IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 December 2022, accepted 19 January 2023, date of publication 31 January 2023, date of current version 10 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3241242

RESEARCH ARTICLE

Recovery of High Order Statistics of PSK Signals Based on Low-Rank Matrix Completion

ZHENGLI XING, JIE ZHOU, ZHAN GE, GUANQIN HUANG, AND MAOHAI HU
Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang 621900, China

Corresponding author: Maohai Hu (humaohai@163.com)

This work was supported in part by the National Science Foundation of China under Grant U1530154.

ABSTRACT High order statistics are useful for automatic modulation recognition and parameter estimations. In this paper, we cast the problem of recovering high order statistics of PSK signals taken from nonuniform compressive samples as a one of recovering a low-rank matrix from missing or corrupted observations. This is a new model to describe the high order statistics of PSK signals. Unlike traditional uniformly Nyquist samples, our method uses the advanced optimization technique, which is guaranteed to find the low-rank matrix by simultaneously fixing the missing entries. Simulation results demonstrate that our method achieves accurate estimates of the major portion of the high order statistics. The new technique can be used to fulfil automatic modulation recognition (AMR) and rough estimations of parameters. More specifically, low-rank structure of PSK signals is studied. In contrast to the existing l_1 optimization criteria, our method proposed here is computationally more efficient and provides high accuracy.

INDEX TERMS Compressed Sensing, high order statistics, low-rank matrix completion, modulation recognition, parameter estimations, PSK signals.

I. INTRODUCTION

For many applications, such as spectrum surveillance and cognitive radio (CR), automatic modulation recognition (AMR) is a basic task [1]. As a well-studied research arena, there is a wide variety of methods available [1], [2], [3], e.g., Wavelet/Fourier transform, Cumulants, and Cyclic features. To implement most of the classical AMR methods, Nyquist-rate (or faster) samples are required, which brings a heavy burden for Analog to Digital Converter (ADC), especially for wide band signals. Whereas, AMR only extracts quite little information from the sampling data, compared with the amounts of samples.

Recently, Compressed Sensing (CS) has emerged as a new sampling theory [4], [5], [6]. In CR, CS has been introduced to relieve the significant burden of ADC [7], which explores the sparsity of wireless signals.

Currently, some attention has been paid to CS based AMR schemes. Lim and Wakin [6] tried to fulfil AMR by estimating the Nth power spectrum without reconstruction, which does not have strong anti-noise ability and can only discriminate BPSK, QPSK and 8PSK signals. Besides, Liu et al. [8] proposed a new method to estimate the compressive high order cyclostationary statistics, which requires a significant amount of specialized hardware. Based on Tian's work [6], Zhou and others derived some other techniques through cyclic spectrum reconstruction with Sub-Nyquist samples [9], [10], whose calculation is usually complicated.

In our previous work [11], [12], utilizing CS, we established the relationship between Sub-Nyquist samples with high order statistics. Based on the reconstructed Nth power spectrum, we fulfilled the tasks of AMR and rough estimations of parameters for PSK signals. The use of CS could reduce the complexity of signal receiver, as well as the amount of data sampled.

To achieve the goal of both accuracy and efficiency, we establish a new model of high order statistics recovery with low-rate nonuniform samples based on recent advances in the area of low-rank matrix completion [13], [14], [15], [16], [17], [18]. We model the recovery of high order statistics of PSK signal as one of completing a low-rank matrix. Thanks to the modified Singular Value Threshold (SVT) algorithm proposed in this paper, which uses very minimal storage space and keeps the computation cost of each iteration low,

The associate editor coordinating the review of this manuscript and approving it for publication was Li Zhang.

we have a good solution method to the model. In contrast to the previous l_1 optimization criteria models of AMR, our new method is much more computationally efficient and can deal with data of very large scale, which can improve the accuracy further. Due to the limited space, we only analyze the model for PSK signals, but the model can be easily extended to some other signals, such as ASK, FSK, et al.

Actually, there are many other matrix completion researches. Mohimani et al. [21] presented a fast algorithm based on smooth l_0-norm. Ghasemi et al. [22] and Malek-Mohamma [23] proposed a smooth rank function method. Wang put forward a novel approach to approximate the rank function and l_0-norm [24]. Li proposed the S-NMF-EC method [25]. These are good works to solve different low-rank completion problems. However, extensive comparisons are beyond the scope of this paper. In this paper, we put forward a simple and effective low-rank matrix completion method, modified from SVT. And comparisons are made between the l_1-norm method when dealing with high order statistics recovery of PSK signals.

The remainder of this paper is organized as follows. In Section II, we introduce the high order statistics recovery model based on low-rank matrix completion. In Section III, we describe the efficient modified SVT algorithm. In Section IV, we provide the AMR and parameter rough estimation strategies. In Section V, we give some comparisons with the l_1 minimization criteria. We show the efficiency of our method by simulations in Section VI. Finally, we conclude this paper in Section VII.

The main contributions of this work lie in each section of this paper:

(1) We look into the problem of recovering high order statistics of PSK signals with compressive samples.

(2) We bring up the low-rank completion model for problem A, which is new and creative.

(3) We put forward a simple and effective algorithm (TSVT, modified from SVT) to solve the model of problem B. And give some analyses of the algorithm.

(4) We testify the effectiveness of the model and algorithm with simulation and used the model to finish the task of automatic modulation recognition and parameter estimations.

II. RECOVERY OF HIGH ORDER STATISTICS BASED ON LOW-RANK MATRIX COMPLETION

In this section, we will formulate the problem of recovering high order statistics of PSK signals as a rank minimization problem. First, we give some fundamental analysis of PSK signals' high order statistics. Based on the analysis, we discuss the low-rank structure of the high order statistics matrix. In the end, we model the recovery of high order statistics as low-rank matrix completion problem.

A. ANALYSIS OF PSK SIGNALS' HIGH ORDER STATISTICS

It has been long known that the Nth Power Nonlinear Transformation (NPT) is an efficient technique to distinguish some types of signals [3], e.g., MPSK (M=2, 4, 8), MSK, OQPSK, FSK and ASK. Different nonlinearities applied to the signals under classification are used to exploit differences in the high order moment-spaces. These differences manifest themselves in spikes of the spectrum of the transformed signals, with locations determined by carrier frequency and symbol rate Rs [3]. For simplicity, we only research the characteristics of MPSK, OQPSK and MSK signals, which are collectively called PSK signals in this article.

In [3], Reichert analyzed the locations of the spikes generated by different nonlinearities, but the author failed to give out the exact amplitudes and locations of the spikes with Square-Root Raised Cosine (SRRC) Pulse Shape filter. In the following steps, we compute the analytic expressions for the amplitudes of the spikes and continuous part of the power spectral density (PSD) by the example of squared BPSK signals. For signals with other shape filters, the conclusions are similar and the low-rank matrix model still applies.

First, we define the signal model. For MPSK, the signal models are [10]:

$$s_{MPSK}(t) = \sum_{n=-L}^{L} Ag(t - nT_s) \exp\left(j2\pi \frac{m_n - 1}{M} + j2\pi f_c\right) \tag{1}$$

where A is the amplitude. $T_s = 1/R_s$ is the symbol period, and R_s is the symbol rate. $M \in \{2, 4, 8\}$ is the number of unique phases; m_n is the n^{th} transmitted symbol; f_c is the carrier frequency; $g(t)$ is SRRC filter, with the roll-off factor α $(0 < \alpha < 1)$. The frequent property of g(t) is as follows:

$$\mathrm{G}(f) = \begin{cases} 1 & |f| < \dfrac{1-\alpha}{2T_s} \\ \cos\left[\dfrac{\pi T_s}{2\alpha}\left(|f| - \dfrac{1-\alpha}{2T_s}\right)\right] & \dfrac{1-\alpha}{2T_s} < |f| \le \dfrac{1+\alpha}{2T_s} \\ 0 & |f| > \dfrac{1+\alpha}{2T_s} \end{cases} \tag{2}$$

According to (1), it is easy to know that, the expression of squared BPSK signals is as follows:

$$\begin{aligned} s^2_{BPSK}(t) &= b_n^2 \exp(j4\pi f_c t + j2\phi_n) + s^2_{cross} \exp(j4\pi f_c t) \\ &= b_n^2 \exp(j4\pi f_c t) + s^2_{cross} \exp(j4\pi f_c t) \end{aligned} \tag{3}$$

where

$$b_n^2(t) = A^2 \sum_{n=-L}^{L} g^2(t - nT_s) \tag{4}$$

and

$$\begin{aligned} s^2_{cross}(t) &= A^2 \sum_n \sum_{m \neq n} \mathrm{g}(t - nT_s)\mathrm{g}(t - mT) \exp(j\phi_n + j\phi_m) \\ &= A^2 \sum_n \sum_{m \neq n} a_n a_m \mathrm{g}(t - nT_s)\mathrm{g}(t - mT) \end{aligned} \tag{5}$$

$\phi_k \in \{0, \pi\}$ is the k-th transmitted symbol's phase, and $a_k = \exp(j\phi_k) \in \{1, -1\}$ is the k-th transmitted symbol.

According to the property of Fourier transform, the item $\exp(j4\pi f_c t)$ works as frequency shift, which means we only need to analyze the items $b_n^2(t)$ and $s_{cross}^2(t)$.

Since $b_n^2(t)$ is the periodical extension of $g^2(t)g^2(t)$, the Fourier Transform of $b_n^2(t)$ should be Fourier Series. As the non-zero bandwidth of $g(t)$ is $\left(-(1+\alpha)/2T_s, (1+\alpha)/2T_s\right)$, according to the convolution property of Fourier Transform, the non-zero bandwidth of $b_n^2(t)$ should be $\left(-(1+\alpha)/T_s, (1+\alpha)/T_s\right)$ $(0 < \alpha < 1)$. Thus, we only need to analyze the coefficients corresponding to $f = 0$ and $f = \pm 1/T_s$.

According to the convolution property of Fourier transform, the Fourier transform maps the product of two signals into their convolution. If we denote the Fourier transform of $b_n^2(t)$ as $B(f)$. Then, it can be calculated that:

$$\begin{aligned} B(f)\big|_{f=0} &= \mathrm{F}\left[b_n^2(t)\right]\Big|_{f=0} \\ &= A^2\left[\mathrm{G}(f) * \mathrm{G}(f)\right]\Big|_{f=0} = \frac{2LA^2}{T_s} \end{aligned} \tag{6}$$

Proof: See Appendix A.

$$\begin{aligned} B(f)\big|_{f=\pm 1/T_s} &= \mathrm{F}\left[b_n^2(t)\right]\Big|_{f=\pm 1/T_s} \\ &= A^2\left[\mathrm{G}(f) * \mathrm{G}(f)\right]\Big|_{f=\pm 1/T_s} = \frac{2LA^2}{T_s}\cdot\frac{\alpha}{\pi} \end{aligned} \tag{7}$$

Proof: See Appendix B.

According to the periodogram method [26], when calculating the PSD of corresponding frequency, we only need to raise the values to the power of 2.

As to the continuous contribution, $s_{cross}^2(t)$, it is caused by the stochastic nature of the modulation signal (self-noise). Denoting the Fourier transform of $s_{cross}^2(t)$ as $S(f)$, we can calculate its PSD with the definition [27]:

$$P(f) = \lim_{L\to\infty} \frac{E\left[|S(f)|^2\right]}{(2L+1)T_s} \tag{8}$$

It can be calculated that:

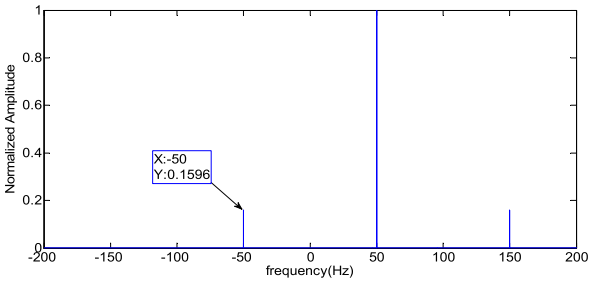
$$P(f) = \lim_{L\to\infty} \frac{E\left[|S(f)|^2\right]}{(2L+1)T_s} = 0 \tag{9}$$

Proof: See Appendix C.

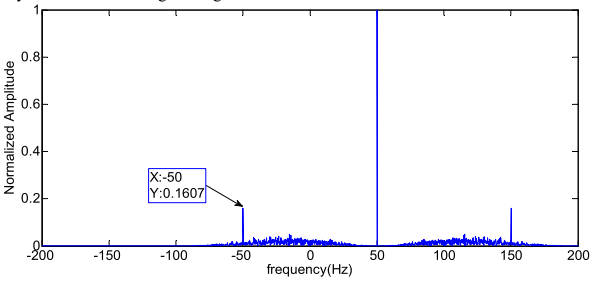
Based on (9), we know that, as the number of symbols increases, the stochastic value of the continuous part tends to zero, which is clearly shown in Fig. 1 (a).

Based on all computations above, it is safe to reach the conclusion that, for BPSK signal, the spectrum of BPSK raised to power 2 is a standard sparse signal, which is applicable to CS theory. And the 'Information', which is concerned by CS theory, is the spikes of the spectrum. In the following part, we can see that, the data matrices generated by the samples are standard low-rank matrices, and the rank of the matrices are also determined by the spikes.

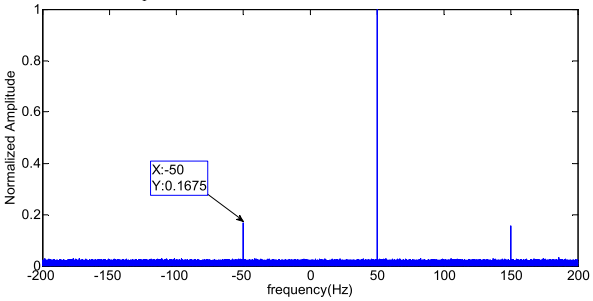
Other simulation parameters of Fig. 1 are set as follows: $f_c = 25\text{Hz}, f_s = 400\text{Hz}, R_s = 100\text{Hz}, \alpha = 0.5$.



(a) Symbol number $N_s = 2^{20}$. There exists barely non continuous part when the symbol number is large enough.



(b) $N_s = 2^9$. The continuous part exists, but their amplitudes are much lower than that of the discrete spikes.



(c) $N_s = 2^{20}$, SNR=-10dB. All aliases brought by AWGN are shaped into constant floor along the whole spectrum.

FIGURE 1. Spectrum of BPSK signal raised to power 2.

In the simulations, we set $\alpha = 0.5$. From (6) and (7), we know that,

$$\lambda = \frac{B(f)\big|_{f=\pm 1/T_s}}{B(f)\big|_{f=0}} = \frac{\alpha}{\pi} = \frac{0.5}{\pi} \approx 0.1592 \tag{10}$$

In Fig. 1(a)-Fig.1(c), we can see, the simulation results of λ approximate the theoretic value.

Moreover, what needs to be pointed out is that, the continuous part of the spectrum will not approximately equal to zero, when the number of symbols is not large enough. But empirically, as long as the number of symbols is no less than several hundreds, which is not a strict restraint in practice, the spikes of the spectrum would be obviously of larger amplitudes than that of the continuous part. This situation is shown is Fig. 1(b). And only the spikes are what we concern, the continuous part is actually redundant structure. As shown below in Table 1 and Table 2, since we only need to extract the spikes of the spectrum, the sequence of the spectrum can be regarded a compressible signal, where CS theory can be applied.

An additional advantage of high order statistics is that when mixed with noise, the features of spikes are 'unbiased' in the sense that the noise, if uncorrelated, does not cause additional spikes which could be regarded as modulation. For the commonly seen Additive White Gaussian Noise (AWGN), as the noise obeys uniformly random distribution, is statistically independent with the signals, all aliases are shaped into constant floor along the whole spectrum. In this case, as long as the Signal to Noise Ratio (SNR) is high enough, the spikes are still much higher than the floor, and the CS theory still applies. BPSK signal mixed with AWGN is shown in Fig. 1 (c).

The computations for the other modulations and nonlinearities of higher order are more tedious but follow in similar way. The spectrum features of PSK signals are shown in Table 1 and Table 2.

B. LOW-RANK MATRIX STRUCTURE

From Table 1, we can see that, the number of spikes (denoted as r) in Nth power spectrum is no more than 7, that means we the signals of nonlinearities can be characterized mainly by complex sinusoidal signals. For a given complex sinusoidal signal,

$$u(t) = e^{j\omega t} \tag{11}$$

we can see that,

$$u(t+1) = u(t) \cdot e^{j\omega} \tag{12}$$

From (12), we can see that, it is an order-1 procedure for a given complex sinusoidal signal.

For a given PSK signal $\mathbf{s} = [s(1), \cdots, s(L)]^T$, we denote s_N as the N$^{\text{th}}$ power transformed forms:

$$\mathbf{s}_N = [s^N(1), \cdots, s^N(L)]^T \tag{13}$$

Then, $\mathbf{s}_N$ is an order-r procedure, where r is decided by Table 1.

Consider the matrices $\mathbf{O}_1$ and $\mathbf{O}_2$ constructed by stacking all the entries of $\mathbf{s}_N$ defined as below.

Suppose $L = 2l_0 (l_0 \in \mathbb{N})$ (if not, we can always discard some entries of $\mathbf{s}_N$ to satisfy the requirement), we define $\mathbf{O}_1$ as,

$$\mathbf{O}_1 = \begin{bmatrix} s^N(1) & s^N(2) & s^N(l_0+1) \\ s^N(2) & s^N(3) & \cdots s^N(l_0+2) \\ \vdots & \vdots & \vdots \\ s^N(l_0) & s^N(l_0+1) & s^N(L) \end{bmatrix} \tag{14}$$

Similarly, suppose $L = l_1^2 (l_1 \in \mathbb{N})$ (if not, we can always discard some entries of $\mathbf{s}_N$ to satisfy the requirement), we define $\mathbf{O}_2$ as,

$$\mathbf{O}_2 = \begin{bmatrix} s^N(1) & s^N(l_1+1) & s^N(l_1^2-l_1+1) \\ s^N(2) & s^N(l_1+2) & \cdots s^N(l_1^2-l_1+2) \\ \vdots & \vdots & \vdots \\ s^N(l_1) & s^N(2l_1) & s^N(L) \end{bmatrix} \tag{15}$$

Considering the sequence of entries in the same row or line in stacking matrices, we can see that, they are also an order-r procedure.

Clearly, irrespective of the number of entries L and the stacking method, the rank of $\mathbf{O}_1$ and $\mathbf{O}_2$ is r.

Here, we can name the stacking methods of $\mathbf{O}_1$ and $\mathbf{O}_2$ as crossover stacking method and non-crossover stacking method. Empirically, for a given L, we find the crossover stacking method has stronger anti-noise ability, as most elements are used many times and it has much more redundant information; while for a given stacking matrix size, which determines the algorithm complexity, the non-crossover method can stack a larger L, and this would bring a more precise result in turn.

Of course, there are many other methods of data stacking. In our discussions and simulations below, we simply use the non-crossover stacking method.

Take BPSK signal as an example,

$$\mathbf{s}^{BPSK} = [s_1, \cdots, s_L]^T \tag{16}$$

BPSK signal raised to power 2 is donated as,

$$\mathbf{s}_2^{BPSK} = [s_1^2, \cdots, s_L^2]^T \tag{17}$$

Stacking matrices $\mathbf{O}_1$ and $\mathbf{O}_2$ are defined as,

$$\mathbf{O}_1 = \begin{bmatrix} s^2(1) & s^2(2) & s^2(l_0+1) \\ s^2(2) & s^2(3) & \cdots s^2(l_0+2) \\ \vdots & \vdots & \vdots \\ s^2(l_0) & s^2(l_0+1) & s^2(L) \end{bmatrix} \tag{18}$$

$$\mathbf{O}_2 = \begin{bmatrix} s^2(1) & s^2(l_1+1) & s^2(l_1^2-l_1+1) \\ s^2(2) & s^2(l_1+2) & \cdots s^2(l_1^2-l_1+2) \\ \vdots & \vdots & \vdots \\ s^2(l_1) & s^2(2l_1) & s^2(L) \end{bmatrix} \tag{19}$$

According to Table 1, $\mathbf{O}_1$ and $\mathbf{O}_2$ both are rank-3 matrices.

In Section II-C, we will cast the rank minimization problem as nuclear norm minimization problem of the data matrices, which is the tightest convex relaxation of the former. Thus, we can validate the rank of the data matrices by the singular values. Take BPSK signal raised to power 2 as an example, the singular values of data matrix with non-crossover stacking method are simulated.

In Fig. 2 (a) and Fig. 2 (b), the singular values are sorted according to their amplitudes. Parameters of simulations are as follows: $f_c = 0$Hz, $f_s = 2.4$kHz, $R_s = 800$Hz, $N_s = 2^{16}$, $\alpha = 0.5$. These two figures illustrate the statement above that, there are three major singular values of the data matrix. Judging from their amplitudes and the calculation results of (6) -(7), we know that, the largest singular value corresponds to the spike of carrier frequency, while the second and the third singular values correspond to the two spikes of symbol rate. The rest singular values correspond to the continuous part of the spectrum.

Since we cast the rank of matrix as nuclear norm, the simulations validate the data of Table 1 that, the spikes'

number in spectrum of $\mathbf{s}_2^{BPSK}$ is 3. Besides, from Fig. 2 (b), we can calculate the ratio of the second/ third singular value to the first singular:

$$\lambda_1 = {}^{47.8491}/_{295.2675} = 0.1620 \tag{20}$$

$$\lambda_2 = {}^{47.5128}/_{295.2675} = 0.1609 \tag{21}$$

Based on the discussion in the last two paragraphs, we know the theoretic value of λ_1 and λ_2 should be identical with λ, which is calculated in (10). We can see the simulation results indeed approximate that of λ.

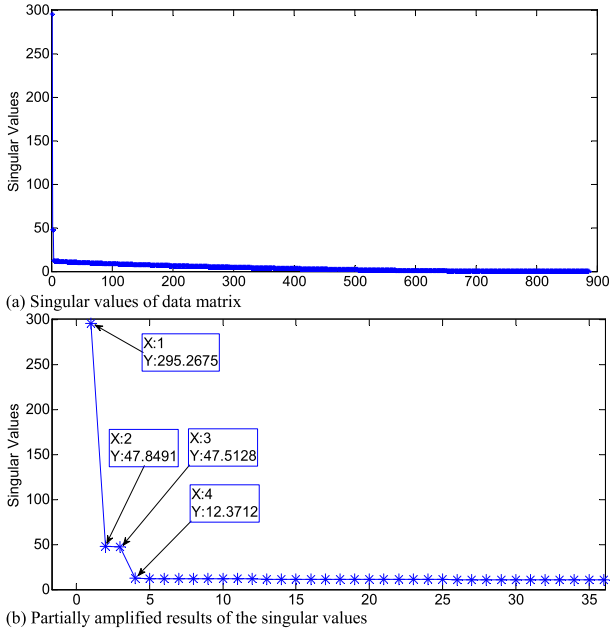


FIGURE 2. Singular values of data matrix with non-crossover stacking method for BPSK signal raised to power 2.

C. MODELING OF LOW-RANK MATRIX COMPLETION

1) LOW-RATE NONUNIFORM COMPRESSIVE SAMPLING AND LOW-RANK MATRIX COMPLETION PRELIMINARIES

As discussed above, the stacking matrices $\mathbf{O}_1$ and $\mathbf{O}_2$ have low-rank structures, which can be recovered from limited observations of s_N. This may arise when data deficiency occurs on the samples. However, one more significant application is for nonuniform compressive sampling (NCS) [6], [8], [28].

NCS was first proposed by Candès et al. [28], the low-rate NCS is applicable to the sparse or compressible signals in frequency domain. In NCS strategy, an analog signal is sampled nonuniformly with random intervals between the sampling times. Usually, these random intervals are integer multiples of a fixed finer sampling interval (which is at least as fine as the Nyquist-rate sampling interval).

If we let $\mathbf{s}$ denote a hypothetical length-L vector of Nyquist-rate samples of some underlying analog signal $r(t)$:

$$\mathbf{s} = [s_1, s_2, \cdots, s_L]^T = [r(t_0), r(t_1), \cdots, r(t_{L-1})]^T \tag{22}$$

where $t_l = lT_s (l \in \mathbb{N})$ and T_s is a sampling interval that is no larger than the Nyquist-rate sampling interval. In previous papers [6], [10], the authors directly collect a random subset of M samples from the L entries of $\mathbf{s}$. The NCS procedure can be represented as:

$$\mathbf{y} = \mathbf{\Phi s} \tag{23}$$

where $\mathbf{\Phi}$ is a randomly generated matrix containing a single 1 in each row. For example, if $M = 5, L = 10$, one realization of $\mathbf{\Phi}$ is:

$$\mathbf{\Phi} = \begin{bmatrix} 1&0&0&0&0&0&0&0&0&0 \\ 0&0&1&0&0&0&0&0&0&0 \\ 0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&0&1 \end{bmatrix} \tag{24}$$

then, we would have the NCS sequence:

$$\mathbf{y} = [y_1, y_2, y_3, y_4, y_5]^T = [s_1, s_3, s_6, s_8, s_{10}]^T \tag{25}$$

As can be seen, the measurement matrix $\mathbf{\Phi}$ fulfils the function of randomly sampling of $\mathbf{s}$.

It is proven that, if we denote Ψ as the $L \times L$ DFT matrix, then $A = \sqrt{L/M} \cdot \Phi\Psi$ satisfies the RIP of order $M/\log^4 L$ [39]. This is the reason why we can recover the sparse spectrum of PSK signals' high order statistics in our previous paper [13], with low-rate nonuniform samples.

In this section, we will cast the problem of recovering high order statistics of PSK signals as a sparse low-rank matrix completion model with low-rate nonuniform samples.

Recently, there is rapidly growing studies in the recovery of partial unknown low-rank or approximately low-rank matrix from limited entries. This problem has widespread uses, such as machine learning [16], [17], [18], control [19] and so on [20].

In previous studies [22], Candès and Recht proved that, most low-rank matrices can be recovered exactly by solving a convex optimization problem. If the unknown matrix is $M \in \mathbb{R}^{n_1 \times n_2}$, $n = \max(n_1, n_2)$, with available m sampled entries $\{M_{ij} : (i,j) \in \Omega\}$, and Ω is a random subset of cardinality m. It has been proven [29] that, most low-rank matrices can be perfectly recovered by solving the following optimization problem.

$$\begin{aligned} &\min \|\mathbf{X}\|_* \\ &\text{s.t.} \quad \mathbf{X}(i,j) = \mathbf{M}(i,j), (i,j) \in \mathbf{\Omega} \end{aligned} \tag{26}$$

provided that the samples number satisfies $m \geq Cn^{6/5} r \log n$, where C is a positive numerical constant.

In (26), $\|\cdot\|_*$ is the nuclear norm, which is the sum of singular values. In some sense, (26) is the tightest convex relaxation of the NP-hard rank minimization problem:

$$\begin{aligned} &\min \ \ \text{rank}(\mathbf{X}) \\ &\text{s.t.} \quad \mathbf{X}(i,j) = \mathbf{X}(i,j), (i,j) \in \mathbf{\Omega} \end{aligned} \tag{27}$$

TABLE 1. Number of discrete peaks for PSK type signal.

Nonlinearity	BPSK	QPSK	8PSK	OQPSK	MSK
None	0	0	0	0	0
$(.)^2$	3	0	0	2	2
$(.)^4$	5	5	0	3	2
$(.)^8$	7	7	7	3	2

TABLE 2. Peak lines in different nonlinearities for PSK type signals N: None. Y: Exist ($n \in \mathbb{Z}$).

Nonlinearity	Frequency	BPSK	QPSK	8PSK	OQPSK	MSK
None	f_c	N	N	N	N	N
None	$f_c + nR_s$	N	N	N	N	N
$(.)^2$	f_c	Y	N	N	N	N
$(.)^2$	$f_c + (n+0.5)R_s$	N	N	N	N	Y
$(.)^2$	$f_c + nR_s$	Y	N	N	Y	N
$(.)^4$	f_c	Y	Y	N	Y	N
$(.)^4$	$f_c + (n+0.5)R_s$	N	N	N	N	N
$(.)^4$	$f_c + nR_s$	Y	Y	N	Y	Y

2) MODELING RECOVERY OF HIGH ORDER STATISTICS AS LOW-RANK MATRIX COMPLETION

Having got the NCS sequence $\mathbf{y}$, we can get the corresponding NPT forms of $\mathbf{y}$, which is donated as $\mathbf{y}_N$:

$$\mathbf{y}_N = [y_1^N, y_2^N, \cdots, y_M^N]^T \tag{28}$$

Thanks to the structural property of $\mathbf{\Phi}$, we have actually finished randomly sampling of $\mathbf{s}_N$:

$$\mathbf{y}_N = \mathbf{\Phi} \boldsymbol{s}_N \tag{29}$$

Then, we can stack the sequence $\mathbf{y}_N$ into the stacking matrices $\mathbf{O}_{y_N}$, with crossover or non-crossover methods. In order for computation and expression simplicities, we set the sampled entries of $\mathbf{O}_{y_N}$ as $y_k^N (k = 1, 2, \cdots, M)$, and the unsampled entries as zeros. So that, $\mathbf{O}_{y_N}$ is of the same size as $\mathbf{O}_{s_N}$ (the stacking matrix of $\mathbf{s}_N$). In fact, we can discover that, the non-zero entries of $\mathbf{O}_{y_N}$ is a random sampling subset of $\mathbf{O}_{s_N}$.

As discussed above, $\mathbf{O}_{s_N}$ is a rank-r matrix. Therefore, we come to the conclusion that, recovering $\mathbf{O}_{s_N}$ from $\mathbf{O}_{y_N}$ is a low-rank matrix completion problem:

$$\begin{aligned} P_r: \ & \min \ rank(\mathbf{O}_{s_N}) \\ & \text{s.t.} \quad \mathbf{O}_{s_N}(i,j) = \mathbf{O}_{y_N}(i,j), (i,j) \in \mathbf{\Omega} \end{aligned} \tag{30}$$

where $\mathbf{\Omega}$ is a random subset of cardinality m. As to the NCS method, $\mathbf{\Omega}$ is determined by the measurement matrix $\mathbf{\Phi}$. Of course, we can also implement the randomly sampling by directly sampling the signals after NPT, and the low-rank matrix completion model is the same as (30).

In fact, (30) is a NP-hard and non-convex problem, whose tightest convex relaxation form is as follows:

$$\begin{aligned} P_*: \ & \min \ \left\| \mathbf{O}_{s_N} \right\|_* \\ & \text{s.t.} \mathbf{O}_{s_N}(i,j) = \mathbf{O}_{y_N}(i,j), (i,j) \in \mathbf{\Omega} \end{aligned} \tag{31}$$

where $\|\cdot\|_*$ is the nuclear norm.

It has been proven by [29] that, most matrices $\mathbf{O}_{s_N}$ of rank r can be perfectly recovered by solving the optimization problem (31).

In practice, the low-rank structure of the stacking matrix $\mathbf{O}_{s_N}$ is seldom observed. This is due to the presence of noise in practical data as well as the continuous part of the spectrum. In this condition, the stacking matrix $\mathbf{O}_{s_N}^1$ of practical data can be discomposed as follows:

$$\mathbf{O}_{s_N}^1 = \mathbf{O}_{s_N} + \mathbf{E} \tag{32}$$

where $\mathbf{E}$ is the noise matrix, and $\mathbf{O}_{s_N}$ is the signal data matrix.

Considering the existence of noise, the model proposed by (31) should be rewritten as follows:

$$\begin{aligned} P_{*,1}: \ & \min \ \left\| \mathbf{O}_{s_N} \right\|_* \\ & \text{s.t.} \left\| \mathrm{P}_\Omega(\mathbf{O}_{s_N} - \mathbf{O}_{y_N}) \right\|_F \le \sigma \end{aligned} \tag{33}$$

where σ is the parameter controlled by the noise level. P_Ω is the orthogonal projector onto the span of matrices vanishing outside of $\mathbf{\Omega}$, so that the (i,j)th component of $\mathrm{P}_\Omega(\mathbf{X})$ is equal to $\mathbf{X}_{ij}$ if $(i,j) \in \mathbf{\Omega}$ and zero otherwise.

However, in practical situation, the noise level of σ is unknown. Based on the discussion above, we know that, the matrix $\mathbf{O}_{s_N}$ is of low-rank structure. However, thanks to the existence of $\mathbf{E}$, the low-rank structure of $\mathbf{O}_{s_N}^1$ has been

destroyed. In fact, there are few theories of low-rank matrix recovery from noisy data at the moment, and the related study is obviously beyond the scope of this paper. In this paper, we give an alternative method by proposing a modified model.

D. MODIFIED MODEL OF DATA RECOVERY

It is clearly stated in (32) that, while getting the practical data matrix $\mathbf{O}_{s_N}^1$, only the signal data matrix $\mathbf{O}_{s_N}$ is low-rank, which is desired by us. Moreover, the rank of $\mathbf{O}_{s_N}$ is determined by Table 1, which is known to us. From this point of view, we can see model of (31) inversely.

On account of the analysis, we can rewrite (33) into a more reasonable form:

$$P_{*,2}: \quad \min \left\| \mathrm{P}_\Omega(\mathbf{O}_{s_N} - \mathbf{O}_{y_N}) \right\|_F \quad \text{s.t.} \left\| \mathbf{O}_{s_N} \right\|_* \le r_* \tag{34}$$

where r_* is a preset parameter.

Now, we have established the basic model used to recover high order statistics of PSK signals. In the following part, we will discuss how to solve (34) efficiently.

III. EFFICIENT SOLUTION VIA CONVEX OPTIMIZATION

In [23], Cai, Candès, and Shen gave the SVT algorithm for the solution of matrix completion problem of large scale. However, the authors failed to handle the noisy sampled situation properly with unknown level of noise energy. Here, based on the model (34), we propose Truncated Singular Value Threshold (TSVT) algorithm.

As discussed in Section II, when the signal sampled is mixed with AWGN, or existing continuous part of spectrum as to the number of the symbols is not large enough, the stacking matrices are not strictly low-rank matrices and there may be no low-rank matrix matching the noisy data. When utilizing the SVT algorithm directly, we find the rank of recovered stacking matrix is divergent. That means we can't get the stopping criterion by only exploring the recovering error or the maximum iteration count, as the SVT algorithm proposed.

In Fig. 3, BPSK signal raised to power 2 is simulated. Simulation parameters are as follows: $f_c = 0$Hz, $f_s = 2.4$kHz, R_s=800Hz, SNR=10dB, $N_s = 2^{12}$, $L = 12288$. Maximum iteration number is set 200. We only randomly sample 1296 entries (about 10.5%) of the signal.

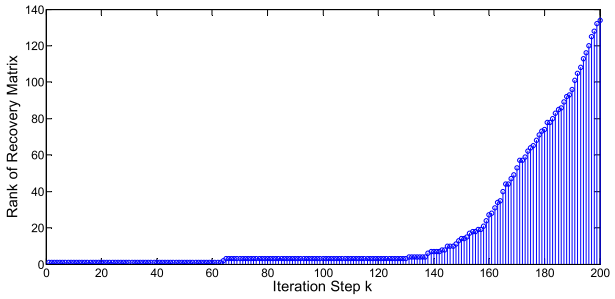


FIGURE 3. Rank of recovered matrix is divergent in practical applications.

Although we know the rank of the recovered data matrix is 3, the rank of the recovered matrix increases with the iteration. When the iteration stops at the set count, the final rank of the data matrix is much larger than 3, as Fig. 3 shown.

In order to describe TSVT algorithm, we give some definitions, cited from SVT [29], which are also used in TSVT algorithm.

In SVT algorithm, the shrinkage iterations are as follows [29]:

$$\begin{cases} \mathbf{O}_{s_N}^k = \mathrm{D}_\tau(\mathbf{Y}^{k-1}) \\ \mathbf{Y}^k = \mathbf{Y}^{k-1} + \delta \mathrm{P}_\Omega(\mathbf{O}_{y_N} - \mathbf{O}_{s_N}^k) \end{cases} \tag{35}$$

where D_τ is the soft-thresholding operator defined in [29].

Algorithm 1 Truncated Singular Value Thresholding (TSVT) Algorithm

Input: sampled set Ω and sampled entries $\mathrm{P}_\Omega(\mathbf{O}_{y_N})$, step size δ, tolerance ò, parameter τ, increment ℓ, maximum iteration count k_{max}, and maximum rank r_{max}.
Output: $\mathbf{U}$, Σ, $\mathbf{V}$and$\mathbf{O}_{s_N}^{opt}$ ($\mathbf{O}_{s_N}^{opt} = \mathbf{U}\Sigma\mathbf{V}^*$), iteration count k.
Description: Recover a low-rank matrix $\mathbf{O}_{s_N}^{opt}$ from a subset of randomly sampled entries.

Set $\mathbf{Y}^0 = k_0 \delta \mathrm{P}_\Omega(\mathbf{O}_{y_N})$ (k_0 is defined below)
Set $r_0 = 0$
for $k = 1$ to k_{max}
  Set $s_k = r_{k-1} + 1$
  repeat
    Compute $[\mathbf{U}^{k-1}, \Sigma^{k-1}, \mathbf{V}^{k-1}]_{s_k}$
    Set $s_k = s_k + \ell$
  until $\sigma_{s_k-\ell}^{k-1} \le \tau$
  Set $r_k = \max\{j : \sigma_j^{k-1} > \tau\}$
  Set $\mathbf{O}_{s_N}^k = \sum_{j=1}^{r_k} (\sigma_j^{k-1} - \tau) u_j^{k-1} v_j^{k-1}$
  if $\|\mathrm{P}_\Omega(\mathrm{O}_{\mathrm{s_N}}^{\mathrm{k}} \mathrm{O}_{\mathrm{y_N}})\|_\mathrm{F} / \|\mathrm{P}_\Omega \mathrm{O}_{\mathrm{y_N}}\|_\mathrm{F} \varepsilon$ or $r_k \ge r_{max}$ **then break**
  Set $\mathbf{Y}_{ij}^k = \begin{cases} 0 \text{if}(i,j) \notin \Omega \\ \mathbf{Y}_{ij}^{k-1} + \delta(\mathbf{O}_{y_N}(i,j) - \mathbf{O}_{s_N}^k(i,j)) \text{ if } (i,j) \in \Omega \end{cases}$
end *for* k
Set $\mathbf{O}_{s_N}^{opt} = \mathbf{O}_{s_N}^k$

Note: $\mathbf{k_0}$ is an integer defined in SVT algorithm, which is relate to the rank of the data matrix. The reviewer can refer to reference [23], pp18.

For convenience, we define for each nonnegative integer $s \le \min\{n_1, n_2\}$,

$$[\mathbf{U}^k, \Sigma^k, \mathbf{V}^k]_s, \ k = 1, 2, \cdots \tag{36}$$

where $\mathbf{U}^k = [u_1^k, \cdots, u_s^k]$ and $\mathbf{V}^k = [v_1^k, \cdots, v_s^k]$ are the first s singular vectors of the matrix $\mathbf{Y}^k$, and Σ^k is a diagonal matrix with the first s singular values $\sigma_1^k, \cdots, \sigma_s^k$ on the diagonal.

For more information of SVT algorithm, readers can refer to paper [30] for details.

As mentioned above, there are few low-rank matrix recovery algorithms from noisy data, and the relative theory is far beyond the scope of this paper. However, by exploring

the intrinsic properties of high order statistics of PSK signals, we propose the Truncated SVT (TSVT) algorithm here, which is described in Algorithm 1.

The most significant difference between TSVT and SVT lies in the stopping criterion. For the original SVT algorithm, the iteration stops when the relative recovering error is small enough or the iteration count comes to the set number. As to the existence of unknown level of noise in practice condition, the relative recovering error criterion is barely satisfied, and the iteration will always come to the largest set number, while the rank of recovered $\mathbf{O}_{s_N}^{opt}$ is far beyond the theoretical rank r.

We know that, only the first r singular vectors and first r singular values contain useful information. Considering this, we set the rank of the recovered matrix, $\mathbf{O}_{s_N}^{k}$, as one stopping criterion. Just as Algorithm 1 shows, once the rank of $\mathbf{O}_{s_N}^{k}$ has reached the set count r_{max}, the algorithm will break.

IV. AMR AND PARAMETER ROUGH ESTIMATIONS STRATEGIES

As highlighted in our introduction and paper of Reichert [3], we can use the high order statistics to fulfill the tasks of AMR and parameter rough estimations.

After recovering the stacking matrix $\mathbf{O}_{s_N}^{opt}$, we can get the original signal sequence $\mathbf{s}_N$. Then, we can get the spectrum of $\mathbf{s}_N$, donated as $\mathbf{f}_N$ (N=2, 4, 8), with which we can finish the tasks of classification and estimations based on Table 1 and Table 2.

In TSVT algorithm, there is a key parameter, r_{max}, which should be set manually. For different applications, the values of r_{max} should be set differently.

In fact, from the discussion above, we know that, every singular vector and the corresponding singular represent a complex sinusoidal signal, which are used for classification and estimations.

A. AMR STRATEGY

As shown in Fig. 1(c), when mixed with noise, the energy of NPT signals is spread in the whole spectrum. To decrease the computation complexity of classification, we need to reduce the dimension of the data. There are many classical methods to realize the purpose, such as PCA, MDS, et al. [31]. Considering the TSVT algorithm arranges the singular values in descent order, the TSVT algorithm can automatically finish the task of principal component analysis.

There many criteria to choose r_{max} based on different AMR strategies. Here, we simply set $r_{max} = 8$, and use Support Vector Machine (SVM) to implement AMR efficiently.

Besides, according to Table 1, to distinguish BPSK signals, we only need $\mathbf{f}_2$. With $\mathbf{f}_2$ and $\mathbf{f}_4$, we can distinguish QPSK, OQPSK and MSK. As for 8PSK, it needs to make full use of $\mathbf{f}_2$, $\mathbf{f}_4$ and $\mathbf{f}_8$. Thus, we can design the AMR strategy as Fig. 4. We use a random sampler at Sub-Nyquist rate to sample the analog signals. After N^{th} Power Transform of the samples, we can get the recovered NPT signal $\mathbf{s}_N$ with maximum rank r_{max}. Having gotten the corresponding spectrum $\mathbf{f}_N$, it is sent to SVM, which will make the final decision as a classifier. In our simulations in Section VI, we use the SVM toolbox (libsvm) provided by Chih [41] as the classifier.

B. ESTIMATIONS OF f_c AND R_s

For the tasks of carrier frequency and symbol rate estimations, we only need to extract the first r singular values and the corresponding singular vectors, which means $r_{max} = r + 1$. Thus, the corresponding $\mathbf{f}_N$ would contain mainly r discrete peaks.

Take QPSK signals for example. If we donate A_1, A_2, A_3 as the locations of three dominant peaks of $\mathbf{f}_4$, and it can be calculated that:

$$\hat{f}_c = {}^{A_1}\!/_4 \left(\mathrm{or}^{(A_2 + A_3)}\!/_8 \right) \tag{37}$$

$$\hat{R}_s = |A_3 - A_2|/2 \tag{38}$$

The parameters estimation methods for other modulations follow in the same way. Authors can refer to Table 2 for the specialized estimation methods.

V. COMPARISON WITH l_1 MINIMIZATION

In our previous studies [13], [14], we have researched the tasks of AMR and parameter estimations of PSK signals based on CS theory. In this section, we will give some comparisons between l_1 minimization model and nuclear-norm minimization model.

A. MINIMUM SAMPLING NUMBER

For simplicity, we can define the compression factor as:

$$\beta = M/L \tag{39}$$

where M is the number of NCS.

It is easy to understand that, the compression ratio increases as β decreases.

In our previous paper [13], we use NCS just as Section II discussed. After getting $\mathbf{y}_N$, we can recover the spectrum of $\mathbf{y}_N$ by the following model:

$$P_1 : \hat{\mathbf{f}}_N = \arg\min \|\mathbf{f}_N\|_1 \quad \text{s.t.} \mathbf{y}_N = \mathbf{\Phi}\mathbf{\Psi}\mathbf{f}_N \tag{40}$$

where $\mathbf{\Psi}$ is the $L \times L$ DFT synthesis matrix, working as the sparsifying matrix.

It has been discussed in Section II-C that, the minimum sampling number M_1 should be no less than $r \cdot \log^4(L)$. Here, we denote the model as Model I.

In our previous paper [14], we realized the purpose of AMR based on CS theory by using compressive sampling of signals after NPT. The measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times L}$ used is Gaussian random matrix. In this case, the minimum sampling number M_2 is of order $r \log\left(L/r\right)$ [40]. The recovering model is the same as (40). And we denote this model as Model II.

As discussed in Section II, the minimum sampling number M_3 required by the low-rank model should be more than $n^{6/5} r \log n = 1/2 L^{3/5} r \log L$ ($n = L^{1/2}$ for non-crossover stacking).

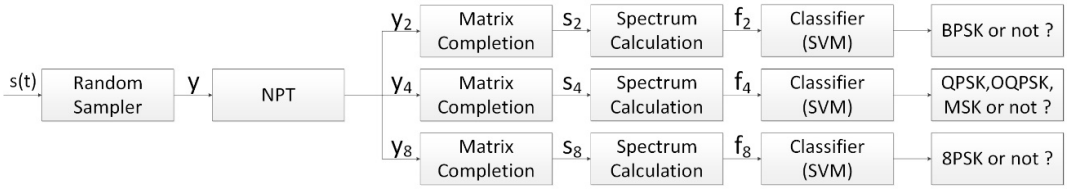


FIGURE 4. The proposed AMR scheme.

Here, we have to point out that, the minimum sampling number stated above is a sufficient but not necessary condition.

Intuitively, we can get the conclusion that,

$$M_1, M_3 \gg M_2 \tag{41}$$

But as what will be discussed below, the current l_1 optimization method can hardly deal with $\mathbf{s}_N$ (or $\mathbf{f}_N$) of the scale larger than several thousands. In this condition, $\mathbf{f}_N$ is not strictly sparse (the corresponding stacking matrix is not strictly of low-rank structure). Thus, the minimum sampling numbers would be larger than the theoretical values. In our simulations previously [13], [14], the compression factor is usually set 0.2-0.3. For the low-rank model proposed in the paper, we denote it as Model III. When dealing with small number of symbols, the compression factor is 0.2-0.3, too. But the compression factor can be much smaller (i.e., the compression ratio can be much larger) if the symbol number is large enough, in which condition, the l_1 minimization algorithm is hard to deal with. In our simulations, we can set the compression factor of Model III as 0.05 or smaller.

B. COMPUTATION COMPLEXITY

As for the convex optimization models, one important problem is the computation complexity.

Traditionally, P_1 is formulated as a linear programming (LP) problem, basis pursuit (BP) for example [32], or approximately solved by heuristic greedy algorithms, which are much faster than LP methods. The greedy algorithms, however, may fail in some conditions [33]. Readers are referred to [34] and [35] for more details of the techniques.

Homotopy method is a fast algorithm that is suitable for large-scale applications to solve P_1. The complexity of the homotopy algorithm is $O(M^2 + ML)$ for each iteration [36], [37], [38].

As to Algorithm 1, each iteration requires a SVD computation that has a complexity of $O(n_1 n_2^2)$. However, to speed up the algorithm, we can compute only partial SVDs in each iteration since we expect the optimal solution to have rank at most 7.

Thus, the complexity of each iteration reduces to $O(ln_1n_2)$ [30]. With non-crossover method, we know that $O(ln_1n_2) = O(lL)$ $(n_1 = n_2 = L^{1/2})$, where l is the number of singular vectors computed in each iteration. It can be seen easily from Algorithm 1 that, $l < r_{max}$ and l is much smaller than M and L. That means, the complexity of $P_{*,2}$ is only linearly dependent on the dimension of signal sequence, and is much smaller than that of 2^{20}, i.e., $O(lL) \ll O(M^2 + ML)1024$.

Since the computation cost of $P_{*,2}$ is much smaller, it means the low-rank model we proposed here can deal with large-scale problems. In our simulations, the proposed algorithm can solve problems, in Matlab, involving sequences of size 2^{20} (stacking matrices of size 1024×1024) in 30 seconds having 70% unknowns on a standard desktop computer with a 2.6 GHz CPU (i-5 4300M) and 16 GB of memory. As to the model of P_1, the computation time would more than 3 minutes for a sequence of size 2^{14}, and would be out of memory for larger scale.

C. COMPUTATION PRECISION

It is difficult to estimate the computation precision exactly. But the statistical characters we use here for AMR and parameters estimations concern much about the symbol number. The ability of dealing with data of much larger scale means we can get much more precise results.

VI. SIMULATION RESULTS

In this section, we verify the efficacy of the low-rank model proposed in the paper. We would give the simulation results of AMR and parameters estimations.

In this section, we use BPSK, QPSK, 8PSK, OQPSK and MSK signals under different scenarios to evaluate the performance of our model. The simulation results contain the correct ratio of recognition and carrier frequency estimation of QPSK signal with $\mathbf{f}_4$.

In the following simulations, the simulation results with uniform samples corresponding to the 'Nyquist' curves. And the 'TSVT' curves represent the results of NCS.

A. VALIDATION OF TSVT

Simulation parameters are set as follows: $f_c = 0$Hz, $R_s = 800$Hz, $f_s = 2.4$kHz, $\alpha = 0.5$, symbol number $N_s = 2^{12}$, SNR=10dB. Length of $\mathbf{s}$ is L= 12288. We only randomly sample 1296 entries of the signal. Thus, $\beta \approx 0.105$.

Fig. 5 (a) depicts the original spectrum of BPSK signal raised to power 2. The amplitudes of spikes are also given in the figures.

The ranks of recovered data matrix by original SVT algorithm are shown in Fig. 3. In the simulation, the maximum iteration number is 200. We can see the rank is divergent.

Fig. 5 (b) depicts the partially amplified result of Fig. 3. We can see that, the rank of recovered data matrix quickly comes to 3, which is the theoretical value. But as the practical

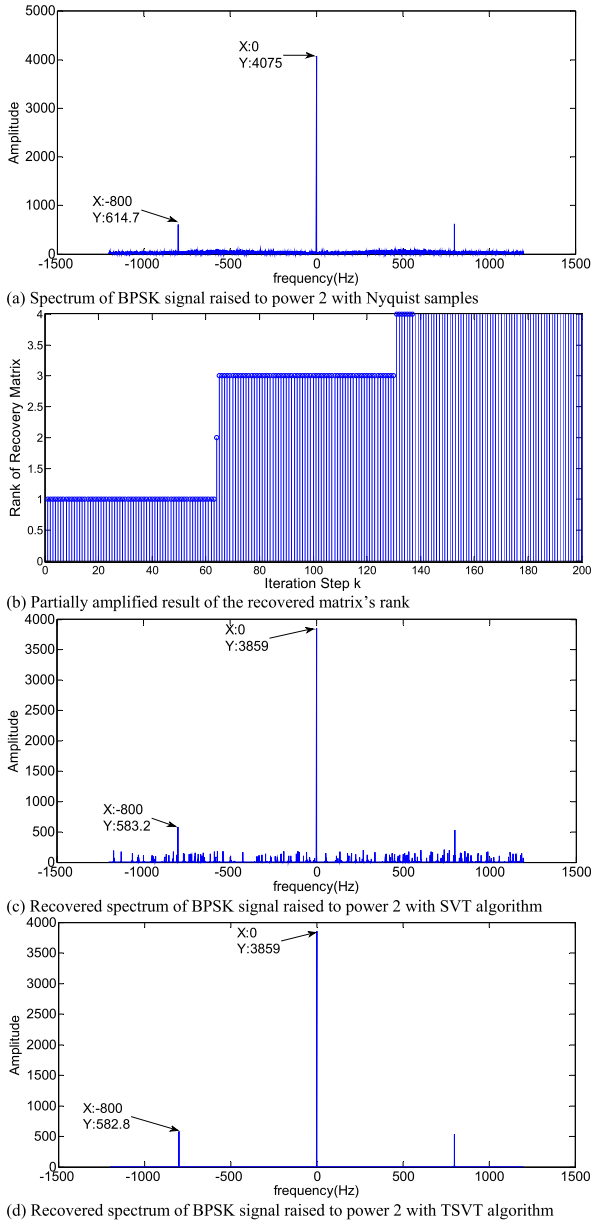


FIGURE 5. Simulations results of TSVT algorithm with BPSK signal raised to power 2.

data sampled is not strictly low-rank, the recovery algorithm won't stop.

For PSK signal recognition and parameter estimations, we only need to retain the first 3 singular values and the corresponding singular vectors.

Fig. 5 (c) shows the recovered spectrum of BPSK signal raised to power 2, with the result of 200 iterations, which is the calculation result with original SVT algorithm. In Fig. 5 (c), apart from 3 major spikes, there are many other noise spikes. This is due to the concentration of noises' energy in a few spikes. This causes that the spikes corresponding to noises are higher than the noise floor of Fig. 5 (a). Obviously, this is not an ideal result for AMR and parameter estimations.

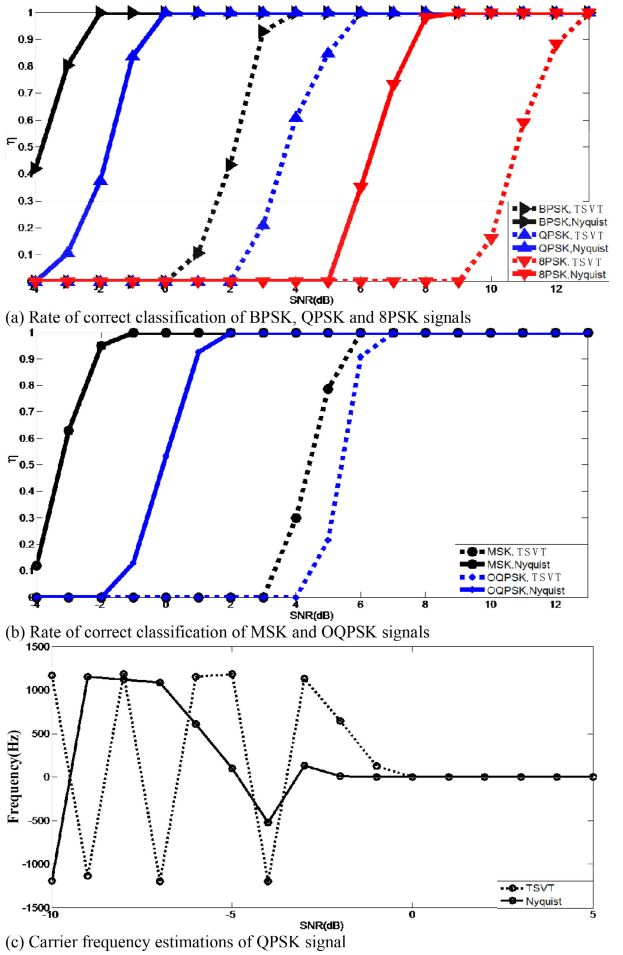


FIGURE 6. Simulation results of AMR and estimation of carrier frequency with small scale of data.

Fig. 5 (d) shows the recovered spectrum with TSVT algorithm. In the algorithm, we simply set $r_{\max} = 4$. (For BPSK signal raised to power 2, the rank of data matrix is 3.) We can see that, the spectrum shown in Fig. 5 (d) recovered the 3 major spikes without other noise spikes. This result is more suitable for AMR and parameter estimations.

B. PSK SIGNAL CLASSIFICATION AND PARAMETER ESTIMATIONS WITH SMALL SCALE OF DATA

In this part, we testify our model and algorithm with small scale of data. Parameters are set as follows: $f_c = 0$Hz, $R_s = 800$Hz, $f_s = 2.4$kHz, $N_s = 2^{12}$, $L = 12288$. As to the data of this scale, the P_1 minimization model can still address, even though a little laborious with the computer configuration features mentioned in Section V-B. The simulation results of P_1 minimization model is shown in [13] and [14].

1) VARY SNR

In this part, we set $\beta = 0.3$, which means $M = 0.3L \approx 3686$.

In the following simulations, we define η as the rate of correct classification.

Fig. 6 (a) and Fig. 6 (b) depict the rate of correct classification of PSK signals versus varying signal-to-noise

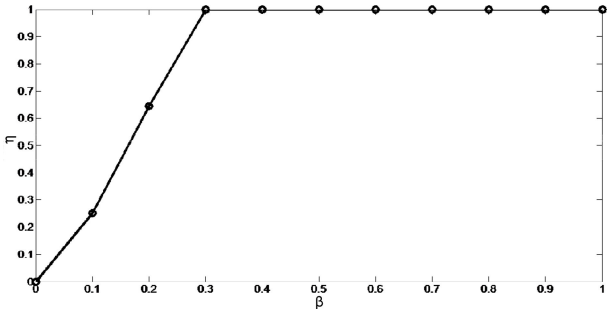


FIGURE 7. Rate of correct classification of BPSK signals with small scale of data.

ratios (SNR). Fig. 6(c) shows the estimation results of carrier frequency of QPSK signal. From Fig. 6(a) and Fig. 6(b), for a given accuracy rate of classification, AMR using NCS requires about 4∼6 dB (7dB for MSK) of SNR more than uniform sampling. From Fig. 6(c), it is about 2dB for the estimation of carrier frequency.

2) VARY COMPRESSION FACTOR

In this part, we set SNR=3dB, and all other parameters are the same with part 1). We validate our model under different compression ratio for BPSK signal.

From Fig. 7, we can see that, with the set parameters, we can achieve almost 100 percent correct classification rate when β is greater than 0.3.

We make the following observations with regards to the simulations. First, as SNR increases, the correct classification rate and parameter estimation accuracy increase. Second, the correct classification rate increases as β increases (i.e., the compression ratio decreases).

Analyzing the simulation results, we hold the conclusion to be obvious that, when SNR is too low, the model we proposed is inferior to the Nyquist uniformly sampling method. We ascribe the reason to the high-level energy of noise and existence of the continuous part of the spectrum. All these factors contribute to the inaccuracy of the sparse model. When SNR goes high enough, our model becomes more applicable.

On the other hand, compared with Nyquist uniformly sampling method, the model we proposed here has a compression factor of 0.3 for simulations of Fig. 6. This is quite a benefit.

C. PSK SIGNAL CLASSIFICATION AND PARAMETER ESTIMATIONS WITH LARGE SCALE OF DATA

In this part, we testify our model and algorithm with larger scale of data. The parameters are as follows: $f_c = 0$Hz, $R_s = 800$Hz, $f_s = 2.4$kHz, $N_s = 2^{17}$, $L = 393216$. As to data of this scale, the P_1 minimization model can hardly address, due to too much memory required. However, our model proposed here can still solve the problem easily.

1) VARY SNR

In this part, we set $\beta = 0.3$. Thus, the number of compressive samples is $M = 0.3L \approx 117965$.

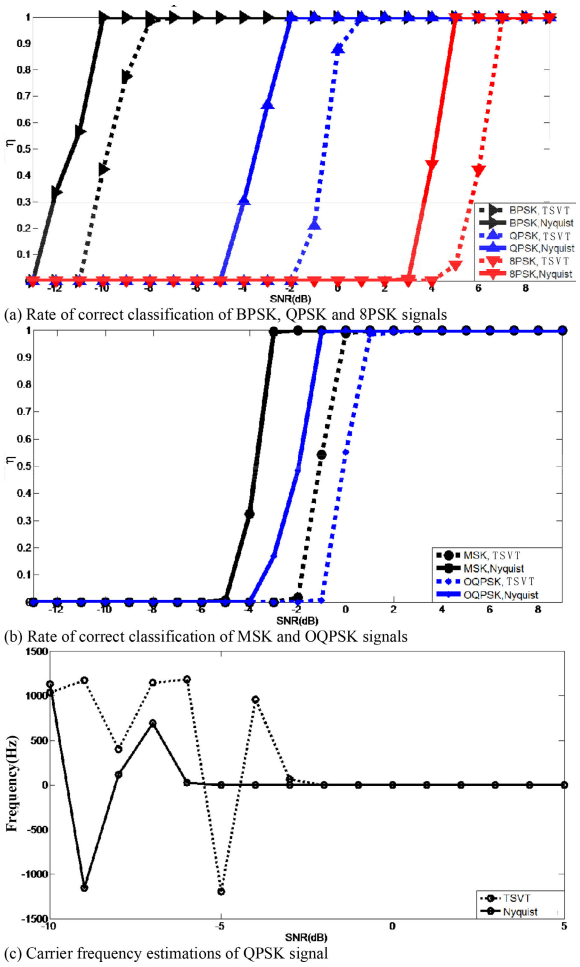


FIGURE 8. Simulation results of AMR and estimation of carrier frequency with large scale of data.

Fig. 8 (a) and Fig. 8 (b) depict the rate of correct classification of PSK signals versus varying SNR. Fig. 8 (c) shows the estimation results of carrier frequency of QPSK signal. From Fig. 8 (a) and Fig. 8 (b), for a given accuracy rate of classification, AMR using NCS requires about 2 dB more of SNR for BPSK, 8PSK and OQPSK than uniformly sampling. While for QPSK and MSK, the requirement is about 4dB. From Fig. 8 (c), it is about 4dB for the estimation of carrier frequency.

From the simulations, we can see that, as the number of symbols increase, the classification and parameter estimation accuracies also improve a lot.

2) VARY COMPRESSION FACTOR

In this part, we set SNR=1dB, and all other parameters are the same with part 1). We validate our model under different compression ratio for BPSK signal.

From Fig. 9, we can see that, with the set parameters, we can achieve almost 100 percent correct classification rate when the compression factor is greater than 0.027.

Compared with Fig. 7, we can see that, even we set a worse SNR here, the compression ratio can improve 10 times more $(0.3/0.027 \approx 11.1)$.

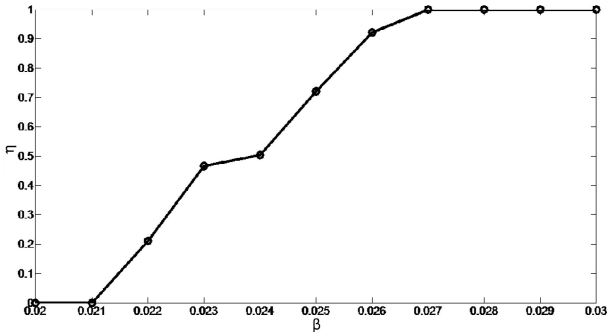


FIGURE 9. Rate of correct classification of BPSK signals with large scale of data.

Compared with the simulation results of Section VI-B, we can see that, when the symbol number is large enough, our model can achieve quite good effects.

We make the following observations with regards to the simulations. First, as the symbol number processed increases, the classification and parameter estimation performances upgrade for a given SNR. This is due to the statistical characters we use here concern much about the symbol number. Second, as the symbol number processed increases, the compression factor can be quite low. Besides, the performance differences between our model and that of the Nyquist uniform sampling decrease a lot. This is because the low-rank structures become more distinct.

We hold the conclusion to be self-evident that, when the symbol number is large, the data matrices have better low-rank structures. Thus, our model here works better.

VII. CONCLUSION

In extending the method of AMR from CS to low-rank matrix completion theory, we proposed the recovery method of high order statistics of PSK signals based on low-rank matrix completion and an efficient solution algorithm. We also considered specifically the applications of AMR and parameter estimations. Simulations provide promising classification and parameter estimation results even when the compression factor is very low. Our work can also extend to some other kinds of signals, and potentially enable AMR for high bandwidth signals with much lower sampling rate.

In this paper, we did not discuss the classification strategy, which is quite fundamental for the accurate classifications. In practice, it is necessary to design an excellent classification scheme. We have used SVM to classify the signals.

As discussed in the introduction, there are many other low-rank matrix completion methods. TSVT(SVT) is only a simple testification of the low-rank model we proposed in the paper. In our future work, more accurate and efficient low-rank matrix completion will be explored.

APPENDIX A
PROOF OF (6)

$$
\begin{aligned}
B(0) &= B(f)\big|_{f=0} \\
&= \mathrm{F}\left[b_n^2(t)\right]\big|_{f=0} \\
&= 2A^2L\left[\mathrm{G}(f) * \mathrm{G}(f)\right]\big|_{f=0} \\
&= 2A^2L\int_{-\frac{1+\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}} \mathrm{G}^2(f)\,\mathrm{d}f \\
&= 2A^2L\left(\begin{array}{l}\int_{-\frac{1+\alpha}{2T_s}}^{-\frac{1-\alpha}{2T_s}} \frac{1+\cos\frac{\pi(-2T_sf-1+\alpha)}{2\alpha}}{2}\mathrm{d}f \\ +\int_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}} \frac{1+\cos\frac{\pi(2T_sf-1+\alpha)}{2\alpha}}{2}\mathrm{d}f + \int_{-\frac{1-\alpha}{2T_s}}^{\frac{1-\alpha}{2T_s}} 1\mathrm{d}f\end{array}\right) \\
&= 2A^2L\left(\mathrm{A}_0+\mathrm{B}_0+\mathrm{C}_0\right)
\end{aligned} \tag{42}
$$

where F[·] denotes Fourier transform, and A_0, B_0, C_0 are defined as follows:

$$
\begin{aligned}
\mathrm{A}_0 &= \int_{-\frac{1+\alpha}{2T_s}}^{-\frac{1-\alpha}{2T_s}} \frac{1+\cos\frac{\pi(-2T_sf-1+\alpha)}{2\alpha}}{2}\mathrm{d}f \\
\mathrm{B}_0 &= \int_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}} \frac{1+\cos\frac{\pi(2T_sf-1+\alpha)}{2\alpha}}{2}\mathrm{d}f \\
\mathrm{C}_0 &= \int_{-\frac{1-\alpha}{2T_s}}^{\frac{1-\alpha}{2T_s}} 1\mathrm{d}f
\end{aligned} \tag{43}
$$

Now, we will calculate A_0, B_0, C_0 separately:

$$
\begin{aligned}
\mathrm{A}_0 &= \frac{1}{2}\left[\frac{1+\alpha}{2T_s} - \frac{1-\alpha}{2T_s} + \int_{-\frac{1+\alpha}{2T_s}}^{-\frac{1-\alpha}{2T_s}} \cos\frac{\pi(2T_sx+1-\alpha)}{2\alpha}\mathrm{d}x\right] \\
&= \frac{1}{2}\left[\frac{\alpha}{T_s} + \frac{\alpha}{\pi T_s}\sin\frac{\pi(2T_sx+1-\alpha)}{2\alpha}\Big|_{-\frac{1+\alpha}{2T_s}}^{-\frac{1-\alpha}{2T_s}}\right] \\
&= \frac{1}{2}\left[\frac{\alpha}{T_s} + \frac{\alpha}{\pi T_s}\sin\pi\right] \\
&= \frac{\alpha}{2T_s}
\end{aligned} \tag{44}
$$

$$
\begin{aligned}
\mathrm{B}_0 &= \frac{1}{2}\left[\frac{1+\alpha}{2T_s} - \frac{1-\alpha}{2T_s} + \int_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}} \cos\frac{\pi(2T_sx+1-\alpha)}{2\alpha}\mathrm{d}x\right] \\
&= \frac{1}{2}\left[\frac{\alpha}{T_s} + \frac{\alpha}{\pi T_s}\sin\frac{\pi(2T_sx+1-\alpha)}{2\alpha}\Big|_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}}\right] \\
&= \frac{1}{2}\left[\frac{\alpha}{T_s} + \frac{\alpha}{\pi T_s}\sin\pi\right] \\
&= \frac{\alpha}{2T_s}
\end{aligned} \tag{45}
$$

$$
\mathrm{C}_0 = \frac{1-\alpha}{2T_s} + \frac{1-\alpha}{2T_s} = \frac{1-\alpha}{T_s} \tag{46}
$$

Then, we come to the result that,

$$
B(0) = B(f)\big|_{f=0} = 2A^2L\left(\mathrm{A}_0+\mathrm{B}_0+\mathrm{C}_0\right) = \frac{2LA^2}{T_s} \tag{47}
$$

APPENDIX B
PROOF OF (7)

According to the symmetry, we know that,

$$
B(-1/T_s) = B(1/T_s) \tag{48}
$$

and we can calculate their values as follows:

$$\begin{aligned}
&B\left(-1/T_s\right)\\
&\quad= B\left(1/T_s\right)\\
&\quad= B\left(f\right)\big|_{f=1/T_s}\\
&\quad= \mathrm{F}\left[b_n^2\left(t\right)\right]\big|_{f=1/T_s}\\
&\quad= 2LA^2\left[\mathrm{G}\left(f\right)*\mathrm{G}\left(f\right)\right]\big|_{f=1/T_s}\\
&\quad= 2LA^2\int_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}}\cos\left[\frac{\pi T_s}{2\alpha}\left(f-\frac{1-\alpha}{2T_s}\right)\right]\\
&\qquad\cdot\cos\left[\frac{\pi T_s}{2\alpha}\left(-f+\frac{1}{T_s}-\frac{1-\alpha}{2T_s}\right)\right]\mathrm{d}f\\
&\quad= 2LA^2\int_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}}\cos\left[\frac{\pi T_s}{2\alpha}\left(f-\frac{1-\alpha}{2T_s}\right)\right]\\
&\qquad\cdot\cos\left[\frac{\pi T_s}{2\alpha}\left(f-\frac{1+\alpha}{2T_s}\right)\right]\mathrm{d}f\\
&\quad= 2LA^2\int_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}}\frac{1}{2}\left\{\cos\left[\frac{\pi T_s}{2\alpha}\left(2f-\frac{1}{T_s}\right)\right]+\cos\frac{\pi}{2}\right\}\mathrm{d}f\\
&\quad= 2LA^2\int_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}}\frac{1}{2}\cos\left(\frac{\pi T_s}{\alpha}f-\frac{\pi}{2\alpha}\right)\mathrm{d}f\\
&\quad= \frac{\alpha LA^2}{\pi T_s}\sin\left(\frac{\pi T_s}{\alpha}f-\frac{\pi}{2\alpha}\right)\Big|_{\frac{1-\alpha}{2T_s}}^{\frac{1+\alpha}{2T_s}}\\
&\quad= \frac{2LA^2}{T_s}\cdot\frac{\alpha}{\pi}
\end{aligned}\tag{49}$$

APPENDIX C
PROOF OF (9)

Firstly, we should give out the calculation of $S(f)$, which is the Fourier transform of $s_{cross}^2(t)$:

$$\begin{aligned}
S\left(f\right) &= \int_{-\infty}^{\infty} s_{cross}^2\left(t\right)e^{-j2\pi ft}dt\\
&= A^2\sum_n\sum_{m\neq n}a_na_m\int_{-\infty}^{\infty}\mathrm{g}(t-nT_s)\mathrm{g}(t-mT_s)e^{-j2\pi ft}dt\\
&= A^2\sum_n\sum_{m\neq n}a_na_me^{-j2\pi fnT_s}\int_{-\infty}^{\infty}\mathrm{g}(t)\mathrm{g}(t-(n-m)T_s)\\
&\quad\times e^{-j2\pi ft}dt\\
&= A^2\sum_n\sum_{m\neq n}a_na_me^{-j2\pi fnT_s}G_{n-m}\left(f\right)
\end{aligned}\tag{50}$$

where

$$\mathrm{G}_{n,m}\left(f\right)=\int_{-\infty}^{\infty}\mathrm{g}(t)\mathrm{g}(t-(n-m)T_s)e^{-j2\pi ft}dt\tag{51}$$

Then, we have,

$$\begin{aligned}
&|S\left(f\right)|^2\\
&\quad= S\left(f\right)S^*\left(f\right)\\
&\quad= \sum_n\sum_m\sum_k\sum_l a_na_mb_kb_le^{j2\pi f(k-n)T_s}\mathrm{G}_{n-m}(f)\mathrm{G}_{k-l}^*(f)
\end{aligned}\tag{52}$$

Thus, the corresponding expectation is as follows:

$$\begin{aligned}
&E[|S(f)|^2]\\
&= \sum_n\sum_{m\neq n}\sum_k\sum_{l\neq k}E(a_na_mb_kb_l)e^{j2\pi f(k-n)T_s}\mathrm{G}_{n-m}(f)\mathrm{G}_{k-l}^*(f)
\end{aligned}\tag{53}$$

Since a_n, a_m, b_k, b_l are IID random sequences in $\{1,-1\}$, it is easy to know,

$$d=a_na_mb_kb_l=\begin{cases}1 & \text{P=0.5}\\ -1 & \text{P=0.5}\end{cases}\tag{54}$$

Then, we can conclude that,

$$E(a_na_mb_kb_l)=0\tag{55}$$

That is to say,

$$P\left(f\right)=\lim_{L\to\infty}\frac{E\left[|S\left(f\right)|^2\right]}{(2L+1)T_s}=0\tag{56}$$

REFERENCES

[1] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Personal Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[2] O. A. Dobre, A. Abdi, and Y. Bar-Ness, "Survey of automatic modulation classification techniques: Classical approaches and new trends," *IET Commun*, vol. 1, no. 2, pp. 137–156, 2007.

[3] J. Reichert, "Automatic classification of communication signals using high order statistics," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, Mar. 1992, p. 221.

[4] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[5] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Sparse signal detection from incoherent projections," in *Proc. IEEE Int. Conf. Acoust. Speed Signal Process.*, May 2006, p. 3.

[6] C. W. Lim and M. B. Wakin, "Automatic modulation recognition for spectrum sensing using nonuniform compressive samples," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2012, pp. 3505–3510.

[7] S. Liu, Y. D. Zhang, T. Shan, and R. Tao, "Structure-aware Bayesian compressive sensing for frequency-hopping spectrum estimation with missing observations," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2153–2166, Apr. 2018.

[8] S. Liu, Z. Mao, Y. D. Zhang, and Y. Huang, "Rank minimization-based Toeplitz reconstruction for DoA estimation using coprime array," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2265–2269, Jul. 2021.

[9] Z. Tian, "Cyclic feature based wideband spectrum sensing using compressive sampling," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2011, pp. 1–5.

[10] C. W. Lim and M. B. Wakin, "Compressive temporal high order cyclostationary statistics," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2942–2956, Mar. 2015, doi: 10.1109/TSP.2015.2415760.

[11] L. Zhou and H. Man, "Wavelet cyclic feature based automatic modulation recognition using nonuniform compressive samples," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–6.

[12] L. Zhu, C. Luo, and J. H. McClellan, "Cyclostationarity-based wideband spectrum sensing using random sampling," in *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 1202–1205.

[13] Z. Xing, J. Zhou, J. Ye, J. Yan, L. Zou, and Q. Wan, "Automatic modulation recognition of PSK signals using nonuniform compressive samples based on high order statistics," in *Proc. IEEE Int. Conf. Communiction Problem-Solving*, Dec. 2014, pp. 611–614.

[14] Z. Xing, J. Zhou, J. Ye, J. Yan, J. Zou, L. Zou, and Q. Wan, "Automatic modulation recognition of PSK signals with sub-nyquist sampling based on high order statistics," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2014, pp. 1–7.
[15] *ACM SIGKDD and Netflix, Proceedings of KDD Cup and Workshop*. Accessed: 2022. [Online]. Available: http://www.cs.uic.edu/∼liub/KDD-cup-2007/proceedings.html
[16] J. d. Abernethy, F. Bach, T. Evgeniou, and J.-P. Vert, "Low-rank matrix factorization with attributes," 2006, *arXiv:cs/0611124*.
[17] Y. Amit, M. Fink, N. Srebro, and S. Ullman, "Uncovering shared structures in multiclass classification," in *Proc. 24th Int. Conf. Mach. Learn.*, Providence, RI, USA, Jun. 2007, pp. 17–24.
[18] A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task feature learning," in *Proc. Adv. Neural Inform. Process. Syst.*, 2007, pp. 41–48.
[19] M. Mesbahi and G. P. Papavassilopoulos, "On the rank minimization problem over a positive semidefinite linear matrix inequality," *IEEE Trans. Autom. Control*, vol. 42, no. 2, pp. 239–243, Feb. 1997.
[20] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, Nov. 1992.
[21] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, "A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 289–301, Jan. 2009.
[22] H. Ghasemi, M. Malek-Mohammadi, M. Babaie-Zadeh, and C. Jutten, "SRF: Matrix completion based on smoothed rank function," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 3672–3675.
[23] M. Malek-Mohammadi, M. Babaie-Zadeh, A. Amini, and C. Jutten, "Recovery of low-rank matrices under affine constraints via a smoothed rank function," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 981–992, Feb. 2014.
[24] H. Wang, R. Zhao, Y. Cen, and F. Zhang, "Low-rank matrix recovery based on smooth function approximation," in *Proc. IEEE 13th Int. Conf. Signal Process. (ICSP)*, Nov. 2016, pp. 729–732.
[25] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, and L. Fang, "Cloud removal in remote sensing images using nonnegative matrix factorization and error correction," *ISPRS J. Photogramm. Remote Sens.*, vol. 148, pp. 103–113, Feb. 2019.
[26] S. M. Kay and S. L. Maple, "Spectrum analysis—A modem perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.
[27] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1997, ch. 1.
[28] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[29] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Dec. 2009.
[30] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
[31] S. Chauhan and K. V. Prema, "Effect of dimensionality reduction on performance in artificial neural network for user authentication," in *Proc. 3rd IEEE Int. Advance Comput. Conf. (IACC)*, Feb. 2013, pp. 22–23.
[32] S. Shaobing, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.
[33] M. Plumbley, "Recovery of sparse representations by polytope faces pursuit," in *Proc. Int. Conf. Independ. Compon. Anal. Blind Source Separat.*, 2006, pp. 206–213.
[34] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approx.*, vol. 13, no. 1, pp. 57–98, 1997.
[35] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, Apr. 2004.
[36] D. M. Malioutov, M. Cetin, and A. S. Willsky, "Homotopy continuation for sparse signal representation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2005, p. 733.
[37] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA J. Numer. Anal.*, vol. 20, no. 3, pp. 389–404, Jul. 2000.
[38] A. Yang, A. Ganesh, S. Sastry, and Y. Ma. *Fast L1-Minimization Algorithm and Application in Robust Face Recognition: A Review*. Technical Report No. UCB/EECS-2010-3. Accessed: 2022. [Online]. Available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-13.html
[39] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Commun. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008.
[40] E. J. Candés, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
[41] C. C. Chang and C. J. Lin. (2001). *LIBSVM: A Library for Support Vector Machines*. [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm

ZHENGLI XING received the B.Eng. degree in engineering physics from Tsinghua University, in 2012, and the M.Eng. degree in communication and information system from the China Academy of Engineering Physics (CAEP), in 2015. Since 2015, he has been an Assistant Researcher with CAEP. His current research interests include communication signal detection, classification, parameter estimation, and compressed sensing.

JIE ZHOU received the B.S. degree from Sichuan University, Chengdu, China, in 1995, and the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, in 2013. He is currently with the China Academy of Engineering Physics, Mianyang, China, as a Research Fellow. His main research interests include wireless communications and tracking telemetry command and communication.

ZHAN GE received the B.S. degree in automation from Sichuan University, Chengdu, China, in 2009, and the M.S. degree from Xi'an Jiaotong University, Xi'an, China, in 2012. He is currently with the Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang, China, as a Research Associate. His current research interests include intelligent signal processing and machine learning.

GUANQIN HUANG received the B.S. degree in information countermeasure technique and the M.S. degree in underwater acoustical engineering from Northwestern Polytechnical University, Xi'an, China, in 2010 and 2013, respectively.

From 2014 to 2022, he was a Research Assistant with the China Academy of Engineering Physics, Mianyang, China. His research interests include digital signal processing and wireless communication network technology.

MAOHAI HU received the B.S. degree in communication and information system and the M.S. degree in communication and information system from Harbin Engineering University, Harbin, China, in 2005 and 2008, respectively.

From 2008 to 2022, he was a Research Fellow with the China Academy of Engineering Physics, Mianyang, China. His research interest includes digital signal processing.