论文

交替加权四线性分解算法用于处理非五线性
五维数阵卿湘东^{①②}, 吴海龙^{①*}, 张喜华^①, 李勇^①, 谷惠文^①, 文瑾^②, 申湘忠^②, 俞汝勤^①^① 化学生物传感与计量学国家重点实验室; 湖南大学化学化工学院, 长沙 410082^② 湖南人文科技学院化学与材料科学系, 娄底 417000

*通讯作者, E-mail: hlwu@hnu.edu.cn

收稿日期: 2015-06-12; 接受日期: 2015-08-15; 网络出版日期: 2016-01-22

国家自然科学基金(编号: 21175041, 21221003)和湖南人文科技学院引进人才启动项目(编号: 8250139)资助

摘要 针对获得的非五线性五维数阵, 采用一种先拓展后处理的思路, 即将该数阵沿着不具有线性的一维铺展成四线性四维数阵, 然后采用四维平行因子分析(4-PARAFAC)、交替加权残差约束四线性分解(AWRCQLD)和新发展的交替加权四线性分解(AWQLD)算法来对其进行解析. 在处理实验数据之前, 还通过 2 组模拟数据对提出的方法进行了验证. 结果表明, 新发展的 AWQLD 算法给出了与 AWRCQLD 和 4-PARAFAC 相似甚至更好的结果, 为处理非多线性多维数阵又提供了一种很有潜力的分析方法.

关键词 交替加权四线性分解, 四维校正, 非五线性五维数阵, 平行因子分析, 色谱

1 引言

随着现代分析仪器的飞速发展和数据采集技术的不断涌现, 三线性结构的数阵较易获取^[1,2]. 现在常用的两种方式是多样本激发发射矩阵荧光(EEMF)^[3]和二维液相/气相色谱或液相-气相色谱(LC×LC^[4]/GC×GC^[5]或 LC×GC^[6])与多通道检测器联用. 但对于四线性或五线性结构的数阵则相对较难, 主要是因为获取的方式比较有限. 而且在多维数阵的测量过程中, 由于受到环境因素如温度、压力和时间等影响, 容易导致一维或多维偏离线性. 常见的是色谱保留时间漂移导致色谱一维的偏离和光谱中瑞利散射导致激发发射光谱两维的偏离. 如果引进温

度和 pH 两维, 由于这两个实验条件的不易控制, 也可能造成两维轮廓变动而产生线性偏离. 此时所得到的多维数阵则是非多线性结构的, 直接采用基于多线性成分模型的算法来处理, 所获得的结果往往不准确.

对于处理非多线性数阵体系, 现已报道了 3 类方法: (1) 第一类是对原始数阵扣除非线性因素之后再采用多线性分解方法进行分析. 例如, 聂瑾芳等^[7]运用三线性分解方法来处理偏离三线性成分模型较大的瑞利散射, 有效地扣除了体系中的一阶瑞利散射干扰, 为后续三维荧光光谱定量分析奠定了良好的基础. (2) 第二类方法则是直接采用基于非多线性模型的算法来进行处理. 现已报道的此类算法有平行因

引用格式: 卿湘东, 吴海龙, 张喜华, 李勇, 谷惠文, 文瑾, 申湘忠, 俞汝勤. 交替加权四线性分解算法用于处理非五线性五维数阵. 中国科学: 化学, 2016, 46: 401–408
Qing XD, Wu HL, Zhang XH, Li Y, Gu HW, Wen J, Shen XZ, Yu RQ. A new alternating weighted quadrilinear decomposition algorithm with application for analysis of non-quinquelinear five-way data arrays. *Sci Sin Chem*, 2016, 46: 401–408, doi: 10.1360/N032015-00093

子分析 2 (PARAFAC2)^[8]、平行线性判别(PARALIND)^[9]、多元曲线分辨结合交替最小二乘(MCR-ALS)^[10]、非双线性秩消失(NBRA)^[11]、双线性最小二乘扩展至线性相关体系(BLLS extended to linearly dependent systems)^[12]、展开的、非线性核或 N 维偏最小二乘结合残差多线性(U-PLS/RML^[13,14], non-linear kernel-PLS/RML^[15]或 N -PLS/RML^[14])和人工神经网络结合 RML (ANN/RML)^[16]等。(3) 第三类方法是将原始数阵进行降维, 沿非线性的一维铺展成符合多线性结构的低维数阵, 再采用基于多线性成分模型的算法进行处理。例如, Maggio 等^[13]采用四维平行因子分析(4-PARAFAC)处理动力学与 pH 相关的五维荧光数阵, 发现结果明显好于直接使用五维平行因子分析(5-PARAFAC)处理相应五维数阵时的结果。

本文通过使用 HPLC-DAD 来监测除草剂莠草胺在不同温度下的水解动力学过程, 第一次获得了色谱-光谱-动力学-温度形式的非五线性五维数阵。然后采用先拓展后处理的思路, 即将该数阵沿着不具有线性的动力学一维方向铺展成四线性四维数阵, 再采用四维平行因子分析(4-PARAFAC)^[17]、交替加权残差约束四线性分解(AWRCQLD)^[18]和新发展的交替加权四线性分解(alternating weighted quadrilinear decomposition, AWQLD)算法来对其进行解析。在处理实际数据之前, 还模拟了 1 组荧光数据和 1 组色谱数据对提出的新算法进行了验证。结果显示, 针对非五线性五维数阵, 该方法能有效地分辨出色谱或(和)光谱和动力学等有用信息, 且获得了比较满意的定量结果。另外, 通过比较 3 种算法所获得的结果, 发现新发展的 AWQLD 算法给出了与 AWRCQLD 和 4-PARAFAC 相似的甚至更好的结果。

2 理论部分

2.1 多线性成分模型

在多线性成分模型的 N 维数阵 $\underline{X}_n$ 中, 每个元素 $x_{ijklm\dots}$ 有以下总的标量表达式:

$$x_{ijklm\dots} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} d_{lf} e_{mf} \cdots + e_{ijklm\dots} \quad (1)$$

式中 $i=1,2,\dots,I$, $j=1,2,\dots,J$, $k=1,2,\dots,K$, $l=1,2,\dots,L$, $m=1,2,\dots,M$, $\dots$ 。这里 $x_{ijklm\dots}$ 是 $\underline{X}_n$ 中的元素, a_{if} 、 b_{jf} 、

c_{kf} 、 d_{lf} 和 e_{mf} 分别是大小为 $I \times F$ 的矩阵 $\mathbf{A}$ 、 $J \times F$ 的矩阵 $\mathbf{B}$ 、 $K \times F$ 的矩阵 $\mathbf{C}$ 、 $L \times F$ 的矩阵 $\mathbf{D}$ 和 $M \times F$ 的矩阵 $\mathbf{E}$ 中的元素。 $e_{ijklm\dots}$ 是 N 维残差阵 $\underline{E}_n$ 中的元素, n 代表维数, F 为因子数。

2.2 四维 PARAFAC 算法

四维 PARAFAC (4-PARAFAC)作为其三线性成分模型的拓展, 在分析化学界受到广泛关注^[17,19]。它通过交替最小二乘方案求解。具体数学公式推导可参考文献[20]。

2.3 交替加权残差约束四线性分解(AWRCQLD)算法

AWRCQLD 算法由付海燕等^[18]于 2011 年提出, 该算法首次采用基于四线性成分模型的不完全扩展矩阵形式, 是为数不多的四线性数据分解算法之一, 继承了已有算法的优势, 分解结果也较稳定。但因为加了一些约束条件, 降低了其运算效率。详细的数学推导可参考文献[18]。

为了既保持已有算法的优势, 又能显著提高算法的运算效率, 本文提出了以下的新算法。

2.4 交替加权四线性分解(AWQLD)算法

多线性成分模型亦可采用不完全扩展矩阵的形式来表述, 参考文献[20]。

基于四线性成分模型的不完全扩展矩阵形式^[20], 能获得以下 4 个等式:

$$F_L = \mathbf{E}_{\dots} ((\mathbf{C} \odot \mathbf{B})^T)^+ = \mathbf{X}_{\dots} ((\mathbf{C} \odot \mathbf{B})^T)^+ - \mathbf{A} \text{diag}(\mathbf{d}_{(l)}) \quad (2)$$

$$F_I = \mathbf{E}_{i\dots} ((\mathbf{D} \odot \mathbf{C})^T)^+ = \mathbf{X}_{i\dots} ((\mathbf{D} \odot \mathbf{C})^T)^+ - \mathbf{B} \text{diag}(\mathbf{a}_{(i)}) \quad (3)$$

$$F_J = \mathbf{E}_{\dots j\dots} ((\mathbf{A} \odot \mathbf{D})^T)^+ = \mathbf{X}_{\dots j\dots} ((\mathbf{A} \odot \mathbf{D})^T)^+ - \mathbf{C} \text{diag}(\mathbf{b}_{(j)}) \quad (4)$$

$$F_K = \mathbf{E}_{\dots k\dots} ((\mathbf{B} \odot \mathbf{A})^T)^+ = \mathbf{X}_{\dots k\dots} ((\mathbf{B} \odot \mathbf{A})^T)^+ - \mathbf{D} \text{diag}(\mathbf{c}_{(k)}) \quad (5)$$

该算法利用以上方程, 并结合 $\mathbf{D}_A$ 、 $\mathbf{D}_B$ 、 $\mathbf{D}_C$ 和 $\mathbf{D}_D$ 作为权重矩阵, 基于交替加权最小二乘(alternating weighted least squares, AWLS), 并采用加权残差来构建新的目标函数, 这里:

$$\mathbf{D}_A = \text{diag}(\text{sqrt}(1/\text{diagm}(\mathbf{A}^+(\mathbf{A}^T)^+)))$$

$$\mathbf{D}_B = \text{diag}(\text{sqrt}(1/\text{diagm}(\mathbf{B}^+(\mathbf{B}^T)^+)))$$

$$\mathbf{D}_C = \text{diag}(\text{sqrt}(1/\text{diagm}(\mathbf{C}^+(\mathbf{C}^T)^+)))$$

$$\mathbf{D}_D = \text{diag}(\text{sqrt}(1/\text{diagm}(\mathbf{D}^+(\mathbf{D}^T)^+)))$$

获得以下4个目标函数:

$$\sigma(\mathbf{D}) = \sum_{l=1}^L \left\| \left(\mathbf{X}_{\dots l} ((\mathbf{C} \odot \mathbf{B})^T)^+ - \mathbf{A} \text{diag}(\mathbf{d}_{(l)}) \right) \mathbf{D}_A \right\|_F^2 \quad (6)$$

$$\sigma(\mathbf{A}) = \sum_{i=1}^I \left\| \left(\mathbf{X}_{i \dots} ((\mathbf{D} \odot \mathbf{C})^T)^+ - \mathbf{B} \text{diag}(\mathbf{a}_{(i)}) \right) \mathbf{D}_B \right\|_F^2 \quad (7)$$

$$\sigma(\mathbf{B}) = \sum_{j=1}^J \left\| \left(\mathbf{X}_{\dots j} ((\mathbf{A} \odot \mathbf{D})^T)^+ - \mathbf{C} \text{diag}(\mathbf{b}_{(j)}) \right) \mathbf{D}_C \right\|_F^2 \quad (8)$$

$$\sigma(\mathbf{C}) = \sum_{k=1}^K \left\| \left(\mathbf{X}_{\dots k} ((\mathbf{B} \odot \mathbf{A})^T)^+ - \mathbf{D} \text{diag}(\mathbf{c}_{(k)}) \right) \mathbf{D}_D \right\|_F^2 \quad (9)$$

根据以上提出来的目标函数, 一种新的算法——交替加权四线性分解(AWQLD)算法被提出. AWQLD 类似于 AWRCQLD, 以交替的方式最小化加权残差, 即固定 $\mathbf{A}$ 、 $\mathbf{B}$ 和 $\mathbf{C}$, 以 $\mathbf{D}$ 为变量, 最小化 $\sigma(\mathbf{D})$; 固定 $\mathbf{B}$ 、 $\mathbf{C}$ 和 $\mathbf{D}$, 以 $\mathbf{A}$ 为变量, 最小化 $\sigma(\mathbf{A})$; 固定 $\mathbf{A}$ 、 $\mathbf{C}$ 和 $\mathbf{D}$, 以 $\mathbf{B}$ 为变量, 最小化 $\sigma(\mathbf{B})$; 固定 $\mathbf{A}$ 、 $\mathbf{B}$ 和 $\mathbf{D}$, 以 $\mathbf{C}$ 为变量, 最小化 $\sigma(\mathbf{C})$.

由于通常情况下, $\mathbf{A}$ 、 $\mathbf{B}$ 、 $\mathbf{C}$ 和 $\mathbf{D}$ 满足列满秩, 固定 $\mathbf{A}$ 、 $\mathbf{B}$ 和 $\mathbf{C}$, 获得更新的 $\mathbf{D}$ 的每一行为:

$$\mathbf{d}_{(l)} = (\text{diagm}((\mathbf{C} \odot \mathbf{B})^+ \mathbf{X}_{\dots l}^T \mathbf{A}) ./ \text{diagm}(\mathbf{A}^T \mathbf{A}))^T \quad (10)$$

固定 $\mathbf{B}$ 、 $\mathbf{C}$ 和 $\mathbf{D}$, 获得 $\mathbf{A}$ 的每一行为:

$$\mathbf{a}_{(i)} = (\text{diagm}((\mathbf{D} \odot \mathbf{C})^+ \mathbf{X}_{i \dots}^T \mathbf{B}) ./ \text{diagm}(\mathbf{B}^T \mathbf{B}))^T \quad (11)$$

固定 $\mathbf{A}$ 、 $\mathbf{C}$ 和 $\mathbf{D}$, 获得 $\mathbf{B}$ 的每一行为:

$$\mathbf{b}_{(j)} = (\text{diagm}((\mathbf{A} \odot \mathbf{D})^+ \mathbf{X}_{\dots j}^T \mathbf{C}) ./ \text{diagm}(\mathbf{C}^T \mathbf{C}))^T \quad (12)$$

固定 $\mathbf{A}$ 、 $\mathbf{B}$ 和 $\mathbf{D}$, 获得 $\mathbf{C}$ 的每一行为:

$$\mathbf{c}_{(k)} = (\text{diagm}((\mathbf{B} \odot \mathbf{A})^+ \mathbf{X}_{\dots k}^T \mathbf{D}) ./ \text{diagm}(\mathbf{D}^T \mathbf{D}))^T \quad (13)$$

AWQLD 算法的步骤描述如下: (1) 选取组分数 N ; (2) 随机初始化矩阵 $\mathbf{A}$ 、 $\mathbf{B}$ 和 $\mathbf{C}$; (3) 用式(10)计算 $\mathbf{D}$; (4) 用式(11)计算 $\mathbf{A}$, 并归一化 $\mathbf{A}$; (5) 用式(12)计算 $\mathbf{B}$, 并归一化 $\mathbf{B}$; (6) 用式(13)计算 $\mathbf{C}$, 并归一化 $\mathbf{C}$; (7) 用式(10)计算 $\mathbf{D}$; (8) 重复步骤(4)~(7)更新 $\mathbf{A}$ 、 $\mathbf{B}$ 、 $\mathbf{C}$ 和 $\mathbf{D}$, 直到满足停止准则:

$$\left| \frac{\text{SSR}^{(m)} - \text{SSR}^{(m-1)}}{\text{SSR}^{(m-1)}} \right| < \varepsilon$$

或者迭代次数大于 500; (9) 输出矩阵 $\mathbf{A}$ 、 $\mathbf{B}$ 、 $\mathbf{C}$ 和 $\mathbf{D}$ 的结果. 其中 SSR 为残差平方和, $\text{SSR}^{(m)} =$

$$\|\mathbf{E}\mathbf{q}\|_F^2 = \left(\sum_i \sum_j \sum_k \sum_l e_{ijkl}^2 \right), \quad m \text{ 为迭代次数, } \varepsilon = 10^{-6}.$$

3 实验部分

为了验证上述发展方法的优良性能, 我们通过 2 组模拟数据和 1 组真实数据对该方法的性能进行详细阐述. 模拟数据的产生与分析均在 MATLAB 平台上运行. 采用 MATLAB 语言来编译算法程序.

3.1 模拟数据

3.1.1 模拟的激发-发射-动力学-温度光谱数据集

首先模拟了一组包含 3 组分 16 样本的荧光光谱数据. 3 组分的激发光谱通过以下等式模拟产生:

$$\mathbf{A}(:,1) = \text{peak}(0.5, 17, 9, i)$$

$$\mathbf{A}(:,2) = \text{peak}(0.5, 20, 15, i)$$

$$\mathbf{A}(:,3) = \text{peak}(0.5, 20, 20, i)$$

式中 $i=1, 2, \dots, 40$. $\text{peak}(a, b, c, x)$ 表示中心值是 c , 标准偏差是 b , 采样点数是 x 的高斯分布函数, a 代表对高斯分布曲线的缩放系数.

发射光谱通过以下的等式模拟产生:

$$\mathbf{B}(:,1) = \text{peak}(0.3, 22, 12, j)$$

$$\mathbf{B}(:,2) = \text{peak}(0.4, 25, 15, j)$$

$$\mathbf{B}(:,3) = \text{peak}(0.4, 18, 17, j)$$

式中 $j=1, 2, \dots, 30$.

3 组分在 16 个样本中($K=16$)的浓度矩阵根据需要进行设计. 5 个温度下 3 组分的动力学曲线通过以下等式产生:

$$\mathbf{D}_f(:,1) = \exp(-k_f t)$$

$$\mathbf{D}_f(:,2) = 1 - \exp(-k_f t)$$

$$\mathbf{D}_f(:,3) = 0.5$$

式中 $k_f = [0.07, 0.1, 0.13, 0.16, 0.19]$, 为不同温度下的降解动力学速率常数.

同方差噪声由以下等式生成:

$$\text{homo_noise} = a_{\text{homo}} \times \text{RAND} \times \max(\mathbf{X}_{\text{IJKLM}})$$

式中 a_{homo} 表示控制均方差噪声的参数, RAND 表示随机分布函数. 本文中 $a_{\text{homo}} = 0.005$.

3.1.2 模拟的保留时间-光谱-动力学-温度 HPLC-DAD 数据集

我们又模拟了一组包含 3 组分 16 样本的色谱数据. 3 组分的保留时间曲线通过以下等式模拟产生:

$$\mathbf{A}(:,1) = 100 \text{ CHI}(0.25, 9, i)$$

$$\mathbf{A}(:,2) = 100 \text{ CHI}(0.25, 15, i)$$

$$\mathbf{A}(:,3) = 100 \text{ CHI}(0.25, 20, i)$$

式中

$$i=1,2,\dots,40, \text{CHI}(a,b,i)=[a \times i^{(b/2-1)} \times \exp(-i/2)] / \{2^{(b/2)}[(n/2-1)!!]\}.$$

3 组分在 16 个样本中的光谱、浓度矩阵、5 个温度下 3 组分的动力学曲线和同方差噪声的生成与 3.1.1 小节中结果相同。

3.2 实验数据

3.2.1 试剂和仪器

实验所有的化学品都是分析纯级。萘草胺(NAP)和萘乙酰胺(NAD)购自阿拉丁化学试剂公司(中国)。吲哚乙酸(IAA)购自 Sigma 化学试剂公司(St. Louis, 美国)。NAP、NAD 和 IAA 工作液的浓度分别为 401.6、207.2 和 53.4 $\mu\text{g mL}^{-1}$ 。

所用仪器为 LC-20AT 色谱系统(Shimadzu, 日本)。

3.2.2 样品制备

在 4 个 25.0 mL 容量瓶中分别加入适量的 NAP 工作液, 再加入 3.0 mL 磷酸缓冲液(pH 1.9), 然后以超纯水稀释至刻度, 获得了 4 个 NAP 最终浓度分别为 8.03、16.06、24.10 和 32.13 $\mu\text{g mL}^{-1}$ 的校正集。加超纯水稀释 6 个分别加了适量 NAP 工作液和恒量干扰化合物 IAA 和 NAD 混合液和 3.0 mL 磷酸缓冲液(pH 1.9)的 25.0 mL 容量瓶至刻度, 获得了 6 个 NAP 最终浓度不同的预测集(如表 1 第二列所示)。这些溶液的 HPLC-DAD 信号形式的动力学过程被记录, 如下所述。

对每个样品, 制备后分别置于 50、60 和 70℃ 的恒温水浴锅中持续加热, 其在反应开始后一系列时刻(0、5、10、15、20、25 min), 用巴氏吸管定时移取 1.0 mL 反应液至 10 mL 圆底离心管中, 然后立刻置于冰水中冷却 30 s 后进样。水解反应在 pH 1.9 的酸

性条件下进行。对于每个样品得到一个四维数阵, 所有的样品最终会产生一个大小为 291×76×6×3×10 (保留时间×波长×反应时间×温度×样品)的非五线性五维数阵, 即可用于多维分析。

4 结果与讨论

4.1 模拟数据的分析

4.1.1 光谱数据集

为了阐述和验证发展方法的可靠性, 我们首先模拟了一组 3 组分 16 样本的荧光光谱数据(数据集 I)。对于该非五线性五维数阵, 采用已有的五维校正算法如五维 PARAFAC 算法来进行处理是完全不可能的。因此, 我们尝试着将五维数阵沿着非线性的一维展开, 形成一个扩展的四维数阵。

对于该四维数阵, 首先采用核一致诊断法(CORCONDIA)^[21]来估计其因子数, 结果为 3。然后采用 4-PARAFAC、AWRCQLD 和 AWQLD 分别来解析所构建的 3 组分模型。图 1 展示了采用 3 种算法分别处理扩展的四维数阵所获得的归一化激发、发射、浓度和动力学-温度曲线。从图中可以看出, 3 种算法分辨所获得各个组分的轮廓曲线完全重叠在一起。值得说明的是, 动力学-温度曲线(图 1(d))中, 反应物与产物的动力学曲线在 5 个温度下完全不一样, 这说明在这一维中不具有线性, 然而发展的方法正确地分辨出了 5 个温度下各个组分的动力学曲线, 且 3 种算法分辨的轮廓完全重合在一起。在分辨的浓度(图 1(c))中, 3 种算法分辨的浓度曲线也完全重合, 说明该方法能够在动力学一维存在非线性时, 对感兴趣分析物进行定量分析。从这些结果可以看出, 新发展的算法获得了与 4-PARAFAC 和 AWRCQLD 完全

表 1 采用 3 种算法分别对样品中分析物萘草胺与两种植物激素存在严重干扰时分辨所获得的定量和统计结果

Sample No.	Actual value ($\mu\text{g mL}^{-1}$)	Predicted concentration ($\mu\text{g mL}^{-1}$) [Recovery (%)]		
		AWQLD	AWRCQLD	PARAFAC
I	8.03	8.02[99.9]	8.16[101.5]	8.16[101.5]
II	12.85	12.82[99.7]	12.89[100.3]	12.89[100.3]
III	17.67	17.84[100.9]	17.85[101.0]	17.85[101.0]
IV	22.49	22.27[99.0]	22.19[98.7]	22.19[98.7]
V	27.31	27.04[99.0]	27.02[98.9]	27.02[98.9]
VI	32.13	31.88[99.2]	31.60[98.4]	31.60[98.4]
AR±S.D. (%)		99.6±0.6	99.8±1.2	99.8±1.2
T-test		1.10	0.31	0.31
RMSEP ($\mu\text{g mL}^{-1}$)		0.013	0.020	0.020

图1 采用3种算法分别处理模拟的荧光数据所获得的归一化激发光谱(a)、发射光谱(b)、浓度(c)和动力学-温度(d)曲线. 其中+代表AWQLD的结果, 圆圈代表AWRCQLD的结果, 实线代表PARAFAC的结果(网络版彩图)

相似的结果.

采用3种算法分别对模拟的数据集I进行分辨所获得的定量和统计结果列于表2中. 从表中可以看出, 3种算法给出的结果都比较合理可靠. 在对预测均方根误差(RMSEP)进行比较后, 发现AWRCQLD给出的结果稍微较好, 4-PARAFAC的结果次之, 而AWQLD稍微较差. 但这种差异非常小, 且3种算法总的预测能力都较好. 充分说明了基于这3种算法的二维校正方法都可以用于在干扰存在下的目标物的定量分析, 且结果准确可靠. 另外, 还从分辨所获得的动力学曲线计算出了降解动力学速率常数(k_1), 列于表S1第3~5列(补充材料)中, 第2列为真实值. 通过比较计算值与真实值可以发现, 3种算法得到的计算值与真实值非常接近, 吻合得较好, 充分说明了该方法准确可靠. 又比较3种算法得出的计算值发现, AWQLD的结果稍好于另外两种算法.

4.1.2 色谱数据集

为了进一步验证新方法对于不同类型数据的解析能力, 又模拟了一组非五线性五维色谱数据(数据

集II)来进行考察. 采用与分析数据集I相同的步骤对色谱数据进行解析. 当组分数为3时, 解析数据集II同样获得了满意的分析结果. 采用3种算法所获得的所有曲线如图S1(见网络版补充材料)所示, 分辨结果与荧光数据基本一致.

采用3种算法对数据集II进行分辨所获得的定量和统计结果见表2. 3种算法所获得的目标分析物的平均回收率分别为 $100.03\% \pm 0.01\%$ 、 $100.03\% \pm 0.007\%$ 和 $100.02\% \pm 0.02\%$. 3种算法所获得的 T 值分别为4.11、6.47和1.88, 此处 $n=7$, $t_{0.025}^7=2.36$. 3种算法所获得的RMSEP分别为 3.3×10^{-4} 、 1.9×10^{-4} 和 1.8×10^{-4} . 这些结果显示, 3种算法都获得了比较可靠的结果, 说明3种算法都可以用于干扰存在下的目标分析物的定量分析. 但比较3种算法的结果可知, AWQLD性能要明显好于另外两种.

同样, 所获得的降解动力学速率常数(k_1)列于表S1第6~8列. 通过比较可以发现, 3种算法得到的计算值与真实值都非常接近, 吻合得较好, 充分说明了该方法准确可靠. 又比较了3种算法的计算值, 发现3种算法所得结果接近, AWQLD获得了与另外两种

算法几乎完全相似的结果.

以上结果都显示, 在解析不同来源类型的模拟数据时, 新方法都表现了良好的分析能力. 接下来, 我们将对真实的实验数据进行解析, 进一步说明新方法的可靠性.

4.2 实验数据的分析

运用新方法, 我们将获得的实际五维数阵沿着动力学一维方向扩展成一个四维数据阵. 然后采用 3 种算法来对其进行解析.

采用基于 CORCONDIA 的结果显示, 对于所构建的模型, 4 个因子就足够了. 采用 3 种算法来解析该

4 组分模型所获得的所有曲线如图 2 所示. 从图 2(a~c)可知, IAA 和 NAD 对 NAP 降解过程的分析产生了严重的干扰. 但结果显示, 3 种算法所获得的 NAP 的保留时间(图 2(a))和光谱(图 2(b))曲线与真实的保持一致. 另外, 与 AWRCQLD 和 4-PARAFAC 有负的保留时间峰和浓度不同, AWQLD 因加入了非负限制, 分辨得到的色谱和浓度曲线都未出现负值, 得到了更合理的谱图.

采用 3 种算法所获得的平均回收率(AR)、标准偏差(S.D.)、RMSEP 和 T 检验见表 1. 所有的预测结果都比较满意, 显示 3 种方法都可用于在存在 IAA 和 NAD 干扰下定量分析 NAP 的水解过程. 值得指出的是, 在对

表 2 采用 3 种算法分别对模拟的光谱(EEMF)和色谱(HPLC)数据集进行分辨所获得的定量和统计结果

数据集	算法	AR±S.D. (%) ^{a)}	T -test ^{b)}	RMSEP ^{c)}
EEMF	PARAFAC	100.02±0.1	0.23	0.0012
	AWRCQLD	99.98±0.06	0.82	0.00034
	AWQLD	100.23±0.2	1.70	0.13
HPLC	PARAFAC	100.03±0.01	4.11	0.00033
	AWRCQLD	100.03±0.007	6.47	0.00019
	AWQLD	100.02±0.02	1.88	0.00018

a) AR=平均回收率(average recovery), S.D.=标准偏差(standard deviation); b) $T = (\bar{X} - \mu_0) / (S / \sqrt{n})$, 此处 $\bar{X}$ 是平均回收率, μ_0 为 100%; c) RMSEP: 预测均方根误差

图 2 采用 3 种算法分别处理实验数据所获得的归一化保留时间(a)、光谱(b)、浓度(c)和动力学-温度曲线(d). 其中+代表 AWQLD 的结果, 圆圈代表 AWRCQLD 的结果, 实线代表 PARAFAC 的结果; 深绿色点线代表真实的萘草胺保留时间和光谱曲线(网络版彩图)

表 3 萘草胺水解反应的回归等式、相关系数、速率常数和半衰期

	$T (^{\circ}\text{C})$	Regression equation	Correlation coefficient	Rate-constant (min^{-1})	Half-life (min)
AWRCQLD PARAFAC	50	$y=0.0278x+0.9363$	0.9973	0.0278	24.94
	60	$y=0.0696x+0.9142$	0.9979	0.0696	9.96
	70	$y=0.1682x+0.8792$	0.9986	0.1682	4.12
AWQLD	50	$y=0.0277x+0.9371$	0.9973	0.0277	25.02
	60	$y=0.0692x+0.9168$	0.9980	0.0692	10.02
	70	$y=0.1605x+0.9168$	0.9992	0.1605	4.32

图 3 采用 3 种算法分析实验数据集所获得的萘草胺的相对浓度的倒数的自然对数对时间($t-t_1$)所作的回归图. (a) 50°C; (b) 60°C; (c) 70°C. 其中+代表 AWQLD 的结果, 圆圈代表 AWRCQLD 的结果, 实线代表 PARAFAC 的结果(网络版彩图)

该实验数据的解析中, AWQLD 获得了更好的结果.

从所获得的动力学-温度曲线还可以获得不同温度下该反应的动力学信息. 图 3 给出了 3 种算法分析实验数据集所获得的萘草胺浓度的倒数的自然对数对时间所作的回归图. 在不同温度下, 两者都具有很好的线性关系. 所获得的动力学参数见表 3. 这些结果显示, NAP 的水解反应遵循伪一阶动力学模型. 分析结果表明, 在解析实际非五线性五维数阵时, 基于四线性成分模型的新方法给出了比较满意的结果, 且新提出的 AWQLD 算法给出了与 4-PARAFAC 和 AWRCQLD 相似甚至更好的结果.

5 结论

针对非五线性五维数阵, 本文提出了一种新的

处理方法. 该方法的基本思路是将非五线性五维数阵沿着不具有线性的一维铺展成四线性四维数阵, 然后采用基于四维平行因子分析(4-PARAFAC)、交替加权残差约束四线性分解(AWRCQLD)和新提出的交替加权四线性分解(AWQLD)算法的四维校正方法来处理. 通过解析 2 组模拟数据和 1 组实验数据, 发展的方法有效地分辨出了色谱或(和)光谱和动力学等有用信息, 且获得了比较满意的定量结果. 通过比较 3 种算法所获得的结果显示, 新发展的 AWQLD 算法给出了与 AWRCQLD 和 4-PARAFAC 算法相似甚至更好的结果. 这些结果说明, 发展的方法在处理非五线性五维数据时发挥了巨大的作用, 即保持了“二阶优势”, 又探索了“高阶优势”, 为处理非多线性多维数阵提供了一种新的有价值的分析方法.

补充材料

本文的补充材料见网络版 chem.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

参考文献

- Escandar GM, Faber NM, Goicoechea HC, Muñoz de la Peña A, Olivieri AC, Poppi RJ. *Trends Anal Chem*, 2007, 26: 752–765
- Olivieri AC, Escandar GM, Muñoz de la Peña A. *Trends Anal Chem*, 2011, 30: 607–617
- Fang DM, Wu HL, Ding YJ, Hu LQ, Xia AL, Yu RQ. *Talanta*, 2006, 70: 58–62
- Nägele E, Vollmer M, Hörth P, Vad C. *Expert Rev Proteomic*, 2004, 1: 37–46

- 5 Zeng ZD, Hugel HM, Marriott PJ. *Anal Bioanal Chem*, 2011, 401: 2373–2386
- 6 Quigley WW, Fraga CG, Synovec RE. *J Microcolumn Sep*, 2000, 12: 160–166
- 7 聂瑾芳, 吴海龙, 张云, 刘亚娟, 王建瑶, 李勇, 俞汝勤. 中国科学: 化学, 2011, 41: 1740–1749
- 8 Kiers HAL, ten Berge JMF, Bro R. *J Chemom*, 1999, 13: 275–294
- 9 Bahram M, Bro R. *Anal Chim Acta*, 2007, 584: 397–402
- 10 Tauler R. *Chemom Intel Lab Syst*, 1995, 30: 133–146
- 11 Reis MM, Gurden SP, Smilde AK, Ferreira MMC. *Anal Chim Acta*, 2000, 422: 21–36
- 12 Goicoechea HC, Olivieri AC. *Appl Spectrosc*, 2005, 59: 926–933
- 13 Maggio RM, Muñoz de la Peña A, Olivieri AC. *Chemom Intel Lab Syst*, 2011, 109: 178–185
- 14 Allegrini F, Olivieri AC. *Anal Chem*, 2012, 84: 10823–10830
- 15 Zhang XH, Wu HL, Wang JY, Tu DZ, Kang C, Zhao J, Chen Y, Miu XX, Yu RQ. *Food Chem*, 2013, 138: 62–69
- 16 Olivieri AC. *J Chemom*, 2005, 19: 615–624
- 17 Nahorniak ML, Cooper GA, Kim YC, Booksh KS. *Analyst*, 2005, 130: 85–93
- 18 Fu HY, Wu HL, Yu YJ, Yu LL, Zhang SR, Nie JF, Li SF, Yu RQ. *J Chemom*, 2011, 25: 408–429
- 19 Kim YC, Jordan JA, Nahorniak ML, Booksh KS. *Anal Chem*, 2005, 77: 7679–7686
- 20 Qing XD, Wu HL, Yan XF, Li Y, Ouyang LQ, Nie CC, Yu RQ. *Chemom Intel Lab Syst*, 2014, 132: 8–17
- 21 Bro R, Kiers HAL. *J Chemom*, 2003, 17: 274–286

A new alternating weighted quadrilinear decomposition algorithm with application for analysis of non-quinquelinear five-way data arrays

Xiangdong Qing^{1,2}, Hailong Wu^{1*}, Xihua Zhang¹, Yong Li¹, Huiwen Gu¹, Jin Wen², Xiangzhong Shen², Ruqin Yu¹

¹ State Key Laboratory of Chemo/Biosensing and Chemometrics; College of Chemistry and Chemical Engineering, Hunan University, Changsha 410082, China

² Department of Chemistry and Materials Science, Hunan University of Humanities, Science and Technology, Loudi 417000, China

*Corresponding author (email: hlwu@hnu.edu.cn)

Abstract: In the work, a new method based on firstly expanding before processing was proposed to investigate the non-quinquelinear five-way data produced in the study. The five-way data array was firstly rearranged into an expanded four-way data array along the non-linear way, then it was decomposed by four-way calibration based on four way-PARAFAC, alternating weighted residue constraint quadrilinear decomposition (AWRCQLD) and newly presented alternating weighted quadrilinear decomposition (AWQLD). Before processing real data, a simulated fluorescence data and a simulated chromatographic data were used to verify the proposed method. It was found that the results of AWQLD were similar to or even better than the ones of PARAFAC and AWRCQLD. These results indicated that the developed method possessed great potential to deal with non-multilinear multidimensional data.

Keywords: alternating weighted quadrilinear decomposition, four-way calibration, non-quinquelinear five-way data, PARAFAC, chromatographic

doi: 10.1360/N032015-00093