Study on Centrifugal Pump Rotor Fault Feature Extraction Based on Fractal Theory

Wencai Liu[1,2](✉) and Qiyong Peng[1]

[1] CNPC Research Institute of Safety & Environment Technology, Beijing, China
{wencailiu, pengqiyong}@cnpc.com.cn
[2] HSE Key Laboratory of CNPC, Beijing, China

Abstract. This paper introduces the fundamental theory of fractal and the definition of correlation dimension firstly, then provides the process solving the correlation dimension by utilizing G-P algorithm and the determination method of relevant parameters. Conducting the feature extraction of rotor fault signal by utilizing the combined method of second generation wavelet noise reduction and fractal, the fault feature vector based on correlation dimension for four states of rotor system are analyzed, which are normal, imbalance, collision-friction and misalignment states.

Keywords: Centrifugal pump · Rotor fault · Feature extraction

1 Introduction

Under normal circumstances, different rotor systems faults correspond to different dynamical mechanisms, and the specific fault information is difficult to be obtained by utilizing traditional time–frequency domain (such as spectral analysis) [1]. Many studies are shown that the fractal dimension can highlight the irregularity and instability of signals to reflect the operating states of mechanical equipment and its parts. Due to the correlation dimension which is sensitive to the non-uniform reflection of system attractor, it can quantitatively provide the amount of independent variables required by dynamical system description and reflect the fractal features very well, and

This paper was prepared for presentation at the 2018 International Field Exploration and Development Conference in Xi'an, China, September 18-20, 2018.

J. Lin (ed.), *Proceedings of the International Field Exploration and Development Conference 2018*, Springer Series in Geomechanics and Geoengineering, https://doi.org/10.1007/978-981-13-7127-1_67

its correlation dimension calculation is more reliable comparing to other methods, and therefore it is feasible that the fault can be diagnosed by measuring the correlation dimension in theory.

2 Calculations of Fractal Dimension and Fractal Parameter

The fractal theory is a nonlinear systems theory, and its research objects are the disordered systems with self-similarity which extensively existed in nature and social activities, and this theory is first proposed by French mathematician Mandelbrot in the 1970s.

2.1 Fractal Dimension

The fractal dimension is a very important concept in the fractal theory, and for a complex system, the complexity and irregularity of set can be quantitatively described by the fractal dimension [2]. There are a variety of fractal dimension types, such as topological dimension, self-similarity dimension, box dimension, capacity dimension, information dimension, and correlation dimension; they can describe the fractal features of fractal set from different aspects [3, 4]. Among them, due to the correlation dimension is sensitive to the non-uniform reflection of system attractor, it can quantitatively provide the amount of independent variables required by dynamical system description and reflect the fractal feature exactly. The calculation method of correlation dimension is relatively simple which can be directly calculated and obtained by utilizing the phase-space reconstruction method through the observed one dimensional time sequence. Thus, the calculation method based on correlation dimension is adopted to conduct the fractal feature analysis of rotor vibration signal [5].

Assume the $\{X_k,\ k = 1, 2, \ldots, N\}$ is one-time sequence data of system and conduct the phase-space reconstruction on this time sequence, the reconstruction result is $X_n(m, \tau) = (x_n, x_{n+1}, \ldots, x_{n+(m-1)\tau})$ in which $n = 1, 2, \ldots, N - m + 1$; $\tau = k\Delta t$ is time delay; Δt is the sampling interval of data; k is arbitrary integer; m is the spatial dimension of reconstruction phase.

Then, the correlation dimension of reconstruction phase-space attractor is:

$$D_{\mathrm{h}} = \lim_{\varepsilon \to 0} \frac{\ln C(\varepsilon)}{\ln(\varepsilon)} \tag{1}$$

In which $C(\varepsilon)$ is correlation integral function:

$$C(\varepsilon) = \frac{1}{N(N-1)} \sum_{i=1}^{N-m+1} \sum_{j=1}^{N-m+1} H\left(\varepsilon - \|y_i - y_j\|\right) i \neq j \tag{2}$$

In which ε is the radius of m dimension hypersphere; H is the Heaviside function and that is

$$H\left(r - \left\|Y_j - Y_k\right\|\right) = \begin{cases} 1 & \left(r - \left\|Y_j - Y_k\right\|\right) \geq 0 \\ 0 & \left(r - \left\|Y_j - Y_k\right\|\right) < 0 \end{cases} \quad (3)$$

2.2 Calculations of Correlation Dimension

Due to the correlation dimension can reflect the fractal feature very well, its calculation method always is a key point in fractal theory study. P. Grassberger et al. have proposed a calculation method which has been called as the embedding space method in 1983 based on thoughts of embedding theorems and reconstruction phase space [6], that is, the G-P algorithm, and this algorithm can directly calculate its correlation dimension according to time sequence data, its calculation is simple and reliable, it has high practicability, and it is the most important method to calculate the correlation dimension.

The calculation steps of solving correlation dimension by G-P algorithm are as follows:

(1) Conduct the normalization processing on the time sequence $\{X_k, k = 1, 2, \ldots, N\}$ and select the reconstruction phase-space embedding dimension m based on the principle of $m \geq 2D_{\mathrm{h}} + 1$. Assume the $y_1 = (x_1, x_2, \ldots, x_m)$ is the first vector in the m dimension phase space and $y_2 = (x_{1+\tau}, x_{2+\tau}, \ldots, x_{m+\tau})$ is the second vector in the m dimension phase space and keep construction in a similar fashion; then, the kth vector is $y_k = \left(x_{1+(k-1)\tau}, x_{2+(k-1)\tau}, \ldots, x_{m+(k-1)\tau}\right)$.

(2) Calculate the distance $r_{ij} = \left|y_i - y_j\right|$ between two vectors of y_i, y_j.

(3) As for arbitrary given point $0 < r \leq 1$, compare the r_{ij} value of all the point pair with r. Assume the $N_1(r) = \left|\text{the number of points } r_{ij} \leq r\right|$, $N_2(r) = \left|\text{the number of points } r_{ij} \leq r\right|$ and the total point pair is $N(\varepsilon) = N_1(\varepsilon) + N_2(\varepsilon)$, calculate the ratio C_r of point pair whose distance is less than r in all the point pair according to Eq. (3). Appropriately adjust the value range of r to make $C_r \propto r^{D_{\mathrm{h}}}$.

(4) Calculate D_{h} according to Eq. (1) and that is the correlation dimension of corresponding time sequence.

The block diagram of calculation steps is shown in Fig. 1.

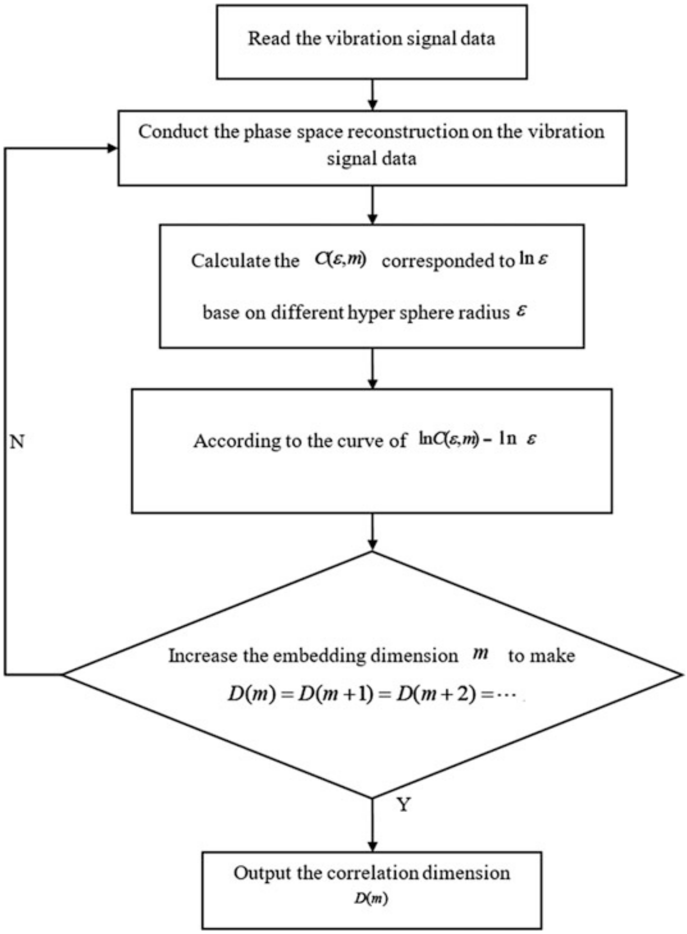

Fig. 1. Correlation dimension calculation block diagram

2.3 Selection of Fractal Parameter

2.3.1 Selection of Time Delay

The selection of time delay τ is very important, and the quality of phase-space reconstruction method will be affected if τ is too large or too small. The common methods used to calculate the time delays mainly are auto-correlation function method, average displacement method (debasing), complex auto-correlation method, and mutual information method. This paper adopts the auto-correlation function method, the calculation of this method is simple, and it has good application result.

The method utilizing the auto-correlation function to calculate the time delay τ method is very mature; it mainly is to calculate the linear auto-correlation coefficient of one sequence based on the principle feature of auto-correlation and to determine the time delay based on the auto-correlation coefficient. Assume one-time sequence is $\{x_i, i = 1, 2, 3, \ldots, N\}$ and the time delay is τ; then, its auto-correlation function can be expressed as:

$$C = \sum_{i=1}^{N-\tau} x_i' x_{i+\tau}' \Bigg/ \sum_{i=1}^{N-\tau} \left(x_i'\right)^2 \tag{4}$$

In which N is the length of sequence, $\bar{x}$ is the average value of sequence and $x_i^{'} = x_i - \bar{x}$. According to the auto-correlation principle, the smaller the C value of auto-correlation function, the smaller the correlation between each element in the sequence and the auto-correlation is smallest when $C = 0$, so take the corresponding time delay τ when C is 0 for the first time as time delay τ of reconstruction phase space, so the useful information of original system included in the time sequence can be ensured not to lose.

2.3.2 Determination of Embedding Dimension

During the phase-space reconstruction, in order to ensure the feature of original state space attractor of this phase space, the embedding dimension m needs to be large enough. In 1980, Takens used mathematical methods to prove the theorem about the size of the embedded dimension:

$$m \geq 2d + 1 \tag{5}$$

In which m is the spatial dimension of embedding phase; d is the dimension of original state space attractor locating space.

According to Eq. (5), the embedding theorem just provides the essential condition to select the embedding dimension but the specific determined method is not provided.

Two methods commonly used to determine the minimum dimension m are the false nearest neighbor method and Cao method. The false nearest neighbor method needs the artificial assumption to determine the threshold value of false neighbor points, and different assumptions may lead to different results and it is sensitive to noise and data length, so it is difficult to be practically applied. In view of this, the improved algorithm which has been called as Cao method is proposed to solve above-mentioned problems. The definition of Cao method is shown as follows:

Take $\{s_i, i = 1, 2, \ldots, N\}$ to express the time sequence to be processed and one group vector $x(i)$ can be obtained by time delay reconstruction; the relational expression between those two is shown as follows.

$$\begin{aligned} x(i) &= [s(i), s(i+\tau), \ldots, s(i+(d-1)\tau)] \\ i &= 1, 2, \ldots, N-(d-1)\tau \end{aligned} \tag{6}$$

In which d is the embedding dimension, and τ is the embedding delay.

In which the nearest neighbor point of ith reconstruction vector $x(i)$ can be expressed by $x\left(i_{NN}^d\right)$ and its definition is shown as follows:

$$a(i,d) = \frac{\left\| x_{d+1}(i) - x_{d+1}\left(i_{NN}^{d}\right) \right\|}{\left\| x(i) - x(i_{NN}^{d}) \right\|} \quad i = 1,2,\ldots,N - d\tau \tag{7}$$

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i,d) \tag{8}$$

In which $E(d)$ is the function about embedding dimension d and time delay τ, the Cao method assumption has obtained the optimum time delay τ by other methods and then only the variation of $E(d)$ along with embedding dimension needs to be considered and its definition is shown as follows:

$$E1(d) = \frac{E(d+1)}{E(d)} \tag{9}$$

When d is bigger than a value d_0, $E1(d)$ does not significantly change anymore and closes to 1, and the dimension of this moment is the minimum embedding dimension. But the determination of optimum embedding dimension by utilizing the Cao method takes the determination of optimum time delay as the premise, and the results are a little different; other methods usually are combined to determine the minimum embedding dimension.

Generally speaking, after the system has been determined to be convergent, the correlation dimension tends to be stable along with the increase in embedding dimension m and it would not have big change anymore, it is shown on the correlation integral curve graph that the curves tend to be parallel along with the increase in m. The system feature saturation method can be utilized to solve the embedding dimension, and the specific method is taken correlation dimension as the feature vector. Conduct the phase-space reconstruction method in low embedding spatial dimension firstly, and calculate the correlation dimension of phase space at this moment; then, increase the value of embedding spatial dimension m, conduct the phase-space reconstruction method and calculate its correlation dimension, repeat in this way until the correlation dimension stops changing (or the change is very little), and the embedding dimension at this moment is the desired value.

2.3.3 Selection of Data Length *N*

Theoretically speaking, the longer the data length N, the calculated correlation dimension will be more in line with fractal feature. But in practice, the longer the data length, the longer the required time of computer calculation

2.3.4 Determination of No-Scale Interval

The deterministic criteria are not given in the G-P algorithm to determine the no-scale interval domain of correlation integral curve, the correlation integral curve is generally observed at present, and it is a section close to straight line taken in the double logarithmic curve. In fact, it is impossible to accurately determine the no-scale interval only from correlation integral curve due to the observation which has great difference with reality. In order to solve this problem, the method calculating the local scaling exponent curve (or called as local slope) of correlation integral curve is adopted, the smaller the slope, the double logarithmic curve is closer to straight line; the local scaling exponent curve can be obtained by conducting the sliding five-point linear regression on correlation integral curve. After the local scaling exponent is obtained, the no-scale interval of correlation integral curve can be accurately determined by referring to local scaling exponent curve.

3 Rotor System Fault Feature Extraction Based on Fractal Theory

The noise can greatly affect the calculation of correlation dimension, it will cause the scale interval of correlation integral curve to decrease and the estimation value of correlation dimension to increase, it also disappears the inherent system determinacy feature, and it even causes the correlation dimension to be not convergent in a severe case. Thus, in order to make the calculation result of correlation dimension to be more accurate, the signal must be processed with noise reduction pretreatment.

There is a centrifugal pump in a one oil field gas station, the rated power is 230 KW, and the rotate speed is 2970 r/m. The misalignment fault has been detected in July 2016, the equipment has been successively monitored four times before the fault has been detected, and the collection position of signal is right above the bearing pedestal of motor output end. The specific times of four measures successively are May 2015,

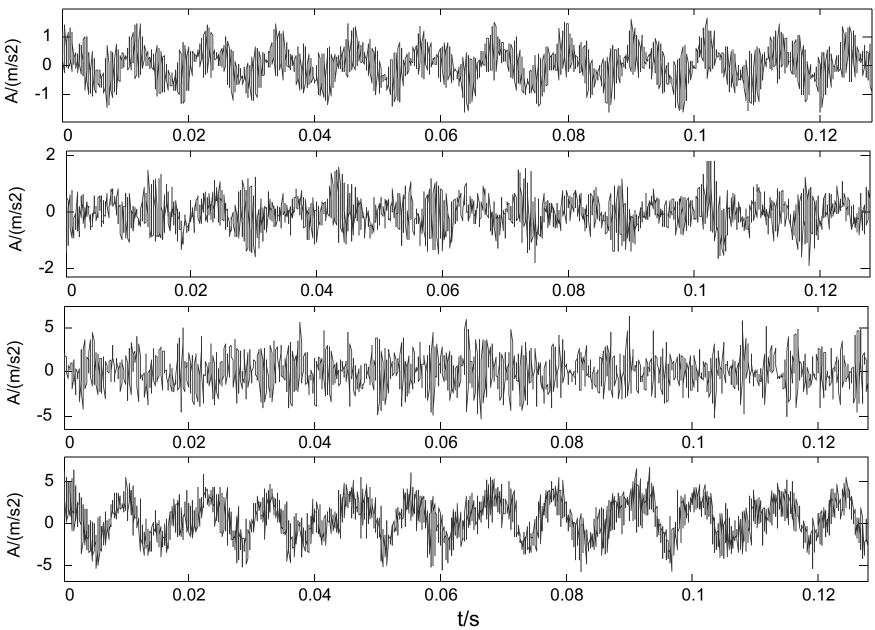

Fig. 2. Vibration signal time domain waveform diagram of each stage

November 2015, May 2016, and July 2017; the measured signals of the first two times are the signal under normal condition. The domain waveform diagrams of vibration signals that have been measured for each time are successively shown in Fig. 2.

Take the data measured in May 2015 for example and utilize the calculation method of correlation dimension to conduct the fractal characterization on collected vibration signal.

First, utilize the auto-correlation principle to calculate and obtain the time delay of this signal sequence $\tau = 44$ s. After the time delay, τ is determined and utilize the Cao algorithm to determine the estimated value of optimum embedding dimension, and results are shown in Fig. 3.

According to Cao algorithm principle, the corresponding embedding dimension

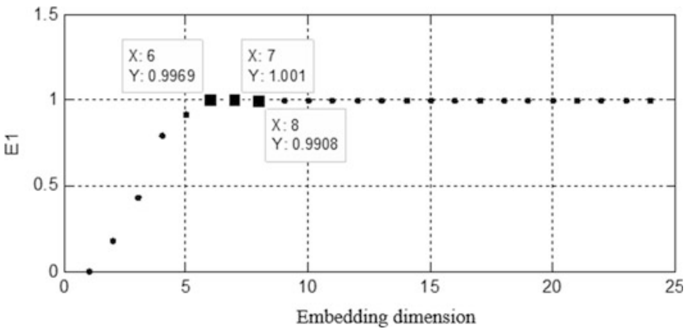

Fig. 3. Relational graphs between E1 and embedding dimension

when the value $E1(d)$ is close to 1 and there is no obvious change occurred is the minimum embedding dimension, so the minimum embedding dimension of this signal sequence can be preliminary determined as $m = 6$.

Then conduct phase-space reconstruction method on the signal sequence and the embedding dimension increases from 4 to 10, and draw the double logarithmic curves graph about the change of $\ln C(\varepsilon)$ along with $\ln \varepsilon$ under different embedding dimension m. Along with the increase in m, the double logarithmic curves tend to be parallel and the correlation dimension D does not change anymore or only has little change; then, the correlation dimension at this moment can be considered as the fractal dimension of signal. The optimum fitting straight-line section domain in the double logarithmic curves graph is the no-scale interval of fractal, and the slope of this straight line is the desired correlation dimension. Conduct the linear fitting on each double logarithmic curves on the correlation integral curve and calculate the slope of its optimum fitting straight-line section; the results are shown in the following table:

Table 1. Correlation dimensions of signal

Embedding dimension	4	5	6	7	8	9	10
Correlation dimension	0.6946	0.7293	0.7501	0.7631	0.7739	0.7742	0.7741

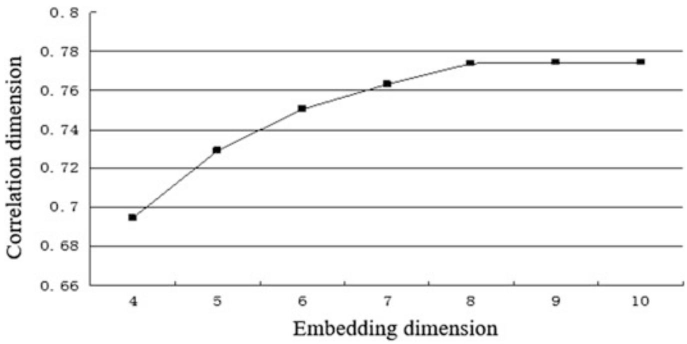

Fig. 4. Correlation dimension broken line graphs

Draw the correlation dimension broken line graph according to Table 1, and it is shown in Fig. 4.

According to Fig. 4, the correlation dimension tends to be stable along with the increase in embedding dimension. When the embedding dimension is 8, the correlation dimension is approximately stable which means the fractal dimension $D = 0.7739$ when the minimum embedding dimension of signal is $m = 8$.

Calculate the fractal parameters of signal of other three time measurement according to above-mentioned method, and the results are shown in Table 2.

Table 2. Fractal parameters of each phase of the signal

Type time	Vibration severity (mm/s)	Time delay τ	Minimum embedding dimension m	Correlation dimension D
May 2015	1.788	44	6	0.7739
November 2015	1.082	18	9	1.005
May 2016	1.701	26	20	2.6624
July 2017	2.523	29	14	2.8949

According to Table 2, the correlation dimension increases constantly as time goes by and there is huge discrimination degree existed between the data fractal dimension of first two time measurements and the data fractal dimension of last two time measurements which can identify the equipment state and fault type more easily, but its vibration severity does not have such discrimination degree. Thus, it can be seen that the operating state of equipment cannot be measured only by vibration severity, and the fractal dimension is another important parameter to measure the equipment operating state.

In order to further research the relationships of fractal dimension in various rotor states, build up the rotor vibration test stand to simulate three fault modes of imbalance, misalignment, and collision–friction.

The hardware system of test stand consists of SYSTEM 1 server, rotating equipment vibration monitoring, and fault virtual experiment platform and so on; the distributing structure is shown in Fig. 5.

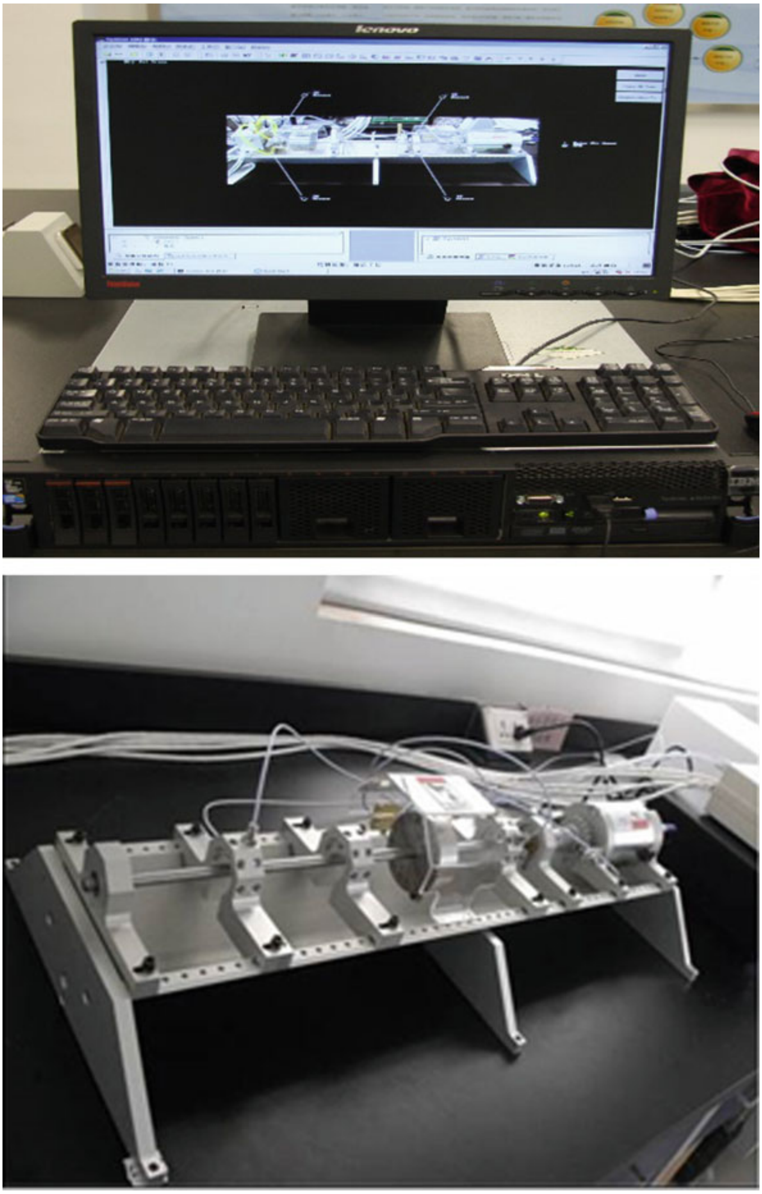

Fig. 5. Rotor fault simulation test stand

Respectively measure the rotor vibration signals under the four states of normal, imbalance, misalignment, and collision–friction. Measure 20 groups data for each state. When the balance weight is simulated, the added balance weight is 0.8 g and the rotating speed is 15000 r/min.

Then respectively conduct the fractal characterization calculation on this data according to above-mentioned correlation dimension calculation method and obtain their correlation dimensions, the results are shown in Table 3.

Table 3. Correlation dimension of each signal

Type no.	Normal	Imbalance	Misalignment	Collision-friction
1	0.6857	1.0764	1.8654	1.4316
2	0.6886	1.0634	1.7985	1.3884
3	0.6873	1.1043	1.8564	1.4168
4	0.7214	1.1464	1.8135	1.3985
5	0.6991	1.0343	1.8346	1.4019
6	0.6832	1.1036	1.7991	1.4324
7	0.6898	0.9987	1.8003	1.4023
8	0.6856	1.0362	1.8444	1.4121
9	0.7356	1.1022	1.8164	1.4001
10	0.7314	1.0234	1.8464	1.42
11	0.6846	1.0892	1.8137	1.4068
12	0.7523	1.0346	1.8434	1.4110
13	0.7104	1.1001	1.8641	1.3987
14	0.7321	0.9990	1.8214	1.3990
15	0.6867	1.1010	1.8098	1.4187
16	0.7162	1.0386	1.8198	1.4312
17	0.6844	1.0943	1.8139	1.4215
18	0.7014	1.1382	1.8298	1.4201
19	0.6833	1.1024	1.8347	1.403
20	0.7313	1.0039	1.8263	1.4197

It can be clearly seen from Table 3 that the value range of signal correlation dimension when the rotor system in the normal state is 0.6832–0.7523, the value range of signal correlation dimension of imbalance fault is 0.9987–1.1464, the value range of signal correlation dimension of misalignment fault is 1.7985–1.8654, and the value range of signal correlation dimension of collision–friction fault is 1.3884–1.4324. Their broken line graphs based on Table 3 are shown in Fig. 6.

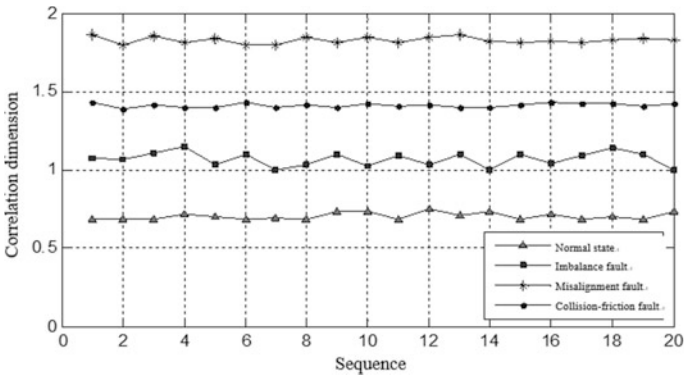

Fig. 6. The broken line graphs of rotor system correlation dimension

According to Fig. 6, good discrimination degree is existed between the correlation dimensions of four states. The operating state of rotor can be intuitively obtained by observing the correlation dimension, and it means utilizing the correlation dimension to characteristic the equipment state is feasible. The rotor fault feature extraction based on fractal theory is an effective method.

4 Conclusion

This paper mainly utilizes the second-generation wavelet noise reduction to calculate the correlation dimension of fault state vibration signal and conduct the fault diagnosis. This paper analyzes the correlation dimensions of rotor system in the normal, imbalance, collision–friction, and misalignment states. A lot of analyses and calculations have shown that the correlation dimensions of different states are significantly different, the fault diagnosis of rotor system can be effectively conducted by utilizing the difference of correlation dimension, and this method has been proved as an effective and intuitive diagnosis method.

Acknowledgements. This work is supported by the national key research and development program of China "Typical storage of dangerous chemicals safety warning and protection integration key technology research and application demonstration (2016YFC0801200)."

References

1. McCormick AC, Nandi AK. Cyclostationarity in rotation machine vibrations. Mech Syst Signal Process. 1998;12(2):225–42.
2. David R. Retrospective on diagnostic reasoning based structure and behaviour. Artif Intell. 1993;59:149–57.
3. Sweden's W. The lifting scheme: a custom design construction of orthogonal wavelets. Appl Comput Harmon Anal. 1996;3(2):186–200.

4. Sweden W. The lifting scheme: a construction of second generation wavelets. SIAMJ Math Anal. 1997;29(2):511–46.
5. Daubechies I, Sweden W. Factoring wavelet transform into lifting steps. J Fourier Anal App. 1998;4(3):247–69.
6. Kantz H, Sehreiber T. Nonlinear time series analysis. Zed Edition, Cambridge University Press, pp. 33–136; 1997.