The Characteristic Frequency Extraction of Helicopter Fault Signal Based on AR Model Estimation

Penghui Niu\(^{1,a}\)^*, Zhen Wang\(^{2,b}\), Ke Jia\(^{3,c}\), Suyong Liu\(^{4}\)

\(^{1}\)Department of Aviation Services Army Aviation Institute of PLA Beijing, China
\(^{2}\)Department of Aeronautical Mechanical Engineering Army Aviation Institute of PLA Beijing, China
\(^{3}\)Department of Aeronautical Mechanical Engineering Army Aviation Institute of PLA Beijing, China
\(^{4}\)Support Department Army Aviation Institute of PLA Beijing, China

\(*\)niu_penghui@lhxy.edu.cn
\(\text{a-mail: }\) niu_penghui@163.com, \(\text{b-mail: }\) 249940319@qq.com, \(\text{c-mail: }\) 5857487@qq.com

**Abstract**—It is very important to extract the characteristic frequency of helicopter mechanical fault signal, which is directly related to the accuracy of fault diagnosis and the reliability of early fault prediction. The traditional characteristic frequency extraction of fault signal mostly uses the spectrum analysis method based on Fourier analysis. The algorithm of this method is simple, but has high requirements on SNR (signal-to-noise ratio) and sampling time. In this paper, researchers use the AR model spectral estimation method based on the modern spectral estimation theory to study the characteristic frequency extraction of helicopter mechanical fault signal. The simulation results show that the effect is better.

1. **INTRODUCTION**

In helicopter mechanical fault diagnosis, one of the most important, critical and difficult problems is the characteristic frequency extraction of fault signal, which is directly related to the accuracy of fault diagnosis and the reliability of early fault prediction [1,2]. The traditional characteristic frequency extraction of fault signal mostly uses the spectral analysis method based on Fourier analysis. The algorithm of this method is simple, but has high requirements on SNR (signal-to-noise ratio) and sampling time. In order to improve the frequency resolution, the sampling time must be increased. These conditions are different from the actual working conditions, resulting in the limitations of this method. However, the parameter model method of modern spectral estimation can greatly improve the frequency resolution of power spectrum estimation without increasing the sampling time. In this paper, AR model spectral estimation method is used to study this problem, and simulation is carried out to verify that the diagnosis effect is better [3-5].

2. **FREQUENCY ESTIMATION BASED ON AR SPECTRUM**

Parameter model estimation is an important part of modern spectral estimation. It considers the random sequence to be generated by a source model. The parameters of the source model calculated from the
random sequence can predict or deduce the data outside the window. The more correct the selection of the model is, the more accurate the calculation of the parameters is, and the more accurate the prediction of the data outside the window is, then the more accurate the spectral estimation is and the higher resolution is [6-8].

2.1. Parameter model and power spectrum

In Figure 1, \( H(z) \) is a causal linear invariant discrete-time system, which is stable and its unit sampling response \( h(n) \) is deterministic. The output sequence \( x(n) \) can be a stationary random sequence or a deterministic time sequence.

![Figure 1. Parameter model.](image)

The relationship between input \( u(n) \) and output \( x(n) \) can be expressed by difference equation as follows:

\[
x(n) + \sum_{k=1}^{p} a_k x(n - k) = \sum_{r=0}^{q} b_r u(n - r).
\]

its transfer function is

\[
H(z) = \frac{\sum_{r=0}^{q} b_r z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}} = \frac{\sum_{r=0}^{q} b_r z^{-k}}{\sum_{k=0}^{p} a_k z^{-k}}.
\]

In the formula, \( a_0 = 1 \), \( b_0 = 1 \). If \( b_r = 0 (r = 1, 2, \ldots, q) \), then the above model is called Autoregressive model of order \( p \) or AR (P) model for short. This model is also called all-pole model. If \( a_r = 0 (r = 1, 2, \ldots, p) \), then the above model is called Moving Average model of order \( q \) or MA (q) model for short, this model is also called all-zero model. If \( b_r = 0 (r = 1, 2, \ldots, q) \) and \( a_r = 0 (r = 1, 2, \ldots, p) \), then the above model is called ARMA (p, q) model.

The power spectrum of the output signal \( P_x(e^{j\omega}) \) is calculated by the following formula

\[
R_x(e^{j\omega}) = P_u(e^{j\omega}) |H(e^{j\omega})|^2.
\]

In the formula, \( P_u(e^{j\omega}) \) is the power spectrum of the input signal, \( H(e^{j\omega}) \) is the transfer function of the system. The modern spectral estimation method based on AR model, MA model and ARMA model all consider that the stationary random signal \( x(n) \) is generated by discrete linear system under white noise excitation, so the power spectrum of AR model, MA model and ARMA model can be calculated.

If the input signal is white noise whose variance is \( \sigma_0^2 \), then the power spectrum of AR model, MA model and ARMA model are as follows.

AR model:

\[
\frac{P_x(e^{j\omega})}{|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}|^2} = \frac{\sigma_0^2}{\sigma_0^2}.
\]

(4a)

MA model:

\[
P_x(e^{j\omega}) = \sigma_0^2 |\sum_{r=0}^{q} b_r e^{-j\omega r}|^2 .
\]

(4b)

ARMA model:

\[
P_x(e^{j\omega}) = \frac{\sigma_0^2 |\sum_{r=0}^{q} b_r e^{-j\omega r}|^2}{|1 + \sum_{k=1}^{p} a_k e^{-j\omega k}|^2} = \frac{\sigma_0^2}{\sigma_0^2}.
\]

(4c)

AR, MA and ARMA are the most important parameter models in power spectrum estimation. Among them, ARMA model needs the least number of parameters, but the parameter estimation needs to use the nonlinear equations, whose operation is far more complex than using AR model; MA model is suitable for expressing the wave trough signal in the power spectrum, but generally needs a higher order to represent the narrow spectrum, and the parameter estimation needs to use the nonlinear operation; AR model is especially suitable for expressing signals with peak in the power spectrum, and the parameters of AR model use linear operation, the arithmetic labor is much smaller than using MA model and ARMA model. Considering that arbitrary ARMA or MA signal model can be approximated by AR model with
infinite order or order which is large enough, so the modern power spectral estimation method based on AR model has been widely studied and applied at present.

2.2. Spectral estimation algorithm based on AR model

There are many algorithms to solve the coefficient of autoregressive model: Yule-Walker algorithm, Burg algorithm, covariance algorithm and modified covariance algorithm and so on.

2.2.1. Yule-Walker algorithm. Yule-Walker algorithm uses the symmetry of autocorrelation matrix and Toeplitz property to solve AR model parameters with the help of Levison Durbin recursive algorithm. Yule-Walker algorithm calculates the parameters of the first-order AR model from the autocorrelation matrix, and then takes them as the initial conditions to calculate the parameters of the second-order AR model; it recurs in a similar fashion until the parameters of the p-order AR model are calculated.

Since the Yule-Walker algorithm needs to estimate the autocorrelation function, the estimation error of autocorrelation function may be very large when observation data is short. The large estimation error will result in the performance degradation of spectral estimation, even the phenomenon of peak shift and spectral line splitting. These problems can be solved by Burg algorithm.

2.2.2. Burg algorithm. The basic idea of Burg algorithm is that according to the principle that make the power average value of forward and backward prediction error minimized, the reflection coefficient \( K_m \) can be calculated and estimate the model parameters with Levenson Durbin recurrence formula. The Burg algorithm includes the following steps:

1. Calculate the initial value
   \[
   P_0 = R(0) = \frac{1}{N} \sum_{k=1}^{N} |x(k)|^2 .
   \]
   \[
   f_0(n) = b_0(n) = x(n) .
   \] (5a)
   In the formula, \( f_0(n) \) is the initial value of forward prediction error \( f_m(n) \), \( b_0(n) \) is the initial value of backward prediction error \( b_m(n) \), and \( m \) assigned from 1, i.e. \( m = 1 \).

2. Calculate the reflection coefficient \( K_m \)
   \[
   K_m = \frac{-\sum_{m=m+1}^{N} f_{m-1}(n) b_{m-1}^*(n-1)}{\frac{1}{2} \sum_{m=m+1}^{N} |f_{m-1}(n)|^2 + |b_{m-1}^*(n-1)|^2} .
   \] (5c)

3. Calculate the filter coefficient
   \[
   a_i^m = a_i^{m-1} + K_m a_{m-i}^{*(m-1)} \quad i = 0,1, \ldots, m
   \]
   (5d)

4. Calculate the prediction error power \( P_m \)
   \[
   P_m = (1 - |K_m|^2) P_{m-1} .
   \] (5e)

5. Calculate the forward and backward forecasting errors
   \[
   f_m(n) = f_{m-1}(n) + K_m b_{m-1}(n-1) .
   \]
   \[
   b_m(n) = b_{m-1}(n - 1) + K_m^* f_{m-1}(n) .
   \] (5f, 5g)

6. If \( m < p \), then \( m = m + 1 \), return to step 2; if \( m = p \), then recursive end.

2.2.3. Covariance algorithm and modified covariance algorithm. Covariance algorithm is similar to Yule-Walker algorithm. Both of them calculate model parameters according to the principle that make prediction error power minimized. However, covariance algorithm does not need to fill zero at both ends of the data to calculate model parameters. The modified covariance algorithm calculates the model parameters according to the principle that make the power average value of forward and backward prediction error minimized. When estimating the frequency of sine wave signal, the modified covariance algorithm is not very sensitive to the phase of sine signal, and the peak shift caused by noise is small, so this algorithm has more advantages than other AR spectrum estimation algorithms. However, the arithmetic labor of the modified covariance algorithm is large, and when the signal-to-noise ratio is low, there will be large deviation and variance.
2.3. The order selection of the model

In the modern spectral estimation based on AR model, MA model and ARMA model, the order selection of the model is very important. The so-called order selection of the model here is the determination of the parameter $p$ of AR model, parameter $q$ of MA model and parameter $p$ and $q$ of ARMA model. For example, when AR model is used to fit a random signal, if the order of the selected AR model is lower than the actual order of the signal to be fitted, the estimated power spectrum will smooth the power spectrum of the signal itself, thus reducing the frequency resolution; if the order is selected too high, false peak and line splitting will occur.

Although there is no ideal method to determine the order, many criteria for order selection can be used for reference. Among them, FPE criterion, AIC criterion and BIC criterion are well-known.

2.3.1. FPE (Final Predict Error) criterion. In 1969, Japanese researcher Akaike proposed the minimum FPE (final prediction error) criterion to identify the order of AR model. The order of AR model is determined by the final prediction error of the model. The final prediction error of AR ($P$) model is defined as

$$FPE(p) = \hat{\sigma}_{wp}^2 \frac{(N+p+2)}{(N-p-2)}.$$  (6)

In the formula, $\hat{\sigma}_{wp}^2$ is the variance of linear prediction error.

When in use, first determine the highest fitting order $M$, it usually an integer between $N/3$ and $N/2$; then establish the model $AR(1), \cdots, AR(M)$ and calculate the corresponding FPE value, if $p_0$ satisfies

$$FPE(p_0) = \min_{1 \leq p \leq M} FPE(p).$$  (7)

Then $p_0$ is the order of AR model, and $AR(p_0)$ is called the best model under FPE criterion.

2.3.2. AIC criterion. In 1974, Akaike put forward the most famous minimum information criterion, also known as AIC criterion, which can be applied to order analysis of AR model and ARMA model. The performance index function of AIC criterion is

$$AIC(p, q) = \ln \sigma_{wp}^2 + (p + q) \frac{\ln N}{N}.$$  (8)

If $(p_0, q_0)$ satisfies

$$AIC(p_0, q_0) = \min_{1 \leq p \leq M} AIC(p, q).$$  (9)

Then $(p_0, q_0)$ is the order of ARMA model.

2.3.3. BIC criterion. The performance index function of BIC criterion is

$$BIC(p, q) = \ln \sigma_{wp}^2 + (p + q) \ln \frac{N}{N}.$$  (10)

2.3.4. MDL criterion. In 1983, Rissanen proposed the MDL criterion. Its performance index function is

$$MDL(p, q) = N \ln \sigma_{wp}^2 + (p + q) \ln N.$$  (11)

2.3.5. CAT function criterion. In 1974, Parzen proposed the CAT function criterion. Its performance index function is

$$CAT(p, q) = \left( \frac{1}{N} \sum_{k=1}^{N} \frac{1}{\sigma_{k,q}^2} \right) - \frac{1}{\sigma_{p,q}^2}.$$  (12)

2.4. Properties of AR model spectral estimation

2.4.1. Smoothness properties of AR spectrum. Because AR model is a rational fraction, the estimated spectrum is smoother than that of classical method. Figure 2 is the power spectrum estimation of a sine signal in the background of Gaussian white noise whose SNR is 30 dB. In the graph, the thick line is the AR spectrum, and the thin double line is the periodic spectrum. Obviously, the AR spectrum is much smoother than the periodic spectrum.
2.4.2. Resolution of AR spectrum. If the sampling interval is $\Delta t$, then the resolution of the signal whose length is $N$ will be $\frac{f_s}{N}$ roughly when DFT is used for spectrum analysis. The resolution of classical spectral estimation is inversely proportional to the length of the signal used. The resolution of modern spectral estimation is not limited by this limitation. This is because, for a given data $x_N(n)$, $n = 0, 1, \cdots, N - 1$, although the estimated autocorrelation function is also limited in length, i.e. $m = -(N - 1) \sim (N - 1)$, some methods of modern spectral estimation imply the extrapolation of data and autocorrelation function, make its possible length exceeds the given length.

AR model is the fitting of given data in the sense of least mean square (LMS), i.e.

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n - k).$$  \hspace{1cm} (13)

In this way, the possible length of $\hat{x}(n)$ is from 0 to $(N - 1 + p)$. In addition, if you replace $x(n)$ with $\hat{x}(n)$, you can continue to extrapolate.

AR spectrum $P_{AR}(e^{j\omega})$ corresponds to an autocorrelation function of infinite length, which is recorded as $r_a(m)$, i.e.

$$P_{AR}(e^{j\omega}) = \frac{\rho_p}{1 + \sum_{k=1}^{p} a_k e^{-j\omega k}} = \sum_{m=-\infty}^{\infty} r_a(m) e^{-j\omega m}. \hspace{1cm} (14)$$

It can be proved that $r_a(m)$ has the following relations with the true autocorrelation function $r_x(m)$:

$$r_a(m) = \begin{cases} r_x(m) & |m| \leq p \\ -\sum_{k=1}^{p} a_k r_a(m - k) & |m| > p \end{cases}. \hspace{1cm} (15)$$

Equation (15) is called "autocorrelation function" matching property of AR model. These $p + 1$ values $r_x(0), r_x(1), \cdots, r_x(p)$ can be used to characterize a $p$-order AR model. Thus the spectrum $P_{AR}(e^{j\omega})$ obtained from AR (P) model corresponds to an infinite autocorrelation sequence $r_a(m)$, which is completely equal to $r_x(m)$ when $m = 0, 1, \cdots, p$, while when $m > p$, $r_a(m)$ is extrapolated from equation (15). In the autocorrelation method of classical spectral estimation, there are

$$\hat{P}_{BT}(e^{j\omega}) = \sum_{m=-p}^{p} \hat{r}_x(m) e^{-j\omega m}. \hspace{1cm} (16)$$

It regards all autocorrelation functions as zero except when $|m| > p$, and its resolution is inevitably limited by the width $(-p \sim p)$ of window function. However, the autocorrelation function corresponding to AR model spectrum is not equal to zero when $|m| > p$. It can be extrapolated from equation (15), so the influence of window function is avoided. This is the main reason for the high resolution of AR model spectrum.

3. SIMULATION AND ANALYSIS

For the fault signal of a certain mechanical part of a helicopter, a real sine signal with single frequency under the background of Gauss white noise is used for simulation. But there is inevitably vibration noise in the actual scene. At the same time, in order to verify the high resolution of the frequency estimation method based on AR spectrum, another harmonic signal is added to the simulation signal used, which is in the following form:
\[ x_1(n) = s_1(n) + s_2(n) + z_1(n), \quad n = 0, 1, \ldots, N - 1. \quad (17) \]

Among them, \( s_1(n) \) and \( s_2(n) \) are two single frequency real sine signal with frequency of 80Hz and 80.3Hz respectively, and their amplitude are 10V and 6V respectively; \( z_1(n) \) is a random noise which obeys to normal distribution and removes mean, its sampling point is 1024 and sampling frequency is 1000Hz. When SNR = 30dB, the signal waveform used for simulation is shown in Figure 3.

Figure 4 is the simulation result when SNR = 30dB. Figure 4 (a) is the partial spectrum obtained by FFT method; Figure 4 (b) - (d) is the partial power spectrum based on AR model, using Yule-Walker algorithm, Burg algorithm and modified covariance method respectively. The FFT method uses 100 times zero filling. Figure 5 and Figure 6 are the simulation situations when SNR = 15dB and 3dB respectively.
It can be seen from Figure 4, Figure 5 and Figure 6 that under the three SNR conditions, FFT method and Yule-Walker algorithm can't distinguish the signal with the frequency of 80.3Hz. At this time, these signals that can't be distinguished will be "combined", thus frequency estimation error will be generated. Burg algorithm and modified covariance method can distinguish signals with the frequency of 80 Hz and 80.3 Hz, but when the SNR is low, the effect is bad. With the increase of SNR, the resolving effect is also improved.

Although the modified covariance method has the best resolution, considering that the Burg algorithm is a more general method, the calculation is not too complex, and also gives a better spectral estimation quality. Therefore, in this paper, Burg algorithm is used to calculate the coefficient of autoregressive model. According to the power spectrum, the frequency of fault signal can be obtained accurately.

![Figure 6. Comparison of several methods when SNR = 3dB.](image)

4. CONCLUSION
Helicopter is a complex mechanical system. The prediction and diagnosis of fault signals in advance is of great significance to flight safety. The traditional power spectrum analysis method has the limitations of high SNR and long sampling time in practical application. In this paper, based on the modern spectral estimation theory, the method of AR model spectral estimation is used to extract the characteristic frequency of helicopter mechanical fault signal. The simulation results show that the effect is better.

REFERENCES
[1] W.L. Zhang, “Vibration fault analysis of helicopter based on vibration prediction technology,” Modern Machinery, No.5, pp.26-30,2019.
[2] B. Cao, “Fault diagnosis research and management system design based on helicopter transmission system,” Electronic Test, No.21, pp.19-21,2019.
[3] F. Gao, H.Y. Li and F. Wu, “Health status monitoring and fault diagnosis based on spectrum analysis for centrifugal pump,” Process Automation Instrumentation, No.7, pp. 24-28, 2019.
[4] C.P. Shang and S.H. Zhang, “Research on bearing fault diagnosis method based on EMD and AR model”, Food and Machinery, No.7, pp.117-120,2019.
[5] M.X. Liu and X.G. Wang, “The realization of AR model power spectrums estimation based on MATLAB,” Electronic Design Engineering, Vol.25, No.17, pp.129-132,2017.
[6] F. Lu, “Spectrum estimation method and simulation based on AR model,” Journal of Xi ’ an University ( Natural Science Edition), Vol.22, No.5, pp. 29-32,2019.
[7] Z. Meng and Z. Liang, “The fault diagnosis for rotating machinery based on BSS and AR Spectrum Estimation,” ACTA METROLOGICA SINICA, Vol.36, No.3, 29-32,2015.
[8] L. Qiao, L.G. Zhang, X.G. Chen, and Y.K. Xiao, “Feature Extraction of Engine Faults Based on Order Ratio Analysis and AR Spectral Estimation,” Automotive Engineering, Vol.35, No.5, pp.462-466, 2013.