

# 基于主成分分析的特征频率提取算法及应用\*

李 振, 李伟光, 赵学智, 郑相立

(华南理工大学机械与汽车工程学院 广州, 510640)

**摘要** 通过研究主成分分析(principal component analysis, 简称 PCA)中有效特征值与信号频率和幅值之间的关系, 发现有效特征值的数量是由原始信号中频率成分的个数决定, 与幅值、频率和相位的大小无关。信号中每个频率产生两个有效的特征值, 且幅值决定协方差矩阵  $C$  的特征值在其分布图中的排列顺序。提出了一种基于 PCA 的特征频率提取算法, 该算法可实现对单个或多个特征频率的准确提取。将此方法应用于大型转子系统轴心轨迹的提纯上, 效果优于谐波小波和小波包算法。

**关键词** 故障诊断; 信号分析; 主成分分析; 特征提取; 轴心轨迹

**中图分类号** TH113.1; TH165.3; TN911.7

## 1 问题的引出

近年来, 主成分分析<sup>[1-2]</sup>在图像处理<sup>[3]</sup>、数据压缩<sup>[4]</sup>、故障诊断<sup>[5]</sup>、神经网络<sup>[6]</sup>、模式识别<sup>[7]</sup>及小波变换<sup>[8]</sup>等领域应用广泛。Kirby 等<sup>[3]</sup>运用 PCA 算法对图像进行压缩并提取图像主要特征, 结合 BP 神经网络算法进行人脸识别。Xi 等<sup>[4]</sup>和 Malhi 等<sup>[9]</sup>分别利用 PCA 算法对原始数据降噪并提取过程特征信息, 然后利用神经网络作为分类器, 实现滚动轴承的故障诊断。Xu 等<sup>[10]</sup>将 PCA 算法应用到离心式压缩机的叶轮故障诊断中, 通过 PCA 实现多时程数据的压缩。Sun 等<sup>[11]</sup>分析了常规故障诊断方法的缺陷后, 将数据挖掘技术引入故障诊断领域, 提出了基于 C4.5 决策树和 PCA 的旋转机械故障诊断方法, 其中 PCA 算法用于减少数据特征。

这些应用均是利用 PCA 进行信号降噪或数据压缩。通过累积贡献率  $L_l$  确定有效主成分个数, 即

$$L_l = \sum_{i=1}^l \lambda_i / \sum_{j=1}^m \lambda_j \quad (1)$$

虽然这些方法在信号降噪及数据降维等方面取得了良好效果, 但对于如何采用 PCA 算法提取或消除指定的特征频谱(单个频率)少有报道。例如, 轴心轨迹的提纯需要提取原始信号的基频、二倍频等特征频率进行轴心轨迹的合成, 对于这一问题的解决大多采用小波包变换<sup>[12-13]</sup>, 谐波小波<sup>[14-15]</sup>, 经验

模态分解<sup>[16]</sup>及稀疏分解等方法, 而 PCA 算法在这方面的应用少有研究。文献[17]虽然提出了基于奇异值分解算法在轴心轨迹提纯上的应用, 但该方法是通过累积贡献率进行信号降噪, 合成轴心轨迹。此方法并不具有提取特征频率的能力。

笔者探讨的关键问题是如何利用 PCA 算法从原始信号中提取指定的单个或多个特征频率。通过探究主成分分析中协方差矩阵  $C$  的特征值  $\lambda_i$  分布规律与特征频率的对应关系, 提出了一种基于 PCA 分离特定频率的算法, 并将该方法应用到大型转子振动试验台轴心轨迹的提纯上, 取得了良好的提纯效果, 明显优于谐波小波和小波包算法。

## 2 PCA 处理信号的基本原理与方法

设有  $m$  个随机向量  $x_1, x_2, \dots, x_m$ , 每个向量  $x_i$  含有  $n$  个样本, 即  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ , 则构造一个  $m \times n$  阶的矩阵  $X$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (2)$$

其中:  $X = [x_1, x_2, \dots, x_m]^T$ 。

若  $x_1, x_2, \dots, x_m$  表示原变量指标, 则对矩阵  $X$  进行主成分分析, 可得  $l$  个新的变量指标  $y_i (i = 1, 2, \dots, l (l \leq m))$  的表达式为

\* 国家高技术研究发展计划(“八六三”计划)资助项目(2015AA043005);国家自然科学基金资助项目(51375178)  
收稿日期: 2017-09-17; 修回日期: 2018-02-13

$$y_i = \alpha_{i1}x_1 + \alpha_{i2}x_2 + \dots + \alpha_{im}x_m = \alpha_i^T X \quad (3)$$

其中:  $y_i \in \mathbf{R}^{1 \times n}$ 。

由主成分分析的定义可知  $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im})^T (i=1, 2, \dots, l)$  为  $X$  的协方差矩阵中第  $i$  个特征值(由大到小排列)对应的特征向量,且  $\alpha_i$  满足

$$\begin{cases} \alpha_i^T \alpha_j = 1 & (i = j) \\ \alpha_i^T \alpha_j = 0 & (i \neq j) \end{cases} \quad (4)$$

式(2)所示的矩阵  $X$  的协方差矩阵为

$$C = \begin{bmatrix} \text{cov}(x_1, x_1) & \text{cov}(x_1, x_2) & \dots & \text{cov}(x_1, x_m) \\ \text{cov}(x_2, x_1) & \text{cov}(x_2, x_2) & \dots & \text{cov}(x_2, x_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(x_m, x_1) & \text{cov}(x_m, x_2) & \dots & \text{cov}(x_m, x_m) \end{bmatrix} \quad (5)$$

其中:  $\text{cov}(x_i, x_j) = E[(x_i - E(x_i))(x_j - E(x_j))^T]$ 。

由主成分分析原理可知,协方差矩阵  $C$  的特征方程为

$$C\alpha_i = \lambda_i \alpha_i \quad (6)$$

其中:  $\lambda_i$  为矩阵  $C$  的特征值;  $\alpha_i$  为与  $\lambda_i$  对应的特征向量。

把特征值按由大到小的顺序排列  $\lambda_1 > \lambda_2 > \dots > \lambda_l$ 。根据式(1)和式(3)可实现数据降维,由原来的  $m$  个变量转换成新的  $l$  个变量。若要进行信号处理,则根据式(3)进行信号重构,两边左乘  $\alpha_i$  并进行求和,得到

$$\sum_{i=1}^m \alpha_i y_i = \sum_{i=1}^m \alpha_i \alpha_i^T X = L_m X \quad (7)$$

根据累积贡献率  $L_l$  选择前  $l$  个主成分进行重构,得到一个近似的矩阵

$$\hat{X} = \sum_{i=1}^l \alpha_i y_i = \sum_{i=1}^l \alpha_i \alpha_i^T X \quad (8)$$

其中:  $\hat{X} \in \mathbf{R}^{m \times n}$ 。

与原矩阵  $X$  相比,近似重构的矩阵  $\hat{X}$  保留了原始矩阵的大部分信息,却消除了原始矩阵的冗余信息<sup>[4, 7]</sup>,例如噪声与工频干扰。

获得  $\hat{X}$  后需按矩阵的构成方式恢复信号,将  $\hat{X}^T$  的行向量首尾相连变成  $\tilde{X}$ ,  $\tilde{X} \in \mathbf{R}^{1 \times mn}$ ,通过矩阵的逆变换,从重构矩阵  $\tilde{X}$  得到恢复后的信号  $\hat{x}$ ,即

$$\hat{x} = \tilde{X}M^+ \quad (9)$$

其中:  $M^+ = (M^T M)^{-1} M^T \in \mathbf{R}^{mn \times L}$ ,  $M^+$  称为  $M$  的伪逆,  $M \in \mathbf{R}^{L \times mn}$ ;  $L$  为信号长度,  $M$  是由  $m$  个  $n$  阶单位矩阵构成。

以  $m=3, n=3$  的矩阵为例,表示为

$$M = \begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & 1 & & & & \\ & & & 1 & & & \\ & & & & 1 & & \\ & & & & & 1 & \\ & & & & & & 1 \end{bmatrix} \quad (10)$$

通过伪逆矩阵恢复信号  $\hat{x}$ ,实际是对矩阵  $\hat{X}$  各副对角线上的元素求平均,这与文献[18]中采用平均法恢复信号的结果一致。由于矩阵  $M$  为稀疏矩阵,尤其当  $m$  与  $n$  的取值较大时,计算伪逆矩阵  $M^+$  所需要的内存和时间会成倍增加。因此,笔者采用平均法从矩阵  $\hat{X}$  中恢复信号  $\hat{x}$ ,表达式为

$$\hat{x}(k) = \begin{cases} \sum_{i=1}^k \hat{x}_{i, k+1-i} / k & (1 \leq k < m) \\ \sum_{i=1}^m \hat{x}_{i, k+1-i} / m & (m \leq k < n) \\ \sum_{i=k-n+1}^m \hat{x}_{i, k+1-i} / (m+n-k) & (n < k \leq m+n-1) \end{cases} \quad (11)$$

其中:  $\hat{x}_{i,j}$  为近似重构矩阵  $\hat{X}$  中第  $i$  行、第  $j$  列的元素。根据式(11)可以恢复出信号  $\hat{x}$ 。

### 3 有效特征值与频率内在规律

#### 3.1 有效特征值与信号特征的关系

有效特征值就是非零特征值。非零特征值和信号中的频率数量有重要联系。为了说明这一联系,首先构造不同幅值、频率和相位的无量纲信号,即

$$x(t) = \sum_{i=1}^k A_i \sin(2\pi f_i t + \varphi_i) \quad (12)$$

以采样频率为 1 024 Hz 采集 4 096 点,对  $x(t)$  信号构造 Hankel 矩阵。构造矩阵行数为  $m$ ,列数为  $n$ ,采用 PCA 分解算法进行信号分解与重构。文献[18]研究了构造 Hankel 矩阵对信号处理的影响,指出当信号长度  $L$  为偶数时,矩阵的行列分别为  $m=L/2, n=L/2+1$ ;当  $L$  为奇数时,  $m=(L+1)/2, n=(L+1)/2$ 。故可得  $m=2\ 048, n=2\ 049$ 。

首先,探讨当  $k=1, A_i=1$  时的情况。构造 3 个信号,各个信号的主成分特征值分布如图 1 所示,特征值个数  $q$  的取值范围为  $[1, 2\ 048]$ ,本例仅显示前 50 个特征值的分布情况。

当  $k=1$  时,构造的每个信号中仅含有一个频率成分,幅值  $A_i$  相同,频率  $f_i$  和相位  $\varphi_i$  不同。由图 1 可看出,每个信号产生两个非零的特征值  $\lambda_1$  与  $\lambda_2$  一前一后紧密排列,尽管信号不同,但各个信号的非零特征值却是相等的。

其次,探讨当  $k=2, A_i$  分别取 1 和 0.8 时的情况。各信号的第 1 个频率成分的幅值均为 1,第 2 个频率的幅值均为 0.8,在这种情况下得到这 3 个信号的主成分特征值分布如图 2 所示。可以看出,

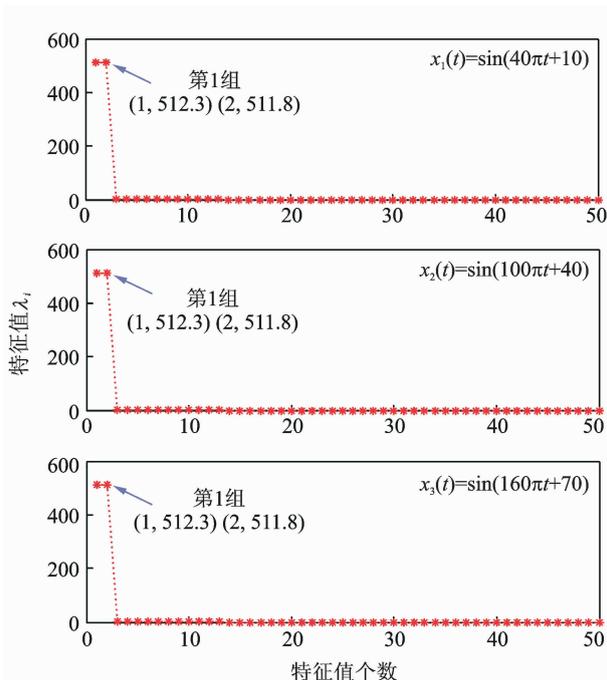


图1 含有1种频率成分的协方差矩阵  $C$  的特征值分布  
Fig.1 Eigenvalues distribution of covariance matrix  $C$  with one frequency component

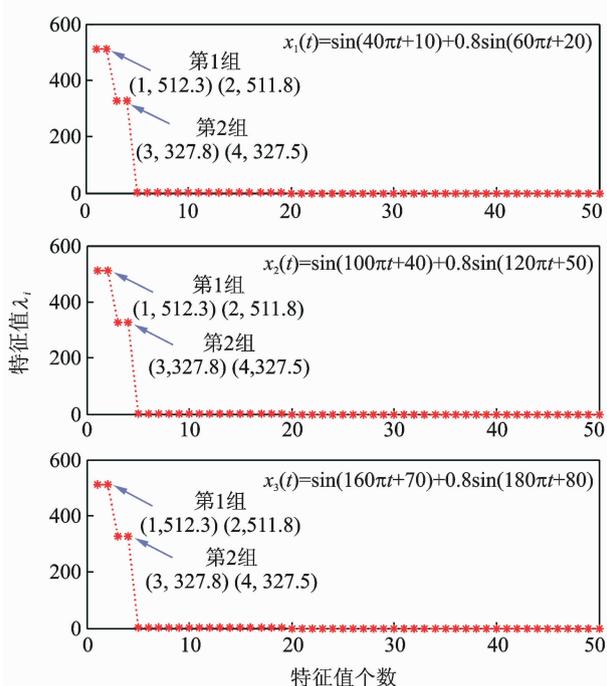


图2 含有2种频率成分的协方差矩阵  $C$  的特征值分布  
Fig.2 Eigenvalues distribution of covariance matrix  $C$  with two frequency component

每个信号  $x_i(t)$  产生两组非零的特征值, 每组特征值包含两个特征值  $\lambda_i$  与  $\lambda_{i+1}$ , 一前一后紧密排列。每个信号产生 4 个非零特征值, 尽管信号不同, 但各个信号的非零特征值却依次对应相等。

对比图 2 与图 1 发现, 两幅图中第 1 组特征值的大小完全相同, 因为图 2 是在图 1 的基础上增加了一个幅值为 0.8 的频率成分, 故可以确定图 2 中的第 1 组非零特征值  $\lambda_1$  与  $\lambda_2$  是由幅值  $A_1$  为 1 的频率成分产生的, 而图 2 中的第 2 组非零特征值是由幅值  $A_2$  为 0.8 的频率成分产生的。

最后, 当  $k=3$ ,  $A_i$  取 1, 0.8 和 0.6 时, 构造 3 个信号。每个信号含有 3 个频率成分, 各信号的第 1 个频率幅值均为 1、第 2 个频率的幅值均为 0.8、第 3 个频率的幅值均为 0.6。在这种情况下得到各信号的特征值分布如图 3 所示。

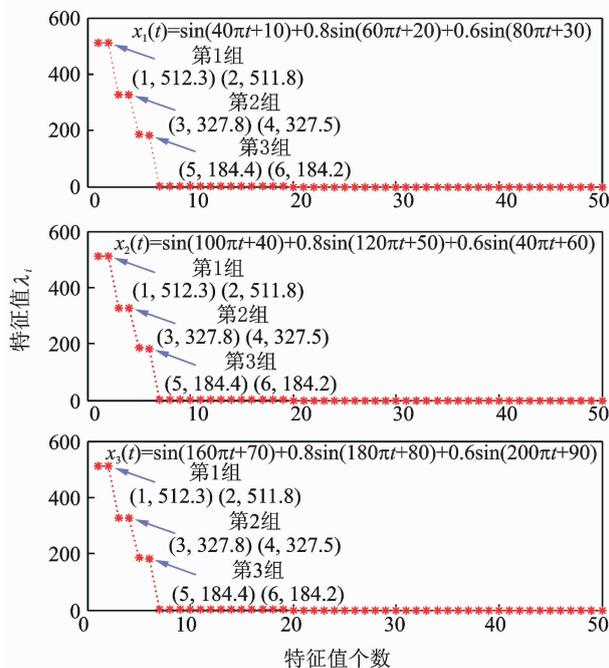


图3 含有3种频率成分的协方差矩阵  $C$  的特征值分布  
Fig.3 Eigenvalues distribution of covariance matrix  $C$  with three frequency component

从图 3 可以看出, 每个信号  $x_i(t)$  均产生 3 组非零的特征值, 每组特征值包含两个临近的特征值  $\lambda_i$  与  $\lambda_{i+1}$ 。每个信号产生的 3 组特征值大小与另外两个信号产生的特征值的大小对应相等。

对比图 3 与图 2、图 1 可以发现, 第 1 组特征值的大小完全相同。由图 1 可知, 图 3 中的第 1 组非零的特征值  $\lambda_1$  与  $\lambda_2$  是由幅值  $A_1$  取 1 的频率成分产生的。对比图 2 与图 3 可以发现, 两图中的第 2 组特征值大小完全相同。因此, 图 3 中第 2 组非零特征值  $\lambda_3$  与  $\lambda_4$  是由幅值为 0.8 的频率成分产生的。最后, 可得图 3 中的第 3 组频率成分  $\lambda_5$  与  $\lambda_6$  是由幅值为 0.6 的频率成分产生的。

若继续增加信号的有效频率成分, 可得到同样

结果。因此可得出以下结论:信号  $x(t)$  在 Hankel 矩阵的构造模式下  $l = \min(m, n)$ , 其中:  $m$  为行数;  $n$  为列数;  $k$  为信号  $x(t)$  中的有效频率个数。

在满足采样定理的条件下,如果  $l > 2k$ , 则有:a. 信号中的每个频率成分产生两个非零的特征值,一前一后紧密排列;b. 信号有效特征值的个数仅与频率个数相关,与幅值  $A_i$ 、频率  $f_i$  和相位  $\varphi_i$  的大小均无关;c. 信号非零特征值在协方差矩阵  $C$  的特征值分布图中的排列顺序由信号频率的幅值  $A_i$  决定,某一频率的幅值越大,产生的两个特征值就越大,排列越靠前。

证明: 设某一信号为  $x(t) = a \sin(\omega t + \varphi)$ , 将信号构造一个  $m \times n$  的 Hankel 矩阵  $X$ , 且矩阵  $X$  可表示为  $X = [x_1, x_2, \dots, x_m]^T$ , 其中,  $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ 。由 PCA 定义可知, 矩阵  $X$  的协方差矩阵  $C$  为

$$C = E[(X - E(X)) \cdot (X - E(X))^T] \quad (13)$$

根据式(6), 右乘  $\alpha_i^T$  并求和, 得到

$$\sum_{i=1}^l C \alpha_i \alpha_i^T = \sum_{i=1}^l \lambda_i \alpha_i \alpha_i^T \quad (14)$$

由于一个频率产生两个有效特征值, 即  $l = 2$ , 得到

$$C I_m = \lambda_1 \alpha_1 \alpha_1^T + \lambda_2 \alpha_2 \alpha_2^T \quad (15)$$

则  $C$  的能量为

$$\begin{aligned} |C|^2 &= \lambda_1^2 \sum_{i=1}^m \alpha_{i1}^2 \sum_{i=1}^m \alpha_{i1}^2 + \lambda_2^2 \sum_{i=1}^m \alpha_{i2}^2 \sum_{i=1}^m \alpha_{i2}^2 + \\ &2\lambda_1 \lambda_2 \sum_{i=1}^m \alpha_{i1} \alpha_{i2} \sum_{i=1}^m \alpha_{i1} \alpha_{i2} \end{aligned} \quad (16)$$

根据式(4), 式(16)可转换为

$$|C|^2 = \lambda_1^2 + \lambda_2^2 \quad (17)$$

将信号  $x(t)$  构造 Hankel 矩阵后代入式(13), 得到协方差矩阵  $C$  的能量与  $a^2$  成正比,  $a^2$  越大, 矩阵  $C$  的能量就越大, 根据式(17)可知  $\lambda_i$  也越大。因此, 某频率成分的幅值越大, 其对应的特征值就越大。

可见, 如果确定了原始信号幅值谱中某一指定频率成分幅值的排序, 便可以根据特征值分布重构出对应的频率成分, 从而实现单个或多个特征频率的提取。例如在幅值谱中, 某一频率的幅值大小排序为  $k$ , 那么选择协方差矩阵特征值分布图中的第  $2k-1$  与第  $2k$  个特征值对应的特征向量进行信号重构, 便可提取出此频率成分。

### 3.2 基于 PCA 的特征频率提取算法

根据频率与非零特征值的关系, 提出一种基于 PCA 分离特征频率的方法, 步骤如下:

1) 对于一个确定的信号  $x(t)$ , 通过 FFT 滤出原始信号的直流分量, 利用消除直流分量的信号构造 Hankel 矩阵  $X$ ;

2) 求矩阵  $X$  的协方差矩阵  $C$ , 并求其特征值  $\lambda_i$ , 按从大到小排列为  $\lambda_1, \lambda_2, \dots, \lambda_m$ , 求得对应特征向量为  $\alpha_1, \alpha_2, \dots, \alpha_m$ ;

3) 根据特征值的  $\lambda_i$  分布, 选择与某一具体频率对应的两个特征值及相应的特征向量进行重构。若某一频率在原始信号幅值谱中的幅值大小排序为  $k$ , 则选择协方差矩阵特征值分布图中的第  $2k-1$  与第  $2k$  个特征值对应的特征向量进行信号重构, 得到一个新的矩阵;

4) 将重构后的矩阵叠加原矩阵的均值获得矩阵  $\hat{X}$ ;

5) 从矩阵  $\hat{X}$  中采用平均法重构信号  $\hat{x}$ , 该信号就是需要提取的特征频率分量。

在式(7)中,  $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im})^T$  为  $X$  协方差矩阵中的第  $i$  个特征值对应的特征向量, 则  $\alpha_1 \alpha_1^T X$  和  $\alpha_2 \alpha_2^T X$  对应幅值最大的频率成分,  $\alpha_3 \alpha_3^T X$  和  $\alpha_4 \alpha_4^T X$  对应幅值第二大的频率成分, 后面依次类推。式(8)表明, 在本研究算法中, 原始信号与被提取的各分量信号之间为一种加减法的关系, 对一个特定分量信号的提取等同于采用减法运算将这一分量信号从原始信号中分离出来, 这种减法运算不会影响被提取分量信号的相位, 这就决定了 PCA 算法是一种零相移频率提取算法。这种特性有别于一般由卷积运算实现的滤波器, 如果滤波器非线性相位的特性, 则其提取出来的分量信号会出现相位失真的现象。PCA 算法的这种简单的加减法运算特性, 决定了 PCA 算法不存在相位失真的问题。

### 3.3 频率提取实例

通过一实例来验证本研究算法的正确性, 构造含噪信号为

$$\begin{aligned} y &= \sin(40\pi t + 20) + 0.5 \sin(100\pi t + 40) + \\ &0.7 \sin(98\pi t + 50) + 0.8 \sin(160\pi t + 60) + e(n) \end{aligned} \quad (18)$$

其中:  $e(n)$  是标准差为 1.2 的高斯白噪声。

对此信号以采样频率 1 024 Hz 采集 4 096 个数据点, 结果如图 4 所示。利用本研究方法分离这 4 个频率, 原始信号的特征值分布如图 5 所示。除含有 4 组比较大特征值外, 由于噪声  $e(n)$  的存在,  $y$  的大部分特征值都是非零的特征值, 噪声信号幅值相对较小, 产生的特征值分布在有效特征值之后。

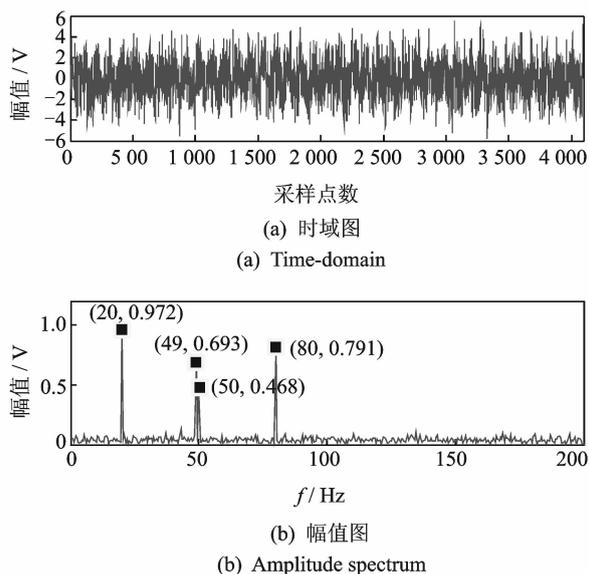


图4 原始信号时域图和幅值谱

Fig. 4 Time-domain and amplitude spectrum of raw signal

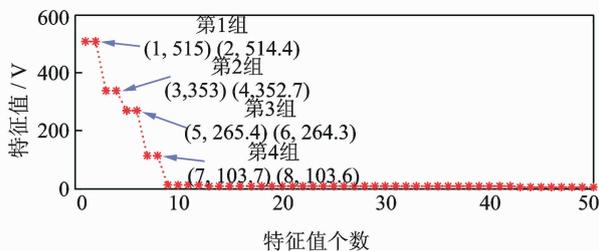


图5 信号 y 的协方差矩阵 C 的特征值分布

Fig. 5 Eigenvalues distribution of the covariance matrix C of y

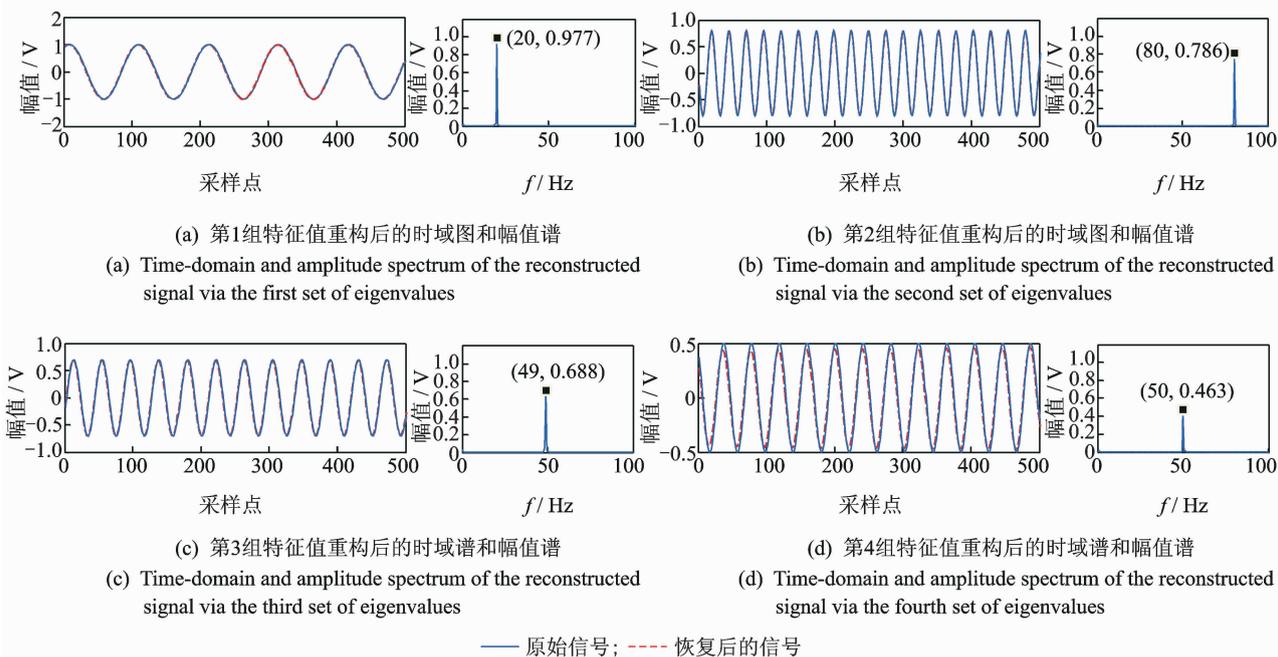


图6 重构信号的幅值谱

Fig. 6 The time and frequency domains of reconstructed signal

将图5中的第1组特征值 $\lambda_1$ 与 $\lambda_2$ 进行重构,结果如图6(a)所示。可见,重构信号的频率为20 Hz,幅值为0.977,这与原始信号中的 $\sin(40\pi t + 20)$ 频率成分一致。

利用第2组特征值 $\lambda_3$ 与 $\lambda_4$ 进行重构,得到信号的时域图和幅值谱如图6(b)所示。可见,重构信号的频率为80 Hz,幅值为0.786,与理想信号的相位完全重合。这说明恢复的信号与原始信号 $0.8\sin(160\pi t + 60)$ 频率成分一致。

采用第3组特征值 $\lambda_5$ 与 $\lambda_6$ 重构得到的结果如图6(c)所示。其频率为49 Hz,幅值为0.688,这与原始信号中 $0.7\sin(98\pi t + 50)$ 频率成分一致。利用最后一组特征值重构得到信号的时域图和幅值谱如图6(d)所示。其频率为50 Hz,幅值为0.463,这与原始信号中 $0.5\sin(100\pi t + 40)$ 频率成分一致。

可见,本研究算法可准确地将原信号中待提取的各个频率成分分离出来,且得到的频率分量和理想的频率成分非常吻合,没有相位失真,这也表明PCA算法是一种零相移频率提取算法。另外,分离出来的各频率分量的幅值非常接近理想成分的幅值,二者相差极小,这种幅值误差可以忽略,不会对实际应用造成影响。

上述结果也证实了原始信号中频率的幅值大小决定特征值在其分布图中的排列顺序。原始信号中某一频率的幅值越大,其特征值在分布图中的排列顺序也就越靠前,且特征值成对出现。同时,对于

49 和 50 Hz 这两个相差仅仅 1 Hz 的频率分量,本算法也能近乎完美地将二者分离开,这是一般带通滤波器难以实现的。

### 4 轴心轨迹提纯

在旋转机械设备中,通常不同形状的轴心轨迹反映不同的故障类型,可根据轴心轨迹形状判断系统对应的故障类型。例如:正常情况下,轴心轨迹通常为长短轴相差不大的椭圆;规则或不规则的花瓣状轴心轨迹通常是由动静件碰磨的故障引起的;内“8”字形的轴心轨迹可反映出油膜涡动的故障;外“8”字或香蕉形的轴心轨迹一般对应的是不对中故障<sup>[14-17, 19]</sup>等。

虽然轴心轨迹的形状是判断设备故障的重要依据,但大型旋转机械运行时工况复杂,其主轴位移信号受电磁干扰或噪声干扰严重,从未提纯的轴心轨迹中难以得到任何有效信息。对于如何提纯轴心轨迹,目前使用较多的方法有小波变换、小波包、谐波小波以及经验模态分解等。这里利用本研究算法提纯轴心轨迹,并与现有方法进行对比。

#### 4.1 试验设备

轴心轨迹的试验数据来自实验室自主研制的大型转子振动测试试验台,如图 7 所示。其中:A 端为近电机端;B 端为远电机端。



图 7 大型转子振动测试试验台

Fig. 7 Large rotor vibration test bed

为实时监测转子的工作情况,在试验台转子 A/B 两侧的轴颈处,以相互垂直的方式安装两个 Kaman KD2306-1S 电涡流位移传感器,如图 8 所示。采用 LMS Test. Lab 进行数据采集。

#### 4.2 轴心轨迹特征的提取

笔者仅对 A 端面处的  $D_{11}$  与  $D_{12}$  号电涡流位移传感器采集的数据进行分析。其中:试验台转速为

1 080 r/min,采样频率为 2 048 Hz,采样点为 4 096 点,得到的时域信号和幅值谱如图 9 和图 10 所示。

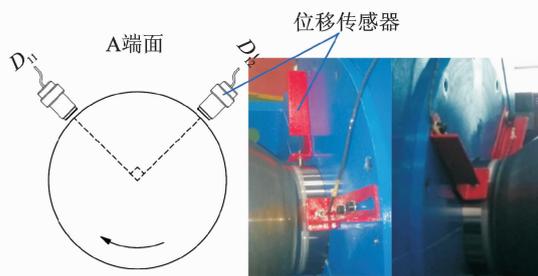
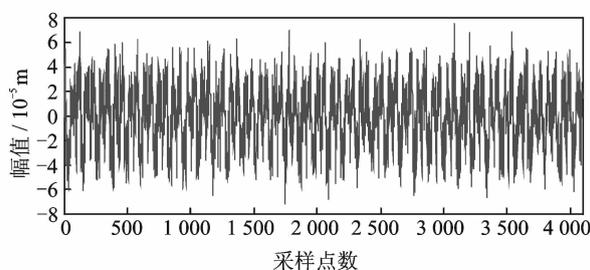
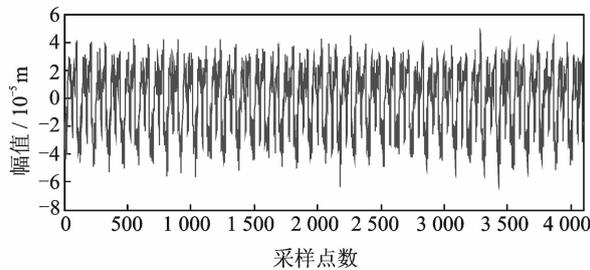


图 8 电涡流位移传感器安装示意图

Fig. 8 Installation of eddy current displacement sensor



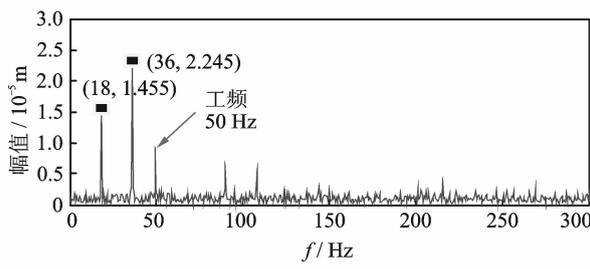
(a)  $D_{11}$



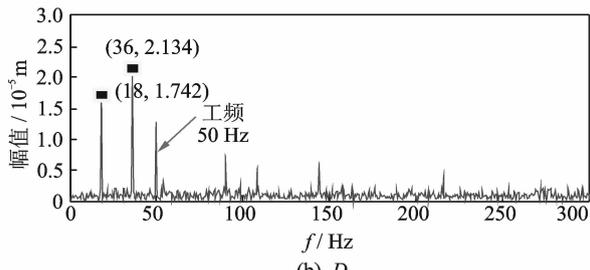
(b)  $D_{12}$

图 9 信号的时域波形

Fig. 9 Time-domain diagram



(a)  $D_{11}$



(b)  $D_{12}$

图 10  $D_{11}$  和  $D_{12}$  信号的幅值谱

Fig. 10 Frequency spectrum of  $D_{11}$  and  $D_{12}$  signal

利用图9中的两个位移信号合成的轴心轨迹如图11所示。其中:信号 $D_{11}$ 作为 $x$ 轴; $D_{12}$ 信号作为 $y$ 轴。在未作任何处理的情况下,根本无法看出轴心轨迹的形状。现采用笔者提出的PCA算法对轴心轨迹进行提纯。

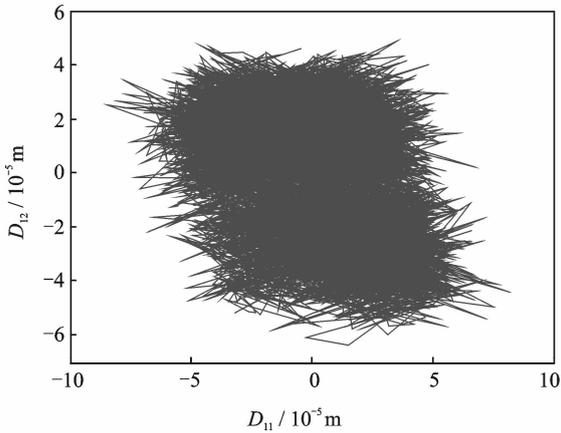


图11 A端面 $D_{11}$ 与 $D_{12}$ 原始轴心轨迹  
Fig. 11 Axis orbit of A-end face

一般说来,利用基频和二倍频合成的轴心轨迹比较可靠,高次谐波会使轴心轨迹变得复杂,甚至变得凌乱,故提取原始信号中的基频与二倍频进行轴心轨迹的合成。信号 $D_{11}$ 与 $D_{12}$ 的协方差矩阵特征值分布如图12所示。采用PCA分离特征频率的方法,根据图10中 $D_{11}$ 与 $D_{12}$ 各个频率幅值的排序,可知基频对应第2组特征值(第3,4个特征值),二倍频对应第1组特征值(第1,2个特征值)。为了提取基频和二倍频,利用图12中两个特征值序列中的前4个特征值进行重构,可提取出信号 $D_{11}$ 与 $D_{12}$ 中的基频和二倍频,其频域结果如图13所示。可见,PCA算法的频率提取效果非常好,不仅滤掉了噪声干扰,还滤除了工频50 Hz及其他分量的干扰。提取的频率幅值与原始信号的基频和二倍频幅值十分接近,说明频谱泄漏可忽略不计。

利用PCA提取到的基频和二倍频信号合成的轴心轨迹如图14所示。采用PCA滤波后合成的轴心轨迹为明显的外8字状,说明近电机端(A端)存在不对中现象。可见,PCA算法在轴心轨迹的提纯上有十分理想的效果。

#### 4.3 与现有轴心轨迹提纯方法的对比

谐波小波算法为一种改进小波算法<sup>[14-15, 20]</sup>,是旋转试验台常用的信号处理方法之一,能够对整个频带进行无限细分,规避了二进小波隔二抽一采样

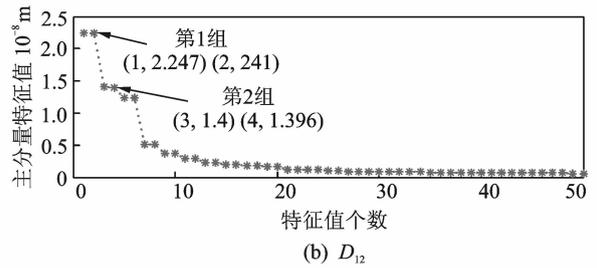
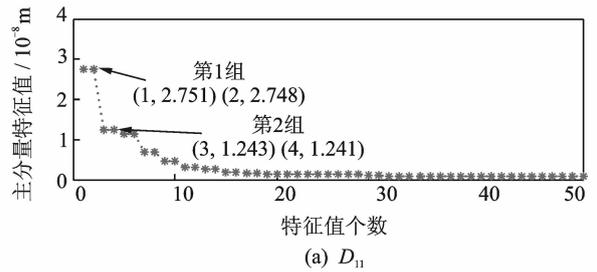


图12  $D_{11}$ 与 $D_{12}$ 信号的协方差矩阵的特征值分布  
Fig. 12 Eigenvalues distribution of  $D_{11}$  and  $D_{12}$  signal

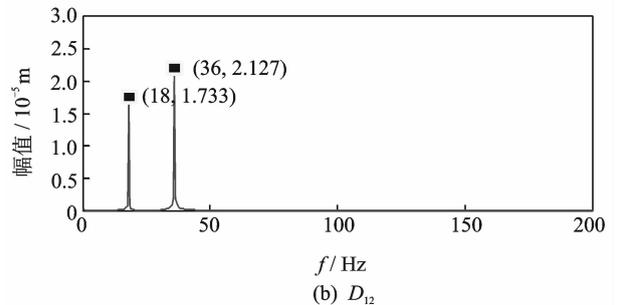
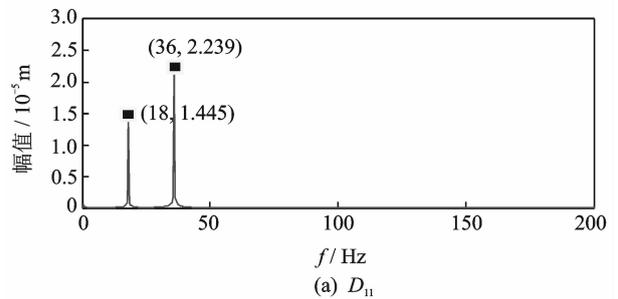


图13  $D_{11}$ 与 $D_{12}$ 信号采用PCA算法提取前的两倍频域  
Fig. 13 Extracting the time and frequency spectrums of the first two frequencies of  $D_{11}$  and  $D_{12}$

的不足,在旋转试验台轴心轨迹的提纯上,可以方便地提取基频及其倍频信号。

采用谐波小波算法对本例的轴心轨迹进行提纯,如图15所示。与图14的轴心轨迹相比,PCA算法提纯的轨迹更清晰,轨迹更紧凑,8字更明显,而图15中看不出8字的小头。

除了谐波小波外,小波包也常用来提纯轴心轨迹。对于本例的轴心轨迹,Daubechies小波包的提纯结果如图16所示。可见,这种轴心轨迹比谐波小波的结果更差,很难看出轴心轨迹形状。

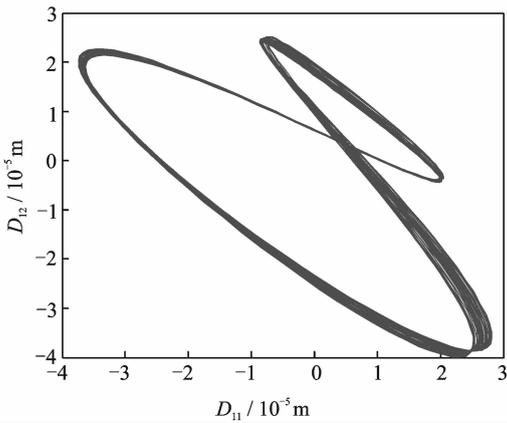


图14 采用前2倍频合成的轴心轨迹

Fig. 14 Axis orbit of the first two frequency syntheses

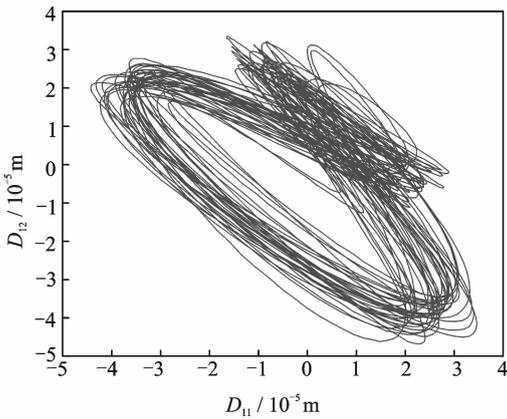


图15 谐波小波提取前2倍频合成的轴向轨迹

Fig. 15 Axis orbit of the first two frequency syntheses by harmonic wavelets

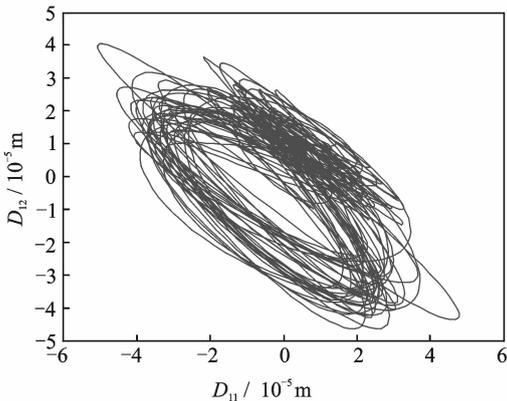


图16 小波包提取前2倍频合成的轴向轨迹

Fig. 16 Axis orbit of the first two frequency syntheses by wavelet packet

的紧密排列;信号有效的个数与信号的频率个数相关,与  $f_i$ ,  $A_i$  和  $\varphi_i$  的取值无关;信号有效特征值在其分布图中的排列顺序仅与信号频率对应的幅值  $A_i$  相关,与  $f_i$  和  $\varphi_i$  的取值无关,对应频率的幅值越大,所产生的特征值也越大,排列也就越靠前。

2) 基于信号中每个频率产生两个有效的特征值和频率对应的幅值大小决定特征值的排列顺序这些特性,提出了采用 PCA 算法提取指定的特征频谱(单个频率),或消除指定特征频率的方法,指出即使两个频率相差 1Hz 也能够不受影响地提取出特征频率。

3) 将 PCA 算法应用到轴心轨迹的提纯上,并与谐波小波和小波包的提纯结果进行对比,证明了本算法在轴心轨迹提纯上的优越性。

## 参 考 文 献

- [1] Abdi H, Williams L J. Principal component analysis [J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4): 433-459.
- [2] 付云晓,贾利民,秦勇,等. 基于 LMD-CM-PCA 的滚动轴承故障诊断方法[J]. 振动、测试与诊断, 2017, 37(2): 249-255.  
Fu Yunxiao, Jia Limin, Qin Yong, et al. Roller bearing fault diagnosis method based on LMD-CM-PCA [J]. Journal of Vibration, Measurement & Diagnosis, 2017, 37(2): 249-255. (in Chinese)
- [3] Kirby M, Sirovich L. Application of the karhunen-loeve procedure for the characterization of human faces [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(1): 103-108.
- [4] Xi Jianhui, Han Yanzhe, Su Ronghui. New fault diagnosis method for rolling bearing based on PCA [C]// Control and Decision Conference, 2013 25th Chinese. New York: IEEE, 2013: 4123-4127.
- [5] 刘永斌,何清波,孔凡让,等. 基于 PCA 和 SVM 的内燃机故障诊断[J]. 振动、测试与诊断, 2012, 32(2): 250-255.  
Liu Yongbin, He Qingbo, Kong Fanrang, et al. Fault diagnosis of internal combustion engine using PCA and SVM [J]. Journal of Vibration, Measurement & Diagnosis, 2012, 32(2): 250-255. (in Chinese)
- [6] Shao Keyong, Cai Miaomiao, Zhao Guofeng. Rolling bearing fault diagnosis based on wavelet energy spectrum, PCA and PNN [C]// Control and Decision Conference. 2014 26th Chinese. New York: IEEE, 2014: 800-804.
- [7] 陈安华,周博,张会福,等. 基于 PCA 与蚁群算法的机

## 5 结 论

1) 在 Hankel 矩阵方式下,信号中每个有效的频率成分最多只产生两个非零的特征值,一前

- 械故障聚类诊断方法[J]. 中国机械工程, 2013, 24(24): 3333-3337.
- Chen Anhua, Zhou Bo, Zhang Huifu, et al. Clustering method of mechanical fault diagnosis based on PCA and ant colony algorithm [J]. China Mechanical Engineering, 2013, 24(24): 3333-3337. (in Chinese)
- [8] 王莉莉,沈月,陈德运,等. PCA与小波变换的ECT图像融合方法[J]. 哈尔滨理工大学学报, 2016, 21(4): 30-35.
- Wang Lili, Shen Yue, Chen Deyun, et al. Image fusion of PCA and wavelet transform in electrical capacitance tomography [J]. Journal of Harbin University of Science and Technology, 2016, 21(4): 30-35. (in Chinese)
- [9] Malhi A, Gao R X. PCA-based feature selection scheme for machine defect classification [J]. IEEE Transactions on Instrumentation and Measurement, 2004, 53(6): 1517-1525.
- [10] Xu Shengli, Jiang Xiaomo, Huang Jinzhi, et al. Bayesian wavelet PCA methodology for turbomachinery damage diagnosis under uncertainty [J]. Mechanical Systems and Signal Processing, 2016, 80(1): 1-18.
- [11] Sun Weixiang, Chen Jin, Li Jiaqing. Decision tree and PCA-based fault diagnosis of rotating machinery [J]. Mechanical Systems and Signal Processing, 2007, 21(3): 1300-1317.
- [12] 段吉安,张小栋. 基于小波变换的轴心轨迹特征提取[J]. 振动、测试与诊断, 1997, 17(1): 31-34.
- Duan Jian, Zhang Xiaodong. Feature extraction from the orbit of the center of a rotor based on wavelet transform [J]. Journal of Vibration, Measurement & Diagnosis, 1997, 17(1): 31-34. (in Chinese)
- [13] 杨建国,夏松波,刘永光,等. 小波降噪在轴心轨迹特征提取中的应用[J]. 哈尔滨工业大学学报, 1999, 31(5): 52-54.
- Yang Jianguo, Xia Songbo, Liu Yongguang, et al. Wavelet denoising and its application in extraction of shaft center orbit features [J]. Journal of Harbin Institute of Technology, 1999, 31(5): 52-54. (in Chinese)
- [14] 张文斌,周晓军,林勇,等. 采用谐波窗方法提纯碰摩转子轴心轨迹[J]. 振动、测试与诊断, 2010, 30(1): 87-90.
- Zhang Wenbin, Zhou Xiaojun, Lin Yong et al. refinement of rotor center's orbit by a harmonic window method [J]. Journal of Vibration, Measurement & Diagnosis, 2010, 30(1): 87-90. (in Chinese)
- [15] 李舜酩,吕国志,许庆余. 转子轴心轨迹的谐波小波提纯[J]. 西北工业大学学报, 2001, 19(2): 220-224.
- Li Shunming, Lü Guozhi, Xu Qingyu. On obtaining accurate rotor sub-frequency signal with harmonic wavelet [J]. Journal of North Western Polytechnical University, 2001, 19(2): 220-224. (in Chinese)
- [16] 陈仁祥,汤宝平,吕中亮. EEMD滤波的转子轴心轨迹提纯方法[J]. 重庆大学学报, 2012, 35(11): 15-20.
- Chen Renxiang, Tang Baoping, Lü Zhongliang. A method of rotor orbit purification based on ensemble empirical mode decomposition filter [J]. Journal of Chongqing University, 2012, 35(11): 15-20. (in Chinese)
- [17] 刘维国. 基于奇异值分解的轴心轨迹提纯方法研究[J]. 声学技术, 2016, 35(5): 30-35.
- Liu Weiguo. Research on method of purification of shaft orbit based on singular value decomposition [J]. Technical Acoustics, 2016, 35(5): 30-35. (in Chinese)
- [18] 赵学智,叶邦彦. 分量形成方式对奇异值分解信号处理效果的影响[J]. 上海交通大学学报, 2011, 45(3): 368-374.
- Zhao Xuezhi, Ye Bangyan. The influence of formation manner of component on signal processing [J]. Journal of Shanghai Jiaotong University, 2011, 45(3): 368-374. (in Chinese)
- [19] 王腾. 基于信号稀疏表征的轴心轨迹提取方法研究[D]. 重庆:重庆大学, 2014.
- [20] Kougioumtzoglou I A, Spanos P D. An identification approach for linear and nonlinear time-variant structural systems via harmonic wavelets [J]. Mechanical Systems and Signal Processing, 2013, 37(1): 338-352.



**第一作者简介:**李振,男,1988年4月生,博士生。主要研究方向为故障诊断和信号处理。曾发表《基于迭代选差与延拓算法的MP稀疏分解研究》(《振动与冲击》2018年第37卷第17期)等论文。  
E-mail:scutlz@163.com

**通信作者简介:**李伟光,男,1958年4月生,博士、教授。主要研究方向为转子动力学、故障诊断与信号处理。  
E-mail:wguangli@scut.edu.cn

