Determination of Equilibrium Constants by Chemometric Analysis of Spectroscopic Data

Mikael Kubista,* Robert Sjöback, and Bo Albinsson

Department of Physical Chemistry, Chalmers University of Technology, S-412 96 Gothenburg, Sweden

We describe a chemometric method to determine equilibrium constants with high accuracy from spectroscopic titrations. Knowledge of component spectra is not required for the analysis, and titrations can be analyzed even when the titration end points are not reached. In fact, the component spectra are determined in the analysis. The analysis is based on a decomposition of the recorded spectra into a product of target and projection matrices using NIPALS. The matrices are then rotated to give the correct concentrations and spectral profiles of the components utilizing the functional form of the equilibrium expression. The output of the analysis is the equilibrium constant and the component spectral profiles. The analysis is highly accurate and can be performed on a standard personal computer. As examples, we determine the protolytic constant for the equilibrium between the anionic and dianionic forms of fluorescein in aqueous solution and the dimerization constant of benzoic acid in *n*-heptane.

INTRODUCTION

Spectroscopic methods are in general highly sensitive and are as such suitable for studying chemical equilibria in solution. When the components involved in the chemical equilibrium have distinct spectral responses, their concentrations can be measured directly, and the determination of the equilibrium constant is trivial. However, in many cases, the spectral responses of two and sometimes even more components overlap considerably and the analysis is no longer straightforward. The common approach has been single-point measurements at a wavelength where one component dominates the spectral response and the contributions from other components are neglected. The single-point measurements are usually made at the edge of an absorption band, where the spectral overlap is least. However, here the spectral response is much lower than at the absorption maximum, and the noise level may be considerable.

By chemometric methods (see, for example, ref 1) one can analyze whole spectra, thereby utilizing all spectral information. The approach is superior to any single-point measurement since several hundreds of data points per spectrum can be treated simultaneously. A problem with chemometric methods, though, is that they primarily provide nonphysical results, which may be difficult and sometimes impossible to transform into something physically meaningful. The special case with two overlapping spectral components was discussed by Lawton and Sylvestre,[2] who provided a graphical method to limit the solutions to those where spectra and concentrations are all positive. Their approach has been extended to more components,[3] which, however, requires rather advanced programming. A problem with the nonnegative criterion is that the analysis is very sensitive to noise, which can always be negative, and methods to relax this criterion have been developed.[4] Another disadvantage is that spectra that may be negative, such as dichroism and difference spectra, cannot be analyzed. Finally, we recall that these methods do not provide a unique answer, but merely limit the number of solutions to those with nonnegative elements.

For systems where two correlated spectra can be recorded on each sample, we have developed a method based on Procrustes rotation to determine the component spectral profiles and concentrations in all samples.[5] The Procrustes rotation approach is highly accurate[6] and has been successfully applied to various systems (see, for example, ref 7). However, two suitably correlated spectra cannot always be recorded on the samples, and additional information is required to determine the component spectral profiles and concentrations.

For components in a chemical equilibrium, this additional information may be the functional form of the equilibrium expression. In this paper we describe how knowledge of the equilibrium expression can be used to transform the result of a chemometric analysis into concentrations and spectral profiles of the components. The method is particularly powerful for systems of lower complexity such as protolytic equilibria, dimerization reactions and complex formation between two molecules.

THEORY

Spectral responses are in general linear, and the spectrum of a component mixture is the sum of the contributions from the components (scalars are shown in lower case letters, vectors in bold lower case letters, and matrices in bold upper case letters. Transpose of a matrix is indicated by ′):

$$\mathbf{a}(\lambda) = \sum_{i=1}^{r} c_i \mathbf{v}_i(\lambda) \qquad (1)$$

where $\mathbf{a}(\lambda)$ is the sample spectrum, r is the number of components, c_i is the concentration of component i, and $\mathbf{v}_i(\lambda)$ is its spectrum normalized to unit concentration. In a titration experiment a series of spectra are recorded:

$$\mathbf{a}_j(\lambda) = \sum_{i=1}^{r} c_{ij} \mathbf{v}_i(\lambda) \quad (j = 1, n) \quad \text{or} \quad \mathbf{A} = \mathbf{CV} \qquad (2)$$

where $\mathbf{a}_j(\lambda)$ is the spectrum of sample, j, c_{ij} is the concentration of component i in sample j, and n is the number of samples. In matrix notation $\mathbf{A}$ is an $n \times m$ matrix containing the n recorded spectra, each digitized into m data points, as rows. $\mathbf{C}$ is an $n \times r$ matrix containing the concentration profiles of the components as columns, and $\mathbf{V}$ is an $r \times m$ matrix containing the normalized spectral responses as rows. The problem is to determine the c_{ij}'s.

* Author to whom correspondence should be addressed.

(1) Mardia, K. V.; Kent, J. T.; Bibby, J. M. *Multivaiate Analysis*; Academic Press: London, 1979.
(2) Lawton, W.; Sylvestre, E. *Technometrics* 1971, *13*, 617-633.
(3) Borgen, O.; Davidsen, N.; Mingyang, Z.; Oyen, O. *Mikrochim. Acta* 1986, *2*, 63-73.
(4) Burdick, D.; Tu, X. *J. Chemomet.* 1989, *3*, 431-441.
(5) Kubista, M. *Chemom. Intell. Lab. Syst.* 1990, *7*, 273-279.
(6) Scarminio, I.; Kubista, M. *Anal. Chem.* 1993, *65*, 409-416.
(7) Chiesa, M.; Domini, I.; Samori, B.; Eriksson, S.; Kubista, M.; Nordén, B. *Gazz. Chim. Ital.* 1990, *120*, 667-670.

0003-2700/93/0365-0994$04.00/0

If the spectral responses of the pure components, the $\mathbf{v}_i(\lambda)$'s, are known, which is the case when the titration end points are reached, the determination of the c_{ij}'s is trivial. However, if they are not known, the information in the recorded spectra is not sufficient to determine the concentrations.

Matrix **A** can be decomposed into an orthonormal basis set using, for example, NIPALS (see Appendix):

$$\mathbf{A} = \mathbf{TP}' + \mathbf{E} \tag{3}$$

T has the same dimensions as **C**; its columns are referred to as target vectors, and they are orthogonal linear combinations of the columns in **C**. **P**′ has the same dimensions as **V**; its rows are referred to as projection vectors, and they are orthonormal linear combinations of the rows in **V**. Matrix **E** is a residual matrix containing the difference between the experimental data and the features accounted for by the first principle components (the number of principle components should be the number of spectroscopically distinguishable components in the samples). Provided that the spectral responses of the components dominate the recorded spectra (i.e., being significantly larger than the experimental noise and any base-line artifacts), matrix **E** contains only noise. It is therefore discarded and not used in further analysis.

The correlations between **T** and **C** and between **P**′ and **V** are related. Neglecting noise and experimental artifacts

$$\mathbf{TP}' = \mathbf{CV} = \mathbf{A} \tag{4}$$

In matrix notation, the correlations are

$$\mathbf{T} = \mathbf{CR} \tag{5}$$

$$\mathbf{P}' = \mathbf{R}^{-1}\mathbf{V} \tag{6}$$

where **R** is a square $r \times r$ matrix. Its elements are unknown and cannot be determined without further information.

Protolytic Equilibria. Consider the protolytic equilibrium:

$$\mathrm{A} + \mathrm{H}^+ \overset{K}{\rightleftarrows} \mathrm{AH}^+ \tag{7}$$

The elements of **R**, which in this case is a 2×2 matrix, can be determined by requiring that the concentrations should satisfy the equilibrium expression

$$[\mathrm{AH}^+]/[\mathrm{A}][\mathrm{H}^+] = K = K_a^{-1} \tag{8}$$

Using the equilibrium relation, the concentrations of the two protolytic forms can be written

$$[\mathrm{A}] = \frac{1}{1 + K[\mathrm{H}^+]}[\mathrm{A}]_t \tag{9}$$

$$[\mathrm{AH}^+] + \frac{K[\mathrm{H}^+]}{1 + K[\mathrm{H}^+]}[\mathrm{A}]_t \tag{10}$$

where $[\mathrm{A}]_t = [\mathrm{A}] + [\mathrm{AH}^+]$ is the total concentration of A. Inserting the concentrations of the protolytic forms into eq 5

$$\mathbf{t}_1 = r_{11}\mathbf{c}_1 + r_{12}\mathbf{c}_2 = r_{11}\frac{1}{1 + K\mathbf{c}_{\mathrm{H}^+}}\mathbf{c}_t + r_{12}\frac{K\mathbf{c}_{\mathrm{H}^+}}{1 + K\mathbf{c}_{\mathrm{H}^+}}\mathbf{c}_t \tag{11}$$

$$\mathbf{t}_2 = r_{21}\mathbf{c}_1 + r_{22}\mathbf{c}_2 = r_{21}\frac{1}{1 + K\mathbf{c}_{\mathrm{H}^+}}\mathbf{c}_t + r_{22}\frac{K\mathbf{c}_{\mathrm{H}^+}}{1 + K\mathbf{c}_{\mathrm{H}^+}}\mathbf{c}_t \tag{12}$$

where $\mathbf{t}_1$ and $\mathbf{t}_2$ are the first and second columns of matrix **T** and $\mathbf{c}_{\mathrm{H}^+}$ and $\mathbf{c}_t$ are vectors containing the known proton concentrations and total concentrations of A, respectively. Equations 11 and 12 are rearranged to

$$\mathbf{t}_1 = r_{11}\mathbf{c}_t + r_{12}K\mathbf{c}_{\mathrm{H}^+}\mathbf{c}_t - K\mathbf{c}_{\mathrm{H}^+}\mathbf{t}_1 \tag{13}$$

$$\mathbf{t}_2 = r_{21}\mathbf{c}_t + r_{22}K\mathbf{c}_{\mathrm{H}^+}\mathbf{c}_t - K\mathbf{c}_{\mathrm{H}^+}\mathbf{t}_2 \tag{14}$$

Linear regressions of the vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ with respect to the vectors $\mathbf{c}_t$, $(\mathbf{c}_{\mathrm{H}^+}\mathbf{c}_t)$ and $(\mathbf{c}_{\mathrm{H}^+}\mathbf{t}_1)$, and $\mathbf{c}_t$, $(\mathbf{c}_{\mathrm{H}^+}\mathbf{c}_t)$ and $(\mathbf{c}_{\mathrm{H}^+}\mathbf{t}_2)$, respectively, where the product vectors are multiplied element by element, gives the regression coefficients r_{11}, $r_{12}K_a$, $-K_a$ and r_{21}, $r_{22}K_a$, $-K_a$. From these the four elements of matrix **R** and the equilibrium constant K_a can be calculated. The fit is preferably made globally, ensuring that a unique value of K_a is determined (see Appendix).

The component concentrations and spectral profiles are then calculated using matrix **R**:

$$\mathbf{C} = \mathbf{TR}^{-1} \tag{15}$$

$$\mathbf{V} = \mathbf{RP}' \tag{16}$$

Other Equilibria. For a general equilibrium the equilibrium expression may be more complex. Still, one can relate the component concentrations to total concentrations and the equilibrium constant and analyze the titration by the same approach. Consider the dimerization reaction

$$2\mathrm{A} \overset{K_D}{\rightleftarrows} \mathrm{A}_2 \tag{17}$$

where

$$[\mathrm{A}_2]/[\mathrm{A}]^2 = K_D \tag{18}$$

The component concentrations expressed in terms of the total concentration of A, $[\mathrm{A}]_t = [\mathrm{A}] + 2[\mathrm{A}_2]$, and the equilibrium constant, are equations of second order:

$$[\mathrm{A}] = -(4K_D)^{-1} \overset{+}{(-)} ((4K_D)^{-2} + [\mathrm{A}]_t/2K_D)^{1/2} \tag{19}$$

$$[\mathrm{A}_2] = \frac{[\mathrm{A}]_t + (4K_D)^{-1}}{2} \overset{(+)}{-} \left(\left(\frac{[\mathrm{A}]_t + (4K_D)^{-1}}{2}\right)^2 - \frac{[\mathrm{A}]_t^2}{4}\right)^{1/2} \tag{20}$$

The concentration vectors $\mathbf{c}_A$ and $\mathbf{c}_{A_2}$ are inserted into eq 5 to give

$$\mathbf{t}_1 = r_{11}\mathbf{c}_A + r_{12}\mathbf{c}_{A_2} \tag{21}$$

$$\mathbf{t}_2 = r_{21}\mathbf{c}_A + r_{22}\mathbf{c}_{A_2} \tag{22}$$

The regression with respect to the total concentration is nonlinear, and K_D is preferably determined by a stepwise search. Concentration vectors, $\mathbf{c}_A$ and $\mathbf{c}_{A_2}$, are calculated for different values of K_D and fitted to the target vectors. The best value of K_D will give the best fit, as determined by χ^2, and the elements of **R** are the regression coefficients for this value of K_D.

The same method can be applied when more components are involved in the equilibrium, though the number of linear equations to be solved increases. Consider the complex formation of two molecules:

$$\alpha\mathrm{A} + \beta\mathrm{B} \overset{K}{\rightleftarrows} \mathrm{A}_\alpha\mathrm{B}_\beta \tag{23}$$

where

$$[\mathrm{A}_\alpha\mathrm{B}_\beta]/[\mathrm{A}]^\alpha[\mathrm{B}]^\beta = K \tag{24}$$

If the three components have overlapping spectra the relation between the target and concentration vectors are

$$\mathbf{t}_1 = r_{11}\mathbf{c}_A + r_{12}\mathbf{c}_{AB} + r_{13}\mathbf{c}_B \tag{25}$$

$$\mathbf{t}_2 = r_{21}\mathbf{c}_A + r_{22}\mathbf{c}_{AB} + r_{23}\mathbf{c}_B \tag{26}$$

$$\mathbf{t}_3 = r_{31}\mathbf{c}_A + r_{32}\mathbf{c}_{AB} + r_{33}\mathbf{c}_B \tag{27}$$

As previously, the concentration vectors are calculated from the equilibrium expression and the known total concentrations, $[\mathrm{A}]_t = [\mathrm{A}] + \alpha[\mathrm{A}_\alpha\mathrm{B}_\beta]$ and $[\mathrm{B}]_t = [\mathrm{B}] + \beta[\mathrm{A}_\alpha\mathrm{B}_\beta]$, for different values of the equilibrium constant and fitted to the

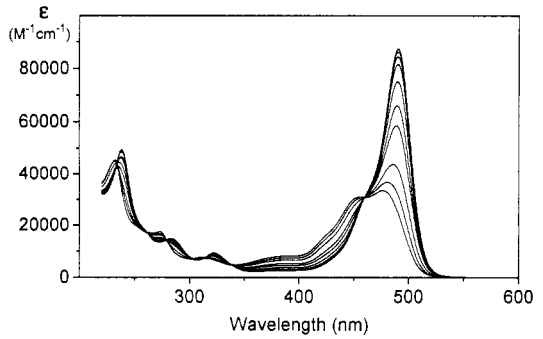

Figure 1. Absorption spectra of fluorescein at pH 5.56, 5.88, 6.19, 6.60, 6.84, 7.12, 7.46, 7.78, 8.09, and 8.93 (from low to high absorption of the main peak around 490 nm). The total fluorescein concentration in all samples was 11.7 μM, assuming a molar absorptivity of 87 600 M^{-1} cm^{-1} of the dianionic species.[14]

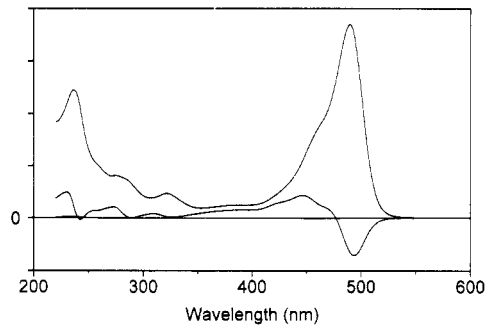

Figure 2. The three most significant projection vectors calculated by NIPALS for the titration data in Figure 1. The orthonormal projection vectors have been scaled with the lengths of the corresponding target vectors to obtain magnitudes that reflect their significance. The third projection vector is seen to be insignificant relative to the first two (it follows the base line), providing evidence that only two spectroscopically distinguishable components contribute to the absorption spectra.

target vectors. The best fit determines K and the elements of **R**.

It is worth noting that if there is a spectral region where only two components contribute, the problem will simplify to two linear equations if the analysis is restricted to this region.

EXAMPLES

Determination of a Protolytic Constant. As an example we determine the protolytic constant for the equilibrium between the fluorescein anion and dianion from absorption spectra of a protolytic titration. The protolytic forms of fluorescein in aqueous solution are the cation, neutral, anion, and dianion.[8] The different pK_a values are rather close, and the neutral and anionic species cannot be obtained in pure forms. Consequently, in a titration attempting to determine the protolytic constant between the anionic and dianionic species, the lower titration end point cannot be reached. Figure 1 shows absorption spectra of fluorescein measured between pH 5.56 and 8.93. In this pH range the concentration of the neutral species is negligible, as is evident from the many isosbestic points.

The recorded spectra were digitized into 951 data points each and arranged in a data matrix **A**. The two most significant target ($\mathbf{t}_1$ and $\mathbf{t}_2$) and projection vectors ($\mathbf{p}_1'$ and $\mathbf{p}_2'$) were calculated by NIPALS (Figure 2). These vectors have no immediate physical significance, since they are unknown linear combinations of the component concentration and spectral profiles.

(8) Zanker, V.; Peter, W. *Chem. Ber.* **1958**, *91*, 572.

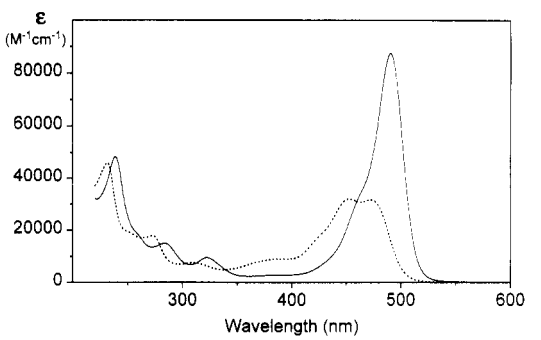

Figure 3. Calculated spectral profiles for fluorescein anion (···) and dianion (—).

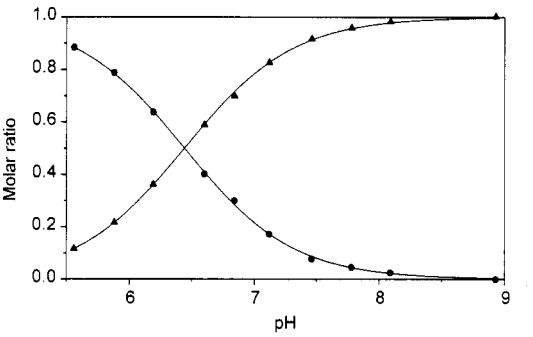

Figure 4. Best fit of calculated concentrations (normalized to unit total concentration) of fluorescein anion (●) and dianion (▲) to the concentration dependence predicted by the protolytic equilibrium equation (solid lines, pK_a = 6.44).

The two target vectors were fitted globally to the vectors $\mathbf{c}_t$, ($\mathbf{c}_H$+$\mathbf{c}_t$), ($\mathbf{c}_H$+$\mathbf{t}_1$), and ($\mathbf{c}_H$+$\mathbf{t}_2$) to give K_a and the elements of matrix **R** (eqs 13 and 14). K_a was determined to 3.6 × 10^{-7} M^{-1}, which corresponds to pK_a = 6.44. Matrix **R**,

$$\mathbf{R} = \begin{bmatrix} 0.26 & 0.47 \\ 0.18 & -0.17 \end{bmatrix} \quad (28)$$

was used to calculate the component spectral and concentration profiles (eqs 5 and 6). The spectrum of the dianionic species (Figure 3) was essentially identical to the spectrum recorded at the highest pH of the titration, as expected, since the upper titration end point was reached in the experiment. This, however, was not assumed in the analysis, and the good correspondence is an indication of a successful result. The calculated spectrum of the monoanionic species is significantly different from the spectrum recorded at the lowest pH of the titration, which was also expected, since the lower titration end point could not be reached. The calculated concentrations vary with pH, as expected for a protolytic equilibrium (Figure 4). This comparison of calculated concentrations with those predicted by the equilibrium expression is a very sensitive control of the analysis. In Figure 4 the calculated concentrations (for pK_a = 6.44) deviate randomly from the predicted ones. However, for a somewhat smaller (6.4) or larger pK_a value (6.5), which give only a slightly larger χ^2, the deviations between calculated and predicted concentrations are systematic (not shown).

Determination of a Dimerization Constant. As a second example we determine the dimerization constant of benzoic acid in *n*-heptane at 30 °C. With standard equipment it is difficult to measure on samples that differ in concentration by more than a factor of ~100, and we were unable to record spectra of the benzoic acid in pure monomeric and pure dimeric forms without changing other factors, such as temperature or solvent. As we shall see, however, with this new method of analysis, the dimerization constant as well as

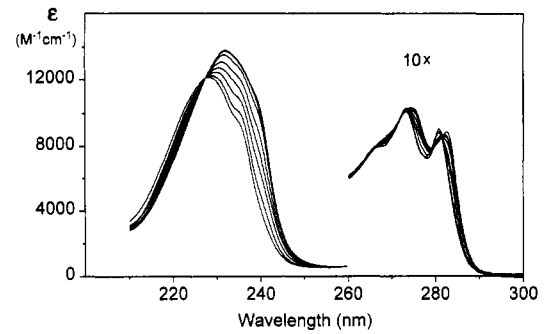

Figure 5. Absorption spectra of benzoic acid in *n*-heptane at 30.0 °C. Total concentrations: 1.02×10^{-5}, 2.01×10^{-5}, 4.77×10^{-5}, 9.64×10^{-5}, 2.01×10^{-4}, 5.11×10^{-4}, 1.01×10^{-3}, and 1.26×10^{-3} M in the order of shift toward higher wavelengths. Cell lengths of 0.1–5 cm were used. Spectra are expressed in molar absorptivities.

the spectra of the two forms can readily be determined, even though the titration end points were not reached.

Figure 5 shows absorption spectra of benzoic acid in *n*-heptane at 30 °C measured at total benzoic acid concentrations between 10^{-5} and 10^{-3} M. With increasing concentration, both the charge-transfer band around 230 nm and the benzene L_b band around 275 nm shift to longer wavelengths as a result of dimer formation.[9] Several isosbestic points are observed in the spectra, indicating that no additional components but the monomer and dimer of benzoic acid are present in the samples.

The spectra were digitized into 450 data points each, and the two most significant target and projection vectors were calculated. Concentration vectors $\mathbf{c}_A$ and $\mathbf{c}_{A_2}$ were calculated for various values of the dimerization constant K_D (eqs 19 and 20) and fitted to eqs 29 and 30:

$$\mathbf{t}_1 = r_{11}\left(\frac{\mathbf{c}_A}{\mathbf{c}_A + 2\mathbf{c}_{A_2}}\right) + r_{12}\left(\frac{2\mathbf{c}_{A_2}}{\mathbf{c}_A + 2\mathbf{c}_{A_2}}\right) \quad (29)$$

$$\mathbf{t}_2 = r_{21}\left(\frac{\mathbf{c}_A}{\mathbf{c}_A + 2\mathbf{c}_{A_2}}\right) + r_{22}\left(\frac{2\mathbf{c}_{A_2}}{\mathbf{c}_A + 2\mathbf{c}_{A_2}}\right) \quad (30)$$

These differ from eqs 21 and 22 by being normalized to unit total concentration (the concentrations in eqs 21 and 22 are replaced by molar ratios). The reason for this modification is that the experimental spectra were recorded using different path lengths and they have similar signal-to-noise ratios when normalized (if compared in absorption units, the noise level of the most concentrated sample is of the same magnitude as the spectral responses of the most diluted samples, making the analysis unstable).

The dependence of the goodness of fit, described by the sum of squared residuals, χ^2, on log K_D is shown in Figure 6. A distinct minimum is seen around log K_D = 4.20, which corresponds to $K_D = 15.9 \times 10^3$ M^{-1}. The regression coefficients obtained with this value of K_D define matrix **R**, and the concentrations and spectral responses of the benzoic acid monomer and dimer are calculated from eqs 15 and 16. The value of K_D is somewhat larger than determined earlier by spectroscopic analysis (8.7×10^3 M^{-1}).[9] Our value is likely to be more correct owing to the higher accuracy of the present analysis, and also we have analyzed more samples than in the previous study (eight, Figure 5, compared to only four in ref 9). The calculated spectra of benzoic acid monomer and dimer (Figure 7) have very little noise owing to the relatively large number of samples analyzed. They are consistent with the spectral shifts in the experimental data, and they reveal

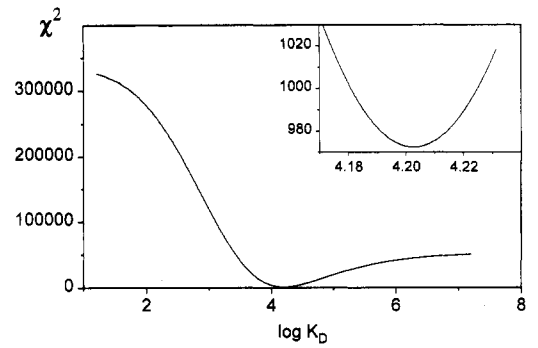

Figure 6. χ^2 (squared residuals) of fitting calculated molar ratios of benzoic acid monomer and dimer using different values of K_D to the two target vectors for the absorption spectra in Figure 5 (eqs 29 and 30). Inset: enlargement of the region $4.17 < \log K_D < 4.24$.

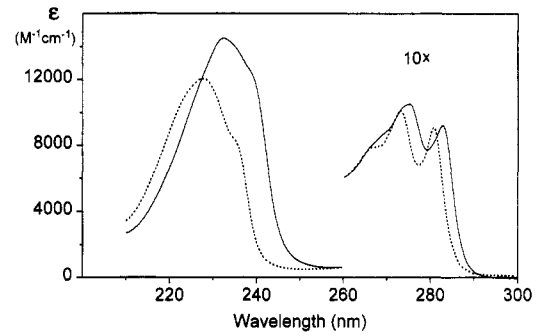

Figure 7. Calculated spectra of benzoic acid monomer (···) and dimer (—) in *n*-heptane.

particular spectral features of the components, such as the shoulder around 233 nm of the benzoic acid monomer. The calculated concentrations are in good agreement with those predicted by the equilibrium expression (not shown) and reveal that the molar ratios in the experimental spectra range from about 15 to 85%. Clearly, the analysis works well even though neither of the samples contained a pure component.

DISCUSSION

We have shown how knowledge of the functional form of the equilibrium expression can be utilized to determine the equilibrium constant and the concentrations and spectral profiles of the components in chemical equilibrium from spectroscopic titrations. As demonstrated with two examples, the approach works very well on data that can be obtained with standard instrumentation.

Number of Components. A problem common to all chemometric methods is to determine the number of independent components, which here should be the number of principal components, *r*, retained from NIPALS.[6,10,11] Fortunately, this is rarely a problem for the systems considered here, since the number of components participating in the equilibrium is usually known beforehand. Still possible complications might be base-line variations, contributions from solvent, and formation of unexpected species. These problems can usually be recognized directly in the experimental data. For example, in systems with only two components, such as in the examples above, isosbestic points will appear in the spectra at wavelengths where the components

(9) Hosoya, H.; Tanaka, J.; Nagakura, S. *J. Mol. Spectrosc.* **1962**, *8*, 257–275.
(10) Malinowski, E. R. *Anal. Chem.* **1977**, *49*, 606–612.
(11) Malinowski, E. R. *Anal. Chem.* **1977**, *49*, 612–617.
(12) Fisher, R.; MacKenzie, W. *J. Agric. Sci.* **1923**, *13*, 311–320.
(13) Wold, H. *Research Papers in Statistics*; Daved, F., Ed.; Wiley: New York, 1966; pp 411–444.
(14) Leonard, H.; Gordon, L.; Livingston, R. *J. Phys. Chem.* **1971**, *75*, 245–249.

have identical spectral responses. Any spectrum that deviates from an isosbestic point must contain an additional component, either a real spectral contribution from a physical species or an artificial component, and should be discarded. For example, in the protolytic titration of fluorescein, spectra recorded at pH below 5.56 (not shown) deviated at the isosbestic points owing to significant contributions from an third protolytic form (the neutral species) and were omitted in the analysis.

If the two components do not have identical responses at any wavelength, or in systems with more than two components, isosbestic points will not appear. In such cases it is wise to calculate a few additional pairs of target and score vectors and inspect them: only those that contribute substantially and have spectral features, in contrast to random noise, should be used. As illustrated in Figure 2, where the third nonrelevant projection vector can hardly be distinguished from the base line, the decision of how many components to use is usually unambiguous.

Stability of the Method. The stability of the approach depends on several factors: the number of spectra analyzed, the number of data points in each spectrum, the degree of spectral overlap between the components, the signal-to-noise ratios of the spectra, and how close to the titration end points one can reach. The effect of these factors was recently extensively investigated for the related Procrustes rotation method,[6] and the general conclusions should also be valid for this approach: (i) the errors in calculated concentrations and spectral profiles increase linearly with increasing noise; (ii) increasing the number of samples increases mainly the accuracy in the calculated spectral profiles, and increasing the number of data points per spectrum increases the accuracy in the determined concentrations; (iii) the analysis is very sensitive to changes in spectral profiles, and the experiments should be designed to minimize artificial spectral shifts; (iv) the larger the spread in relative concentrations the better. Simulations of the latter dependence (not shown) indicate that the analysis in general is very stable; if the degree of spectral overlap is not too extensive and the experimental signal-to-noise ratio is high, reasonable estimates of, for example, a protolytic constant can be obtained in a pH range as narrow as 2 units, even if the range does not include the pK_a value.

Applications of the Method. Essentially any equilibrium system should be possible to analyze by the proposed method. When the equilibrium expression is not known, the approach can be used to test different models. The only requirement is that the spectral response is linear, and the approach is therefore applicable to most spectroscopic techniques. The analysis is also fast; even when the equilibrium constant must be determined by stepwise search, complete analysis is made in a few minutes on an standard personal computer. The used algorithms are a standard iteration (NIPALS) and a linear least-squares fit (linear regression), requiring neither advanced programming nor access to mathematical libraries. A program for the analysis in ASYST code can be obtained from the authors.

ACKNOWLEDGMENT

This project is supported by Magn. Bergvalls Stiftelse, Stiftelsen Wilhelm och Martina Lundgrens Vetenskapsfond and the Swedish Research Council for Engineering Sciences.

APPENDIX

NIPALS. Since only the first few most significant target and projection vectors are required for the analysis, the NIPALS method is the optimum choice. The NIPALS algorithm is as follows:[12,13]

(1) Chose the column in matrix **A** with the largest variance as a starting value for $\mathbf{t}_1$.

(2) Calculate the corresponding projection vector as

$$\mathbf{p}_1' = \mathbf{t}_1'\mathbf{A}/\mathbf{t}_1'\mathbf{t}_1 \tag{31}$$

(3) Normalize $\mathbf{p}_1$ to unit length by multiplying with c

$$c = 1/(\mathbf{p}_1'\mathbf{p}_1)^{1/2} \tag{32}$$

(4) Calculate a new target vector as

$$\mathbf{t}_t = \mathbf{A}\mathbf{p}_1 \tag{33}$$

(5) Check for convergence. If convergence has been achieved go on with step 6; otherwise repeat from step 2. (6) Form the residual matrix

$$\mathbf{E} = \mathbf{A} - \mathbf{t}_1\mathbf{p}'_1 \tag{34}$$

Use **E** as a new **A** and calculate the next pair of target and projection vectors by repeating the procedure. It is recommended to calculate an extra pair of vectors than required by the analysis to check the validity of the assumed equilibrium reaction. If the extra projection vector contains spectral features, it is a strong indication that an additional spectroscopic component is present in significant amounts and the assumed equilibrium may be incorrect.

Global Analysis. Some of the linear equations encountered in the analysis have common parameters and are preferably solved simultaneously. A simple approach is to suitably catenate the vectors to obtain a single linear equation. The equation system below is equivalent to eqs 13 and 14:

$$\mathbf{y}_1 = c_{11}\mathbf{x}_1 + \mathbf{c}_{12}\mathbf{x}_2 + \mathbf{c}_c\mathbf{x}_3 \tag{35}$$

$$\mathbf{y}_2 = c_{21}\mathbf{x}_1 + c_{22}\mathbf{x}_2 + c_c\mathbf{x}_4 \tag{36}$$

The vectors are catenated as follows:

$$(\mathbf{y}_1{:}\mathbf{y}_2) = c_{11}(\mathbf{x}_1{:}0) + c_{12}(\mathbf{x}_2{:}0) + c_{21}(0{:}\mathbf{x}_1) + c_{22}(0{:}\mathbf{x}_2) + c_c(\mathbf{x}_3{:}\mathbf{x}_4) \tag{37}$$

where **0** is the null vector. This equation can be written in matrix form as

$$\mathbf{y} = \mathbf{X}\mathbf{c} \tag{38}$$

where **y** is the catenated vector $(\mathbf{y}_1{:}\mathbf{y}_2)$, **c** is a vector containing the coefficients c_{11}, c_{12}, c_{21}, c_{22}, and c_c. and **X** is a matrix having the vectors $(\mathbf{x}_1{:}0)$, $(\mathbf{x}_2{:}0)$, $(0{:}\mathbf{x}_1)$, $(0{:}\mathbf{x}_2)$, and $(\mathbf{x}_3{:}\mathbf{x}_4)$ as columns. The matrix equation is solved for **c**:

$$\mathbf{c} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{39}$$

RECEIVED for review October 5, 1992. Accepted January 8, 1993.