CORRESPONDENCE

Canonical Correlation Technique for Rank Estimation of Excitation–Emission Matrices

Sir: One of the major problems that has challenged analytical chemists for a long time is to determine the number of components in a multicomponent sample. Often the sample data occur in the form of a matrix whose rank, in the absence of noise, is equal to the number of components. The presence of random noise in the data, however, will generally cause the rank of the mixture matrix to exceed the number of components in the mixture. The problem then is to estimate what the rank would be if the noise were removed.

An eigenanalysis in the form a singular value decomposition of the measured mixture matrix provides information that is useful for estimating rank. Most methods that have been proposed utilize information from the eigenvalues and ignore the information contained in the eigenvectors (*1–6*). An exception is the frequency analysis method proposed by Rossi and Warner (*7*), which relies on the prior expectation that the Fourier transforms of noise eigenvectors will have heavily weighted coefficients for the high frequencies.

In this communication we will report a new approach to this problem which incorporates eigenvector information without requiring prior expectations about the nature of the eigenvectors. It utilizes the multivariate statistical technique of canonical correlations. We will present the methodology in the context of fluorescent mixtures and will also assume that the readers are familiar with the excitation–emission matrix (EEM). For information on the formulation of the EEM and its basic properties, see ref 8 and 9.

GENERAL CONSIDERATION

The mathematical formulation of the EEM for a k-component sample in the noise-free case can be expressed as

$$\mathbf{S} = \mathbf{X}\mathbf{D}\mathbf{Y}^T \tag{1}$$

where $\mathbf{S}$ is an m by n matrix, $\mathbf{X}$ is the m by k matrix whose jth column is the excitation spectrum for the jth component, $\mathbf{Y}$ is the n by k matrix whose jth column is the emission spectrum for the jth component, and $\mathbf{D}$ is a k by k diagonal matrix whose elements depend on the concentrations of the fluorescence emitters, fluorescence quantum efficiency, molar absorptivity, etc. We assume that both the columns of $\mathbf{X}$ and the columns of $\mathbf{Y}$ are linearly independent, which implies that the rank of $\mathbf{S}$ is equal to k, the number of components in the mixture. If $\mathbf{M}$ is the measured EEM, then $\mathbf{M}$ will typically contain a noise term $\mathbf{N}$ in addition to the signal $\mathbf{S}$, i.e.

$$\mathbf{M} = \mathbf{S} + \mathbf{N} \tag{2}$$

A common method for estimating k (the rank of $\mathbf{S}$) is to inspect the singular values from the singular value decomposition of $\mathbf{M}$. If the noise is low compared to the signal, one would expect to see a large drop between the kth and $(k + 1)$st singular values and much smaller drops between the $(k + 1)$st and later singular values. When the noise level is not low, however, this method is likely to fail, especially when the mixture contains more than two or three individual components and/or the individual components are highly correlated, i.e. have high spectral overlap.

We propose an alternative method for estimating k. This alternative approach utilizes canonical correlations and can be applied whenever there are at least two measured EEMs for mixtures containing the same k constituents. The method does not require that the constituents be present in the same proportions in both mixtures. However, for simplicity and ease of exposition, we will illustrate the method for replicate measurements on the same mixture.

Let $\mathbf{M1}$ and $\mathbf{M2}$ be m by n EEM's of a k-component mixture obtained under the same experimental conditions. We assume that $k < m \leq n$. Write $\mathbf{M1}$ and $\mathbf{M2}$ as

$$\begin{aligned} \mathbf{M1} &= \mathbf{X}\mathbf{D}\mathbf{Y}^T + \mathbf{N1} \\ &= \mathbf{U}_1\mathbf{L}_1\mathbf{V}_1^T \end{aligned} \tag{3}$$

$$\begin{aligned} \mathbf{M2} &= \mathbf{X}\mathbf{D}\mathbf{Y}^T + \mathbf{N2} \\ &= \mathbf{U}_2\mathbf{L}_2\mathbf{V}_2^T \end{aligned}$$

where $\mathbf{N1}$ and $\mathbf{N2}$ denote random noise and the $\mathbf{U}_i\mathbf{L}_i\mathbf{V}_i^T$ $(1 \leq i \leq 2)$ are the singular value decompositions for $\mathbf{M1}$ and $\mathbf{M2}$.

Traditional methods of rank estimation utilize the information contained in the singular values, which appear in the diagonal matrices $\mathbf{L}_1$ and $\mathbf{L}_2$. However, there is also useful information in the eigenvectors, the columns of the $\mathbf{U}_i$ and $\mathbf{V}_i$ matrices. We will show how that information can be extracted, using the statistical technique of canonical correlations.

In the absence of noise $\mathbf{L}_1 = \mathbf{L}_2$, $\mathbf{U}_1 = \mathbf{U}_2$, $\mathbf{V}_1 = \mathbf{V}_2$, and the singular values from $k + 1$ to m are equal to zero. Furthermore, the first k columns of $\mathbf{U}$ span the same subspace of the m-dimensional space as the columns of $\mathbf{X}$ do, and the first k columns of $\mathbf{V}$ span the same subspace of the n-dimensional space as the columns of $\mathbf{Y}$ do.

In the presence of noise we expect that $\mathbf{L}_1$ and $\mathbf{L}_2$ will differ and that in each case all m singular values will be positive. However, if the noise is low, all singular values from $k + 1$ to m should be close to zero. Traditional methods for estimating k exploit this expectation by looking for a pattern in the singular values in which all those from $k + 1$ to m are near zero.

With noise we also expect differences between $\mathbf{U}_1$ and $\mathbf{U}_2$ and between $\mathbf{V}_1$ and $\mathbf{V}_2$. If the noise is low, we expect that the subspace spanned by the first k columns of $\mathbf{U}_1$ and the subspace spanned by the first k columns of $\mathbf{U}_2$ will each be close to the subspace spanned by the columns of $\mathbf{X}$, and therefore close to each other. When we include the $(k + 1)$st columns of $\mathbf{U}_1$ and $\mathbf{U}_2$, the contribution is mainly the result of noise. Since the noise matrices $\mathbf{N}_1$ and $\mathbf{N}_2$ are independent, there is no reason to expect any relationship between the directions they contribute to an eigenanalysis.

These considerations lead us to an expectation about the structure of the subspaces spanned by the first $k + 1$ columns of $\mathbf{U}_1$ and $\mathbf{U}_2$, i.e. there are k dimensions in which these two subspaces nearly coincide, but their $(k + 1)$st dimensions are nearly orthogonal. Canonical correlations (*10–12*) have an interpretation as the cosines of the angles between two subspaces. If we calculate canonical correlations between the first $k + 1$ columns of $\mathbf{U}_1$ and $\mathbf{U}_2$, we expect to see k high correlations and one low correlation. If this pattern does in fact appear in the data, it provides us with an estimate of k.

Similar statements apply to the columns of $\mathbf{V}_1$ and $\mathbf{V}_2$. We would therefore expect that an estimate of k obtained from a canonical correlation analysis of the eigenvectors of $\mathbf{M1}\mathbf{M1}^T$ and $\mathbf{M2}\mathbf{M2}^T$ would be confirmed by a canonical correlation analysis of the eigenvectors of $\mathbf{M1}^T\mathbf{M1}$ and $\mathbf{M2}^T\mathbf{M2}$.

This methodology might be described as rank estimation by canonical correlation analysis of matrix pairs (RECCAMP). It may be helpful to point out some of the similarities and

0003-2700/89/0361-2219$01.50/0

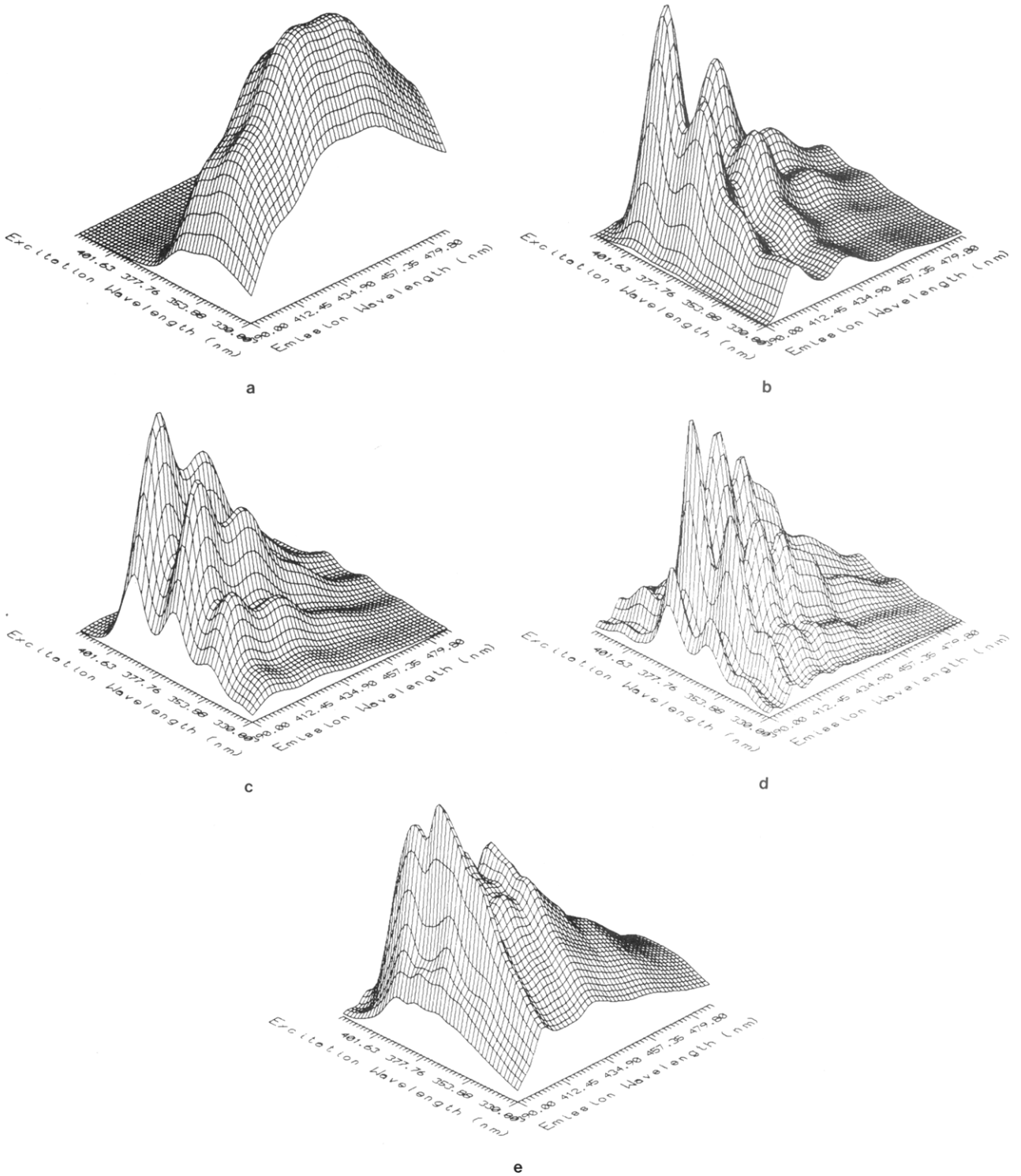

Figure 1. Plots of the noise-free EEM's for (a) B(b)F; (b) B(k)F; (c) BgP; (d) DPA, and (e) mixture of B(b)F, B(k)F, BgP, and DPA.

Table I. Uncorrected Matrix Correlations between the EEM's of the Individual Components in the Simulated Mixture

	B(b)F	B(k)F	BgP	DPA
B(b)F	1	0.364	0.538	0.45
B(k)F		1	0.742	0.816
BgP			1	0.763
DPA				1

differences between RECCAMP and GRAM, the generalized rank annihilation method (*13*, *14*), before proceeding to the next section. On the surface the methods may appear quite similar because both require two data matrices and both rely on eigenanalyses. The similarity ends there, however. GRAM is a calibration method, not a method for rank estimation. The two matrices used in GRAM need not have the same rank as long as all of the components in one of the matrices (the sample) appear also in the other (calibration or calibration + sample). It is pointless to apply GRAM to replicate measures of the same sample. By contrast RECCAMP requires that the two matrices have the same constituent components and necessarily, therefore, the same rank. It is sensible to use RECCAMP on replicate measures of the same sample.

RESULTS AND DISCUSSION

The canonical correlation technique described here was

Table II. Eigenvalues of $c^2\mathbf{S}^T\mathbf{S}$ and the Seven Leading Eigenvalues of $\mathbf{M1}^T\mathbf{M1}$, $\mathbf{M2}^T\mathbf{M2}$, and $\bar{\mathbf{M}}^T\bar{\mathbf{M}}$, Where $\bar{\mathbf{M}} = (1/2)(\mathbf{M1} + \mathbf{M2})$[a]

	eig1	eig2	eig3	eig4	eig5	eig6	eig7
			$c = 25$ $r = 0.6\%$				
$c^2\mathbf{S}^T\mathbf{S}$	6995.7	85.723	5.035	4.093			
$\mathbf{M1}^T\mathbf{M1}$	7000.2	86.043	5.068	4.275	0.179	0.152	0.1374
$\mathbf{M2}^T\mathbf{M2}$	6995.2	87.49	5.033	4.085	0.152	0.142	0.1365
$\bar{\mathbf{M}}^T\bar{\mathbf{M}}$	6997.6	86.712	5.005	4.129	0.08	0.0713	0.0706
			$c = 10$ $r = 1.7\%$				
$c^2\mathbf{S}^T\mathbf{S}$	1119.3	13.72	0.806	0.655			
$\mathbf{M1}^T\mathbf{M1}$	1121.2	13.9	0.861	0.791	0.178	0.152	0.137
$\mathbf{M2}^T\mathbf{M2}$	1119.2	14.48	0.874	0.709	0.149	0.141	0.135
$\bar{\mathbf{M}}^T\bar{\mathbf{M}}$	1120.1	14.14	0.821	0.696	0.0797	0.0712	0.0702
			$c = 6$ $r = 2.9\%$				
$c^2\mathbf{S}^T\mathbf{S}$	402.95	4.938	0.290	0.2357			
$\mathbf{M1}^T\mathbf{M1}$	404.11	5.093	0.372	0.3443	0.177	0.151	0.137
$\mathbf{M2}^T\mathbf{M2}$	402.896	5.438	0.387	0.3067	0.146	0.139	0.133
$\bar{\mathbf{M}}^T\bar{\mathbf{M}}$	403.46	5.2103	0.324	0.2745	0.079	0.071	0.0695
			$c = 5$ $r = 3.5\%$				
$c^2\mathbf{S}^T\mathbf{S}$	279.83	3.43	0.201	0.1637			
$\mathbf{M1}^T\mathbf{M1}$	280.81	3.576	0.2989	0.2599	0.176	0.151	0.137
$\mathbf{M2}^T\mathbf{M2}$	279.796	3.863	0.3088	0.2402	0.1458	0.141	0.132
$\bar{\mathbf{M}}^T\bar{\mathbf{M}}$	280.26	3.664	0.2426	0.2008	0.0787	0.071	0.069
			$c = 4$ $r = 4.3\%$				
$c^2\mathbf{S}^T\mathbf{S}$	179.09	2.195	0.1289	0.1048			
$\mathbf{M1}^T\mathbf{M1}$	179.89	2.334	0.2423	0.1935	0.173	0.1493	0.136
$\mathbf{M2}^T\mathbf{M2}$	179.08	2.562	0.2493	0.1889	0.145	0.1393	0.131
$\bar{\mathbf{M}}^T\bar{\mathbf{M}}$	179.44	2.392	0.1781	0.1407	0.078	0.0708	0.0686
			$c = 3$ $r = 5.8\%$				
$c^2\mathbf{S}^T\mathbf{S}$	100.74	1.234	0.073	0.059			
$\mathbf{M1}^T\mathbf{M1}$	101.36	1.367	0.205	0.172	0.149	0.141	0.135
$\mathbf{M2}^T\mathbf{M2}$	100.75	1.536	0.209	0.155	0.143	0.138	0.131
$\bar{\mathbf{M}}^T\bar{\mathbf{M}}$	101.01	1.395	0.131	0.0978	0.075	0.0699	0.068

[a] r is calculated as the ratio of the standard deviation of the simulated normal noise to the maximum element of $c\mathbf{S}$.

applied to a four-component simulated mixture. The signal spectra were composed of four individual components obtained from the steady-state EEMs of benzo[*b*]fluoranthene [B(b)F], benzo[*k*]fluoranthene [B(k)F], benzo[*ghi*]perylene (BgP) and 9,10-diphenylanthracene (DPA). Each EEM has 50 rows and 50 columns, corresponding to excitation wavelengths from 330 to 420 nm and emission wavelengths from 390 to 490 nm. The normalized rank-one fits of these EEMs are plotted in Figure 1. The normalization condition was $Tr(\mathbf{MM}^T) = 1$, i.e. the Frobenius matrix norm $\|\cdot\|_F$ (*15*, *16*). The uncorrected matrix correlations (*17*) between the EEMs for the individual components are listed in Table I. The noise-free EEM of the simulated mixture is the sum of these rank-one fits and is also plotted in Figure 1.

The data used for the simulation were obtained in the following form:

$$\mathbf{M1} = c\mathbf{S} + \mathbf{N1}, \text{ and} \qquad (4)$$

$$\mathbf{M2} = c\mathbf{S} + \mathbf{N2}$$

where **S** is the noise-free EEM of the mixture, c is a scalar used to control the signal-to-noise ratio, and **N1** and **N2** are simulated white noise matrices. The entries of **N1** and **N2** were generated independently from a normal distribution function with mean 0 and variance 0.001. By varying c in (4), we can also see how the methodology is affected by changes in the signal-to-noise ratio.

Any reasonable method of rank estimation will work well when the noise level is very low and poorly when the noise level is very high. To discern meaningful differences between competing methods, we must evaluate their performances at intermediate levels of noise. By simulating the noise as we have done, we can explore these critical intermediate levels of the signal-to-noise ratio. Notice, though, that our simulated noise has been applied to a signal derived from real data.

For the six selected values of c (25, 10, 6, 5, 4, and 3) the four eigenvalues of $c_2\mathbf{S}^T\mathbf{S}$ (or $c_2\mathbf{S}\mathbf{S}^T$) and the leading seven eigenvalues of $\mathbf{M1}^T\mathbf{M1}$ (or $\mathbf{M1}\mathbf{M1}^T$) and $\mathbf{M2}^T\mathbf{M2}$ (or $\mathbf{M2}\mathbf{M2}^T$) are listed in Table II. These eigenvalues are the squares of the singular values of **S**, **M1**, and **M2**. The EEMs of **M1** for these values of c are plotted in Figure 2 (the EEMs of **M2** are similar to those of **M1** and therefore are not plotted).

Rank estimation based on eigenvalues (or singular values) requires only a single measured EEM, not two. If two measured EEMs are available, it is logical to average them to reduce the noise. Therefore, Table II also contains the eigenvalues obtained from $\bar{\mathbf{M}}$, the average of **M1** and **M2**.

A visual inspection of the eigenvalues in Table II shows that the selected range for c is appropriate for the rank estimation problem based on eigenvalues. At $c = 25$ the break between the fourth and fifth eigenvalues is dramatic, suggesting a confident estimate of 4 for the value of k. At $c = 3$ the break between the fourth and fifth eigenvalues is scarcely noticeable.

To implement the canonical correlation procedure, let $\mathbf{C}_1^i$ and $\mathbf{C}_2^i$ denote the subspaces spanned by the respective i leading eigenvectors of $\mathbf{M1}\mathbf{M1}^T$ and $\mathbf{M2}\mathbf{M2}^T$. Then, for $i = 1, 2, 3$, etc., calculate canonical correlations between $\mathbf{C}_1^i$ and $\mathbf{C}_2^i$. These correlations will necessarily fall in the range between 0 and 1. A correlation not significantly different from 0 is an indication that we have reached the noise. Therefore, at the ith step we test the statistical significance of the ith canonical correlation coefficient (*18*, *19*). If the significance level, i.e. P value is sufficiently small, continue. If not, stop and declare $k = i - 1$. Confirm by repeating the process for $\mathbf{R}_1^i$ and $\mathbf{R}_2^i$, the subspaces spanned by the respective i leading eigenvectors of $\mathbf{M1}^T\mathbf{M1}$ and $\mathbf{M2}^T\mathbf{M2}$.

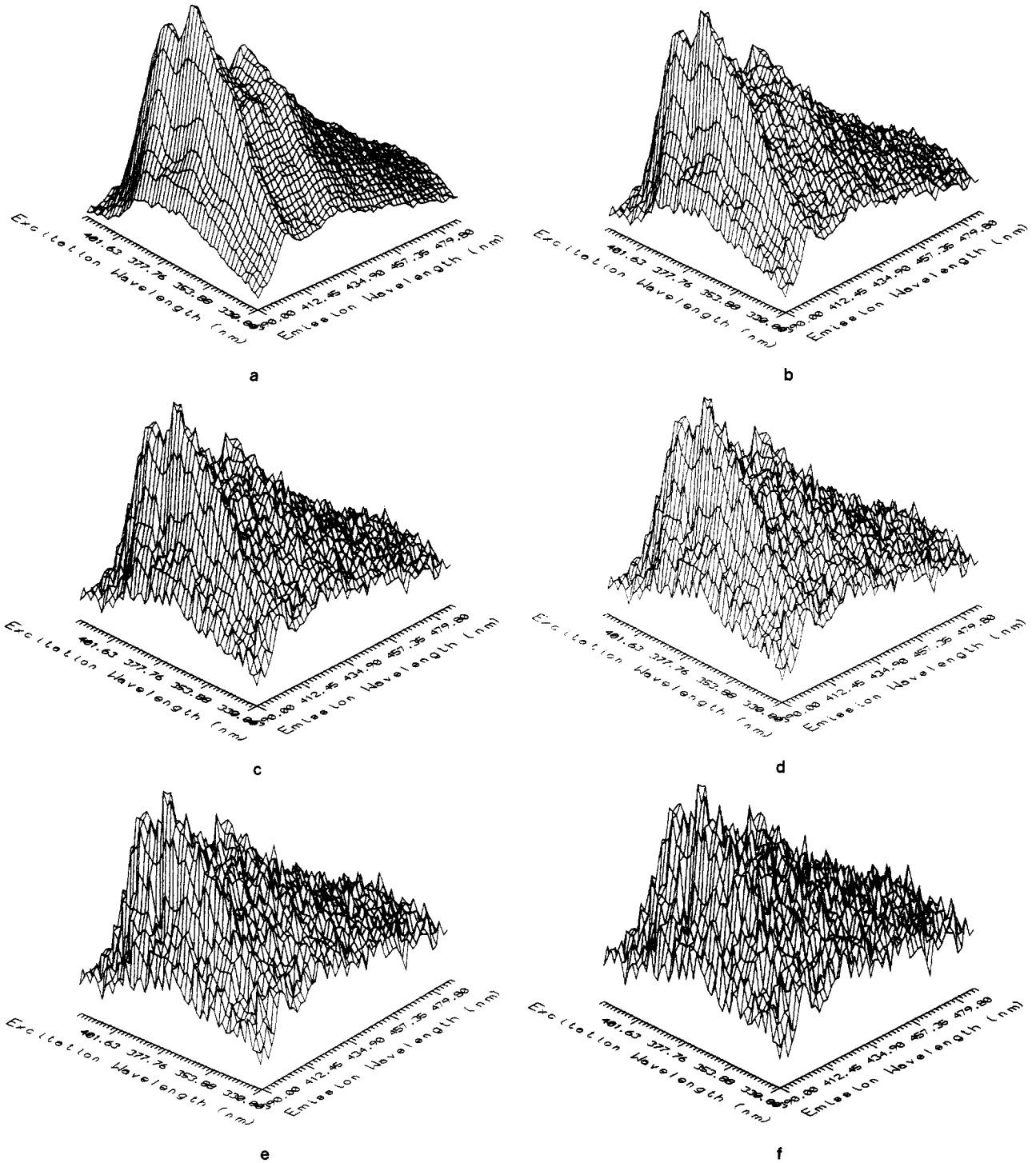

Figure 2. Plots of the noise-simulated EEM **M1** at (a) $c = 25$; (b) $c = 10$; (c) $c = 6$; (d) $c = 5$; (e) $c = 4$, and (f) $c = 3$.

These computations were performed for our simulated data using the SAS statistical program PROC CANCORR. The results for $c = 3, 4, 5$ are presented in Table III and Table IV. Exactly what significance level to use as a cutoff criterion is a fine-tuning question, the answer to which will have to wait for more experience. However, the results are clear in this example. In the general practice of statistical significance testing, the most commonly used cutoff levels for declaring "statistical significance" are 0.10, 0.05 and 0.01. If any of these are applied here the algorithm correctly identifies $k = 4$ when $c = 4$ or 5. At $c = 3$ it picks up three components and misses the fourth.

Is there any reason to continue the calculations for larger values of i once a nonsignificant results has been obtained? Again, a definitive answer will have to wait for more experience, but under certain conditions the answer may be yes. Notice that the third and fourth eigenvalues derived from the signal are not very different. When these eigenvalues are equal, or nearly so, it is possible in the presence of noise that the third canonical correlation would be insignificant at $i = 3$, but that the third and fourth canonical coefficients both become significant when we move to $i = 4$. Despite the closeness of the third and fourth eigenvalues in our example, that phenomenon did not occur in our simulation. It could still happen for some future data set. For completeness the canonical correlation calculations at $i = 6$ and $i = 7$ are included in Tables III and IV. The first seven canonical correlations for all the selected values of c are listed in Table V.

Table III. Canonical Correlations between the Column Subspaces C_2^i and C_2^i and the Corresponding *P* Values for the Hypothesis That the Current and the Smaller Canonical Correlations Are 0

	c = 3		*c* = 4		*c* = 5	
	can. corr	*P* value	can. corr	*P* value	can. corr	*P* value
i = 1	0.9996	0.0001	0.9997	0.0	0.9998	0.0
i = 2	0.9996	0.0	0.9997	0.0	0.9998	0.0
	0.9554	0.0001	0.9743	0.0001	0.9833	0.0001
i = 3	0.9996	0.0	0.9997	0.0	0.9998	0.0
	0.9585	0.0001	0.9766	0.0001	0.9850	0.0001
	0.4300	0.0020	0.6124	0.0001	0.7161	0.0001
i = 4	0.9996	0.0	0.9997	0.0	0.9998	0.0
	0.9601	0.0001	0.9771	0.0001	0.9852	0.0001
	0.5207	0.0059	0.7433	0.0001	0.8445	0.0001
	0.0462	0.7550	0.4708	0.0007	0.6546	0.0001
i = 5	0.9996	0.0001	0.9998	0.0	0.9998	0.0
	0.9606	0.0001	0.9774	0.0001	0.9853	0.0001
	0.5336	0.0706	0.7586	0.0001	0.8528	0.0001
	0.1404	0.9211	0.5195	0.0073	0.6686	0.0001
	0.0287	0.8476	0.0002	0.9987	0.0151	0.9193
i = 6	0.9996	0.0001	0.9998	0.0	0.9998	0.0
	0.9617	0.0001	0.9784	0.0001	0.9862	0.0001
	0.6096	0.0550	0.7770	0.0001	0.8553	0.0001
	0.3075	0.7645	0.5196	0.0773	0.6709	0.0010
	0.1722	0.8386	0.2002	0.7629	0.2026	0.7598
	0.0527	0.7276	0.0404	0.7894	0.0324	0.8304
i = 7	0.9996	0.0001	0.9998	0.0	0.9998	0.0
	0.9656	0.0001	0.9808	0.0001	0.9876	0.0001
	0.6482	0.1478	0.7894	0.0001	0.8656	0.0001
	0.3580	0.9022	0.5684	0.2024	0.7147	0.0049
	0.2051	0.9444	0.2196	0.9213	0.2328	0.9114
	0.1912	0.8068	0.1977	0.7827	0.1901	0.8028
	0.0257	0.8669	0.0340	0.8241	0.0398	0.7949

Table IV. Canonical Correlations between the Row Subspaces R_2^i and R_2^i and the Corresponding *P* Values for the Hypothesis That the Current and the Smaller Canonical Correlations Are 0

	c = 3		*c* = 4		*c* = 5	
	can. corr	*P* value	can. corr	*P* value	can. corr	*P* value
i = 1	0.9994	0.0001	0.9997	0.0	0.9998	0.0
i = 2	0.9995	0.0	0.9997	0.0	0.9998	0.0
	0.9620	0.0001	0.9787	0.0001	0.9864	0.0001
i = 3	0.9995	0.0	0.9997	0.0	0.9998	0.0
	0.9657	0.0001	0.9807	0.0001	0.9877	0.0001
	0.3407	0.0166	0.5482	0.0001	0.6744	0.0001
i = 4	0.9995	0.0	0.9997	0.0	0.9998	0.0
	0.9676	0.0001	0.9811	0.0001	0.9877	0.0001
	0.3712	0.1485	0.6268	0.0001	0.7750	0.0001
	0.0228	0.8774	0.4993	0.0003	0.6593	0.0001
i = 5	0.9995	0.0001	0.9997	0.0	0.9998	0.0
	0.9729	0.0001	0.9818	0.0001	0.9886	0.0001
	0.4327	0.2011	0.6556	0.0001	0.7783	0.0001
	0.2350	0.5576	0.5292	0.0049	0.6636	0.0001
	0.1026	0.4922	0.0779	0.6026	0.0468	0.7545
i = 6	0.9995	0.0001	0.9997	0.0	0.9998	0.0
	0.9773	0.0001	0.9866	0.0001	0.9913	0.0001
	0.4706	0.1945	0.6858	0.0001	0.7903	0.0001
	0.4174	0.3732	0.5745	0.0232	0.7071	0.0002
	0.1621	0.8439	0.1599	0.7654	0.1543	0.7918
	0.0746	0.6218	0.1273	0.3991	0.1213	0.4219
i = 7	0.9995	0.0001	0.9997	0.0	0.9998	0.0
	0.9820	0.0001	0.9900	0.0001	0.9937	0.0001
	0.5581	0.2062	0.7059	0.0007	0.7970	0.0001
	0.4253	0.5477	0.5875	0.0802	0.7189	0.0022
	0.2977	0.7181	0.2957	0.6909	0.2810	0.7456
	0.2178	0.6859	0.2377	0.6276	0.2335	0.6540
	0.0695	0.6497	0.0536	0.7264	0.0388	0.8002

For comparison purposes Table VI presents the results for our data from the imbedded error function and factor indicator function rank estimation methods described by Malinowski (*4*, *5*). Both these methods are based exclusively on the eigenvalues. As Table VI shows, both methods correctly identify $k = 4$ when $c = 25$ or $c = 10$. At $c = 6$ the IE method misses two of the components, and at lower values of c both methods incorrectly give 2 as the value of k.

In summary, when at least two data matrices are available for linear mixtures containing the same constituents, infor-

Table V. Canonical Correlations between the Column Subspaces C_1^7 and C_2^7 and between the Row Subspaces R_1^7 and R_2^7 and the Corresponding *P* Values for the Hypothesis That the Current and the Smaller Canonical Correlations Are 0

column space		row space		column space		row space	
can. corr	*P* val	can. corr	*P* val	can. corr	*P* val	can. corr	*P* val
	c = 25				*c* = 10		
1.0000	0.0	1.0000	0.0	1.0000	0.0	1.0000	0.0
0.9995	0.0	0.9997	0.0	0.9969	0.0001	0.9984	0.0001
0.9950	0.0001	0.9909	0.0001	0.9686	0.0001	0.9448	0.0001
0.9883	0.0001	0.9879	0.0001	0.9261	0.0001	0.9245	0.0001
0.2611	0.8970	0.2938	0.8641	0.2525	0.9009	0.2894	0.8443
0.1647	0.8765	0.1372	0.9371	0.1729	0.8519	0.1642	0.8835
0.0308	0.8407	0.0066	0.9654	0.0390	0.7991	0.0105	0.9451
	c = 6				*c* = 5		
0.9999	0.0	0.9999	0.0	0.9998	0.0	0.9998	0.0
0.9914	0.0001	0.9957	0.0001	0.9876	0.0001	0.9937	0.0001
0.9092	0.0001	0.8540	0.0001	0.8656	0.0001	0.7970	0.0001
0.8005	0.0001	0.7995	0.0001	0.7147	0.0049	0.7189	0.0022
0.2406	0.9059	0.2795	0.7925	0.2328	0.9114	0.2810	0.7456
0.1845	0.8182	0.2130	0.7348	0.1901	0.8028	0.2335	0.6540
0.0420	0.7837	0.0269	0.8603	0.0398	0.7949	0.0388	0.8002
	c = 4				*c* = 3		
0.9998	0.0	0.9997	0.0	0.9996	0.0001	0.9995	0.0001
0.9808	0.0001	0.9900	0.0001	0.9656	0.0001	0.9820	0.0001
0.7894	0.0001	0.7059	0.0007	0.6482	0.1478	0.5581	0.2062
0.5684	0.2024	0.5875	0.0802	0.3580	0.9022	0.4253	0.5477
0.2196	0.9213	0.2957	0.6909	0.2051	0.9444	0.2977	0.7181
0.1977	0.7827	0.2377	0.6276	0.1912	0.8068	0.2178	0.6859
0.0340	0.8241	0.0536	0.7264	0.0257	0.8669	0.0695	0.6497

Table VI. Values of the Imbedded Error Function (IE) and the Factor INDicator Function (IND) Calculated from the Eigenvalues of $T^T\tilde{M}$, Where Asterisk Indicates the Number Determined by the Respective Criterion

IE	IND	IE	IND
$c = 25$		$c = 10$	
0.071 452 22	0.000 549 225 7	0.029 573 8	0.000 227 322 4
0.033 220 82	0.000 195 297 7	0.016 358 37	0.000 096 167 17
0.029 432 07	0.000 153 292 1	0.016 651 65	0.000 086 727 36
0.014 836 16*	0.000 072 864 8*	0.014 822 12*	0.000 072 795 83*
0.016 207 69	0.000 077 816 7	0.016 194 21	0.000 077 751 94
0.017 387 01	0.000 083 635 9	0.017 372 44	0.000 083 565 79
0.018 341 97	0.000 090 057 3	0.018 330 81	0.000 090 002 48
$c = 6$		$c = 5$	
0.018 839 17	0.000 144 809 5	0.016 279 75	0.000 125 136 2
0.012 901 66*	0.000 075 845 9	0.012 214 1*	0.000 071 803 95*
0.014 299 05	0.000 074 474 2	0.013 838 56	0.000 072 075 85
0.014 798 52	0.000 072 679 9*	0.014 783 17	0.000 072 604 55
0.016 171 71	0.000 077 643 9	0.016 157 43	0.000 077 575 34
0.017 347 47	0.000 083 445 6	0.017 331 64	0.000 083 369 52
0.018 310 1	0.000 089 900 8	0.018 296 46	0.000 089 833 82
$c = 4$		$c = 3$	
0.013 826 32	0.000 106 277 6	0.011 547 13	0.000 088 758 38
0.011 624 5*	0.000 068 377 8*	0.011 148 95*	0.000 065 542 2*
0.013 438 51	0.000 069 992 2	0.013 104 95	0.000 068 254 94
0.014 754 58	0.000 072 464 1	0.014 683 68	0.000 072 115 93
0.016 131 74	0.000 077 452	0.016 072 22	0.000 077 166 27
0.017 303 53	0.000 083 234 3	0.017 242 63	0.000 082 941 38
0.018 271 4	0.000 089 710 8	0.018 213 31	0.000 089 425 57

mation useful for rank estimation can be extracted from canonical correlations of the eigenvectors. The example presented here suggests that an algorithm incorporating canonical correlations of the eigenvectors will outperform traditional methods, but more experience is needed to fine tune the parameters of the algorithm.

LITERATURE CITED

(1) Rasmussen, G. T.; Isenhour, T. L.; Lowry, S. R.; Ritter, G. L. *Anal. Chim. Acta* **1978**, *103*, 213.
(2) Tway, P. C.; Cline Love, L. J.; Woodruff, H. B. *Anal. Chim. Acta* **1980**, *117*, 45.
(3) Gold, H. S.; Rasmussen, G. T.; Mercer-Smith, J. A.; Whitten, D. G.; Buck, R. P. *Anal. Chim. Acta* **1980**, *122*, 171.
(4) Malinowski, E. R.; Howery, D. G. In *Factor Analysis in Chemistry*; Wiley: New York, 1980.
(5) Malinowski, E. R. *Anal. Chem.* **1977**, *49*, 612.
(6) Gabriel, K. R. *Biometrika* **1971**, *58*, 453.
(7) Rossi, T. M.; Warner, I. M. *Anal. Chem.* **1986**, *54*, 810–815.
(8) Warner, I. M. In *Contemporary Topics in Analytical and Clinical Chemistry*; Hercules, D. M., Hieftje, G. M., Snyder, L. R., Evenson, M. A., Eds.; Plenum Press: New York, 1982; Vol. 4.
(9) Warner, I. M.; Neal, S. L.; Rossi, T. M. *J. Res. Natl. Bur. Stand. (U.S.)* **1985**, *90*, 487.
(10) Mardia, K. V.; Kent, J. T.; Bibby, J. M. *Multivariate Analysis*; Academic Press: London, 1972.
(11) Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*; 2nd ed.; Prentice Hall: Engelwood Cliff, NJ, 1988.
(12) Tatsuoka, M. M. *Multivariate Analysis*; John Wiley and Sons: New York, 1971.
(13) Sanchez, E.; Kowalski, B. R. *Anal. Chem.* **1986**, *58*, 496–499.
(14) Sanchez, E.; Kowalski, B. R. *J. Chemom.* **1988**, *2*, 247–263 and 265–280.
(15) Golub, G.; Van Loan, C. *Matrix Computations*; The Johns Hopkins University Press: Oxford, 1983.
(16) Stewart, G. W. *Introduction to Matrix Computations*; Academic Press: New York, 1973.
(17) Burdick, D. S.; Tu, X. M. *J. Chemom.* **1989**, *3*, 431–441.
(18) Rao, C. R. *Linear Statistical Inference*; John Wiley and Sons: New York, 1973.
(19) Kshirsagar, A. M. *Multivariate Analysis*; Marcel Dekker: New York, 1972.

Xin M. Tu
Donald S. Burdick*
Institute of Statistics and Decision Sciences
Department of Mathematics
Duke University
Durham, North Carolina 27706

David W. Millican
Linda B. McGown
Department of Chemistry
Duke University
Durham, North Carolina 27706

RECEIVED for review April 10, 1989. Accepted June 14, 1989. L. B. McGown is grateful for support for this project from the Department of Energy (Grant Number DE-FG05-88ER13931).

CORRECTION

Separation of Two Components of Horse Myoglobin by Isoelectric Focusing Field-Flow Fractionation

J. Chmelík, M. Deml, and J. Janča (*Anal. Chem.* **1989**, *61*, 912–914).

The manuscript received and accepted dates for this correspondence were inadvertently omitted on publication. The correct dates are as follows: Received for review April 11, 1988. Revised manuscript received December 30, 1988. Accepted January 17, 1989.