

DOI: 10.11830/ISSN.1000-5013.201708005



采用 PARAFAC 的欠定盲分离中 机械振源数估计方法

杨 诚, 李志农

(南昌航空大学 无损检测技术教育部重点实验室, 江西 南昌 330063)

摘要: 针对复杂机械系统振源数未知的欠定盲源分离(UBSS)问题,为提高欠定盲源分离的性能,提出一种基于平行因子分析(PARAFAC)和核一致诊断(CORCONDIA)的欠定盲源数估计算法.该算法利用二阶非平稳源分离的基本思想,将中心化传感器数据分成不重叠的数据块,计算各数据块的单一时延协方差矩阵并叠加成三阶张量,即平行因子模型.利用核一致诊断算法估计 PARAFAC 模型的最佳组分数,从而得到机械系统的振源数.仿真实验结果表明:该算法可从非平稳欠定混合信号中准确估计振源数目.将所提算法应用于多机振动源实验,结果进一步验证了该方法的有效性.

关键词: 振源估计; 平行因子分析; 核一致诊断; 欠定混合; 盲源分离

中图分类号: TN 911.6 **文献标志码:** A **文章编号:** 1000-5013(2018)03-0337-06

Mechanical Vibration Source Number Estimation of Underdetermined Blind Source Separation Using PARAFAC

YANG Cheng, LI Zhinong

(Key Laboratory of Nondestructive Testing of Ministry of Education,
Nanchang Hangkong University, Nanchang 330063, China)

Abstract: Considering the existing problem of underdetermined blind separation (UBSS) in the complex mechanical system with unknown number of vibration sources, a new estimation algorithm of source number in the UBSS method based on parallel factor analysis (PARAFAC) and the core consistency diagnostic (CORCONDIA) is proposed to improve the performance of UBSS method. The main idea of this algorithm is that the centralized sensor data are firstly divided into some non-overlapping data blocks. Then single time-delay covariance matrices of each data block are calculated and stacked into a third-order tensor, which is constructed into the PARAFAC model. CORCONDIA is used to estimate the optimal number of components in PARAFAC model. Thus the obtained number of components is the number of vibration sources. The simulation results show that the proposed algorithm can accurately estimates the number of vibration sources from the underdetermined mixtures of non-stationary signal. It has been successfully applied to the test of multi-source mechanical vibration, and the experiment results further verify the effectiveness of the proposed algorithm.

收稿日期: 2017-08-08

通信作者: 李志农(1966-),男,教授,博士,主要从事智能检测与信号处理、机械设备状态检测与故障诊断的研究. E-mail: lizhinong@tsinghua. org. cn.

基金项目: 国家自然科学基金资助项目(51675258, 51261024, 51075372); 机械传动国家重点实验室开放基金资助项目(SKLMT-KFKT-201514); 南昌航空大学研究生创新专项基金资助项目(YC2016050)

Keywords: source number estimation; parallel factor analysis; core consistency diagnostic; underdetermined mixture; blind source separation

随着机械系统结构越来越复杂,精密程度越来越高,机械故障诊断基础研究亟待由从前的零件故障研究向机械系统多故障研究突破^[1].盲源分离(BSS)作为一种新颖的信号分析与处理工具,被众多学者应用于机械系统多故障复合诊断领域,并取得了一定的研究成果.在实际工程中,复杂机械系统的振源数往往是未知的,盲源分离难以精确有效地从观测信号中分离出所有源信息,这严重限制了盲源分离在机械多故障诊断中的应用.因此,准确估计振源数对盲源分离在机械系统多故障研究中的进一步应用具有重要意义.在过去十几年的研究中,国内外专家学者研究了未知信源条件下源数估计方法^[2-4].然而,在机械振动信号盲源分离中,对于观测信号数小于源信号数情况下的源数估计问题并没有考虑.近年来,平行因子分析(PARAFAC)因其在宽松约束条件下模型分解具有唯一性而备受关注,众多学者对 PARAFAC 模型分解唯一性进行了研究,现已广泛应用于环境与资源利用^[5]、化学计量^[6]、通信与信号处理^[7]等领域.其中,值得关注的是在化学计量领域中,学者们利用平行因子分析对未知干扰系统进行组分定量分析时,一般需要利用核一致诊断法先对系统组分数进行预估并得到最佳组分数,从而使 PARAFAC 模型分解得到有实际意义的解.就源信号与传感器的相对位置和信号的传播速度而言,文中研究的机电系统可忽略传感器到不同振源的延时差.基于此,本文针对线性瞬时混合模型,提出一种基于平行因子分析和核一致诊断的欠定盲源数估计算法.

1 平行因子分析的理论基础

平行因子模型^[8]是一种多维矩阵低秩分解模型,又称为三线形模型,其模型如图 1 所示.

三维矩阵 $\mathbf{X} \in \mathbb{C}^{I \times J \times M}$ 的 PARAFAC 分解标量形式可表示为

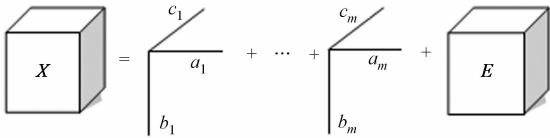


图 1 平行因子模型
Fig. 1 Parallel factor model

$$X_{i,j,n} = \sum_{m=1}^M A_{i,m} B_{j,m} C_{n,m} + E_{i,j,n}. \tag{1}$$

式(1)中: $i=1,2,\cdots,I; j=1,2,\cdots,J; n=1,2,\cdots,N; m=1,2,\cdots,M; \mathbf{A} \in \mathbb{C}^{I \times M}, \mathbf{B} \in \mathbb{C}^{J \times M}$ 和 $\mathbf{C} \in \mathbb{C}^{N \times M}$ 分别为该模型的 3 个承载矩阵; $\mathbf{E} \in \mathbb{C}^{I \times J \times M}$ 为噪声矩阵.

三线形模型亦可写成切片形式,即

$$\left. \begin{aligned} \mathbf{X}_{i::} &= \mathbf{B} \mathbf{D}_i(\mathbf{A}) \mathbf{C}^T + \mathbf{E}_{i::}, \\ \mathbf{X}_{:,j:} &= \mathbf{A} \mathbf{D}_j(\mathbf{B}) \mathbf{C}^T + \mathbf{E}_{:,j:}, \\ \mathbf{X}_{::,n} &= \mathbf{A} \mathbf{D}_n(\mathbf{C}) \mathbf{B}^T + \mathbf{E}_{::,n}. \end{aligned} \right\} \tag{2}$$

式(2)中: $\mathbf{D}_i(\mathbf{A})$ 为由矩阵 \mathbf{A} 的第 i 行元素构成的对角矩阵.

与二维矩阵低秩分解有所不同,PARAFAC 模型分解在没有任何额外的约束条件下,具有唯一性^[9-10].

2 欠定盲源数估计算法

2.1 构建 PARAFAC 模型

考虑线性瞬时混合的盲源分离模型为

$$\mathbf{X}(t) = \mathbf{A} \cdot \mathbf{S}(t) + \mathbf{E}(t). \tag{3}$$

式(3)中: $\mathbf{S}(t)=[s_1(t), s_2(t), \cdots, s_M(t)]^T$ 为 M 个源信号构成的向量; $\mathbf{X}(t)=[x_1(t), x_2(t), \cdots, x_N(t)]^T$ 为 N 个传感器所感测到的信号构成的向量; \mathbf{A} 为 $N \times M$ 维混合矩阵; $\mathbf{E}(t)$ 为附加的噪声信号.

在实际应用中,受到传感器个数的限制,源信号的个数往往多于混合信号的个数,而且源信号的个数常是未知的.因此,仅考虑欠定的盲源分离情形,即 $N < M$.

假设每一个观测信号包含 K 个采样点,利用二阶非平稳源分离(SONS)的主要思想^[11],首先,对各

传感器数据进行中心化预处理. 然后, 将预处理后的信号分成不重叠的 L (可整除) 段数据块, 每段数据块包含 $N_l = K/L$ 个数据点, 用 $l = 1, \dots, L$ 标志数据块. 因此, 式(3)也可表示为

$$\mathbf{X}(l, t) = \mathbf{A} \cdot \mathbf{S}(l, t) + \mathbf{E}(l, t). \quad (4)$$

式(4)中: $\mathbf{X}(l, t) = [x_1(l, t), \dots, x_N(l, t)]^T$; $\mathbf{S}(l, t) = [s_1(l, t), \dots, s_M(l, t)]^T$.

同时, 估计观测信号各段数据块的时滞协方差矩阵 \mathbf{R}_x , 有

$$\mathbf{R}_x(T_l, \tau_i) = \mathbf{A} \mathbf{R}_s(T_l, \tau_i) \mathbf{A}^T, \quad \forall l, i. \quad (5)$$

式(5)中: $\mathbf{A} \in \mathbf{R}_{n \times m}$ 为混合矩阵; τ_i 为时间延迟, $i = 1, 2, \dots, I$; T_l 为第 l 个时间窗口.

针对单一时延 $\tau_1 = 0$ 协方差矩阵 (即 $I = 1$), 有

$$\mathbf{R}_x(T_l, \tau_1) = \mathbf{A} \mathbf{R}_s(T_l, \tau_1) \mathbf{A}^T, \quad \forall l. \quad (6)$$

式(6)中: 源时滞协方差矩阵为

$$\mathbf{R}_s(T_l, \tau_1) = \begin{bmatrix} R_{1,1}(T_l, \tau_1) & R_{1,2}(T_l, \tau_1) & \cdots & R_{1,M}(T_l, \tau_1) \\ R_{2,1}(T_l, \tau_1) & R_{2,2}(T_l, \tau_1) & \cdots & R_{2,M}(T_l, \tau_1) \\ \vdots & \vdots & & \vdots \\ R_{M,1}(T_l, \tau_1) & R_{M,2}(T_l, \tau_1) & \cdots & R_{M,M}(T_l, \tau_1) \end{bmatrix}. \quad (7)$$

假设源信号 $\mathbf{S}(t) = [s_1(t), s_2(t), \dots, s_M(t)]^T$ 之间互不相关, 则源时滞协方差矩阵 \mathbf{R}_s 为对角矩阵.

将各段数据块时滞协方差 $\mathbf{R}_x(T_l, \tau_1)$ 叠加成三阶张量, 记为 $R_{n_1, n_2, l}$, 其中, $n_1 = 1, \dots, N$; $n_2 = 1, \dots, N$; $l = 1, \dots, L$. \mathbf{A} 中的元素记为 $a_{n \times m}$; 构建一个 $M \times L$ 维矩阵 \mathbf{C} , 其表达式为

$$\mathbf{C} = \begin{bmatrix} R_{1,1}(T_1, \tau_1) & R_{1,2}(T_2, \tau_1) & \cdots & R_{1,L}(T_L, \tau_1) \\ R_{2,1}(T_1, \tau_1) & R_{2,2}(T_2, \tau_1) & \cdots & R_{2,L}(T_L, \tau_1) \\ \vdots & \vdots & & \vdots \\ R_{M,1}(T_1, \tau_1) & R_{M,2}(T_2, \tau_1) & \cdots & R_{M,L}(T_L, \tau_1) \end{bmatrix}. \quad (8)$$

根据式(6)可改写为

$$R_{n_1, n_2, l} = \mathbf{A} \cdot \text{diag}(\mathbf{C}) \mathbf{A}^T, \quad \forall l. \quad (9)$$

记其标量形式为

$$R_{n_1, n_2, l} = \sum_{m=1}^M a_{n_1, m} c_{l, m} a_{n_2, m}^*. \quad (10)$$

式(10)中: $n_1 = 1, 2, \dots, N$; $n_2 = 1, 2, \dots, N$. 与式(1)对比可知, 式(10)为 PARAFAC 分解模型.

2.2 核一致诊断法

在 PARAFAC 分解模型具有可辨识性条件下, PARAFAC 模型中因子数也具有唯一性. 为了对模型的组分数进行确定, 通常采用由 Bro 等^[12]提出的核一致诊断算法进行测定. 利用核一致诊断法对上述构建的 PARAFAC 模型进行求解, 得到的最佳组分数与机械振动系统的振源数相对应. 核一致诊断法通过计算 PARAFAC 模型中的超对角阵 \mathbf{T} 和 truck3 模型中的立方阵 \mathbf{G} 之间的相似程度, 即通过核一致 ξ 估计组分数, 有

$$\xi = \left[1 - \frac{\sum_{a=1}^M \sum_{b=1}^M \sum_{c=1}^M (g_{abc} - t_{abc})^2}{\sum_{a=1}^M \sum_{b=1}^M \sum_{c=1}^M t_{abc}^2} \right] \times 100\%. \quad (11)$$

式(11)中: M 为模型的组分数; g_{abc} 为立方阵 \mathbf{G} 的元素; t_{abc} 为超对角阵 \mathbf{T} 的元素.

对于可辨识的 PARAFAC 模型, 超对角阵 \mathbf{T} 和 truck3 模型的立方阵 \mathbf{G} 非常相似, 此时的核一致值等于 100%. 一般来说, 当核一致值 ξ 大于或等于 60% 时, 可认为其符合三线性模型; 当核一致值 ξ 低于 40% 时, 可认为偏离三线性模型. 因此, 可以根据核一致值的变化确定最佳组分数, 即振动源个数.

3 仿真试验与分析

为验证该算法的有效性, 源信号生成函数为

$$\mathbf{S}(t) = \begin{bmatrix} s_1(t) \\ s_2(t) \\ s_3(t) \\ s_4(t) \end{bmatrix} = \begin{bmatrix} 0.1(t+5) \cdot (3 + \sin(80\pi t)) \sin(60\pi t) \\ 0.2(t+4)^{1.25} \cdot (1 + \sin(80\pi t)) \sin(40\pi t) \\ 0.1(t+5)^{1.5} \sum_{n=1}^{101} \sin(500(t-0.1n)) \exp(\frac{-50(t-0.1n)}{2}) u(t-0.1n) \\ 0.1(t+5)^{0.8} \sum_{n=1}^6 \sin(n20\pi t) \end{bmatrix} \cdot \quad (12)$$

仿真源信号时间波形及幅值谱如图 2,3 所示。图 2,3 中: H 为幅值; f 为频率。采样频率 $f_s=1\ 000$ Hz,采样点数 $N=10\ 240$,设置混合矩阵取值范围在 $[0\sim1]$ 区间。混合矩阵 \mathbf{A} 与源信号 $\mathbf{S}(t)$ 相乘得到的混合信号波形和幅值谱,如图 4,5 所示。由图 4 可知:无法从混合信号中直接评估独立组成成分波形及数目。由图 5 可知:观测信号中均存在源信号的特征频率,四路振动源信号完全相互干扰混在一起。

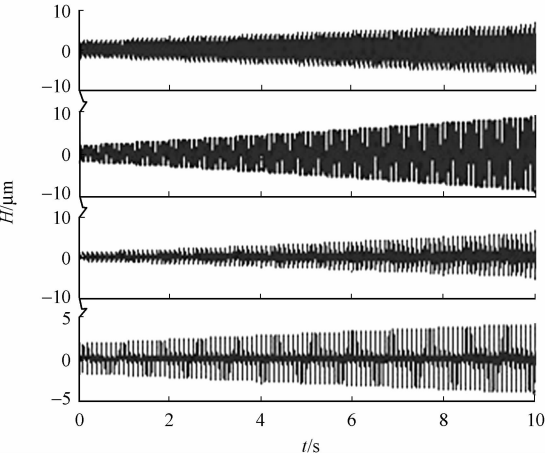


图 2 源信号时间波形

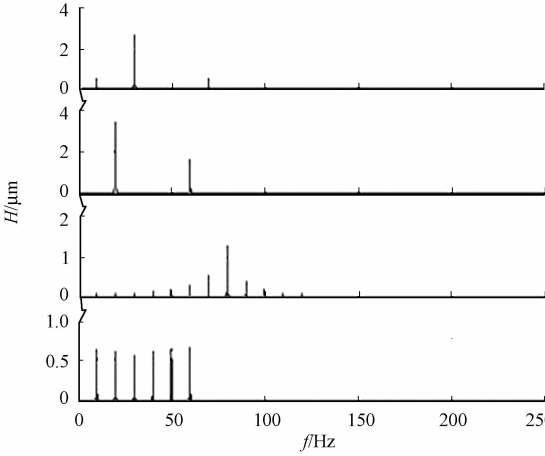


图 3 源信号幅值谱

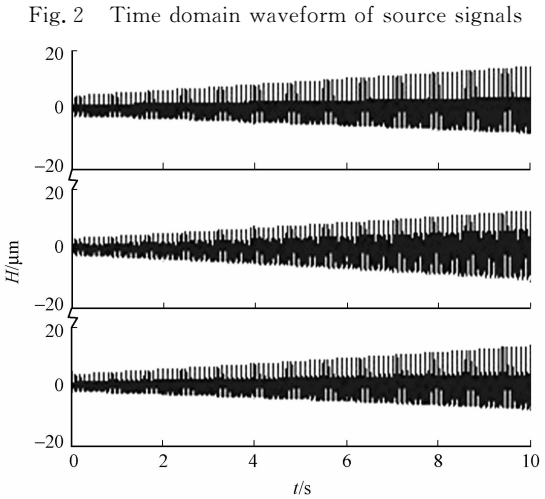


图 4 混合信号时间波形

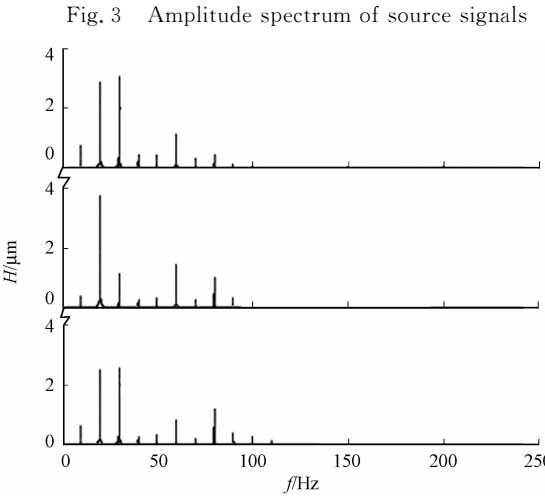


图 5 混合信号幅值谱

Fig. 2 Time domain waveform of source signals

Fig. 3 Amplitude spectrum of source signals

由于观测信号成分复杂,无法预知机械系统中的振源数,故采用核一致诊断法估计系统中的振源数。通过 PARAFAC 和 SONS 联合构造 PARAFAC 模型,并利用核一致诊断对 PARAFAC 模型求解,得到模型因子数,即振动源数。首先,对各观测信号的 10 240 个采样点进行去均值化预处理;然后,将其分割成 20 块不重叠的数据段,每个数据段包含 512 个数据点,并在此基础上,得到每一块数据块相应的 3×3 维时滞协方差矩阵;接着,将 20 块数据段各自的时滞协方差矩阵叠加成 3 阶张量形式,合理初始化载荷矩阵,采用 PARAFAC 模型的三线性交替最小二乘算法(TALS)对模型进行低秩分解。

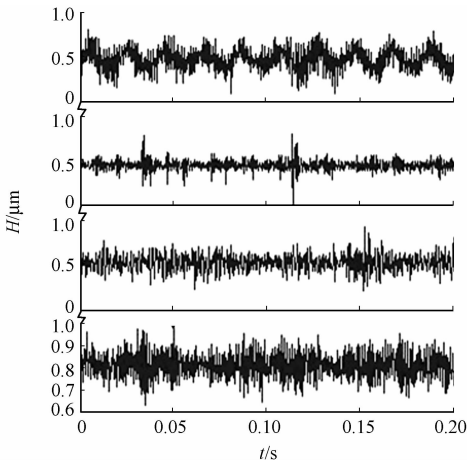
通过核一致诊断算法,估计最佳组分数,从而得到振动源数 M 。研究模型不同组分数对核一致值的影响,如图 6 所示。由图 6 可知:可根据核一致值的变化确定最佳组分数。即当组分数为 $1\sim4$ 时,对应的

核一致值均大于 80%, 表明符合模型; 当组分数大于 4 时, 对应的核一致值迅速趋近于 0, 表明偏离模型. 由此得到最佳因子数为 4, 即振动源数 $M=4$.

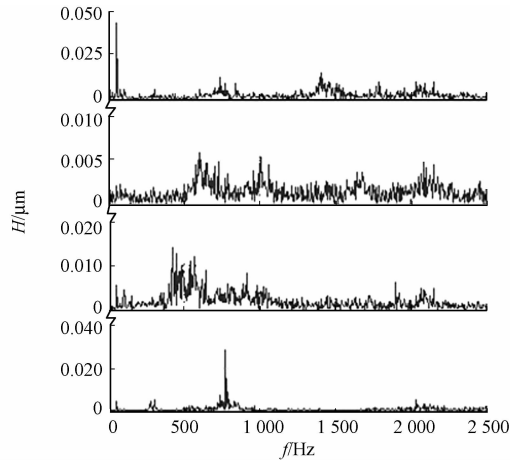
4 试验台试验与分析

采用电机-减速箱耦合实验装置(图 7), 验证所提算法在机械多振源系统中源数估计的有效性. 设置 3 个振源装置: 主机组(大电机-减速箱)中的大电机、干扰电机 1 和干扰电机 2. 在实验装置内, 本着传感器尽量靠近振源点的原则, 将传感器 1 安装于大电机(2.2 kW)壳体上; 将传感器 2, 3 安装于靠近齿轮箱输入轴位置; 将传感器 4 安装于齿轮箱顶盖位置; 将传感器 5 安装于干扰电机 1 上; 将传感器 6 安装于干扰电机 2 上. 在具体实验中, 采用定频率、等间隔采样策略, 采样频率为 5 kHz(为清楚起见, 仅给出前 1 024 个样本点).

当各电机同时运行时, 由传感器 1, 2, 5, 6 共同获得 4 通道的混合振动观测信号, 其时域及频域波形, 如图 8 所示. 由图 8 可知: 观测混合信号波形复杂, 难以有效评估出蕴涵其中的振源波形和数目信息. 文献[13]对参考振源信号的独立性进行校验, 不相关源数估计为 5.



(a) 波形图



(b) 幅值谱

图 8 观测信号波形图与幅值谱

Fig. 8 Time domain waveform and amplitude spectrum of mechanical observed signals

利用 PARAFAC 的核一致诊断法估计 4 通道壳体结构实验台观测混合信号的振源数目, 分析结果如图 9 所示. 由图 9 可知: 当组分数为 1~5 时, 对应的核一致值均大于 80%, 表明符合模型; 当组分数大于 5 时, 对应的核一致值迅速趋近于 0, 表明偏离模型. 由此得到最佳因子数为 5, 即独立振动源个数为 5. 与文献[13]估计的振源数一致, 进一步证实了该方法的有效性. 在机械振源数估计中, 将 PARAFAC 模型的唯一性与核一致诊断算法相结合, 能够准确地得到振源数目, 为下一步欠定混合信号的盲源分离提供可靠的先验信息.

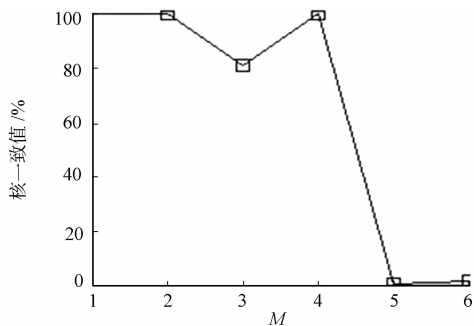


图 6 基于核一致诊断的信源数估计
Fig. 6 Source number estimation based on CORCONDIA

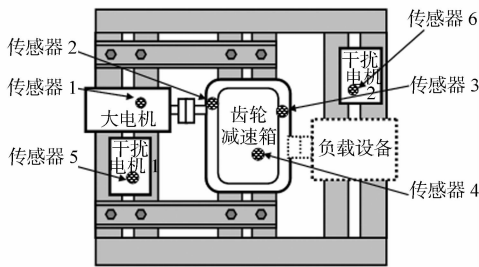


图 7 电机-减速箱耦合实验装置
Fig. 7 Motor reducer coupling experiment

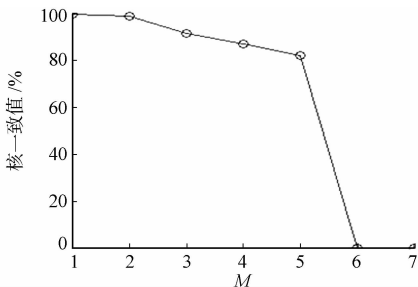


图 9 基于核一致诊断的实验台振源数估计
Fig. 9 Vibration source number estimation of experiment table based on CORCONDIA

5 结论

1) 提出一种基于平行因子分析和核一致诊断的欠定盲源数估计算法. 该算法利用二阶非平稳源分离的基本思想, 首先, 将中心化预处理后的传感器数据分成不重叠的数据块. 然后, 计算各数据块的单一时延协方差矩阵, 并叠加成平行因子模型. 最后, 利用核一致诊断算法估计平行因子模型的最佳组分数, 从而得到振源数的估计.

2) 根据振动信号的非平稳性, 该 PARAFAC 模型在宽松约束条件下具有唯一可辨识性. 此外, 该模型在欠定条件下仍满足可辨识性条件, 故该算法能够解决欠定盲源分离下的源数估计问题. 仿真实验结果表明, 该算法可从非平稳欠定混合信号中准确估计振源数目. 将所提算法应用于多机振动源实验中, 实验结果进一步验证了该方法的有效性.

参考文献:

- [1] 王国彪, 何正嘉, 陈雪峰, 等. 机械故障诊断基础研究“何去何从”[J]. 机械工程学报, 2013, 49(1): 63-72. DOI:10.3901/JME.2013.01.063.
- [2] LEI Huang, TENG Long, WU Shunjun. Source enumeration for high-resolution array processing using improved Gerschgorin radii without eigendecomposition[J]. IEEE Transactions on Signal Processing, 2008, 56(12): 5916-5925. DOI:10.1109/TSP.2008.929331.
- [3] 成玮, 何正嘉, 张周锁. 基于独立分量分析的壳体结构振源数目估计方法[J]. 机械工程学报, 2014, 50(19): 73-79. DOI:10.3901/JME.2014.19.073.
- [4] WU H T, YANG J F, CHEN F K. Source number estimators using transformed Gerschgorin radii[J]. IEEE Transactions on Signal Processing, 1995, 43(6): 1325-1333.
- [5] SEREDYŃSKA-SOBECKA B, STEDMON C A, BOE-HANSEN R, *et al.* Monitoring organic loading to swimming pools by fluorescence excitation-emission matrix with parallel factor analysis (PARAFAC)[J]. Water Research, 2011, 45(6): 2306-2314. DOI:10.1016/j.watres.2011.01.010.
- [6] MOUNIER S, RAYNAUD M, OURSEL B, *et al.* Fluorescence indexes and PARAFAC components-what is the best way[C]//Egu General Assembly Conference. Vienna: [s. n.], 2015: 15639.
- [7] ZHANG Xiaofei, CAO Rencheng, ZHOU Ming. Noncircular-PARAFAC for 2D-DOA estimation of noncircular signals in arbitrarily spaced acoustic vector-sensor array subjected to unknown locations[J]. EURASIP Journal on Advances in Signal Processing, 2013, 2013(1): 1-10. DOI:10.1186/1687-6180-2013-107.
- [8] 张小飞. 平行因子分析理论及其在通信和信号处理中的应用[M]. 北京: 电子工业出版社, 2014.
- [9] SIDIROPOULOS N D, BRO R, GIANNAKIS G B. Parallel factor analysis in sensor array processing[J]. IEEE Transactions on Signal Processing, 2000, 48(8): 2377-2388. DOI:10.1109/78.852018.
- [10] SIDIROPOULOS N D, GIANNAKIS G B, BRO R. Blind PARAFAC receivers for DS-CDMA systems[J]. IEEE Transactions on Signal Processing, 2000, 48(3): 810-823. DOI:10.1109/78.824675.
- [11] CHOI S, CICHOCKI A, BELOUCHRANI A. Blind separation of second-order nonstationary and temporally colored sources[C]//IEEE Signal Processing Workshop on Statistical Signal Processing. Piscataway: IEEE Press, 2001: 444-447. DOI:10.1109/SSP.2001.955318.
- [12] BRO R, KIERS H A. A new efficient method for determining the number of components in PARAFAC models[J]. Journal of Chemometrics, 2003, 17(5): 274-286. DOI:10.1002/cem.801.
- [13] 焦卫东. 基于独立分量分析的旋转机械故障诊断方法研究[D]. 杭州: 浙江大学, 2003.

(责任编辑: 钱筠 英文审校: 崔长彩)