

Strategies for the Reduction and Interpretation of Multicomponent Spectral Data

Isiah M. Warner, S. L. Neal, and T. M. Rossi
Emory University, Atlanta, GA 30322

Accepted: July 1, 1985

Fluorescence data can be rapidly acquired in the form of an emission-excitation matrix (EEM) using a novel fluorometer called a video fluorometer (VF). An EEM array of 4096 data points composed of fluorescence intensity measured at 64 different emission wavelengths and excited at 64 different excitation wavelengths can be acquired in less than one second. The time-limiting factor in using this information for analytical measurement is the interpretation step. Consequently, sophisticated computer algorithms must be developed to aid in interpretation of such large data sets. For “*r*” number of components, the EEM data matrix, **M**, can be conveniently represented as

$$\mathbf{M} = \sum_{i=1}^r \alpha_i \mathbf{x}(i) \mathbf{y}(i)'$$

where $\mathbf{x}(i)$ and $\mathbf{y}(i)'$ are the observed excitation and emission spectra of the i^{th} component and α_i is a concentration dependent parameter. Such a data matrix is readily interpreted using linear algebraic procedures.

Recently a new instrument has been described which rapidly acquires fluorescence detected circular dichroism (FDCD) data for chiral fluorophores as a function of multiple excitation and emission wavelengths. The FDCD matrix is similar in form to EEM data. However, since the FDCD matrix may have legitimate negative entries while the EEM is theoretically non-negative, different assumptions are required. This paper will describe the mathematical algorithms developed in this laboratory for the interpretation of the EEM in various forms. Particular emphasis will be placed on linear algebraic and two-dimensional Fourier Transform procedures.

Key words: circular dichroism; eigenvectors; fluorescence; pattern recognition.

Introduction

Advances in computer technology and developments in multiparametric detection devices have had a profound effect on developments in chemical analysis. These developments have made it possible to expand the applicability of several analytical methods to more complex systems. Multicomponent analysis by fluorescence spectroscopy is one example of such a method.

About the Authors, Paper: Isiah M. Warner, S. L. Neal, and T. M. Rossi are with the Department of Chemistry at Emory. The work they describe was supported in part by the Office of Naval Research and the National Science Foundation. Isiah M. Warner also acknowledges support from an NSF Presidential Young Investigator Award.

In the conventional fluorescence experiment, the sample solution is irradiated with monochromatic light which produces molecules in the excited state when absorbed. As these molecules return to the ground state, light with a characteristic wavelength distribution is emitted. This distribution of the emitted light is known as the fluorescence emission spectrum. The fluorescence excitation spectrum is the fluorescence intensity as a function of absorption wavelength. At low absorbance (<0.01), the intensity of the fluorescence emitted at a given wavelength is directly proportional to the amount of light absorbed and therefore to the concentration of the analyte in the sample solution.

These characteristic spectra make fluorescence spectroscopy inherently more selective than absorption methods and provide qualitative as well as quantitative measurements. For example, when the excitation spectrum of the analyte only partially overlaps the excitation spectrum of the

other components in the sample, the sample can be irradiated with light of a wavelength which is only absorbed by the analyte. Since only the analyte absorbs the incident light, fluorescence will only be emitted by the analyte, in the absence of synergistic effects such as energy transfer. This technique is called selective excitation [1]¹ and can be used qualitatively by acquiring the complete emission spectrum of the analyte or quantitatively by determining the analyte concentration using the emission of a sample of known concentration as a standard.

This approach to multicomponent analysis can be expanded by acquiring fluorescence spectra at several excitation and emission wavelengths. Acquiring multiple spectra with conventional instrumentation is a time-consuming process, even with a microprocessor-controlled instrument. It can require more than one hour to acquire the complete emission spectrum of a sample at 64 excitation wavelengths using conventional instrumentation. Many samples would undergo significant photodecomposition over such a period of time. The development of the videofluorometer (VF) which rapidly acquires two dimensional fluorescence data has resolved this problem [2]. This instrument uses polychromatic excitation and a silicon intensified target vidicon detector (television camera) to acquire data in matrix format without mechanical scanning. The VF can acquire 64 emission spectra generated at 64 excitation wavelengths in less than 1 second. Therefore, data processing of this emission-excitation matrix (EEM) is now the time-limiting step in the analysis and requires the use of a computer for data reduction.

The rows of the EEM of a pure sample are multiples of the emission spectrum of that compound, while the columns are multiples of the excitation spectrum, indicating that the EEM is of bilinear form. Such matrices are ideally suited for data reduction techniques such as factor analysis [3] and pattern recognition [4].

Another technique whose applicability is expanded by multiparametric detection is fluorescence detected circular dichroism (FD CD) [5]. The FD CD spectrum is the difference in the intensity of fluorescence produced by excitation with left and right circularly polarized light recorded as a function of the wavelength of the exciting light. Chiral molecules preferentially absorb one form of circularly polarized light. The FD CD spectrum reflects the structure of the chiral fluorophore in a solution. A two-dimensional rapid scanning FD CD spectrometer has been developed which measures the FD CD at several excitation and emission wavelengths [6]. When these data are collected in matrix format, they are also of bilinear form. However, the algorithms designed to resolve the spectra of components from the EEM cannot be applied to this ellipticity matrix since it may contain legitimate negative values. This

manuscript will discuss the strategies developed for the qualitative reduction of multicomponent EEMS as well as the alternative techniques developed for multicomponent ellipticity matrices.

Eigenvector Analysis [7]

When the absorbance of a sample is less than 0.01, the intensity of the fluorescence, I_f , can be approximated by the expression

$$I_f = 2.303 I_0 \phi_f \epsilon b c \quad (1a)$$

where I_0 is the intensity of the incident radiation, ϕ_f is the fluorescence quantum efficiency (the fraction of absorbed photons emitted as fluorescence), ϵ is the molar extinction coefficient, b is the thickness of sample cell and c is the concentration of the fluorophore in the sample solution. Each element, m_{ij} , of the emission-excitation matrix, \mathbf{M} , represents the fluorescence intensity at wavelength λ_j that was generated by excitation at wavelength λ_i . Therefore, each of these elements can be generally expressed as

$$m_{ij} = 2.303 \phi_f I_0(\lambda_i) \epsilon(\lambda_i) \gamma(\lambda_j) \delta(\lambda_j) b c \quad (2a)$$

where $\gamma(\lambda_j)$ reflects the dependence of I_f on the monitored emission wavelength and $\delta(\lambda_j)$ is a parameter which incorporates instrumental artifacts like sensitivity and signal collection geometry. Combining these terms based on excitation and emission wavelength related variables results in a simpler expression:

$$m_{ij} = \alpha x_i y_j \quad (3a)$$

where α is a scalar that equals $2.303 \phi_f b c$, x_i is the excitation term given by

$$x_i = I_0(\lambda_i) \epsilon(\lambda_i) \quad (4a)$$

and y_j is the emission term which is expressed

$$y_j = \gamma(\lambda_j) \delta(\lambda_j) \quad (5a)$$

When the x_i are properly sequenced, the array of x_i is a representation of the excitation spectrum and can be denoted \mathbf{x} in vector notation. Likewise, when the y_j are properly sequenced, the vector \mathbf{y} represents the emission spectrum. Since the emission profile is independent of the exciting wavelength, and the excitation profile is independent of the monitored emission wavelength, the matrix \mathbf{M} can clearly be expressed as the vector product of \mathbf{x} and \mathbf{y} , multiplied by a scalar concentration term, i.e.

$$\mathbf{M} = \alpha \mathbf{x} \mathbf{y}^T \quad (6a)$$

¹Figures in brackets indicate literature references.

In a sample containing "n" fluorescent compounds, the matrix \mathbf{M} is the sum of the EEMs of the individual components provided synergistic effects are negligible. Thus, the n component matrix can be expressed as

$$\mathbf{M} = \sum_{k=1}^n \alpha_k \mathbf{x}_k \mathbf{y}_k^T \quad (7a)$$

A more convenient notation for \mathbf{M} is

$$\mathbf{M} = \mathbf{X}\mathbf{Y} \quad (8a)$$

in which the columns of the matrix \mathbf{X} are the excitation spectra, \mathbf{x}_k , of the n components, and the rows of \mathbf{Y} are the emission spectra, \mathbf{y}_k , of the components. The concentration term can be considered to be absorbed into either of the matrices.

Qualitative analysis of the matrix \mathbf{M} requires a determination of the number of independently emitting compounds, i.e., the rank of the matrix, and the set of basis vectors \mathbf{x}_k and \mathbf{y}_k which are the excitation and emission spectra of the components. Eigenanalysis plays a significant role in both determinations. Therefore, it is appropriate to preface the discussion of rank estimation and spectral resolution with a brief presentation of pertinent eigenanalysis principles.

An eigenvector is defined as any vector \mathbf{x} which is a solution of the equation

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad (9a)$$

in which λ is a scalar called the eigenvalue. The magnitude of the eigenvalue is a consequence of the importance of the information reflected in the eigenvector to the data in the matrix. When the eigenvalue is large, then the factor represented by the eigenvector makes a large contribution to the data. If it is small, then the contribution of the factor is small. Therefore, when a matrix has a rank n which is greater than 1, it has n eigenvectors and n eigenvalues.

Since the matrix \mathbf{M} is bilinear, the covariance matrices, $\mathbf{M}^T\mathbf{M}$ and $\mathbf{M}\mathbf{M}^T$ can be used to generate the eigenvectors. The eigenvalues of the covariance matrices are the squares of the eigenvalues of \mathbf{M} , but apart from this small detail, it is more expedient to use the covariance matrices to generate the eigenvectors since the covariance matrices are always square and symmetric.

As the preceding discussion indicates, rank estimation of an ideal matrix (one that is noise-free) is straightforward, simply determined by the number of non-zero eigenvalues. However, experimental data matrices are not free of noise. They can contain systematic or random errors superimposed on the signal. For these matrices, there are several methods of rank estimation [3,8]. When the signal-to-noise ratio is high, it is possible to differentiate the eigenvalues associated

with fluorescence (primary eigenvectors) from those associated with noise (secondary eigenvectors) by a direct comparison of their magnitudes. As the signal-to-noise ratio decreases, this approach becomes more difficult. Another method which will be described in a later section of this manuscript is differentiation of eigenvectors based on their frequency distributions, realizing that random noise in spectral data usually will have higher frequency than the signal.

The other part of the analysis of \mathbf{M} is the resolution of the basis vectors \mathbf{x}_k and \mathbf{y}_k from the matrix. An infinite number of basis vectors exists for a given matrix, and resolving the spectral vectors from the matrix is usually not possible without *a priori* knowledge of the components. However, the eigenvectors are an orthonormal basis for \mathbf{M} and are easily generated. Since the eigenvectors reproduce the matrix, \mathbf{M} can be expressed as

$$\mathbf{M} = \mathbf{U}\mathbf{V} \quad (10a)$$

in which the columns of \mathbf{U} are the excitation eigenvectors, \mathbf{u}_k , and the rows of \mathbf{V} are the emission eigenvectors, \mathbf{v}_k . The eigenvalues have been absorbed into one of the matrices. Since the eigenvectors are an orthonormal basis, they often contain negative elements even though \mathbf{M} is theoretically a non-negative matrix. Emission and excitation spectra are also theoretically non-negative. Therefore, the eigenvectors can be transformed to possible spectral vectors by transforming them to a non-negative basis set.

To perform this transformation, the values of the matrix \mathbf{K} and its inverse, \mathbf{K}^{-1} , which transform \mathbf{U} and \mathbf{V} to non-negative matrices, must be found. This is algebraically sound, since it is equivalent to multiplying \mathbf{U} and \mathbf{V} by the identity matrix. This transformation is mathematically expressed as

$$\mathbf{M} = \mathbf{U}\mathbf{K}\mathbf{K}^{-1}\mathbf{V} \quad (11a)$$

This condition also ensures that the transformed vectors are also a basis for \mathbf{M} . The transformed excitation vectors are the columns of the matrix \mathbf{U}' , and the transformed emission vectors are the rows of the matrix \mathbf{V}' . These matrices are given by the equations

$$\mathbf{U}' = \mathbf{U}\mathbf{K} \geq 0 \quad (12a)$$

and

$$\mathbf{V}' = \mathbf{K}^{-1}\mathbf{V} \geq 0 \quad (13a)$$

The values of the elements of \mathbf{K} and \mathbf{K}^{-1} can be found from the expressions for the elements of \mathbf{U}' and \mathbf{V}' . In the two component case, the elements are given by

$$u'_{1i} = k_{11}u_{1i} + k_{21}u_{2i} \geq 0, \quad (14a) \quad \text{and}$$

$$u_{2i} = k_{22}u_{2i} + k_{12}u_{1i} \geq 0, \quad (15a)$$

$$v'_{1j} = 1/k(k_{22}v_{1j} - k_{12}v_{2j}) \geq 0, \quad (16a)$$

and

$$v'_{2j} = 1/k(k_{11}v_{2j} - k_{21}v_{1j}) \geq 0. \quad (17a)$$

It can be assumed that $k_{11} = k_{22} = 1$ without loss of generality. The values for k_{12} and k_{21} can be found by solving these expressions. It's clear from these expressions why this method is only applicable to two component matrices since the elements of K^{-1} become non-linear for more than two components. The boundaries for the values of k_{12} and k_{21} given by these expressions are

$$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} \geq k_{12} \geq \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}} \quad (18a)$$

$$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} \geq k_{21} \geq \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}} \quad (19a)$$

The accuracy of transformations performed with values meeting these criteria has been shown to be a function of the overlap of the spectra of the components. This is illustrated in the ambiguity table in figure 1 which summarizes the results of the transformation for the 16 possible spectral overlap combinations for a binary mixture. It should be noted that in 7 out of the 16 possible cases, at least one spectrum of each component is correctly resolved. Close inspection of the table shows that in cases where the spectra of neither component are enveloped in both dimensions, at least one of the spectra of each component is resolved unambiguously. Clearly, if both the emission and excitation spectra of the two compounds have only partial overlap, all four spectra will be resolved unambiguously from their mixture matrix.

		Emission Overlap				
		comp. 1		comp. 2		
Excitation Overlap	comp. 1	$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} = k_{12} = \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$	$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} = k_{12} = \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$	$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} \geq k_{12} > \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$	$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} \geq k_{12} > \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$	
	comp. 2	$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} = k_{21} = \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$	$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} \geq k_{21} > \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$	$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} = k_{21} = \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$	$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} \geq k_{21} > \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$	
	All spectra certain.		1 spectrum of ea. comp. certain. 1 spectrum of ea. component given by extreme k.		1 spectrum of ea. comp. certain. 1 spectrum of ea. component given by extreme k.	
	All spectra uncertain.		Both spectra of ea. component given by extreme k's.		All spectra uncertain. Both spectra of ea. component given by extreme k's.	
	1 spectrum of ea. comp. certain 1 spectrum of ea. component given by extreme k.		All spectra uncertain. Both spectra of ea. component given by extreme k's.		1 spectrum of ea. comp. certain. 1 spectrum of ea. component given by intermediate k.	
$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} > k_{12} \geq \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} > k_{12} \geq \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} > k_{12} > \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		
$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} = k_{21} = \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} \geq k_{21} > \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} = k_{21} = \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		
1 spectrum of ea. comp. certain 1 spectrum of ea. component given by extreme k.		All spectra uncertain. Both spectra of ea. component given by extreme k's.		1 spectrum of ea. comp. certain. 1 spectrum of ea. component given by intermediate k.		
$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} = k_{12} = \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} = k_{12} = \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} \geq k_{12} > \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		
$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} \geq k_{21} > \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} > k_{21} > \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} \geq k_{21} = \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		
1 spectrum of ea. comp. certain. 1 spectrum of ea. component given by extreme k.		1 spectrum of ea. comp. certain. 1 spectrum of ea. component given by intermediate k.		All spectra uncertain. Both spectra of ea. component given by extreme k's.		
$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} > k_{12} \geq \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} > k_{12} \geq \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		$\min_{v_{2j} > 0} \frac{v_{1j}}{v_{2j}} > k_{12} > \max_{u_{1i} > 0} \frac{-u_{2i}}{u_{1i}}$		
$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} > k_{21} \geq \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} > k_{21} > \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		$\min_{v_{1j} > 0} \frac{v_{2j}}{v_{1j}} \geq k_{21} = \max_{u_{2i} > 0} \frac{-u_{1i}}{u_{2i}}$		
All spectra uncertain. Both spectra of ea. component given by extreme k's.		All spectra uncertain. 1 spectrum of ea. comp. given by ext. k, other by int. k.		All spectra uncertain. Both spectra of ea. component given by intermediate k's.		

Figure 1—Ambiguity Table of Two-Component Excitation-Emission Matrix.

Rank Estimation by Frequency Analysis of the Eigenvectors [9]

The need to correctly determine the rank of a matrix for qualitative analysis has been demonstrated previously. Most rank estimation methods are statistical in nature and depend on the correct evaluation of the variance in the data. The method of rank analysis presented here uses Fourier Transform Image Analysis and is based on the assumption that the primary eigenvectors of an EEM (i.e., those associated with the fluorescence) will contain information that is weighted toward the low frequency Fourier coefficients since spectra tend to be broad-banded. Similarly, the transforms of the secondary eigenvectors will have larger high frequency coefficients. However, this manuscript is not a suitable medium for a comprehensive presentation of the theory of Fourier transforms. The reader is referred to any introductory text on the subject of Fourier analysis [10].

The Fourier transform is fairly simple to implement since fast Fourier transform algorithms are routinely available for use even on small computers. Since any continuous function can be reproduced by addition of a series of sine and cosine functions with various frequencies, amplitudes, and phases, the forward Fourier transform is a method of determining these frequencies, amplitudes, and phases from the function. The inverse transform reconstructs the time domain function from the frequencies, amplitudes, and phases. The discrete Fourier transform equation used to transform the eigenvector to the frequency domain is

$$V(u) = 1/N \sum_{x=0}^{N-1} v(x) \exp(-2\pi i(xu)/N) \quad (1b)$$

where u is the frequency domain coordinate, $V(u)$ is the Fourier transform of $v(x)$, and N is the number of points in the discrete approximation of the function. The complex frequency domain function, $V(u)$, is frequently represented by the Fourier spectrum which is given by

$$|V(u)| = [V(u)_{\text{real}}^2 + V(u)_{\text{imag}}^2]^{1/2} \quad (2b)$$

The area of the Fourier spectrum, T , is the sum of all the frequency coefficients and is given by

$$T = \sum_{u=0}^{N-1} |V(u)| \quad (3b)$$

The segment of the Fourier spectrum bounded by $\pm u_{\text{lim}}$ is denoted $A_{u_{\text{lim}}}$ which is defined as

$$A_{u_{\text{lim}}} = \sum_{u=-u_{\text{lim}}}^{u=u_{\text{lim}}} |V(u)| \quad (4b)$$

which is simply the sum of the frequency coefficients in the section of the Fourier spectrum bound by u_{lim} and $-u_{\text{lim}}$. The relative importance of this frequency region in reproducing the time domain eigenvector, $v(x)$, can be expressed by calculating the percent of T which lies in this frequency range. The parameter which represents the importance of the range from u_{lim} to $-u_{\text{lim}}$ is called $\%A_{u_{\text{lim}}}$ and is calculated from the expression

$$\%A_{u_{\text{lim}}} = A_{u_{\text{lim}}}/T \times 100 \quad (5b)$$

When the value of u_{lim} is well chosen, there is a marked drop in the $\%A_{u_{\text{lim}}}$ for secondary eigenvectors. Table 1 shows a table comparing the rank estimation by frequency analysis to four statistical methods for matrices with ranks greater than or equal to 3. The frequency analysis method was the most accurate on the data tested here.

Eigenvector Ratioing [11]

This method was developed to resolve the spectra of compounds from the ellipticity matrices of two component mixtures. The presence of legitimate negative values in this matrix produced a need for a different algorithm to analyze this matrix despite its similarity to the EEM. The ellipticity matrix, F , is also bilinear and can be expressed as

$$F = ST \quad (1c)$$

where the columns of the matrix S are the circular dichroism (CD) spectra, s_k of the fluorophores in the sample, and the rows of the matrix T , symbolized by the vectors t_k , are the emission spectra of those chiral fluorophores.

The CD and emission eigenvectors of F are also an orthonormal basis which span the matrix, F . The CD eigenvectors, q_k , may be represented as the columns of the matrix Q and the emission eigenvectors, p_k , as the rows of the matrix P . Thus, F is also given by the equation

$$F = OP \quad (2c)$$

assuming that the eigenvalues have been absorbed into one of the matrices.

These eigenvectors must also be transformed to possible spectral vectors. However, the CD eigenvectors should not be transformed to a non-negative vector because the CD spectral vectors are not always non-negative. However, since the matrix F contains a finite amount of information, if the correct emission spectral vectors are found, the corre-

Table 1. Estimated numbers of components in text mixtures.

Mixture No.	No. of Components		Estimated No. of Components			
	Known	REFAE	Stat1	Stat2	Stat3	Stat4
1	3	3*	3*	2	3*	4
2	3	3*	2	2	3*	4
3	3	3*	2	2	2	2
4	2	2*	1	1	2*	2*
5	2	2*	1	1	2*	3
6	2	2*	2*	2*	2*	5
7	3	3*	3*	2	2	5
8	4	4*	3	3	3	5
9	5	5*	3	3	3	5
10	6	5	3	3	4	6*

* denotes a correctly estimated rank

Stat1 = Eigenvalue perturbation

Stat2 = Average error method

Stat3 = Chi squared test

Stat4 = Standard error test

This is valid since

$$F = Q'P' \tag{9c}$$

and

$$I = P'P'^T(P'P')^{-1} \tag{10c}$$

This means that eq (8c) is another representation of the equation

$$Q' = Q'I \tag{11c}$$

sponding CD spectral vectors are fixed and easily obtained.

In this algorithm, the elements of the matrix, **K**, which transforms the emission eigenvectors to non-negative vectors are sought. Then, the possible CD spectral vectors are generated using the transformed emission vectors. The transformed emission vectors are the rows of the matrix **P'** which is given by

$$P' = KP \geq 0 \tag{3c}$$

For a two component mixture the values of the elements of **K** can be determined by expressing eq (3c) in terms of the elements of the matrices and solving the resulting expressions for the elements of **K** algebraically:

$$P_{1i}' = k_{11}p_{1i} + k_{12}p_{2i} \geq 0 \tag{4c}$$

$$P_{2i}' = k_{21}p_{1i} + k_{22}p_{2i} \geq 0 \tag{5c}$$

Again, assuming that $k_{11} = k_{22} = 1$, the elements of **K** are within the ranges defined by

$$\min_{p_{2i} < 0} \frac{-p_{1i}}{p_{2i}} \geq k_{12} \geq \max_{p_{2i} > 0} \frac{-p_{1i}}{p_{2i}} \tag{6c}$$

$$\min_{p_{1i} < 0} \frac{-p_{2i}}{p_{1i}} \geq k_{21} \geq \max_{p_{1i} > 0} \frac{-p_{2i}}{p_{1i}} \tag{7c}$$

These ranges are generated because the sense of an inequality changes when both sides of the inequality are divided by a negative number.

The possible CD spectral vectors are generated by solving the following equation:

$$Q = FP'^T(P'P')^{-1} \tag{8c}$$

Due to the weaker constraints on the transformation elements in this algorithm, non-converging ranges are generated for the values of the transformation elements. However, it was found that if either of the components is a sole emitter in the monitored emission range, the value of the transformation element needed to transform an eigenvector to the spectra of the other component is usually given at an extreme of one of the ranges in eqs (6c) and (7c). This is because the values from the regions of sole emission best meet the criteria expressed in eqs (6c) and (7c). Therefore, except when the spectra are totally coincident at the baselines at least one spectrum of each component is often retrievable using this technique. Figure 2 shows an ambiguity table that was generated to illustrate the usefulness of this algorithm.

This algorithm has also been tested on EEMs and it was found that this technique can be used with either set of eigenvectors: the emission eigenvectors (as it is with the ellipticity matrix) or the excitation eigenvectors. The results can be summarized by an ambiguity table similar to the one in figure 2 which had excitation overlap on the vertical axis. The ambiguity table using only one axis is more ambiguous than the earlier table (fig. 1) using both axes. These results do not conflict with those found using eigenvector analysis on the EEM. "Multiplying" the two ambiguity tables generated by this algorithm yields the table generated for eigenvector analysis, verifying the validity of both methods.

Emission Overlap

$\min_{p_1 < 0} \frac{-p_2}{p_1} \geq k_{12} \geq \max_{p_1 > 0} \frac{-p_2}{p_1}$ $\min_{p_2 < 0} \frac{-p_1}{p_2} \geq k_{21} \geq \max_{p_2 > 0} \frac{-p_1}{p_2}$ <p style="font-size: small;">All spectra uncertain. Both spectra of ea. component given by extreme k's</p>	$\min_{p_1 < 0} \frac{-p_2}{p_1} \geq k_{12} \geq \max_{p_1 > 0} \frac{-p_2}{p_1}$ $\min_{p_2 < 0} \frac{-p_1}{p_2} \geq k_{21} \geq \max_{p_2 > 0} \frac{-p_1}{p_2}$ <p style="font-size: small;">All spectra uncertain. 1 spectrum of ea. comp. given by ext. k, other by int. k.</p>	$\min_{p_1 < 0} \frac{-p_2}{p_1} \geq k_{12} \geq \max_{p_1 > 0} \frac{-p_2}{p_1}$ $\min_{p_2 < 0} \frac{-p_1}{p_2} \geq k_{21} \geq \max_{p_2 > 0} \frac{-p_1}{p_2}$ <p style="font-size: small;">All spectra uncertain. 1 spectrum of ea. comp. given by ext. k, other by int. k.</p>	$\min_{p_1 < 0} \frac{-p_2}{p_1} \geq k_{12} \geq \max_{p_1 > 0} \frac{-p_2}{p_1}$ $\min_{p_2 < 0} \frac{-p_1}{p_2} \geq k_{21} \geq \max_{p_2 > 0} \frac{-p_1}{p_2}$ <p style="font-size: small;">All spectra uncertain. Both spectra of ea. component given by intermediate k's.</p>
--	---	---	--

Figure 2—Ambiguity Table of Two-Component Ellipticity Matrix.

Conclusions

This manuscript has provided an overview of qualitative analysis techniques developed for matrix formatted fluorescence data. Qualitative analysis of matrices was shown to generally consist of three basic procedures: rank estimation, determination of unknown component spectra, and screening of expected compounds. The techniques outlined here only addressed the first two phases of the problem but they represent only a portion of the methods that have been developed to fill these requirements. The methods presented here successfully attack the stated problems within the framework of the limitations described.

Rank estimation by frequency analysis can sometimes be more accurate than statistical methods for evaluation of spectral data. It would be useful to develop a criterion where the algorithm will automatically select a useful range for differentiating the secondary eigenvectors from the primary.

In the present approach, eigenvector analysis and eigenvector ratioing are limited to binary mixtures. Few real samples are binary. These methods must be extended to higher order mixtures.

These techniques have been developed for use with fluorescence data, but are generally applicable to other forms of matrix formatted data. Some of the algorithms require that the data matrix be bilinear in form; however this is a characteristic of many types of data. For example, diode array detection of liquid chromatography and absorption kinetic

data using a diode array detector are bilinear in form. Clearly, there is a need for further development in this area.

References

- [1] Parker, C.A., *Photoluminescence of Solutions with Applications to Photochemistry and Analytical Chemistry*, American Elsevier, New York (1968).
- [2] Warner, I.M.; J.B. Callis, E.R. Davidson, M. Gouterman, and G.D. Christian, *Anal. Lett.*, **8**, 665 (1965).
- [3] Malinowski, E.R., and D.G. Howery, *Factor Analysis in Chemistry*, John Wiley, New York (1980).
- [4] Jurs, P.C., and T. L. Isenhour, *Chemical Applications of Pattern Recognition*, John Wiley, New York, NY, 1975.
- [5] Tinoco, I. Jr. and D.H. Turner, *J. Am. Chem. Soc.*, **98(21)**, 6453 (1976).
- [6] Thomas, M.; G. Patonay and I.M. Warner, work in progress.
- [7] Warner, I.M., "The Analysis of Matrix Formatted Multicomponent Data," *Contemporary Topics in Analytical and Clinical Chemistry*, D.M. Hercules, G.M. Hieftje, L.R. Snyder and M.A. Evenson, eds., Plenum, New York (1982).
- [8] Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press, New York (1972).
- [9] Rossi, T.M., and I.M. Warner, "Rank Determination of Emission Excitation Matrices via Fourier Transformation of Eigenvectors," American Chemical Society, 187th National Meeting, April 8-13, 1984, St. Louis, MO.
- [10] Bloomfield, P., *Fourier Analysis of Time Series: An Introduction*, John Wiley and Sons, New York (1976).
- [11] Neal, S.L.; C.-N. Ho and I.M. Warner, "Qualitative Analysis of Multicomponent Fluorescence Detected Circular Dichroism Data," American Chemical Society, 187th National Meeting, April 8-13, 1984, St. Louis, MO.