Table III. ^{6}Li NOE of **1** in Diethyl Ether: % ^{6}Li NMR Enhancement

proton irrad	CH_3	H(syn)	H(anti) less assoc	H(anti) more assoc
3 s	12	25	29.5	29.5
10 s		28	39	39

sured unless the line-shape calculation takes account of both processes. Further, NMR studies of allyllithium compounds must be carried out as a function of RLi concentration, ligand, concentration, and temperature, the latter as low as is consistent with good NMR resolution.

This work emphasizes the utility of NOE measurements to determine the relative positions of hydrogens relative to lithium. Further, we establish the necessity of obtaining NMR data at as low a temperature as is practical in order to minimize the effects of exchange processes.

Experimental Section

NMR spectroscopy was carried out using Bruker equipment, AM-500 for protons and MSL 300 for ^{6}Li and ^{13}C. Relevant instrumental parameters are listed in Table II.

NOE effects were measured for ^{6}Li NMR spectra by selectively irradiating hydrogen resonances. In these experiments, first both proton and lithium NMR spectra were recorded. A proton frequency list was then created including the following resonances: methyl, CH(syn), CH(anti) both species, and a frequency approximately 10 ppm removed from all resonances, selected as null.

In a typical accumulation the proton resonance was irradiated for 3 s, then the decoupler turned off and the ^{6}Li spectrum recorded (see Table III). These results establish the relative proximities of ^{6}Li and different protons. Because of a slow but significant interspecie exchange rate, the above results are probably somewhat attenuated.

Tetrakis(2-methylallyl)tin. The general procedure follows that of Fishwich and Wallbridge.[17] A 2-L three-neck flask fitted with an overhead stirrer, a reflux condenser, and an addition funnel was flame-dried while flushing with argon. Magnesium (25 g) was added along with a crystal of iodine and 25 mL of THF. Methallyl chloride (10 g) was added and the solution stirred without further addition until the yellow/brown color of the iodine solution was dissipated. Once this occurred the solution was cooled with ice, and a solution of allyl chloride (85 g) and tin tetrachloride (25 mL, 55.7 g, 0.21 M) in hexane (200 mL) was added dropwise. After approximately 15 mL an additional liter of THF was added to dissolve the Grignard reagent. After complete addition the mixture was stirred at room temperature overnight and then heated to reflux for 0.5 h. After cooling to room temperature this reaction mixture was poured into an ice/ammonium chloride slurry and the organic layer extracted with hexane. The combined organic layers were dried over $MgSO_4$ and filtered; then the solvent was removed in vacuo, yielding 36 g of the title compound, bp 134 °C (4 Torr), 50% yield; proton NMR (250 MHz, $CDCl_3$) δ 1.71 (s, 3 H), 1.91 (s, 2 H), 4.54, 5.24.

2-Methylallyllithium.[18] All of the following procedures were performed in a Vacuum-Atmospheres inert atmosphere box, with argon. Glassware was oven-dried at 130 °C and cooled in vacuo.

Freshly distilled and degassed tetrakis(2-methylallyl)tin (1.7 g) was weighed into a 150-mL Schlenk flask. Diethyl ether (25 mL, freshly vacuum transferred from Na/K benzophenone ketyl) was added and stirring begun. *n*-Butyllithium-^{6}*Li* (4 mL, 2.6 M in hexane) was added via syringe over 15 min. The yellow solution was stirred 1 h. The diethyl ether was removed in vacuo (in the glove box). A yellow paste formed together with a layer of dibutyldiallyltin. Three times pentane (10 mL) was added; the phases were mixed for 30 min and the pentane was poured off. The viscous allyllithium etherate was evacuated for a short period to remove pentane and then dissolved in ether (5 mL). This reaction is essentially quantitative. An NMR sample of this solution showed no impurities.

Acknowledgment. This research was partly supported by the National Science Foundation, Grant No. CHE 8304636, which also helped finance the high-field NMR equipment used in this work. We thank Dr. Charles Cottrell, Chemical Instrumentation Center, The Ohio State University, for assistance with the NMR technology.

Registry No. 2-Methylallyllithium, 61777-16-0; tetrakis(2-methylallyl)tin, 39482-35-4; methallyl chloride, 563-47-3; tin tetrachloride, 7646-78-8; *n*-butyllithium-^{6}*Li*, 20535-67-5.

---

(17) Fishwick, M.; Wallbridge, M. G. H. *J. Organomet. Chem.* **1970**, *25*, 69.

(18) Seyferth, D.; Weiner, M. A. *J. Org. Chem.* **1961**, *26*, 4797.

Multicomponent Quantitative Analysis Using Second-Order Nonbilinear Data: Theory and Simulations

Bruce E. Wilson,[‡] Walter Lindberg,[†] and Bruce R. Kowalski*

Contribution from the Laboratory for Chemometrics, Department of Chemistry, BG-10, University of Washington, Seattle, Washington 98195. *Received July 22, 1988*

Abstract: The application of rank annihilation to a class of instruments generating data which can be classified as second-order nonbilinear is described. This method permits determining the concentration of an analyte of interest even in the presence of spectral interferents unknown and unaccounted for during calibration. The primary difficulty in applying rank annihilation to nonbilinear data is that multiple, different concentration estimates are obtained; it is shown that, for two-dimensional mass spectrometry and absolute value mode two-dimensional nuclear magnetic resonance spectroscopy, the correct concentration estimate should be the smallest of those obtained. The method provides a back-estimate of the calibration matrix, which can be used as an indication of the accuracy of the resulting concentration estimate, as well as a diagnostic for the presence of matrix effects between the calibration and unknown samples. In this paper, the theory and motivation for rank annihilation is presented, and the application to nonbilinear data is presented using computer simulations and a test system consisting of 2D *J*-coupled NMR spectra (COSY) of six sugars in D_2O.

In modern analytical chemistry, one of the fastest growing areas for practical applications is the use of multivariate calibration for multicomponent quantitative analysis.[1] Multivariate methods are necessary in order to utilize the tremendous amounts of data that can be collected with modern analytical instruments and laboratory computers, and because fast, inexpensive calibration methods can largely replace slower, more expensive direct determination methods. A good example of this is the use of near-infrared reflectance spectroscopy to routinely measure the protein content of wheat, while the Kjeldahl nitrogen analysis is

---

[‡] Present address: Eastman Chemicals Research Labs, P.O. Box 1972, Building 95A, Lincoln, Rd., Kingsport, TN 37662.

[†] Current address: Swedish National Agricultural Laboratory, Box 5097, 900 05 Umeå, Sweden.

* Corresponding author.

(1) Beebe, K. R.; Kowalski, B. R. *Anal. Chem.* **1987**, *59*, 1007A-1017A.

0002-7863/89/1511-3797$01.50/0

necessary only to determine the protein content for calibration samples.[2,3]

The simplest type of calibration, so-called univariate calibration, regresses a single measured property, such as the sample's absorbance at a single wavelength, against one or more standards of known analyte concentrations.[4,5] Such a method has the advantage that it is comparatively easy to understand and implement—most modern scientific calculators have this capability built in. However, such a calibration method is extremely limited in practice, because it assumes that the measured response is influenced solely by the analyte of interest. This requirement is termed the unique signal requirement, and there is, unfortunately, no a priori method of checking its validity for a given sample.[6] Further, even if the unique signal requirement is met, it is not possible to use a univariate calibration model to determine the concentration of more than one chemical component.

By contrast, when using multivariate methods, such as measuring the absorbance at several wavelengths and using principal components regression (PCR) to correlate the absorbances with the concentrations of the analyte(s) of interest, it is possible to simultaneously determine several chemical components, as well as components for which no unique signal channel exists.[1,4,5] It is also possible to use residual analysis in a multivariate calibration model to test whether the unknown sample contains chemical components not present in the calibration samples.[4,7,8] These additional components are called spectral interferents, and they render the usual univariate and multivariate models invalid for predicting analyte concentrations in the unknown sample.[6] At least for multivariate methods, it is generally possible to detect spectral interferents and thereby know that the model is invalid; for univariate methods the same facts which preclude verifying the unique signal requirement preclude the detection of spectral interferents.[6]

Certain multivariate methods, notably curve resolution, can sometimes be used to estimate the concentrations of one or two analytes when an unknown spectral interferent is present, but these methods generally require assumptions which cannot be verified a priori. For many of these methods, the principal requirement is that one of the signal channels must respond solely to the analyte of interest.[9–11] Specifically which channel is unique for the analyte of interest does not need to be known, only that such a channel exists. However, as for univariate methods, it is not possible to verify that this assumption holds.[6]

One solution to the problem of spectral interferents was developed by Ho and co-workers,[12] a calibration method which they called rank annihilation factor analysis and which is applicable only to data from certain types of analytical instruments. Examples of data to which rank annihilation may be applied include those from chromatography with a multichannel (wavelength) detector (e.g., LC–UV, GC–MS) and fluorescence excitation–emission matrices. The important characteristics of these instruments are that (a) each sample yields a matrix of data (as opposed to a single scalar or a vector of data), termed the response matrix, and (b) the rank of a response matrix for a pure chemical component is one, in the absence of noise. Methods and the data thereby produced which meet these two requirements are classified as second-order bilinear.[12,13] To date, rank annihilation methods have been successfully applied to fluorescence excitation–emission,[12,14,15] thin layer chromatography with multichannel UV detection,[16] and LC–UV.[17–19]

Furthermore, using an advanced formulation of rank annihilation—the generalized rank annihilation method (GRAM)—it is possible to quantitate for multiple analytes using a single calibration sample,[20] even in the presence of unknown spectral interferents. GRAM is also more powerful in that it provides the pure component profiles (e.g., UV spectra and elution profiles in LC–UV) as a consequence of solving for the concentration estimates, which is useful for comparison against library spectra as well as for qualitative analysis.

The successes in applying rank annihilation to real chemical problems are most remarkable in that quantitation for the analytes of interest in the presence of unknown spectral interferents is possible, but without the requirement of having a unique signal channel. Unfortunately, the requirements of rank annihilation have restricted the analytical instruments to which it may be applied. In particular, it has not been possible to apply rank annihilation to two-dimensional mass spectrometry (MS/MS) or two-dimensional nuclear magnetic resonance (2D NMR). These instruments are both capable of generating a matrix of data for each sample, but the rank of one of these pure component response matrices is generally substantially larger than one, and so MS/MS and 2D NMR (and the data they generate) are classified as second-order nonbilinear.

In this paper it will be demonstrated that rank annihilation based methods can be used with nonbilinear data to solve the background problem and estimate single analyte concentrations, at least under certain circumstances. However, the advantages GRAM offers for multiple analyte quantitation from a single calibration sample and for estimating the pure component response matrices from that single calibration sample are lost.

Theory

The terminology for classifying types of instruments and the data they produce is as suggested by Sanchez and Kowalski[21,22] and is based upon the order of the tensor that can be used to represent the data generated by the instrument for a single sample. A pH meter and a single wavelength spectrophotometer both generate a single datum for each sample, and so are classified as zeroth-order instruments. A spectrometer or sensor array[23] generates a vector of data (e.g., absorbance at multiple wavelengths) which can be represented as a first-order tensor, and so are both first-order instruments. Note that first order is also sometimes used to refer to instruments which have a linear response with concentration (i.e., a first-order polynomial); in this paper the term first order will always refer to instruments which produce a vector of data per sample or to the data itself. Hyphenated chromatographic methods, such as LC–UV, and other so-called two-dimensional experiments such as 2D NMR and MS/MS generate a matrix of data per sample and so are second-order instruments. However, not all hyphenated techniques are second order; for example, plasma emission spectroscopy

---

(2) Naes, T.; Martens, H. *J. Chemometrics* **1988**, *2*, 155–167.

(3) Williams, P.f Norris, K., Eds. *Near-infrared Technology in the Agricultural and Food Industries*; American Association of Cereal Chemists: St. Paul, MN, 1987.

(4) Weisburg, S. *Applied Linear Regression*, 2nd ed.; Wiley: New York, 1985.

(5) Draper, N. R.; Smith, H. *Applied Regression Analysis*, 2nd ed.; Wiley: New York, 1981.

(6) Osten, D. W.; Kowalski, B. R. *Anal. Chem.* **1985**, *57*, 908–917.

(7) Chatterjee, S. *Stat. Sci.* **1986**, *1*, 379–416.

(8) Lorber, A.; Veltkamp, D.; Kowalski, B. R., Outliers Analysis in Multivariate Calibration, submitted to *Anal. Chem.*

(9) Caruna, R. A.; Searle, R. B.; Heller, T.; Shupack, S. I. *Anal. Chem.* **1986**, *58*, 1162–1167.

(10) Kawata, S.; Komeda, H.; Sasaki, K.; Minima, S. *Appl. Spectrosc.* **1985**, *39*, 610–614.

(11) Malinowski, E. R. *Anal. Chim. Acta* **1982**, *134*, 129–137.

(12) Ho, C.-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1978**, *50*, 1108–1113.

(13) Hirschfeld, T. *Anal. Chem.* **1980**, *52*, 297A–312A.

(14) Ho, C.-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1980**, *52*, 1071–1079.

(15) Ho, C.-N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1981**, *53*, 92–98.

(16) Gianelli, M. L.; Burns, D. H.; Callis, J. B.; Christian, G. D.; Anderson, N. H. *Anal. Chem.* **1983**, *55*, 1858–1862.

(17) McCue, M.; Malinowski, E. R. *J. Chromatogr. Sci.* **1983**, *21*, 229–234.

(18) Sanchez, E.; Ramos, L. S.; Kowalski, B. R. *J. Chromatogr.* **1987**, *385*, 151–164.

(19) Ramos, L. S.; Sanchez, E.; Kowalski, B. R. *J. Chromatogr.* **1987**, *385*, 165–180.

(20) Sanchez, E.; Kowalski, B. R. *Anal. Chem.* **1986**, *58*, 496–499.

(21) Sanchez, E.; Kowalski, B. R. *J. Chemometrics* **1988**, *2*, 247–263.

(22) Sanchez, E.; Kowalski, B. R. *J. Chemometrics* **1988**, *2*, 265–280.

(23) Carey, W. P.; Beebe, K. R.; Kowalski, B. R. *Anal. Chem.* **1987**, *59*, 1529–1534.

coupled with mass spectrometry to form ICP–AES–MS is a first-order method because one does not measure a mass spectrum of a specific atomic emission line or vice versa. In LC–UV, by comparison, one does measure the UV spectrum at a specific chromatographic retention time.[13]

Within second order, a distinction is made between bilinear and nonbilinear data, based upon the mathematical rank of the response matrix for a pure component. For a bilinear second-order instrument, such as hyphenated chromatography, the rank of a pure component response matrix is unity. For a nonbilinear second-order instrument, such as MS/MS or most common 2D NMR experiments (e.g., COSY, NOESY), the rank of a pure component response matrix is larger than one. Another way of expressing the difference between bilinear and nonbilinear instruments is that for a bilinear instrument, the response function $\mathrm{R}(x_1,x_2)$ is separable into two independent response functions, $\mathrm{R}_1(x_1)\mathrm{R}_2(x_2)$, each a function of one variable only. For a nonbilinear instrument, $\mathrm{R}(x_1,x_2)$ is not separable. In chemical terms, in LC–UV the elution profile and the UV spectra are independent of one another, but in MS/MS the daughter spectrum is a function of the parent ion mass.

For a second-order bilinear instrument, the m rows by n columns response matrix $\mathbf{N}_k$ for a pure chemical component k can be represented as

$$\mathbf{N}_k = \mathbf{x}_k c_{k,\mathrm{N}} \mathbf{y}_k^{\mathrm{T}} + \mathbf{E} \tag{1}$$

where $\mathbf{x}_k$ and $\mathbf{y}_k$ are column vectors which are the pure component response in the row and column spaces, $c_{k,\mathrm{N}}$ is the scalar concentration in some appropriate set of units so that $\|\mathbf{x}_k\| = \|\mathbf{y}_k\| = 1$, where $\|\cdot\|$ represents the Euclidian vector norm, $\mathbf{E}$ represents unmodelled experimental error, and T indicates vector (or matrix) transposition. To the extent that $\mathbf{E}$ can be neglected, $\mathbf{N}_k$ has rank one. In the case of LC–UV, $\mathbf{x}_k$ would be the pure component UV spectrum and $\mathbf{y}_k$ would be the pure component chromatographic profile. Note also that the matrix formed from the outer product of two vectors, e.g., $\mathbf{x}_k y_k^{\mathrm{T}}$, is sometimes referred to as a diad.

If $\mathbf{M}$ is a bilinear response matrix for a mixture containing K chemical components then, assuming linear superposition of spectra, the bilinear expansion of $\mathbf{M}$ is given by

$$\mathbf{M} = \sum_{k=1}^{K} \frac{c_{k,\mathrm{M}}}{c_{k,\mathrm{N}}} \mathbf{N}_k = \mathbf{X}\mathbf{C}_{\mathrm{M}}\mathbf{Y}^{\mathrm{T}} \tag{2}$$

where $c_{k,\mathrm{M}}$ and $c_{k,\mathrm{N}}$ are the concentrations of the kth chemical component in the mixture and calibration samples, respectively, the $\mathbf{N}_k$ are the pure component response matrices, $\mathbf{X}$ is a matrix whose columns are the pure component row space responses (e.g., spectra), $\mathbf{C}_{\mathrm{M}}$ is a diagonal matrix whose elements are the concentrations in appropriate units so that the columns of $\mathbf{X}$ and $\mathbf{Y}$ are normalized, and $\mathbf{Y}$ is a matrix whose columns are the pure component column space responses (e.g., chromatographic profiles). Note that the columns of $\mathbf{X}$ and $\mathbf{Y}$ are not necessarily orthogonal, as would be the case for the singular value decomposition (SVD) of $\mathbf{M}$. However, if the responses of the K components are linearly independent, then the rank of $\mathbf{M}$ is nonetheless equal to K.

The original rank annihilation method[12] is based on considering a difference matrix $\mathbf{B}$ of the form

$$\mathbf{B}(\alpha) = \mathbf{M} - \alpha\mathbf{N}_p \tag{3}$$

where $\mathbf{N}_p$ is the response matrix of the analyte of interest, designated as component p, where $p \in \{1 \ldots K\}$, and α is a nonnegative scalar constant. Note that when α is equal to the ratio $\alpha_0 = c_{p,\mathrm{M}}/c_{p,\mathrm{N}}$, then the rank of $\mathbf{B}$ is reduced by one unit:

$$\mathbf{B}(\alpha_0) = \mathbf{M} - \frac{c_{p,\mathrm{M}}}{c_{p,\mathrm{N}}}\mathbf{N}_p = \sum_{\substack{k=1 \\ k \neq p}}^{K} \frac{c_{p,\mathrm{M}}}{c_{p,\mathrm{N}}}\mathbf{N}_k \tag{4}$$

The bilinear expansion for $\mathbf{B}(\alpha_0)$ contains $K - 1$ linearly independent terms, and its rank is therefore equal to $K - 1$.

As originally proposed by Ho and co-workers,[12] the implementation of the concept embodied in eq 3 and 4 involves finding the minimum in the eigenvalues of $\mathbf{B}$ as a function of α. However, this method is iterative and computationally intensive, despite later algorithmic improvements.[14,15] Lorber[24,25] then noted that minimizing the rank of $\mathbf{B}$ in eq 3 is equivalent to solving a generalized eigenproblem of the form

$$\mathbf{M}\mathbf{z} = \mathbf{N}_p\mathbf{z}\lambda \tag{5}$$

where λ is the eigenvalue and $\mathbf{z}$ is the eigenvector. Because $\mathbf{N}_p$ is a rank one matrix, there is exactly one nonzero eigenvalue, which is equal to the concentration ratio $\lambda = \alpha_0 = c_{p,\mathrm{M}}/c_{p,\mathrm{N}}$. Therefore, the noniterative rank annihilation solution to the calibration problem for the bilinear second-order case is found by solving the corresponding eigenvalue–eigenvector problem.

An interesting point, noted by Sanchez and Kowalski[21] and related to later work by Lorber,[26] is that $\mathbf{z}$ is also chemically significant; it is the portion of $\mathbf{y}_p$ which is orthogonal to the remaining $\mathbf{y}_k$, an assertion which can be verified by noting that the solution to the eigenproblem is unique. In the nomenclature of Sanchez and Kowalski's work,[21,22] $\mathbf{z}$ is the contravariant of $\mathbf{y}_p$; in Lorber's nomenclature $\mathbf{z}$ is the net analyte signal of $\mathbf{y}_p$.

However, rank annihilation of the form in eq 5 is limited because $\mathbf{N}_p$ is the response of the pure analyte of interest. A more general formulation of rank annihilation, one which does not require pure component response matrices for calibration, was derived by Sanchez and Kowalski[20] by considering a representation of the form

$$\mathbf{M} = \mathbf{X}\mathbf{C}_{\mathrm{M}}\mathbf{Y}^{\mathrm{T}} \tag{6a}$$

$$\mathbf{N} = \mathbf{X}\mathbf{C}_{\mathrm{N}}\mathbf{Y}^{\mathrm{T}} \tag{6b}$$

where the calibration spectrum, $\mathbf{N}$, may contain more than one chemical component. Equation 5 then generalizes to the generalized eigenproblem

$$\mathbf{N}\mathbf{Z} = \mathbf{M}\mathbf{Z}\boldsymbol{\Lambda} \tag{7}$$

where $\mathbf{Z}$ is now a matrix whose columns are eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues. As in eq 5, the eigenvalues contain the concentration ratios: $\boldsymbol{\Lambda} = \mathbf{C}_{\mathrm{N}}\mathbf{C}_{\mathrm{M}}^{-1}$, but even more powerful is the fact that $\mathbf{Z}$ is the pseudo-inverse of $\mathbf{Y}^{\mathrm{T}}$—the pure component responses in the column space (e.g., chromatographic elution profiles). This assertion can be quickly verified by noting that the solution to eq 7 is unique and substituting $\mathbf{M} = \mathbf{X}\mathbf{C}_{\mathrm{M}}\mathbf{Y}^{\mathrm{T}}$, $\mathbf{N} = \mathbf{X}\mathbf{C}_{\mathrm{N}}\mathbf{Y}^{\mathrm{T}}$, $\boldsymbol{\Lambda} = \mathbf{C}_{\mathrm{N}}\mathbf{C}_{\mathrm{M}}^{-1}$, and $\mathbf{Z} = (\mathbf{Y}^{\mathrm{T}})^{+}$ into eq 7:

$$(\mathbf{X}\mathbf{C}_{\mathrm{N}}\mathbf{Y}^{\mathrm{T}})(\mathbf{Y}^{\mathrm{T}})^{+} = (\mathbf{X}\mathbf{C}_{\mathrm{M}}\mathbf{Y}^{\mathrm{T}})(\mathbf{Y}^{\mathrm{T}})^{+}(\mathbf{C}_{\mathrm{N}}\mathbf{C}_{\mathrm{M}}^{-1}) \tag{8}$$

Because $\mathbf{C}_{\mathrm{N}}\mathbf{C}_{\mathrm{M}}^{-1}$ is a diagonal matrix within the space of $\mathbf{Y}$, it is unchanged by the projection matrix $\mathbf{Y}^{\mathrm{T}}(\mathbf{Y}^{\mathrm{T}})^{+}$, where + indicates the generalized inverse, so that eq 8 yields an identity. In practice, it is possible for some of the diagonal elements of $\mathbf{C}_{\mathrm{M}}$ to be zero, if the calibration sample (N) contains chemical components not present in the mixture sample (M), so that it may be necessary to replace $\mathbf{M}$ in eq 7 with $\mathbf{M} + \mathbf{N}$, so that $\boldsymbol{\Lambda}$ becomes $\mathbf{C}_{\mathrm{N}}(\mathbf{C}_{\mathrm{N}} + \mathbf{C}_{\mathrm{M}})^{-1}$, which does not have zero elements in the denominator.[20]

Wilson and co-workers[27] then noted that eq 7 can be solved by using the QZ algorithm if projection matrices are used to reduce $\mathbf{M}$ and $\mathbf{N}$ to square matrices. This method has the advantage that the diagonal elements of $\boldsymbol{\Lambda}$ are calculated in the form $\Lambda_{ij} = \alpha_i/\beta_i$, where α_i and β_i are separately calculated scalar quantities, which eliminates the need to replace $\mathbf{M}$ with $\mathbf{M} + \mathbf{N}$ when some of the $c_{k,\mathrm{M}}$ may be zero. Using the QZ algorithm is also computationally more robust and faster than the transformations proposed by Sanchez and Kowalski[20] for solving eq 7.

Given that $\mathbf{Z} = (\mathbf{Y}^{\mathrm{T}})^{+}$, then the pure component responses in the other order $\mathbf{X}$ may be estimated as

$$\hat{\mathbf{C}}_{\mathrm{N}}\hat{\mathbf{X}} = \mathbf{M}\mathbf{Z}\boldsymbol{\Lambda} \tag{9}$$

---

(24) Lorber, A. *Anal. Chim. Acta* **1984**, *164*, 293–297.
(25) Lorber, A. *Anal. Chem.* **1985**, *57*, 2395–2397.
(26) Lorber, A. *Anal. Chem.* **1986**, *58*, 1167–1172.
(27) Wilson, B. E.; Sanchez, E.; Kowalski, B. R. *J. Chemometrics*, in press.

where $\hat{\mathbf{C}}_N$ and $\hat{\mathbf{X}}$ denote estimated quantities. $\hat{\mathbf{X}}$ may be isolated by normalizing the columns of $\mathbf{MZ\Lambda}$; the resultant normalization factors are the diagonal elements of $\hat{\mathbf{C}}_N$. The columns of $\hat{\mathbf{X}}$ may then be checked against library spectra for the pure components as a check on the quality of the quantitation.[17,18] Alternatively, if $\mathbf{M}$ and $\mathbf{N}$ both contain unidentified chemical components, the estimated spectra in $\hat{\mathbf{X}}$ and elution profiles in $\hat{\mathbf{Y}} = (\mathbf{Z}^+)^T$ may be used for qualitative analysis.

For a second-order nonbilinear instrument, the response matrix $\mathbf{N}_k$ for a pure component can be represented as

$$\mathbf{N}_k = \sum_{l=1}^{L} \mathbf{p}_l s_l \mathbf{t}_l + \mathbf{E} = \mathbf{PST}^T + \mathbf{E} \qquad (10)$$

where L is the rank of $\mathbf{N}$, the $(\mathbf{p}_l,\mathbf{t}_l)$ are a set of vectors (possibly orthogonal) which span the column and row spaces, respectively, of $\mathbf{N}$, the s_l are scaling factors to normalize the $\mathbf{p}_l,\mathbf{t}_l$, and $\mathbf{E}$ is unmodelled experimental error. For real data, L is generally not known; a number of methods for estimating L exist,[28-30] but the actual estimate of L is not critical for this application.

Note that in the expansion of a pure component for bilinear data (eq 1), only one pair of vectors $\mathbf{x}_k$ and $\mathbf{y}_k$ exists, but in the nonbilinear expansion (eq 10), an infinite number of sets of $\mathbf{p}_l,\mathbf{t}_l$ exist, each differing by a rotation matrix. Furthermore, the bilinear expansion contains chemically meaningful information; recall for the LC–UV case the $\mathbf{x}_k$ and $\mathbf{y}_k$ vectors were the pure component UV spectrum and elution profile, respectively. For the nonbilinear case, there may be chemically meaningful decompositions, but the most readily obtainable decompositions, such as the singular value decomposition, do not contain this sort of unambiguous chemical information.[31,32]

To apply rank annihilation to nonbilinear data, consider utilizing a nonbilinear pure component response matrix $\mathbf{N}_p$ in eq 5. Because a nonbilinear $\mathbf{N}_p$ has rank larger than one, eq 5 becomes formally the same as the GRAM formulation in eq 7: $\mathbf{N}_p\mathbf{Z} = \mathbf{MZ\Lambda}$. The diagonal elements of Λ, Λ_{ii}, are again concentration ratio estimates, but there are two possibilities for the Λ_{ii} that result: either (a) all the Λ_{ii} will be the same (in the absence of noise), in which case the eigenproblem is ill-posed for the eigenvectors, or (b) the Λ_{ii} will be substantially different from each other, in which case the problem is to select the Λ_{ii} which is the true concentration estimate.[25] In practice, we have observed case a only for mathematical additions of artificial spectra; for real data the overlap between the row and columns spaces of pure component response matrices and the presence of noise inevitably cause case b to be the result, as discussed below. Consequently, applying rank annihilation to nonbilinear data generates a series of concentration estimates and the problem is to choose which of these estimates, if any, yields the true concentration.

To deal with the problem of multiple, possibly different Λ_{ii}, Lorber suggested two solutions:[25] (a) average all of the concentration estimates or (b) approximate a nonbilinear $\mathbf{N}_p$ by a bilinear matrix $\bar{\mathbf{N}}_p = \mathbf{u}_1\sigma_1\mathbf{v}_1$, where $\mathbf{u}_1$, σ_1, and $\mathbf{v}_1$ are the most significant terms in some decomposition of $\mathbf{N}_p$ in the form of eq 10. An example of (b) would be to use the diad formed by the first row and column principal components of $\mathbf{N}_p$ from the singular value decomposition. As will be demonstrated in the Results, neither of these methods is particularly satisfactory.

A third method is now proposed for dealing with multiple concentration estimates, applicable only when the response matrices contain only nonnegative elements, which is to use the smallest concentration estimate produced by solving the GRAM eigenproblem: $\mathbf{N}_p\mathbf{Z} = \mathbf{MZ\Lambda}$. The rationale for this method is as follows. Consider an expansion of $\mathbf{N}_p$ in the form of eq 10, $\mathbf{N}_p = \mathbf{PST}^T$. Let $\mathbf{U}$ and $\mathbf{V}$ be a normalized basis set for the remaining response matrices so that $\mathbf{N}_k = \mathbf{UU}^T\mathbf{N}_k = \mathbf{N}_k\mathbf{VV}^T \ \forall\ k = 1...K$, $k \neq p$. Suppose that for some $\mathbf{p}_i,\mathbf{t}_i$, where $i \in \{1, ..., L\}$, the following relationships hold: $\mathbf{p}_i \neq \mathbf{UU}^T\mathbf{p}_i$ and $\mathbf{t}_i \neq \mathbf{TT}^T\mathbf{t}_i$, meaning that $\mathbf{p}_i$ and $\mathbf{t}_i$ are not completely included in the space spanned by the remaining $\mathbf{N}_k$, and are therefore linearly independent of the columns of $\mathbf{U}$ and $\mathbf{V}$, respectively. Another way of expressing this requirement is that if $\mathbf{M}_p$ is the response of a mixture containing all of the components except the pth, then rank $\mathbf{M}$ must be larger than rank $\mathbf{M}_p$; i.e., adding component p to the mixture must increase the rank of the response matrix. If rank $\mathbf{M}$ is greater than rank $\mathbf{M}_p$, then $\mathbf{N}_p$ is said to increase the rank of $\mathbf{M}$. Then, if the diad formed by $\mathbf{p}_i s_i \mathbf{t}_i^T$ is rank annihilated against the mixture response matrix, by solving $\mathbf{Mz} = (\mathbf{p}_i s_i \mathbf{t}_i^T)\mathbf{z}\Lambda$, then the resultant concentration estimate will be correct, in the absence of noise. The proof of this statement is the proof for noniterative rank annihilation for bilinear data, as given by Lorber,[24] because the assumptions above are equivalent to linear independence for bilinear pure components.

Next, suppose instead that for all $\mathbf{p}_i,\mathbf{t}_i$, $i \in \{1, ..., L\}$, the above assumption does not hold; that is, $\mathbf{p}_i = \mathbf{UU}^T\mathbf{p}_i$ or $\mathbf{t}_i = \mathbf{VV}^T\mathbf{t}_i$, meaning either $\mathbf{p}_i$ is not linearly independent of the columns of $\mathbf{U}$ or $\mathbf{t}_i$ is not linearly independent of the columns of $\mathbf{V}$. In this case, it can be shown that the estimated concentrations will be biased by an amount which depends on the amount of correlation between $\mathbf{p}_i$ and $\mathbf{U}$, between $\mathbf{t}_i$ and $\mathbf{V}$, and on s_i—the relative contribution of $\mathbf{p}_i\mathbf{t}_i^T$ to $\mathbf{N}_p$. Futher, it can be shown that for absolute value mode 2D NMR and for MS/MS, for which the $\mathbf{N}_k$ can be diagonalized to a positive frame of reference,[33] then the concentration estimate will be larger than the true value.

It is important to note that the above conclusions indicate that rank annihilation based methods will only be applicable to nonbilinear systems when the rank of the response matrices is substantially less than the smaller of the number of rows and the number of columns. The reason for this is the requirement above that addition of $\mathbf{N}_p$ must increase the rank of $\mathbf{M}$. The maximum rank for a matrix is the minimum of the number of rows and columns, and if one or more components have a rank which is at or near the maximum, it will not be possible for additional components to increase the rank of $\mathbf{M}$, and the concentration estimates will necessarily be biased.

In the event that the response matrices are at or near full rank for a particular instrument or system, it is still possible to use these data for multicomponent quantitative analysis by treating the response matrix as a very long vector and utilizing a first-order data calibration method such as Partial Least Squares regression.[34] The nonbilinear rank annihilation requirement that $\mathbf{N}_p$ increase the rank of $\mathbf{M}$ is a stronger requirement than the linear independence which is required for first-order calibration.[21] Any second-order instrument can therefore always be treated as a first-order instrument, although any advantages for quantitation in the presence of unknown spectral interferents will necessarily be lost.

Experimental Section

All 2D NMR experiments were performed on a Bruker WM-500 (Brucker Instruments) spectrometer operating at 500.13 MHz. Stock sugar solutions were prepared by dissolving known masses of the pure sugars (Sigma Chemicals, St. Louis, MO) in known masses of D_2O (Cambridge Isotope Laboratories, Woburn, MA). Calibration samples were obtained by dilution by mass with D_2O; mixture samples were obtained by mixing, and in some cases diluting, stock solutions by mass.

Data analysis and simulations were performed on a MicroVAX II (Digital Equipment Corp., Maynard, MA), primarily using the Ctrl-C data analysis environment (Systems Control Technology, Palo Alto, CA) to implement the generalized rank annihilation algorithm, as described by Wilson and co-workers.[27]

Results and Discussion

In order to test the application of rank annihilation to nonbilinear data in cases where sources of variation could be con-

---

(28) Malinowski, E. R.; Howry, D. G. *Factor Analysis in Chemistry*; Wiley-Interscience: New York, 1980.

(29) Wold, S. *Technometrics* **1978**, *20*, 397–405.

(30) Eastment, H. T.; Krzanowski, W. J. *Technometrics* **1982**, *24*, 73–77.

(31) Mandel, J. *Am. Stat.* **1982**, *36*, 15–24.

(32) Klema, V. C.; Laub, A. J. *IEEE Trans. Autom. Control* **1980**, AC-25, 164–176.

(33) Wilson, B. E. Ph.D. Dissertation, University of Washington, Seattle, WA, 1988.

(34) Wold, S.; Geladi, P.; Esbensen, K.; Ohman, J. *J. Chemometrics* **1987**, *1*, 41–55.

Table I. Effect of Noise in Nonbilinear Rank Annihilation[a]

noise, %	mean	std dev	min	max
0.00	0.5000	0.0000	0.5000	0.5000
1.00	0.4997	0.0004	0.4985	0.5002
2.50	0.4994	0.0007	0.4974	0.5007
3.75	0.4994	0.0011	0.4962	0.5011
5.00	0.4988	0.0014	0.4945	0.5006
7.50	0.4988	0.0026	0.4893	0.5019
10.00	0.4984	0.0034	0.4873	0.5030
12.50	0.4980	0.0031	0.4901	0.5047
15.00	0.4976	0.0037	0.4862	0.5032
20.00	0.4984	0.0058	0.4756	0.5059

[a] Mean, standard deviation, minimum, and maximum of the concentration estimate from 30 repetitions at each of 10 levels of added noise for rank annihilation analysis of mathematical mixtures of constructed pure component spectra. The true concentration is 0.5000.

trolled, sets of response matrices ("spectra") were constructed for hypothetical pure components using sine and cosine functions, and these spectra were mathematically added to generate mixture spectra. By construction, these spectra were diagonalizable to a positive frame of reference, so that rank annihilation could be used, with the concentration estimate taken from the smallest of the eigenvalues (the method proposed in the theory section). For these simulation experiments reported in this paper, none of the analytes had a unique signal channel, so that these simulation experiments are examples of quantitation problems involving completely overlapped, unknown spectral interferents.

For the first set of experiments, each pure component response matrix was constructed from two diads, $\mathbf{x}_i s_i \mathbf{y}_i^T$, where each $\mathbf{x}_i$ contained 60 elements and each $\mathbf{y}_i$ contained 40, to generate a system with $m = 60$ rows and $n = 40$ columns. Initially, all four diads (two for each pure component) were constrained to be orthogonal, and rank annihilation was performed on each component separately, in a manner formally equivalent to generalized rank annihilation: $\mathbf{MZ}_1\mathbf{\Lambda}_1 = \mathbf{N}_1\mathbf{Z}_1$ and $\mathbf{MZ}_2\mathbf{\Lambda}_2 = \mathbf{N}_2\mathbf{Z}_2$, where $\mathbf{N}_1$ and $\mathbf{N}_2$ are the two pure component response matrices, $\mathbf{M}$ is a mixture response matrix generated from $\mathbf{N}_1$ and $\mathbf{N}_2$, and $\mathbf{Z}_1$, $\mathbf{Z}_2$, $\mathbf{\Lambda}_1$, and $\mathbf{\Lambda}_2$ are the corresponding eigenvector and eigenvalue matrices. The result of this analysis was the correct concentration estimate for the pure components, which is hardly surprising considering that all four diads were constrained to be orthogonal. The two diads in $\mathbf{N}_1$ behave as two independent bilinear pure components, each with the same concentration ratio between the mixture and the pure component.

In order to test the effect of noise on this simple case, normally distributed random numbers with zero mean and a standard deviation equal to a percentage of the largest absolute value in the matrix were added to the mixture spectra, to simulate white noise (where the standard deviation of the noise is equal for all channels, as opposed to shot noise where the standard deviation increases with the square root of the signal intensity). Rank annihilation was performed as above on the mixture spectra containing added noise using the original pure components, and the analysis was repeated 30 times for each of 10 levels of added noise. In each experiment, the estimated concentration was taken as the smallest of the concentration estimates obtained; the results are given in Table I. As the amount of added noise was increased, the standard deviation of the estimated concentration increased, and the estimated concentrations began to be noticeably biased low. This bias is predictable based on Jenson's inequality,[35] given that the estimate chosen is the smallest of the estimates obtained, and is a drawback of the proposed method of selecting a concentration estimate. One point to consider in Table I is that adding noise equal to 20% of the tallest peak in the mixture spectrum corresponds, in this system, to an average signal-to-noise ratio of approximately 2.5:1, where the average S/N is calculated as the average of the signal intensities, divided by the noise standard deviation. It is reasonable to consider the S/N calculated in this fashion, because these mathematically constructed spectra do not contain any regions of base-line signal. Even with this extreme level of noise, the advantages of multichannel signal averaging give a mean concentration uncertainty of 2.3% if the uncertainty region is taken as ±2 standard deviations, and a worst case prediction error of 7% over 30 repetitions.

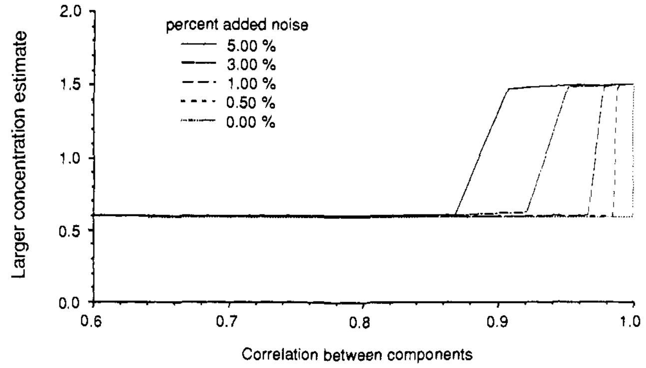


Figure 1. Effect of correlation on rank annihilation eigenvalues. The graph shows the effect of intercomponent collinearity on the largest eigenvalue from simulation experiments. The percentage of noise is the standard deviation of the added noise expressed as a percentage of the tallest peak in the mixture spectrum.

To test the effect of collinearity between diads within a given pure component the orthogonality constraint for the two diads forming a pure component was relaxed, while still holding the other two diads to be orthogonal to each other and to the first two diads. The result of these experiments was that, in the absence of noise, two correct concentration estimates were obtained even when the diads were almost completely collinear, with a correlation of 0.99. When the correlation reached 1.0, $\mathbf{N}_p$ became a rank one matrix, and only one significant eigenvalue was obtained, which yielded the true concentration ratio.

To test the effect of collinearity between components, the orthogonality constraint for two diads, one in each component, was relaxed. This experiment was repeated for a number of different levels of added noise; the results are summarized in Figure 1. For all values of correlation between the two diads and of added noise to the mixture spectrum, the smallest concentration estimate remained at the true value, to within the limits of error. As the correlation between the diads approached 1.0, the larger estimated concentration rose sharply and leveled off at an elevated value relative to the true value. For this simple experiment, the expectation of the larger eigenvalue, $E[c_+]$, can be shown to be equal to

$$E[c_+] = c_1 = \frac{s_4}{s_2}c_2 \tag{11}$$

where $\mathbf{M} = c_1\mathbf{N}_1 + c_2\mathbf{N}_2$, $\mathbf{N}_1 = \mathbf{x}_1 s_1 \mathbf{y}_1^T + \mathbf{x}_2 s_2 \mathbf{y}_2^T$, $\mathbf{N}_2 = \mathbf{x}_3 s_3 \mathbf{y}_3^T + \mathbf{x}_4 s_4 \mathbf{y}_4^T$, and $\mathbf{x}_2\mathbf{y}_2^T$ and $\mathbf{x}_4\mathbf{y}_4^T$ are the diads for which the orthogonality constraint was relaxed.

Of particular importance in the results in Figure 1 is the fact that the larger eigenvalue does not increase gradually away from the true value, but rather remains at the true value and then breaks sharply upward. The break becomes less sharp with added noise and occurs at lower amounts of correlation, but nonetheless remains sharp in comparison with the behavior which might be expected if the eigenvalue was linearly dependent on the amount of correlation. This result is important because for real data, the diads for a given pure component, in whatever basis set might be chosen, will not be completely orthogonal to the remaining components. If the eigenvalues grew gradually away from the true value as the diads became even partially correlated, it would not be possible to apply rank annihilation to any real nonbilinear case.

In the Theory section, it was demonstrated in eq 10 that it is possible, when using GRAM, to back estimate the **X** and **Y** matrices. Sanchez and co-workers[18,19] noted in applying GRAM to LC–UV data that the correlation between the estimated UV spectra in $\hat{\mathbf{X}}$ and library spectra of the pure components was

---

[35] Rao, C. R. *Linear Statistical Inference and its Applications*; Wiley: New York, 1973.

Table II. Concentration Estimates from GRAM-Based Analysis of 2D NMR Spectra of Sugar Solutions[a]

	analyte					
true value	arabinose	cellibiose	lactose	melibiose	sorbose	xylose
true value	23.121	11.397	39.103	17.800	3.671	23.069
concn estimates	20.643	9.767	40.534	17.189	3.347	24.264
	20.643	12.942	46.994	19.098	3.347	25.264
	24.160	26.253	117.439	24.644	4.349	30.351
	25.335	37.192	165.920	32.308	5.271	30.946
	35.638	41.486	168.034	44.828	9.100	39.571
mean spectral residual × 10^4	9.283	7.099	2.453	2.799	2.379	3.110
concn error, %	11.11	14.30	3.66	3.43	8.82	5.18

[a] For each of the six sugars, the true concnetration ratio $c_{p,\mathrm{M}}/c_{p,\mathrm{N}}$ is listed first, followed by the five smallest concentration estimates from rank annihilation of that sugar's pure component spectrum against a mixture containing all six sugars. All concentrations are in mg solution per mL D_2O. The concentration error is based on using the smallest concentration estimate.

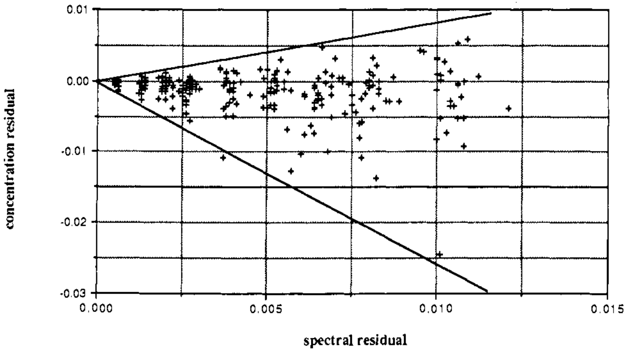


Figure 2. Correlation between spectral and concentration residuals. The spectral residuals are the average of the absolute values of the elements of the matrix $\hat{N}_p - N_p$, divided by the average absolute value of the elements of N_p for scale. The concentration residuals are the difference between the estimated concentration and the true concentration, divided by the true concentration for scale.

Table III. Concentration Estimates for Noniterative Rank Annihilation Using the First Principal Component Diad from the Calibration Spectrum[a]

	analyte					
	arabinose	cellibiose	lactose	melibiose	sorbose	xylose
true value	23.12	11.40	39.10	19.20	3.67	23.07
est value	32.65	15.20	45.84	21.75	6.26	45.66

[a] For each of the six sugars, the true concentration ratio is listed first, followed by the concentration estimate which results from using the first principal component diad from the singular value decomposition of the pure component spectrum and applying noniterative rank annihilation as per eq 5. All concentrations are in mg solute per mL D_2O.

indicative of the accuracy of the concentration. That is, inaccurate concentration estimates generally occurred at the same time as poor correlations of the estimated spectra in $\hat{X}$ with the library UV spectra of the pure components taken using the same LC-UV equipment and solvents. Furthermore, Sanchez[36] demonstrated a closed form connection between the errors in the estimated spectra and the errors in the predicted concentration for a $m = 2$ rows by $n = 2$ columns case. In the nonbilinear case, the **X** and **Y** matrices are replaced by **P** and T, which do not themselves contain unambiguous chemical information. However, estimates for **P**, T, and S_p can be obtained as

$$\hat{T} = (Z^+)^T \text{ and } \hat{P}\hat{S}_p = MZ\Lambda \qquad (12)$$

so that $\hat{N}_p = (Z^+)^T MZ\Lambda$. By comparing $\hat{N}_p$ with N_p an estimate of the maximum error in the concentration estimate $\hat{c}_p$ can be obtained. This correlation is demonstrated in Figure 2, which shows a scatter plot of the concentration residual versus the average absolute residual of $\hat{N} - N_p$ for a total of 300 experiments applying rank annihilation to mathematical mixtures of the constructed pure component spectra. As the average spectral residual increases, the deviations of the concentration estimates about the true value also increase, in an approximately linear fashion.

To test the applicability of the results thus far obtained on more complicated systems, the experiments relaxing orthogonality constraints were also performed on the following four systems: three components, each of rank two; two components, each of rank three; three components, each of rank three; and three components, each of rank five. As above, there were $m = 60$ rows and $n = 40$ columns. Various experiments were performed to study the effects of inter- and intracomponent collinearity, both separately and together. In all cases, the results were consistent with those achieved above, indicating that (a) intracomponent collinearity causes some of the eigenvalues to become undefined by reducing the dimensionality of the problem; (b) high intercomponent collinearity causes the relevant concentration estimates to be biased; (c) the sign of the intercomponent collinearity bias ($c_{\text{predicted}} - c_{\text{true}}$) is positive if the spectra are diagonalizable to a positive frame of reference; (d) random noise causes the distribution of concentration estimates (over a series of experiments) to be skewed negative relative to the true concentration (a consequence of choosing the smallest eigenvalue as the concentration estimate); and (e) so long as N_p increases the rank of **M**, the effects of (a)-(d) on the smallest eigenvalue are not synergistic.

Based on these results, rank annihilation as above was then applied to 2D proton J-correlated magnitude NMR spectra (COSY) of six sugars in D_2O: arabinose, cellibiose, lactose, melibiose, sorbose, and xylose, and a mixture containing all six sugars. These spectra were stored as 256 by 256 matrices with estimated ranks of 3 (cellibiose), 6 (sorbose), 7 (arabinose and lactose), 8 (melibiose), and 12 (xylose). The results of this analysis are summarized in Table II. Note that while the errors in the predicted concentration are larger than might be expected based on the simulated data, this analysis of real data is nonetheless an example of quantitation for an analyte of interest in the presence of five unknown spectral interferents. The calibration for each analyte was performed separately, using only that analyte's pure component spectrum as the N_p matrix. In addition, as shown in Table II, those analytes with larger concentration error tend also to have larger average spectral residuals.

The larger than expected errors in the concentration estimates are interpreted to be primarily a consequence of T_1 noise and the fact that this noise was eliminated by symmetrization.[37] For a given spectrum, **R**, the procedure employed for removing T_1 noise was to set the elements R_{ij} and R_{ji} both equal to the smaller of their original values. This procedure, while better than leaving the T_1 noise in the spectra, causes sufficient systematic differences between the mixture and calibration spectra to account for much of the observed prediction error. Furthermore, there were particular problems quantitating for the two analytes with the greatest prediction error, arabinose and cellibiose, as discussed below.

In applying rank annihilation to the NMR data, it was necessary to zero out the portion of the spectrum corresponding to the

---

(36) Sanchez, E. Ph.D. Dissertation, University of Washington, Seattle, WA, 1987.

(37) Mehlkopf, A. F.; Korbee, D.; Tiggelman, T. A. *J. Magn. Reson.* **1984**, *58*, 315-323.

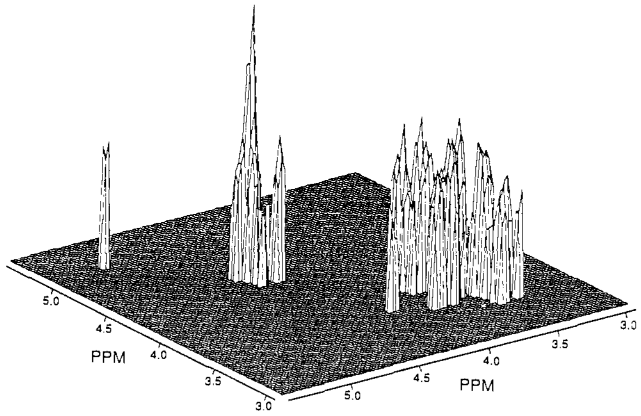


Figure 3. Arabinose calibration spectrum: absolute value mode plot of COSY spectrum of arabinose.

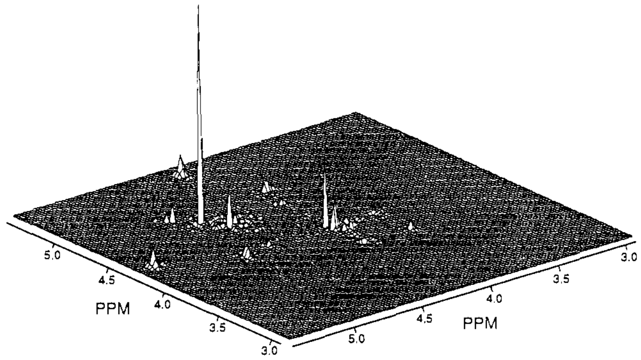


Figure 4. Residual matrix from arabinose calibration: calibration spectrum ($\mathbf{N}_p$) minus estimated spectrum ($\hat{\mathbf{N}}_p$) for arabinose. Note the large systematic deviation near $\delta = 4.7$ ppm.

presence of partially protonated water, HOD. This fact was noticed after applying rank annihilation to the data without any such pretreatment and comparing the estimated calibration matrix $\hat{\mathbf{N}}_p$ with the true calibration matrix for each of the six components. The quantitation for lactose, melibiose, sorbose, and xylose presented no unexpected difficulties, but the $\hat{\mathbf{N}}_p$ estimate for arabinose (Figure 4) and cellibiose exhibited systematic differences from the calibration spectra (Figure 3), much like a peak was missing from the estimated spectra at $\delta = 4.7$ ppm along the diagonal. We interpret this missing peak as being due to residual HOD in the samples; other 2D NMR spectra we have obtained indicate that the residual HOD peak in arabinose and cellibiose can occur at a chemical shift different from that in the remaining four sugars and in the mixture spectrum. The differences in chemical shift were on the order of 0.1 ppm, and so a region from $\delta = 4.2$ ppm to $\delta = 4.8$ ppm and 0.05 ppm wide along the diagonal was zeroed out for the analysis. The results in Table II were obtained after this region of the spectrum was suppressed. Interestingly, the errors in the estimated concentrations for the other four sugars (lactose, melibiose, sorbose, and xylose) decreased from an average of approximately 10% to an average of approximately 5%. We interpret this as signifying that the quantitation for these components was also affected by residual HOD, although not to the same degree as arabinose and cellibiose, possibly because of the large number of other peaks along the diagonal in the region of $\delta = 4.7$ ppm.

When the HOD peak moves, the principal components of the arabinose and cellibiose spectra which correspond to the HOD peak do not lie in the space spanned by the mixture spectrum, nor do they lie in the space spanned by $\mathbf{N}_p + \mathbf{M}$, which is used to form the basis for the projection.[27] Consequently, none of the variance structure associated with these principal components was retained when the mixture and calibration spectra were projected into P–T coordinate space, and so no such variance structure could be present in $\hat{\mathbf{N}}$, which explains the absence of the HOD peaks from these estimates. The estimated concentrations for arabinose and cellibiose were substantially smaller than the true values because the effective concentration of the arabinose and cellibiose residual HOD peaks in the mixture spectrum is zero; they simply are not present in the mixture. The same type of behavior would be expected if the analyte of interest was not present in the mixture sample, except that the estimated calibration spectrum would exhibit more serious deviations from the true spectrum.

In essence, the problem with the shift in the HOD peak is indicative of a potential drawback to the application of rank annihilation to nonbilinear data, which is that quantitation will be biased when matrix effects are present, an inherent difficulty when calibrating with pure component spectra.

As was mentioned in the Theory section, two other methods have been proposed, but not previously tested, for applying rank annihilation to nonbilinear data:[25] (a) averaging the resultant concentration estimates, and (b) approximating a nonbilinear $\mathbf{N}_p$ by a bilinear $\tilde{\mathbf{N}}_p$ constructed, for example, from the first row and column principal components of the singular value decomposition of $\mathbf{N}_p$. It is clear from the data presented in Table II that method a will not yield a satisfactory concentration estimate. The first problem that would have to be addressed is the number of concentration estimates to average. Table II only gives the five smallest estimates; for each of the sugars there are easily another ten concentration estimates that would have to be considered, all of which are larger than those given in Table II. As for method b, Table III gives the results when the singular value decomposition of $\mathbf{N}_p$ is calculated, and a diad formed from the first row and column principal components is used for rank annihilation on the same data as was used for Table II, as per eq 5. In this analysis, the same diagonal region in the spectra was zeroed out as in the Table II analysis, despite the fact that using only the first principal component gives no diagnostic information by which a problem such as the HOD peak shifts could be detected. With this advantage, method b gives some correlation with the true concentration values, although not as good as the data in Table II. Analysis using the first terms from a PLS-type decomposition[27] of $\mathbf{N}_p$ was also attempted, but the results were not significantly different from that obtained via the SVD. As a further test, the pure component spectra were mathematically added together to give an ideal case mixture spectrum, with concentration values equal to those in the real mixture. The GRAM-based analysis used in Table II gave the concentration estimates equal to the true values to within the limits of machine precision, but method b failed to give the exact answers for five of the sugars, instead yielding concentration estimates off by as much as 4%. The failure of method b to correctly solve such an ideal case is seen as conclusive evidence of its inappropriateness for applying rank annihilation to nonbilinear data, particularly since it can be shown that there are no circumstances under which method b would succeed and the GRAM-based analysis would fail to correctly estimate the concentrations, assuming that appropriate attention is paid to the $\hat{\mathbf{N}}_p$ estimate as a diagnostic tool.

Given two of the methods proposed for selecting the best concentration estimate from GRAM-based analysis of nonbilinear data, (a) approximating $\mathbf{N}_p$ by the first principal component diad and (b) selecting the smallest of the eigenvalues from using GRAM and the full $\mathbf{N}_p$, an additional method warrants consideration: (c) selecting the smallest concentration estimate from using successive principal component diads to approximate $\mathbf{N}_p$. While method c would generate the correct concentration estimates under the same circumstances as method b, it has the disadvantages of being computationally more intensive and of not generating the diagnostic estimate $\hat{\mathbf{N}}_p$.

In conclusion, it has been shown that it is possible to apply rank annihilation to nonbilinear data, generating an eigenproblem formally the same as the generalized rank annihilation method. The requirements for nonbilinear rank annihilation are that (a) the pure component second-order spectra must be available, (b) no serious matrix effects or similar nonlinearities in response can be present, (c) the spectra must be nonnegative, and (d) the responses for the analytes of interest must increase the rank of the mixture response matrix. Further, when using rank annihilation as per eq 7, one obtains a back estimate, $\hat{N}_p$, of the pure component spectrum, N_p. Systematic differences between $\hat{N}_p$ and N_p are indicative of matrix effects and other violations of condition b above, while random differences between the two matrices are indicative of the accuracy of the concentration estimate. Using rank annihilation and nonbilinear second-order data, it is then possible to quantitate for an analyte of interest even in the presence of unknown spectral interferents.

Acknowledgment. The authors wish to thank Tom Pratum for assistance acquiring the NMR spectra as well as Avraham Lorber and Eugenio Sanchez for their valuable inputs regarding rank annihilation. B.E.W. acknowledges the support of a National Science Foundation Graduate Fellowship and a Center for Process Analytical Chemistry Graduate Fellowship. This work was partially supported by the Office of Naval Research under Contract No. N00014-84-K-0270.

DEWAR-PI Study of Electrophilic Substitution in Selected Polycyclic Fluoranthene Hydrocarbons

Michael J. S. Dewar* and Roy D. Dennington, II

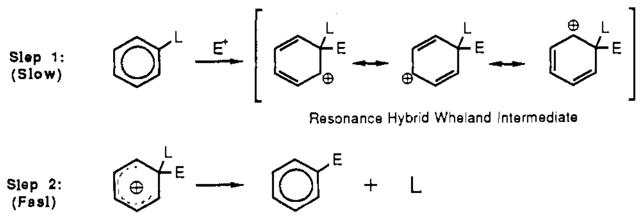
Contribution from the Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712. *Received August 23, 1988*

Abstract: DEWAR-PI molecular orbital calculations are reported for 14 nonalternant polycyclic aromatic hydrocarbon derivatives of fluoranthene and for the Wheland intermediates (arenium ions) to study electrophilic substitution at all the methine groups in them. Calculations were also carried out for related alternant aromatic hydrocarbons. The species studied were indeno[1,2,3-*hi*]chrysene (**1**), benz[*def*]indeno[1,2,3-*hi*]chrysene (**2**), fluoreno[3,2,1,9-*defg*]chrysene (**3**), benz[*a*]aceanthrylene (**4**), benz[*e*]acephenanthrylene (**5**), indeno[1,2,3-*cd*]pyrene (**6**), fluoreno[9,1,2,3-*cdef*]chrysene (**7**), benz[*def*]indeno[1,2,3-*qr*]chrysene (**8**), dibenz[*a,e*]aceanthrylene (**9**), dibenz[*a,j*]aceanthrylene (**10**), dibenz[*e,k*]acephenanthrylene (**11**), dibenz[*a,l*]aceanthrylene (**12**), and benzo[*k*]fluoranthene (**13**), benzo[*j*]fluoranthene (**14**), fluoranthene (**15**), phenanthrene (**16**), pyrene (**17**), chrysene (**18**), benzo[*def*]chrysene (**19**), benz[*a*]anthracene (**20**), anthracene (**21**), and naphthalene (**22**).

Interest in higher polyaromatic hydrocarbon derivatives of fluoranthene stems from their widespread prevalence as environmental pollutants. Some of these hydrocarbons can be mutagenic and/or carcinogenic, stimulating cancer research as well as synthetic work. Surprisingly little is known about the chemistry of these polycyclic fluoranthenes or their patterns of electrophilic substitution. Because they are nonalternant hydrocarbons, simple theoretical arguments describing substitution do not exist.

The majority of aromatic electrophilic substitutions occur by the *arenium ion mechanism* as depicted in Scheme I. The electrophile attacks in the first step, generating a positively charged intermediate (the so-called arenium ion or Wheland intermediate), and the leaving group departs in the second step. For the hydrocarbons studied, the leaving group is always hydrogen, and the results discussed in this paper are valid only for electrophiles that follow this type of mechanism.

Scheme I



The rate-determining step is almost always the formation of the Wheland intermediate (WI). Species of this kind are well-known, e.g., the species formed by protonation of aromatic hydrocarbons by strong acids. It can be shown by PMO theory[1] that the activation energy (E_a) for substitution at a given site in an even alternant hydrocarbon (AH) is proportional to the heat of reaction (ΔH_{Rx}) for the formation of the corresponding WI:

$$E_a = A\Delta H_{Rx} \tag{1}$$

where A is a constant. Thus the difference (ΔE_{XY}) in activation energies for substitution at different positions (X and Y) is proportional to the difference in energies of their respective WIs (E_X and E_Y):

$$\Delta E_{XY} = E_a(X) - E_a(Y) = B(E_X - E_Y) \tag{2}$$

where B is another constant. Thus the quantities of ΔE_{XY} should serve as a measure of the relative reactivities of various positions to electrophilic substitution. In order to predict the favored point of attack in a given aromatic compound, the energies of the WIs have to be evaluated for all possible positions, 14 in the case of indeno[1,2,3-*hi*]chrysene (**1**).

Calculations of this kind for the compounds of interest here by even the simplest ab initio procedures would require an unreasonable amount of computer time. Indeed, the use of even all-valence-electron semiempirical methods would be very expensive. These difficulties can be avoided by using a π-SCF treatment[2] that was developed here some years ago, based on the Pople[3] method together with the Pariser–Parr[4] procedure for allowing for electron correlation. This Pariser–Parr–Pople (PPP) approach, as it is commonly termed, was originally developed for studies of light absorption. Our group was able to show that it could be parametrized to reproduce ground-state properties of conjugated molecules very effectively,[2,5] heats of formation of conjugated and aromatic hydrocarbons being reproduced to within the limits of error of the best thermochemical measurements.

---

(1) (a) Dewar, M. J. S. *J. Am. Chem. Soc.* **1952**, *74*, 3341, 3345, 3350, 3353, 3355, 3357. (b) Dewar, M. J. S.; Dougherty, R. C. *The PMO Theory of Organic Chemistry*; Plenum: New York, 1975.

(2) Dewar, M. J. S.; Chung, A. *J. Chem. Phys.* **1965**, *42*, 756.

(3) (a) Pople, J. A. *Trans. Faraday Soc.* **1953**, *49*, 1375. (b) Pople, J. A.; Brickstock, A. *Trans. Faraday Soc.* **1954**, *50*, 901.

(4) Pariser, R.; Parr, R. G. *J. Chem. Phys.* **1953**, *21*, 466, 767.

(5) (a) Dewar, M. J. S.; Gleicher, G. *J. Am. Chem. Soc.* **1965**, *87*, 685, 692. (b) Dewar, M. J. S.; de Llano, C. *Ibid.* **1969**, *91*, 789. (c) Dewar, M. J. S.; Marita, T. *Ibid.* **1969**, *91*, 796. (d) Dewar, M. J. S.; Thompson, C. C. *Ibid.* **1965**, *87*, 4414. (e) Dewar, M. J. S.; Gleicher, G. *Tetrahedron* **1965**, *21*, 1817, 3423. (f) Dewar, M. J. S.; Gleicher, G.; Thompson, C. C. *J. Am. Chem. Soc.* **1966**, *88*, 1349. (g) Dewar, M. J. S.; Hashmall, J.; Venier, C. G. *Ibid.* **1968**, *90*, 1953. (h) Dewar, M. J. S.; Kelemen, J. *J. Chem. Phys.* **1968**, *49*, 499.

0002-7863/89/1511-3804$01.50/0