

基于非负 Tucker 3 分解的稀疏分量 分析在故障信号提取中的应用

王海军 许飞云

(东南大学机械工程学院, 南京 211189)

摘要: 针对初始故障信号不稀疏难于判断的问题,在非负 Tucker 3 分解(NTD)的基础上,提出了一种基于 NTD 的稀疏分量分析(SCA)处理二次特征信号的方法.同时,为了克服 NTD 算法收敛慢、易陷入过拟合等局限性,对分解因子增加了非负约束,并提出了对分解因子一次更新的算法.对比传统的最小交替二乘法,该更新算法能一次性地计算所有分解因子,避免了计算大规模的 Jacobian 矩阵,从而较大地提高了算法的效率.实验结果表明:NTD 和 SCA 相结合的方法(SCA_NTD)只需迭代约 150 步可达到收敛,而且在频谱稀疏性处理方面优于 NTF 等传统的方法;在分解相同维数张量的条件下,SCA_NTD 的最高精度达到了 97.16%.因此,SCA_NTD 不仅能够改善信号特征的稀疏性,而且对提高算法的收敛速度和精度也具有重要的意义.

关键词: 非负 Tucker 3 分解;稀疏分量分析;更新算法;交替最小二乘法

中图分类号: TH17; TP206 **文献标志码:** A **文章编号:** 1001-0505(2013)04-0758-05

Sparse component analysis based on nonnegative Tucker 3 decomposition for fault signal extraction

Wang Haijun Xu Feiyun

(School of Mechanical Engineering, Southeast University, Nanjing 211189, China)

Abstract: Aiming at the problem of non-sparseness of original diagnosis signal and being difficult to distinguish, a method of sparse component analysis(SCA) based on nonnegative Tucker 3 decomposition(NTD) is proposed to process the quadratic feature of faults for improving the sparseness. Meanwhile, a new updating algorithm with nonnegative constraints is put forward to overcome the limitation of slow convergence and data overfitting in NTD. Compared with the conventional alternative least squares(ALS), the updating algorithm can simultaneously compute all the factors in one time that avoids calculating the large-scale Jacobian matrix and therefore improves the efficiency. The experimental results show that the method of combining NTD and SCA(SCA_NTD) only needs 150 steps to achieve convergence which is superior to other typical methods like NTF, etc; SCA_NTD also gets the highest accuracy of 97.16% under the same dimension of a tensor. Therefore, SCA_NTD not only improves the sparseness but also is significant to improve the convergence and efficiency.

Key words: nonnegative Tucker 3 decomposition; sparse component analysis; updating algorithm; alternative least squares method

Tucker 3 分解模型是 Tucker^[1] 针对多维数据降维而提出的一种高效数学分解模型.随着计算机技

术的发展,在交替最小二乘算法的基础上,Paatero 等^[2]建立了三维 CANDECOMP/PARAFAC(CP)并

收稿日期: 2012-12-06. 作者简介: 王海军(1982—),男,博士生;许飞云(联系人),男,博士,教授,博士生导师, fyxu@seu.edu.cn.

基金项目: 国家自然科学基金资助项目(50875048, 51175079, 51075069).

引文格式: 王海军,许飞云.基于非负 Tucker 3 分解的稀疏分量分析在故障信号提取中的应用[J].东南大学学报:自然科学版,2013,43(4):758-762. [doi:10.3969/j.issn.1001-0505.2013.04.016]

行计算模型,CP模型对非负分解算法的发展起到了很大的推动作用.从实体数据的有效性出发,Lee等^[3]提出并证实了非负矩阵分解方法对图像的局部特征具有良好的解释性.自此,非负矩阵分解方法在盲信号处理、图像特征提取、神经系统和化学计量学等领域中得到了广泛的应用.目前,对三维目标特征的局部化处理和分类仍然是研究中的热点和难点.这样便引出了对算法的稀疏性控制优化等方面的研究.在特征稀疏性处理上比较典型的研究主要包括利用拉格朗日优化、增加并行因子补偿项和稀疏性控制项等方法来表达固有分量的特征^[4-9].在较小规模数据的计算上,这些方法能达到预期的目的.但是,对于高维大规模数据的计算往往会带来收敛速度慢和局部过拟合问题,而这些问题会直接影响到计算的精度和二次特征的有效表达.因此,针对过拟合和特征稀疏性的问题,本文提出了非负 Tucker 3 分解(NTD)结合稀疏分量分析(SCA)的方法来提高分解的效率以及二次信号特征的稀疏性;同时,对 Tucker 3 的分解因子进行非负约束和一次性更新计算以提高计算的精度和效率.这样提取出的二次特征在稀疏性和精度方面也会得到有效的控制.因而,该方法的研究从理论上来说显得很有必要,在实践中也具有重要意义.

1 Tucker 3 分解模型

作为一种成功的张量分解方法,Tucker提出的分解方法可以简单地概括为3种模型,即 Tucker 1、Tucker 2 和 Tucker 3 分解模型.其中,Tucker 3 分解模型又是前2种方法研究的延续和发展.在非负张量分解(NTF)的特征局部化处理与应用中具有十分重要的意义^[10-12].根据 NTF 的优越性,本文主要以 Tucker 3 分解模型作为研究对象.

根据图1,设一个实体的三维数据元 $Y \in \mathbf{R}^{I_1 \times I_2 \times I_3}$,则分解因子主要由1个核张量 $G \in \mathbf{R}^{J_1 \times J_2 \times J_3}$ 和3个模矩阵 $A^{(n)} \in \mathbf{R}^{I_n \times J_n}$, $n=1, 2, 3$ 组成;模矩阵由对应的 J_n 个列向量组成,即 $A^{(n)} = [a_1^{(n)}, a_2^{(n)}, \dots, a_{J_n}^{(n)}] \in \mathbf{R}^{I_n \times J_n}$. Tucker 3 分解模型的数学表达式为

$$Y \approx G \times_1 A^{(1)} \times_2 A^{(2)} \times_3 A^{(3)} = G \times \{A\} = \hat{Y} \quad \text{s. t. } \|A^{(n)}\| = 1 \quad (1)$$

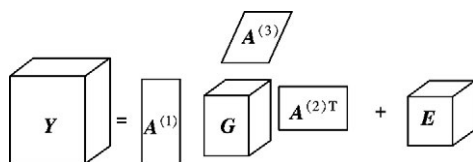


图1 Tucker 3 分解模型

式中, \times_1 表示矩阵形式的模数积; $\{A\}$ 表示所有模矩阵的 Kronecker 积; \hat{Y} 为 Y 被分解后的近似值.

对式(1)的最佳近似进行分解,可转化为求解以下最优化方程:

$$\min \frac{1}{2} \|Y - \hat{Y}\|_F^2 \quad (2)$$

对式(2)中的 $A^{(n)}$ 分别逐个求偏导,便可得到模矩阵 $A^{(1)}, A^{(2)}, A^{(3)}$ 的计算式.对 $A^{(n)}$ 进行交替迭代计算,并增加对 $A^{(n)}$ 非负约束,可得到张量核 G .非负约束对计算陷于局部过拟合起到了较大的抑制作用^[13].但是,这种迭代计算方式会产生巨大的 Jacobian 矩阵,同时也带来了收敛慢和效率低的问题^[12].因此,研究一种更高效合理的计算方法是本文的研究目的之一.

2 牛顿-高斯梯度下降的迭代方法

对分解因子 $A^{(1)}, A^{(2)}, A^{(3)}$ 以及 G 进行重新组合,得到一个新的矩阵 $M = [A^{(1)T}, A^{(2)T}, \dots, A^{(N)T}, \text{vec}(G)]$.其中, $\text{vec}(\cdot)$ 表示展开堆叠的张量 G (关于张量的展开,可参见文献[10]).将各堆叠的矩阵重新排列成一行.根据牛顿-高斯梯度下降迭代法,得到矩阵 M 的更新方程为

$$M \leftarrow M - H^{-1}r \quad (3)$$

式中 H 为海森矩阵; r 为梯度矩阵,计算公式为

$$r = K^T(\hat{Y} - y), \quad H = K^T K \quad (4)$$

s. t. $K = [K_1, K_2, \dots, K_{N+1}]$, $K \in \mathbf{R}^{\prod_{n=1}^N I_n \times \sum_{n=1}^N R_n J_n}$

$$K_n = \begin{cases} P_n^T(\{A\}_{-n} C_{(n)}^T \otimes I_{I_n}) & n = 1, 2, \dots, N \\ \{A\} & n = N + 1 \end{cases}$$

式中, $y = \text{vec}(Y)$, $\hat{Y} = \text{vec}(\hat{Y})$; P 为转置矩阵; $\{A\}_{-n}$ 表示除 $A^{(n)}$ 之外所有模矩阵的 Kronecker 积; K 为一个 Jacobian 矩阵.

为了避免海森矩阵 H 出现极值为零而影响计算的效率,取 H 的近似矩阵 \hat{H} ,以保证极值不为零,令 $\hat{H} = H + uI$,其中 $0 < u \ll 1$.

将新的海森矩阵 \hat{H} 分块,则必然存在 $\hat{H} = \lambda I$, $\lambda > 0$.在给定矩阵 M 附近,点 $M^{t+1} = M^t - \eta(\nabla f(M^t))$ 以及下降的步长 $0 < \eta < 1/\lambda$,必有 $f(M^{t+1}) < f(M^t)$,证明从略.

3 SCA 的二次特征处理

3.1 信号的稀疏处理

高斯-笛卡尔密度核函数是由 Khoromskij 等^[14]针对三维 PARAFAC 分解因子计算而提出,并应用于冗余化学原子库信号稀疏化的方法.高

<http://journal.seu.edu.cn>

斯-笛卡尔密度核函数不仅能处理离散化信号,而且还具有滤波降噪的作用.因此,本文主要根据NTD各因子间叉积的特点,结合高斯-笛卡尔积,建立联合核函数为

$$\Phi(Y) = c(Y - G \times \{A\}) \exp(-\mu(Y - G \otimes \{A\})^2) \quad (5)$$

$$\text{s. t. } c = \frac{1}{(2\pi)^{1/2} |\sigma|}, \mu = \frac{1}{2\sigma^2}$$

式中, σ 为 Y 与 \hat{Y} 之间的协方差. 在这里, $c \exp(-\mu(Y - G \otimes \{A\})^2)$ 与离散信号的窗函数作用类似.

假如 $\Phi(Y)$ 为冗余完备库, Y 为其观测信号, 则令 $\Phi(Y)$ 为一高斯原子, 与 Y 信号长度相同, 均进行归一化处理. 两者间的最优化核函数求解可转化为解决以下映射内积的优化问题:

$$|\langle Y, \Phi(Y_0) \rangle| = \alpha \sup |\langle Y, \Phi(Y) \rangle| \quad (6)$$

$$\text{s. t. } 0 < \alpha < 1$$

式中, $\langle Y, \Phi(Y_0) \rangle$ 表示 Y 与 $\Phi(Y_0)$ 的内积; $\Phi(Y_0)$ 为 Y 与 $\Phi(Y)$ 间的最佳原子库; α 为充分考虑信号长度和 $\Phi(Y_0)$ 损失等原因的常数. 则信号最终分解为 2 部分: 一部分为最佳高斯核函数; 另一部分为分配后的残余信号. 其数学表达式为

$$\langle Y, \Phi(Y) \rangle = \langle Y, \Phi(Y_0) \rangle \Phi(Y_0) + rY \quad (7)$$

式中, $\langle Y, \Phi(Y_0) \rangle \Phi(Y_0)$ 表示观测信号在 $\Phi(Y_0)$ 上的最佳映射; rY 为映射后的残余信号. 实际上, 高斯核函数与 NTD 后各因子参数直接相关. 而根据交替迭代计算方式, 核张量为 $G = Y \times \{A\}^T$. 因此, 模矩阵的计算直接影响着高斯核函数的质量. 这也充分说明了更新算法在迭代计算中很重要.

3.2 混合矩阵估计

混合矩阵是稀疏分量处理的重要组成部分, 其精确度直接决定着信号分离的结果. 假设稀疏化处理后的混合信号由 Y_1, Y_2, \dots, Y_N 子信号组成, 将各分量进行分层处理, 令 $y_n = Y_n(:, j) \in Y_n, 1 \leq n \leq N$. 其中 y 表示 $m \times N$ 的观测矩阵. SCA 类似独立分量分析^[15], 主要解决以下线性信号分解问题:

$$Y = HS \quad (8)$$

式中, S 为 $n \times N$ 稀疏源信号矩阵, N 为信号样本; $H = (h_{ij}) (i=1, 2, \dots, m; j=1, 2, \dots, n)$ 为未知的混合矩阵. 为了保证在信号特征依然稀疏的情况下能得到最佳混合矩阵, 演化成解的最优化问题, 即

$$\min \|Y - HS\| \quad (9)$$

当信号损失足够小时, 对方程(9)求最优解, 得到

$$H = YS^\dagger \quad (10)$$

式中, \dagger 表示伪逆. 估算出 H 矩阵后, 用最小交替迭代二乘法对稀疏源信号矩阵进行逼近计算. 整个信

号特征提取的流程如图 2 所示.

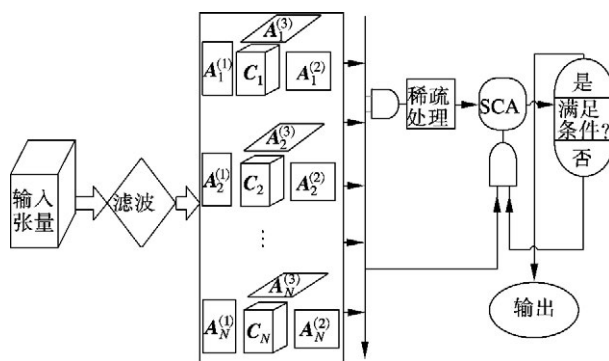


图 2 混合信号的 SCA 处理流程

4 齿轮箱信号分解验证

为了验证该算法的稀疏性和可靠性, 采用东南大学故障诊断研究所的 3 种齿轮箱的故障数据. 实验设备包括 3 套齿轮箱-电机系统, 将位于中间的单级齿轮箱作为测试对象. 齿轮箱的内部结构原理如图 3 所示. 3 个齿轮箱中的主动轮分别设置为正常、齿面点蚀和均匀磨损 3 种故障状态. 齿轮箱的输入轴通过刚性联轴器与电机相连, 转速可由 Siemens MicroMaster420 控制器进行调节. 主动轮与从动轮间的传动比为 31:46. 在电机转速约为 4 000 r/min 的情况下, 通过分别安装在齿轮箱上垂直和水平方向上的压电传感器采集振动信号. 如图 3 所示, 传感器灵敏度为 100 mV/g, 误差范围为 ± 3 dB, 采样频率为 3 838 Hz. 分别对 3 种故障状态每种采集 20 组振动信号, 每组长度为 4 096 点. 齿轮的啮合频率和滚动轴承外圈通过频率分别为 310 和 99.7 Hz. 采样频率约为 10 kHz.

取双谱 2 个正频率轴的频率点数为 64, 将分别采集到的信号加噪后组成 255 组包含 3 种故障

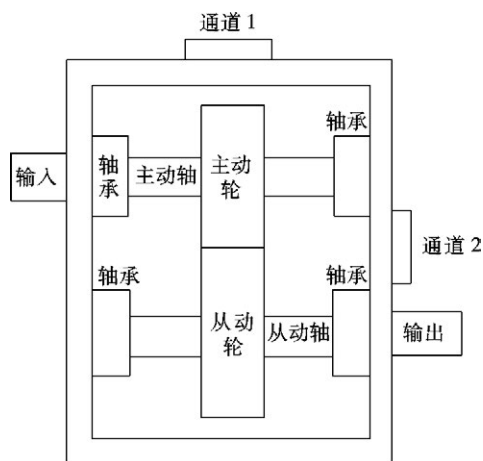


图 3 齿轮箱结构图

状态的数据,即构成一个 $\Omega \times \Omega \times S$ 的张量,其中 $\Omega = 64$, $S = 255$. 取 NTD 后的张量核为 $32 \times 32 \times 64$, 设定 NTF 分解因子维数为 32^3 迭代中的收敛误差为

$$E_t = \frac{\|\hat{Y} - Y\|_F}{\|Y\|} \times 100\% \quad (11)$$

将本文的迭代收敛误差与传统的 NTF 进行比较,如图 4 所示.

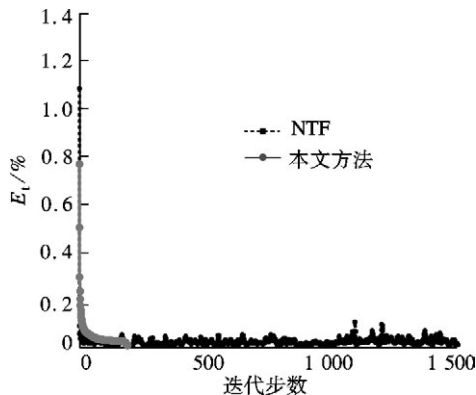


图 4 迭代收敛误差比较

设定迭代停止误差为 10^{-3} ,在训练张量维数相同的情况下,NTD 达到目标精度所需的迭代步数约为 150,收敛效率明显高于 NTF. 另一方面,在迭代误差计算过程中,NTD 的收敛误差较平滑,从而说明了该算法具有良好的健壮性. 因此,从收敛效率和健壮性两方面看,本文算法均优于 NTF.

为了让特征信息更加容易识别,本实验需对故障信号进行时频变换. 随机选取 3 种故障数据,进行快速 FFT 变换后得到的初始状态频谱图如图 5 所示.

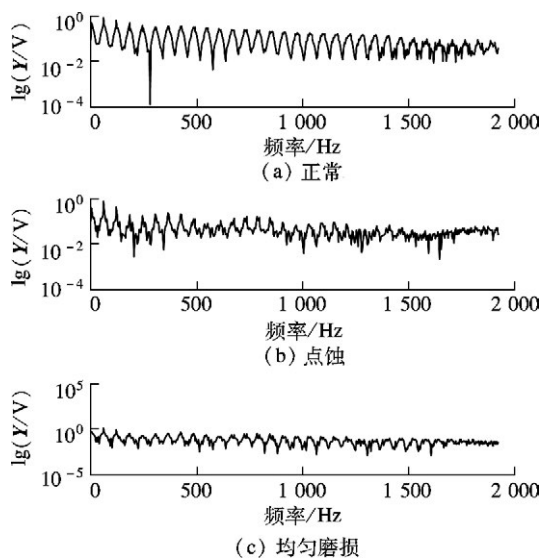


图 5 初始 3 种故障状态的频谱图

由图 5 可见 3 种状态对应的振动频率分布在

99.7 Hz 倍频时的概率较大. 对于正常状态和均匀磨损状态,初始信号的二次特征并不容易判断识别. 在电机高速旋转情况下,点蚀状态振动明显,频谱特征相对容易判断. 但是,如果在噪声干扰下,频谱特征分布不均匀,稀疏性差,信号的优势频率并不突出. 类似地,正常状态和均匀磨损状态的信号特征稀疏性更需要改进. 对此,本实验将采用上面提出的 SCA 与 NTD 相结合(SCA_NTD)的方法提取信号的二次故障特征,其频域特征如图 6 所示.

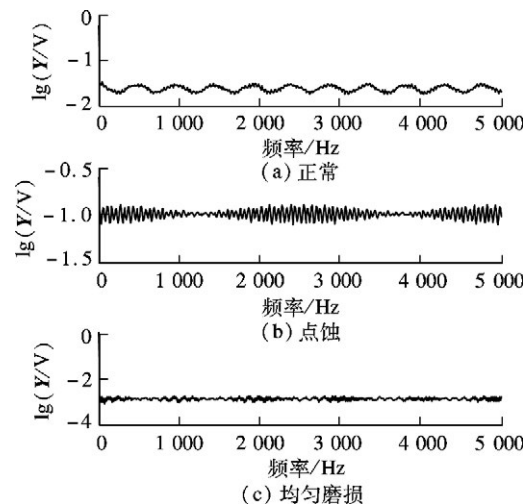


图 6 SCA_NTD 提取出的齿轮故障频谱图

与初始信号的二次特征相比,SCA_NTD 提取出的特征频率能满足周期性的特点,也与齿轮减速箱啮合频率和轴承通过外圈频率相符合;点蚀状态的振幅值对应于基频 99.7 和 310 Hz 的多倍频;均匀状态对应基频为 99.7 Hz 的倍频,与齿轮啮合频率也相符. 另外,特征信号的稀疏性比较好,容易观测,易于判断. 在此,将特征值小于 10^{-6} 近似作为特征信号的稀疏值. 初始状态与 SCA_NTD 方法处理后的特征稀疏值个数的结果如表 1 所示. 不难发现,处理后的特征稀疏值个数较多,从而说明了 SCA_NTD 处理后的盲信号具有良好的稀疏性.

表 1 齿轮箱故障特征的稀疏值个数

SCA_NTD	正常	点蚀	均匀磨损
处理前	7	79	68
处理后	204	291	200

为了证明 SCA_NTD 的可靠性,将其与经典的交替最小二乘法的 NTD(ALS_NTD) 和非负张量分解 NTF 算法进行比较. 计算精度为 $A_c = (1 - E_t) \times 100\%$. 实验结果如表 2 所示.

由表 2 中的精确度分布可见,相同计算方法的精度随着张量维数的增大而增高. 总体上看,SCA_NTD 计算得到的精度要比 ALS_NTD 和 NTF 高.

<http://journal.seu.edu.cn>

随着张量核维数的调整,SCA_NTD 的最高精度达到了 97.16% 相比 ALS_NTD 与 NTF 的最高精度 93.93% 和 88.81% ,优势明显. 从张量核维数组合

情况可看出,当核张量维数约为张量维数的一半时,精度最高. 因此,根据这一规律合理选择张量核维数,SCA_NTD 的可靠性将会进一步得到保证.

表 2 SCA_NTD 与 ALS_NTD 在不同张量核维数下的精确度

张量维数	方法	张量核维数				%
		(16 ,16 ,16)	(32 ,32 ,32)	(32 ,32 ,64)	(64 ,64 ,64)	
64 × 64 × 128	ALS_NTD	86.28	89.43	89.84	87.32	
	SCA_NTD	92.56	95.64	96.38	93.41	
	NTF	84.93	85.61	86.24	88.56	
64 × 64 × 255	ALS_NTD	88.34	93.26	93.93	89.42	
	SCA_NTD	90.26	96.98	97.16	94.35	
	NTF	87.39	88.69	88.81	88.00	

5 结语

针对 NTD 算法提取的二次特征信号不稀疏问题,结合 SCA 二次分离的方法得到了更加稀疏的特征信号. 在处理 SCA 的混合矩阵问题时,采用了交替迭代计算稀疏源伪逆矩阵的方法. 同时,为了避免 NTD 在迭代过程中陷于局部过拟合而导致误差增大和效率降低的问题,提出了一次更新所有分解因子的方法. 实验结果表明,SCA_NTD 达到了改善二次特征信号的稀疏性以及提高了计算精度和效率的目的.

参考文献 (References)

- [1] Tucker L R. Some mathematical notes on three-mode factor analysis[J]. *Psychometrika*, 1966, **31**(3): 279-311.
- [2] Paatero P, Tapper U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values [J]. *Environmetrics*, 1994, **5**(2): 111-126.
- [3] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization [J]. *Nature*, 1999, **401**(6755): 788-791.
- [4] Hazan T, Polak S, Shashua A. Sparse image coding using a 3D non-negative tensor factorization[C]//10th IEEE International Conference on Computer Vision. Beijing, 2005: 50-57.
- [5] Morup M, Hansen L K, Arnfred S M. Algorithms for sparse nonnegative Tucker decompositions [J]. *Neural Computation*, 2008, **20**(8): 2112-2131.
- [6] Cichocki A, Zdunek R, Phan A H, et al. *Alternating least squares and related algorithms for NMF and SCA problems in nonnegative matrix and tensor factorizations* [M]. Chichester, UK: John Wiley & Sons, Ltd, 2009: 203-266.
- [7] Peng S, Xu F, Jia M, et al. Sparseness-controlled non-negative tensor factorization and its application in machinery fault diagnosis [J]. *Journal of Southeast University: English Edition*, 2009, **25**(3): 346-350.
- [8] Karoui M S, Deville Y, Hosseini S, et al. Blind spatial unmixing of multispectral images: new methods combining sparse component analysis, clustering and non-negativity constraints [J]. *Pattern Recognition*, 2012, **45**(12): 4263-4278.
- [9] Asaei A, Davies M E, Bourlard H, et al. Computational methods for structured sparse component analysis of convolutive speech mixtures [C]//IEEE International Conference on Acoustics, Speech and Signal Processing. Kyoto, Japan, 2012: 2425-2428.
- [10] Kolda T G. *Multilinear operators for higher-order decompositions* [M]. California, USA: Sandia National Laboratories, 2006.
- [11] Cai X J, Chen Y N, Han D R. Nonnegative tensor factorizations using an alternating direction method [J]. *Frontiers of Mathematics in China*, 2013, **8**(1): 3-18.
- [12] Jiang L L, Yin H Q. Bregman iteration algorithm for sparse nonnegative matrix factorizations via alternating l (1) -norm minimization [J]. *Multidimensional Systems and Signal Processing*, 2012, **23**(3): 315-328.
- [13] Albright R, Cox J, Duling D, et al. Algorithms, initializations, and convergence for the nonnegative matrix factorization [R]. Raleigh, USA: Carolina State University, 2006.
- [14] Khoromskij B, Khoromskaia V, Chinnamsetty S, et al. Tensor decomposition in electronic structure calculations on 3D Cartesian grids [J]. *Journal of Computational Physics*, 2009, **228**(16): 5749-5762.
- [15] 刘海林, 姚楚君. 欠定混叠稀疏分量分析的超平面聚类算法 [J]. 系统仿真学报, 2009(7): 1826-1828.