Quantitative Diagnosis Method of Gearbox Under Varying Conditions Based on ARX Model and Generalized Canonical Correlation Analysis

LIUBANG HAN1,2, QIDONG WANG1,3, KUOSHENG JIANG1,2,4,5, XUANYAO WANG1,2,5, AND YUANYUAN ZHOU1,2

1School of Mechanical Engineering, Anhui University of Science and Technology, Huainan 232001, China
2State Key Laboratory of Mining Response and Disaster Prevention and Control in Deep Coal Mines, Anhui University of Science and Technology, Huainan 232001, China
3Hefei University, Hefei 232000, China
4State Key Laboratory for Manufacturing System Engineering, School of Mechanical Engineering, Xi’an Jiaotong University, Xi’an 710049, China
5Shaanxi Automobile Holding Group Huainan Special Purpose Vehicle Co., Ltd., Huainan 232001, China

This work was supported in part by the National Natural Science Foundation Project in China under Grant 51705003, in part by the Anhui Provincial Natural Science Foundation Project under Grant 1808085QE130 and Grant 1908085ME159, in part by the Anhui Provincial Education Department Fund under Grant 11673, in part by the State Key Laboratory of Mechanical Manufacturing System Engineering Open Research Fund under Grant sklms2018009, in part by the Research on Key Technology and Industrialization Application of Unmanned Intelligent Patrol Vehicle under Grant 1804a09020016, in part by the Anhui University of Science and Technology Innovation Fund under Grant 2019CX1004 and Grant 2019CX2030, and in part by the Anhui Provincial Key Laboratory of Intelligent Equipment and Technology Open Fund under Grant 201901004.

ABSTRACT Fault diagnosis of gearboxes under the condition of varying speed and varying load is a hotspot and difficulty in the research of gearboxes. The response signals of gearbox under varying conditions exhibit non-linear and non-stationary characteristics, which increase the complexity of quantitative diagnosis of gearbox faults. A quantitative diagnosis method of gearbox faults based on the improved autoregressive with exogenous (ARX) model and generalized canonical correlation analysis (GCCA) is proposed in this paper. The ARX model is improved based on incremental recursive identification of Kalman filter to build system transfer characteristic models using the excitation and response signals of gearboxes. ARX models of gearboxes are nonlinear and the GCCA is proposed to build the quantitatively relationship between models with faulty status and healthy status. Simulation and experiment results indicate that the proposed method can effectively identify the severity of the gearbox failures under varying conditions and provides a promising method for the quantitative diagnosis of rotating machinery.

INDEX TERMS Quantitative diagnosis, varying conditions, ARX, generalized canonical correlation analysis.

I. INTRODUCTION

The fault diagnosis of gearboxes under varying operation conditions has attracted the extensive attention of researchers [1]–[4]. The development trend of most gearboxes tends to be more integrated, and gearboxes will be replaced when the fault reaches a certain level. Therefore, diagnosis of the existence and severity of a fault is more in line with the engineering requirements. And, there are two key factors that need to be solved in the quantitative diagnosis method of gearbox faults under the variable working conditions. The first is the model of the gearbox system, and the second is a quantitative evaluation of the model.

Traditional gearbox models are typically built in two ways: the first one is a mass-spring-vibrator model [5]–[7], in which a linear approximation research method is used, and the other one is built via a subsystem modelling method [8], [9] to simplify the torque, force, and other factors in the relationship between the force and reaction in the system. An external description is often referred to as an output–input description. The starting point is to treat the gearbox system as a “black box”. Based on this premise, an external description avoids...
the characteristic information inside the system and directly reflects the dynamic causal relationship between the external variable groups of the system. A differential equation, transfer function model, and time-series model are common external models of the system.

The canonical correlation analysis (CCA) is a common pattern recognition method, which can realize the recognition of linear correlation components in random vectors [10], [11]. However, if there is a nonlinear relation between the variables, a CCA does not always extract useful features [12], [13]. A generalized canonical correlation analysis (GCCA) can highlight the nonlinear relationship between vectors, and can solve this contradiction. It is therefore more universal and more suitable for a fault diagnosis based on the system characteristics.

A typical dynamical system is composed of three main subsystems from the perspective of the transfer path of system response, namely: active sub-system; generating response signals such as noise and vibration, mechanical and airborne transfer paths and passive subsystem; receiving the noise and vibration [14]–[16]. The fault diagnosis methods can be divided into two categories: excitation-based and response-based methods. From the perspective of the load effect of the system for fault diagnosis and condition monitoring, an excitation-based fault diagnosis method mainly includes a method based on the stator current [17], an instantaneous speed analysis [18], [19], and a torsional vibration analysis [20], which mainly makes fault diagnosis from the perspective of system input, without considering the impact of system input on system output. A response-based fault diagnosis method mainly includes diagnosis method based on vibration application [21], [22], lubrication performance [23], acoustic signal [24], temperature signal [25] and the Hertz deformations [26] which are mainly used for a fault diagnosis and condition monitoring from the system output perspective, and do not care about the impact of the transfer path of response signals on the fault diagnosis.

The diagnosis method based on the system characteristics is studied in this paper, in order to overcome the lack of excitation-based and response-based methods. From the perspective of system characteristics, the system-based diagnostic method comprehensively considers the effects of input signals, output signals and signal transmission paths on fault characteristics. The transfer function is an inherent property of the system, which will not change under varying conditions, and only change when the gearbox fails. The stator current and vibration acceleration signals are respectively selected as the excitation and response signals, and a transfer function is introduced to model the gearbox system characteristics under varying conditions.

The variational mode decomposition (VMD) is recognized as a powerful tool for analyzing nonlinear and nonstationary signals and hence is extensively employed for mechanical vibration analysis [22], [27]. Combined with order analysis it can effectively diagnose the gearbox failure [28]–[30]. It can be used as a powerful reference for the diagnostic effect of the method proposed in this paper. In this study, an ARX model of incremental recursive identification (IRI) based on a Kalman filter (KF) is proposed, which can better reflect the real status of the system, improve the stability of the model, and achieve a strong anti-interference capability under varying starting and stopping conditions. Aiming at the problem of a non-linear quantitative identification of the ARX model, a nonlinear quantitative evaluation method based on GCCA is proposed to extract the non-linear correlativity of the system transfer function between gearbox features under a faulty status and those under a healthy status, and the quantitative evaluation results is expressed using the canonical correlation coefficients.

A flow chart of the proposed method is shown in Figure 1. To elaborate on this method, the input and output signals of the gearbox under varying conditions are first determined in Section 2.1. In Section 2.2, a model of the system characteristics of the gearbox is established from the perspective of a transfer function and a Laplace transform. Based on this, a gearbox system characteristic model based on ARX is described in Section 2.3. And in Section 2.4, an improved ARX model for an incremental recursive system parameter identification based on the application of a KF is proposed. The improved model verified through a simulation is described in Section 3.1, and a nonlinear quantitative evaluation method based on a GCCA is described and compared it with the analysis results of CCA, in Sections 3.2 and 3.3. An experimental verification of the proposed method is provided and compared with the traditional fault diagnosis method in Section 4. Finally, the conclusions are drawn in Section 5.

**II. MODELLING OF GEARBOX SYSTEM UNDER VARYING CONDITIONS**

Under variable operating conditions, the dynamic signals representing mechanical faults not only have non-linear...
characteristics, they also have time-varying characteristics, creating many difficulties in the feature extraction and fault diagnosis of a gearbox [31]. On the one hand, nonlinear coupling may occur between certain frequencies of the excited dynamic signals owing to mechanical component failures or defects [32]. On the other hand, the fault frequency under variable operating conditions is a function of time and follows changes with the working conditions [33]. Therefore, both changes in the operating conditions and mechanical failures can cause changes in the nonlinear coupling relationship and coupling degree of the signal, which is the result of their combination, rather than a simple superposition, which brings about new challenges to the modelling of a gearbox.

**A. SELECTION OF INPUT AND OUTPUT SIGNALS**

1) **SELECTION OF INPUT SIGNAL FOR VARYING CONDITIONS GEARBOX SYSTEM IDENTIFICATION**

A monitoring and diagnosis method based on a current signal was first applied to a fault diagnosis of a motor body structure. Numerous research results have shown that a diagnosis method based on a motor stator current can be used to effectively diagnose and analyze equipment faults.

Because the current contains torsional vibration and frequency information of the power supplied to the AC motor, torsional vibration information of the motor can be reflected in the stator current. Any torsional vibrations with a specific frequency will produce a sideband around the fundamental frequency component of the current signal. Moreover, the size of the sideband reflects the magnitude of the input torque, and the width of the sideband represents the frequency of the input torsional vibration. The existence of a gear fault and its location can be qualitatively distinguished by identifying the torsional vibration frequency. Therefore, the stator current is selected as the input signal of the ARX model of the gearbox under the starting and stopping conditions.

2) **SELECTION OF OUTPUT SIGNAL FOR VARYING CONDITIONS GEARBOX SYSTEM IDENTIFICATION**

The vibration acceleration signal contains the characteristic signals related to the faults and working conditions, and is widely used in fault diagnosis and condition monitoring. The vibration response measured at the surface of the gearbox housing is mainly composed of the gear meshing frequency, the characteristic frequency of faulty gears, the bearing frequency, the motor, and the interference caused by the complicated transmission path between the excitation source and the sensor. Because the response function of the gearbox transmission path is a function of frequency, the vibration response and side frequency feature of a gearbox vary with the gearbox speed.

Figure 2 shows a time–frequency diagram of the vibration acceleration with a sinusoidal change of the given velocity, which further verifies that the vibration acceleration signal contains the speed information. In addition, the vibration-based acceleration signal contains the characteristic information not only related to the fault, but also related to the working conditions.

**B. MODELLING OF GEARBOX SYSTEM**

The gearbox is a continuous-time dynamic system. For a gearbox system with a continuous single-input and single-output, if the input of the system is $u(t)$ and the output is $y(t)$, the corresponding differential equation can be expressed as follows:

$$
\frac{d^n y(t)}{dt^n} + a_{n-1} \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots + a_1 \frac{d y(t)}{dt} + a_0 y(t) = b_m \frac{d^m u(t)}{dt^m} + b_{m-1} \frac{d^{m-1} u(t)}{dt^{m-1}} + \cdots + b_1 u(t) + b_0 u(t) \tag{1}
$$

Taking the Laplace transform on both sides,

$$
G(s) = \frac{Y(s)}{U(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_m}{a_0 s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n} \tag{2}
$$

Therefore, the characteristic transfer function of the gearbox system can be uniquely determined by two coefficient vectors.

$$
P = [b_0, b_1, \ldots b_{m-1}, b_m] ; \quad Q = [a_0, a_1, \ldots a_{n-1}, a_n] \tag{3}
$$

From the perspective of system diagnostics, vectors $P$ and $Q$ represent the inherent characteristic of the gearbox. In the case of a normal gearbox during the starting and stopping process, the vectors $P$ and $Q$ should remain unchanged. Once the vectors $P$ and $Q$ change significantly, the gearbox becomes faulty.

**C. MODELLING OF GEARBOX SYSTEM BASED ON ARX**

To establish a system model, it is necessary to select different system identification methods. Auto regressive (AR) and auto regressive and moving average model (ARMA) are frequently used methods for establishing the parameter models. ARX is a controlled autoregressive model that contains not only the autoregressive term but also the inputs that the AR and ARMA models do not have. Therefore, the ARX models can fuse the function of analyzing the influence of the input on the output, which is compatible with the condition monitoring, and a fault diagnosis based on the system characteristics [2], [34]–[36].

According to the model described in Section 3.1, to establish a system ARX model reflecting the status of the gearbox. The motor stator current and vibration response signals are...
used as the input and output variables of the ARX system, respectively. It is assumed that the known output and input variables of the gearbox are $y(1), y(2) \cdots y(N + n)$ and $u(1), u(2) \cdots u(N + m)$, respectively, and the output and input orders of the gearbox system characteristics are $n$ and $m$.

The error of the system model is $\Delta(t)$, which assumes that the model parameters are $A_i$ and $B_j$. A series of systems composed of $N$ equations can be obtained by combining the time-series relation of the input and output of the gearbox:

$$
y = \Phi \Delta + \theta$$

- $\Phi = \begin{bmatrix} \phi(n + 1) \\ \phi(n + 2) \\ \vdots \\ \phi(N) \end{bmatrix}$
- $\Delta = \begin{bmatrix} -y(n) & \cdots & -y(1) & u(m) & \cdots & u(1) \\ -y(n + 1) & \cdots & -y(2) & u(m + 1) & \cdots & u(2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -y(N - 1) & \cdots & -y(N - n) & u(N - 1) & \cdots & u(N - m) \end{bmatrix}$

- $y = [y(n + 1), y(n + 2), \ldots, y(N)]^T$
- $\Delta = [\varepsilon(n + 1), \varepsilon(n + 2), \ldots, \varepsilon(N)]^T$

where $\phi$ is the matrix-vector of the gearbox system characteristic, $\theta$ is matrix-vector of the gearbox, and $\theta = [A_1, \ldots, A_n, B_1, \ldots, B_m]^T$.

The traditional ARX identification minimises its loss function according to the least-squares principle. The loss function is shown through Equation (7):

$$
\delta = \sum_{t=n+1}^{N} \Delta^2(t) = \sum_{t=n+1}^{N} [y(t) - \Phi(t)\theta]^2
$$

The estimated value of parameter $\theta$ can then be obtained using Equation (8).

$$
\hat{\theta} = [\Phi^T \Phi]^{-1} \Phi^T y
$$

Measuring the input time-series $\{u_t\}$ and output time-series $\{y_t\}$ by using the system, the model of the gearbox system characteristics can be established, and the system model parameters of the gearbox can be estimated as $\hat{\theta} = [A_1, \ldots, A_n, B_1, \ldots, B_m]^T$, that is, the system characteristics and health status of the gearbox are contained in these model parameters. Therefore, the system analysis and fault diagnosis can be conducted based on the established model.

### D. IMPROVED ARX MODEL OF IRI BASED ON KF

The parameter estimation of the transmission characteristics of the traditional ARX model of a gearbox uses the least-squares method, which has the advantage of being easy to understand. However, it has a large identification error, particularly when the operating conditions are complicated, and the external interference will seriously affect the recognition speed and increase the error, which is difficult to be used for online identification. Therefore, an improved ARX model of IRI based on KF method is proposed in this study, which can reduce the memory usage, improve the accuracy of the algorithm, and realise the real-time status monitoring of the gearbox under varying conditions.

The IRI algorithms generally take the form shown in Equation (9) [37].

$$
\hat{\theta}(t) = \hat{\theta}(t - 1) + K(t)(y(t) - \hat{y}(t))
$$

where $\hat{\theta}(t)$ and $y(t)$ are the parameter estimate and system response, respectively, at time $t$; $\hat{y}(t)$ is the predicted system response obtained from the identification model at time $t$; and $K(t)$ is the recursive gain matrix of ARX identification model, where $K(t) = Q(t)\Phi(t)$ is generally chosen.

Because ARX can realise a modelling with a linear regression form, $\Phi(t)$ is the specific change of $\hat{y}(t)$ for parameter $\theta$.

$$
y(t) = \Phi^T(t)\hat{\theta}(t) + \epsilon(t)
$$

where $\hat{\theta}(t)$ is the response parameter of a gearbox, $\epsilon(t)$ is the noise signal for the excitation and response of the gearbox system. Supposing a change in $\theta(t)$, the actual excitation and response variable of the gearbox satisfy Equations (11) and (12).

$$
\hat{\theta}(t) = \hat{\theta}(t - 1) + \omega(t)
$$

$$
E(\omega(t)\omega^T(t)) = S_1
$$

The optimal incremental matrix of the system identification algorithm can be calculated, and the KF can be further applied. The IRI algorithm based on a KF can be completely expressed as follows.

$$
\hat{\theta}(t) = \hat{\theta}(t - 1) + K(t)(y(t) - \hat{y}(t))
$$

$$
\hat{y}(t) = \Phi^T(t)\hat{\theta}(t - 1)
$$

$$
K(t) = Q(t)\Phi(t)
$$

$$
Q(t) = \frac{P(t - 1)}{S_2 + \Phi^T(t)P(t - 1)\Phi(t)}
$$

$$
P(t) = P(t - 1) + S_1 - \frac{P(t - 1)\Phi^T(t)P(t - 1)\Phi(t)}{S_2 + \Phi^T(t)P(t - 1)\Phi(t)}
$$

where $S_2$ is the identification variance matrix of the system. To improve the application of the incremental recursive algorithm, a KF was used, and the forgetting factor $\lambda$ is added during the identification process. The recursive algorithm of system can then be modified as follows.

$$
Q(t) = \frac{P(t - 1)}{\lambda + \Phi^T(t)P(t - 1)\Phi(t)}
$$

$$
P(t) = \frac{1}{\lambda} \left[ P(t - 1) + \frac{P(t - 1)\Phi^T(t)P(t - 1)\Phi(t)}{\lambda + \Phi^T(t)P(t - 1)\Phi(t)} \right]
$$

ARX not only contains the regression of the ARX model itself, it also contains the influence of the input. Therefore, it has a strong anti-interference capability for noise signals. When a noise is small, the model has fine robustness and a
simple algorithm. When the noise is large, the influence of noise on the identification accuracy can be compensated by increasing the order of the model. The use of the IRI model based on a KF to replace the one-time least-squares operation can greatly reduce the memory occupancy and calculation time of the system, which is suitable for online identification.

III. SIMULATION OF IMPROVED MODEL AND QUANTITATIVE EVALUATION BASED ON GCCA

A. SIMULATION OF IMPROVED ARX MODEL

For a signal with noise interference, the formula can be expressed as Equation (20).

\[ y(f, \tau, t) = \begin{cases} e^{-\frac{t}{5\tau}} - e^{-\frac{t}{5\tau}} \frac{\sin(2\pi f t + \phi)}{\sin(\phi)} + q(t), & \tau \leq t \leq \tau + W_s \\ 0, & \text{else} \end{cases} \]  

(20)

where \( f \), \( \tau \) and \( q(t) \) are the signal oscillation frequency, delay factor, and random noise, respectively, \( W_s \) is the transient waveform period, and \( \xi \) is the damping coefficient \((\xi \in [0, 1) \subset R^+ ) \).

The formula of the signal without noise interference can be expressed through Equation (21), and its waveform is shown in Figure 3.

\[ y(f, \tau, t) = \begin{cases} e^{-\frac{t}{5\tau}} - e^{-\frac{t}{5\tau}} \frac{\sin(2\pi f t + \phi)}{\sin(\phi)}, & \tau \leq t \leq \tau + W_s \\ 0, & \text{else} \end{cases} \]  

(21)

After adding random noise \( q(t) \), the output after identification using the traditional ARX model is as shown in Figure 4.

The simulation results in both Figures 5 and 6 show the existence of boundary effects. Ignoring the boundary effect, the improved ARX model clearly has more anti-noise interference ability and can better reflect the real status of the system.

B. PRINCIPLE OF GENERALISED CANONICAL CORRELATION ANALYSIS

Through a GCCA, a more accurate template of the health of a gearbox can be established to better realise a fault identification and fault quantitative diagnosis. The following is a description of the GCCA method applied. It is assumed that a random variable has the following nonlinear relationship [13].

\[ y \sim \sum_{i=1}^{n} a_i f_i(x) \]  

(22)

where \( f_i(x) \) denotes the nonlinear correlation form of random variables \( y \) and \( x \). The random variable \( x \) is mapped to the high-dimensional vector \( m \) in the form of \( f_i(x) \).

\[ m = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{bmatrix} \]  

(23)

where \( m \) denotes a mapped high-dimensional random vector, and \( f_i(x) \) denotes the mapping function corresponding to \( m \).

Through the above mapping, the nonlinear relationship between \( y \) and \( x \) is transformed into a linear relationship between \( y \) and \( m \).

\[ y \sim \sum_{i=1}^{n} a_i m_i = a^T m \]  

(24)

After high-dimensional mapping, the linear correlation between \( y \) and \( m \) can be extracted using the CCA. By analysing the correlation between \( y \) and \( m \), the nonlinear relationship between \( y \) and \( x \) can be obtained.
C. OF QUANTITATIVE EVALUATION BASED ON GCCA

The output of the transfer function obtained by the ARX model described in Section 3 is a nonlinear curve. To evaluate a non-linear curve quantitatively, a method for applying a GCCA is proposed. Through this method, an accurate quantitative relationship between the transfer function and the health template of the gearbox under varying conditions can be obtained, and the gearbox under varying conditions can thus be quantitatively diagnosed. To verify the effectiveness of the GCCA further, the simulation data are used to verify its validity in extracting the nonlinear correlation. The specific steps are as follows:

1) GENERATE THE SIMULATION DATA

Firstly, two columns of random variables for samples $x_1$ and $x_2$ are generated, and a sufficient number of samples, 5,000, is used to meet the needs of the correlation analysis, and thus two columns of 5,000 rows of random sample vectors are generated.

Secondly, the samples $y_1$ and $y_2$ are generated according to the random samples $x_1$ and $x_2$, where $y_1 = x_1^2 + \text{randn}$, and $y_2 = x_2^3 + \text{randn}$. Thus, two $5000 \times 2$ datasets, X and Y, are obtained, in which $y_1$ is squared with $x_1$ and $y_2$ is linear with $x_2$. The correlation between the sample sets X and Y is extracted using the GCCA to verify whether it can extract the nonlinear relationship mentioned above.

2) ANALYSE THE RELATIONSHIPS BETWEEN DATASETS USING THE CCA

The traditional CCA was directly applied to the sample sets X and Y, and after the analysis, it was found that only one pair of canonical correlation variables (CCV) has a correlation coefficient of greater than 0.01. The correlation coefficient of the pair of variables reached 0.8922, and the coefficients of each variable are as shown in Table 1.

As can be seen from Table 1, if the CCA is used directly, only the linear correlation between $x_2$ and $y_2$ can be extracted, and the nonlinear correlation between $x_1^2$ and $y_1$ is not extracted.

3) ANALYSE RELATIONSHIPS BETWEEN DATASETS USING THE GCCA

For the generated sample sets $X_{5000 \times 2}$ and $Y_{5000 \times 2}$, the polynomials are selected to map them at high dimensions.

$$m = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{bmatrix} = \begin{bmatrix} x \\ x^2 \\ x^3 \end{bmatrix}$$ (25)

A nonlinear mapping of random variable $x$ is applied to a higher dimensional vector $m$. After mapping, $m$ is extracted according to the generalised eigenvalues of the Lagrange matrix, which contains six random variables, and its sample matrix form is $[x_1,x_1^2,x_2,x_3^2,3,2,3]_{5000 \times 6}$. The relationship between $m$ and $y$ is extracted using the CCA to analyse the nonlinear relationship between $x$ and $y$.

The GCCA is used to extract the correlation between $y$ and $m$, and the six pairs of generalised CCVs are extracted according to the eigenvalues and there are only 2 correlation coefficients larger than 0.01. The correlation coefficient of the first pair of CCVs reached 0.9996, which shows a strong correlation. The specific coefficients of the variables are provided in Table 2.

As can be seen from Table 2, the CCV reflects the correlation between $-0.001x_1 - 0.0278x_1^2 + 0.0008x_2 - 0.0001x_3^2$ and $0.0278y_1$. It can be determined that $x_1^2$ is the main provider of the correlation of dataset $M$, and $y_1$ is the primary provider of the correlation of dataset $Y$. Both have the same symbol, representing a positive correlation. Therefore, the CCV successfully extracted the correlation between $x_1^2$ and $y_1$.

The correlation coefficient of the second pair of CCVs reached 0.8922, and the coefficients of each variable are as shown in Table 3.

From Table 3, we can see that the generalized CCV obtained can accurately reflect the correlation between $0.0034x_1 - 0.0005x_1^2 - 0.0001x_1^3 - 0.4984x_2 + 0.0002x_3^2 - 0.0003x_3^3$ and $-0.0003y_1 - 0.4465y_2$. An analysis of its coefficient composition clearly shows that $x_2$ and $y_2$ are the main providers of the correlation of datasets $M$ and $Y$, respectively. The two symbols are identical, representing a positive correlation. Therefore, the CCV extracted the correlation between $x_2$ and $y_2$.

To summarise, the GCCA can extract the nonlinear correlation between vectors, and establish a more accurate template of the health of the gearbox through the GCCA to better

| $X$ dataset variable | $x_1$ | $x_2$ |
|----------------------|-------|-------|
| Variable coefficient | 0.0011 | -0.5023 |

| $Y$ dataset variable | $y_1$ | $y_2$ |
|----------------------|-------|-------|
| Variable coefficient | 0.0002 | -0.4464 |

| $M$ dataset variable | $x_1$ | $x_1^2$ | $x_2$ | $x_2^2$ |
|----------------------|-------|---------|-------|---------|
| Variable coefficient | -0.0001 | -0.0278 | 0.0008 | 0.0000 |
| Y dataset variable   | $y_1$ | $y_2$   |
| Variable coefficient | -0.0278 | 0.0000 |

| $M$ dataset variable | $x_1$ | $x_1^2$ | $x_2$ | $x_2^2$ |
|----------------------|-------|---------|-------|---------|
| Variable coefficient | 0.0034 | -0.0005 | -0.4984 | 0.0002 |
| Y dataset variable   | $y_1$ | $y_2$   |
| Variable coefficient | -0.0003 | -0.4465 |
realise a fault quantitative diagnosis, and establish a more accurate correlation model of a complex system.

IV. EXPERIMENTAL VERIFICATION OF QUANTITATIVE EVALUATION OF GEARBOX SYSTEM UNDER VARYING CONDITIONS

The gearbox is a vital transmission component in rotating machinery and is a component with a high failure rate in the equipment, so it has great significance to diagnose the fault. Nowadays, the development trend of most gearboxes tends to be more integrated, and when the fault reaches a certain level, it is replaced. Therefore, diagnosis of the existence and degree of a fault is more in line with the engineering requirements. Thus, this paper aims to study the quantitative diagnosis of gearboxes. From the perspective of the characteristics of a gearbox system, according to the improved ARX model of IRI based on a KF and a nonlinear GCCA, the input and output signals are comprehensively analysed. And the transmission characteristic curve (TCC) of the gearbox and the correlation coefficient between the faulty gear and the normal gear are obtained, thereby quantitatively evaluating the fault state of the gearbox.

A. EXPERIMENT DESIGN

In this study, normal, worn, and broken tooth gears were further tested under the starting and stopping conditions on an SQ fixed-axis gearbox test bench to verify the effectiveness of the proposed method in terms of its gearbox fault diagnosis.

The layout and mechanical structure of the experimental table are shown in Figure 7. To facilitate the experimental comparison, three gears, namely, a normal gear, gear #1 with a broken tooth, and gear #1’ with a worn teeth surface, were mounted to the input shaft of the gearbox, and the three gears could be manually switched without disassembly. Figure 8 shows photographs of gears #1’ and #1, in which a broken tooth on gear #1 and a generally worn failure on each tooth on gear #1’.

In this experiment, a PCB acceleration sensor with a sensitivity of 100 mV/g and an AEMC current plier with a sensitivity of 10 mV/A were used to collect the vibration and current signals of the gearbox, respectively, and the sensor installation position is as shown in Figure 7(b). The sampling frequency of the vibration acceleration signal is 10 kHz, and the sampling length is 22 s. The sampling frequency of the current signal is 1 kHz and the sampling length is also 22 s, and both are sampled simultaneously. To ensure the comparability of the test data, the faulty tooth and normal tooth of the gearbox experienced the same nonlinear speed-up process presupposed using a programmable computer, where the speed increased from 800 to 1,800 rpm, as shown in Figure 9. Because the characteristic frequency of the gearbox is time-varying under a varying speed, the frequency is given in an ordered form for analysis, as shown in Table 4.

B. DIAGNOSTIC EFFECT OF TRADITIONAL METHODS

Traditional gearbox fault diagnosis methods can be divided into excitation-based and response-based fault diagnosis methods. In this paper, an original method is proposed to use...
stator current and vibration acceleration signals as excitation and response signals respectively, and comprehensively analyse excitation and response signals to monitor the failure status of gearbox from the perspective of system characteristics, as shown in Figure 10.

Response based fault diagnosis method is widely used in traditional diagnosis methods, especially the fault diagnosis method based on vibration signal, which is the earliest method used in fault diagnosis. Moreover, time-domain analysis and order analysis are the most used analysis methods in fault diagnosis. The VMD is recognized as a powerful tool for analysing nonlinear and nonstationary signals, and hence is extensively employed for mechanical vibration analysis [22], [27]. Therefore, VMD combined with order analysis is used in this section as the traditional response-based fault diagnosis method to analyse the vibration signal of the gearbox with a broken tooth, and the analysis results are used to verify the effectiveness of the proposed method.

Figures 11(a) and 11(b) show the time-domain waveform under the start of a faulty gear and for a normal gear, respectively. However, it is difficult to identify the waveform anomaly caused by the fault from Figure 11(a). As the root cause of this phenomenon, although gear #1 has a partially broken tooth, the tooth surface is basically intact along the involute, and thus the disengagement of the tooth surface does not occur during the meshing process. That is, the fault only affects the local meshing stiffness of the gear, but does not cause a severe tooth surface impact. Therefore, for such a failure, a time domain analysis often has difficulty achieving the desired diagnostic results.

The vibration signal of the gearbox with a broken tooth is decomposed by VMD, and the result is shown in Figure 12. It can be noticed in Figure 12 that there are five intrinsic modal functions (IMF) were correctly recovered by the VMD approach. The IMFs contain the broken tooth fault information of the gearbox. Reconstructing the gearbox fault signal using the IMFs, the time-domain waveform of the reconstructed signal is shown in Figure 13.

The change in speed causes many frequency components are mixed together and cannot distinguish the characteristic frequency of the fault. The order analysis is able to correlate the frequency spectrum and time history with the rotational speed of rotating components. Moreover, after VMD and signal reconstruction, some unnecessary signals are removed, and the fault is more easily identified. So the order spectrum analysis of the reconstructed signal data is carried out in this section, the analysis results are shown in Figure 14.

It can be seen from Figure 14, the order spectral curve has prominent peaks at the 32 and 38 orders. According to order analysis theory and gearbox characteristics, order 32 correspond to the meshing frequency of #1 and #2, which is independent of the gearbox faults. However, the peak corresponding to the order 38 represents the meshing frequency of the faulty gear, which characterizes the broken tooth fault feature of the gear.

Therefore, the VMD combined with order spectrum analysis can effectively diagnose the fault type of the gearbox. However, Order analysis distorts the amplitude of signals, which will affect quantitative diagnostic results. Many papers have further improved the performance of order analysis,
such as tacho-less order analysis [38] et al., which has significantly improved the defect.

This proves that the experimental scheme designed in this paper is reasonable and the fault of gear can be detected effectively. It provides a powerful reference for the diagnostic effect of the method proposed in this paper.

C. DIAGNOSTIC EFFECT OF THE PROPOSED METHOD

In this paper, a new quantitative diagnostic method of gearbox under variable working conditions is proposed independently. The method comprehensively analyses the excitation and response signals from the perspective of system characteristics. To compare the diagnostic effect with the VMD and order analysis, and further verify the effectiveness of the quantitative diagnosis of the proposed method. The stator current and vibration acceleration signal of gearboxes with a broken tooth gear, a worn gear and a normal gear under the starting and stopping conditions were analyzed by the method.

If the order of ARX model selected is too small, the transfer characteristics of the system cannot be fully expressed. However, if the order of the ARX model selected is too large, a signal redundancy will occur and many unnecessary burrs will appear on the characteristic curve of the system, which may affect the recognition performance of the subsequent identification systems. According to the principle of the AIC information criterion combined with the characteristics of the experimental objects and experimental experience, the order of the ARX model was determined to be 18 in this study. From the perspective of the gearbox system characteristics according to the improved gearbox ARX model and GCCA, the transmission characteristics of the gearboxes with a normal, a worn gear, and a broken tooth gear are analysed using the stator current signal and the vibration acceleration signal. And the TCCs of three kinds of gearboxes are obtained, respectively, as shown in Figure 15.

As is clear from Figure 15, the TCCs of the gearboxes of a worn gear, a broken tooth gear and normal gear have respective remarkable characteristics and are easily distinguished. By comparing with the TCC of the normal gearbox, it can be judged whether a gearbox has faults and fault types. And, the diagnosis results are consistent with the results of VMD and order analysis in section 5.2. Therefore, the method can effectively identify the fault type of the gearbox.

Through the analysis of TCCs of gearboxes, it can be found that, the TCC of the gearbox with a worn gear is very similar to that of the normal gearbox, except that a significant peak appears near the frequency of 500 Hz. Therefore, for the worn gear used in this experiment, as long as its operating frequency is greater than 500Hz, it can still be used normally, but this practice is not recommended in practical engineering applications. And the TCC of the gearbox with a broken tooth maintains a high amplitude throughout the entire frequency band, which indicates that the broken tooth fault is a more serious fault and must be replaced once it occurs.

From the perspective of the characteristics of the gearbox system, according to the improved gearbox ARX model and GCCA, the canonical correlation coefficients of the gearboxes with a worn gear and the broken tooth gear relative to the normal gearbox are 0.8925 and 0.3454, respectively, as shown in Table 5. Clearly, the correlation coefficient between a broken tooth and normal gearbox is less than that between a worn gear and normal gearbox. It demonstrates numerically that the fault of broken tooth is more serious than that of wear gear, which is consistent with the previous analysis. This indicates that the smaller the correlation coefficient between a faulty gearbox and a normal gearbox, the more serious the failure. It is proved that the proposed method can quantitatively diagnosis the fault severity using the correlation coefficient.

V. CONCLUSION

The basic idea of system characteristic diagnosis is that the system characteristic is an inherent property of the gearbox system and has nothing to do with the running state. When the system characteristics change, it indicates that the gearbox system is malfunctioning. The system characteristic model
of gearbox is established based on the improved ARX model. Considering the quantitative diagnosis of the severity of gearbox faults, GCCA was used to evaluate the quantitative relationship between the faulty gearbox and the healthy gearbox. After normalization, the proposed method pays more attention to the differences between morphological features of the TCC, and does not care about the relationship between the amplitude or energy of the TCC. According to the different TCC forms and the correlation coefficients with the TCC form of the normal gear, the type and the severity of the gears fault can be diagnosed. Simulation and experimental results verify the effectiveness of the system characteristic diagnosis method. Comparing with VMD and order analysis, the proposed method can realize the quantitative diagnosis of gearbox faults under varying condition, and has more engineering value which can provide reference guidance for the replacement of the gearbox.

REFERENCES

[1] I. Antoniadou, G. Manson, W. Staszewski, T. Barszcz, and K. Worden, “A time–frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions,” Mech. Syst. Signal Pro- cess., vols. 64–65, pp. 188–216, Dec. 2015.

[2] M. Zhao, J. Jiao, and J. Lin, “A data-driven monitoring scheme for rotating machinery via self-comparison approach,” IEEE Trans. Ind. Informat., vol. 15, no. 4, pp. 2435–2445, Apr. 2019.

[3] J. Wang, F. Cheng, W. Qiao, and L. Qu, “Multiscale filtering reconstruction for wind turbine gearbox fault diagnosis under varying-speed and noisy conditions,” IEEE Trans. Ind. Electron., vol. 65, no. 5, pp. 4268–4278, May 2018.

[4] M. Zhao and J. Lin, “Health assessment of rotating machinery using a rotary encoder,” IEEE Trans. Ind. Electron., vol. 65, no. 3, pp. 2548–2556, Mar. 2018.

[5] X. Liang, M. J. Zuo, and Z. Feng, “Dynamic modeling of gearbox faults: A review,” Mech. Syst. Signal Proces., vol. 98, pp. 852–876, Jan. 2018.

[6] X. Liang, M. J. Zuo, and M. Pandey, “Analytically evaluating the influence of crack on the mesh stiffness of a planetary gear set,” Mechanism Mach. Theory, vol. 76, pp. 20–38, Jun. 2014.

[7] K. Peng, P. Borghesani, W. A. Smith, R. B. Randall, Z. Y. Chin, J. Ren, and Z. Peng, “Vibration-based updating of wear prediction for spur gears,” Wear, vol. 426–427, pp. 1410–1415, Apr. 2019.

[8] K. Jiang, Y. Zhou, Q. Chen, and L. Han, “In processing fault detection of machinery based on instantaneous phase signal,” IEEE Access, vol. 7, pp. 123535–123543, 2019.

[9] M. Marzebali, S. Kia, H. Henao, G. Capolino, and J. Faiz, “Planetary gearbox torsional vibration effects on winding–rotor induction gener- ator electrical signatures,” IEEE Trans. Ind. Appl., vol. 52, no. 6, pp. 4770–4780, Aug. 2016.

[10] Z. Li, Y. Jiang, C. Hu, and Z. Peng, “Difference equation based empirical mode decomposition with application to separation enhancement of multi-fault vibration signals,” J. Difference Equ. Appl., vol. 23, nos. 1–2, pp. 457–467, Feb. 2017.

[11] Z. Li, Y. Jiang, Q. Guo, C. Hu, and Z. Peng, “Multi-dimensional variational mode decomposition for bearing-crack detection in wind turbines with large driving-speed variations,” Renew. Energy, vol. 116, pp. 55–73, Feb. 2018.

[12] K. Jiang, L. Han, and Y. Zhou, “Quantitative evaluation of the impurity content of grease for low-speed heavy-duty bearing using an acoustic emission technique,” Meas. Control, vol. 52, nos. 7–8, pp. 1159–1166, Sep. 2019.

[13] J. Yoon and D. He, “Planetary gearbox fault diagnostic method using acoustic emission sensors,” IET Sci., Meas. Technol., vol. 9, no. 8, pp. 936–944, Nov. 2015.

[14] T. Touret, C. Changenet, F. Ville, M. Lalmi, and S. Becquerelle, “On the use of temperature for online condition monitoring of geared systems— A review,” Mech. Syst. Signal Proces., vol. 101, pp. 197–210, Feb. 2018.

[15] K. Jiang, L. Li, L. Han, and S. Guo, “In-process quality inspection of rolling element bearings based on the measurement of microelastic deformation of outer ring,” Shock Vib., vol. 2019, pp. 1–12, May 2019.

[16] K. Dragomiretskiy and D. Zosso, “Variational mode decomposi- tion,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, Feb. 2014.

[17] Y. Lei, J. Lin, M. J. Zuo, and Z. He, “Condition monitoring and fault diagnosis of planetary gearboxes: A review,” Measurement, vol. 48, pp. 292–305, Feb. 2014.

[18] A. Sapena-Bano, J. Burriel-Valencia, M. Pineda-Sanchez, R. Puche-Panadero, and M. Riera-Guasp, “The harmonic order tracking analysis method for the fault diagnosis in induction motors under time-varying conditions,” IEEE Trans. Energy Convers., vol. 32, no. 1, pp. 244–256, Mar. 2017.

[19] Y. Wang, P. W. Tse, B. Tang, Y. Qin, L. Deng, T. Huang, and G. Xu, “Order spectrogram visualization for rolling bearing fault detection under speed variation conditions,” Mech. Syst. Signal Proces., vol. 122, pp. 580–596, May 2019.

[20] Y. Li, K. Ding, G. He, and X. Jiao, “Non-stationary vibration feature extraction method based on sparse decomposition and order tracking for gearbox fault diagnosis,” Measurement, vol. 124, pp. 453–469, Aug. 2018.

[21] Y. Li, G. Li, Y. Wei, B. Liu, and X. Liang, “Health condition identification of planetary gearboxes based on variational mode decomposition and generalized composite multi-scale symbolic dynamic entropy,” ISA Trans., vol. 81, pp. 329–341, Oct. 2018.

[22] K. Feng, K. Wang, Q. Ni, M. J. Zuo, and D. Wei, “A phase angle based diagnostic scheme to planetary gear faults diagnostics under non-stationary operational conditions,” J. Sound Vib., vol. 408, pp. 190–209, Nov. 2017.

[23] C. Junsheng, Y. Dejie, and Y. Yu, “A fault diagnosis approach for roller bearing faults based on EMD method and AR model,” Mech. Syst. Signal Proces., vol. 20, no. 2, pp. 350–362, Feb. 2006.

[24] Y. Lin, U. Kruger, J. Zhang, Q. Wang, L. Lamont, and L. E. Chaar, “Seasonal analysis and prediction of wind energy using random forests and ARX model structures,” IEEE Trans. Control Syst. Technol., vol. 23, no. 5, pp. 1994–2002, Sep. 2015.
K. Jiang, G. Xu, L. Liang, T. Tao, and F. Gu, “The recovery of weak impulsive signals based on stochastic resonance and moving least squares fitting,” *Sensors*, vol. 14, no. 8, pp. 13692–13707, Jul. 2014.

W.-X. Zhao, H.-F. Chen, and W. Xing Zheng, “Recursive identification for nonlinear ARX systems based on stochastic approximation algorithm,” *IEEE Trans. Autom. Control*, vol. 55, no. 6, pp. 1287–1299, Jun. 2010.

M. Zhao, J. Lin, X. Wang, Y. Lei, and J. Cao, “A tacho-less order tracking technique for large speed variations,” *Mech. Syst. Signal Process.*, vol. 40, no. 1, pp. 76–90, Oct. 2013.

LIUBANG HAN received the B.S. degree from the Liaoning University of Technology, Jinzhou, China, in 2016, and the M.S. degree in mechanical engineering from the Anhui University of Science and Technology, Huainan, China, in 2018, where he is currently pursuing the Ph.D. degree in mine electromechanical engineering. His research interests include fault diagnosis, signal processing, and electromechanical control systems.

QIDONG WANG received the M.S. degree from the Jilin University of Technology, Changchun, China, in 1987, and the Ph.D. degree from the Hefei University of Technology, Hefei, China, in 2002. He studied at Aalborg University, Aalborg, Denmark, from August 1996 to October 1997. He is currently a Professor with the Anhui University of Science and Technology and Hefei University. His research interests include automotive dynamics control and modern design methods.

XUANYAO WANG received the B.S. and M.S. degrees in mechanical engineering from the Anhui University of Science and Technology, Anhui, China, in 2003 and 2006, respectively, and the Ph.D. degree in automotive engineering from the Hefei University of Technology, Anhui, in 2017. He is currently an Associate Professor with the School of Mechanical Engineering, Anhui University of Science and Technology. He has authored about 20 publications in journals and conference proceedings. His research interests mainly focus on vehicle system dynamics and control, automotive active safety technology, and advanced driving assistance technology.

YUANYUAN ZHOU received the B.S. degree from the Anhui University of Science and Technology, Huainan, China, in 2018, where he is currently pursuing the M.S. degree in mechanical engineering. His research interests include signal processing and fault diagnosis.

KUOSHENG JIANG received the M.S. degree from the Shaanxi University of Technology, Hanzhong, China, in 2010, and the Ph.D. degree from the State Key Laboratory for Manufacturing System Engineering, Xi’an, China, in 2016. He is currently working with the Anhui University of Science and Technology, Huainan, China. His research interests include design of sensor measurement and control systems, signal processing, and fault diagnosis.