Application of Fractional Fourier Transform for Prediction of Ball Mill Loads Using Acoustic Signals

JIANQUAN SHI, GANGQUAN SI, AND YANBIN ZHANG, (Member, IEEE)
State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi’an Jiaotong University, Xi’an 710049, China
Corresponding author: Gangquan Si (sigangquan@mail.xjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61304118, in part by the Program for New Century Excellent Talents in University under Grant NCET-13-0456, and in part by the Specialized Research Fund for Doctoral Program of Higher Education of China under Grant 20130201120011.

ABSTRACT
In cement plant and power plant, ball mills remain in current and widespread use. The load parameter inside a ball mill directly impacts the stability of the production process, the grinding production rate, and the quality of the product in the grinding process. Accurately predicting the load from acoustic signals remains a challenging problem because of the nonlinearity and high dimensions of spectral data. In this paper, the application of fractional Fourier transforms on acoustic signals for estimating mill load parameter was researched. A fractional Fourier transform can give intermediate time–frequency representations by controlling an additional order, and the acoustical frequency spectra in the fractional Fourier domain can provide more information about the load parameters. According to the distribution of acoustic frequency spectra in the fractional Fourier domain, the strategies of predicting ball mill loads were divided into three segments, namely feature extraction, offline modeling, and online monitoring. These techniques included an acoustic signal analysis in different fractional orders, feature extracted based on mutual information and kernel principal component analysis, offline soft measuring modeling compared with other regression models, and online adaptive monitoring based on the optimal fractional order. The experimental investigation of the proposed method demonstrates its effectiveness for estimating mill loads in the fractional Fourier domain by comparing with the result in the Fourier domain.

INDEX TERMS Feature extraction, modeling, parameter estimation, signal processing, soft sensor.

I. INTRODUCTION
Large tumbling coal ball consumes 40% of their energy in their pulverizing system. However, most of the energy is wasted in rotating the heavy mill shell and balls. No more than 1% of the supplied energy is used for comminution [1]. The mill load is defined as the ratio of instantaneous mill size to maximum mill load volume. Overloading causes “mill blockage”, “belly be filled full”, and production process breaks off. Conversely, low loading results in “mill running with only ball mill”, leads to power energy waste, increases steel consumption, and damages grinding devices [2]. Consequently, it is important to predict the mill loads to maintain their optimal working conditions.

In practice, direct measurements of the parameters require costly instrumentation. Many researchers have taken an interest in “soft” measurement of mill loads, and many signals are used to estimate the loads such as the motor current signal, the acoustic signals, the vibration signals, and the negative pressure at the inlet of the mill shell [3]–[6]. For acoustic signals, the main characteristics are non-stationary, nonlinear, and multiple components since the complexity of powder mechanism in pulverizing, and finding mill load information in the time domain from it have proved difficult. However, the relation between mill load information and acoustic signals can be obtained in Fourier frequency domain.

The fast Fourier transform (FFT) is the most widely used for time-frequency transformation [7], but it is proposed to process stationary and linear signals. Because wavelet transforms can offer resolutions for prolonged
duration disturbances and short duration disturbances, they are also used to monitor and diagnose working conditions [8]. Recently, Hilbert-Huang transforms have been studied as a means to eliminate the redundant and irrelevant components of acoustic signals in the time domain [9], [10]. Motivated by the fractional Fourier transform (FrFT) used for solving non-stationary signals in many fields, such as image retrieval [11], speech recognition [12], and signal encryption [13], the relation between loads and acoustic signals in the fractional Fourier domain is researched in this paper.

In contrast with other proposed FrFT application methods, multiple orders are selected instead of the optimal order. Acoustic signals are transformed into the fractional Fourier domain with fixed orders, obtaining the acoustic spectrum data. Since the acoustic spectrum has the dimensionality problem, multi-scale, and multiple component characteristics [2], there are still many challenges on soft sensing mill load parameters, such as how to extract the candidate features from acoustic spectrums, how to select the modeling technology, and how to achieve the online monitoring. Therefore, three parts, such as feature selection and extraction, offline modeling, and online tracking, has been researched based on the research of the predecessors.

To overcome the dimensionality problem, the dimensional reduction is done to obtain sensitive frequencies or informative features [14]. Some relevant methods for the selection of characteristic frequencies and feature extraction have already been proposed [15]–[19]. A practical method of feature selection and extraction based on those methods is put forward to obtain the candidate features in the fractional domain. The proposed method employs some data mining techniques, such as mutual information (MI), statistical characteristics, and kernel principal component analysis (KPCA).

After candidate features are obtained, offline models can be trained for predicting mill loads. There are many modeling technology methods, including least squares support vector machine (LSSVM), neural networks (NN), partial least squares (PLS), and support vector machines (SVMs) [20]–[23]. Kernel mapping and structural risk minimization provide the common modeling strategy used in SVMs, PLS, and LSSVMs. However, in LSSVMs, the quadratic programming problem is formulated as a general equality constraint, instead of inequality constraints, resulting in higher generalization and less complexity. Therefore, we select the LSSVM as the offline modeling strategy, and some comparison experiments are compared to prove its feasibility.

In online operation, noise signals, such as the starting and stopping of adjacent mills, the action of ventilation doors, the quality and moisture of the coal, and the wear of steel balls [24], [25], unavoidably interfere with acoustic signals. As a result, the prediction of offline models may be inaccurate. More importantly, those interferences are not monitored in real time, making it impossible to build compensation models to correct for offline models’ inaccuracy. However, this problem can be solved in the fractional Fourier domain by using a fractional spectrum subtraction algorithm and multilevel filters.

Therefore, the study of perspective for acoustic signals, feature selection and extraction methods, and the methods for disturbance cancelling are the main innovation of this paper. To testify the effectiveness of these proposed methods, some experiments are investigated on the acoustic signals obtained from an industrialized ball mill (DTM 350/700) at the QinLing Power Plant in China. The experimental results show the effectiveness of estimating the mill load by acoustic signals in the fractional Fourier domain.

The remaining parts are organized as follows: Section II mainly introduces the relevant knowledge of pulverizing systems, the FrFT, and the algorithm of fractional spectrum subtraction. Section III describes the processing method for acoustic signals. Section IV verifies the application of the proposed method and the experimental results. Section V presents the conclusion and next future work.

II. RELEVANT KNOWLEDGE

A. DESCRIPTION OF MILL LOAD AND ACOUSTIC SIGNAL

The mill load represents the output ability and the running condition of a ball mill pulverizing system, and usually has three statuses: low load, normal load, and overload. A graphic representation of a pulverizing system demonstrates in Fig. 1. The inputs of a coal ball mill are hot air, recycled air, raw coal, and unqualified powder. In the grinding process, the balls are thrown up by the rotating mill, and then fall striking the coals. Finally, the coal powder is transferred into a coarse-pulverized-coal separator and a fine-pulverized-coal separator, and the unqualified powder is returned to the ball mill [26]. Therefore, the mill load is a key parameter for achieving closed-loop control of the pulverizing system. Detailed information on the acoustic signal was provided in [27]. According to the crashing mechanism, acoustic signals are classified as acceleration noise and ringing noise. The acceleration noise is caused by the changing speed and the pressure disturbance, which is defined as

\[ L_p = 10 \log v^2 + 7 \log v - 20 \log r - 10 \log \left( \frac{c t_0}{V^{1/3}} \right) + 4 \]  

where \( v \) is the impact velocity, \( V \) is the volume, \( c \) is the velocity of sound, \( r \) is the measuring distance, and \( t_0 \) is the impact time. The ringing noise is caused by the crack of the balls against the lining plate, and can be obtained by

\[ L_{eq} = 10 \log \left[ \frac{d F(f_0)}{dt} \right]^2 + 10 \log \left[ H_c(f_0) e^{-j 90} \right] + 10 \log \left( \frac{\alpha_{rad}}{f_0} \right) - 10 \log \eta_s - 10 \log d + B \]  

where \( F(f_0) \) is the impact force caused by the crack of the balls and the coal in the cylinder, \( H_c(f_0) \) is the mill shell-response function, \( a \) is the weighting coefficient of the A sound level, \( \alpha_{rad} \) is the sonic amplitude coefficient, \( \eta_s \) is the damping coefficient, \( d \) is the thickness of mill, and \( B \) is a constant.
B. FRACTIONAL FOURIER TRANSFORM

The Fourier transform is applied widely in the signal processing field and can be regarded as a linear transformation operator. For time-frequency transformation, if the Fourier transform is considered as a counterclockwise rotation $\frac{\pi}{2}$, the FrFT can be explained as a linear operator, counterclockwise rotation of the real axis of $\theta = \alpha \times \frac{\pi}{2}$. The parameter $\alpha$ is called the order of the FrFT. Therefore, acoustic signals may be analyzed in the fractional Fourier frequency domain by controlling the order from 0 to 1, not just in the time domain or the Fourier frequency domain.

The $\alpha^{th}$ order FrFT is an integral transform

$$F^\alpha[f(\mu)] = \int_{-\infty}^{+\infty} K_\alpha(\mu, x)f(x)dx$$ (3)

where $K_\alpha(\mu, x)$ is the transform kernel of the continuous FrFT given by

$$K_\alpha(\mu, x) = \begin{cases} f(\theta) & 0 < |\alpha| < 2 \\ \delta(\mu - x) & \alpha = 0 \\ \delta(\mu + x) & \alpha = \pm 2 \\ \end{cases}$$ (4)

$$f(\theta) = \sqrt{1 - i \cot(\theta)} \exp(i\pi(\cot(\theta)x^2 + \mu^2) - 2\csc(\theta)\mu x))$$ (5)

where $i$ is the imaginary unit. When $\alpha = 1$, the FrFT is the same as the FFT. The FrFT has many properties; for example, additivity, linearity, commutativity, and inversion. In this study, additivity was mainly used, and $\alpha \in (0, 1]$ was selected as the fractional Fourier domain.

For computation convenience, the discrete FrFT is defined through the spectral expansion analogous to the kernel:

$$F^\alpha[m, n] = \sum_{k=0}^{N-1} p_k[m](\lambda_k)^\alpha p_k[n]$$ (6)

where $p_k[n]$ is an orthonormal eigenvector set of the $N \times N$ discrete Fourier transform matrix, and $\lambda_k$ are the corresponding eigenvalues. Therefore, index additivity will be demonstrated by the orthonormality of $p_k[n]$, that is $F^{\alpha_1}F^{\alpha_2} = F^{\alpha_1+\alpha_2}$ [25].

C. FRACTIONAL SPECTRUM SUBTRACTION

Spectrum subtraction has been extensively used for enhancing speech; it can remove additive white Gaussian noise and improve the quality of the speech signal. In recent years, concerning the additivity, fractional spectral subtraction has been explored to estimate and remove the speech noise spectrum [28].

Let $l(n)$ be the frames of noisy speech samples:

$$l(n) = x(n) + d(n)$$ (7)

where $x(n)$ is the clean signals and $d(n)$ is the additive noise. When $x(n)$ and $d(n)$ are relatively independent, the equation...
can be depicted by the FrFT as

\[ L_{\alpha}(u) = X_{\alpha}(u) + D_{\alpha}(u) \]  

(8)

Let \( L_{\alpha}(u) = |L_{\alpha}| \exp(j\varphi_{\alpha,1}) \), \( X_{\alpha}(u) = |X_{\alpha}| \exp(j\varphi_{\alpha,x}) \), and \( D_{\alpha}(u) = |D_{\alpha}| \exp(j\varphi_{\alpha,d}) \), the following equation can be obtained:

\[ |L_{\alpha}| \exp(j\varphi_{\alpha,1}) = |X_{\alpha}| \exp(j\varphi_{\alpha,x}) + |D_{\alpha}| \exp(j\varphi_{\alpha,d}) \]  

(9)

where \( |L_{\alpha}|, |X_{\alpha}|, \) and \( |D_{\alpha}| \) are the \( \alpha \)th amplitude of frequency spectrum; \( \varphi_{\alpha,1}, \varphi_{\alpha,x}, \) and \( \varphi_{\alpha,d} \) denote the \( \alpha \)th phase characteristics of \( L_{\alpha}, X_{\alpha} \) and \( D_{\alpha} \) respectively. The amplitude of the clean signal is:

\[ |X_{\alpha}| = |L_{\alpha}| - |D_{\alpha}| \]  

(10)

If the noise signal is estimated correctly, the optimal order will be obtained by the minimum mean square error criterion [28]. Therefore, it is decisive for estimating the noise signal. Common noise estimating algorithms (e.g., minimum statistics (MS), minima controlled recursive averaging (MCRA), and improved MCRA) are based on the minimum statistic theory, which uses the minimum value instead of the additive noise [29]. Acoustic signals of a ball mill are different from those of a speech signal, whose minimum value cannot be regarded as the noise signal because a mill runs nonstop. A flowchart of the fractional spectrum subtraction concept is shown in Fig. 2.

**III. MEASURING METHOD BASED ON FRFT**

In this paper, acoustic signals can be regarded as additive signals composed of clean signals and “noise signals”. The clean signals are related directly to the mill load, and the noise signals are irrelevant to the mill load. Sometimes there is no assessment method for extracting the clean signals in the time domain or the Fourier domain because of signal coupling. But, in the fractional Fourier domain, the clean signals and the noise signals can be separated with a particular order. Visual representations of the coupling relation are shown in Fig. 3. If the noise signals can be estimated, the clean signals will be obtained by using a fractional spectrum subtraction algorithm. Unfortunately, the noise signals can’t be estimated in advance owing to the complexity of the compositions, so the optimal order cannot be calculated.

To address this problem, a series of orders were set up ahead, and multilevel offline models were established. The optimal model with the best performance was identified according to the real mill loads, and the optimal order corresponded to the optimal model. Therefore, the noise signals had been removed in the output of the optimal model. Considering the robustness of the prediction values, the final mill load was the data fusion result of these offline models based on adaptive weighting fusion algorithm.

In the online application, when the noise signals changed, the coefficients of these offline models must be recalculated. However, it was difficult to know when the noise signals had changed. The main disturbances influencing the noise signals were summarized: the starting or stopping of adjacent mills, the action of ventilation doors, the quality and moisture of the coal, and the wear of the steel balls. For the disturbance of the starting or stopping of adjacent mills, which can be monitored with human intervention. An additional acoustic sensor was installed to measure the background noise in the opposite direction, the location as shown in Fig. 1. When the adjacent mills are starting or stopping, the measured background noise was regarded as noise signals and removed using the fractional spectrum subtraction algorithm. For the other disturbances, there was no method to monitor or measure. According to the multilevel filter in the fractional domain, the optimal order was used to monitor whether a disturbance had occurred. If the optimal order were changed, the disturbances would have happened. The deviation caused by these disturbances was revised by recomputing the offline model weight coefficients.

The flowchart of the proposed method is displayed in Fig. 4, which consists of three units. The first one is candidate features extraction based on FrFT, characteristic frequencies selected, and KPCA. The second one is offline modeling based on LSSVM. The last one is online monitoring based on adaptive weighting fusion method and the optimal order.

**A. FEATURE SELECTION AND EXTRACTION**

Feature selection is done on the fractional frequency spectra, which can describe the magnitude and phase of acoustic signals. Many researchers are quite dedicated to mining acoustic signals in the Fourier domain to extract the most informative features, and they do many training tasks to estimate a mill load. Based on such research, the proposed method of feature selection and extraction first selects the characteristic frequency based on MI values. Then, as many statistical features as possible are extracted from the characteristic frequency. Finally, KPCA is used to mine the nonlinear features from the characteristic frequency and all statistical features. For comparison with the other methods, the basic methods and equations used are presented as follows:

1. Frequency sub-band selection based on MI: In this paper, one sub-band frequency contains a single frequency point, and two thresholds are given for selecting the
characteristic frequency. This is different from the method in [15], which directly selected candidate features based on a threshold.

2) Statistical characteristics: The main statistical features are selected: equal-weighting energy ($power$), mean square frequency ($msf$), peak to rms ratio ($prr$), mean frequency ($mf$), characteristic frequency to power ($cfp$), frequency variance ($var$), and root mean square ($rms$). The equations for the statistical features are

\[
\text{power}_i = \sum_{j=1}^{N} (S_{\alpha, ij})^2
\]  \hspace{1cm} (11)

\[
\text{mf}_i = \sum_{j=0}^{N} \frac{j \cdot S_{\alpha, ij}}{\sum_{j=0}^{N} S_{\alpha, ij}}
\]  \hspace{1cm} (12)

\[
\text{var}_i = \frac{1}{N} \sum_{j=1}^{N} (S_{\alpha, ij} - \bar{S}_{\alpha, i})^2
\]  \hspace{1cm} (13)

\[
\text{rms}_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (S_{\alpha, ij})^2}
\]  \hspace{1cm} (14)

\[
\text{prr}_i = \frac{\|S_{\alpha, i}\|_\infty}{\sqrt{\frac{1}{N} \sum_{j=1}^{N} |S_{\alpha, ij}|^2}}
\]  \hspace{1cm} (15)

\[
\text{msf}_i = \frac{\sum_{j=0}^{N} j^2 \cdot S_{\alpha, ij}^2}{\sum_{j=0}^{N} S_{\alpha, ij}^2}
\]  \hspace{1cm} (16)

\[
\text{cfp} = \frac{\text{power}_{j \in \text{feature}}}{\text{power}}
\]  \hspace{1cm} (17)
where \( i \) is the number of samples, \( S_a \) is the result of the FrFT with the order \( \alpha \), \( j \) is the number of frequency points, and the maximum value of \( j \) is \( N \). \( k_{\text{feature}} \) are the points belonging to characteristic frequencies, and these characteristics have to be analyzed further by KPCA.

(3) Feature extraction based on KPCA: The dimensions of the statistical characteristics and characteristic frequencies are compressed, and the nonlinear features from the statistical characteristics and characteristic frequencies are extracted.

B. OFFLINE MODELING

The acoustic signals \( l_1(t) \), \( l_2(t) \), and \( n(t) \) are obtained from the forepart sensor, the posterior part sensor, and the background noise sensor respectively. The background noise can be removed by the fractional spectrum subtraction algorithm, yielding the clean signals \( x_1(t) \), and \( x_2(t) \). After assigning a series of order \( \alpha \in (0,1] \), the frequency spectra \( S_{a,1} \), \( S_{a,2} \) are obtained by using the FrFT. Next, the candidate features, the principal components of the statistical characteristics (\( PC_{\text{stat}} \)) and the characteristic frequencies (\( PC_{\text{sub}} \)), are extracted based on feature selection and an extraction algorithm. The offline models are defined as

\[
\hat{y}_a = f_a(PC_{\text{stat}}, PC_{\text{sub}}) = \omega^T \varphi(x_{i,j-1,j}) + b \tag{18}
\]

where \( x_{i,j-1,j} \) represents \( PC_{\text{stat}}, PC_{\text{sub}} \). \( \varphi(x) \) maps \( x \) to a high-dimensional feature space, and \( \omega \) and \( b \) are obtained by solving the following optimization problem:

\[
\min J(\omega, e) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{k=1}^{N} e_k^2
\]

\[\text{s.t. } y_k = \omega^T \varphi(x_k) + b + e_k \tag{19}\]

where \( \gamma \) is the regularization parameter. After introducing an unconstrained Lagrangian function,

\[
L(\omega, b, e, \alpha) = J(\omega, e) - \sum_{k=1}^{N} \alpha_k (\omega^T \varphi(x_k) + b + e_k - y_k) \tag{20}
\]

where \( \alpha_k \) are the Lagrangian multipliers. The derivatives are set to equal zero.

\[
\begin{align*}
\frac{\partial L}{\partial \omega} &= 0 \rightarrow w = \sum_{k=1}^{N} \alpha_k \varphi(x_k) \\
\frac{\partial L}{\partial b} &= 0 \rightarrow \sum_{k=1}^{N} \alpha_k = 0 \\
\frac{\partial L}{\partial e_k} &= 0 \rightarrow \alpha_k = \gamma e_k \\
\frac{\partial L}{\partial \alpha_k} &= 0 \rightarrow \omega^T \varphi(x_k) + b + e_k - y_k = 0
\end{align*} \tag{21}
\]

After eliminating \( \omega \) and \( e_k \), the solution is represented by the set of linear equations:

\[
\begin{bmatrix}
0 \\
1
\end{bmatrix}
\begin{bmatrix}
\bar{T} \\
\Omega + \gamma^{-1} I
\end{bmatrix}
\begin{bmatrix}
b \\
\alpha
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
y
\end{bmatrix} \tag{22}
\]

where \( y = [y_1; \ldots; y_N] \), \( \bar{T} = [1; \ldots; 1] \), \( \alpha = [\alpha_1; \ldots; \alpha_N] \), and \( \Omega_{mn} = \varphi(x_m)^T \varphi(x_n) = K(x_m, x_n), m, n = 1, \ldots, N \), and the RBF kernel \( K(x_m, x_n) = \exp(-|x_m - x_n|^2/2\sigma^2) \) is chosen. Based on Mercer’s theorem, the LSSVM regression model can be obtained:

\[
f(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{23}
\]

where \( \alpha, b \) are solutions of Equation (21); All off-line models of different orders are trained by using Equation (23).

C. ONLINE MONITORING

The candidate models were selected from all offline models by setting an appropriate threshold for the mean square error. Then the estimated value of the mill load was calculated with these candidate models by using an adaptive weighting algorithm, and the formulas are as follow:

\[
\hat{y} = \sum \alpha_k \times \hat{y}_a \tag{24}
\]

\[
dy_a = \hat{y}_a - y \tag{25}
\]

\[
D_t = \{\alpha | MSE_a < \lambda\} \tag{26}
\]

\[
t_a = 1/\sigma^2_{\alpha} \sum_{\alpha} \frac{1}{\sigma^2_{\alpha}} \tag{27}
\]

where \( \hat{y} \) is an estimate of the mill load, \( t_a \) is the coefficient of the \( \alpha \)th prediction, \( \sigma^2_{\alpha} \) is the variance of \( dy \), and \( D \) is the set of candidate models. Commonly, if the \( \hat{y}_a \) is more adjacent to the real mill load, the weighting coefficient will be more significant. The optimal order corresponds to the largest value of the weighting coefficient. To strengthen the weight of the optimal order in the estimation, the weighting coefficient of the optimal order was multiplied by 2 when the values were \( \leq 0.5 \). The other order coefficient was halved.

For an online application, the operating state of adjacent mills can be monitored manually, because the starting or
stopping of adjacent mills does not occur frequently and requires a human operation. When the operating condition of the adjacent mill was updated, the spectrum subtraction algorithm filtered out the background noise generated by state changes. For the other disturbances, a semi-supervised method was used. Providing the real mill load periodically and verifying whether the optimal order changed during a certain time interval. When the optimal order changed, it was considered that the noise signals caused by interference were not negligible. This situation indicated that the noise signals could not be completely filtered out in the FrFT of the optimal order. Therefore, the reliability of the estimation result was improved by reselecting the candidate models and recalculating the weighting coefficients.

IV. APPLICATION RESEARCH
An industrial ball mill (DTM350/700) at the QinLing Power Plant in China was used to obtain the acoustic signals. It is a continuous grinding mill with rated revolution 17.57 per minute. Three acoustic sensors were MPA206 with the 20 ~ 10KHz frequency response range. The forepart and posterior part sensors were positioned toward the mill to obtain the mill acoustic signals, and the middle sensor faced the opposite direction to sample the environmental noise. Because the response time of the acoustic sensors was ≤200ms, the acoustic signals were extracted at sampling points within the region from 500 to 700ms with a sampling frequency of 51.2KHz. The acoustic signals, consisting of 1,200 groups, covered three states: low load, medium load, and high load. The details of the data processing are as follows: First, the acoustic signals were normalized in the time domain. Each signal was divided into 19 frames with each frame containing 1,024 samples, and the overlap of each frame was 512 samples; Next, these frames were processed using a Hamming window; Finally, the frequency spectra were averaged over several frames to overcome the problem of data acquisition error [19].

A. APPLICATION OF FEATURE SELECTION AND EXTRACTION
To display the performance of the fractional spectrum subtraction algorithm, acoustic signals with the condition of 70% mill load are represented in Fig. 5. The results of the spectrum subtraction compared with the unprocessed signals according to Equation. (10) are shown in Figs. 5a and 5b. The frequency spectra of the acoustic signals and the noise signal are shown in Figs. 5c and 5d. Fig. 5 shows that finding mill load information in the time domain is difficult, and the fractional spectrum subtraction algorithm filters out background noise signals and improves the frequency characteristics of the initial signals.

Based on the FrFT properties mentioned above, \( \alpha \in (0, 1] \) was chosen, and the fractional orders \( \alpha = [0.1, 0.2, \ldots, 1] \) were firstly preset. However, when the order was \( \leq 0.8 \), the characteristic frequencies would be \( \geq 10 \) KHz, which was the limit of the response frequency for acoustic sensors. Thus, the fractional orders were reselected from 0.8 to 1 and spaced by 0.02, making a total of 11 orders. For simplicity, only three results with \( \alpha = [0.8, 0.9, 1] \) are shown in Fig. 6a, and some partially enlarged views are shown in Fig. 6b.

There were a total of 10 groups of acoustic signals with different mill loads, such as 20%, 30%, 40%, 45%, 50%, 55%, 60%, 70%, 80%, 85%. It covered the low states, medium states, and high states. From Fig. 6a, those mill loads are undistinguished from the full spectrum. But in the part 4k - 10k Hz, different mill loads can be distinguished. Three

**FIGURE 6.** (a) Curves of acoustical frequency spectrum for different mill loads with different order; (b) Partial enlarged views corresponding with the same order of (a).
conclusions can be drawn from Fig. 6: (1) Different mill loads could be distinguished in the fractional Fourier domain. (2) For all orders, only a partial frequency (the “characteristic frequency”) could distinguish different mill loads. (3) The sets of characteristic frequencies were different for different orders.

Next, characteristic frequencies were selected based on MI values. Firstly, the acoustical frequency spectrum was split to 1,024 sub-bands with space out 50 Hz. Then, calculating the MI values about these sub-bands and the mill load, the set of characteristic frequencies contained those sub-bands with MI values greater than the defined threshold. To achieve consistency of the displayed results, the results with $\alpha = [0.8, 0.9, 1]$ were selected for illustration. The MI values were calculated using the method in [30], which is shown in Fig. 7a.

We ran some experiments to obtain the optimal threshold. The relation between the threshold and the prediction errors are exhibited in Fig. 7b, and the corresponding order and data in Fig. 7 are consistent with those in Fig. 6. Fig. 7 shows that: (1) the values of MI for different orders (e.g., the maximum MI and the amplitude) were different, and (2) the thresholds between different orders were different. In the modeling process, the dimension of the training data directly correlated to the time complexity, and the number of dimensions was decided by the threshold. Therefore, in Fig. 7b the optimal threshold is indicated by the arrow. The threshold and the number of characteristic frequencies for all orders are listed in Table 1. However, the threshold of MI was set to 0.2 in this study to extract statistical features.

Finally, candidate features are obtained as follows: (1) Characteristic frequencies were obtained by selecting the value of MI greater than the threshold of 0.2. (2) The statistical characteristics were calculated using Eq. (11-17). (3) The features were extracted based on KPCA, and the accumulative contribution rates of the three principal

<table>
<thead>
<tr>
<th>Order</th>
<th>0.8</th>
<th>0.82</th>
<th>0.84</th>
<th>0.86</th>
<th>0.88</th>
<th>0.9</th>
<th>0.92</th>
<th>0.94</th>
<th>0.96</th>
<th>0.98</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>0.45</td>
<td>0.45</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.45</td>
<td>0.55</td>
<td>0.5</td>
<td>0.55</td>
<td>0.55</td>
<td>0.6</td>
</tr>
<tr>
<td>Numbers</td>
<td>17</td>
<td>27</td>
<td>18</td>
<td>22</td>
<td>22</td>
<td>48</td>
<td>14</td>
<td>23</td>
<td>9</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>MSE</td>
<td>0.069</td>
<td>0.073</td>
<td>0.076</td>
<td>0.060</td>
<td>0.048</td>
<td>0.072</td>
<td>0.074</td>
<td>0.072</td>
<td>0.069</td>
<td>0.074</td>
<td>0.089</td>
</tr>
</tbody>
</table>
components (PCs) are ≥85%. For comparison, we selected features extracted from the all-bands frequency as comparative data. Scatter plots of those features for $\alpha = 0.9$ and $\alpha = 1$ are shown in Fig. 8.

As shown in Fig. 8, each class contained 50 samples, and a Gaussian function whose parameter equaled to 1 was used as the kernel of the KPCA. Fig. 8 shows that (1) three PCs were enough to distinguish the different mill loads, and (2) features extracted from characteristic frequency sub-bands and statistical characteristics could be used as the input of offline models. Therefore, the candidate features were constituted by three PCs of $PC_{\text{sub}}$ and $PC_{\text{stat}}$.

**B. APPLICATION OF OFFLINE MODELING**

The offline models were estimated with the LSSVM, PLS, SVM, and radial basis function (RBF) methods for predicting the mill loads from the extracted features. In the experiments, the datasets were divided into a training set and a testing set. The number of data in training set varied between 50 and 1,000 in steps of 50, and the number in the testing set was 100. All samples were randomly selected and normalized.

The LSSVM, PLS, SVM, and RBF methods were run in MATLAB version R2016b with default parameters used unless otherwise specified. For the RBF, three-layer topologies were used, and the RBF was used as the nonlinear activation function. A Gaussian function was used as the kernel function in the LSSVM and the SVM, and the hyperparameters were optimized by using a 10-fold cross-validation method. The performance metric was selected as the MSE and the correlation coefficient $R$. The MSE was computed concerning the testing sets, and the $R$ was computed with respect to the training sets. The modeling process was done 20 times for all orders.

Consistent with the results presented above, the results of orders 0.8, 0.9 and 1 are illustrated in Fig. 9. The MSE values are shown in the interval $[0, 1]$, and all the values are the average values of 20 experimental runs. These conclusions can be drawn: (1) The MSE value obtained with the LSSVM
FIGURE 9. The MSE and R values for different training samples by PLS, RBF, SVM and LSSVM modeling methods in 0.8, 0.9 and 1 order.

FIGURE 10. (a) The estimation results with the operation of ventilation door; (b) the estimation results with the start of adjacent mill.

on different support vectors was lower than those of PLS, SVM, and RBF, which confirms that the LSSVM has better predictive ability for other algorithms. (2) The R values of the LSSVM became closer to 1 with increasing sample size. Because nonlinear inequality constrained optimization problems were solved in the SVM, the sparse solution was trained by some support vectors, instead of all training samples. Consequently, the R values of the SVM in all training sets tended to become smaller as the sample size increased.

(3) When the number of training data was 600, the LSSVM algorithm maintained a correct estimation as high as 85%.

LSSVM, PLS, SVM, and RBF experiments on 600 training samples were run to confirm the effectiveness of the LSSVM in the fractional domain. In consideration of the hyperparameters being obtained by the cross-validation method and the randomly selected training samples, all experiments were trained ten times to mitigate the introduction of any accidental factors. The results of all orders are given in Table 2. For each
order, the MSE values decided the candidate models, and the threshold was set to 0.14. The weighting coefficients (WCs) for all orders were calculated using Eq. (27); these are also shown in Table 2. The experimental results verify that the LSSVM model is the most effective regression model for the extracted features. The results of the orders (0.8 $\leq$ 0.98) in a fractional domain, comparing with the result of the order 1 (FFT), have been mixed. But the result of the optimal order is better than the result of the FFT. For the fractional orders, the optimal order was considered to be the 0.96 order. The candidate models were in the orders 0.84, 0.88, 0.96, and 0.98.

C. APPLICATION OF ONLINE MONITORING

Based on the above analysis, the mill load was the fusion of the offline candidate models. It is clear that the optimal order corresponded to the largest weighting coefficient. The offline models mentioned above were established with no adjacent mill working. In this section, five groups of acoustic signals with a ventilation operation and an adjoining mill running were selected to evaluate the proposed method. In this experiment, five groups of signals with different mill loads were used as the testing data. The prediction results after feature selection and extraction are illustrated in Fig. 10, and the results of the LSSVM in all fractional orders are shown in Table 3.

Experimental results reveal the following: (1) The online monitoring method was more sensitive to the disturbance of the ventilation operation than to the start of an adjacent mill. This also shows that the spectral subtraction algorithm was effective. (2) The results of online monitoring were better than the results of offline models when disturbances occurred. (3) The results of the optimal model were the closest to the real mill load. In particular, when the weight value of the optimal model was $\lesssim$0.5, there were multiple models with similar precision. Therefore, the weight values of the optimal model could be increased to improve the performance of online method. The fusion result would improve the robustness of the predictive algorithm.

V. CONCLUSION

In this paper, we propose a series of measurement methods in the fractional Fourier domain for predicting the loads of tumbling ball mills. The methods comprise feature selection and extraction, offline modeling, and online monitoring. For feature extraction, in this study, the acoustic signals were displayed and researched in the fractional Fourier domain, and informative extracted features could split the different mill loads in high-dimensional space. The performance of the LSSVM, PLS, SVM, and RBF were demonstrated for a series of training data sets. The results verify the efficiency of the LSSVM for building offline models. For the disturbances in online monitoring, clean signals or extracted features without noise signals were obtained according to the additive property of the FrFT. When a disruption occurred, the predictive models were corrected by reselecting the offline candidate models and recalculating the adaptive coefficients, instead of retraining the offline models. Experimental results show that the online monitoring method was simple and effective in coping with the ventilation operation and the action of an adjacent mill.

REFERENCES


