

多约束鲁棒非负矩阵分解的冲击特征频带 自适应分解方法

梁霖¹, 崔旭军², 胡文昊², 代普森²

(1. 西安交通大学仪器科学与技术学院, 710049, 西安; 2. 西安交通大学机械工程学院, 710049, 西安)

摘要: 针对非负矩阵分解(NMF)在轴承故障的冲击振动频带分解中存在能量导向引起的频带混叠问题,提出了一种面向冲击故障频带自适应分解的多约束鲁棒非负矩阵分解方法。首先,采用 β 散度自适应加权的误差函数来避免信号未知分布引起的风险,以信号频带数量作为分解秩的选择参考,通过对基矩阵施加正交约束实现自适应频带划分;其次,结合周期冲击响应时频谱的光滑频域和稀疏时域的特性,引入具有良好物理意义的光滑、稀疏约束,构建了面向振动时频谱的多约束鲁棒非负矩阵分解模型;最后,借助正则化技术和 Stiefel 流形优化方法设计了求解算法。仿真和实验结果表明,与多种 NMF 方法和典型频带选择方法相比,在面对缺陷引起的微弱冲击时,低频区间中往往存在着多种干扰源影响,所提分解模型能准确提取出高频区间的冲击响应频带,避免了能量导向的传统频带分解方式不足,约束项的引入则有效地提升了 NMF 的求解结果,增强了 NMF 方法在冲击特征频带微弱时的辨识能力。

关键词: 冲击特征频带;非负矩阵分解;多约束;频带分解

中图分类号: TH133.33 **文献标志码:** A

DOI: 10.7652/xjtuxb202504016 **文章编号:** 0253-987X(2025)04-0171-10

Adaptive Decomposition Method for Impact Characteristic Frequency Bands of Multi-Constrained Robust Non-Negative Matrix Factorization

LIANG Lin¹, CUI Xujun², HU Wenhao², DAI Pumiao²

(1. School of Instrument Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China;

2. School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: In response to the issue of frequency band mixing caused by energy orientation in the decomposition of impact vibration frequency bands related to bearing faults using non-negative matrix factorization (NMF), a multi-constrained robust non-negative matrix factorization method for adaptive decomposition of impact fault bands is proposed. Firstly, an error function with adaptive weighted β -divergence is utilized to mitigate risks arising from unknown signal distributions, with the number of signal frequency bands serving as a reference for selecting the decomposition rank. Adaptive band division is achieved by imposing orthogonal constraints on the basis matrix. Secondly, considering the smooth frequency domain of the periodic impact response spectrogram and the sparse time domain characteristics, physically meaningful smooth and sparse constraints

are introduced to establish a multi-constrained robust non-negative matrix factorization model tailored to vibration time-frequency spectra. Finally, a solution algorithm is designed leveraging regularization techniques and Stiefel manifold optimization methods. Simulation and experimental results demonstrate that, compared to various NMF methods and typical band selection methods, when facing weak impacts caused by defects, the low-frequency range often harbors multiple sources of interference. The proposed decomposition model accurately extracts the impact response bands in the high-frequency range, circumventing the shortcomings of traditional energy-oriented band decomposition methods. The introduction of constraints effectively enhances the solution results of NMF, reinforcing its discriminative capability when dealing with weak impact characteristic frequency bands.

Keywords: impulse band; non-negative matrix decomposition; multiple constraints; band decomposition

装备的安全稳定运行需要准确可靠的特征提取和诊断技术,而在装备零部件局部故障,如点蚀、剥落等引起的冲击振动中,由于冲击激发的共振频带个数及位置未知,当冲击激励微弱、其他干扰源或背景噪声较大时,冲击特征的提取和分离较为困难^[1]。

以峭度谱^[2]为代表的自谱^[3]等频域方法,通常使用峭度、稀疏度、Gini 指标等评估指标进行冲击特征频带的选取,具有一定的效果,但由于选择的频带固定,仅依据指标容易产生误判。而考虑冲击特征的时频表示方法,如短时傅里叶变换、循环相干谱的频频表示方法^[4],虽然可以更全面地描述冲击响应信号的非平稳特性,但在面对由二维数据组成的矩阵时,进一步分析处理的方法仍有不足。

作为矩阵分析的一种手段,非负矩阵分解(NMF)方法基于加性组合来获得局部化的信号特征,具有较好的可解释性,已经在语音、图像处理等领域得到应用^[5-6]。在状态监测诊断中,Wodecki 等^[7]将 NMF 应用于时频谱的分解,并将基矩阵作为滤波器进行冲击频带的选择。王华庆等^[8]使用 IS(itakura-saito)散度结合行列式约束构造的 NMF 模型,对时频谱进行分解,并通过加权峰值因子进行重构信号筛选,实现了轴承复合故障分离。薛红涛等^[9]使用 IS 散度的稀疏 NMF,对时频矩阵进行分解,实现了轮毂电机轴承复合故障的特征提取。

为了对问题精准建模,对 NMF 的约束条件以及低秩表示的泛化模型被不断提出和应用,Liang 等^[10-11]将卷积 NMF 模型应用于轴承故障诊断,针对冲击故障特征的非线性结构,在稀疏非负矩阵分解的基础上加入核函数进行高维映射,提出了稀疏核非负矩阵分解算法。Yang 等^[12]提出采用判别非负矩阵分解提取柴油发动机振动信号时频谱中的故障特征。由时频谱构成的三阶张量分解被用于轴承

和齿轮故障诊断^[13-14],循环谱与 NMF 相结合的故障诊断方法相继被提出^[15-16]。

然而,作为低秩表示方法,NMF 方法主要依据其数值特性或对应最大似然估计的噪声分布经验地选取误差函数,即为了方便计算和理解,大多数 NMF 采用了欧式距离,但会导致以能量为导向的频带分解,从而造成了低幅值频带被忽略。其次,NMF 方法常以冲击源数量作为分解秩的参考,试图直接提取各冲击响应的特征信息,在多共振频带及噪声干扰下,其频带分离效果十分有限。

为此,本文通过基矩阵的正交约束实现频带的自适应划分,并引入 β 散度自适应加权的分布鲁棒模型,将其作为误差函数以缓解能量为导向的划分方法不足,同时基于周期冲击响应的时频表示特征,结合光滑和稀疏约束,设计了多约束下的分布鲁棒矩阵分解模型及其求解算法,有效增强了冲击故障特征的提取能力。

1 非负矩阵分解模型

低秩矩阵近似是提取高维数据中隐藏信息的常用方法,然而分解结果具有较大的不确定性。通过引入非负性约束得到的 NMF 模型不仅能够很好地缓解该问题,同时还具有非常好的可解释性,其数学表达如下

$$\mathbf{V} \approx \mathbf{WH} = \sum_i \mathbf{w}_i \mathbf{h}_i \quad (1)$$

式中: $\mathbf{V} \in \mathbb{R}^{m \times n}$ 为原始矩阵; $\mathbf{W} \in \mathbb{R}^{m \times r}$ 和 $\mathbf{H} \in \mathbb{R}^{r \times n}$ 分别为非负基矩阵和非负权矩阵; r 为分解秩,通常情况下, r 的选择需要满足 $(n+m)r < mn$ 。由于噪声等因素的干扰,原始矩阵只能做到近似低秩分解,则分解模型的目标函数为

$$J = D(\mathbf{V}, \mathbf{WH}) \quad (2)$$

式中: D 为误差函数,用以描述两矩阵的近似程度,通常采用 α 散度、 β 散度、 γ 散度等距离函数来描述。其中 β 散度包含欧式距离、KL 散度和 IS 散度等,可表示为

$$D_{\beta}(\mathbf{x}, \mathbf{y}) = \sum_i \frac{1}{\beta(\beta-1)} (x_i^{\beta} + (\beta-1)y_i^{\beta} - \beta x_i y_i^{\beta-1}), \beta \in \mathbb{R} \quad (3)$$

当 $\beta = 2$ 时, β 散度退化为欧式距离; 当 $\beta = 1$ 时, β 散度等价于 KL 散度; $\beta = 0$ 时, β 散度则等价于 IS 散度。不同 β 散度构建的 NMF 误差函数所提取特征不仅会受到样本中对应特征幅值影响,同时受到样本中特征大幅值的数量以及其余特征的影响。因此,在面对实际特征模式和信号中未知干扰的情况,误差函数的选择会对分解结果产生一定的影响。

为了防止模型过拟合和提高算法泛化性能,并引导问题向期望解的方向收敛。对 NMF 模型需要通过惩罚函数在原优化问题上引入额外约束,即使使用正则化方法实现,对应的正则化 NMF 目标函数可表示为

$$J(\mathbf{V}, \mathbf{WH}) = D(\mathbf{V}, \mathbf{WH}) + \alpha_W R_W(\mathbf{W}) + \alpha_H R_H(\mathbf{H}) \quad (4)$$

式中: $R(\mathbf{W})$ 、 $R(\mathbf{H})$ 为 $\mathbf{W}$ 和 $\mathbf{H}$ 的正则函数(罚函数); α_W 、 α_H 为 $\mathbf{W}$ 和 $\mathbf{H}$ 的正则化参数,用于权衡代表数据重构效果的误差项和反映约束效果的正则项,其值均不应小于 0。

典型的正则项有正交约束,光滑约束和稀疏约束等,其中正交约束条件可以表达为 $\mathbf{W}^T \mathbf{W} = \mathbf{I}_r$ [17]。而光滑约束主要有局部方差正则项 [18]、噪声频带惩罚正则项 [19]、全变分正则项 [20-22] 等,其中全变分正则项通过前项与后项之差序列的平方和构建,即

$$R(\mathbf{W}) = \frac{1}{2} \|\mathbf{\Gamma}_W \mathbf{W}\|_F^2 \quad (5)$$

式中: $\mathbf{\Gamma}_W$ 为一阶差分矩阵。在稀疏约束中,除了典型的 L_1 范数, L_q 范数可以诱导出更加稀疏的解,其表达式为

$$R(\mathbf{H}) = \|\mathbf{H}\|_q = \sum_{i,j} |\mathbf{H}(i,j)|^q, 0 < q \leq 1 \quad (6)$$

2 多约束联合的分布鲁棒 NMF 模型

2.1 面向冲击响应时频谱的约束设计

滚动轴承局部故障所引发的周期冲击振动信号具有明显的非平稳性,而其时频表达不仅很好地保留了冲击特征的时域和频域信息,同时具有很好的低秩特性。当 $\mathbf{V}$ 为包含周期冲击振动的信号时频谱

时,分解的基矩阵 $\mathbf{W}$ 对应周期冲击响应的频域分布,权矩阵 $\mathbf{H}$ 则是周期冲击响应的时域包络。对于振动响应来说,在频域中的高频段,冲击响应幅值往往较低,而低频段中的冲击响应幅值较高,但会包含其他源的干扰成分,不利于频带分离。

根据周期冲击响应的时频特性,在正则项中,可以在对应频带的 $\mathbf{W}$ 上施加正交约束,来避免同一频带被分解到不同子空间。在时间方向的 $\mathbf{H}$ 向量上,增加稀疏约束可有效抑制非冲击响应时段的噪声。

2.2 多约束分布鲁棒非负矩阵分解模型

由于不同误差函数对应着不同的样本统计分布,而振动信号的真实分布,尤其是故障状态下的分布往往未知,因此为了降低单一 β 值选择带来的风险,引入了包含 $\beta=0,1,2$ 的分布鲁棒非负矩阵分解模型 [23],通过对 3 种误差函数赋予权重,并在优化过程中自适应调整权重参数,使得 3 种散度均参与目标函数的优化,并且向误差函数添加约束正则项构成新的目标函数,其优化问题的数学表达为

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} [\max_{\beta \in \Omega} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH})] + \alpha_H R_H(\mathbf{H}) + \alpha_W R_W(\mathbf{W}) \quad (7)$$

式中: $\bar{D}_{\beta}$ 为归一化的误差函数。归一化会使得不同的 D_{β} 在优化过程中具有相同权重,避免了仅针对有较大误差评估的误差函数进行优化,通常以 $D(\mathbf{V}, \mathbf{WH})$ 对应的最小值进行归一化。

根据冲击响应时频特性和分解需求,对基矩阵 $\mathbf{W}$ 引入全变分光滑正则项,即 $R_W(\mathbf{W}) = \|\mathbf{\Gamma}_W \mathbf{W}\|_F^2$, 来保证分解的频谱连续,并对权矩阵 $\mathbf{H}$ 引入 $L_{1/2}$ 范数 [24] 和光滑约束项,即 $R_H(\mathbf{H}) = \|\mathbf{H}\|_{1/2} + \|\mathbf{H}^T \mathbf{\Gamma}_H\|_F^2$, 来保证周期冲击包络的光滑稀疏性。

对于正交约束来说,将满足正交约束条件 $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ 的 $\mathbf{W}$ 组成一个集合

$$S = \{\mathbf{W} \in \mathbb{R}^{m \times r} \mid r \leq m, \mathbf{W}^T \mathbf{W} = \mathbf{I}_r\} \quad (8)$$

相较于在欧式空间进行正交约束的优化求解,在 Stiefel 流形空间中进行优化可以得到更好的正交性结果和更快的收敛速度。

将解空间限定在 Stiefel 流形上,进行优化求解,则可得到具有光滑、正交和稀疏约束的约束分布鲁棒非负矩阵分解(SmODRNMFLq),表达式为

$$\min_{\mathbf{W}, \mathbf{H} \geq 0, \mathbf{W} \in S} \left\{ \begin{aligned} & [\max_{\beta \in \Omega} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH})] + \frac{\alpha_{SmW}}{2} \|\mathbf{\Gamma}_W \mathbf{W}\|_F^2 + \\ & \alpha_{sp} \|\mathbf{H}\|_{1/2} + \frac{\alpha_{SmH}}{2} \|\mathbf{H}^T \mathbf{\Gamma}_H\|_F^2 \end{aligned} \right\} \quad (9)$$

式(9)可以通过最小化不同权重的多个目标函数进行求解,即

$$\min_{\mathbf{W}, \mathbf{H} \geq 0, \mathbf{W} \in \mathbf{S}} \left\{ \sum_{\beta \in \Omega} \lambda_{\beta} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) + \frac{\alpha_{\text{Sm}}}{2} \|\mathbf{F}_W \mathbf{W}\|_{\text{F}}^2 + \alpha_{\text{sp}} \|\mathbf{H}\|_{\frac{1}{2}} + \frac{\alpha_{\text{SmH}}}{2} \|\mathbf{H}^T \mathbf{F}_H \mathbf{H}\|_{\text{F}}^2 \right\} \quad (10)$$

上述优化目标关于矩阵 $\mathbf{H}$ 的梯度可表示为

$$\nabla_{\mathbf{H}} J = \sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{H}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) + \frac{1}{2} \alpha_{\text{sp}} \mathbf{H}^{-\frac{1}{2}} + \alpha_{\text{SmH}} (\mathbf{H}^T \mathbf{F}_H \mathbf{F}_H^T \mathbf{H}) \quad (11)$$

式中

$$\mathbf{H}^T \mathbf{F}_H \mathbf{F}_H^T \mathbf{H} = (\mathbf{H}_3 + \mathbf{H}) - (\mathbf{H}_1 + \mathbf{H}_2) \quad (12)$$

$$\nabla_{\mathbf{H}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) = \mathbf{W}^T (\mathbf{WH})^{(\beta-1)} - \mathbf{W}^T ((\mathbf{WH})^{(\beta-2)} \cdot \mathbf{V}) \quad (13)$$

$$\mathbf{H}_1 = \begin{bmatrix} h_{1,2} & h_{1,3} & \cdots & h_{1,n} & 0 \\ h_{2,2} & h_{2,3} & \cdots & h_{2,n} & 0 \\ \vdots & \vdots & & \vdots & 0 \\ h_{r,2} & h_{r,3} & \cdots & h_{r,n} & 0 \end{bmatrix} \quad (14)$$

$$\mathbf{H}_2 = \begin{bmatrix} 0 & h_{1,1} & h_{1,2} & \cdots & h_{1,n-1} \\ 0 & h_{2,1} & h_{2,2} & \cdots & h_{2,n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & h_{r,1} & h_{r,2} & \cdots & h_{r,n-1} \end{bmatrix} \quad (15)$$

$$\mathbf{H}_3 = \begin{bmatrix} 0 & h_{1,2} & \cdots & h_{1,n-1} & 0 \\ 0 & h_{2,2} & \cdots & h_{2,n-1} & 0 \\ \vdots & \vdots & & \vdots & 0 \\ 0 & h_{r,2} & \cdots & h_{r,n-1} & 0 \end{bmatrix} \quad (16)$$

将式(11)关于矩阵 $\mathbf{H}$ 的梯度拆分为正梯度之差的形式

$$\nabla_{\mathbf{H}} J = \sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{H}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) + \frac{1}{2} \alpha_{\text{sp}} \mathbf{H}^{-\frac{1}{2}} + \alpha_{\text{SmH}} (\mathbf{H} + \mathbf{H}_3) \quad (17)$$

$$\nabla_{\mathbf{H}} J = \sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{H}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) + \alpha_{\text{SmH}} (\mathbf{H}_1 + \mathbf{H}_2) \quad (18)$$

根据 Stiefel 流形在点 $\mathbf{X}$ 处的黎曼梯度^[25], 矩阵 $\mathbf{W}$ 的黎曼梯度为

$$\nabla_{\mathbf{W}} J = \sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{W}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) - \mathbf{W} \left(\sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{W}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) \right)^T \mathbf{W} + \alpha_{\text{SmW}} \mathbf{F}_W^T \mathbf{F}_W \mathbf{W} - \alpha_{\text{SmW}} \mathbf{W} \mathbf{W}^T \mathbf{F}_W^T \mathbf{F}_W \mathbf{W} \quad (19)$$

式中

$$\mathbf{F}_W^T \mathbf{F}_W \mathbf{W} = (\mathbf{W} + \mathbf{W}_3) - (\mathbf{W}_1 + \mathbf{W}_2) \quad (20)$$

$$\nabla_{\mathbf{W}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) = (\mathbf{WH})^{(\beta-1)} \mathbf{H}^T - ((\mathbf{WH})^{(\beta-2)} \cdot \mathbf{V}) \mathbf{H}^T \quad (21)$$

按照 $\mathbf{H}_1$ 、 $\mathbf{H}_2$ 和 $\mathbf{H}_3$ 的构造方法, 同理可得 $\mathbf{W}_1$ 、 $\mathbf{W}_2$ 和 $\mathbf{W}_3$ 。将式(19)拆分为正项之差的形式

$$\nabla_{\mathbf{W}} J = \alpha_{\text{SmW}} (\mathbf{W}_3 + \mathbf{W}) + \sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{W}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) + \mathbf{W} \left(\sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{W}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) + \alpha_{\text{SmW}} (\mathbf{W}_1 + \mathbf{W}_2) \right)^T \mathbf{W} \quad (22)$$

$$\nabla_{\mathbf{W}} J = \alpha_{\text{SmW}} (\mathbf{W}_1 + \mathbf{W}_2) + \sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{W}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) + \mathbf{W} \left(\sum_{\beta \in \Omega} \lambda_{\beta} \nabla_{\mathbf{W}} \bar{D}_{\beta}(\mathbf{V}, \mathbf{WH}) + \alpha_{\text{SmW}} (\mathbf{W}_3 + \mathbf{W}) \right)^T \mathbf{W} \quad (23)$$

通过缩放梯度法, 可以得到在 λ 固定时矩阵 $\mathbf{W}$ 、 $\mathbf{H}$ 的更新公式, 而矩阵 $\mathbf{W}$ 、 $\mathbf{H}$ 固定时 λ 的更新公式可使用 Frank-Wolfe 方法实现。在对矩阵 $\mathbf{W}$ 、 $\mathbf{H}$ 进行块坐标下降前, 需要分别进行归一化以满足不同约束项的归一化条件, 求解算法步骤如下。

输入: $\mathbf{V} \in \mathbb{R}_+^{r \times n}$, 分解秩 r , 有限数集 Ω , 光滑正则化参数 α_{Sm} , 稀疏正则化参数 α_{sp} , 迭代次数 m ;

初始化: $\mathbf{W}_0 \in \mathbb{R}_{++}^{m \times r}$, $\mathbf{H}_0 \in \mathbb{R}_{++}^{n \times r}$, $\lambda_{\beta} = 1/|\Omega|$;

对 $k = 1, 2, \dots, m$ 执行迭代

$$\mathbf{W}_k = \max \left\{ \epsilon, \mathbf{W}_{k-1} \cdot \left(\frac{[\nabla_{\mathbf{W}_{k-1}} J]}{[\nabla_{\mathbf{W}_{k-1}} J + \epsilon]} \right) \right\}$$

归一化 $\mathbf{H}_k$;

$$\mathbf{H}_k = \max \left\{ \epsilon, \mathbf{H}_{k-1} \cdot \left(\frac{[\nabla_{\mathbf{H}_{k-1}} J]}{[\nabla_{\mathbf{H}_{k-1}} J + \epsilon]} \right) \right\}$$

归一化 $\mathbf{W}$, 使得 $\|\mathbf{W}(:, j)\|_{\text{F}}^2 = 1, \forall j = 1, 2, \dots, r$

找到在所有 $\bar{D}_{\beta}$ 中, 使得 $\bar{D}_{\beta^*}(\mathbf{V} | \mathbf{W}_k \mathbf{H}_k)$ 最大的 β^* , 为 $\lambda_{k,\beta}^*$ 赋值

$$\lambda_{k,\beta}^* = \begin{cases} 0, & \beta = \beta^* \\ 1, & \beta \neq \beta^* \end{cases}$$

定义步长 $\nu_k = 1/(k+1)$, 对各 $\beta \in \Omega$ 更新 $\lambda_{k,\beta}$

$$\lambda_{k,\beta} = \lambda_{k-1,\beta} + \nu_k (\lambda_{k,\beta}^* - \lambda_{k-1,\beta});$$

直至迭代收敛或达到最大迭代值, 循环结束;

输出: SmODRNMFLq 的近似解 $\mathbf{W}_k$ 、 $\mathbf{H}_k$ 。

对于 NMF 的参数分解秩 r , 多数 NMF 方法以源数量作为其选择的依据, 因此数值较小。但在复杂的振动加速度频谱中, 较少的分解秩 r 会使能量微弱的频带和其他频带合并在同一子空间中, 从而极大降低了包络谱的效果。

为此, 对于本文提出的 SmODRNMFLq 模型,

选取较大的分解秩 r ,在基矩阵的频谱光滑性约束下,可以通过频带的充分分解,使得弱能量的频带被有效分解出来。在基向量和权向量组成的系列子空间中,通过每个子空间的逆变换来获得相应频带的信号,并依据希尔伯特变换解调出包络成分,组成子空间包络谱的分布。根据各子空间重构波形的包络谱,进行周期冲击特征的识别与故障诊断分析。

3 仿真分析

为了展示本文所提方法的优越性,首先通过数值仿真来进行分析与验证。参照滚动轴承振动模型,模拟的局部故障振动信号如下

$$x(t) = u(t) + v(t) + n(t) \tag{24}$$

式中: $u(t)$ 为局部故障激发的周期性瞬态冲击成分; $v(t)$ 为随机冲击分量; $n(t)$ 为高斯噪声。其中故障周期 $T_F = 0.02\text{ s}$,振幅为 4,故障冲击激发了 3 个共振频带,其中心频率分别为 3、6.5、7 kHz,冲击衰减的阻尼取 0.01。其次,考虑到非高斯噪声成分,模拟了随机冲击分量 $v(t)$,其包含了 2 个随机冲击,分别位于 0.3、0.8 s 时,其共振频带分别为 2 500、6 750 Hz,且峰值能量之比为 2 : 1,对信号施加了高斯噪声 $n(t)$ 。

模拟信号的时域波形、频谱和时频谱如图 1 所示,由图 1(b)可知,周期冲击引发的 3 个共振带能量递减,低频带的 3 kHz 频带受到 2.5 kHz 随机冲击频带的干扰,而高频区的 6.5、7 kHz 能量较小,受到另一个随机冲击的影响。图 1(c)中,周期冲击序列对应的频带和 2 个随机冲击成分的频带存在一定的干扰。

为了对比分析,除本文所提稀疏光滑正交鲁棒模型 SmODRNMFLq,还选择了标准 NMF 模型、基于欧式距离的正交模型 OEDNMF、以及光滑正交模型 SmODRNMF,其中 SmODRNMF 模型在分布鲁棒非负矩阵分解模型的基础上施加了光滑和正交约束。考虑到模拟的冲击频带数,时频谱的分解秩 r 设为 10,4 种方法的分解结果如图 2~5 所示,为了使得各分离基向量和权向量清晰,对基矩阵和权矩阵进行了独立的归一化。

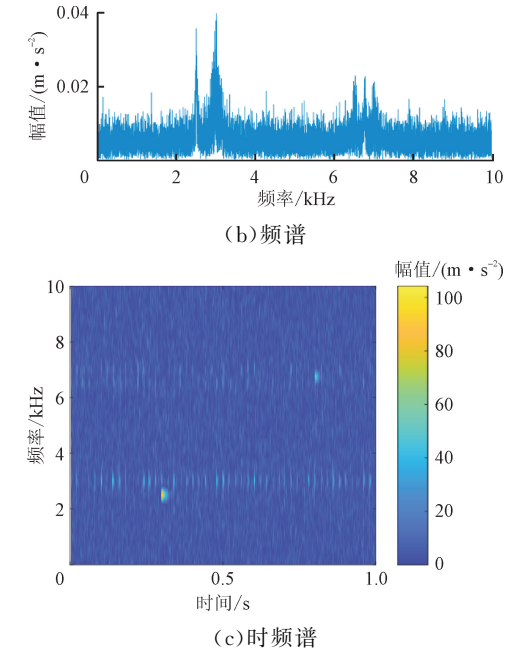


图 1 周期性冲击振动响应的仿真信号
Fig. 1 Simulated signals of periodic impact vibration response

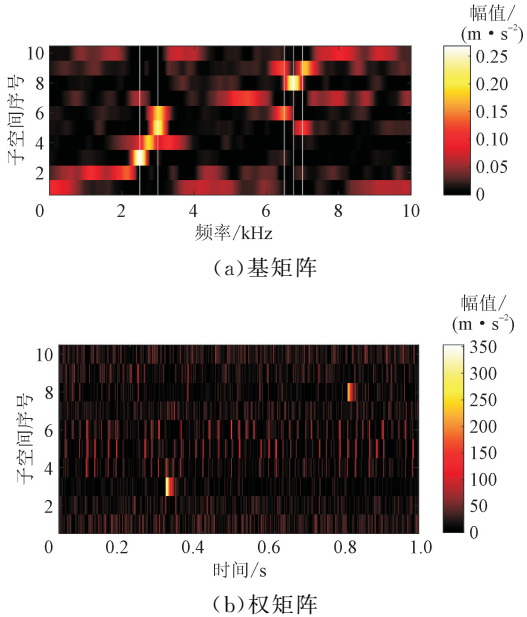
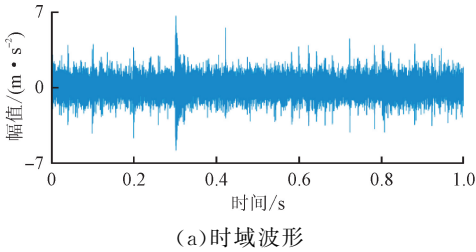
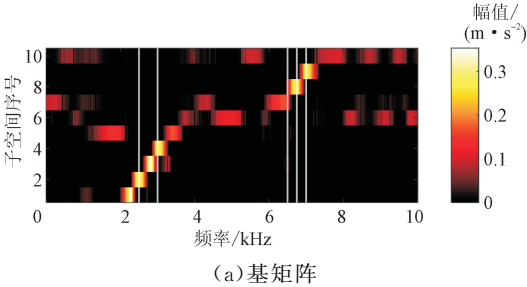
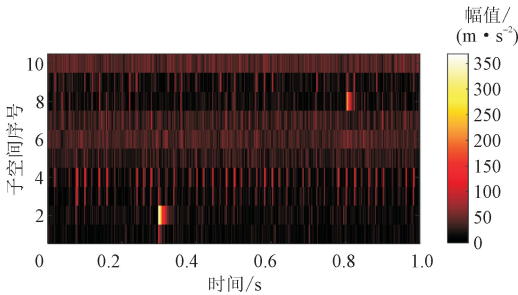


图 2 标准 NMF 的分解结果
Fig. 2 Decomposition result for standard NMF

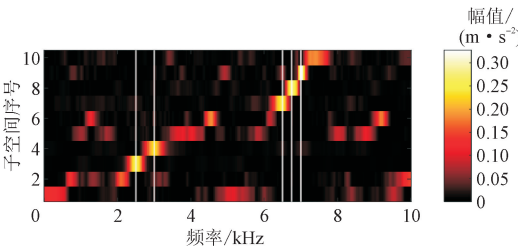




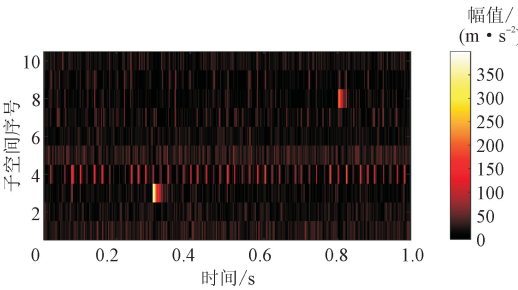
(b) 权矩阵

图 3 OEDNMF 的分解结果

Fig. 3 Decomposition result for OEDNMF



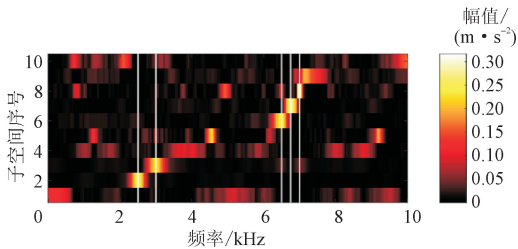
(a) 基矩阵



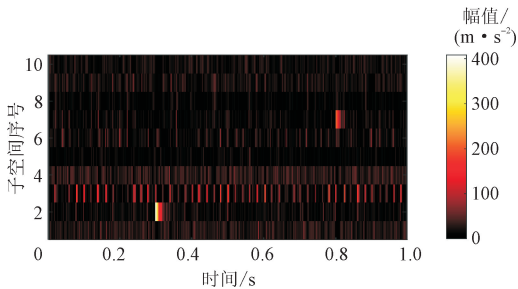
(b) 权矩阵

图 4 SmODRNMFLq 的分解结果

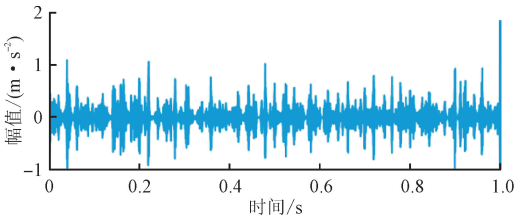
Fig. 4 Decomposition result for SmODRNMFLq



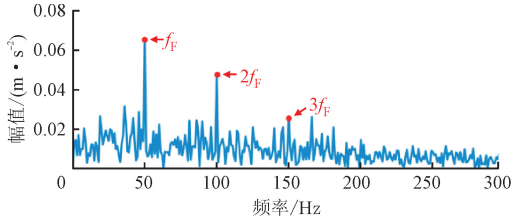
(a) 基矩阵



(b) 权矩阵



(c) 第 6 子空间重构波形



(d) 第 6 子空间重构波形包络谱

图 5 SmODRNMFLq 的分解结果

Fig. 5 Decomposition result for SmODRNMFLq

从图 2(a)可知,标准 NMF 的分解结果中,能量较大的低频区占有多个子空间,相同的故障冲击频带被分离到相近的子空间中,3 kHz 频带成分被同时分到第 4、5 和 6 子空间中。由于没有正交约束的限制,高频区的冲击共振带也被分解在不同的空间中。

从 OEDONMF 模型分解结果可见,较标准 NMF 的结果有一定提升。由于分解模型中施加了正交约束项,分解得到的各基向量近似正交,频带得以独立分解。但基于欧式距离的误差函数具有能量导向的选择特性,在分解秩较多时,能量较大的 3 kHz 频带会被拆分到多个子空间中,导致明显不光滑现象,且在高频部分出现欠分解现象,6.5 kHz 的部分频带被分解至其他频带中。

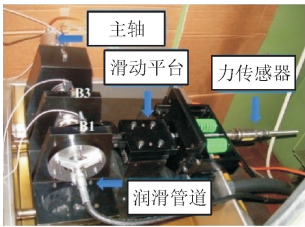
对于 SmODRNMFLq 模型,在分布鲁棒误差函数的影响下,冲击故障频带和随机冲击频带有效分离,同时获得较好的正交效果。SmODRNMFLq 模型分解得到的权向量较 SmODRNMFLq 模型有明显的稀疏性提升,从而准确地提取出了故障冲击序列的时域特征和频率特征,第 6 子空间重构波形包络谱,准确提取出了故障特征频率。

本文所提模型在正交约束的基础上,利用了光滑和稀疏约束,使得频带分解时充分考虑了高频区域的信息,对幅值微弱的频带具有很好的辨识和分离能力。

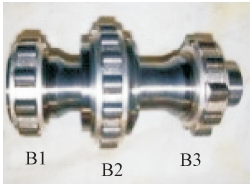
4 实例应用

本节使用轴承局部故障的振动数据进行应用验证,所选数据来源于文献[26]。DIRG 轴承实验台如

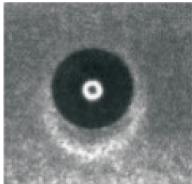
图 6 所示,实验台主要由高速主轴、空心转轴、滚动轴承、滑动平台等组成,滑动平台的移动方向与轴相互垂直,通过静力传感器测量径向力。实验过程中,采样率为 51 200 Hz,采样时长为 10 s。图 6(b)中 3 个滚动轴承 B1、B2 和 B3 内圈固定于空心转轴上,且 B1、B3 轴承的外圈固定在底座上,用于支承空心转轴,B2 轴承外圈与滑动平台连接,用于施加径向载荷。



(a)实验台整体结构



(b)轴承结构



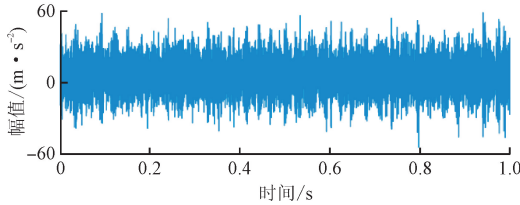
(c)模拟的内圈故障

图 6 DIRG 轴承实验台

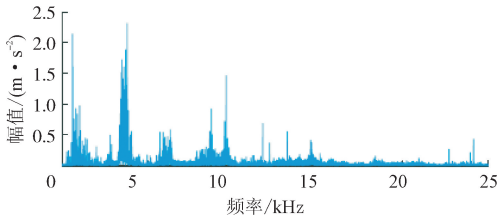
Fig. 6 DIRG bearing test bench schematic

现选择内圈故障数据(编号“C1A_100_502_2”)中的振动信号进行分析,轴承 B1 内圈上的圆锥形压痕直径近似为 450 μm 。根据轴承结构参数可知内圈故障 BPF1 的阶次为 6.112,高速主轴实际转速为 88.9 Hz。

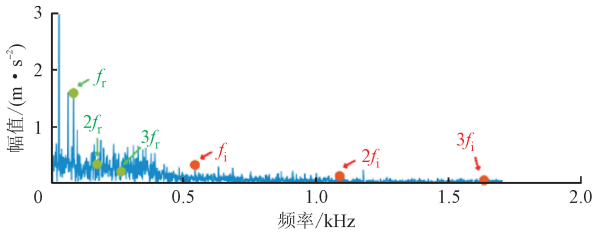
信号波形、频谱、包络谱以及时频谱如图 7 所示,可知频谱中包含了较多共振频带,其中低频段的能量较大,而高频段(1 kHz 以上)的振动能量普遍较低,在图 7(c)的包络谱中,除了转频 f_r 及谐波分量外,几乎无法观察到内圈故障的特征频率分量 f_i 及其谐波。



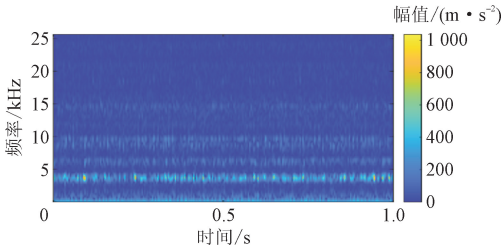
(a)时域波形



(b)频谱



(c)包络谱

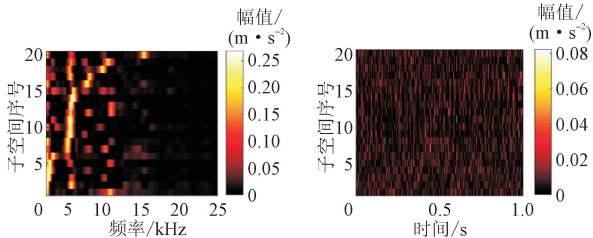


(d)时频谱

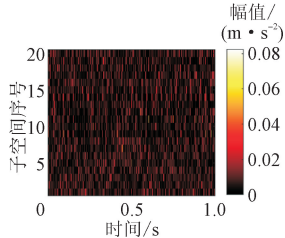
图 7 轴承振动信号的波形及频谱

Fig. 7 Waveforms and spectra of bearing vibration signals

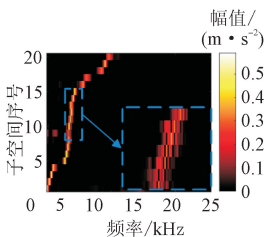
对时频谱阵的分解中,为了保证正交频带的完全分离,分解秩设定为 20,图 8 给出了标准 NMF、OEDNMF 以及本文所提 SmODRNMFLq 模型分解结果,可知标准 NMF 分离得到的基向量几乎都包含有 4 kHz 左右的低频共振峰,并带有其他频带信息,而不同基向量对应的权向量的结果不尽相同,这说明在无约束情况下,标准 NMF 方法无法实现频带的准确划分。图 8(c)中各基向量近似正交,然而在分解中,仍有较多基向量对应特征集中在能量较大的 4 kHz 附近,其放大图中的各基矩阵出现了明显的不光滑情形,这不符合冲击频响函数的频谱表现。图 8(e)中对应划分的基向量在高频和低频分布较为均匀,且在光滑约束下各基向量光滑,由此验证了所提方法进行自适应频带划分的合理性。



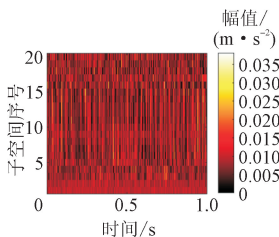
(a)标准 NMF 的基矩阵



(b)标准 NMF 的权矩阵



(c)OEDNMF 的基矩阵



(d)OEDNMF 的权矩阵

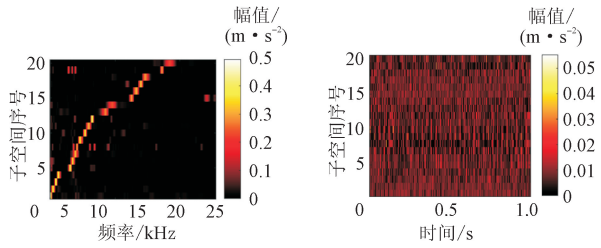


图 8 不同 NMF 模型的分解结果

Fig. 8 Decomposition result for different NMF models

SmODRNMFLq 模型的分解结果如图 9 所示,对分解的 20 个子空间进行重构可得相应频带下的波形,且各频带波形的包络谱如图 9(a)所示。根据实验台运行的转频和内圈故障特征 BPFI,高频段的第 14 子空间中,含有明显的内圈故障特征频率及其谐波。为了显示方便,该频带下的重构波形及包络谱如图 9(b)、9(c)所示,其中内圈故障的特征频率 f_i 及其倍频清晰可见,故障特征的提取效果明显。

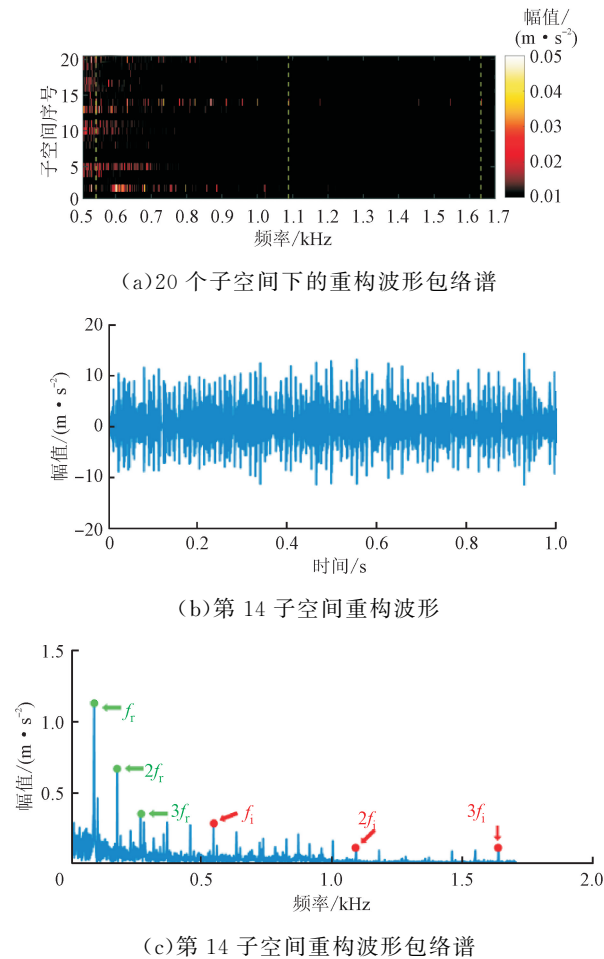


图 9 SmODRNMFLq 模型的分解结果

Fig. 9 Decomposition result for SmODRNMFLq

为了对比分析,本文选择了 Autogram 方法进行最优频带选择与冲击特征提取,其中频带划分等级为 5,计算结果如图 10 所示。在图 10(a)所示的各频带包络相关峭度中,第 4 级中的第 12 个频带是最优频带(图中红色方框所示),其位置与本文所提方法的第 20 子空间频带接近,该频带对应的滤波信号及包络谱如图 10(b)、10(c)所示,可知波形中存在着明显的转频成分,且包络谱中的背景噪声较大,内圈故障特征频率 f_i 非常微弱。由于转频及其谐波成分的误导,使得 Autogram 方法没有选择到更好的频带。

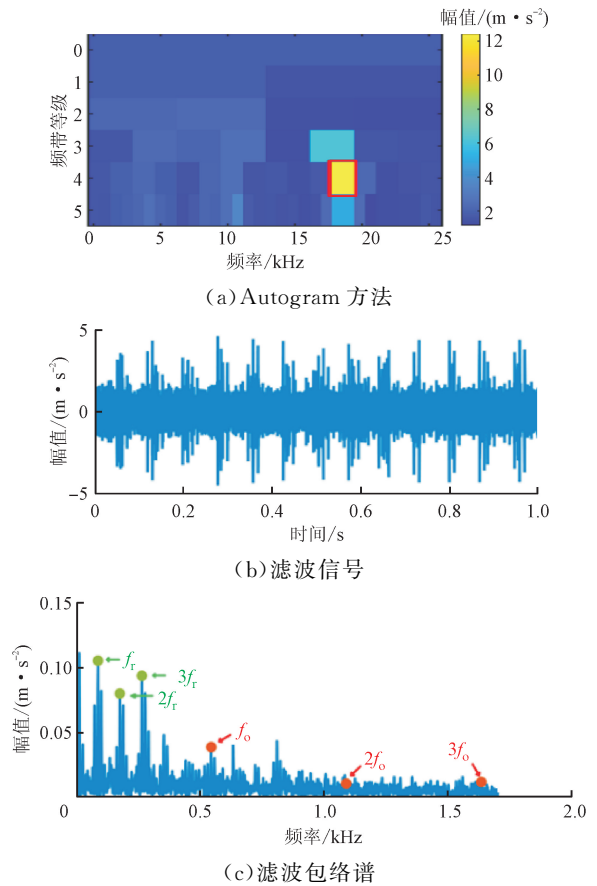


图 10 Autogram 方法的频带选择结果

Fig. 10 Selection result for Autogram

5 结 论

结合约束项和误差函数对 NMF 模型求解结果的影响以及冲击振动信号的时频分布特性,本文提出了一种面向冲击频带自适应分解的多约束鲁棒非负矩阵分解模型与算法,得到如下结论。

(1) 结合分布鲁棒模型误差函数的构建方式,利用故障周期冲击响应的时频特征引入光滑和稀疏约束,并与正交约束结合,建立了多约束鲁棒非负矩阵

分解模型,借助正则化技术和 Stiefel 流形优化方法推导出了优化求解算法。

(2)所提模型具有良好的物理意义,能够获得同时具有基矩阵正交光滑、权矩阵光滑稀疏,且对频带特征幅值大小不敏感的效果,从而减少频带混叠,实现冲击多频带的有效分离。

(3)通过仿真数据和轴承故障实验,验证了所提方法在冲击频带分解中的有效性,将该模型与自谱方法对比,验证了本文方法在冲击故障特征提取中的优越性。

参考文献:

- [1] SMITH W A, BORGHESEANI P, NI Qing, et al. Optimal demodulation-band selection for envelope-based diagnostics: a comparative study of traditional and novel tools [J]. *Mechanical Systems and Signal Processing*, 2019, 134: 106303.
- [2] ANTONI J. Fast computation of the kurtogram for the detection of transient faults [J]. *Mechanical Systems and Signal Processing*, 2007, 21(1): 108-124.
- [3] MOSHREFZADEH A, FASANA A. The autogram: an effective approach for selecting the optimal demodulation band in rolling element bearings diagnosis [J]. *Mechanical Systems and Signal Processing*, 2018, 105: 294-318.
- [4] MAURICIO A, SMITH W A, RANDALL R B, et al. Improved envelope spectrum via feature optimisation-gram (IESFOgram): a novel tool for rolling element bearing diagnostics under non-stationary operating conditions [J]. *Mechanical Systems and Signal Processing*, 2020, 144: 106891.
- [5] SUN Jing, WANG Zhihui, SUN Fuming, et al. Sparse dual graph-regularized NMF for image co-clustering [J]. *Neurocomputing*, 2018, 316: 156-165.
- [6] LEPLAT V, GILLIS N, ANG A M S. Blind audio source separation with minimum-volume beta-divergence NMF [J]. *IEEE Transactions on Signal Processing*, 2020, 68: 3400-3410.
- [7] WODECKI J, KRUCZEK P, BARTKOWIAK A, et al. Novel method of informative frequency band selection for vibration signal using nonnegative matrix factorization of spectrogram matrix [J]. *Mechanical Systems and Signal Processing*, 2019, 130: 585-596.
- [8] 王华庆,王梦阳,宋浏阳,等. 双约束非负矩阵分解的复合故障信号分离方法 [J]. *振动工程学报*, 2020, 33(3): 590-596.

et al. Method of compound fault signal separation using double constraints non-negative matrix factorization [J]. *Journal of Vibration Engineering*, 2020, 33(3): 590-596.

- [9] 薛红涛,丁殿勇,李沛铖,等. 基于分量加权重构和稀疏 NMF 的轮毂电机轴承复合故障特征提取方法 [J]. *机械工程学报*, 2023, 59(9): 146-156.
- XUE Hongtao, DING Dianyong, LI Ruicheng, et al. Feature extraction method based on component weighted reconstruction and sparse NMF for bearing compound faults of in-wheel motor [J]. *Journal of Mechanical Engineering*, 2023, 59(9): 146-156.
- [10] LIANG Lin, SHAN Lei, LIU Fei, et al. Impulse feature extraction of bearing faults based on convolutive nonnegative matrix factorization [J]. *IEEE Access*, 2020, 8: 88617-88632.
- [11] LIANG Lin, DING Xingyun, LIU Fei, et al. Feature extraction using sparse kernel non-negative matrix factorization for rolling element bearing diagnosis [J]. *Sensors*, 2021, 21(11): 3680.
- [12] YANG Yongsheng, MING Anbo, ZHANG Youyun, et al. Discriminative non-negative matrix factorization (DNMF) and its application to the fault diagnosis of diesel engine [J]. *Mechanical Systems and Signal Processing*, 2017, 95: 158-171.
- [13] GABOR M, ZDUNEK R, ZIMROZ R, et al. Non-negative tensor factorization for vibration-based local damage detection [J]. *Mechanical Systems and Signal Processing*, 2023, 198: 110430.
- [14] LIANG Lin, WEN Haobin, LIU Fei, et al. Feature extraction of impulse faults for vibration signals based on sparse non-negative tensor factorization [J]. *Applied Sciences*, 2019, 9(18): 3642.
- [15] LEE J H. Enhancement of decomposed spectral coherence using sparse nonnegative matrix factorization [J]. *Mechanical Systems and Signal Processing*, 2021, 157: 107747.
- [16] LIANG Lin, DING Xingyun, WEN Haobin, et al. Impulsive components separation using minimum-determinant KL-divergence NMF of bi-variable map for bearing diagnosis [J]. *Mechanical Systems and Signal Processing*, 2022, 175: 109129.
- [17] DING C, LI Tao, PENG Wei, et al. Orthogonal non-negativematrix t-factorizations for clustering [C]// *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2006: 126-135.

(下转第 192 页)

- dynamic loads on spur gear teeth and on bearings [J]. *Journal of Sound and Vibration*, 2003, 267(5): 1065-1084.
- [26] CHANG Lehao, LIU Geng, WU Liyan. A robust model for determining the mesh stiffness of cylindrical gears [J]. *Mechanism and Machine Theory*, 2015, 87: 93-114.
- [27] 丁长安, 张雷, 周福章, 等. 线接触弹性接触变形的解析算法 [J]. *摩擦学学报*, 2001, 21(2): 135-138.
DING Chang'an, ZHANG Lei, ZHOU Fuzhang, et al. Theoretical formula for calculation of line-contact elastic contact deformation [J]. *Tribology*, 2001, 21(2): 135-138.
- [28] FERNANDEZ DEL RINCON A, VIADERO F, IGLESIAS M, et al. A model for the study of meshing stiffness in spur gear transmissions [J]. *Mechanism and Machine Theory*, 2013, 61: 30-58.
- [29] 吕宏展, 王羽达, 王君政. 含安装误差的鼓形齿轮承载接触分析 [J]. *机械工程学报*, 2022, 58(21): 126-133.
LÜ Hongzhan, WANG Yuda, WANG Junzheng. Loaded tooth contact analysis of spur gears with longitudinal modification considering assembly errors [J]. *Journal of Mechanical Engineering*, 2022, 58(21): 126-133.

(编辑 李慧敏)

(上接第 179 页)

- [18] CHEN Z, CICHOCKI A. Nonnegative matrix factorization with temporal smoothness and/or spatial decorrelation constraints [J]. *Signal Processing*, 2005, 68: 1-6.
- [19] SALEHANI Y E, GAZOR S. Smooth and sparse regularization for NMF hyperspectral unmixing [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(8): 3677-3692.
- [20] ESSID S, FÉVOTTE C. Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring [J]. *IEEE Transactions on Multimedia*, 2013, 15(2): 415-425.
- [21] VIRTANEN T. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria [J]. *IEEE Transactions on Audio Speech and Language Processing*, 2007, 15(3): 1066-1074.
- [22] FABREGAT R, PUSTELNIK N, GONÇALVES P, et al. Solving NMF with smoothness and sparsity constraints using PALM [EB/OL]. (2021-03-18) [2024-10-10]. <https://arxiv.org/abs/1910.14576>.
- [23] GILLIS N, HIEN L T K, LEPLAT V, et al. Distributionally robust and MULTI-OBJECTIVE nonnegative matrix factorization [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(8): 4052-4064.
- [24] QIAN Yuntao, JIA Sen, ZHOU Jun, et al. Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 49(11): 4282-4297.
- [25] YOO J H, CHOI S J. Nonnegative matrix factorization with orthogonality constraints [J]. *Journal of Computing Science and Engineering*, 2010, 4(2): 97-109.
- [26] DAGA A P, FASANA A, MARCHESIELLO S, et al. The Politecnico di Torino rolling bearing test rig: description and analysis of open access data [J]. *Mechanical Systems and Signal Processing*, 2019, 120: 252-273.

(编辑 赵炜)