Adaptation of Characteristic Vector Analysis and Titration Curve Analysis for Calculations of pK_{BH^+} from Ultraviolet–Visible Spectral Data

By **Romuald I. Zalewski**, Division of General and Analytical Chemistry, Academy of Economics, Poznań, Poland

Equilibrium constants for protonation, K_{BH^+}, have been extensively investigated by u.v.–visible spectrophotometry in sulphuric acid. To overcome difficulties in calculating the pK_{BH^+} values, experimental data were used for estimating two main orthogonal characteristic vectors contributing to the total variability in absorption. The largest amount of variability was associated with the spectral change accompanying protonation and the reconstituted absorption curves were calculated. A new, one-step computer procedure is proposed for calculation of pK_{BH^+} by the use of characteristic vector analysis and titration curve analysis.

It is well known that, despite many successful methods for calculating pK values for organic bases from spectrophotometric data,[1-4] medium effects in concentrated acids (hydration and solvation by species existing in more concentrated acids,[5] overlapping of different spectral bands, and dehydration) can still cause difficulties. To overcome these difficulties Zalewski and Dunn [4] have used a least-squares computer program for titration curve analysis (TCA).

This paper deals with a less familiar approach, that of characteristic vector analysis (CVA), which has been described in detail,[6] summarized by Simonds with reference to optical data,[7] and used by Reeves [8] and Woud [9] to investigate problems dealing with protonation.

In spite of fairly encouraging results the CVA procedure has not been used extensively. There is considerable interest in pK_{BH^+} constants owing to their utility in structural studies and studies of reaction rates. In connection with our pK investigations,[10] we examined the utility of CVA for the treatment of spectral data. Since we obtained encouraging results [11] we now suggest a new treatment of spectral data by the use of CVA and TCA in a one-step procedure.

METHODS

Characteristic Vector Analysis.—The method can be used for estimating the number of independent orthogonal vectors contributing to the total variation observed in a family of absorption curves, and for treatment of a family of absorption curves for the largest amount of variability. The absorptivities A_i taken at n wavelengths, $i = 1,2 \ldots n$, constitute a 1-row, n-column data vector. For m acidities, the m vectors can be arranged to form an $m \times n$ data matrix. The CVA procedure involves the computation of the variance–covariance matrix from the data matrix of the absorbance vectors.

The characteristic vectors V_n are uniquely determined for a family of absorption curves and apply to all data vectors. They form sets of n numbers. Mathematically, the absorption curves may be represented by equation (1), where the Y values are the amounts of characteristic

$$A_{m,i} = \bar{A}_{m,i} + Y_1V_{1,i} + Y_2V_{2,i} + \ldots + Y_kV_{k,i}\ (k \leqslant n) \quad (1)$$

vectors which must be added to the mean absorbance vector $\bar{A}_{m,i}$ in order to obtain the sample vector. The Y values vary from one absorption curve to another.

According to equation (1), the characteristic vectors $V_{m,i}$, the mean absorbance vector $\bar{A}_{m,i}$, and the Y values give sufficient information to reconstruct the absorbance at wavelength i for each acidity m.

The number of characteristic vectors required to represent all the variability among a family of spectral curves will be equal to or less than n. The power of the CVA method arises, however, from the fact that there is a high probability that the greater part of the variability within a family of curves may be explained by the use of only a few characteristic vectors.[7] Reeves [8] reported only two characteristic vectors, the first of which always accounted for 96% or more of the total variability. As characteristic vectors are orthogonal, they represent statistically independent types of variability.

It was assumed as a first approximation that the large first vector can be associated with the spectral change accompanying protonation.[8,9] The reconstituted absorption curves will reflect the acid–base equilibria.

Titration Curve Analysis.—This method [4] has often been used to estimate pK_{BH^+} values for a wide variety of carbonyl bases.[10] Following the general procedure for basicity determination, a wavelength, λ_B, characteristic of the base, and a wavelength, λ_{BH^+}, characteristic of the ion may be chosen. A plot of absorbance *vs.* acidity function gives a titration curve. If the total concentration of the base, free as well as ionised, is [C], then the measured absorbance, A, in a solution where the base is only partly ionised, can be shown to be given by equation (2). Equation (2) is an adequate representation of the

$$A_i = [C](A_BK_{BH^+} + h_0A_{BH^+})/(K_{BH^+} + h_0) \quad (2)$$

titration curve. Since A, h_0, and [C] are measurable quantities, equation (2) contains three unknowns: K_{BH^+}, A_{BH^+}, and A_B, which in general might be approximated experimentally. It was shown previously that A_B and A_{BH^+} may vary across the range of acidities where both species contribute to the absorbance. If the variability is not dramatic in the range of acidities given by $pK_{BH^+} \pm 2H_x$ units, a least-squares computer program can be used to solve equation (2), taking estimated values of unknowns, and ordinates of the points forming the titration curve.

New Procedure for Calculation of pK_{BH^+} *from Spectral Data.*—The new procedure was tested with many aromatic and alicyclic carbonyl bases and we now discuss cinnamic acid as a detailed example. Figure 1(a) shows an experimental family of curves for cinnamic acid in

TABLE

pK_{BH^+} values calculated for wavelengths i and H_A and H_0 acidity function

	H_A corrected		H_0 corrected		H_A uncorrected		H_0 uncorrected	
i	$-pK_{BH^+}$	Slope	$-pK_{BH^+}$	Slope	$-pK_{BH^+}$	Slope	$-pK_{BH^+}$	Slope
1	4.116 ± 0.014	1.002	6.442 ± 0.139	0.783	3.691 ± 0.017	1.184	5.635 ± 0.219	0.634
2	4.117 ± 0.014	1.002	6.412 ± 0.154	0.751	3.842 ± 0.051	1.063	5.838 ± 0.203	0.696
3	4.117 ± 0.018	1.018	6.392 ± 0.161	0.739	3.959 ± 0.040	1.014	6.213 ± 0.219	0.808
4	4.130 ± 0.040	1.054	6.405 ± 0.157	0.748	4.266 ± 0.065	0.987	6.586 ± 0.150	0.657
5	4.118 ± 0.013	1.005	6.449 ± 0.134	0.780	4.575 ± 0.089	1.012	7.273 ± 0.168	0.809
8	4.119 ± 0.013	1.002	6.431 ± 0.146	0.767	3.446 ± 0.075	1.004	4.842 ± 0.199	0.867
9	4.114 ± 0.013	1.002	6.377 ± 0.177	0.709	3.831 ± 0.028	1.006	5.712 ± 0.180	0.706
10	4.113 ± 0.013	1.002	6.311 ± 0.185	0.698	4.092 ± 0.032	1.003	6.165 ± 0.181	0.702
11	4.113 ± 0.014	1.001	6.307 ± 0.186	0.697	4.231 ± 0.035	1.004	6.421 ± 0.165	0.846
12	4.114 ± 0.014	1.001	6.312 ± 0.184	0.699	4.320 ± 0.032	0.999	6.663 ± 0.160	0.768
13	4.114 ± 0.013	1.001	6.332 ± 0.178	0.707	4.485 ± 0.045	0.986	6.916 ± 0.153	0.691
14	4.112 ± 0.013	1.001	6.436 ± 0.144	0.766	4.615 ± 0.038	0.982	7.238 ± 0.142	0.683

40—96% sulphuric acid solution. Figure 1(b) shows a family of curves calculated by characteristic vector analysis,* the first vector accounting for 94.93% of the total variability. For all the substances studied [11] the first vector always accounts for 94—96%, and the sum of the first and the second characteristic vectors for >99% of the total variability. Calculations of further vectors is both time-consuming and ineffective, so it was not performed.

The family of curves calculated with the first vector may be associated with the spectral change accompanying protonation. The family of curves has an isosbestic point and the $\lambda_{max.}$ value and absorptivity differ slightly from the experimental data. The advantages of the corrected family of curves over the experimental ones are shown by the pK_{BH^+} values calculated for all the wavelengths i and H_A and H_0 acidity functions. The results of TCA calculations are summarized in the Table. Values of pK_{BH^+} in the first cases (with correction) are *ca.* −4.11 or *ca.* −6.35 whereas in the second cases (without correction) deviations are too high to be acceptable. The slope of the plot of log[B]/[BH^+] *vs.* acidity function is near unity for H_A, and thus this acidity function is appropriate for description of the protonation of cinnamic acid.

There are problems in proceeding from the two-step procedure of CVA followed by TCA, to a one-step procedure. The basic problems are: (i) how to choose characteristic wavelengths, and (ii) which data and in which form for A_B, A_{BH^+}, and pK_{BH^+} should be used as input.

On the basis of many curves [11] obtained for both vectors (*e.g.* Figure 2 for cinnamic acid) it was assumed that for λ_B, the first and the second vectors should reach minimum and zero values, respectively, whereas for

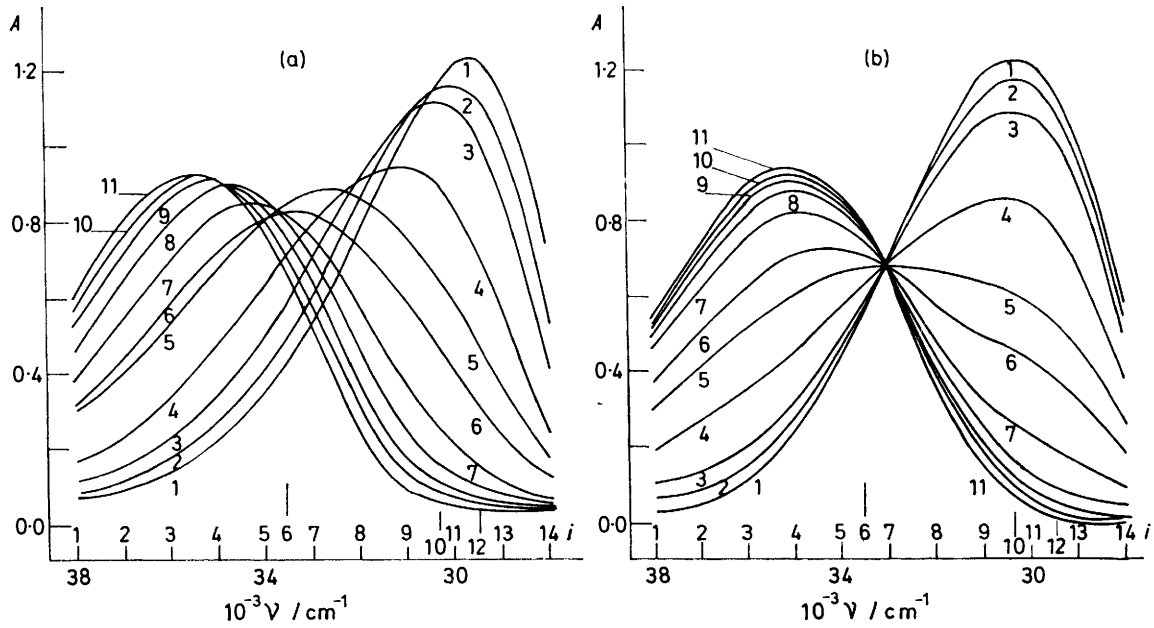

FIGURE 1 Absorbance of cinnamic acid as a function of sulphuric acid concentration (1), 95.0; (2), 88.5; (3), 84.1; (4), 78.6; (5), 74.2; (6), 70.6; (7), 65.7; (8), 60.1; (9), 55.5; (10), 47.7; and (11), 39.2%, (a) for experimental data, and (b) calculated by CVA with the $V_{1,i}$ accounting for 94.93% of the total variability

* The basic program for CVA was kindly provided by Dr. J. T. Edward, McGill University, and was adopted for operation on an ODRA 1205 computer in FORTRAN IV ICL.

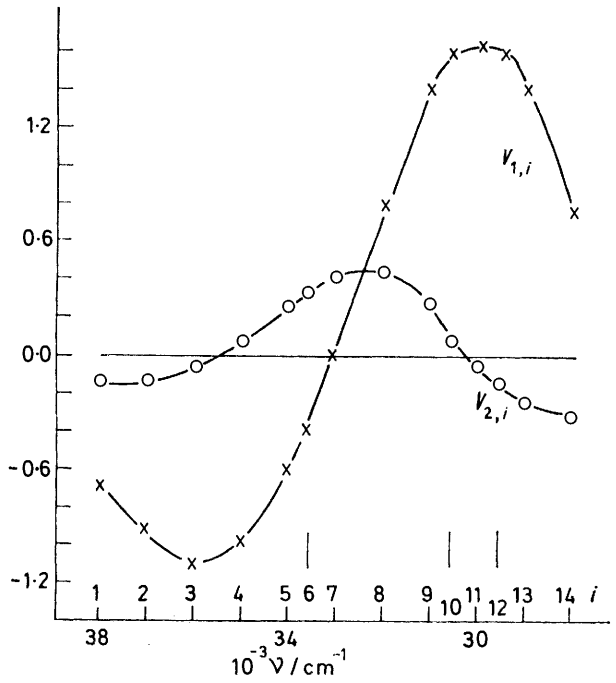

FIGURE 2 Characteristic vectors $V_{1,i}$ and $V_{2,i}$ *vs.* frequency

λ_{BH^+} the first and the second vectors should reach maximum and zero values, respectively. It was observed that for the wavelengths λ_B and λ_{BH^+} extension of the titration curve leads to its maximum value being obtained. On the basis of this observation the criteria C_1—C_3 were established (for $i = 1,2 \ldots n$), and used in the computer program. Since the number of wavelengths (columns), n, in our program was $\leqslant 25$ there were cases when no column in the reconstituted data matrix would follow all three criteria. For this reason all the columns following at least one of the conditions were used in the pK_{BH^+} calculations.*

$$C_1\ i = \max_i .\ V_{1,i} \text{ and/or } i = \min_i .\ V_{1,i}$$

$$C_2\ i = \min_i .\ (V_{2,i} - 0)$$

$$C_3\ i = \max_i .\ (A_{1,i} - A_{m,i})$$

The first and the last elements of each column were established as A_B and A_{BH^+}, respectively. According to the requirements of the program, the range of the magnitude of pK_{BH^+} was roughly estimated from experimental absorption curves by a graphical method.

The program is written in FORTRAN for a 25 × 25 data matrix and can be expanded for a larger number of columns. Program and operating instructions are available on request.

[8/1828 *Received*, 18*th October*, 1978]

REFERENCES

[1] L. A. Flexer, L. P. Hammett, and A. Dingwall, *J. Amer. Chem. Soc.*, 1935, **57**, 2103.
[2] R. Stewart and M. Granger, *Canad. J. Chem.*, 1961, **39**, 2508.
[3] Ch. T. Davies and T. A. Geisman, *J. Amer. Chem. Soc.*, 1959, **76**, 3507.
[4] R. I. Zalewski and G. E. Dunn, *Canad. J. Chem.*, 1969, **47**, 2263.
[5] R. A. Cox, *J. Amer. Chem. Soc.*, 1974, **96**, 1059.
[6] T. W. Anderson, 'Introduction to Multivariate Analysis,' Wiley, New York, 1958.
[7] J. L. Simonds, *J. Opt. Soc. Amer.*, 1963, **53**, 968.
[8] R. L. Reeves, *J. Amer. Chem. Soc.*, 1966, **88**, 2240.
[9] H. Woud, Ph.D. Thesis, McGill University, Montreal, 1974.
[10] R. I. Zalewski, *Bull. Acad. polon. Sci., Sér. Sci. Chem.*, 1970, **18**, 353; 1971, **19**, 351; 1972, **20**, 853; R. I. Zalewski, Z. Geltz, V. D. Orlov, J. Borovoj, and W. F. Lavrousin, *Zhur. obshchei Khim.*, 1978. **48**, 1802.
[11] E. Dawidziak, J. Niedbata, and R. I. Zalewski, *Bull. Acad. Polon. Sci. Sér. Sci, Chem.*, 1979, **27**, 45; R. I. Zalewski and Z. Geltz, *ibid.*, submitted for publication; R. I. Zalewski, V. D. Orlov, J. Borovoj, and W. F. Lavrousin, *Zhur. obshchei Khim.*, submitted for publication.

* The maximum and the minimum numbers of columns used in the pK calculation are 6 (3 around λ_B, and 3 around λ_{BH^+}) and 2, respectively. The pK_{BH^+} values should be independent of the column used.