Computation of the Range of Feasible Solutions in Self-Modeling Curve Resolution Algorithms

Paul J. Gemperline

Department of Chemistry, East Carolina University, Greenville, NC 27858

Self-modeling curve resolution (SMCR) describes a set of mathematical tools for estimating pure-component spectra and composition profiles from mixture spectra. The source of mixture spectra may be overlapped chromatography peaks, composition profiles from equilibrium studies, kinetic profiles from chemical reactions and batch industrial processes, depth profiles of treated surfaces, and many other types of materials and processes. Mathematical solutions are produced under the assumption that pure-component profiles and spectra should be nonnegative and composition profiles should be unimodal. In many cases, SMCR may be the only method available for resolving the composition profiles and pure-component spectra from these measurements. Under ideal circumstances, the SMCR results are accurate quantitative estimates of the true underlying profiles. Although SMCR tools are finding wider use, it is not widely known or appreciated that, in most circumstances, SMCR techniques produce a family of solutions that obey nonnegativity constraints. In this paper, we present a new method for computation of the range of feasible solutions and use it to study the effect of chromatographic resolution, peak height, spectral dissimilarity, and signal-to-noise ratios on the magnitude of feasible solutions. An illustration of its use in resolving composition profiles from a batch reaction is also given.

1. BACKGROUND

Self-modeling curve resolution (SMCR) was first introduced by Lawton and Sylvester.[1] The method, as first reported, resolved two-component mixture spectra into nonnegative concentration profiles and nonnegative absorption spectra of the pure constituents without any prior knowledge. With proper scaling, they showed that mathematical solutions for the pure-component spectra which satisfy nonnegativity constraints must lie on two line segments bounded by upper and lower bounds, yielding a range of feasible solutions for each estimated pure-component spectrum. Examples followed several years later showing the resolution of three-component mixtures by use of a new SMCR method called iterative target transformation factor analysis (ITTFA).[2,3] Since then, SMCR has been widely studied in the chemometrics literature. A section dedicated to resolution regularly appears in the biannual Fundamental Reviews in Chemometrics published in *Analytical Chemistry*.[4] At last count, more than 20 different SMCR methods or variations have been reported in the chemical literature. These techniques include evolving factor analysis (EFA),[5,6,7] window factor analysis (WFA)[8,9] simple-to-use interactive self-modeling mixture analysis (SIMPLISMA),[10] orthogonal projections (OP),[11,12] heuristic evolving latent projection (HELP),[13,14,15] eigen-structure tracking analysis (ETA),[16,17] fixed-size window evolving factor analysis (FSW-EFA),[18] latent projection graphs (LPG),[19] and others. Nearly all of these techniques were originally developed and applied to the resolution of overlapped chromatographic peaks from hyphenated instruments such as GC-MS or HPLC with UV/visible diode detectors.

1.1. SMCR Solutions Are Not Unique. Tauler et al.[20] identified two classes of ambiguity in SMCR results: intensity ambiguity and rotational ambiguity. Intensity ambiguity describes how the amplitude of a concentration profile, $\mathbf{c}_i$, can be increased with a corresponding decrease in its matching spectrum, $\mathbf{p}_i$. The product of the two signals remains constant, however, and the sum or integrated signal due to the constituent of interest can provide a quantitative measure of the total amount of the species present in the mixture. Rotational ambiguity refers to the infinite number of nonorthogonal rotations that give estimated composition profiles and spectra that satisfy the nonnegativity constraints. Under special circumstances, it is claimed that unique solutions

* Corresponding author: (tel.) 252-328-6767; (fax) 252-328-6210; (e-mail) gemperlinep@mail.ecu.edu.

(1) Lawton, W. H.; Sylvestre, E. A. *Technometrics* **1971**, *13*, 617−633.
(2) Gemperline, P. J. *J. Chem Inf. Comput. Sci.* **1984**, *24*, 206−212.
(3) Vandeginste, B. G. M.; Derks, W.; Kateman, G. *Anal. Chim. Acta* **1985**, *173*, 253−264.
(4) Brown, S. D.; Sum, S. T.; Despagne, F. *Anal. Chem.* **1996**, *68*, 21R−61R.
(5) Gampp, H.; Maeder, M.; Meyer, J. C.; Zuberbuhler, A. D.; *Talanta* **1985**, *32*, 1133−1139.
(6) Meader, M.; Zuberbuehler, A. D. *Anal. Chim. Acta* **1986**, *181*, 287−291.
(7) Maeder, M. *Anal. Chem.* **1987**, *59*, 527−530.
(8) Malinowski, E. R. *J. Chemom.* **1992**, *6*, 29−40.
(9) Malinowski, E. R. *J. Chemom.* **1996**, *10*, 273−279.
(10) Windig, W.; Guilment, J. *Anal. Chem.* **1991**, *63*, 1425−1432.
(11) Cuesta Sanchez, F.; Toft, J.; van den Bogaert, B.; Massart, D. L. *Anal. Chem.* **1996**, *68*, 79−85.
(12) Cuesta Sanchez, F.; Vandeginste, B. G. M.; Hancewicz, T. M.; Massart, D. L. *Anal. Chem.* **1997**, *69*, 1477−1484.
(13) Kvalheim, O. M.; Liang, Y.-Z. *Anal. Chem.* **1992**, *64*, 936−946.
(14) Kvalheim, O. M.; Liang, Y.-Z. *Anal. Chem.* **1992**, *64*, 946−953.
(15) Liang, Y.-Z.; Kvalheim, O. M. *Chemom. Intell. Lab. Syst.* **1993**, *20*, 115−125.
(16) Toft, J.; Kvalheim, O. M. *Chemom. Intell. Lab. Syst.* **1993**, *19*, 65−73.
(17) Cuesta Sanchez, F.; Toft, J.; Kvalheim, O. M.; Massart, D. L. *Anal. Chim. Acta* **1995**, *314*, 131−139.
(18) Keller, H. R.; Massart, D. L. *Anal. Chim. Acta* **1991**, *246*, 379−390.
(19) Vacque, V.; Dupuy, N.; Sombret, B.; Huvenne, J. P.; Legrand, P. *Appl. Spectrosc.* **1997**, *51*, 407−415.
(20) Tauler, R.; Smilde, A.; Kowalski, B. *J. Chemom.* **1995**, *9*, 31−58.

10.1021/ac990648y CCC: $18.00
Published on Web 00/00/0000

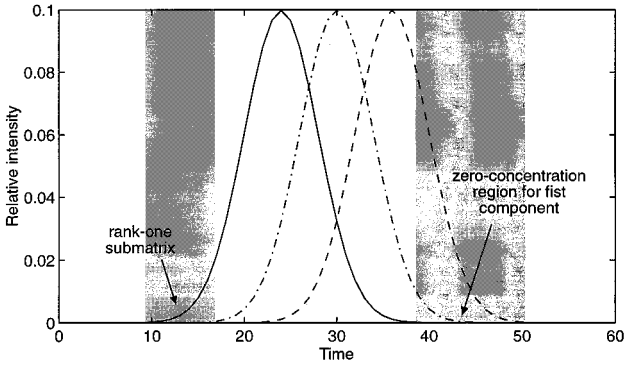

Figure 1. Simulated three-component chromatogram showing regions in time where a rank-one submatrix would be expected and a zero-component region for component one would be expected. Similar regions can be identified for other components of the chromatogram.

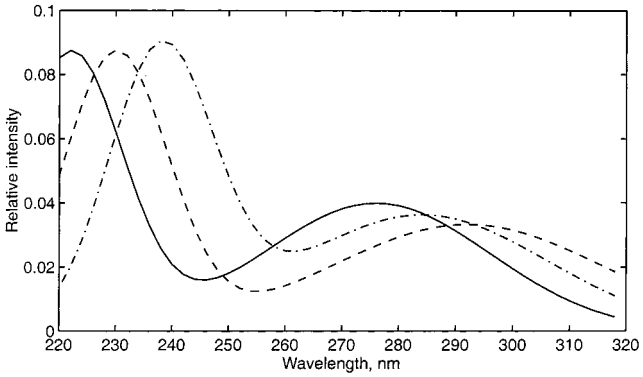

Figure 2. Simulated UV/visible spectra corresponding to the peaks shown in Figure 1.

can be produced when SMCR techniques are applied to resolution of overlapped chromatographic peaks.[9,13,21] These techniques rely upon identification of so-called zero-component regions and rank one windows (See Figure 1) to provide unique resolution. In the strictest sense this claim can only be true in the absence of measurement noise when the positions of window boundaries of zero-component regions are absolutely certain. The situation is even more difficult when embedded peaks are encountered, e.g., a peak which has boundaries that lie completely inside the boundaries of another component.[22] This situation may be frequently encountered when SMCR is applied to batch reactions when reactive intermediates are present.

In this paper we identify four different effects of measurement error that can contribute to the uncertainty in SMCR results: (1) uncertainty in the position of window boundaries, (2) measurement error that propagates into nonnegativity constraints, (3) measurement error in pure-component spectra derived from rank-one submatrixes, and (4) measurement error that propagates when projecting a full set of measurements into subspaces spanning zero concentration regions. Since the nonnegative constraints are derived from the experimental measurements, error in the measurements propagates into the constraints, leading to uncertainty in the location of the range of feasible solutions. Locating the boundaries of zero-component regions requires the determination of local rank, and error in the measurements leads to uncertainty in the determination of these boundaries. For example, consider a three-component simulated spectro-chromatographic data set generated by eq 1 using the pure chromatographic profiles and pure-component spectra shown in Figures 1 and 2

$$\mathbf{A} = \mathbf{C}\mathbf{P}^T \qquad (1)$$

where $\mathbf{A}$ is an $n \times m$ data matrix of mixture spectra with a maximum signal intensity of 1.00 AU, one spectrum per row; $\mathbf{C}$ is an $n \times k$ matrix of pure composition profiles, one profile per column; $\mathbf{P}$ is a $k \times m$ matrix of pure-component spectra; and k is the number of components, in this case three.

After adding random normal deviations with a mean of zero and a standard deviation of 0.0025 (noise level 1) or a standard deviation of 0.00025 (noise level 2), evolving window factor analysis with a window width, $w = 4$, was performed to locate the boundaries of zero-component and rank-one windows. The uncertainty in the position of these boundaries can be clearly seen in Figure 3.

1.2. Nonnegativity Constraints in SMCR. Borgen and Kowalski reported an analytical solution to the two-component constrained-boundary problem and gave a software algorithm for finding the range of feasible nonnegative composition profiles.[23] Borgen et al. later reported an extension of the solution to N components and illustrated its use on a three-component problem.[24] The technique was mathematically difficult to understand and implement. A few years later, Henry and Kim published a solution for finding the feasible ranges for an N-component resolution problem by use of linear programming with constraints formulated in a different eigenspace and applied the technique to simulated airborne particulate composition data.[25] Most recently, Wentzell et al. published a new procedure for finding the permissible bands for the spectra and concentration profiles of mixtures using a simplex optimization procedure and applied it to resolving the composition profiles in the autocatalytic oxidation of oxalic acid by permanganate in the presence of sulfuric acid.[26]

Recently, we developed a conceptually simple approach[27] for finding the constrained-boundary problem that utilizes Matlab's constrained optimizing routine, CONSTR, found in the Matlab Optimization Toolbox.[28] The method works by finding the transformation matrix, $\mathbf{B}$, which maximizes (upper bound) or minimizes (lower bound), individually, the integrated signal of each constituent in the mixture, subject to simultaneous nonnegativity constraints in the concentration domain and the spectral domain. Constraints are automatically generated from the measurements and the SMCR model, consequently, measurement noise propagates into the constraints in varying amounts according

(21) Liang, Y.-Z.; Kvalheim, O.M.; Manne, R. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 235−250.
(22) Manne, R. *Chemom. Intell. Lab. Syst.* **1995**, *27*, 89−94.
(23) Borgen, O. S.; Kowalski, B. R. *Anal. Chim. Acta* **1985**, *174*, 1−26.
(24) Borgen, O. S.; Davidsen, N.; Mingyang, Z.; Oyen, O. *Mikrochim. Acta* **1986**, *2*, 63−73.
(25) Henry, R. C.; Kim, B. M. *Chemom. Intell. Lab. Syst.* **1990**, *8*, 205−216.
(26) Wentzell, P. D.; Wang, J.-H.; Loucks, L. F.; Miller, K. M. *Can. J. Chem.* **1998**, *76*, 1−12.
(27) Quinn, A. C.; Gemperline, P. J.; Baker, B.; Zhu, M.; Walker, D. S. *Chemom. Intell. Lab. Syst.* **1999**, *45*, 199−214.
(28) *Optimization Toolbox, Version 1.5.2*, The Mathworks, Inc.: Natick, MA, 1990, 1996.

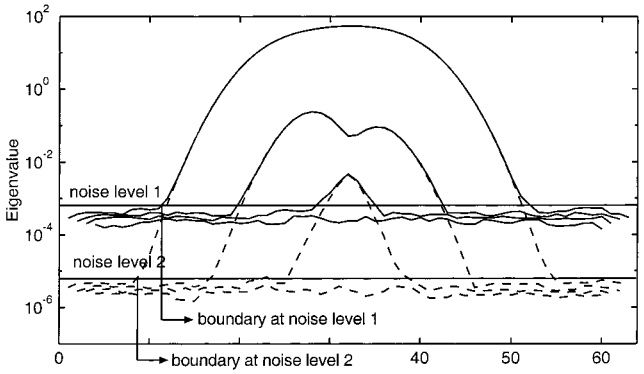

Figure 3. Evolving window factor analysis of a simulated spectrochromatographic data set at two noise levels produces different boundaries.

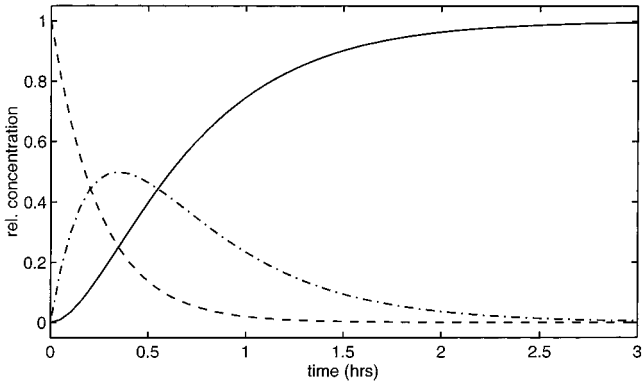

Figure 4. Simulated composition profiles for the sequential nonreversible reaction A + B ⇔ C ⇔ D + E. Reactant A (- - -), intermediate C (-·-), and product D (—) are shown.

to the net analyte signal for the corresponding constituents. Our Matlab code attempts to compensate for this problem by finding an adaptive, robust noise estimate for each profile which is then used to "relax" the constraints slightly to allow small negative deviations. A detailed description of the algorithm is given later.

A simulated data set for a batch reaction was generated by numerical integration of the appropriate rate laws for the sequential reversible reaction A + B ⇔ C ⇔ D + E. Pure-component profiles for species A, C, and D are shown in Figure 4 with species B and E being nonabsorbing. The simulated spectra shown in Figure 5 were generated so that a nearly unique response was obtained for reactant A at 285 nm, intermediate C at 310 nm, and product D at 350 nm. These pure-component profiles and spectra were used to produce a data matrix of mixture spectra that was scaled to a maximum absorbance of 0.500. Random normal deviations with a mean of zero and a standard deviation of 0.001 were added to simulate measurement noise.

Figures 6 and 7 show the SMCR results and range of feasible profiles obtained for the simulated batch reaction data described in Figures 4 and 5. An initial solution to the curve resolution problem was obtained by using ITTFA[2] and used as the starting point for calculating the range of feasible solutions. Any of the previously cited SMCR methods could have been used to initialize the computation. The first measured spectrum, being a spectrum of the pure reactant A, was tagged in the computation and not allowed to change in the search for the boundary solutions. This condition forces the feasible range of the spectrum of reactant A

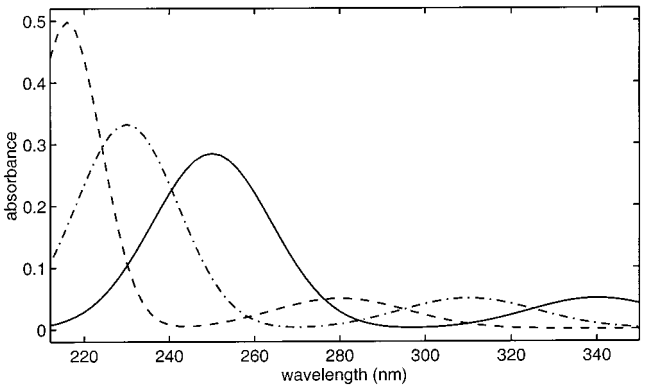

Figure 5. Simulated UV/Visible spectra for the sequential nonreversible reaction shown in Figure 4. Reactant A (- - -), intermediate C (-·-), and product D (—) are shown.

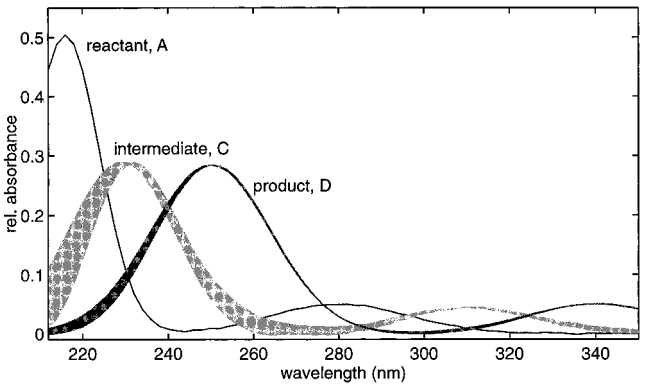

Figure 6. SMCR results and range of feasible profiles for simulated data described in Figures 4 and 5.

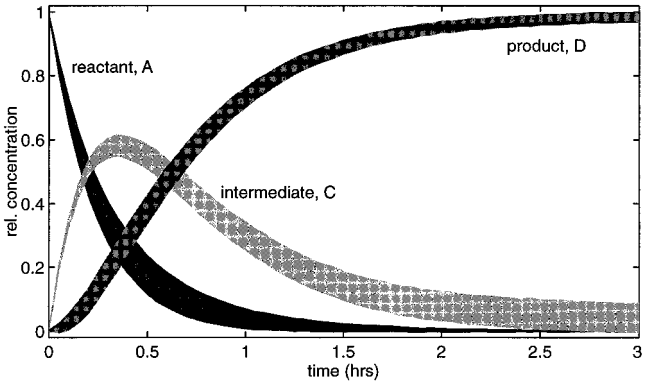

Figure 7. SMCR results and range of feasible pure-component spectra for the simulated data described in Figures 4 and 5.

to zero (see Figure 7). Additionally, the composition profiles for intermediate C and product D are forced to zero at time = 0 by this condition (see Figure 6). Except for the first spectrum in the data matrix, no rank-one subwindows or zero-concentration subwindows may be identified, as the last sequence of spectra may be a steady-state mixture of reactant(s), intermediate(s), and product(s) at thermodynamic equilibrium. Because of the first spectrum's purity restriction and spectral regions of nearly full selectivity at 285, 310, and 350 nm, the feasible bands obtained for the composition profiles are fairly narrow for this difficult embedded-peak problem.

1.3. New Method for Finding Feasible Ranges. The new method for finding feasible ranges begins with a matrix of mixture

spectra, $\mathbf{A}(n \times m)$, with spectra measured at n time intervals and m wavelengths. Before the search for the upper and lower feasible boundaries can be performed, an initial solution to the curve-resolution problem must be found using any one of the SMCR methods cited previously. We used ITTFA to initialize all the calculations described in this paper. ITTFA is performed in the k-factor subspace determined by the singular value decomposition of $\mathbf{A}$

$$\mathbf{A} = \mathbf{U}_k\mathbf{S}_k\mathbf{V}_k^T + \mathbf{E}_k \tag{2}$$

where $\mathbf{U}_k$, $\mathbf{S}_k$, and $\mathbf{V}_k$, are the truncated matrixes of left eigenvectors, singular values, and right eigenvectors of $\mathbf{A}$, respectively, and $\mathbf{E}_k$ is the matrix of residuals at k factors. The number of factors to be retained, k, is determined by use of an appropriate F-test. The initial solution point provided by SMCR can be represented as a least-squares transformation matrix, $\mathbf{B}(k \times k)$, which is used to estimate pure-composition profiles, $\mathbf{C}_p$, and spectral profiles, $\mathbf{E}_p$.

$$\mathbf{C}_p = \mathbf{U}_k\mathbf{S}_k\mathbf{B} \tag{3}$$

$$\mathbf{P}_p = \mathbf{B}^{-1}\mathbf{V}_k^T \tag{4}$$

Substitution of eqs 3 and 4 into eq 1 provides a complete solution to the curve resolution problem.

$$\mathbf{A} = (\mathbf{USB})(\mathbf{B}^{-1}\mathbf{V}^{\mathrm{T}}) + \mathbf{E}_k \tag{5}$$

$$\mathbf{A} = \mathbf{C}_p\mathbf{P}_p + \mathbf{E}_k \tag{6}$$

The new method for computing the boundaries of feasible ranges solves the problem by adjusting the $k \times k$ least-squares solution, $\mathbf{B}$, for each of the k resolved constituents in the mixture spectra. For example, with a $k = 3$ component mixture, six boundary conditions are determined, one pair of upper and lower bounds for each constituent. Each pair of upper and lower boundaries are determined according to the following procedure. A rank-one matrix, $\mathbf{A}_q$, for the constituent of interest, q, is computed

$$\mathbf{A}_q = \mathbf{c}_q\mathbf{p}_q \tag{7}$$

and the integrated signal due to the constituent is approximated by taking the sum over n and m as shown in eq 8.

$$\max\{\sum_{i=1}^{n}\sum_{j=1}^{m}(\mathbf{c}_q\mathbf{p}_q)\} \text{ or } \min\{\sum_{i=1}^{n}\sum_{j=1}^{m}(\mathbf{c}_q\mathbf{p}_q)\} \tag{8}$$

The upper boundary condition for the qth constituent is found by maximizing this sum, and the lower boundary condition is found by minimizing this sum, subject to following simultaneous nonnegativity constraints for all k concentration profiles and spectral profiles.

$$\mathbf{C} > -\alpha E_{\mathbf{C}} \text{ and } \mathbf{P} > -\alpha E_{\mathbf{E}} \tag{9}$$

The constants, $\alpha E_{\mathbf{C}_p}$ and $\alpha E_{\mathbf{E}_p}$, in eq 9 are robust noise estimates (described later) of the composition profiles and spectral profiles, respectively. Instead of setting the constraints in eq 9 greater than zero, $\mathbf{C} > 0$ and $\mathbf{P} > 0$, robust noise estimates allow small negative deviations in the composition profiles and spectral profiles in zero-concentration regions and zero-absorbance regions, respectively. Small, negative deviations from zero occur in these regions because measurement error propagates into the estimated profiles. The factor, α, is typically set to 1.0, but may be adjusted to larger values if the optimization procedure fails to find an initial feasible solution due to constraints that are too narrow. For components that exhibit zero-concentration regions as determined by local rank analysis, additional constraints shown in eq 10 are added to ensure the respective composition profiles approximately equal to zero

$$\mathbf{C}_{(t)} < \alpha E_{\mathbf{C}} \tag{10}$$

where $\mathbf{C}_{(t)}$ represents the portion of the composition profiles at times, t, where the concentration is expected to be zero.

The solution point, $\mathbf{B}(k \times k)$, is updated during each cycle of the iteration; thus, the estimated pure composition profiles and spectral profiles, $\mathbf{C}_p$ and $\mathbf{P}_p$, are also recomputed (eqs 3 and 4) at each cycle of the iterative search process. Random measurement noise propagates into these estimates to varying degrees, depending on the selectivity and sensitivity of the measurements for the different constituents in the mixture. As a consequence, the random error in each estimated profile may change from one cycle of the iterative process to the next. For this reason, robust estimates of the random measurement error are computed during each cycle of the iteration for use in eqs 9 and 10.

Noise levels in the resolved profiles are estimated using 6 × the median absolute deviation from a double smoothing procedure. In the first smoothing step, the estimated concentration profiles and spectral profiles are smoothed by a three-point binomial filter. The difference between the smoothed and unsmoothed profiles is smoothed a second time, leaving only random deviations in the second residual curve. The median absolute deviation of the second difference between the smoothed residual and unsmoothed residuals is used to estimate the noise level in each individual profile.

Since the magnitude of any single constituent's composition profile can be decreased by increasing the magnitude of its corresponding pure spectrum, proper scaling while solving the constrained optimization problem is necessary to give smooth convergence. The estimated boundary spectra, $\mathbf{P}_p$, were rescaled at each iteration of the search so that their maximum absorbance matches the absorbance of the starting (input) spectra. This scaling causes amplitude variation in the boundary solutions to appear in composition profiles, not spectra, which seems to be the most intuitive option for chemists. To further improve convergence, the functions to be minimized and maximized were scaled to 1.0 prior to calling the optimization routine.

1.4. Implementation Details. A subroutine function for computing the double sum in eq 8, the estimated profiles, $\mathbf{C}_p$ and $\mathbf{P}_p$ (eqs 3 and 4), the constraints (eqs 9 and 10), and the robust noise estimates, $E_{\mathbf{C}_p}$ and $E_{\mathbf{E}_p}$ is provided as an input argument to the Matlab CONSTR function. This subroutine function is called

by CONSTR once at each cycle of the iterative process as it searches for an upper or lower boundary condition. A single call to the general purpose Matlab function CONSTR performs the optimization for one boundary condition. For example, the CONSTR optimization function would be called six times for a three-component resolution problem, twice for each constituent to find the pair of maximum and minimum boundary conditions described by eq 8.

2. EXPERIMENTAL SECTION

2.1. Simulated Chromatography Data. The chromatograms and spectra shown in Figures 1 and 2 were used to generate simulated three-component mixtures with overlapping chromatography peaks to illustrate the effects of measurement error on the range of feasible solutions produced by SMCR. Each data matrix was scaled to have a maximum signal intensity of 1.0. Random measurement errors were simulated by adding scaled random numbers with a mean of zero to the simulated data matrix.

The effect of uncertainty in the determination of zero-component boundaries was investigated at a chromatographic resolution of 0.375 and equal peak heights for all three peaks. Simulated random measurement errors were scaled to 0.01% of the maximum signal intensity. The minimum and maximum boundaries of the feasible ranges were computed using different boundaries for the zero-component regions of peak one and peak three.

The effect of random measurement error on the range of feasible solutions was investigated at a chromatographic resolution of 0.50 and equal peak heights for all three peaks. Seven different trials were used with simulated measurement errors scaled to 0.01%, 0.02%, 0.05%, 0.10%, 0.20%, 0.50%, and 1.0% of the maximum signal intensity.

The effect of peak height on the range of feasible solutions was investigated at a chromatographic resolution of 0.50 and peak heights of 1:1:1, 1:1/2:1, 1:1/4:1, and 1:1/8:1. Simulated random measurement errors were scaled to 0.10% of the maximum signal intensity.

The effect of chromatographic resolution on the range of feasible solutions was investigated at resolutions of 0.375, 0.50, 0.625, 0.750, and 0.875 using equal peak heights for all three peaks. Simulated random measurement errors were scaled to 0.10% of the maximum signal intensity.

2.2. Batch Reaction Data. An industrially significant laboratory-scale batch reaction was monitored with UV/visible fiber-optic probe using previously described conditions.[29] An amount of 7.39 g of the starting material, a nucleoside, (Glaxo-Wellcome, Research Triangle Park, NC) was dissolved in 31.5 mL of pyridine (Fisher Scientific, Fair Lawn, NJ) in a 100-mL, three-neck round-bottom flask. The temperature of the reaction was held constant by a heating mantle at 50 °C. An amount of 8.40 g of trityl chloride (triphenylmethyl chloride, Aldrich Chemical Co., Milwaukee, Wisconsin) was added to the mixture and allowed to react for 5 h. The amounts of reagents were chosen so that the reaction mixture would fully cover the tip of a fiber-optic probe. The mixture was stirred continuously using a magnetic stirrer and a Teflon coated stirring bar. The flask was fitted with a thermometer, a drying tube filled with anhydrous calcium sulfate, and a fiber-optic probe in the third opening. Spectra of this reaction were recorded at specified time intervals using a UV/visible fiber-optic spectrograph and a custom built PEEK (polyether-ether ketone) single-fiber transmission probe with a 1-mm path length.[30,31] The spectra were taken at 30-s intervals during the first 15 minutes, 2 min intervals during the next 30 minutes, and 5 min intervals for the remainder of the reaction. The real-time measurements were stored directly into MATLAB binary files using data acquisition software written in-house.

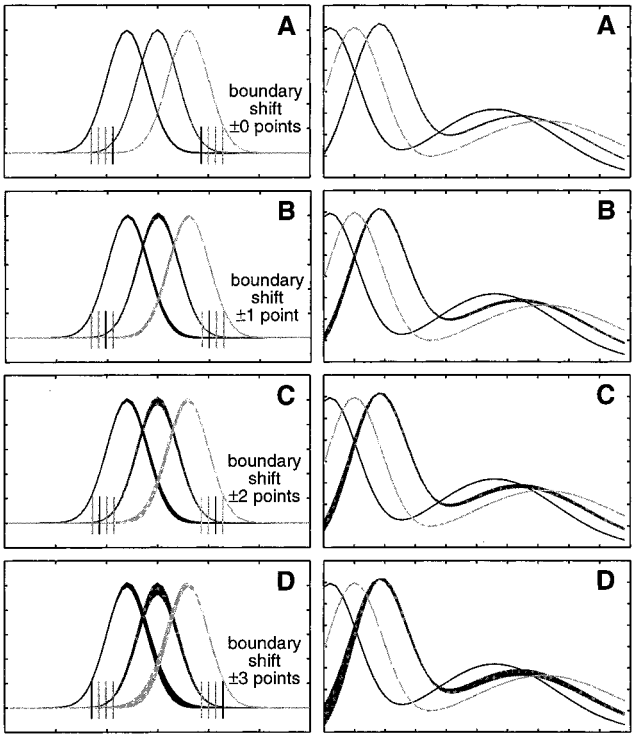

Figure 8. Effect of uncertainty in the position of zero-component regions on the range of feasible solutions produced by SMCR. Estimated composition profiles and corresponding estimated pure-component spectra are shown.

3. RESULTS

3.1. Simulated Chromatography Data. Analysis of the simulated chromatography data produced the range of feasible results shown in Figures 8−9 and Table 1. During the search for the boundaries of the feasible range, the spectra of the first peak and the third peak were not adjusted, since the pure-component spectra of these two peaks could easily be estimated from the rank-one submatrixes at the beginning and end of the peak cluster. Only the spectrum of the middle peak was adjusted. Despite these restrictions, a wide range of variation was observed in the shape and area of the estimated curves for the first and third peaks. At low noise levels, the SMCR results faithfully reproduced the actual peak areas as long as the window boundaries for zero-component regions of the first peak and third peak were accurately specified (see Figure 8A).

(29) Quinn, A.; Gemperline, P. J.; Baker, B.; Zhu, M.; Walker, D. S. Chemom. Intell. Lab. Syst. **1999**, *45*, 199−214.

(30) Cho, J.; Gemperline, P. J.; Walker, D. *Appl. Spectrosc.* **1995**, *49*(12), 1841−1845.

(31) Cho, J. H.; Gemperline, P. J.; Salt, A.; Walker, D. S. *Anal. Chem*. **1995**, *67*, 2858−2863.

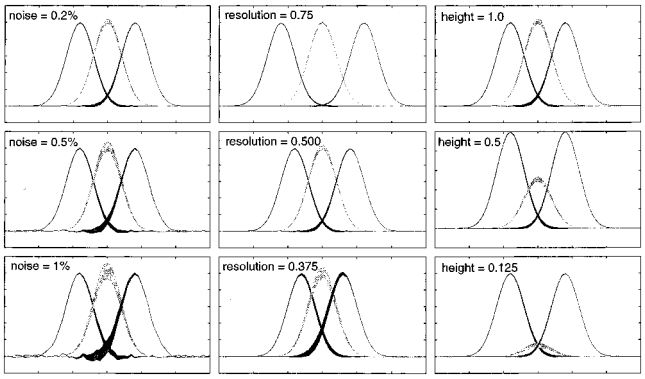

Figure 9. Effect of measurement noise, chromatographic resolution, and peak height on the range of feasible solutions produced by SMCR.

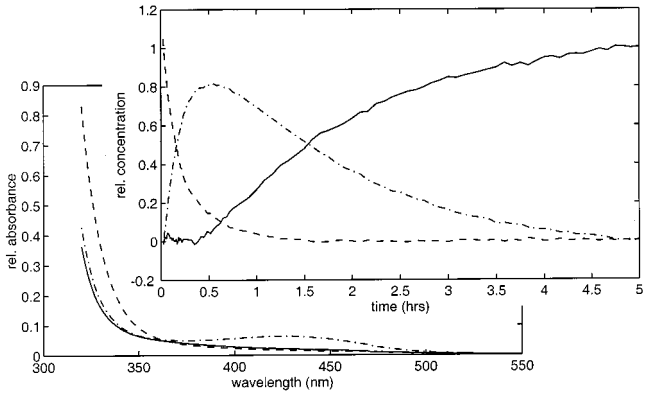

Figure 10. ITTFA results from the analysis of the tritylation of a nucleoside.

Table 1: Summary of Effects Studied and Range of Feasible Solutions Produced

resolution	window adj.	height	% noise	% error in peak area minimum	maximum	total
			Effect of Uncertainty in Boundaries			
0.375	0	1	0.01	−0.25	0.30	0.56
0.375	1	1	0.01	−2.06	3.69	5.74
0.375	2	1	0.01	−5.71	3.39	9.10
0.375	3	1	0.01	−13.06	3.49	16.56
			Effect of Measurement Noise			
0.500	0	1	0.01	0.00	0.00	0.00
0.500	0	1	0.02	−0.10	0.00	0.10
0.500	0	1	0.05	−0.64	2.63	3.26
0.500	0	1	0.1	−1.30	3.58	4.87
0.500	0	1	0.2	−3.39	4.63	8.03
0.500	0	1	0.5	−9.33	7.60	16.93
0.500	0	1	1	−19.11	12.19	31.31
			Effect of Peak Height			
0.500	0	1	0.1	−1.30	3.58	4.87
0.500	0	0.5	0.1	−3.21	7.01	10.22
0.500	0	0.25	0.1	−9.54	11.88	21.42
0.500	0	0.125	0.1	−14.96	29.27	44.23
			Effect of Chromatographic Resolution			
0.875	0	1	0.1	−0.36	0.89	1.26
0.750	0	1	0.1	−0.59	1.30	1.89
0.625	0	1	0.1	−0.72	1.47	2.19
0.500	0	1	0.1	−1.30	3.58	4.87
0.375	0	1	0.1	−7.65	10.12	17.77

The additional constraints described by eq 10 were included in the search for the minimum and maximum boundaries because of the presence of zero-component regions on the left- and right-hand sides of the cluster of peaks. On the left side of the first boundary shown in Figure 8, the composition profile of the third peak was forced to be close to zero. A small allowance for positive and negative deviations was included as specified by the robust noise estimates described earlier. On the right side of the second boundary shown in Figure 8, the composition profile of the first peak was also forced to be close to zero. When the boundaries of these zero-component regions were made wider, the range of feasible solutions increased as shown in Figures 8B−8D and Table 1. At the lower bound for the middle peak, the width of the first and third peaks increased, and the amplitude of the middle peak became smaller. There was a corresponding change in the estimated spectrum of the middle peak characterized by a decrease in its correlation to the spectra of the first and third peaks. The variance captured by the middle peak decreased, and there was a corresponding increase in the total variance captured by the first and third peaks.

Figure 9 shows the range of feasible solutions for some of the conditions described in Table 1. Estimated spectra are not shown in Figure 9; however, the results were visually similar to those shown in Figure 8. As before, the spectra of the first and third peaks were not adjusted in the search for the minimum and maximum boundaries because good estimates of the pure-component spectra of these two peaks were obtained from the rank-one submatrixes at the beginning and end of the peak cluster. As expected, each effect studied caused the range of feasible solutions to increase as the effect became more severe. For example, as the level of measurement noise increased, greater uncertainty in the constraints allowed a wider range of positive and negative deviations in the estimated composition profiles and estimated pure-component spectrum. As the noise level increased, the upper bound for the middle peak increased significantly with a corresponding increase in the correlation of its estimated spectrum with the spectra of the first and third peaks. There was also a corresponding decrease in the width of the estimated profiles of the first and third peaks. As the noise level increased there was also a significant decrease in the lower bound for the middle peak. The effect of reduced chromatographic resolution on the upper and lower bounds of the middle peak was similar to the effect of increased measurement noise. As the height of the middle peak was reduced, the error in the absolute peak area at the upper and lower bounds stayed nearly constant, whereas the relative peak area increased due to the overall reduction in peak area.

3.2. Reaction Composition Profiles. The batch reaction described in the Experimental Section was analyzed by the iterative target transformation factor analysis method of curve resolution[2] using three factors based on inspection of Malinowski's RE function. The selection of three factors was confirmed by performing ITTFA analysis with $k = 2$, 3, and 4 factors. The standard ITTFA procedure converged smoothly when three factors were used but failed to converge to meaningful results with two or four factors.

The ITTFA solution shown in Figure 10 was used as the initial starting point for the computation of the range of feasible profiles shown in Figure 11. The consumption of trityl chloride, the formation of a reactive intermediate which is assumed to be trityl

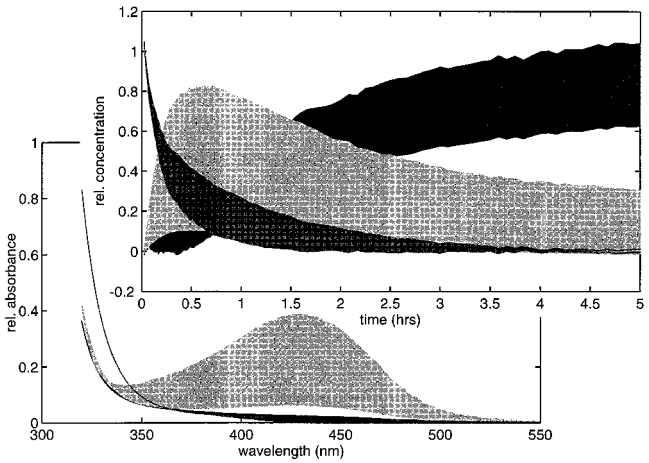

Figure 11. Range of feasible profiles for the SMCR analysis of the tritylation of a nucleoside.

cation,[32] and the formation of product can each be seen in Figure 10. The spectrum of the starting material was taken from the first-recorded spectrum following the addition of trityl chloride to the reaction mixture and was not allowed to change during the search for the boundary conditions. In addition, mass balance was assumed; consequently, the composition profiles were scaled so they summed to a constant. As before, the estimated spectra were scaled so that the maximum absorbance matched the maximum absorbance of the original (starting) spectra. In this problem the maximum absorbance values were found at 320 nm.

The estimated spectrum of the intermediate shows a small absorption band at 430 nm while the estimated spectrum for the product shows a large decrease and change in shape of the absorption band below 320 nm. These changes can also be seen in the original mixture spectra. The absorbance below 350 nm decreases as the reaction progresses, while the band at 430 nm grows in and then disappears.

SMCR profiles from batch reactions will seldom meet the criteria needed for unique solutions, and thus, will usually have very wide ranges of feasible solutions. Unlike the chromatography data sets described above, the location of zero-component regions cannot be deduced in batch reactions by analysis of local rank at the end of the reaction where steady state conditions are achieved. At the end of the reaction there may be an equilibrium mixture of intermediate and product. This leads to fairly wide composition profiles at the end of the batch as can be seen in Figure 11. At the end of the batch where the concentration of starting material is approximately zero, the shapes of the product and intermediate profiles are related to each other because of the mass balance constraints employed. The lower boundary of the product profile occurs when the upper bound of the intermediate profile is reached. Conversely, the upper boundary of the product profile occurs when the lower boundary of the intermediate profile is reached. In the other parts of the reaction the relationship between the shapes of all three profiles is not as easily visualized. In the batch reaction studied here, the spectra of the starting material and product are highly correlated to each other which causes the range of feasible solutions to be very wide, even though the initial spectrum of the pure reactant was known.

(32) Pocker, Y.; Buchholz, R. F. *J. Am. Chem. Soc.* **1970**, *92*, 2075−2084.

4. CONCLUSION

In many circumstances, SMCR methods such as ITTFA may be the only way estimates of composition profiles can be obtained. The advantage of SMCR is that it is suitable for the study of unresolved mixtures when no prior information is available about the nature and composition of these mixtures. For example, reference data obtained by HPLC analysis of quenched reaction mixtures will usually contain destroyed reactive intermediates and will not accurately reflect the composition of the batch mixture at the time of sampling. In some applications, kinetic modeling of the batch measurements may be used; however, this approach cannot be applied in processes in which reagents are added at a slow rate over time.

The new method for computing the feasible bands of SMCR results works well in a wide number of cases; however, there are circumstances which can cause the computation to fail. Small uncompensated negative baseline offsets in the measured mixture spectra will sometimes cause the optimization code to become trapped at an unfeasible solution point. If the negative offset is large enough, no feasible solution may exist. This problem is easily corrected by adding a constant offset to the data matrix of mixture spectra. At each iteration in the boundary search procedure the constraints change slightly because they are derived from the new solution point. Occasionally this causes the optimization code to become trapped outside of the boundary of the feasible space. Straightforward revisions to the algorithm are planned that will allow it to return to the previous feasible point. Finally, redundant solutions are sometimes encountered when the feasible solutions for different components overlap. In this case the optimization code will stray into the local minimum of the overlapping component.[26]

ACKNOWLEDGMENT

The Measurement and Control Engineering Center at the University of Tennessee is acknowledged for financial support of this research. Partial funding from this research was also provided by the National Science Foundation through Grant EEC-9815569.

Received for review June 15, 1999. Accepted September 8, 1999.

AC990648Y