EESTI NSV TEADUSTE AKADEEMIA TOIMETISED. 31. KÖIDE
KEEMIA. 1982, NR. 3

ИЗВЕСТИЯ АКАДЕМИИ НАУК ЭСТОНСКОЙ ССР. ТОМ 31
ХИМИЯ. 1982, № 3

УДК 543.422.6

Ü. HALDNA, A. MURSHAK

FACTOR ANALYSIS APPLIED TO THE UV ABSORPTION SPECTRA

1. Digital simulation models for algorithms tests

(Presented by O. Eisen)

A number of cases exist where a measured set of UV absorption spectra shows a variability due to two or more simultaneous spectral processes. The factor analysis (FA) has often been applied to these sets of spectral data in order to obtain information about the individual processes causing the observed changes in spectra [1–4]. The basic FA procedure involved is the principal component analysis (PCA), which yields a set of abstract factors [5]. As a rule, the abstract factors describe the matrix of experimental data well but have no physical counterparts [5]. It should be noted that always when the PCA has been applied to UV absorption spectra, a special algorithm given by Simonds [1] has been used. This particular PCA method has been claimed to yield physically meaningful factors (vectors and their coefficients [2–4]). The successful use of the first vectors coefficients obtained by this method in protonation studies [2–4] somewhat supports these claims. Indeed, the medium dependence of the first vector coefficient obtained by the Simonds' method on UV spectra strongly resembles that of the corresponding ^1H and ^{13}C NMR chemical shifts [6]. On the other hand, the application of Simonds' version of PCA [1] to the amides protonation was unsuccessful: the coefficient of the first vector did not reach a stable final value in the sulphuric acid solutions where the base is practically in BH^+ form; this coefficient continues to vary slowly with increasing acid concentration [7].

The Simonds' method makes use of the mean corrected data factors obtained by the Simonds' algorithm [7]. If this method [1] yields only one solution to the problem from a number of possible ones, there remains little ground for the belief in the physical meaning of the factors obtained. The goal of this paper is to show how an artificial data matrix may be constructed in order to check whether the Simonds' algorithm [1] is able to reproduce in a two-factor space the initial vectors and their coefficients used for generating the data matrix decomposed.

The vectors obtained from the decomposition of a data matrix by PCA are orthogonal [5], and so the initial vectors must be. If only two factors are involved, the orthogonality of the vectors A_1 and A_2 may be expressed as follows [8]:

$$\sum_{i=1}^{i=k} a_{1,i} \cdot a_{2,i} = 0. \qquad (1)$$

Choosing two arbitrary vectors A_1 and G, the last one can easily be converted into A_2, this being orthogonal to A_1:

$$a_{2,i}=g_i+t \qquad (2)$$

where the constant t added to each g_i is

$$t=-\sum_{i=1}^{i=k} a_{1,i}\cdot g_i \Big/ \sum_{i=1}^{i=k} a_{1,i}. \qquad (3)$$

The vectors A_1 and A_2 have k elements each, where k is the number of wavelengths considered.

The Simonds' method makes use of the mean corrected data matrix, i. e.,

$$e_{j,i}=\bar{e}_i+c_{1,j}\cdot a_{1,i}+c_{2,j}\cdot a_{2,i} \qquad (4)$$

where $e_{j,i}$ is the molar extinction coefficient for the solution j at the wavelength i, $c_{1,j}$ and $c_{2,j}$ are the coefficients of the vectors a_1 and a_2, respectively, for the solution j, and $\bar{e}_i$ is the mean value of e_{ij} at the wavelength i:

$$\bar{e}_i=\sum_{j=1}^{j=l} e_{j,i}/l. \qquad (5)$$

Eq. (4) may be written for all solutions ($j=1, 2, \ldots, l$) at an arbitrary wavelength i. Taking the sum over all these equations above (4), we find that

$$a_{1,i}\sum_{j=1}^{j=l} c_{1,j}+a_{2,i}\sum_{j=1}^{j=l} c_{2,j}=0. \qquad (6)$$

Eq. (6) sets up the conditions to be met by the $c_{1,j}$ and $c_{2,j}$ values. We can write Eq. (6) for each wavelength ($i=1, 2, \ldots, k$) and multiply them by the respective $a_{2,i}$ values ($i=1, 2, \ldots, k$). Taking the sum over all these equatious, we obtain

$$\left(\sum_{i=1}^{i=k} a_{1,i}\cdot a_{2,i}\right)\left(\sum_{j=1}^{j=l} c_{1,j}\right)+\left(\sum_{i=1}^{i=k} a_{2,i}^2\right)\left(\sum_{j=1}^{j=l} c_{2,j}\right)=0. \qquad (7)$$

Due to Eq. (1) the first term in Eq. (7) is zero. The sum of $a_{2,i}^2$-s cannot be zero. So we can conclude that

$$\sum_{j=1}^{j=l} c_{2,j}=0. \qquad (8)$$

An analogous route may be used to show that

$$\sum_{j=1}^{j=l} c_{1,j}=0. \qquad (9)$$

Any arbitrary set of real numbers $d_1, d_2, \ldots, d_l$ can be converted into the respective $c_{1,j}$ ($c_{2,j}$) values:

$$c_{n,j}=d_j+p \qquad (10)$$

where the constant p is

$$p=-\sum_{j=1}^{j=l} d_j/l. \qquad (11)$$

It should be noted that the Simonds' method [1] usually yields $1 > c_{n,j} > -1$.

In order to construct the data matrix by Eq. (4) one needs to add an arbitrary $\bar{e}_{ii}$ value to the sum of products at each wavelength considered.

REFERENCES

1. Simonds, J. L. Application of characteristic vector analysis to photographic and optical response data. — J. Opt. Soc. Amer., 1963, v. 53, N 8, p. 968—974.
2. Edward, J. T., Wong, S. C. Ionization of carbonyl compounds in sulfuric acid. Correction for medium effects by characteristic vector analysis. — J. Amer. Chem. Soc., 1977, v. 99, N 13, p. 4229—4232.
3. Cox, R. A., Smith, C. R., Yates, K. The excess acidity method. The basicities and rates and mechanisms of enolization, of some acetophenones and acetone, in moderately concentrated sulfuric acid. — Canad. J. Chem., 1979, v. 57, N 22, p. 2952—2959.
4. Zalewski, R. I. Adaption of characteristic vector analysis and titration curve analysis for calculations for pK_{BH^+} from ultraviolet-visible spectral data. — J. Chem. Soc., Perkin Trans. II, 1979, p. 1637—1639.
5. Malinowski, E. R., Howery, D. G. Factor Analysis in Chemistry. New York, Chichester, Brisbane, Toronto, 1980, p. 12, 26, 126.
6. Haldna, Ü., Mursbak, A., Kuura, H. Characteristic vector analysis applied to the protonation equilibria. 1. Acetone and 2-butanone in aqueous sulfuric and perchloric acids. — Organic Reactivity, 1980, v. 19, N 4 (64), p. 384—399.
7. Cox, R. A., Yates, K. Medium effects in protonation equilibrium studies. Accurate acidity constants using the excess acidity method. — Canad. J. Chem., 1981, v. 59, N 11, p. 1560—1567.
8. Überla, K. Faktorenanalyse. Heidelberg — New York, 1977, p. 37, 45.

Academy of Sciences of the Estonian SSR,
Institute of Chemistry

Received
Nov. 18, 1981

Ü. HALDNA, A. MURSAK

ULTRAVIOLETSETE ABSORPTSIOONISPEKTRITE FAKTORANALÜÜS

1. Andmemaatriksite konstrueerimine algoritmi testi jaoks

Artiklis on esitatud digitaalse testmudeli koostamise meetod faktoranalüüsi Simondsi algoritmi kontrollimiseks.

Ю. ХАЛДНА, А. МУРШАК

ПРИМЕНЕНИЕ ФАКТОРНОГО АНАЛИЗА К УЛЬТРАФИОЛЕТОВЫМ СПЕКТРАМ ПОГЛОЩЕНИЯ

1. Составление дигитальной модели для испытания алгорифмов

Рассмотрен случай зависимости измеренных УФ-спектров поглощения от двух или более одновременно протекающих спектральных процессов. Для их обработки применен метод факторного анализа. Предложен способ составления дигитальной модели проверки алгорифма Саймондса, позволяющий оценить факторы, определяющие наблюдаемые изменения в спектрах поглощения.