathcal{X})_{y_1,...,y_T} = A_{y_1,...,y_T}^\top \circ B_{y_1,...,y_T}^\top \circ C_{y_1,...,y_T}^\top, \quad f \in \{1, \ldots, F\}, \quad \text{hankel structure constraint} \)
setting of subspace learning can also help to adapt dynamically to the new patterns in the spectrum observations. To this end, we tackle an online tensor completion problem to estimate the CP factor matrices \( \{A, b, C\} \) and the anomaly component matrix \( V \) at each time slot \( t \) by considering the exponential weighted least squares cost function as

\[
\min_{A, b, C, V} \sum_{t=1}^{T-t} \frac{1}{2} P_{b_it}(Y_t - \text{Adiag}(b^t)C' - V_t)^2
\]

Online-based

\[
\frac{\mu_r}{2} \left\| P_{b_{i}\tau},1:W-1 \left( \text{Adiag}(b^{\tau-1})C' \right)_{r,1:W-1} \right\|^2_F + \left\| (Adiag(b^{\tau-1})C')_{r,1:W-1} \right\|^2_F
\]

Hankel structure error

\[
+ \frac{\mu_r \tau_t}{2 \sum_{t=1}^{T-t}} \left\| (A\|b\|_2^2 + \|C\|_F^2) + \frac{\mu_r \tau_t}{2} \right\| b^t \|^2_2
\]

Frobenius and \( l_2 \) norm regularizer

\[
+ \mu_s \tau_t \|V_t\|_1
\]

(8)

where \( P_{b_it}(\cdot) \) represents the sampling operator for the \( \tau \)-th frontal slice matrix of the spectrum tensor. \( \mu_r \geq 0 \) is a hankel structure error-controlling parameter. \( \{Y_t, V_t\} \in \mathbb{R}^{F \times W} \) are the \( \tau \)-th frontal slice matrices of \( \mathcal{Y} \) and \( \mathcal{V} \), respectively. \( 0 < \lambda < 1 \) is the so-called forgetting factor, which enables observations in distant past to be forgotten with exponentially decaying and the more importance to be assigned to recent observations. In the case of infinite memory \( \lambda = 1 \), the formulation (8) coincides with the batch estimator (7). It should be noted that, for the hankel structure constraint of \( S_H \), we consider only two successive slice matrices to avoid model re-construction and diagonal-averaging using \( \{b^t\}^t_{t=1} \) in the past.

### B. SOLUTIONS AND ALGORITHMIC DESIGN

In this subsection, we give the detailed solutions of the minimization problem in (8), where the optimization variables are \( \{A, b, C\} \) and \( V \). It is readily seen that the problem is not convex for the joint variables. However, it is convex for each individual variable when the other variables are fixed. Thus, we take an alternating direction minimization procedure to successively optimize one variable with the others fixed under the framework of ADMM [27].

1) **UPDATE \( b [t] \)**

We calculate \( b [t] \) via an \( l_2 \)-norm regularized least square problem, which has a closed-form solution. With fixed \( \{A [t-1], C [t-1]\} \) derived at time slot \( t-1 \), we carry out the minimization with respect to \( b [t] \) as

\[
b [t] = \arg \min_{b} \frac{1}{2} \left[ \sum_{t=1}^{T-t} \left\| P_{b_it}(Y_t - \text{Adiag}(b^t)C - V_t) \right\|_F^2 + \mu_r \left\| \begin{bmatrix} \omega_t \otimes (Y_t - A[t-1] \text{diag}(b)C[t-1]) - V_t \end{bmatrix} \right\|_F^2 + \mu_s \left\| b^t \right\|_2^2 \right]
\]

(9)

where \( \omega_t \in \mathbb{R}^{F \times W} \) denotes a binary \( \{0, 1\} \)-matrix with \( (\omega_t)_{f,w} = 1 \) if \( (\mathcal{Y})_{f,w,t} \) is observed, and \( (\omega_t)_{f,w} = 0 \) otherwise. Defining \( F (b [t]) \) as the inner objective to be minimized in (9), we can obtain the optimal \( b [t] \) by differentiating \( F (b [t]) \) such that \( \frac{\partial F (b [t])}{\partial b} = 0 \), i.e.,

\[
F (b [t]) = \left[ \sum_{t=1}^{T-t} \left( \mu_r \tau_t \|I_R + \sum_{f=1}^{F} \sum_{w=1}^{W} (\omega_t)_{f,w} (g_{f,w} [t])' g_{f,w} [t] \right)^{-1} \right] \left[ \sum_{t=1}^{T-t} \left( \mu_r \tau_t \|I_R + \sum_{f=1}^{F} \sum_{w=1}^{W} (\omega_t)_{f,w} (g_{f,w} [t])' g_{f,w} [t] \right)^{-1} \right] \left[ \sum_{t=1}^{T-t} \left( \mu_r \tau_t \|I_R + \sum_{f=1}^{F} \sum_{w=1}^{W} (\omega_t)_{f,w} (g_{f,w} [t])' g_{f,w} [t] \right)^{-1} \right]
\]

(10)

where \( g_{f,w} [t] = a^t [t-1] \otimes e^w [t-1] \) and \( I_R \) is an identity matrix of size \( R \).

2) **UPDATE \( V [t] \)**

Introducing \( M [t] = Y_t - A [t-1] \text{diag}(b [t]) (C [t-1])' \) into (8), we further carry out the minimization with respect to \( V [t] \) as

\[
V [t] = \arg \min_{V} \frac{1}{2} \left\| \omega_t \otimes (M [t] - V) \right\|_2^2 + \mu_s \left\| V \right\|_1
\]

(11)

Here, we decompose the inner objective to be minimized in (11) into a parallel set of smaller sub-objectives, one for each row of \( V [t] \) as

\[
v_f [t] = \arg \min_{v_f} \frac{1}{2} \left\| \text{diag} (\omega_f) (M [t] \otimes v_f) + \mu_s \left\| v_f \right\|_1 \right\|_2^2
\]

(12)

Obviously, these subproblems are typical lasso regression problems, and an efficient solver can be found in [34].
3) UPDATE $A [t]$ AND $C [t]$

In the next step of the alternating direction minimization procedure, we update $A [t]$ and $C [t]$ alternatively by addressing a second-order stochastic gradient based on the recursive least square method with forgetting parameters. We first fix $\{b, C, V\}$ and carry out the minimization in (8) with respect to $A$ as

$$
\min_{A} \frac{1}{2} \sum_{\tau=1}^{t} \lambda^{t-\tau} \cdot \left\| (\omega_{\tau} \otimes (Z_{\tau} - A\text{diag}(b[\tau])(C[\tau-1]))) \right\|_{F}^{2} + \mu_{h}[\tau] \left\| (\omega_{\tau})_{:,1:w-1} \otimes \left[ (\text{diag}(b[\tau-1])(C[\tau-1]))_{:,2:w} - (A\text{diag}(b[\tau])(C[\tau-1]))_{:,1:w-1} \right] \right\|_{F}^{2} + \frac{\mu_{r}[\tau]}{2} \| A \|_{F}^{2},
$$

(13)

where $Z_{\tau} = Y_{\tau} - V[\tau]$. The objective function in (13) can be decomposed into a parallel set of smaller problems, one for each row $d'$ of $A$ as [33]

$$
\min_{d'} \frac{1}{2} \sum_{\tau=1}^{t} \sum_{w=1}^{W} \left[ \lambda^{t-\tau} \cdot (\omega_{\tau})_{f,w} (Z_{\tau})_{f,w} - d' \text{diag}(b[\tau])(c^{w}[\tau-1])) \right]^{2} + \mu_{h}[\tau] \sum_{w=1}^{W-1} \lambda^{t-\tau} \left\| (\omega_{\tau})_{f,w} d'[\text{diag}(b[\tau-1])(c^{w+1}[\tau-1])'] - \text{diag}(b[\tau])(c^{w}[\tau-1]) \right\|_{F}^{2} + \frac{\mu_{r}[\tau]}{2} \| d' \|_{F}^{2}, f \in \{1, \ldots, F\}.
$$

(14)

Here we denote $\text{diag}(b[\tau]) (c^{w}[\tau-1])'$ and $\text{diag}(b[\tau]) (c^{w+1}[\tau-1])' - \text{diag}(b[\tau]) (c^{w}[\tau-1])'$ in (14) as $\alpha_{w}[\tau]$ and $\beta_{w}[\tau]$, respectively. By setting the derivative of (14) regarding $d'$ equal to zero, the optimal $d'[\tau]$ can be obtained as

$$
da'[\tau] = \left( (G_{f}[\tau])^{-1} S_{f}[\tau] \right)'.
$$

(15)

where $G_{f}[\tau]$ and $S_{f}[\tau]$ are calculated by updating their previous values as

$$
G_{f}[\tau] = \lambda G_{f}[t-1] + \sum_{w=1}^{W} (\omega_{\tau})_{f,w} \alpha_{w}[\tau](\alpha_{w}[\tau])' + \mu_{h}[\tau] \sum_{w=1}^{W-1} (\omega_{\tau})_{f,w} \beta_{w}[\tau](\beta_{w}[\tau])' + (\mu_{r}[\tau] - \lambda \mu_{r}[\tau-1]) I_{R},
$$

(16)

$$
S_{f}[\tau] = \lambda S_{f}[t-1] + \sum_{w=1}^{W} (\omega_{\tau})_{f,w}(Z_{\tau})_{f,w} \alpha_{w}[\tau].
$$

(17)

The derivations of (15), (16) and (17) are detailed in Appendix A. Likewise, as for $C [t]$, the minimization problem in (8) is reformulated as

$$
\min_{C} \frac{1}{2} \sum_{\tau=1}^{t} \lambda^{t-\tau} \cdot \left( \| \omega_{\tau} \otimes (Z_{\tau} - A[\tau]\text{diag}(b[\tau])(C[\tau])) \|_{F}^{2} + \mu_{h}[\tau] \left\| (\omega_{\tau})_{:,1:w-1} \otimes \left[ (A[\tau]\text{diag}(b[\tau-1])(C[\tau]))_{:,2:w} - (A[\tau]\text{diag}(b[\tau])(C[\tau]))_{:,1:w-1} \right] \right\|_{F}^{2} + \frac{\mu_{r}[\tau]}{2} \| C \|_{F}^{2}. \right.
$$

(18)

Similarly, we decompose the objective function in (18) into a set of smaller problems, one for each row $c^{w}$ of $C$ as given in (19), as shown at the bottom of the next page. Here we denote $a'[\tau] \text{diag}(b[\tau])$ and $a'[\tau] \text{diag}(b[\tau-1])$ in (19) as $y_{\tau}[\tau]$ and $\eta_{\tau}[\tau]$, respectively. It should be emphasized that the second term representing the hankel structure error is not included for the last row of $C[\tau]$. In addition, since $c^{w+1}[\tau]$ is needed for the calculation of $c^{w}[\tau]$, these subproblems cannot be solved in parallel but in the descending order of $w$. Finally, the optimal $c^{w}[\tau]$ is obtained by solving the minimization problem in (19) as

$$
c^{w}[\tau] = \left( (J_{w}[\tau])^{-1} U_{w}[\tau] \right)',
$$

(20)

where $J_{w}[\tau]$ and $U_{w}[\tau]$ are calculated by updating their previous values as (21) and (22) shown at the bottom of the next page.

After the above mathematical derivations, the overall algorithm to solve the minimization problem in (8) is summarized in Algorithm 1, where an estimated signal subspace $\hat{X}$ is finally accomplished by eliminating the negative effects of missing data, and anomalies $V$.

C. COMPUTATIONAL COMPLEXITY ANALYSIS

This subsection is devoted to show the computational complexity per iteration of the proposed algorithm. The calculation of $b[\tau]$ in (10) corresponds to the computation in line 3 of Algorithm 1, where the dominant cost includes a $R \times R$ matrix inversion with a computational cost $O(R^{3})$, the vector multiplication with a computational cost $O(\|\omega_{\tau}\|_{0}R^{2})$, and the vector addition with a computational cost $O(\|\omega_{\tau}\|_{0}R)$. The estimation of $V[\tau]$ in (11) corresponds to the computation in line 4 of Algorithm 1, where a Lasso estimator is used to calculate each row of $V[\tau]$ with a computational cost $O(W^{3})$, thus a total of $O(FW^{3})$ is required. As for the calculation of $A[\tau]$ in lines 5 ~ 10 of Algorithm 1, each row of $A[\tau]$ requires $O(\|\omega_{\tau}\|_{0}R^{2}) + O(\|\omega_{\tau}\|_{0}R) + O(R^{3})$ for vector multiplication, addition, and matrix inverse in (15) ~ (17). Thus, a total of $O(\|\omega_{\tau}\|_{0}R^{2}) + O(\|\omega_{\tau}\|_{0}R) + FW^{3}R^{3}$ is required for the calculation of $A[\tau]$. Similarly, the calculation of $C[\tau]$ in lines 11 ~ 16 of Algorithm 1 requires $O(3\|\omega_{\tau}\|_{0}R^{2}) + O(\|\omega_{\tau}\|_{0}R) + WFW^{3}R^{3}$. Consequently, the total computational complexity at each iteration in Algorithm 1 results in $O((1 + F + W)R^{3}) + O(3\|\omega_{\tau}\|_{0}R^{2}) + O(\|\omega_{\tau}\|_{0}R) + O(FW^{3})$. It indicates that the length of the
TABLE 2. The ‘IN Subband I’ Dataset.

| Name                      | IN Subband I |
|---------------------------|-------------|
| Type of data              | PSD over frequency |
| Frequency center          | 770MHz      |
| Frequency span            | 1500MHz     |
| Number of sampling points | 8192        |
| Number of measurements    | 259666      |
| Sampling frequency resolution | 200kHz    |
| Measurement resolution    | 1.8s        |

IV. EXPERIMENTAL EVALUATIONS

Data preparation: In this section, we use the real-world spectrum dataset from the RWTH Aachen University spectrum measurement campaign [30] for the evaluation of the proposed SDR-HTC algorithm. Specifically, we use the ‘IN Subband I’ dataset for all the following experiments. This dataset records the measured PSD-values over 8192 frequency points for one-week duration. Some fundational measurement parameters are shown in Table 2. To characterize the impact of anomalies and missing data, we artificially inject anomalies to a portion of entries in the original dataset and drop elements in the dataset uniformly and independently using the same way as described in Section II. We use $p_{\text{miss}}$ and $p_{\text{anomaly}}$ to respectively denote the missing data percentage and the anomaly data percentage.

Parameter setup: For the proposed SDR-HTC algorithm, the chosen ranges of the rank-, sparsity-, and Hankel structure error-controlling parameters are derived from the optimality conditions for (8), which indicate that for $\mu_r$, $\mu_s$ and $\mu_h$ selected from the corresponding ranges, an optimal solution can always be found. Along the lines of [23], [33], we empirically set the penalty parameters as $\mu_r = \mu_h = 10^{-3}$ and $\mu_s = 6 \times 10^{-2}$. We also set the forgetting factor $\lambda = 0.7$ to enable observations in the past to be forgotten with exponentially decaying. The estimated rank is set as $R = 10$.

Evaluation metric: To quantitatively evaluate the spectrum data recovery performance of our proposed SDR-HTC algorithm, we adopt one of the most widely recognized evaluation matrices, i.e., the normalized root squared error (RMSE) in dB to quantify the magnitude of error as

$$RMSE = 10 \log_{10} \left( \frac{1}{N} \sum_{n=1}^{N} (\hat{x}_n - x_n)^2 \right)$$

where $x_n$ denotes the ground-truth to the estimated value $\hat{x}_n$.

In the following, we first conduct an experiment to compare our proposed SDR-HTC algorithm with the state-of-the-art spectrum data recovery methods under different missing data percentages and anomaly data percentages. Then, we perform predictability analysis of the recovered spectrum data obtained by our proposed SDR-HTC algorithm compared with the unrecovered spectrum data. Finally,
Algorithm 1 Robust Online Spectrum Data Recovery Based on Hankelized Tensor Completion (SDR-HTC)

Require: \{Y_t, \omega_t\}_{t=1}^{T}, \mu_r, \mu_s, \mu_h, \lambda;
1: Initialize: Set \(A[0]\) and \(C[0]\) at random, while \(b[0] = 0\) and \(V[0] = 0\). The temporally used variables are set as \(G_f[0] = 0, S_f[0] = 0, J_w[0] = 0\) and \(U_w[0] = 0\).
2: for \(t = 1, 2, \ldots, T\) do
3:   Update the projection coefficient vector \(b[t]\); (10)
4:   Calculate the anomaly component \(V[t]\); (12)
5:   Update the subspace factor matrix \(A[t]\);
6:   for \(f = 1, 2, \ldots, F\) do
7:     Update \(G_f[t]\); (16)
8:   end for;
9:   Obtain \(a'[t]\); (15)
10: end for;
11: Update the subspace factor matrix \(C[t]\);
12: for \(w = W, W-1, \ldots, 1\) do
13:   Update \(J_w[t]\); (21)
14: end for;
15: Update \(U_w[t]\); (22)
16: end for;
17: Obtain \((\hat{X})_{::,t} = A[t] \text{diag}(b[t]) \text{ (C[t])'}\);
18: Obtain \((\hat{Y})_{::,t} = V[t]\);
19: end for;
20: Output: The estimated signal subspace \(\hat{X}\). The estimated spectrum anomaly \(\hat{V}\);

we perform spectrum prediction with various predictive algorithms on the recovered spectrum data to verify the effectiveness of our proposed SDR-HTC algorithm for spectrum prediction. We run all the experiments on a laptop computer with 2.20 GHz Intel Core i7 CPU and 16 GB RAM using MATLAB R2018b.

A. COMPARISON WITH THE STATE-OF-THE-ART SPECTRUM DATA RECOVERY METHODS

In this subsection, we compare our proposed SDR-HTC algorithm with several candidate baseline spectrum data recovery methods in the literature, which can be broadly grouped into three classes: 1-D time-series-based spectrum data recovery methods in time domain only; 2-D matrix-based spectral-temporal spectrum data recovery methods and 3-D tensor-based spectral-temporal spectrum data recovery methods. For the 1-D baseline methods, we select the well-known MAF (Moving Average Filter) interpolation method [18], SES (Single Exponential Smoothing) method [19] and CS (Compressive Sensing) method [35]. They are the earliest yet very effective spectrum data recovery methods, but they cannot deal with the corrupted data. For the 2-D baseline methods, we select the MCMR (matrix completion and matrix recovery) algorithm in [23]. MCMR is a robust online spectrum data recovery method, which also utilizes ADMM for solving matrix completion problem. It provides valuable intuitions for the optimization part in this paper. For the 3-D baseline methods, we select the most recent work in literature, i.e., HaLRTC (high accuracy low rank tensor completion) algorithm in [24]. Notably, the proposed SDR-HTC framework is quite different from HaLRTC at least in two aspects: i) SDR-HTC is robust to anomalies, while HaLRTC is designed without considering the impact of anomalies; ii) SDR-HTC works in an online manner, while HaLRTC is a batch algorithm.

1) RECOVERY ACCURACY

Fig. 4 compares the spectrum data recovery error of different methods under a missing data percentage from 0.1 to 0.9 in terms of RMSE. We evaluate the case where the fraction of entries to inject anomalies is \(p_{\text{anomaly}} = 0.1\) and the length of the sliding window for our framework is \(W = 100\). We can see that our proposed SDR-HTC algorithm has a higher recovery accuracy than the other five algorithms. Especially, during the interval between \(p_{\text{miss}} = 0.3\) to \(p_{\text{miss}} = 0.7\), our framework reveals significantly higher recovery capability. We also observe that the recovery error of all methods increase with the percentage of missing data and the SDR-HTC always keeps at a relatively lower level. In the next, we set the missing data percentage as \(p_{\text{miss}} = 0.2\) and investigate the recovery performance under different corruption percentages (from 0.05 to 0.3). The results are shown in Fig. 5. Similar as the observations in Fig. 4 that the proposed SDR-HTC always outperforms the other five algorithms. However, the recovery RMSE of the SDR-HTC is not quite sensitive with an increasing percentage of the injected anomaly data. This result demonstrates the robustness of the proposed framework against sparsely-distributed anomalies.
2) COMPUTATION TIME
Fig. 6 compares the computation time of different spectrum data recovery methods. We evaluate the case where the percentage of missing data is $p_{\text{miss}} = 0.2$ and the percentage of injected anomalies is $p_{\text{anomaly}} = 0.1$. Fig. 6 reveals that the proposed SDR-HTC algorithm is super fast compared to CS and HaLRTC.

In summary, the above experimental results suggest that our algorithm can achieve better recovery performance and computation efficiency with both partial and corrupted observations.

3) WINDOW LENGTH
The impact on the different window lengths $W = [10, 200]$ for the hankel structure of our proposed SDR-HTC algorithm are evaluated under different missing data percentages and different anomaly data percentages in Fig. 7. More concretely, we evaluate the case where $p_{\text{anomaly}} = 0.1$ and $p_{\text{miss}} = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ in Fig. 7(a) and the case where $p_{\text{miss}} = 0.2$ and $p_{\text{anomaly}} = \{0.1, 0.2, 0.3\}$ in Fig. 7(b). It is shown in Fig. 7(a) that our method appears an obvious “thresholding phenomena” that the recovery RMSE is continuously decreased when the window length increases below a certain threshold (e.g. $W = 50$ for the case of $p_{\text{miss}} = 0.1$). It indicates that there exists an optimal window length for the proposed SDR-HTC algorithm to achieve the highest spectrum data recovery accuracy. We can also see that the threshold is bigger when more spectrum observations are lost. A similar “thresholding phenomena” is also observed in Fig. 7(b). However, the threshold only slightly increases when adding more corrupted data. Additionally, as it has been analysed in Section III-D that the window length $W$ is dominant in deciding the computation complexity of our algorithm and a larger window length $W$ is required to achieve the optimal performance for more missing data and anomalies, it can also be concluded that the computation time of our algorithm will increase with the percentage of missing data and anomalies to maintain the optimal performance.

B. COMPARISON OF MAXIMUM PREDICTABILITY
The degree to which any predictive algorithm can correctly predict the evolution of a frequency point is $\Phi_{\text{max}}$. In this subsection, we compare the maximum predictability of the unrecovered spectrum data sequences with that of the recovered spectrum data sequences obtained by our proposed SDR-HTC algorithm. A noteworthy point is that we should ignore the first 10% recovery results when calculating the maximum predictability. This is because the first 10% results
cannot reflect the real effectiveness of SDR-HTC due to the instability of the online subspace learning algorithm at the very beginning. Fig. 8 illustrates the comparison for a special case where $p_{\text{miss}} = 0.5$ and $p_{\text{anomaly}} = 0.2$. We can see that for any frequency point in the $20\text{MHz} \sim 1520\text{MHz}$ frequency region, the maximum predictability of the recovered spectrum data sequence always outperforms that of the unrecovered spectrum data sequence. It indicates that our SDR-HTC algorithm can effectively alleviate the effects of missing values and anomalies for spectrum prediction and improve the predictability of the frequency points. Fig. 9 and Fig. 10 further compare the maximum predictability distribution of the unrecovered spectrum data sequences and the maximum predictability distribution of the recovered spectrum data sequences under different missing data percentages and different anomaly data percentages. We first determine the maximum predictability $\Phi_{\text{max}}$ of each spectrum data sequence and then obtain the distribution of the maximum predictability $\Phi_{\text{max}}$ over all spectrum data sequences in each case. Fig. 9 depicts the maximum predictability distribution of the unrecovered spectrum data sequences and the maximum predictability distribution of the recovered spectrum data sequences under different missing data percentages; whereas Fig. 10 presents the maximum predictability distribution of the unrecovered spectrum data sequences and the maximum predictability distribution of the recovered spectrum data sequences under different anomaly data percentages.

The above conclusions obtained from Fig. 8 are strengthened in Fig. 9, where we also observe higher maximum predictabilities of the recovered spectrum data sequences over that of the unrecovered spectrum data sequences. It indicates that we can reach a higher potential predictive accuracy if using the recovered spectrum data sequences with the predictive algorithms for spectrum prediction. While the maximum predictability of the unrecovered spectrum data sequences gets lower with the increase of the percentage of the missing data, our recovered spectrum data sequences maintains stable and high maximum predictability, especially when $p_{\text{miss}} < 0.5$. A similar result applies to Fig. 10, where higher and stable maximum predictabilities of the recovered spectrum data sequences are observed for any percentage of the injected anomalies. Combining Fig. 8, Fig. 9 and Fig. 10, our proposed SDR-HTC algorithm has shown its effectiveness for approaching the limit of predictability in spectrum prediction tasks.

**C. COMPARISON OF PREDICTIVE ALGORITHMS**

In the last subsection, we compare the prediction accuracy of several predictive algorithms on both the unrecovered spectrum data sequences and the recovered spectrum data sequences obtained via our proposed SDR-HTC algorithm. We employ five predictors: 1) simple forecasting strategy: EWMA (Exponential Weighted Moving Average) [23], 2) time series regression: KNN (K-Nearest Neighbors) [25], 3) autoregressive method: SARIMA (Seasonal Autoregressive Integrated Moving Average) [36], 4) machine learning: MLP (Multilayer Perceptron) [37] and 5) deep recurrent neural network: LSTM (Long Short-Term Memory) [38]...
for predicting the frequency point evolution, and compare their performance given the value of the maximum predictability. We evaluate the case where $p_{\text{miss}} = 0.5$ and $p_{\text{anomaly}} = 0.2$, and randomly select a frequency point within the 20MHz $\sim$ 1520MHz frequency region for the following evaluation. For the selected frequency point, the maximum predictability of its unrecovered spectrum data sequence and the recovered spectrum data sequence obtained by our proposed SDR-HTC algorithm can be easily calculated by (3) as 0.9306 and 0.9563 respectively. Fig. 11 depicts the prediction accuracy of the five predictors on the recovered/unrecovered spectrum data sequences.

We can see that LSTM achieves the highest prediction accuracy among the five predictors by capturing the multiple features or hidden patterns of the frequency point. While LSTM achieves a spectrum prediction accuracy of 0.8689 with the unrecovered spectrum data sequence and a spectrum prediction accuracy as high as 0.9024 with the recovered spectrum data sequence, its performance is always limited by the maximum predictability $\Phi_{\text{max}}$, which corresponds to the definition of maximum predictability in Section II. We also find that a higher prediction accuracy is obtained when performing spectrum prediction with the recovered spectrum data sequence for any predictor. This finding in line with the observations in Section IV-B indicates that our proposed SDR-HTC algorithm can effectively recover the missing values from the corrupted observations so as to approach the limit of predictability as well as provide higher potential prediction accuracy for the spectrum prediction algorithms.

V. CONCLUSION

In this paper, we focus on recovering missing values from corrupted historical spectrum observations to approach the limit of predictability in spectrum prediction tasks. Through analyzing a real-world spectrum observation dataset, we observe that there is a high predictability of frequency bands. We also find that the high predictability is degraded when part of the spectrum observations is missing or corrupted by anomalies. Based on our findings, we propose to formulate a hankelized spectrum tensor to take advantage of the latent time-directional structure of the spectrum data, and design a robust online SDR-HTC algorithm to efficiently recover missing values and separate anomalies. We have done extensive experiments based on real-world spectrum observations to evaluate the performance of our proposed SDR-HTC algorithm. We show that SDR-HTC can achieve a higher accuracy in data recovery with lower computation cost compared with the baseline algorithms. Besides, as for spectrum prediction, SDR-HTC enables to retrieve the high predictability of frequency bands and to improve the performance of any predictor.

ACKNOWLEDGMENT

This article was presented in part at the IEEE/CIC International Conference on Communications, China, in 2020.

DERIVATION OF $A[f]$ BY RECURSIVE LEAST SQUARES

This appendix describes the derivation of $A[f]$. Defining $F(a'[f])$ as the inner objective to be minimized in (14), its derivative with regard to $a'[f]$ is calculated as

$$\frac{\partial F(a'[f])}{\partial (a'[f])} = \sum_{\tau=1}^T \left[ -\sum_{w=1}^W \lambda^{t-\tau}(\omega_t)_f,w (Z_t)_f,w - a' \alpha_w[\tau] \alpha_{w}[\tau] \right] + \mu_h[t] \sum_{w=1}^W \lambda^{t-\tau}(\omega_t)_f,w \beta_{w}[\tau] \beta_{w}[\tau] + \mu_r[t](a')' \right].$$

Then, by setting this derivative equal to zero and using

$$a'[\alpha_w[\tau] \alpha_{w}[\tau]] = \alpha_{w}[\tau]'(a'[f])', \beta_{w}[\tau] \beta_{w}[\tau] = \beta_{w}[\tau]'(\beta_{w}[\tau])'(a'[f])',$$ we get the following:

$$\left( \sum_{\tau=1}^T \sum_{w=1}^W \lambda^{t-\tau}(\omega_t)_f,w \alpha_{w}[\tau] \alpha_{w}[\tau]' \right) + \mu_h[t] \sum_{w=1}^W \lambda^{t-\tau}(\omega_t)_f,w \beta_{w}[\tau] \beta_{w}[\tau]' \right) + \mu_r[t] R(a')'.$$

Let’s define $G_f[t]$ and $S_f[t]$ respectively as

$$G_f[t] = \sum_{\tau=1}^T \sum_{w=1}^W \lambda^{t-\tau}(\omega_t)_f,w \alpha_{w}[\tau] \alpha_{w}[\tau]' \right) + \mu_h[t] \sum_{w=1}^W \lambda^{t-\tau}(\omega_t)_f,w \beta_{w}[\tau] \beta_{w}[\tau]' \right) + \mu_r[t] R(a')'.$$
By introducing (26) and (27) into (25), we get

\[ G_f[t] = \sum_{t=1}^{W} \left[ \sum_{w=1}^{W} \lambda^{t-\tau} \omega_r \alpha_w \lambda^{\tau} (\alpha_w \lambda^{\tau})' + \mu_z \sum_{w=1}^{W} \lambda^{t-\tau} \omega_r \beta_w \lambda^{\tau} (\beta_w \lambda^{\tau})' \right] + \mu_{t} I_R \]

\[ + \sum_{w=1}^{W} (\omega_r)_{f,w} \alpha_w [\alpha_w (\alpha_w)'] + \mu_h \sum_{w=1}^{W} (\omega_r)_{f,w} \beta_w [\beta_w (\beta_w)'] \]

\[ = \lambda \left[ \sum_{t=1}^{W} \sum_{w=1}^{W} \lambda^{t-1-\tau} (\omega_r)_{f,w} \alpha_w \lambda^{\tau} (\alpha_w \lambda^{\tau})' + \mu_h \sum_{w=1}^{W} \lambda^{t-1-\tau} (\omega_r)_{f,w} \beta_w \lambda^{\tau} (\beta_w \lambda^{\tau})' \right] + \mu_{t} [I_R - \lambda] \]

\[ + \sum_{w=1}^{W} (\omega_r)_{f,w} \alpha_w [\alpha_w (\alpha_w)'] + \mu_h \sum_{w=1}^{W} (\omega_r)_{f,w} \beta_w [\beta_w (\beta_w)'] + \mu_{t} [I_R - \lambda] \]

\[ = \lambda G_f [t - 1] + \sum_{w=1}^{W} (\omega_r)_{f,w} \alpha_w [\alpha_w (\alpha_w)'] + \mu_h \sum_{w=1}^{W} (\omega_r)_{f,w} \beta_w [\beta_w (\beta_w)'] + \mu_{t} [I_R - \lambda] \] (28)

By introducing (26) and (27) into (25), we get \( a_f[t] = \left( G_f[t] \right)^{-1} s_f[t] \). Here, \( G_f[t] \) and \( S_f[t] \) can be further transformed as (28), as shown at the top of the page and (29).

It is clear that both \( G_f \) and \( S_f \) can be obtained each time by updating their previous values. With the updated values, we finally get \( a_f[t] \) as (15).

\[ S_f[t] = \sum_{t=1}^{W} \sum_{w=1}^{W} \lambda^{t-\tau} (\omega_r)_{f,w} (Z_r)_{f,w} \alpha_w [\alpha_w \lambda^{\tau}] \] (27)

\[ = \lambda S_f [t - 1] + \sum_{w=1}^{W} (\omega_r)_{f,w} (Z_r)_{f,w} \alpha_w [\alpha_w \lambda^{\tau}] \] (29)

REFERENCES

[1] X. Xing, T. Jing, W. Cheng, Y. Hua, and X. Cheng, “Spectrum prediction in cognitive radio networks,” IEEE Wireless Commun., vol. 20, no. 2, pp. 90–96, Apr. 2013.

[2] G. Ding, Y. Jiao, J. Wang, Y. Zou, Q. Wu, Y.-D. Yao, and L. Hanzo, “Spectrum inference in cognitive radio networks: Algorithms and applications,” IEEE Commun. Surveys Tuts., vol. 20, no. 1, pp. 150–182, 1st Quart., 2018.

[3] Y. Zhao, Z. Hong, Y. Luo, G. Wang, and L. Pu, “Prediction-based spectrum management in cognitive radio networks,” IEEE Syst. J., vol. 12, no. 4, pp. 3303–3314, Dec. 2018.

[4] P. X. Zou, X. Wang, W. Linghu, R. Sun, T. Peng, and W. Wang, “Prediction-based spectrum access optimization in cognitive radio networks,” in Proc. IEEE 29th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC), Bologna, Italy, Sep. 2018, pp. 1–7.

[5] V.-D. Nguyen and O.-S. Shin, “Cooperative prediction-and-sensing-based spectrum sharing in cognitive radio networks,” IEEE Trans. Cognit. Commun. Netw., vol. 4, no. 1, pp. 108–120, Mar. 2018.

[6] H. Eltorn, S. Kandkepan, Y-C. Liang, and R. J. Evans, “Cooperative soft fusion for HMM-based spectrum occupancy prediction,” IEEE Commu. Lett., vol. 22, no. 10, pp. 2144–2147, Oct. 2018.

[7] H. Li, X. Ding, Y. Yang, Z. Xie, and G. Zhang, “Online spectrum prediction with adaptive threshold quantization,” IEEE Access, vol. 7, pp. 173425–173434, 2019.

[8] X. Yu, P. Zheng, C. Chen, Y. Li, and B. Vucetic, “Mobile collaborative spectrum sensing for heterogeneous networks: A Bayesian machine learning approach,” IEEE Trans. Signal Process., vol. 66, no. 21, pp. 5634–5647, Nov. 2018.

[9] L. Yu, Y. Guo, Q. Wang, C. Luo, M. Li, W. Liao, and P. Li, “Spectrum availability prediction for cognitive radio communications: A DCG approach,” IEEE Trans. Cognit. Commun. Netw., vol. 6, no. 2, pp. 476–485, Jun. 2020.

[10] J. Jiang, W. Yin, H. Li, E. Hossain, and Z. Han, “Collaborative spectrum sensing from sparse observations in cognitive radio networks,” IEEE J. Sel. Areas Commun., vol. 29, no. 2, pp. 327–337, Feb. 2011.

[11] L. Zhang, G. Ding, Q. Wu, Y. Zou, Z. Han, and J. Wang, “Byzantine attack and defense in cognitive radio networks: A survey,” IEEE Commu. Surveys Tuts., vol. 17, no. 3, pp. 1342–1363, 3rd Quart., 2015.

[12] P. Thakur, A. Kumar, S. Pandit, G. Singh, and S. N. Satashia, “Analysis of high-traffic cognitive radio network with imperfect spectrum monitoring technique,” Comput. Netw., vol. 147, pp. 27–37, Dec. 2018.

[13] J. Luo, T. Hong, and M. Yae, “Real-time anomaly detection for very short-term load forecasting,” J. Modern Power Syst. Clean Energy, vol. 6, no. 2, pp. 235–243, Mar. 2018.

[14] J. Mei, Y. De Castro, Y. Goude, J.-M. Azaiez, and G. Hebrail, “Nonnegative matrix factorization with side information for time series recovery and prediction,” IEEE Trans. Knowl. Data Eng., vol. 31, no. 3, pp. 493–506, Mar. 2019.

[15] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, “Limits of predictability in human mobility,” Science, vol. 327, no. 5968, pp. 1018–1021, Feb. 2010.

[16] G. Ding, J. Wang, Q. Wu, Y.-D. Yao, R. Li, H. Zhang, and Y. Zou, “On the limits of predictability in real-world radio spectrum state dynamics: From entropy theory to 5G spectrum sharing,” IEEE Commun. Mag., vol. 53, no. 7, pp. 178–183, Jul. 2015.

[17] J. Sun, L. Shen, G. Ding, R. Li, and Q. Wu, “Predictability analysis of spectrum state evolution: Performance bounds and real-world data analytics,” IEEE Access, vol. 5, pp. 22760–22774, 2017.

[18] W. Huang, P. Xu, Z. Zeng, N. J. G. Falkner, X. Li, and W. E. Zhang, “Recovering missing values from corrupted spatio-temporal sensory data via robust low-rank tensor completion,” in Database Systems for Advanced Applications, vol. 10177, Cham, Switzerland: Springer, 2017, pp. 607–622.

[19] X. Li, Y. Zheng, J. Zhang, and T. Li, “St-nvl: Filling missing values in geo-sensory time series data,” in Proc. 25th Int. Joint Conf. Artif. Intell., New York, USA, Jun. 2016, pp. 2704–2710.

[20] J. Li, G. Ding, X. Zhang, and Q. Wu, “Recent advances in radio environment map: A survey,” in Machine Learning and Intelligent Communications, vol. 226, Cham, Switzerland: Springer, 2018, pp. 247–257.
[21] X. Chen, J. Yang, and G. Ding, “Minimum Bayesian risk based robust spectrum prediction in the presence of sensing errors,” *IEEE Access*, vol. 6, pp. 29661–29665, 2018.

[22] A. Zaeemzadeh, M. Joneidi, B. Shahrasbi, and N. Rahmavard, “Missing spectrum-data recovery in cognitive radio networks using piecewise constant nonnegative matrix factorization,” in *Proc. IEEE Mil. Commun. Conf.*, Tampa, FL, USA, Oct. 2015, pp. 238–243.

[23] G. Ding, F. Wu, Q. Wu, S. Tang, F. Song, A. V. Vasilakos, and T. A. Tsiftsis, “Robust online spectrum prediction with incomplete and corrupted historical observations,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8022–8036, Sep. 2017.

[24] J. Sun, J. Wang, G. Ding, L. Shen, J. Yang, Q. Wu, and L. Yu, “Long-term spectrum state prediction: An image inference perspective,” *IEEE Access*, vol. 6, pp. 43489–43498, 2018.

[25] C. Ge, Z. Wang, and X. Zhang, “Robust long-term spectrum prediction with missing values and sparse anomalies,” *IEEE Access*, vol. 7, pp. 16655–16664, 2019.

[26] K. Xie, X. Li, X. Wang, G. Xie, J. Wen, and D. Zhang, “Graph based tensor recovery for accurate Internet anomaly detection,” in *Proc. IEEE Conf. Comput. Commun.*, Honolulu, HI, USA, Apr. 2018, pp. 1502–1510.

[27] N. D. Sidiroopoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, “Tensor decomposition for signal processing and machine learning,” *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

[28] K. Xie, X. Li, X. Wang, G. Xie, J. Wen, J. Cao, and D. Zhang, “Fast tensor factorization for accurate Internet anomaly detection,” *IEEE/ACM Trans. Netw.*, vol. 25, no. 6, pp. 3794–3807, Dec. 2017.

[29] K. Zhao, D. Kryushchev, and H. Vo, “Predicting taxi and uber demand in cities: Approaching the limit of predictability,” *IEEE Trans. Knowl. Data Eng.*, early access, Dec. 9, 2020, doi: 10.1109/TKDE.2019.2955686.

[30] M. Wellens, J. Riithijärvi, and P. Mahonen, “Empirical time and frequency domain models of spectrum use,” *Phys. Commun.*, vol. 2, nos. 1–2, pp. 10–32, Mar. 2009.

[31] Y.-C. Chen, L. Qu, Y. Zhang, G. Xue, and Z. Hu, “Robust network compressive sensing,” in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, Maui, HI, USA, 2014, pp. 545–556.

[32] T. Babaie, S. Chawla, and S. Ardon, “Network traffic decomposition for anomaly detection,” 2014, arXiv:1403.0157. [Online]. Available: http://arxiv.org/abs/1403.0157

[33] H. Kasui, W. Kellerer, and M. Kleinsteuber, “Network volume anomaly detection and identification in large-scale networks based on online time-structured traffic tensor tracking,” *IEEE Trans. Netw. Service Manage.*, vol. 13, no. 3, pp. 636–650, Sep. 2016.

[34] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *J. Roy. Stat. Soc., Ser. B Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996.

[35] Y. Arjouni and N. Kaabouch, “A comprehensive survey on spectrum sensing in cognitive radio networks: Recent advances, new challenges, and future research directions,” *Sensors*, vol. 19, no. 1, p. 126, Jan. 2019.

[36] D. B. Alencar, C. M. Affonso, R. C. L. Oliveira, and J. C. R. Filho, “Hybrid approach combining sarima and neural networks for multi-step ahead wind speed forecasting in brazil,” *IEEE Access*, vol. 6, pp. 55986–55994, 2018.

[37] J. Ge, D. Liu, L. Kong, and X. Chen, “A multilayer perceptron neural network-based spectrum prediction approach with gray decision,” *Commun., Signal Process., Syst.*, vol. 463, pp. 2395–2404, Jul. 2019.

[38] L. Yu, J. Chen, G. Ding, Y. Tu, J. Yang, and J. Sun, “Spectrum prediction based on taguchi method in deep learning with long short-term mem ory,” *IEEE Access*, vol. 6, pp. 45923–45933, 2018.

XI LI (Student Member, IEEE) received the B.E. degree from the School of Information Science and Engineering, Southeast University, China, where she is currently pursuing the Ph.D. degree. She was a Visiting Student with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY, USA, from November 2018 to October 2019. Her research interests include cognitive radio networks, vehicular ad hoc networks, and application of machine learning to wireless communication.

GUOJUN CHEN (Student Member, IEEE) received the B.E. and Ph.D. degrees in information engineering from Southeast University, Nanjing, China, in 2008 and 2016, respectively. He was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, from July 2014 to January 2015. He was also a Research Assistant and a Postdoctoral Fellow with the Institute of Network Coding, The Chinese University of Hong Kong, Hong Kong, from March 2016 to July 2017. Since August 2017, he has been with the School of Information Science and Engineering, Southeast University, where he is currently an Assistant Professor. His research interests include information theory, signal processing, and wireless communications.

YINFENG XU (Member, IEEE) received the B.E. and Ph.D. degrees in information engineering from Southeast University, Nanjing, China, in 2008 and 2016, respectively. He was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, from July 2014 to January 2015. He was also a Research Assistant and a Postdoctoral Fellow with the Institute of Network Coding, The Chinese University of Hong Kong, Hong Kong, from March 2016 to July 2017. Since August 2017, he has been with the School of Information Science and Engineering, Southeast University, where he is currently pursuing the Ph.D. degree. His research interests include vehicular ad hoc networks, source coding, and rate distortion.

XIN WANG (Senior Member, IEEE) received the B.S. degree in telecommunications engineering and the M.S. degree in wireless communications engineering from the Beijing University of Posts and Telecommunications, Beijing, China, and the Ph.D. degree in electrical and computer engineering from Columbia University, New York, NY. She was a Technical Staff Member in mobile and wireless networking with Bell Labs Research and Lucent Technologies. She is currently an Associate Professor with the Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY. Her research interests include sparse signal processing, machine learning, and big data analysis. She served on executive committee and a technical committee of numerous conferences and funding review panels. She received the NSF Career Award in 2005 and the ONR Challenge Award in 2010. She serves as an Associate Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING.

TIECHENG SONG (Member, IEEE) received the B.S. and M.S. degrees from the Department of Radio Engineering, Southeast University, Nanjing, China, in 1989 and 1992, respectively, and the Ph.D. degree from the School of Information Science and Engineering, Southeast University, in 2006. In 1992, he joined the Department of Radio Engineering, Southeast University, as a Research Associate. He was an Executive Vice President with the Nanjing Institute of Communication Technology. He is currently a Professor with the National Mobile Communication Research Laboratory, School of Information Science and Engineering, Southeast University. His research interests include 5G wireless systems, optical wireless communications, and cognitive radio.