

荧光光度法中 Kalman 滤波法指示未知组分的研究

张晶玉* 杨金夫 任 英 高丽云**

(中国科学院长春应用化学研究所 长春 130022)

摘要 以苯丙氨酸、酪氨酸和色氨酸三组分体系的荧光光度分析为例,研究了 Kalman 滤波法指示未知组分的潜力. 根据新息序列的白噪声特性,建立指示函数,并确定指示函数的临界值,指示未知组分存在;并根据相对新息序列的形状可大致估计未知组分的荧光光谱的形状和峰位.

关键词 荧光光度法, Kalman 滤波法, 氨基酸

关于混合物中未知组分的指示,在 70 年代 Lawton 等^[1]提出了自模型曲线分辨法用于解析二元混合体系的光谱. 之后,较多作者也做出了各具特色的工作^[2,3]. 胡鑫尧等^[4]将主成分分析与光谱分峰技术相结合,提出了相关的计算机算法. 梁逸曾等^[5]提出了完全未知混合体系的纯物种光谱辨析.

Kalman 滤波法是一种有效的多变量分析方法,近年来发展迅速,在分析化学的不少领域已有深入研究和应用^[6~13],但大多用于已知重叠光谱的解析. 本工作以苯丙氨酸、酪氨酸和色氨酸三组分体系的荧光光度分析为例,考察了 Kalman 滤波法指示未知组分的潜力. 根据新息序列的白噪声特性建立指示函数,根据荧光信号的测量精度确定指示函数的临界值,不仅能指示未知组分的存在,而且还可估计未知组分光谱的大致峰位.

1 理论

关于 Kalman 滤波方法已在前文^[11]详细介绍. Kalman 滤波法的特点之一是利用其新息序列的白噪声特性来判断滤波结果的准确性. 新息序列 $in(\lambda_K)$ 是观测信号与估计信号之差:

$$in(\lambda_K) = X(\lambda_K) - S^T(\lambda_K)c(\lambda_K / \lambda_{K-1}) \quad (1)$$

其中 $X(\lambda_K)$ 是第 K 个测量波长 λ_K 处样品的观测信号, S 是灵敏度矩阵(上标 T 表示转置), c 是浓度矢量. 对于最佳滤波, $in(\lambda_K)$ 呈零均白噪声特性,其量值与测量过程的随机误差有关. 为便于根据测量精度来确定指示函数的临界值(见下节), $in(\lambda_K)$ 与观测信号之比为相对新息序列 $rin(\lambda_K)$:

$$rin(\lambda_K) = in(\lambda_K) / X(\lambda_K) \quad (2)$$

$rin(\lambda_K)$ 与 $in(\lambda_K)$ 有相同的统计特性,可以根据 $rin(\lambda_K)$ 是否呈零均白噪声特性来直接判断是否有未知组分存在. 但在实际应用中,用由下式定义的指示函数 F 作预检更为方便:

$$F = \frac{1}{n} \sum_{K=1}^n [rin(\lambda_K)]^2 \quad (3)$$

指示函数 F 的临界值由测量精度确定. F 小于临界值,表示滤波结果可靠; F 超过临界值,

就意味着滤波结果的可靠性值得怀疑, 需进一步察看 $rin(\lambda_K)$ 寻找原因. 若 $rin(\lambda_K)$ 在大于通带宽的波长范围内取同一符号 (正号或负号) 且绝对值大于其临界值, 说明样品中存在未知组分, 并可根椐 $rin(\lambda_K)$ 的形状来大概判断未知组分光谱的峰形和峰位. 需要注意的是, F 和 $rin(\lambda_K)$ 的临界值是不同的, 由方程 (2) 和 (3) 比较可知, F 的临界值应为 $rin(\lambda_K)$ 临界值的平方. 另外, 本滤波方法设定浓度的滤波初值为零[见文献[11]方程 (9)]. 这一设定必然导致 $rin(\lambda_K)$ 在滤波初期不呈零均白噪音特性, 所以在计算 F 时最初的 $rin(\lambda_K)$ 不考虑在内; $rin(\lambda_K)$ 在滤波初期超过临界值的波动也不应认为是未知组分的存在而引起的.

2 实验

2.1 主要仪器与试剂

岛津 RF-5000 型荧光分光光度计, 光谱带 $E_x = 5.0\text{nm}$, $E_m = 5.0\text{nm}$. 光谱扫描数据输入 NEC PC-8801B 计算机完成全部计算, 程序用 BASIC 语言编写.

酪氨酸、色氨酸 (生化试剂) 和苯丙氨酸均配成 0.1mg/mL 储备液, 储于冰箱中, 使用时根据需要适当稀释. 磷酸二氢钾-氢氧化钠缓冲溶液 (pH 7.4) 用以 4:5 体积 0.5mol/L NaOH 和 KH_2PO_4 溶液混合而成 (见文献[14]).

实验中所用其它试剂均为分析纯, 水为超纯水.

2.2 实验方法

取适量酪氨酸、色氨酸及苯丙氨酸储备液于 25mL 容量瓶中, 加入 2mL 缓冲溶液, 用水稀释至刻度. 用 1cm 石英池在 261.6nm 波长下激发样品, 在 $276.8 \sim 446.0\text{nm}$ 波段间隔 3.4nm 测定荧光强度. 对纯组分溶液, 扫描三次取平均值计算灵敏度; 样品溶液均扫描八次. 试剂空白作为一个组分考虑, 因此纯组分溶液和样品溶液的扫描光谱均无需扣除试剂空白.

3 结果与讨论

3.1 荧光光谱

图 1 给出了酪氨酸、色氨酸、苯丙氨酸及混合物在 $\lambda_{\text{ex}} = 261.6\text{nm}$ 激发条件下的荧光光谱.

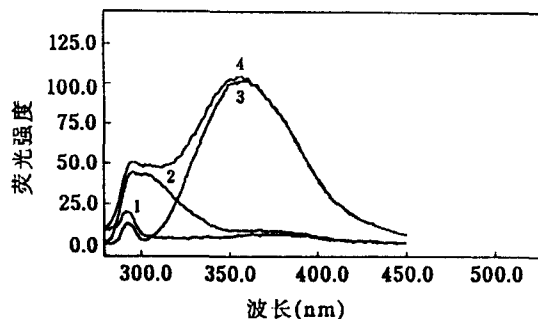


图 1 荧光光谱

1—苯丙氨酸 ($4.0\mu\text{g/mL}$); 2—酪氨酸 ($0.4\mu\text{g/mL}$); 3—色氨酸 ($0.2\mu\text{g/mL}$); 4—3 号混合样品 (见表 1)

3.2 新息序列及指示函数临界值的确定

新息序列的量值决于测量取过程的随机误差即测量精度. 实验中, 通常用测量信号的相

对标准偏差 (RSD) 来衡量测量精度. 不难看出, $rin(\lambda_K)$ 与 RSD 在数值上应当是相当的, 可以根据 RSD 的大小来确定 $rin(\lambda_K)$ 的临界值. 在我们的实验条件下, 大量试验表明, RSD 通常小于 10%. 所以, $rin(\lambda_K)$ 的临界值被确定为 0.1, 相应地 F 的临界值为 0.01. 若 F 大于 0.01, 表明滤波结果不一定可靠, 应进一步察看 $rin(\lambda_K)$ 寻找原因; 若在大于通常带宽的波长范围内 $rin(\lambda_K)$ 大于 0.1 或小于 -0.1, 表明样品中存在未知组分, $rin(\lambda_K)$ 绝对值的最大处就是未知组分光谱的大致峰位.

3.3 未知组分及其峰位的判断

以苯丙氨酸、酪氨酸和色氨酸三组分体系为例, 各组分的荧光光谱见图 1. 表 1 给出了滤波结果及相应的指示函数 F 的数值. 当模型正确时, 即建立系统模型时三组分均考虑在内时, 能得到很好的浓度估计值. 图 2 是分析 1 号样品时的新息序列图, 由于浓度初值假定为零, 所以 $rin(\lambda_K)$ 在滤波初期取值较大, 但很快减小, 在 ± 0.1 范围内变化, 且呈零均值白噪声特性. 除 9 号样品外, F 值均小于 0.01, 说明滤波结果可靠, 实际上无需再察看 $rin(\lambda_K)$ 本

表 1 模型中三组分全考虑及忽略某组分时的 Kalman 滤波结果

样品号	合成样品组成 ($\mu\text{g}/\text{mL}$)			模型中三组分全考虑				模型中忽略某组分				
	苯丙氨酸	酪氨酸	色氨酸	浓度估计值 ($\mu\text{g}/\text{mL}$)				被忽略组分	浓度估计值 ($\mu\text{g}/\text{mL}$)			
				苯丙氨酸	酪氨酸	色氨酸	F		苯丙氨酸	酪氨酸	色氨酸	F
1	4.0	0.50	0.20	3.87	0.50	0.20	0.001	酪氨酸	10.4	—	0.20	0.18
2	4.0	0.20	0.20	3.84	0.21	0.20	0.002	酪氨酸	6.50	—	0.20	0.08
3	4.0	0.40	0.20	3.93	0.39	0.20	7.8×10^{-4}	酪氨酸	9.07	—	0.20	0.15
4	4.0	0.20	0.40	3.76	0.19	0.39	0.005	酪氨酸	6.20	—	0.40	0.05
5	2.0	0.50	0.20	1.81	0.48	0.19	4.5×10^{-4}	苯丙氨酸	—	0.50	0.19	0.014
6	4.0	0.50	0.20	3.87	0.50	0.20	0.001	苯丙氨酸	—	0.53	0.20	0.05
7	4.0	0.20	0.40	3.76	0.19	0.39	0.005	色氨酸	-43	0.20	—	0.47
8	4.0	0.20	0.10	3.83	0.19	0.10	7.9×10^{-2}	色氨酸	-7.5	0.19	—	0.39
9	4.0	0.50	0	3.8	0.50	—	0.02	—	—	—	—	—

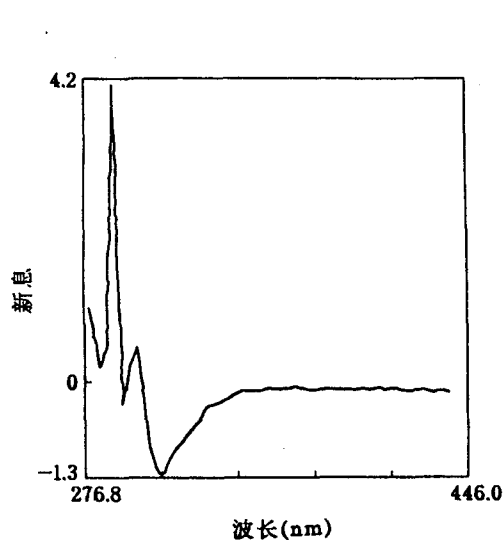


图 2 系统模型正确时分析表 1 中 1 号样品的新息序列

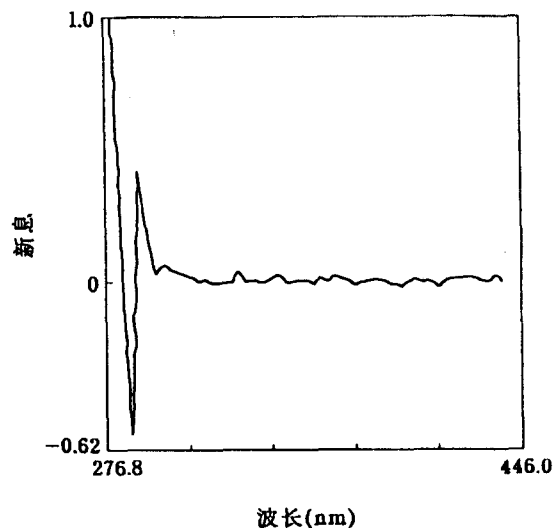


图 3 系统模型中忽略酪氨酸时分析表 2 中 1 号样品的新息序列

身. 另一方面, 若建立模型时忽略了某个组分, 把这个三组分体系当作二组分或单组分体系处理 (相当于样品中存在未知组分), 如表 1 所示, 所得结果的误差就很大, 指示函数 F 也都大于 0.01, 说明结果不可靠, 需要进一步检查 $rin(\lambda_K)$ 本身. 图 3 是以苯丙氨酸-色氨酸两组分建立模型时分析 1 号样品的新息序列. 很显然 $rin(\lambda_K)$ 不呈白噪音特性, 在 300nm 以后 $rin(\lambda_K)$ 均为负值, 且其绝对值均大于 0.1, 所以可以断定存在未知组分, 其光谱的大致峰位在 310nm, 峰形也与 $rin(\lambda_K)$ 形状相似. 与图 1 所示酪氨酸的光谱相比, 上述判断是符合实际的. 图 4 是以苯丙氨酸-酪氨酸两组分模型分析 7 号样品时的新息序列, 同样根据 $rin(\lambda_K)$ 的形状可判断未知组分的存在, 并确定其大致峰形及峰位; 所不同的是, 这是 $rin(\lambda_K)$ 均取正值. 经验表明, 当未知组分的光谱峰位位于滤波前期时, $rin(\lambda_K)$ 就呈现“负峰”; 而当未知组分峰位位于滤波后期时, $rin(\lambda_K)$ 则出现“正峰”.

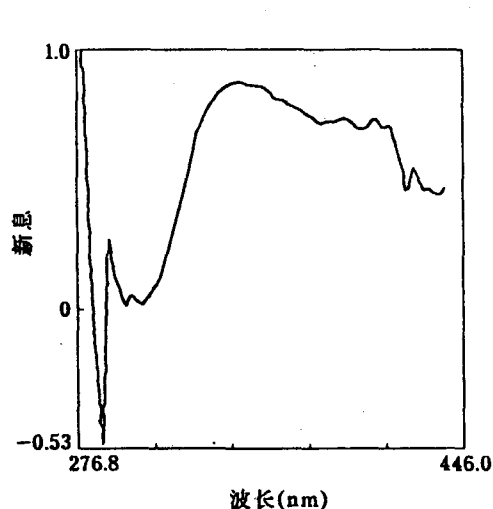


图 4 系统模型中忽略色氨酸时分析表 1 中 7 号样品的新息序列

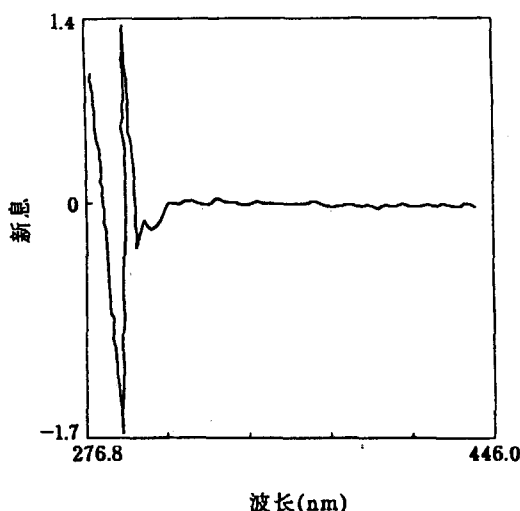


图 5 系统模型中忽略苯丙氨酸时分析表 1 中 5 号样品的新息序列

苯丙氨酸的光谱峰很小, 且处于滤波起始位置. 即使在这种情况下, 若系统模型中忽略苯丙氨酸, F 值仍然大于 0.01 (表 1 中 5、6 号样品), 表明滤波结果的可靠性值得怀疑, 但这时新息序列仍呈白噪音特性 (图 5), 难以判断 F 大于 0.01 是因未知组分而致. 这种情况实际上说明了本方法的局限性, 尤其当光谱峰位位于滤波初期时, 从 $rin(\lambda_K)$ 难以预测该组分

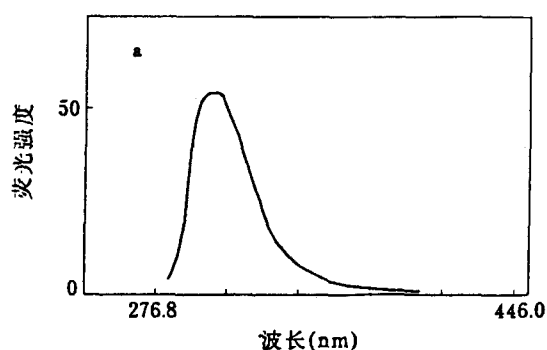


图 6a 苯丙氨酸 (4.0 μg/mL) 和酪氨酸 (0.5 μg/mL) 合成样品的荧光光谱

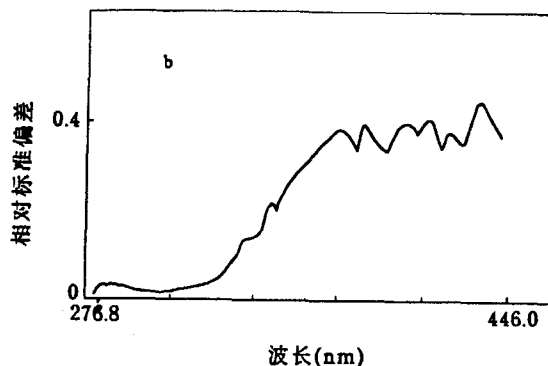


图 6b 与图 6a 相对应的强度测量 8 次的相对标准偏差

的存在, 更无法判断其大致峰形和峰位.

图 6 给出了一种特殊情况. 9 号样品只有苯丙氨酸和酪氨酸两组分, 如图 6a 所示, 在 345nm 以后荧光信号很低, 由此导致信号的相对标准偏差增大, 甚至高达 45% (见图 6b). 由表 2 最后一行可见, 这时苯丙氨酸和酪氨酸的浓度估计值是准确的, 但相应的 F 值仍大于 0.01. 进一步检查新息序列本身, 如图 6c 所示, $rin(\lambda_K)$ 在滤波后期绝对值增大, 但正负号相间, 仍呈白噪音特征. 所以不应认为有未知组分存在, 而应认为是测量噪音过大导致 F 大于 0.01.

综上所述, 正确判断未知组分的存在, 主要应当检验新息序列本身, 但指示函数 F 可以提供结果可靠性的预先性信息, 从而减少不必要的对 $rin(\lambda_K)$ 的检验.

3.4 本检验法与自相关分析法的比较

Kalman 滤波法具有能提供结果可靠性信息的能力. 为此提出了各种不同的判断标准来指示结果的可靠性^[10,15,16]. Seelling 和 Blount^[15]以及刘一鸣和俞汝勤^[17]对新息序列进行自相关分析来判断滤波结果的可靠性. 他们提出了所谓 NAC (归一化自相关系数) 标准. 当大于 5% 的 NAC 值超过置信范围时, 结果就不可靠. 我们试图用这一标准来指示未知组分, 结果如表 2 所示. 对于 1~3 号样品 NAC 标准给出了准确的判断; 对于 4、5 号样品 NAC 标准则给出了错误的判断结果. 原因可能是由于 NAC 标准要求测量的点很多^[16] (如数百个), 但这在实际分析中是困难的. 另一方面, 对于所有的样品, F 标准都给出了准确的指示结果. F 标准不受样品测量点的限制, 临界值可以根据实际测量精度决定. 所以从我们实验结果看在指示未知组分时, F 标准比 NAC 标准更有效.

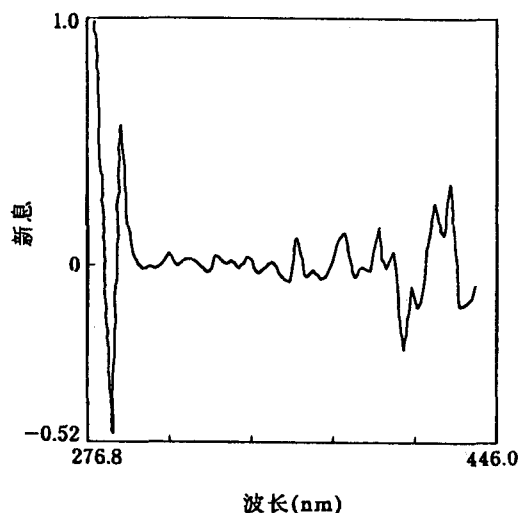


图 6c 对应图 6a 的新息序列

表 2 F 和 NAC 标准指示结果能力的比较

样品号	合成样品组成 ($\mu\text{g}/\text{mL}$)			模型中被忽略组分	NAC 超过置信范围的比率 (%)	F
	苯	酪	色			
1	4.0	0.5	0.2	—	1	0.001
2	4.0	0.2	0.1	色氨酸	24	0.39
3	4.0	0.5	0.2	酪、色氨酸	26	0.47
4	4.0	0.5	0.2	苯丙氨酸	0	0.05
5	4.0	0.5	0.2	酪氨酸	1	0.18

参考文献

- 1 Lawton, W. H.; Sylvester, E. A., *Technometrics*, **1971**, 13, 617.
- 2 Borger, O. S.; Kowalski, B. R., *Anal. Chim. Acta*, **1985**, 174, 1.
- 3 Meister, A., *Anal. Chim. Acta*, **1984**, 161, 149.
- 4 胡鑫尧, 汪国柄, 科学通报, **1985**, 30, 398.
- 5 梁逸曾, 谢玉珑, 俞汝勤, 化学学报, **1991**, 49, 394.
- 6 Poullisse, H. N. J., *Anal. Chim. Acta*, **1979**, 112, 361.
- 7 Didden, C. B. M.; Poullisse, H. N. J., *Anal. Lett.*, **1980**, 13, 921.
- 8 李志良, 李梦龙, 石乐明, 俞汝勤, 中国稀土学报, **1989**, 7(4), 65.
- 9 Brown, T. F.; Brown, S. D., *Anal. Chem.*, **1981**, 53, 1410.
- 10 Van Veen, E. H.; de Loos-Vollebregt, M. T. C., *Anal. Chem.*, **1991**, 63, 1441.
- 11 杨金夫, 朴哲秀, 曾宪津, 陈新海, 张卓勇, 关秋获, 化学学报, **1992**, 50, 1017.
- 12 杨金夫, 朴哲秀, 曾宪津, 陈新海, 张卓勇, 关秋获, 化学学报, **1993**, 51, 283.
- 13 Yang Jinfu, Piao Zhexiu, Zeng Xianjin, Zhang Zhouyong, Cheng Xinhai, *Spectrochim. Acta*, **1992**, 47B, 1055.
- 14 李耀群, 许金钩, 黄贤智, 药物分析杂志, **1991**, 11, 323.
- 15 Seelling, P. F.; Blount, H. N., *Anal. Chem.*, **1979**, 51, 327.
- 16 Mehra, R. K., *IEEE Trans. Autom. Control*, **1970**, 15, 175.
- 17 Liu Yi-Ming; Yu Ru-Qin, *Talanta*, **1988**, 35(9), 707.

A Study on the Indication of Unexpected Components by Kalman Filtering in Fluorospectrophotometry

Zhang, Jing-Yu* Yang, Jin-Fu Ren, Ying Gao, Li-Yun

(Changchun Institute of Applied Chemistry, The Chinese Academy of Sciences, Changchun, 130022)

Abstract The potential of Kalman filtering for indication of unexpected components in a mixture was experimentally evaluated by taking the spectrofluorimetric analysis of the tricomponent system of phenylalanine, tryptophen and tyrosine as an example. According to the whiteness of the innovation sequence, a criterion, which was defined as the mathematical expectation of the sequence of the innovation normalized to the measured signals (normalized innovations), was developed to flag for unexpected emissions. Insights into the relationship between the innovation sequence and the measurement noises revealed that the normalized innovations are numerically equivalent to the relative standard deviations (*RSD*) of the measured signals and the value of the criterion therefore should be of the order of the square of the *RSD*. The threshold of the criterion was thus determined experimentally. An omission of any contributed component in the model destroyed the whiteness of the innovation sequence and hence made the criterion higher than the threshold. Thus, an unexpected component was indicated. Furthermore, the peak position and the shape of the emission from the unexpected component were also approximately estimated by checking the structure of the normalized innovation sequence.