Direct optimization of self-modeling curve resolution: application to the kinetics of the permanganate - oxalic acid reaction

Peter D. Wentzell, Ji-Hong Wang, Leon F. Loucks, and Kelli M. Miller

Abstract: A new procedure to extract spectral and concentration information for unknown components in a series of mixture spectra is described. The algorithm, a form of self-modeling curve resolution (SMCR), requires no assumptions about the mixture data except for linear additivity and non-negativity of component spectra, and non-negativity of concentrations. It provides permissible bands for the spectra and concentration profiles of individual components in the mixtures and is applicable to problems in chromatography, equilibrium studies, and kinetics, among others. The new method employs a nonlinear optimization procedure and is substantially more efficient than similar methods that have been previously reported. The SMCR algorithm is demonstrated by applying it to the autocatalytic oxidation of oxalic acid by permanganate in the presence of sulfuric acid. The reaction was monitored spectroscopically between 200 and 1000 nm under four different reaction conditions. Results verified the presence of two reaction intermediates previously proposed in the literature and confirmed their spectral properties. Concentration profiles for the intermediates are described.

Key words: curve resolution, kinetics, chemometrics, permanganate, oxalic acid.

Introduction

The use of multichannel detectors to study chemical phenomena has become commonplace in recent years. Such detectors are known as first-order detectors in the vernacular of chemometrics and the most common examples are multiwavelength absorption or emission spectrometers and mass spectrometers. When coupled with an experiment that presents a time series of mixtures varying in composition, such as in chromatographic or kinetics studies, a first-order detector produces a sequence of vectors that can be combined to form a matrix, or second-order data set. Even if the detector is not completely selective for each chemical species of interest, such a data matrix can significantly enhance the quality of information available to the investigator over that obtained with a single-channel detector.

The most common model assumed for a data matrix of this type is second-order bilinear, represented mathematically as

\[ D = CS^T + E \]

Here \( D \) is the \( mxn \) matrix of responses measured at \( n \) detector channels for the \( m \) samples in the time series (i.e., each row of \( D \) is a vector of responses for a single mixture). The \( mxp \) matrix \( C \) is the matrix of concentrations of each of \( p \) observable components in the \( m \) mixtures, and the \( nxp \) matrix \( S \) consists of the sensitivities of each of the components

at each of the $n$ detector channels (i.e., the columns of $S$ represent the “spectra” of individual components at unit concentration). The superscript “T” indicates the transpose of the matrix. Finally, $E$ is the $mn$ matrix of residuals (errors) for the measurements. (A complete list of symbols used is given in Table 1.)

Given a multivariate data set of this type, it is generally the objective of the chemist to extract the matrices $C$ and $S$ from the matrix $D$, thereby obtaining the concentration profiles and spectra of the species of interest. Such information can be extremely valuable, for example, in the identification of unknown species or the elucidation of a reaction mechanism. If either $C$ or $S$ is completely known, computation of the companion matrix is straightforward, but more often this information is incomplete. This is the case, for example, in a kinetics experiment where one or more unidentified intermediates may be present, or in chromatography where component peaks may be unresolved, or in equilibrium studies where several unknown species may coexist. In some of these cases, it may be possible to extract spectra and concentration profiles for the individual components if there are regions of the data (spectral channels or samples) that are uniquely associated with particular components, but in many instances this will not be true. In these cases, it is usually impossible to obtain unique solutions for the profiles of interest. However, by imposing some simple general constraints, it is often possible to define upper and lower boundaries for the component spectra (and concentration profiles) in the mixtures, thereby providing some information about what is happening in the mixtures. This procedure is known as self-modeling curve resolution (SMCR).

The principles of SMCR were introduced in 1971 by Lawton and Sylvestre (1), who applied the method to two-component mixtures. There are only three basic constraints applied in the original method: (i) pure component spectra cannot be negative, (ii) component concentrations cannot be negative, and (iii) pure component spectra (and concentration profiles) must lie in the $p$-dimensional subspace defined by the mixture spectra, where $p$ is the number of mixture components. Since the original work, many applications and extensions of the method have appeared, with a substantial amount of effort directed to devising algorithms applicable to more than two components. Unfortunately, SMCR becomes more difficult as the number of components becomes large, and these methods have had mixed success.

This paper describes a new approach to SMCR for mixtures of more than two components. The new method is

### Table 1. Glossary of symbols used.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>True or estimated $mxp$ matrix of component concentration profiles</td>
</tr>
<tr>
<td>$\hat{C}$</td>
<td>Estimate of $C$ with negative values set to zero</td>
</tr>
<tr>
<td>$D$</td>
<td>$mxn$ data matrix of mixture spectra as a function of time</td>
</tr>
<tr>
<td>$\hat{D}$</td>
<td>Truncated PCA estimate of $D$</td>
</tr>
<tr>
<td>$\tilde{D}$</td>
<td>Estimate of $D$ from $\hat{C}$ and $\tilde{S}$</td>
</tr>
<tr>
<td>$E$</td>
<td>$mn$ residual error matrix for $D$</td>
</tr>
<tr>
<td>$H$</td>
<td>Inverse Heaviside (step) function to penalize negative values</td>
</tr>
<tr>
<td>$m$</td>
<td>Number of mixture spectra in the time series</td>
</tr>
<tr>
<td>$n$</td>
<td>Number of wavelength channels for the spectra</td>
</tr>
<tr>
<td>$P$</td>
<td>Penalty function used in Sasaki algorithm</td>
</tr>
<tr>
<td>$P', P'', P'''$</td>
<td>Objective/penalty functions for new algorithm</td>
</tr>
<tr>
<td>$p$</td>
<td>Number of observable components in mixtures; pseudorank of $D$</td>
</tr>
<tr>
<td>$q$</td>
<td>Number of factors retained in PCA/SVD truncation</td>
</tr>
<tr>
<td>$R, \hat{R}$</td>
<td>Scores matrix from PCA (“^” indicates truncation)</td>
</tr>
<tr>
<td>$S$</td>
<td>True or estimated $nxp$ matrix of pure component spectra</td>
</tr>
<tr>
<td>$\hat{S}$</td>
<td>Estimated matrix $S$ with negative values set to zero</td>
</tr>
<tr>
<td>$s_i$</td>
<td>Estimated or true $nx1$ spectrum for component $i$</td>
</tr>
<tr>
<td>$s_i, s_i', s_i''$</td>
<td>Initial estimate and boundary estimates of spectrum for component $i$</td>
</tr>
<tr>
<td>$s_{jk}$</td>
<td>An element of the spectrum for component $k$ at channel $j$</td>
</tr>
<tr>
<td>$T$</td>
<td>$pxp$ transformation matrix for converting eigenvectors ($V$) to pure component spectra</td>
</tr>
<tr>
<td>$T^*$</td>
<td>Initial estimate of $T$</td>
</tr>
<tr>
<td>$t_i$</td>
<td>$px1$ transformation vector to estimate spectrum of component $i$</td>
</tr>
<tr>
<td>$\Delta t$</td>
<td>Step size for simplex optimization</td>
</tr>
<tr>
<td>$U, \hat{U}$</td>
<td>Left singular matrix from SVD (“^” indicates truncation)</td>
</tr>
<tr>
<td>$U_{jk}^{'}, U_{jk}^{''}$</td>
<td>Objective functions used in Sasaki algorithm</td>
</tr>
<tr>
<td>$V, \hat{V}$</td>
<td>Right singular matrix (eigenvectors) from SVD (“^” indicates truncation)</td>
</tr>
<tr>
<td>$w$</td>
<td>Weighting factor used in penalty function</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Scale factor for penalty functions</td>
</tr>
<tr>
<td>$\Lambda, \hat{\Lambda}$</td>
<td>Diagonal matrix of eigenvalues from PCA (“^” indicates truncation)</td>
</tr>
<tr>
<td>$\lambda_i$</td>
<td>Diagonal element $i$ of $\Lambda$ (eigenvalue $i$)</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>Root-mean-square difference between $D$ and $\hat{D}$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>Root-mean-square difference between and $\tilde{D}$</td>
</tr>
</tbody>
</table>
illustrated through its application to an interesting autocatalytic reaction, the oxidation of oxalic acid by permanganate. Like virtually all SMCR methods, this new approach is based on principal component analysis, which will be introduced first.

**Principal component analysis**

The first step in applying SMCR to a second-order bilinear data set is usually the determination of the number of observable components present (note that the term “observable” is used here to indicate that a chemical species must have a measurable response and have concentration and spectral profiles that are sufficiently different from other species in the mixture). This determination is typically carried out using principal component analysis (PCA) (2), normally implemented through singular value decomposition (SVD):

\[ D = U \Lambda^{1/2} V^T = RV^T \]

If D is \( m \times n \) and \( m \leq n \), then U and R are \( m \times m \), \( \Lambda \) is \( m \times m \) diagonal matrix of eigenvalues (\( \Lambda^{1/2} \) is called the matrix of singular values), and \( V \) is \( n \times m \). The matrix \( R (= U \Lambda^{1/2}) \) is sometimes called the scores matrix and columns of \( V \) are called the eigenvectors or loadings. The columns of U and V are orthonormal. If the measurements in D were error-free, then the rank of D would be equal to \( p \), the number of independently observable components in the linear system, and the last \((m - p)\) eigenvalues would be zero. In practice, the presence of experimental errors means that the numerical rank of D is normally full \((\text{min}(m,n))\), so an estimate of the “pseudorank” (also called the chemical rank) of the data matrix is needed in order to determine the number of components. To do this, the SVD results are truncated to the first \( q \) principal components and then used to reconstruct the data matrix:

\[ \hat{D} = \hat{U} \hat{\Lambda}^{1/2} \hat{V}^T = \hat{R} \hat{V}^T \]

In this equation, \( \hat{U} \) and \( \hat{R} \) are \( m \times q \), \( \hat{\Lambda} \) is \( q \times q \), and \( \hat{V} \) is \( n \times q \). Although there are a variety of methods to estimate the pseudorank of D, most are typically based on increasing \( q \) from unity until D reorders matrix and within experimental error (2). At this point, \( q \) is taken to be equal to \( p \), the pseudorank of the data matrix.

**Two-component mixtures**

Although PCA is useful for estimating the number of components in a series of mixtures, it does not generally provide the concentration or spectral profiles of those components. In the context of eq. [1], the \( p \) columns of \( V \) are sometimes called “abstract spectra” and will define the \( p \)-dimensional subspace of the \( n \)-dimensional spectral space containing the pure component spectra. However, the abstract spectra are orthonormal, whereas the real spectra usually are not, and therefore the \( p \)-dimensional subspace can theoretically be defined by an infinite number of combinations of \( p \) vectors. The same arguments hold for the relationship between \( \hat{U} \) (or \( \hat{R} \)) and the concentration profiles of each component.

A solution to this dilemma was provided by Lawton and Sylvestre (1) who proposed SMCR for two-component mixtures. In addition to the restrictions imposed by the SVD, these authors reasoned that most linearly additive spectroscopic methods were bound by two additional constraints: non-negative concentration and spectral profiles for the pure components. Although these constraints do not generally specify unique solutions, they do define boundaries for spectra and concentration profiles that are useful in certain circumstances. By normalizing pure component spectra to unit area, Lawton and Sylvestre were able to describe a simple mathematical procedure to identify the high and low limits of pure component spectra.

One unavoidable problem with this approach, and others like it, is that the absolute scales of the spectra and concentration profiles have no meaning. This “ambiguity of scale” is a consequence of the fact that a spectral response is a product of concentration and sensitivity, and a change in scale for one can be offset by a reciprocal change in scale for the other. Nevertheless, the shape of a spectrum or concentration profile for a given component is consistent for that component and these can be very useful for identifying unknown mixture components or modeling their behaviour. Some applications of this approach have been reviewed by Hamilton and Gemperline (3).

**More than two components**

Although the SMCR method of Lawton and Sylvestre has been widely used, it does not easily extend itself to more than two components. The principles of the method are quite general and it is essentially a problem in linear programming, but the practical difficulties of encoding the constraints become difficult to manage as the number of components increases (4, 5). For this reason, several alternative strategies to the problem have been adopted. These have been reviewed by Hamilton and Gemperline (3) and will be treated only briefly here.

One of the problems in extending SMCR to more than two components is that the technique must define a region of feasible solutions for each component, and this is not as straightforward as in the two-component case, since a higher dimensional space means more boundaries to define. One approach to the problem has been to attempt to define unique solutions by adding additional constraints. These constraints may be based on physical requirements, such as the existence of pure component responses (6, 7) or the presence of the component in a spectral library (8), or may be more general, such as requirements for maximum dissimilarity among spectra (9) or maximum spectral simplicity (10).

One of the most popular methods of generating “unique” solutions by curve resolution is iterative target-transformation factor analysis (ITTFA) (11, 12), which is also known as alternating least squares (ALS). With this method, the spectral or concentration profiles of one of the components is first estimated using some independent method. This estimation can be as simple as choosing a delta function to represent a chromatographic peak (11), or may involve a technique such as varimax rotation (12) or evolving factor analysis (13). The estimate is then projected into the SVD subspace and iteratively refined with its companion constraints.
Fig. 1. Visual representation of optimization problem for a three-component system and the step size (Δt) used in simplex optimization.

**Profile by setting negative values to zero until the process converges on positive solutions. The principle advantages of ITTFA are its speed and simplicity. Among its drawbacks, however, is a sensitivity to the initial estimate of the concentration profile or spectrum and a susceptibility to poorly defined convergence (14). Furthermore, the ITTFA procedure produces a single solution, ignoring the fact that there are usually a range of solutions.**

**True SMCR methods will generate a range of feasible solutions for component spectra and concentration profiles where no unique solutions exist. A direct extension of the original two-component approach can be used (4, 5), but solutions can become complex, especially if there are no additional constraints or the number of components is greater than three. Ohta (15) developed a Monte Carlo method to search for feasible solutions for the spectra that should be generally applicable, but is also likely to be slow. Nonlinear optimization has also been used to address the problem (16, 17) and this is the strategy used in this work. Since the method presented in this paper is derived from the work of Sasaki et al. (16), a brief description of this approach will be given.**

The first step in this procedure developed by Sasaki et al. is decomposition of the data matrix D by SVD followed by truncation of the solution to rank p. Since the pure component spectra must lie in the subspace defined by the eigenvectors, each of these spectra can be represented as a linear combination of eigenvectors, that is:

**[4] S = VT**

or,

**[5] s_1 = V t_1**

Here V is the nxp matrix of eigenvectors (although this is the truncated matrix, the “^” has been removed for notational convenience), S is the nxp matrix of pure component spectral vectors, and s_i is the nx1 spectral vector for component i (S = [s_1 s_2 ... s_n]). The pxp matrix T (or alternatively the px1 vectors t) describes the linear combinations of the eigenvectors in V needed to give the pure component spectra. Of course, the elements of T are unknown and are the parameters sought by the optimization. As was the case with two-component methods, the pure component spectra are normalized to unit area, i.e.,

**[6] \( \sum_{i=1}^{n} t_{ij} = 1 \)**

This places a consequential restriction on T such that, for the kth component,

**[7] \( t_{pk} = \frac{1 - \sum_{j=1}^{p-1} t_{kj} \sum_{j=1}^{n} v_{ji}}{\sum_{j=1}^{n} v_{jp}} \)**

In other words, the last row of T is defined by the first \( p-1 \) rows and only \( p \times (p-1) \) parameters need to be optimized.

With the Sasaki method (16), the optimization proceeds as follows for each component and each spectral channel. First, a set of \( t_{ij} \)'s is selected to define T and, therefore, the pure component spectra in accordance with eq. [4]. The matrix S is then used to estimate the concentration profile according to

**[8] \( C = DS(S^T S)^{-1} = DV(T^{-1})^T \)**

Note that this calculation can only be performed if T is non-singular, ensuring that the spectral profiles are not linear combinations of one another.

Once the spectral and concentration profiles have been calculated, the Sasaki algorithm minimizes the following functions with respect to the \( t_{ij} \)'s:

**[9] \( U_{jk}^{(+)} = +s_{jk} + P \)**

**[10] \( U_{jk}^{(-)} = -s_{jk} + P \)**

where the penalty function, P, is given by:

**[11] \( P(T) = \gamma \left[ \sum_{j=1}^{n} \sum_{k=1}^{p} H(s_{jk})s_{jk}^2 + \sum_{i=1}^{m} \sum_{k=1}^{p} H(c_{ik})c_{ik}^2 \right] \)**

and,

**[12] \( H(y) = \begin{cases} 0 & (y \geq 0) \\ 1 & (y < 0) \end{cases} \)**

The functions \( U_{jk}^{(+)} \) and \( U_{jk}^{(-)} \), which are optimized separately for each component and spectral channel, represent the optimization functions for the upper and lower boundaries. The penalty function, P, ensures that no negative spectral or concentration values are tolerated if the scaling factor, \( \gamma \), is set to a very large value. This function will be zero for non-negative solutions, allowing eqs. [9] and [10] to seek the highest and lowest allowed values for the spectral profiles.

It is useful to provide a visual interpretation of the Sasaki algorithm for a three-component system. In this case, the acceptable solutions can be regarded as being confined (by the unit area constraint) to three regions of the \( t_1-t_2 \) plane as shown in Fig. 1. (The diagram is not intended to represent an actual mixture problem, but is included for pedagogical

© 1998 NRC Canada
purposes.) The shaded regions indicate those cases which give \( P = 0 \) (i.e., they satisfy the non-negativity constraints) and therefore represent candidate spectra for the three components. The stars represent spectra obtained for the mixtures and in this case are shown as a smoothly varying sequence such as that which might be observed in a kinetics experiment or a chromatography run. Note that a selected combination of pure component spectra, such as that indicated by the points connected by the triangle, must encompass all of the mixture spectra. Because of this, it is not possible to simply choose a spectrum from each region arbitrarily, and the three spectra are mathematically interdependent, as suggested by the triangle. For each spectral channel, the Sasaki method seeks to find the points in each component region that give the highest and lowest response values. For more components, the process is essentially the same, except that the feasible regions become volumes and hyper-volumes instead of surfaces.

The Sasaki method is a useful, general approach to solving the problem of SMCR with more than two components, but it also has some drawbacks as well. Some of these are listed below.

1. The discontinuous nature of the penalty function can make the optimization difficult and inhibits the use of gradient techniques.
2. The optimization is done twice for each spectral channel and each component, requiring \( 2np \) repetitions of the process. To further compound this problem, the original work (16) suggests an iterative procedure in which the value of the scaling factor, \( \gamma \), is increased by a factor of 10 at the end of each run. Thus, the proposed procedure can be very computationally demanding, and becomes more so as the number of components is increased.
3. The search strategy needs to be able to locate local optima. The optimization is carried out one component at a time, with the intention of locating a pair of points within each region corresponding to the spectral boundaries for that component. However, it is possible for the search to “jump” to a different region and locate the optima for a different component. This problem can lead to redundant solutions and becomes more serious as the regions in Fig. 1 move closer together and, in some circumstances, overlap. Therefore, any optimization procedure used must be able to locate local optima associated with each component.
4. In some instances, there is no region in which the penalty function, \( P \), is equal to zero. With experimental data, measurement errors can lead to negative values for some responses, which in turn may create a situation that forces either the concentration or spectral profiles to have negative values. If this is the case, the penalty function dominates the optimization and generates a single solution (i.e., no upper and lower bands) for each component. Furthermore, the nature of the solution depends on the scale of the measurements. This is because the spectra in eq. [11] are scaled to unit area, whereas the magnitude of the concentration values will depend on the magnitude of the measurements. The relative magnitudes of the concentration and spectral values will determine which is more likely to be forced to negative values during the course of the optimization, and therefore the shape of the extracted profile.

5. Because each spectral channel is separately optimized, the final bands produced to define the spectral boundaries do not necessarily represent contiguous spectra. In other words, the maximum value found for component 1 at wavelength 1 is not necessarily derived from the same spectrum as the maximum found at wavelength 2. Because of this, the resulting bands may show the limits, but do not themselves represent a spectrum. This may cause problems. For example, the resulting spectra will not necessarily have unit area.

In this work, the algorithm originally proposed by Sasaki et al. has been refined in an attempt to overcome some of these weaknesses.

A new approach

The procedure developed in this paper for treating mixtures of more than two components by SMCR is substantially faster than the Sasaki algorithm discussed in the previous section and, furthermore, helps minimize the influence of negative observations. The first step in the new algorithm is the same as for the Sasaki algorithm: decomposition of the original data matrix by SVD and estimation of the number of components present. Following this, the residual variance is calculated from

\[
\sigma_i^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (d_{ij} - \hat{d}_{ij})^2}{(m - p)(n - p)}
\]

In this equation, \( d_{ij} \) and \( \hat{d}_{ij} \) are elements of the original and reconstructed data matrices, \( \lambda \) represents an eigenvalue, \( p \) is the number of components, and \( m \) and \( n \) are the number of rows and columns in the data matrix. In this work, the residual variance is used as an estimate of the measurement variance.

The second stage of the modified algorithm involves the optimization of a function similar to eq. [11] that was used in the Sasaki algorithm. In this work, the objective function has been modified to

\[
P'(T) = \sum_{j=1}^{n} \sum_{k=1}^{p} H(s_{jk})s_{jk}^2 + \frac{1}{W^2} \sum_{i=1}^{m} \sum_{k=1}^{p} H(c_{ik})c_{ik}^2
\]

where the weighting factor, \( W \), is given by

\[
W = \frac{\max_{i=1,...,m} (\Psi_i^2)^{1/2} \left( \sum_{i=1}^{m} v_i^2 \right)^2}{\max_{j=1,...,n} (\Psi_j^2)}
\]

The purpose of the weighting factor in eq. [14] is to compensate for scale differences between the spectra and concentration profiles. Because the spectra are normalized to unit area, elements of the concentration vectors can be much larger or much smaller than the elements of the spectral vectors, depending on the magnitude of the observations. Although this will not be a problem when it is possible for all of the elements of \( S \) and \( C \) to be positive, some data sets will require negative elements and the disproportionate weighting of the two terms can cause difficulties (i.e., the...
contribution of negative values is not equally balanced between the two parts). The weighting factor is calculated assuming that the first eigenvector is a reasonably good approximation for the “average” spectrum (\( = v_i / \Sigma v_i \)) and the first scores vector estimates the “average” concentration profile (\( = u_i / \Sigma v_i \)). An approximate normalization factor is obtained from the ratio of the maxima of these two profiles.

The optimization of eq. [14] with respect to the elements of \( T \) is carried out using a simplex procedure because of the discontinuities in the objective function. In cases where there are feasible non-negative solutions for \( S \) and \( C \), there is no unique solution for \( T \), but this is not important. The purpose of this step is to locate the general region of potential solutions so that local optimization procedures can be used. The optimization of \( P \) results in a solution that will be designated as \( T^o \), and this serves as an initial guess for the local optimization procedures.

The third stage of the algorithm uses \( T^o \) as a starting point and attempts to locate one of the boundary solutions for each of the components in turn. Again, simplex optimization is used, but care must be taken to ensure that the search for one component does not stray too far into the local minimum corresponding to another component (see Fig. 1) or else redundant solutions will result. To reduce the chances of this happening, \( T^o \) is used to gauge the distance between the component regions in “t”-space and set the step size for the simplex optimization. This could be done in a variety of ways (for example, taking the average distance between the components), but we have found it useful to set the step size to the largest value of the smallest differences among coordinate pairs, i.e.,

\[
[16] \quad \Delta t = \max_{i=1,...,p} \left\{ \min_{j,k=1,...,p;j\neq k} |t_{jk} - t_{ik}| \right\}
\]

Because this equation is somewhat confusing, a visual example is presented in Fig. 1, although only the two dimensions \( t_1 \) and \( t_2 \) are considered, since the \( t_3 \) axis is not shown. Imagining the points shown to represent \( T^o \), it is clear that the smallest separation on the \( t_1 \) axis is between components 1 and 2, i.e., \( (t_{12} - t_{11}) \), and the minimum separation on the \( t_2 \) axis is between components 2 and 3, i.e., \( (t_{23} - t_{22}) \). Since the former of these two minimum distances is larger, it is chosen as the step size for the simplex optimization, \( \Delta t \), as shown in Fig. 1 (again, this ignores the \( t_3 \) axis simply for purposes of illustration). The rationale for this definition is that this is the largest step size that is still likely to confine the search to a particular region. The value of this step size is not critical, however, as long as it sufficiently restricts the search space.

For each of the components present, the initial guess for the simplex optimization is provided by \( T^o \) and this is adjusted to find an optimum \( T \) by the following procedure. First, the spectra are calculated from the trial solution using eq. [4] (subject to the constraints in eqs. [6] and [7]). Any negative values in the calculated spectra are then set to zero:

\[
[17] \quad \tilde{s}_{ij} = \max(s_{ij}, 0)
\]

Here \( s_{ij} \) is an element of the calculated matrix of spectra and \( \tilde{s}_{ij} \) is an element of the adjusted spectra. The adjusted spectra are normalized to unit area as before and then used to calculate the concentration profiles according to:

\[
[18] \quad C = D\tilde{S}(\tilde{S}^T\tilde{S})^{-1}
\]

As with the spectra, negative concentration values are set to zero, giving the new matrix, \( \tilde{C} \). The result is then used to calculate a new estimate for the data matrix:

\[
[19] \quad \tilde{D} = \tilde{C}\tilde{S}^T
\]

The purpose of this procedure is to ensure that negative values are removed from the concentration and spectral profiles. Obviously, if negative values are present, the space spanned by \( \tilde{C} \) and \( \tilde{S} \) will not be the same as the original described by \( U \) and \( V \), which could result in a poor reconstruction by eq. [19]. However, if the negative values are small, resulting only from measurement errors, the perturbations introduced should also be small and may result in an acceptable solution. Following the reconstruction, a new residual variance is calculated from the difference between the spectra reconstructed from eq. [19] (\( \tilde{D} \)) and the PCA reproduced spectra (\( \tilde{D} \)):

\[
[20] \quad \sigma_2^2 = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (\tilde{d}_{ij} - \tilde{d}_{ij})^2}{mn}
\]

The quantity \( \sigma_2 \) is analogous to what Malinowski refers to as “imbedded” error, giving the difference between the proposed “pure” data and the PCA reconstructed data. If there were no negative values in the calculated spectra or concentration profiles, \( \tilde{D} \) should equal \( D \) and therefore \( \sigma_2 \) will be zero. If adjustments for negative values were made, \( \tilde{D} \) and \( D \) will not be equal and the question is whether \( \tilde{D} \) is an acceptable estimate of \( D \). To answer this question, we simply examine the ratio:

\[
[21] \quad \alpha = \sigma_2^2 / \sigma_1^2
\]

In Malinowski’s treatment of errors in PCA (2), the imbedded error should always be less than the real error (estimated by \( \sigma_1 \)), so the solution is considered feasible if \( \alpha \) is less than unity. Because the imbedded error for the correct solution is usually significantly smaller than the real error, this criterion is likely to provide slightly wider boundaries than should be the case, but in practice the boundaries predicted are not particularly sensitive to the critical value of \( \alpha \).

At this stage of the procedure, the actual objective function minimized is

\[
[22] \quad P^*(T) = -\|s_i - s_i^0\| + \gamma H(\sigma_1^2 - \sigma_2^2)
\]

where \( s_i \) is the trial spectrum for component \( i \), \( s_i^0 \) is the spectrum for component \( i \) computed from \( T^o \), and \( \gamma \) is a large constant. The second term of this expression ensures that only solutions that produce reasonable estimates of the original data are accepted. The first term attempts to maximize the difference between the trial spectrum and \( s_i^0 \), thereby pushing the solution towards one of the boundaries of the acceptable range. This procedure is carried out for each component, in each case resulting in a solution which will be designated as \( s_i \).

The final stage of the procedure is similar to the one just described except that the function minimized in this case is
in an earlier study. Further experimental details on the procedure used in obtaining these data are given below.

Data sets

Five sets of data were used to test the new algorithm in this work. The first was that originally given by Lawton and Sylvestre in their initial description of SMCR (1). These data consist of spectra measured for 5 mixtures of 2 dyes at 30 wavelength channels and were used to compare the new algorithm to the original procedure for two-component mixtures. The remaining four data sets were recorded for the autocatalytic reaction between permanganate and oxalic acid in a sulfuric acid medium. The conditions for these reactions were intended to duplicate those used by Pimienta et al. (18) in an earlier study. Further experimental details on the procedures used in obtaining these data are given below.

Procedure

All chemicals used were ACS Reagent Grade or better and solutions were prepared in deionized water. Solutions of oxalate were prepared from the disodium salt (Fisher, primary standard) and were acidified by adding the required amount of 1.000 M sulfuric acid (Caledon). Volumetric transfers were carried out using Class A glassware and Eppendorf pipettes (10–100 µL and 1–10 mL).

The permanganate – oxalic acid reaction was carried out by first adding 2.50 mL of a solution containing sodium oxalate and sulfuric acid to a standard 1 cm quartz cuvette (Hellma). To initiate the reaction, a 60.6 µL aliquot of potassium permanganate solution (0.02113 M) was added to the cuvette to give a starting concentration of 5.00 × 10⁻⁴ M for all of the reaction mixtures. Following this addition, timing was initiated and the solution was mixed by inverting it twice before placing it in the optical path of an HP8453 diode array spectrophotometer (Hewlett-Packard, Palo Alto, Calif.). Spectra were recorded at 1 nm intervals between 200 and 1000 nm with an integration time of 0.5 s and the shutter closed between measurements. The initial reagent concentrations, interval between measurements, and total run time for each of the four data sets are given in Table 2. A blank consisting of 5.00 × 10⁻⁴ M sulfuric acid was recorded prior to each kinetic run. All solutions were brought to 25°C prior to initiating the runs and the thermostatted cell holder of the spectrometer was maintained at 25.0°C. No stirring was done during the run.

Computational aspects

All calculations were carried out using programs written in Matlab 5.0 (The MathWorks, Natick, Mass.) on a Sun Microsystems SparcServer 1000 with 230 MB of memory and four 50 MHz SuperSPARC CPUs.

Results and discussion

Two-component data

The algorithm developed in this work was initially applied to the two-component data of Lawton and Sylvestre (1). Although the two-component case is easily solved analytically, these data were included here to validate the new method. Figure 2 shows the spectral boundaries obtained by the Lawton and Sylvestre (LS) method and the new method for both components in the mixture. From the figure, it is clear that there is excellent agreement between the two methods at the crossover points and the shapes of the spectral profiles are very similar, but the boundaries obtained by the new method are somewhat outside of those found by the LS technique. This is as expected, since the new SMCR method assumes a certain level of experimental error and therefore permits a wider range of solutions. In fact, it should be noted that the LS boundaries, so often employed for two-component SMCR problems, are actually too

<table>
<thead>
<tr>
<th>Run</th>
<th>[H₂SO₄] (M)</th>
<th>[H₂C₂O₄] (M)</th>
<th>[KMnO₄] (M)</th>
<th>Time interval (s)</th>
<th>Delay + run time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.27</td>
<td>0.200</td>
<td>0.00050</td>
<td>10</td>
<td>30+600</td>
</tr>
<tr>
<td>2</td>
<td>0.27</td>
<td>0.025</td>
<td>0.00050</td>
<td>8</td>
<td>25+480</td>
</tr>
<tr>
<td>3</td>
<td>0.27</td>
<td>0.0013</td>
<td>0.00050</td>
<td>4</td>
<td>20+180</td>
</tr>
<tr>
<td>4</td>
<td>0.09</td>
<td>0.0013</td>
<td>0.00050</td>
<td>6</td>
<td>20+360</td>
</tr>
</tbody>
</table>

[23] \[ P^*(T) = -\|s_i - s'_i\| + \gamma H(\sigma_i^2 - \sigma_{i'}^2) \]
switch roles as the strongest absorbers. Permanganate is the and the intermediate complexes of Mn(III) and Mn(IV) isosbestic points arise at wavelengths where permanganate tions). This is shown in the enlarged portion of Fig. 4. The around 730 nm (710–740 nm, depending on reaction condi- tions). In addition, measurements made above 700 nm 390 nm. All of these points shift somewhat with reaction Pimienta et al. (18) reported and used the isosbestic point Fig. 4, which shows superimposed spectra for run number 3. The kinetics of this system are characterized by several isosbestic points that occur in the early to middle part of the reaction and then disappear. Some of these are evident in this study. Although component spectra generated by including wave- lengths below 320 nm were of the same general shape as those generated when these wavelengths were excluded, the spectral boundaries were wider and sometimes truncated to zero due to excursions to negative values. Because of this, they were regarded as less reliable. The next step in preparing for SMCR was to build one large “augmented” array from the four kinetic data available. This was done by first appending the data matrix from the second run (61 time channels × 680 wavelength channels) to the bottom of the matrix from the first run (same di- mensions). Runs 3 and 4 were then appended to this in the same way. This is a common practice (see, for example, ref. 13) that can be used whenever all data sets have the same wavelength channels. Although each of the four data matrices were also treated independently by SMCR, there are certain advantages to using the augmented matrix. First, the number of measurements is increased, providing more reliable results from PCA. Second, because several systems (encompassing different conditions) are analyzed simulta- neously, the restrictions of non-negativity constrained solu- tions to narrower (and probably more reliable) regions than when single runs were analyzed. Finally, the use of the aug- mented data set results in one set of spectral boundaries, whereas four separate runs produce four sets of spectral boundaries for each component. Although these should be consistent with one another, they are likely to have broader limits and differ somewhat from run to run, making their interpretation more difficult. Of course, the use of an aug- mented array makes the implicit assumption that the spectra are independent of the kinetic run, which does not seem un- reasonable in this study.

The SMCR analysis of the augmented data matrix was carried out assuming that there were four components present. Although only permanganate and two intermediates were expected, the assumption of four components allowed for the possibility of an additional intermediate, which, if not present, would result in a meaningless spectrum. The spectral boundaries for each of the four components ex- tracted by SMCR are shown in Fig. 5. As noted earlier, SMCR produces an ambiguity of scale, so all of the spectra are normalized to unit area and there is no significance to the absolute scale of the y axis. The first spectrum is clearly that of the permanganate ion and the upper and lower bounds are so close together as to be indistinguishable, indicating very little ambiguity in the shape of this spectrum. The second spectrum, labeled “Intermediate 1,” corresponds very closely to the spectrum of the Mn(III) complex narrow. We were able to generate, with relatively little ef- fort, realistic cases of simulated chromatographic/spectro- scopic data in which the true spectra were outside of the LS boundaries, but inside the boundaries found by the new al- gorithm. This suggests that the results provided by the new algorithm may be more reliable than those based solely on a linear programming approach.

Permanganate – oxalic acid reaction

Figure 3 shows a three-dimensional mesh plot typical of the kinetic data obtained in this work. The example chosen is run number 3, the same as that shown in the earlier work of Pimienta et al. (18), and this figure exhibits virtually identical features. Characteristic of this reaction are the ac- celearting decay of the permanganate peak between 450 and 650 nm due to the autocatalytic reduction of permanganate by oxalic acid to form Mn(II), and the transient absorbance between 300 and 400 nm, which rises to a maximum before finally decaying to zero. This latter feature is attributed to the presence of one or more intermediates. Pimienta et al. (18) have proposed two intermediates, the first a well-recognized Mn(III) species (bis(oxalato)manganate (III), [Mn(C₂O₄)₂]⁻), and the second a soluble Mn(IV) species. The aim of the present work was to use this autocatalytic reaction system to test the new SMCR algo- rithm and, if possible, confirm the presence of the two inter- mediates with the spectral characteristics reported by Pimienta.

The kinetics of this system are characterized by several isosbestic points that occur in the early to middle part of the reaction and then disappear. Some of these are evident in Fig. 4, which shows superimposed spectra for run number 3. Pimienta et al. (18) reported and used the isosbestic point around 450 nm, and its presence was confirmed in this work, along with similar isosbestic points around 335 nm and 390 nm. All of these points shift somewhat with reaction conditions. In addition, measurements made above 700 nm in this work established an additional early isosbestic point around 730 nm (710–740 nm, depending on reaction condi- tions). This is shown in the enlarged portion of Fig. 4. The isosbestic points arise at wavelengths where permanganate and the intermediate complexes of Mn(III) and Mn(IV) switch roles as the strongest absorbers. Permanganate is the stronger absorber between 335 and 390 nm and between 450 and 730 nm; the complexes are the stronger absorbers below 335 nm, between 390 and 450 nm, and above 730 nm. The shift in isosbestic points has been used to suggest the presence of more than one intermediate.

The analysis of the kinetic data began by removing measure- ments at wavelengths below 320 nm. This was done for two reasons. First, it eliminated possible absorbance by ox- alic acid in the UV region, thereby reducing the number of components by one and simplifying the problem. Second, the absorbances at shorter wavelengths tended to be high, intro- ducing problems with nonlinear response and larger noise levels, both of which complicated the mathematical analysis. Although component spectra generated by including wave- lengths below 320 nm were of the same general shape as those generated when these wavelengths were excluded, the spectral boundaries were wider and sometimes truncated to zero due to excursions to negative values. Because of this, they were regarded as less reliable.
reported by Pimienta et al. (18) and has fairly tight boundaries as well. This is consistent with the fact that this species is a well-recognized intermediate. The third spectrum represents the second intermediate, which has been attributed to a soluble Mn(IV) species. Again, this spectrum is in good agreement with the spectrum reported in the literature (18), although the spectral bands are substantially wider than for the Mn(III) species, particularly at shorter wavelengths. This indicates that there is a fair amount of difficulty determining the spectrum of this species, probably because its contribution is dominated by other absorbing species (although it is also possible that the spectrum is sensitive to reaction conditions). The fourth spectrum in the series, although it exhibits some systematic features, is not considered to support the presence of an additional intermediate. Ideally, this spectrum should appear completely random in the absence of another component, but the features observed, which occur mainly in high-absorbance regions, are likely compensating for
small amounts of nonlinearity in the response. In PCA it is not unusual to see this type of behaviour when the nonideal characteristics of the instrument are compensated for by additional eigenvectors. Similar structure in the residual spectra was observed when the system was modeled with five components, consistent with non-chemical artifacts.

To facilitate a better comparison between the spectral boundaries obtained for the intermediates and those reported by Pimienta et al. (18), the latter results were digitized and the two sets of spectra were scaled to a common area. The results are compared in Fig. 6. It is clear that there is excellent agreement between the two sets of results, with the spectra reported by Pimienta (indicated by the open circles) closely matching one of the spectral boundaries (solid lines) in both cases. Since the results of Pimienta et al. were obtained through an alternative approach, assuming the presence of only two intermediates and making use of isosbestic point data, this serves as validation of the SMCR results. The difference is that the latter results were obtained without any assumptions other than the non-negativity of spectral and concentration profiles.

Also shown in the figure are spectral boundaries obtained by the Sasaki algorithm (dotted lines). For both species, the upper and lower boundaries extracted by the Sasaki algorithm are completely overlapped and so appear as a single line. This demonstrates one of the weaknesses of this algorithm when even small negative values remain in the data after PCA. Because the penalty function in the Sasaki algorithm is essentially absolute in its rejection of negative values, the optimization routine ignores the spectral contribution to the objective functions in eqs. [9] and [10] and simply minimizes the penalty, giving the same result for both boundaries. In other words, whenever negative spectral and concentration contributions cannot be removed, regardless of how small they are, a single solution results. In this case, the Sasaki solutions match one of the boundaries from our algorithm for both species, but it is clear that there is a discrepancy with Pimienta’s results for the first intermediate.

The concentration profiles extracted by SMCR for each of the three species are shown in Fig. 7. As with the spectra, the concentration profiles extracted by SMCR normally have an ambiguity of scale, so while comparisons can be made for the same species in different runs, interspecies comparisons are not generally possible. In this case, however, the spectra reported by Pimienta et al. (18) were used to bring the concentration profiles for the intermediates to an approximately comparable scale. This means that the relative magnitudes of the concentration profiles in the last two columns of Fig. 7 can be compared directly. The concentration profiles for MnO$_4^-$ are on a separate scale, however, and can only be compared within the column.

As in the case of the spectral profiles, the concentration profiles for MnO$_4^-$ show very tight boundaries, to the point of being indistinguishable. These show a very clear picture of the autocatalytic reduction under different reaction conditions. The concentration profiles for the intermediates show generally wider bands and exhibit differing contributions from each species under different reaction conditions. In an attempt to quantify the relative contribution of each species for each run, Pimienta et al. (18) reported the percentage of intermediate 2 (Mn(IV)) to the total intermediate contribution in each run. This assumed a constant ratio of the intermediates, which is probably approximately valid for a significant portion of the reaction. They reported values of 9, 31, 60, and 68% Mn(IV) in runs 1 through 4, respectively. By integrating the mean concentration profiles for our results, we obtained values of 7, 29, 50, and 54%. Given the approximate nature of both calculations, this agreement is quite good. More importantly, however, the results presented in Fig. 7 give a picture of what is happening to species in solution, allowing a better visualization of reaction dynamics.

Conclusions

The principal objective of this work has been to introduce and demonstrate a modified SMCR procedure based on nonlinear optimization. The new procedure is substantially faster than the former procedure on which it is based, requiring many fewer steps in the optimization. In addition, the new method handles experimental measurement error in a more logical way and, unlike other SMCR methods, allows small departures of the solutions from the PCA subspace if these are warranted by measurement uncertainty. The new method requires no pure component regions or constraints other than non-negativity. Application of the method to the two-component data of Lawton and Sylvestre (1) led to results consistent with the analytical solution, but with broader boundaries, as expected when accounting for experimental error. Furthermore, the new method works in certain cases where the analytical solution fails because of the presence of measurement uncertainty. The new method was also applied to kinetic data for the permanganate oxidation of oxalic acid. This was an excellent case study, since (i) there were no clear pure component regions, (ii) there were a wide range of signal amplitudes, and (iii) results from an alternative approach were available for comparison. Results confirmed the presence of the two intermediates proposed by Pimienta et al. (18) and the spectra extracted by SMCR were in excellent agreement with those reported earlier. This clearly indicates the potential for SMCR in kinetic studies. In addition,
our laboratory has applied this algorithm to chromatography with multiwavelength fluorescence detection and to environmental receptor modeling. Given the wide range of studies now producing second-order data, tools such as this are likely to become more important in the future.

The authors gratefully acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada and helpful discussions with Prof. T.S. Cameron.

References
