The Use of Matrix Methods in the Modeling of Spectroscopic Data Sets

Eric R. Henry
Laboratory of Chemical Physics, National Institutes of Health, Bethesda, Maryland 20892-0520

ABSTRACT We describe a general approach to the model-based analysis of sets of spectroscopic data that is built upon the techniques of matrix analysis. A model hypothesis may often be expressed by writing a matrix of measured spectra as the product of a matrix of spectra of individual molecular species and a matrix of corresponding species populations as a function of experimental conditions. The modeling procedure then requires the simultaneous determination of a set of species spectra and a set of model parameters (from which the populations are derived), such that this product yields an optimal description of the measured spectra. This procedure may be implemented as an optimization problem in the space of the (possibly nonlinear) model parameters alone, coupled with the efficient solution of a corollary linear optimization problem using matrix decomposition methods to obtain a set of species spectra corresponding to any set of model parameters. Known species spectra, as well as other information and assumptions about spectral shapes, may be incorporated into this general framework, using parametrized analytical functional forms and basis-set techniques. The method by which assumed relationships between global features (e.g., peak positions) of different species spectra may be enforced in the modeling without otherwise specifying the shapes of the spectra will be shown. We also consider the effect of measurement errors on this approach and suggest extensions of the matrix-based least-squares procedures applicable to situations in which measurement errors may not be assumed to be normally distributed. A generalized analysis procedure is introduced for cases in which the species spectra vary with experimental conditions.

INTRODUCTION

The understanding of biochemical processes often relies on the use of models that describe the underlying mechanisms in terms of interconversions among hypothetical molecular species. The study of these processes using spectroscopic methods then requires that the observed spectra be related to the populations and intrinsic spectra of these model species. A common example of such modeling appears in the analysis of spectrophotometric titrations. In a simple case, the binding of a ligand X to a molecule B—i.e., the interconversion between two molecular species B $\leftrightarrow$ BX—is monitored by measuring spectra $A(\lambda,[X])$ as functions of wavelength λ, and the concentration of ligand [X]. If the spectra of the two pure species, $S_B(\lambda)$ and $S_{BX}(\lambda)$, are known a priori, then the species concentrations $c_B([X])$ and $c_{BX}([X])$ are determined in a "model-free" fashion by performing a least-squares fit

$$A(\lambda, [X]) \approx c_B([X])S_B(\lambda) + c_{BX}([X])S_{BX}(\lambda) \qquad (1)$$

independently at each value of [X]. If, on the other hand, the species spectra are not known, it is still possible to determine both the species concentrations and the species spectra from a set of spectra $\{A\}$ measured for different values of [X], provided some *model* is assumed for the dependence of the species concentrations on [X]. For this titration, we might predict the following such dependence:

$$c_B([X]) = \frac{c_0}{1 + K[X]}, \quad c_{BX}([X]) = \frac{c_0 K[X]}{1 + K[X]} \qquad (2)$$

where c_0 is the (known) total concentration of both species and K is an equilibrium constant that remains to be determined. We would then attempt to estimate the parameter K and the species spectra S_B and S_{BX} by performing a set of simultaneous least-squares fits using the measured spectral amplitudes at different wavelengths λ_i as a function of [X]:

$$A(\lambda_i, [X]) \approx \frac{c_0(S_B(\lambda_i) + K[X]S_{BX}(\lambda_i))}{1 + K[X]} \qquad (3)$$

Here the amplitudes $S_B(\lambda_i)$ and $S_{BX}(\lambda_i)$ of the species spectra at wavelength λ_i, as well as the equilibrium constant K, are treated as adjustable parameters, and the value of K is required to be consistent among all of the single-wavelength fits.

The solution of such a *spectroscopic modeling problem* generally requires some procedure to vary both the spectra of the species and the adjustable parameters in the model, from which the populations of the species as a function of experimental conditions are computed, to reproduce the observed variations of the measured spectra with experimental conditions. Because the model parameter K in the above example is independent of wavelength, the set of least-squares fits represented by Eq. 3 uses the same model species populations at each wavelength. By arranging the measured spectra for different values of [X] as the columns of a matrix, we can separate out the unknown species spectra from the model species populations. The set of

Received for publication 12 August 1996 and in final form 8 November 1996.

Address reprint requests to Dr. Eric R. Henry, National Institutes of Health, Bldg. 5, Room 110, Bethesda, MD 20892-0520. Tel.: 301-496-6031; Fax: 301-496-0825; E-mail: eric@helix.nih.gov.

0006-3495/97/02/652/22 $2.00

simultaneous least-squares fits in Eq. 3 may then be written

$$\begin{pmatrix} A(\lambda_1, [X]_1) & \dots & A(\lambda_1, [X]_n) \\ \vdots & \vdots & \vdots \\ A(\lambda_m, [X]_1) & \dots & A(\lambda_m, [X]_n) \end{pmatrix} \approx \begin{pmatrix} S_B(\lambda_1) & S_{BX}(\lambda_1) \\ \vdots & \vdots \\ S_B(\lambda_m) & S_{BX}(\lambda_m) \end{pmatrix} \begin{pmatrix} p_B(K, [X]_1) & \dots & p_B(K, [X]_n) \\ p_{BX}(K, [X]_1) & \dots & p_{BX}(K, [X]_n) \end{pmatrix} \quad (4)$$

where $p_i(K, [X])$ is the population of species i predicted by the model for equilibrium constant K and ligand concentration [X]. Being able to write the problem in matrix form is more than a notational convenience. Procedures for solving such modeling problems may, in many cases, exploit this matrix organization of spectra and populations to considerable advantage, in terms of both algorithmic efficiency and the compactness of the stored representation of the data (Henry and Hofrichter, 1992). In this paper we describe an approach to the solution of the spectroscopic modeling problem that is built upon the concepts and procedures of matrix analysis. It should be emphasized that a properly constructed matrix-based modeling procedure should be equivalent in all meaningful ways to a "global analysis" approach (Beechem, 1992) as embodied in Eq. 3, which does not rely upon a matrix organization of the data. It is hoped that the efficiencies offered by a matrix viewpoint will make demanding analyses of larger data sets more practical.

Some of the ideas and techniques presented here are not new. Our goal in this paper is to integrate existing ideas with certain new techniques and perspectives, which have been developed in treating specific modeling problems, into a "handbook" that might prove useful to others. The procedures presented here have been applied most recently to the analysis of time-resolved optical absorption spectra of photolyzed hemoglobin using complex kinetic models (Henry et al., 1997). However, the present discussion has been kept as general as possible and makes no direct assumptions about the types of spectra and models being considered.

The organization of the paper is as follows: The spectroscopic modeling problem in matrix form and outline of the general approach to solving it are introduced in Definition and Features and Direct versus SVD-Based Analysis sections. In Incorporating Known Species Spectra we discuss methods to incorporate known or assumed features of species spectra into the matrix-based modeling a priori. More careful consideration is then given in Spectroscopic Modeling to the role of measurement errors in defining the spectroscopic modeling problem, both to support the use of efficient least-squares algorithms wherever possible and to suggest possible efficient methods of solution in cases where least-squares optimization is not appropriate. Finally, a generalization of the original matrix-based specification of the spectroscopic modeling problem is presented in A Generalized Spectroscopic Modeling Problem, which with some assumptions allows the species spectra (rather than the populations alone) to vary with experimental conditions. Detailed mathematical developments in support of the discussion are presented in three appendices at the end of the article.

DEFINITION AND FEATURES OF THE SPECTROSCOPIC MODELING PROBLEM

We consider a general set of spectra $\{A_i\}$ (e.g., optical absorption or Raman spectra) which have been obtained for a system under various sets of measurement conditions, and assume that these spectra are functions of a single spectroscopic parameter λ (e.g., wavelength). The set of spectra may be written in the compact form $\{A_i\} = \{A(\lambda, \{x_i\})\}$, where $\{x_i\}$ is the set of measurement conditions, one for each spectrum. When several experimental conditions are being varied, each x_i may be a set of values of distinct experimental parameters, e.g., $x_i = \{\text{time}(i), \text{pH}(i), \text{temperature}(i), \dots\}$. In the general modeling problem, a set of model "species" is postulated, as well as some prescription $p_j = M_j(\{x_i\}, \{\xi_k\})$ for computing the population of species j as a function of the experimental parameters $\{x_i\}$ and some set of model parameters $\{\xi_k\}$. Associated with each species j is a spectrum $S_j(\lambda)$, which is assumed not to vary with experimental conditions. (This assumption will be relaxed somewhat in A Generalized Spectroscopic Modeling Problem.) Measured spectra must be written as sums of the species spectra weighted by the respective species populations, that is

$$A(\lambda, \{x_i\}) \approx \sum_{j=1}^{N} S_j(\lambda) M_j(\{x_i\}, \{\xi_k\}) \quad (5)$$

To proceed, we arrange the set of scalars $\{A(\lambda, \{x_i\})\}$, measured for a fixed set of n_λ spectroscopic parameter values λ, in the form of a matrix $\mathbf{A}$. Each column of $\mathbf{A}$ is a single measured spectrum, so that each row of $\mathbf{A}$ corresponds to a single fixed λ, and each column is identified with one of n_p sets of experimental parameters x_i:

$$\mathbf{A} = \begin{pmatrix} A(\lambda_1, \{x_1\}) & A(\lambda_1, \{x_2\}) & \dots & A(\lambda_1, \{x_{n_p}\}) \\ A(\lambda_2, \{x_1\}) & A(\lambda_2, \{x_2\}) & \dots & A(\lambda_2, \{x_{n_p}\}) \\ \dots & \dots & \dots & \dots \\ A(\lambda_{n_\lambda}, \{x_1\}) & A(\lambda_{n_\lambda}, \{x_2\}) & \dots & A(\lambda_{n_\lambda}, \{x_{n_p}\}) \end{pmatrix} \quad (6)$$

In a similar fashion we define a $n_\lambda \times n_s$ matrix $\mathbf{S}$ of species spectra, each column of which contains one of the n_s (possibly unknown) species spectra $\{S_j\}$, and a $n_s \times n_p$ population matrix function $\mathbf{M}(\{(\xi_k\})$, each row of which consists of the populations of a single species computed for all n_p experimental parameter sets. The statement (Eq. 5) of the

modeling hypothesis may then be written

$$\mathbf{A} \approx \mathbf{SM}(\{\xi_k\}) \tag{7}$$

The goal of the modeling is to determine sets of model parameters $\{\xi_k\}$ and a matrix **S** of species spectra such that the relation in Eq. 7 is optimal. In this article we discuss general features of the solution of this problem, which do not rely upon specific forms or methods of construction of the matrices **S** and **M**.

Obtaining optimal sets of model parameters $\{\xi_k\}$

For the purposes of this article, the optimization implied by Eq. 7 will most often be taken as the minimization of a sum of unweighted squared residuals, and the relation "≈" will imply such simple least-squares optimization unless otherwise noted. As discussed below, the least-squares minimization in Eq. 7 may be decomposed into separate minimization problems in the space of model parameters $\{\xi_k\}$ and in the space of spectral amplitudes **S**. The solution of the latter problem by efficient matrix methods is a subject of the next section and Appendix C; I consider here only the optimization in the space of (possibly nonlinear) model parameters $\{\xi_k\}$.

A detailed discussion of least-squares minimization algorithms is beyond the scope of this article. However, the basic issue of which algorithm to use should be addressed briefly. Specialized search procedures that exploit the structure of the problem of minimizing sums of squares will generally be more efficient than the naive use of general-purpose minimization algorithms to treat the least-squares problem as a special case (Press et al., 1993). We have found the Marquardt-Levenberg algorithm (Marquardt, 1963) modified to include possible linear equality and inequality constraints (Shrager, 1970) to be quite satisfactory. This algorithm requires derivatives of function values with respect to the adjustable parameters, which may be computed analytically or numerically. For the types of functions encountered in modeling studies, analytical expressions for derivatives are often not available, but when these functions are sufficiently "well-behaved," a straightforward evaluation of the derivatives by finite differences is possible (Brown and Dennis, 1972). When this is not the case, any of the available derivative-free general-purpose minimization algorithms may be used, albeit usually with some loss of performance.

Modeling situations for which least-squares optimization is not appropriate may often still be cast in the form of an iteratively reweighted least-squares optimization (see Appendix C). Alternatively, a general-purpose algorithm may be used to locate parameter sets that minimize the chosen functional. If derivatives of the functional with respect to the parameters are easily computed, then one of the available conjugate-gradient or variable-metric algorithms may be used (Press et al., 1993). Otherwise, Brent's derivative-free method (Brent, 1973) works well for functionals that satisfy certain continuity requirements, and the Nelder-Mead simplex method (Nelder and Mead, 1965; Press et al., 1993) is available for those that do not.

A common feature of many optimization problems is the possible existence of multiple minima. Typical search procedures make use of local properties of the function to be minimized to generate a search strategy; one consequence is that, depending on the starting point, the procedure may become "trapped" in some minimum other than the global minimum. If the number of adjustable parameters is small enough, multiple minima may be located or bracketed using some form of exhaustive search of the accessible parameter space. In more complex cases, it may still be possible to compile a useful, if less exhaustive, survey of parameter space by means of a simulated-annealing procedure combined with a Monte Carlo search (Kirkpatrick et al., 1983; Szu and Hartley, 1987). In either case, low-lying regions of the functional hypersurface in parameter space are identified, which may then serve as starting points for a Marquardt-Levenberg or other minimization.

Reduction in the size of the matrix of species spectra

In the most general case, each of the model species is assigned a unique spectrum or column of **S**. However, in all but the simplest modeling situations, or in the absence of prior spectral assignments for most of the species, the quality of the experimental data (as well as the descriptive power of the model itself) is unlikely to support independently adjustable spectra for all species for which spectra are not known. Therefore, it is necessary to group the species in such a way that all species within a single group are assumed to have the same spectrum, which is distinct from that of other groups. The assumed grouping then becomes an important feature of the model itself, and transforms the relation in Eq. 7 in a simple way: The matrix $\mathbf{S}_m$ now consists of just the minimal set of n_m possible distinct spectra, and the appropriate sets of rows of the population matrix **M** computed from the model must be summed to produce the rows of a new spectral population matrix $\mathbf{M}_m$, which contain total species populations corresponding to each of the spectra in $\mathbf{S}_m$. The reduced representation of the data matrix then takes the form

$$\mathbf{A} \approx \mathbf{S}_m\mathbf{M}_m \tag{8}$$

$$\mathbf{M}_m = \mathbf{GM}$$

$$G_{ij} = \begin{cases} 1 & \text{if species } j \in \text{ group } i \\ 0 & \text{otherwise} \end{cases}$$

The matrix **G** is $n_m \times n_s$; $\mathbf{M}_m$ is therefore $n_m \times n_p$, with rows of populations corresponding to the spectra that comprise the columns of $\mathbf{S}_m$. This reduced representation is formally equivalent to Eq. 7, so the latter will be used in subsequent discussion with the understanding that it implic-

itly reflects this grouping of spectrally distinct species. We will also use n_s to indicate the number of species spectra (columns of **S**) in the representation, independent of the detailed significance of this matrix.

"DIRECT" VERSUS SVD-BASED ANALYSIS OF THE DATA MATRIX

A straightforward least-squares interpretation of Eq. 7 translates to the requirement that the residual

$$\delta^2 \equiv \sum_{i=1}^{n_\lambda} \sum_{j=1}^{n_p} \left(A_{ij} - \sum_{l=1}^{n_s} S_{il} M_{lj}(\{\xi_k\}) \right)^2 \tag{9}$$

be a minimum with respect to some chosen set of adjustable spectral amplitudes (S_{il}) and/or model parameters (ξ_k). One approach is to minimize this expression directly using some appropriate multiparameter minimization algorithm, while treating each of the unknown spectral amplitudes and model parameters as nominally independent adjustable parameters. The resulting large number of adjustable parameters and its possible impact on the performance (and even the choice) of the minimization algorithm is the principal disadvantage of this approach. On the other hand, the direct minimization of δ^2 in Eq. 9 also retains a complete flexibility in specifying the spectroscopic features of the problem, which is often reduced or lost in more efficient or "elegant" procedures. For example, the incorporation of known species spectra, or even parts of spectra, into the analysis is very straightforward, as is the use of explicit constraints on spectral amplitudes for any subset of the spectroscopic indices λ.

Of course, the complete flexibility of this direct analysis may not be especially helpful for many problems, but procedures that retain some of its flexibility while eliminating the more serious of its inefficiencies are often desirable. A particular burden is the need to carry all of the spectral amplitudes S_{il} as independently adjustable parameters. To eliminate this requirement, we note that the individual residuals within the parentheses in Eq. 9 are explicitly linear in the spectral amplitudes, although generally nonlinear in the parameters $\{\xi_k\}$. Thus, with the matrix **M** fixed by specifying values for parameters in the set $\{\xi_k\}$, a corollary set of spectral amplitudes may be determined by solving the *linear* least-squares problem

$$\mathbf{A} \approx \mathbf{SM} \tag{10}$$

It is well known (see, for example, Lawson and Hanson (1974)) that this problem may always be solved using a complete orthogonal decomposition of the matrix **M**. This decomposition produces orthogonal matrices **H** and **K** such that $\mathbf{M} = \mathbf{QRK}^T$, where the matrix **R** has the block form

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{11}$$

and the square matrix $\mathbf{R}_{11}$ is full-rank (and hence invertible; see Note 1). The overall least-squares problem represented by Eq. 9 may then be solved by using an iterative minimization algorithm to vary only the adjustable model parameters $\{\xi_k\}$. At each iteration, the matrix **M** computed from the current values of these parameters is used to solve the linear least-squares problem (Eq. 10) to produce a set of spectral amplitudes that are "optimally consistent" (i.e., produce the smallest residuals) with the current model parameters. In other words, the unknown spectral amplitudes are tied to the model parameters through the linear least-squares requirement (Eq. 10) and do not have to be explicitly varied in the search for a minimum. This technique is in the spirit of so-called variable projection algorithms (Golub and Pereyra, 1973), in which the adjustable parameters are divided into sets of linear and nonlinear parameters, and only the nonlinear parameters are varied explicitly during the minimization. The most mathematically sophisticated implementations exploit this separation to the extent that the linear parameters need not even be computed during the minimization with respect to the nonlinear parameters. This convenience comes, however, at the expense of some algorithmic complexity and specialization. In our discussion we will take the mathematical "low road" by including the direct solution of the corollary linear least-squares problem as an explicit step in the overall optimization procedure.

The analysis of the modeling problem in Eq. 7 may often be facilitated by simplifying the data matrix **A** in some way that does not compromise its essential information content. Beyond the obvious simplifications designed merely to reduce the size of the data matrix (e.g., truncation and/or resampling of the data on a coarser grid of spectroscopic parameters λ) are more sophisticated rank-reduction procedures that attempt to extract minimal descriptions of the "meaningful" content of a data set. Methods based on singular value decomposition (SVD) have become increasingly popular in recent years. The application of SVD to spectroscopic data sets has been reviewed in detail elsewhere (Henry and Hofrichter, 1992). Briefly, the SVD of an arbitrary $m \times n$ ($m > n$) matrix **A** expresses it as the product

$$\mathbf{A} = \mathbf{U\Sigma V}^T \tag{12}$$

where **U** is a $m \times n$ matrix and **V** is a $n \times n$ matrix, the columns of each comprising orthonormal sets of vectors (i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_n$, the $n \times n$ identity matrix), and Σ is a $n \times n$ diagonal matrix with nonnegative diagonal elements σ_i called the *singular values* of **A**. When **A** is a $n_\lambda \times n_p$ matrix of spectroscopic data arranged as in Eq. 6, the columns of **U** are themselves spectra of the same type, the normalized *basis spectra* of **A**. The contribution of a specified basis spectrum to each of the measured spectra (columns of **A**) is given by the elements of the corresponding row of $\mathbf{V}^T$ (column of **V**), scaled by the corresponding singular value. The σ_i, along with the corresponding columns of **U** and **V**, may always be ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. With this ordering, the rank r of **A** is the index of the last (and smallest) nonzero singular value. Moreover, for any $q \leq r$, the truncated matrices $\mathbf{U}_q$ and $\mathbf{V}_q$,

consisting of the first q columns of $\mathbf{U}$ and $\mathbf{V}$, respectively, and Σ_q, the diagonal matrix containing the q largest singular values, provide the best possible least-squares approximation of rank q of the matrix $\mathbf{A}$, $\mathbf{A} \cong \mathbf{A}_q = \mathbf{U}_q\Sigma_q\mathbf{V}_q^T$. The residual of this approximation is given by

$$|\mathbf{A} - \mathbf{A}_q|^2 = \sum_{i=q+1}^{t} \sigma_i^2 \tag{13}$$

This last property helps motivate an overall strategy for the use of SVD in reducing spectroscopic data sets: By considering the distribution of singular values, a "truncated SVD" consisting of a (generally small) subset of the basis spectra and corresponding amplitude vectors is selected that represents the original data matrix within some acceptable tolerance. Additional processing of these retained vectors may be done, including linear combination and further screening based on signal-to-noise considerations. The final result is a pair of matrices $\mathbf{U}'$ and $\mathbf{V}'$ of basis spectra and vectors of amplitude-versus-experimental parameter (into which singular values have been absorbed), respectively, all of which are of acceptable quality for further analysis, such that the product $\mathbf{A}' \equiv \mathbf{U}'\mathbf{V}'^T$ remains close to $\mathbf{A}$. (The reader is referred to Henry and Hofrichter (1992) for a detailed discussion of this process.) Of course, any procedure that satisfactorily produces such a minimal product representation would serve, but the SVD provides an especially convenient mathematical framework for the analysis.

It must be emphasized that the matrices $\mathbf{U}'$ and $\mathbf{V}'$ are still purely mathematical objects at this stage, but they provide a useful starting point for the model-based synthesis of species spectra and populations. To proceed, we note that within this reduced representation (of rank q) of the data set, all measured spectra are linear combinations of the columns of $\mathbf{U}'$. Moreover, a basic assumption of the modeling is that all spectra are linear combinations of the species spectra in $\mathbf{S}$. We therefore make the *ansatz* that the species spectra are linear combinations of the columns of $\mathbf{U}'$, i.e., $\mathbf{S} = \mathbf{U}'\mathbf{C}$ for some matrix of coefficients $\mathbf{C}$. We then have a representation of the data matrix in the form

$$\mathbf{A} \simeq \mathbf{A}' = \mathbf{U}'\mathbf{V}'^T \approx \mathbf{SM} = \mathbf{U}'\mathbf{CM} \tag{14}$$

Therefore, within our chosen basis $\mathbf{U}'$ for all spectra, we can re-cast the least-squares problem in Eq. 10 as a search for a set of model parameters $\{\xi_k\}$ and a matrix of coefficients $\mathbf{C}$ such that the relation

$$\mathbf{V}'^T \approx \mathbf{CM}(\{\xi_k\}) \tag{15}$$

is optimally satisfied, thereby removing direct reference to the species spectra from the problem (see Note 2). As in the direct method described above, the explicit linearity in the coefficients $\mathbf{C}$ allows us to employ a minimization procedure in which only the model parameters $\{\xi_k\}$ are directly varied, with a matrix $\mathbf{C}$ computed as needed from the current model parameters by solving the linear least-squares problem involving $\mathbf{V}'^T$ and the current matrix $\mathbf{M}$.

The formal similarity between the least-squares problems in Eq. 15 and 10 is apparent, and the same types of solution procedures may be applied to both. The obvious advantage of the SVD-based analysis is that the matrices involved are generally much smaller: $\mathbf{V}'^T$ is $q \times n_p$ and $\mathbf{C}$ is $q \times n_s$, whereas $\mathbf{A}$ and $\mathbf{S}$ are $n_\lambda \times n_p$ and $n_\lambda \times n_s$, respectively. In many modeling situations the consequent reduction in computational effort required may be sufficient justification for choosing this method. However, this reduction in the size of the problem comes at the expense of restricting all spectra to be linear combinations of the columns of $\mathbf{U}'$. As a result, at most q linearly independent species spectra may be constructed, and any model requiring $n_s > q$ is likely to encounter mathematical difficulties in the solution of the problem in Eq. 15. Thus, the number of amplitude vectors with sufficient information content to contribute to the fitting operation in Eq. 15 imposes a practical limit on the number n_s of spectroscopically distinct species in any model to be applied to such a data set.

The choice between a direct and an SVD-based treatment of the data during modeling largely depends upon whether or not the restricted spectral basis set places unacceptable constraints on some scientific or procedural aspect of the model. In addition to the above-mentioned possible difficulties related to the number of spectroscopic species in the model, this restriction may also limit the ability to incorporate known features of species spectra into the model to the desired extent. In the absence of such limitations the SVD-based approach appears to be the natural choice based on efficiency considerations. However, if the actual solution of the model to produce the population matrix $\mathbf{M}(\{\xi_k\})$—which is a common feature of both methods—is significantly more expensive than the matrix manipulations required to solve the linear least-squares problems in either method, then the relatively small added cost associated with the direct method may not be prohibitive.

INCORPORATING KNOWN SPECIES SPECTRA AND OTHER SPECTROSCOPIC CONSTRAINTS INTO THE MODEL

The results of modeling efforts tend to be more satisfactory when the known and desired features of the system are introduced a priori to the greatest possible extent. In particular, input of any available information or assumptions about one or more of the species spectra may help accelerate and focus the search for descriptive sets of model parameters. The basic method for incorporating such information in the modeling is the same whatever its level of detail, which may range from complete measured spectra from other sources to general qualitative shape information: The "known" species spectra are assembled in an appropriate way, which may involve direct importation of spectra from other sources, or the evaluation of some parametrized functional form, which reflects known or assumed shape information. In the direct method these spectra are represented

by vectors of actual amplitudes $S(\{\lambda_i\})$ for each of the chosen values of the spectroscopic index λ; in the SVD-based method, the spectra are represented by vectors of coefficients for linear combinations of the chosen basis spectra (columns of $\mathbf{U}'$; see below). The contribution of each such measured or synthetic spectrum to the simulated data set is computed by multiplying its representation vector by the current set of populations for that species taken from its assigned row of the current population matrix $\mathbf{M}$. The resulting contribution is a $n_\lambda \times n_p$ matrix, each column of which is the spectral contribution from that species for a single set of experimental parameters $\{x_i\}$. The sum of all of these "pre-set" contributions is then subtracted from the full data matrix, and this corrected matrix is used in the linear least-squares computation (analogous to that in Eq. 10) with the remaining rows of the population matrix, if any, to produce the remaining previously unspecified spectra.

Within a direct treatment of the data, we proceed as follows: We divide our set of n_s species into sets of m pre-set species with indices p_i and spectra P_i and $n_s - m$ "free" species with indices f_i. At any point where the species spectra are required, the pre-set spectra $P_i(\{\zeta_l\})$ are first assembled (where we explicitly exhibit the possible dependence of these spectra on adjustable parameters $\{\zeta_l\}$), and the following matrices are constructed from these spectra and from the current population matrix $\mathbf{M}(\{\xi_k\})$:

$$\mathbf{P} = (P_1 \quad P_2 \quad \ldots \quad P_m)$$

$$\mathbf{M}_p = \begin{pmatrix} M_{p_1}(\{\xi_k\}) \\ M_{p_2}(\{\xi_k\}) \\ \ldots \\ M_{p_m}(\{\xi_k\}) \end{pmatrix}$$

$$\mathbf{M}_f = \begin{pmatrix} M_{f_1}(\{\xi_k\}) \\ M_{f_2}(\{\xi_k\}) \\ \ldots \\ M_{f_{n_s-m}}(\{\xi_k\}) \end{pmatrix} \tag{16}$$

These matrices are used to compute the residual free data matrix and solve the linear least-squares problem

$$\mathbf{A}_f \equiv \mathbf{A} - \mathbf{P}\mathbf{M}_p \approx \mathbf{F}\mathbf{M}_f \tag{17}$$

for the matrix $\mathbf{F}$ of the remaining free spectra.

The procedure within the SVD-based treatment is similar. The pre-set spectra are computed as before, but must then be expressed as linear combinations of the chosen basis spectra of $\mathbf{U}'$. Rather than a matrix $\mathbf{P}$ of pre-set spectra, we now have the matrix

$$\mathbf{C}_p = \begin{pmatrix} C_{1,1} & C_{1,2} & \ldots & C_{1,m} \\ C_{2,1} & C_{2,2} & \ldots & C_{2,m} \\ \ldots & \ldots & \ldots & \ldots \\ C_{q,1} & C_{q,2} & \ldots & C_{q,m} \end{pmatrix} \tag{18}$$

of coefficients describing each of the m pre-set spectra P_i as linear combinations of the q columns of $\mathbf{U}'$. We then compute the residual "free" $\mathbf{V}'^{\mathrm{T}}$ and solve the linear least-squares problem

$$\mathbf{V}_f'^{\mathrm{T}} \equiv \mathbf{V}'^{\mathrm{T}} - \mathbf{C}_p\mathbf{M}_p \approx \mathbf{C}_f\mathbf{M}_f \tag{19}$$

for the remaining free coefficients.

Although formally equivalent to the procedure in the direct treatment, the SVD-based method of incorporating known spectral information into the modeling requires some additional compromises. The obvious difficulty lies again with the restricted spectral basis set that is imposed. The use of the SVD and subsequent processing is intended to produce sets of basis spectra and corresponding amplitudes which are able to account for all meaningful spectroscopic information contained in the data set. However, even the most careful efforts in this direction may be hindered by the intrinsic mathematical properties of the SVD. For example, in the presence of noise a small signal component may be spread over several basis spectra to maintain orthogonality of the different components. While procedures are available that can partially undo this effect (Henry and Hofrichter, 1992), there is no way to ensure complete separation of signal from noise in the analysis. Of course, even with a perfectly determined set of basis spectra, it is not certain that such a set would be able to perfectly describe some imported spectrum that might have been measured on a different instrument (or even on a different day!) than that for the current data set. The ability of a chosen basis set to describe a spectrum computed from some parametrized functional form can also not be guaranteed. This procedure should therefore be evaluated case-by-case, based on whether the best available basis set is able to represent all of the relevant spectroscopic content of interest, both intrinsic to the data set and imported from other sources, to within acceptable tolerances.

With these caveats, we now discuss possible methods for programming spectroscopic information into the framework just outlined.

Measured spectra from other sources

There are clear advantages to being able to assume that one or more of the spectroscopically distinct species in the model coincides with a species for which the spectrum is available from other sources. For example, if the initial or final state in some reaction being probed by kinetic methods is expected to be a pure model species, then the corresponding spectrum measured by equilibrium techniques may be usable in the modeling. Construction of the associated pre-set spectrum P_i may require some additional processing of the extrinsic spectrum to adjust for differences between measurement conditions under which it was obtained and conditions relevant to the data set being modeled. Obvious corrections of an instrumental nature include the use of interpolation to map measurements made for one set of

spectroscopic indices λ onto another such set, convolution or deconvolution to correct for differences in instrumental resolution, and scaling or other simple transformations to correct for other differences in experimental conditions. If the nature of the required correction is known, but certain quantitative features of the correction are not known with sufficient confidence, then it is often helpful to express the correction operation in terms of some adjustable parameter(s), the determination of which is integrated into the model fitting procedure along with the determination of the other model parameters. A simple example is the use of an adjustable scale factor in cases for which the exact shape of the spectrum is known, but the overall amplitude appropriate for the current modeling situation is not (Henry et al., 1997).

Analytical functional forms and other parametrized spectra

In certain types of spectroscopy the underlying properties of the system being studied may dictate some "natural" shape for spectra of individual species (e.g., Lorentzian or Gaussian lineshapes). It is then only necessary to identify those features of this shape that distinguish the spectra of different species (e.g., amplitude, peak position, linewidth) and provide appropriately parametrized functions to compute the pre-set spectra $P_i = f_i(\lambda, \{\zeta_l\})$ when required (see Note 3). Some or all of the spectroscopic parameters $\{\zeta_l\}$ may be allowed to vary as part of the modeling procedure, optionally under constraints that reflect additional knowledge or assumptions about the properties of the system. Unfortunately, in many cases the true spectral shapes are inadequately approximated by simple "natural" functions, and it is necessary to modify these functions in some way to improve the representation of spectral shapes. Possible modifications include:

a) expanding the parametrization of a "natural" function to introduce additional degrees of freedom. One typical example is the generalization of a Gaussian form through the introduction of a skewness parameter b (e.g., Hoff et al., 1994) via

$$\exp(-((\nu - \nu_0)/\Delta)^2) \rightarrow \exp(-(\ln(1 + b(\nu - \nu_0))/b)^2) \quad (20)$$

b) multiplying a natural function by, or convolving it with, some distorting function (e.g., Jones et al., 1993).

c) "warping" some natural function $f(\lambda)$ by performing a (generally nonlinear) transformation of the λ axis.

This method shares with that in example (a) the goal of generalizing a "natural" shape in some useful way. However, it addresses the problem from a different perspective, by transforming the λ axis by some function g: $\lambda \rightarrow g(\lambda)$—that is, $P_i(\lambda) = f(g(\lambda))$. The function g effectively changes the shape of the spectrum by distorting the λ scale. Perturbing spectral shapes by transforming the λ axis permits an analysis to depart from idealized spectral shapes while preserving important qualitative features of the distorted spectra. For example, modifying the shape of the spectrum by stretching and compressing the λ axis in a continuous fashion preserves the sequence (but may modify the positions) of extrema and zero-crossings. At the same time, with proper choices of the functions f and g it offers the ability to interpolate smoothly between shapes of very different character. Consider the Gaussian function

$$f(\lambda) = \exp(-((\lambda - \lambda_m)/\sigma)^2) \quad (21)$$

This may be rewritten as

$$f\left(\frac{\lambda - \lambda_m}{\sigma}\right) = \exp\left(-\left(g\left(\frac{\lambda - \lambda_m}{\sigma}\right)\right)^2\right)$$

$$g(x) = x \quad (22)$$

which may be viewed as the trivial "identity warping" of $f(\lambda)$. We now observe that by choosing $g(x) = (\ln(1 + x^2))^{1/2}$ we produce a new functional form

$$f\left(g\left(\frac{\lambda - \lambda_m}{\sigma}\right)\right) = \exp\left(-\left(\left(\ln\left(1 + \left(\frac{\lambda - \lambda_m}{\sigma}\right)^2\right)\right)^{1/2}\right)^2\right)$$

$$= \frac{1}{1 + (\lambda - \lambda_m)^2/\sigma^2} \quad (23)$$

which is Lorentzian. Furthermore, the one-parameter family of functions

$$g(x, b) = (\ln(1 + bx^2)/b)^{1/2} \quad (24)$$

can be shown to produce the Gaussian case for $b \rightarrow 0$ (i.e., $g(x, 0) = x$) and reduces to the Lorentzian case for $b = 1$. (This function bears some superficial resemblance to the function in the exponential on the right-hand side of Eq. 20, but the purpose and net effects of the two transformations are quite different.) By varying b between 0 and 1 we can smoothly interpolate between a Gaussian and a Lorentzian lineshape (and b may in fact exceed 1 to produce a "super-Lorentzian" shape). Such an interpolation is shown in Fig. 1*a*. In situations in which a spectral lineshape is only known to be "somewhere between" Gaussian and Lorentzian, it may be advantageous to incorporate a variably warped Gaussian of the form $f(g(\lambda, b))$ (or some further generalization thereof) and let the data set itself help to "decide" the most appropriate shape.

This example represents a reasonable and conservative approach to the use of such warping functions. We first choose a spectral shape function f to be a possible natural shape function for the system. We then select a parametrized "warping" function $g(x, \{\eta_i\})$ such that the family of functions $f(g(\lambda, \{\eta_i\}))$ generated by the warping operation includes this natural shape function as a special or limiting case. As a rule, a suitable warping function g should be

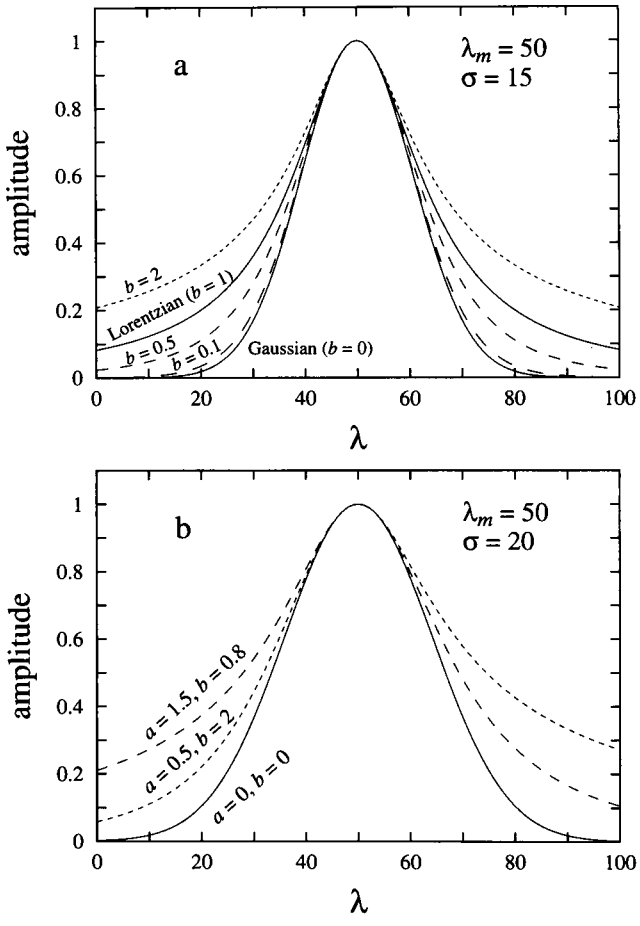

FIGURE 1 Illustration of the use of "warping" functions to perform functional interpolation between distinct spectral lineshapes. (*a*) Functional interpolation between Lorentzian and Gaussian lineshapes using

$$y = \exp\left(-\left(g\left(\frac{\lambda - \lambda_m}{\sigma}, b\right)\right)^2\right)$$

$$g(x, b) = (\ln(1 + bx^2)/b)^{1/2}$$

by varying the single parameter *b*. (*b*) Asymmetric interpolation between Lorentzian and Gaussian lineshapes using

$$y = \exp\left(-\left(g\left(\frac{\lambda - \lambda_m}{\sigma}, a, b\right)\right)^2\right)$$

$$g(x, a, b) = \begin{cases} (\ln(1 + bx^2)/b)^{1/2}, & x > 0 \\ (\ln(1 + ax^2)/a)^{1/2}, & x < 0 \end{cases}$$

by varying the two parameters *a* and *b*.

continuous and at least once differentiable, for both practical and aesthetic reasons. Other restrictions will generally depend on the type of lineshape function *f* with which it is being used. Within these straightforward mathematical constraints, it is possible to tailor the resulting family of spectral shapes to almost any desired degree, generally at the expense of additional parameters. As one final example, consider again the Gaussian spectral shape $f((\lambda - \lambda_m)/\sigma)$ discussed above. If we generalize the family of warping functions to assume the form

$$g(x, a, b) = \begin{cases} (\ln(1 + bx^2)/b)^{1/2}, & x > 0 \\ (\ln(1 + ax^2)/a)^{1/2}, & x < 0 \end{cases} \quad (25)$$

for distinct nonnegative constants *a* and *b*, then we can independently adjust the degree of "Lorentzian-ness" on the two sides of the peak at λ_m (Fig. 1 *b*; see Note 4). Note that λ_m remains the maximum value of the composite function $f(g(\lambda))$ for this entire family of warping functions, so that we retain direct parametric control of the peak position throughout the analysis.

d) replacing a natural function with some purely mathematical representation that has convenient computational properties. The use of parametrized functional forms to represent species spectra is intended to force these spectra to belong to some family of possible shapes. While the overall features of such families of spectra are most often motivated by "natural" properties of spectra for the system in question, there are cases in which family features are dictated by other considerations. One such scenario has already been discussed, namely the SVD-based treatment of species spectra. In that case, species spectra are constrained to belong to the family of spectra that are linear combinations of the basis spectra (columns of **U**′) provided. The basis spectra are in turn chosen based on mathematical properties, including signal-to-noise and their ability to represent the measured data set to within some tolerance. Many such choices of basis are possible, which are motivated more by mathematical requirements (e.g., continuity and smoothness) than by physical considerations. We briefly mention only one such representation—that using B-splines—not as a specific endorsement, but rather as a framework within which to illustrate how *any* assumed basis set for species spectra may be conveniently incorporated into the modeling analysis.

A detailed discussion of B-splines is outside the scope of this article. For present purposes, a spline $S_n(x)$ of order *n* on an interval $a \le x \le b$ is a piecewise polynomial function that may be described as follows: On each of a set of *p* sub-intervals $x_i \le x \le x_{i+1}$ ($a = x_1 \le x_2 \le \cdots \le x_{p+1} = b$), $S_n(x)$ is equal to a polynomial $p_i(x)$ of order *n* (degree $n - 1$)

$$p_i(x) = a_{0i} + a_{1i}x + \cdots + a_{n-1,i}x^{n-1} \quad (26)$$

The coefficients $a_{k,i}$ are chosen so that this piecewise polynomial function and its derivatives up to order $n - 2$ are continuous everywhere on the interval (a, b) and specifically at the chosen *knots* x_i which delimit the sub-intervals. Any such spline function may be written as a linear combination of basis splines (or B-splines) which are constructed for the specified order *n* and set of knots $\{x_i\}$. An example of such a basis is shown in Fig. 2. We require that

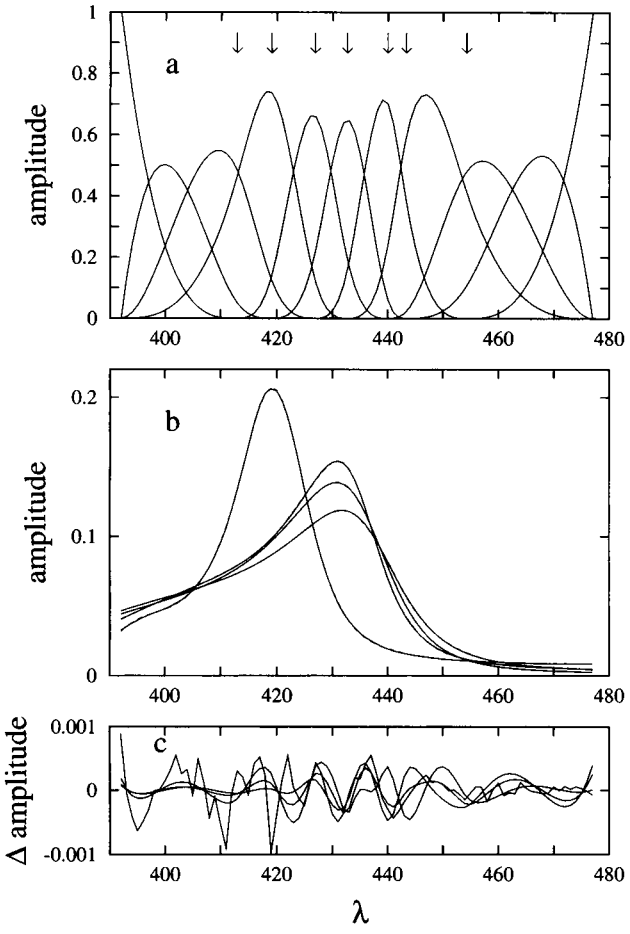

FIGURE 2 The use of B-spline basis functions to represent families of spectral shapes. In this example, a basis set of fourth-order (cubic) B-splines is constructed which optimally represents a specified set of spectral shapes, and therefore also the family of spectra that are linear combinations of these shapes. (*a*) A basis of cubic B-splines defined by seven internal knots. The knot positions were adjusted using a minimization procedure to optimize the ability of the basis set to span a specified set of spectra. The locations of the optimal knots are indicated by arrows. (*b*) The set of spectra used as the target for optimization of the B-spline basis shown in (*a*). The solid curves are the target spectra, and the dashed curves are the final best-fit linear combinations of the optimal B-splines. (*c*) Differences between the target spectra and their optimal representations as linear combinations of the B-spline spectra shown in (*a*).

our species spectra all be linear combinations of such a set of basis "spectra," which may be arranged as columns of a matrix **B**—that is, $\mathbf{S} = \mathbf{BC}$ for some matrix **C** of coefficients. The linear least-squares problem of Eq. 10 then takes the form

$$\mathbf{A} \approx \mathbf{BCM} \tag{27}$$

Here the matrices **B** and **M** are known, and we must solve for the matrix **C**. This problem is a special case of the multi-term linear least-squares problem discussed in Appendix B and may be solved using the methods described therein. (The SVD-based approach described above is clearly a special case of Eq. 27 in which we can exploit the orthonormality of the columns of the matrix $\mathbf{U}'$ (=**B**) of basis spectra to eliminate this matrix from the problem.)

Explicit relationships between species spectra

The methods discussed above allow the shape of a pre-set spectrum to be specified to at least some extent. If two or more pre-set spectra are defined in a consistent manner, it may be possible to enforce desired relationships (e.g., relative peak positions) between them through appropriate choices of—or constraints on—parameters defining the spectral shape. On the other hand, the "free" spectra, which are only determined by solving the linear least-squares problem of Eq. 10, are not subject to such control. Within a general framework, the imposition of constraints relating *global* features of spectra whose shapes are completely unknown a priori, if feasible, may require difficult nonlinear constraints on all the spectral amplitudes. However, our ability to represent spectra sampled on sets of spectroscopic parameters λ as vectors, and to describe the least-squares problem in terms of matrix operations involving these vectors, provides a framework for a possible simpler approach. This method requires that prescribed relationships between otherwise unknown spectra be described in terms of linear transformations applied to the associated vectors. That is, some matrix **T** may be constructed such that the vector associated with free spectrum S_2 is computed from the vector for free spectrum S_1 by the product $S_2 = \mathbf{T}S_1$. (Note that these are linear transformations combining the various elements of a single spectrum, corresponding to different spectroscopic indices λ, to produce the elements of a new spectrum, rather than linear combinations of corresponding elements of several spectra.) We show in Appendix A that, with some generality, such a transformation matrix **T** may be constructed to represent simple shifts along the λ axis; variations of the same procedure allow the construction of approximate matrix representations for other distorting transformations as well (e.g., narrowing/broadening).

To proceed, we assume that the pre-set spectra have been constructed using one of the methods described above. It then remains only to solve the linear least-squares problem of Eq. 10. We define a minimal set of m "prototype" free spectra $\{F_i\}$ (which are columns of a matrix **F**), in terms of which all other free spectra may be defined using linear transformations **T**. For generality, we allow there to be p distinct such transformations $\mathbf{T}_j$ operative simultaneously. That is, we can divide the nonprototype free spectra into p disjoint sets of spectra $\pi_1, \ldots, \pi_p$ (with set π_i containing $m_i \leq m$ spectra), which are columns of corresponding matrices $\Pi_1, \ldots, \Pi_p$, such that each column vector of Π_k is related to a unique column of **F** through the transformation $\mathbf{T}_k$. In matrix notation:

$$\Pi_k = \mathbf{T}_k\mathbf{F}\mathbf{E}_k \tag{28}$$

where the $m \times m_k$ selection matrix $\mathbf{E}_k$ consists of ones and zeroes arrayed in such a way that post-multiplication of **F**

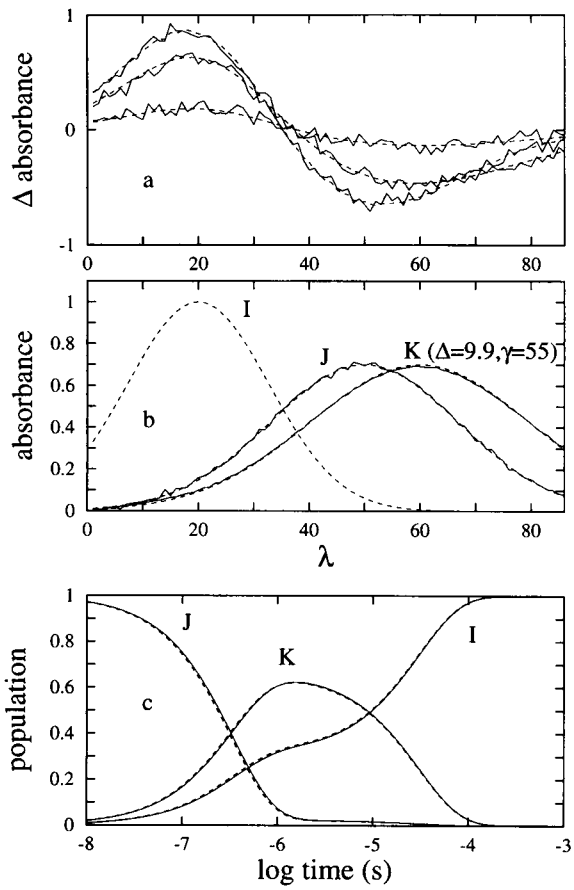

FIGURE 3 An illustration of the use of linear operators to enforce relationships between otherwise unknown spectra during model fitting. A simulated data matrix **A** was constructed from a single set of three prototype species spectra (I, J, K) as a function of spectroscopic index λ and their corresponding populations as a function of time. The species spectra were all defined as Gaussians with differing widths and peak positions. Populations were computed from a simple kinetic scheme I ↔ J ↔ K using prescribed values of the interconversion rates k(I → J), k(J → I), k(J → K) and k(K → J) and initial populations [I] = [K] = 0 and [J] = 1. All spectra were "measured" as differences referenced to the fixed spectrum of the fully populated species I. Normally distributed pseudo-random numbers (of standard deviation 0.04) were added to all elements of the noise-free data matrix to produce the simulated data matrix **A**. The two spectra $\mathbf{S}_J$ and $\mathbf{S}_K$ (represented as column vectors) and selected kinetic parameters were recovered from these data by least-squares fitting. In the fitting, only $\mathbf{S}_J$ was treated as a strictly unknown ("prototype free") spectrum; the free spectrum $\mathbf{S}_K$ was expressed as a scaled, shifted and broadened form of $\mathbf{S}_J$ through the use of the linear operators introduced in Appendix A:

$$\mathbf{S}_K = [\text{scale}][\text{shift}][\text{broaden}]\mathbf{S}_J$$
$$= \mu \exp(-\Delta\mathbf{Z})\exp(\gamma\mathbf{Y})\mathbf{S}_J$$

where **Z** and **Y** are matrix representations of the first- and second-derivative operators, respectively, Δ and γ are adjustable shift and broadening parameters, respectively, and μ is an adjustable scale factor to compensate for the effect of the broadening operator exp(γ**Y**) on the peak amplitude of the spectrum (Appendix A). The least-squares problem is then:

$$\mathbf{A} \approx (\mathbf{S}_J - \mathbf{S}_I)\mathbf{M}_J + (\mathbf{S}_K - \mathbf{S}_I)\mathbf{M}_K$$

$$\mathbf{A} + \mathbf{S}_I(\mathbf{M}_J + \mathbf{M}_K) \approx \mathbf{S}_J\mathbf{M}_J + \mu(\exp(-\Delta\mathbf{Z})\exp(\gamma\mathbf{Y}))\mathbf{S}_J\mathbf{M}_K$$

by this matrix selects and reorders the columns of **F** so that column j of the reordered matrix produces column j of $_k$ through $\mathbf{T}_k$. Corresponding to each matrix Π_k is a $m_k \times n_p$ matrix $\mathbf{M}_k$ consisting of those rows of the current population matrix **M** corresponding to the species spectra making up the columns of Π_k. This operation of selecting rows of **M** may be written

$$\mathbf{M}_k = \mathbf{D}_k\mathbf{M} \tag{29}$$

for appropriate $m_k \times n_s$ selection matrices $\mathbf{D}_k$. We also define matrices $\mathbf{M}_F$ and $\mathbf{D}_F$ for the populations of those states corresponding to the prototype spectra in **F**.

The total of contributions of the sets of spectra **F** and the Π_k and associated populations to the residual simulated data set in Eq. 17 is

$$\mathbf{A} - \mathbf{P}\mathbf{M}_p \approx \mathbf{F}\mathbf{M}_F + \sum_{k=1}^{p} \Pi_k\mathbf{M}_k \tag{30}$$

With the above definitions, we can now write this contribution using only the prototype spectral matrix **F** and the free population matrix $\mathbf{M}_f$ as

$$\mathbf{A} - \mathbf{P}\mathbf{M}_p \approx \mathbf{F}\mathbf{D}_F\mathbf{M}_f + \sum_{k=1}^{p} \mathbf{T}_k\mathbf{F}\mathbf{E}_k\mathbf{D}_k\mathbf{M}_f$$
$$= \sum_{k=0}^{p} \mathbf{B}_k\mathbf{F}\mathbf{C}_k \tag{31}$$

where we define

$$\mathbf{B}_0 = \mathbf{I}, \qquad \mathbf{B}_k = \mathbf{T}_k$$
$$\mathbf{C}_0 = \mathbf{D}_F\mathbf{M}_f, \quad \mathbf{C}_k = \mathbf{E}_k\mathbf{D}_k\mathbf{M}_f \tag{32}$$

We now have a linear least-squares problem of the general form

$$\mathbf{A}_f \approx \sum_{k=0}^{p} \mathbf{B}_k\mathbf{F}\mathbf{C}_k \tag{33}$$

in which the matrices $\mathbf{B}_k$, $\mathbf{C}_k$, and $\mathbf{A}_f$ are provided, and we must determine the unknown matrix **F**. Unlike for the sim-

where $\mathbf{S}_J - \mathbf{S}_I$ and $\mathbf{S}_K - \mathbf{S}_I$ are the column vectors representing the two unknown difference spectra, and $\mathbf{M}_J$ and $\mathbf{M}_K$ are the corresponding species populations (as row vectors) computed from the kinetic model. For fixed Δ, γ, μ, and populations **M**, this is a special case of the linear least-squares problem of Eq. 33 for the unknown single-column matrix $\mathbf{S}_J$, which may be solved using the methods described in Appendix B. During the fit, the kinetic rates k(I → J) and k(K → J) were fixed at the values used to construct the data set, and the rates k(J → I) and k(J → K) were permitted to vary from "random but reasonable" starting values. The final output of the fit consists of optimal values of these two kinetic parameters, the single spectrum $\mathbf{S}_J$, and the shift Δ, broadening γ, and scaling μ required to construct $\mathbf{S}_K$ from $\mathbf{S}_J$. (*a*) Selected spectra from the simulated data set (*solid curves*) with the corresponding spectra from the best-fit data set (*dashed curves*). (*b*) Original Gaussian species spectra (*dashed curves*) and the spectra of J and K extracted from the model fit (*solid curves*). (*c*) Species populations used to create the noise-free data set (*dashed curves*) and the populations derived from the least-squares fit (*solid curves*).

ple case represented by Eq. 10, this problem is not generally solvable by a simple sequence of matrix transformations and/or decompositions. However, there are straightforward methods for solving the problem provided that none of the matrices involved becomes too large. These methods are discussed in Appendix B.

To summarize, this section presents a procedure for solving the restricted linear least-squares problem in which certain of the free spectra are derivable from others by linear transformations. The matrices $\mathbf{E}_k$ and $\mathbf{D}_k$ and their products need to be computed only once, and the matrices $\mathbf{M}_f$ and $\mathbf{M}_p$ are derived in a simple way from the current population matrix $\mathbf{M}(\{\xi\})$. The matrices $\mathbf{T}_k$ may be fixed or defined in terms of adjustable parameters and computed at each iteration. The final solution produces only the minimal set of prototype free spectra $\mathbf{F}$, from which the remaining free spectra are calculated as needed using the current $\mathbf{T}_k$. A simple illustration of this technique is shown in Fig. 3. Here the modeling problem involves two unknown spectra, those of the species J and K. It is assumed that the spectrum of K is derivable from that of J through a product of successive linear transformations, a broadening, followed by a shift in λ (both of which are discussed in Appendix A), followed by a simple scaling. That is, the spectrum of J is the single prototype free spectrum, from which the spectrum of K is determined through the combined operator $\mathbf{T}_1 = \mu \exp(-\Delta\mathbf{Z})\exp(\gamma\mathbf{Y})$; $\mathbf{Z}$ and $\mathbf{Y}$ are constant matrices, and the shift parameter Δ, broadening parameter γ and scale factor μ are varied during the fit.

Formally this procedure may be applied to an SVD-based analysis as well as to a direct treatment of the spectra. However, the SVD-based analysis represents species spectra only using the coefficients that express these spectra as linear combinations of the included columns of $\mathbf{V}$. This level of abstraction makes the task of constructing linear transformations that enforce useful relationships between free species spectra much more difficult.

Other constraints on species spectra

Linear-algebraic techniques also make it possible to solve the linear least-squares problem in Eq. 10 for the matrix $\mathbf{S}$ of spectral amplitudes while imposing a broad class of linear constraints relating the amplitudes for different values of the spectroscopic index λ within and among the species spectra. These constraints have the general form

$$\sum_{\text{spectra j}} \sum_{\text{elements i}} \alpha_{ij} S_{ij} \begin{cases} \geq c \\ = c \end{cases} \tag{34}$$

where the upper alternative ($\geq c$) represents a general inequality constraint and the lower represents an enforced equality. A useful example of such a constraint imposes an equality or inequality relation on the integrated areas under two or more species spectra. With individual spectra represented as vectors, the integrated area of spectrum k is computed (to within a constant factor) as the simple sum over all the elements of the corresponding vector—that is, with all coefficients α_{ik} ($i = 1, \ldots, n_\lambda$) equal to unity. A simple constraint on the numerical value of the area of a single spectrum k would require coefficients

$$\alpha_{ij} = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases} \tag{35}$$

with the constant c in Eq. 34 equal to the required value of the area. A constraint on the numerical difference between the area of species spectrum k and some constant multiple κ of the area of spectrum l would be represented by coefficients of the form

$$\alpha_{ij} = \begin{cases} 1, & j = k \\ -\kappa, & j = l \\ 0, & j \neq k, l \end{cases} \tag{36}$$

In a similar fashion, relations between corresponding elements of different spectra or between different elements of a single spectrum are prescribed using nonzero values of selected coefficients, with the remaining coefficients equal to zero.

Efficient methods for incorporating such linear constraints into the linear least-squares problem are discussed in detail by Lawson and Hanson (1974). These methods generally rely on a formulation of the least-squares problem as $\mathbf{Ex} \approx \mathbf{f}$, where $\mathbf{E}$ is a matrix and $\mathbf{x}$ and $\mathbf{f}$ are vectors; equality and inequality constraints on the amplitudes $\mathbf{x}$ are written as $\mathbf{Cx} = \mathbf{d}$ and $\mathbf{Gx} \geq \mathbf{h}$, respectively, for matrices $\mathbf{C}$ and $\mathbf{G}$ and vectors $\mathbf{d}$ and $\mathbf{h}$. To exploit such techniques to solve Eq. 10 with linear constraints on elements of $\mathbf{S}$, we must restate the problem in expanded form as developed in Appendix B, in which the data matrix $\mathbf{A}$ and the spectral matrix $\mathbf{S}$ are both arranged as single-column vectors, and the coefficient matrix is constructed using the Kronecker product as in Eq. B6. Then each row of the constraint coefficient matrix $\mathbf{C}$ or $\mathbf{G}$ consists of a complete set of coefficients $\{\alpha_{ij}\}$ for a single constraint as written in Eq. 34, with the corresponding position in the column vector $\mathbf{d}$ or $\mathbf{h}$ containing the appropriate value of the consta nt c for that constraint.

As discussed in Appendix B, the solution of the least-squares problem of Eq. B6 may be complicated by the large size of the coefficient matrices that may arise. However, this approach is necessary when inequality constraints, or equality constraints of the most general form (Eq. 34), are required. On the other hand, a certain class of useful equality constraints may be incorporated directly into the solution of Eq. 10 in full-matrix form. Specifically, it may be shown (E. Henry, unpublished work) that the solution of $\mathbf{AX} \approx \mathbf{B}$ subject to the equality constraints $\mathbf{CXD} = \mathbf{E}$, for specified matrices $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{E}$, is

$$\begin{aligned} \mathbf{X} &= \bar{\mathbf{X}} + (\mathbf{I}_n - \mathbf{C}^+\mathbf{C})(\mathbf{A}(\mathbf{I}_n - \mathbf{C}^+\mathbf{C}))^+\mathbf{G}\mathbf{D}\mathbf{D}^+ \\ &\quad + \mathbf{A}^+\mathbf{G}(\mathbf{I}_m - \mathbf{D}\mathbf{D}^+) \\ \bar{\mathbf{X}} &= \mathbf{C}^+\mathbf{E}\mathbf{D}^+, \ \mathbf{G} = \mathbf{B} - \mathbf{A}\bar{\mathbf{X}} \end{aligned} \tag{37}$$

where $\mathbf{M}^+$ is the so-called *pseudoinverse* of the matrix $\mathbf{M}$ (Lawson and Hanson, 1974), and $\mathbf{I}_n$ and $\mathbf{I}_m$ are the identity matrices of the same dimensions as $\mathbf{CC}^+$ and $\mathbf{DD}^+$, respectively. The form $\mathbf{CXD} = \mathbf{E}$ imposes linear constraints (with sets of coefficients given by the columns of $\mathbf{D}$) on specific linear combinations (with sets of coefficients given by the rows of $\mathbf{C}$) of the individual elements of each column of $\mathbf{X}$. This allows equality constraints relating, for example, the integrated areas—or the values of selected elements—of one or more species spectra to be treated within the much more economical framework of Eq. 10.

SPECTROSCOPIC MODELING IN THE PRESENCE OF MEASUREMENT ERRORS

In actual experimental situations, a spectroscopic measurement incurs some uncertainty, arising from both random errors ("noise") and systematic errors. In the presence of such possible errors, the optimization problem in Eq. 7 should reflect that any single determination $\mathbf{A}$ of a $(m \times n)$ spectroscopic data matrix may be written as the product of a "true" spectral matrix $\mathbf{S}$ and a matrix $\mathbf{M}$ of model species populations, plus some error matrix $\mathbf{E}$. That is

$$\mathbf{A} = \mathbf{SM} + \mathbf{E} \equiv \mathbf{A}^0 + \mathbf{E} \tag{38}$$

where $\mathbf{A}^0$ is the "error-free" prediction of the model. Hypothetically, for a given $\mathbf{A}^0$, a large number of independent determinations of $\mathbf{A}$ would produce a set of data matrices distributed according to some conditional probability distribution $\rho(\mathbf{A}|\mathbf{A}^0)\ d\mathbf{A} \equiv \rho(\{A_{11}, A_{12}, \dots\}|\{A^0_{11}, A^0_{12}, \dots\})\ dA_{11}dA_{12}\dots$. If we assume that the variations of each element of the error matrix during successive trials are uncorrelated with those of every other element, then the joint probability density $\rho(\mathbf{A}|\mathbf{A}^0)$ is the product of independent probability densities for each element—i.e., $\Pi\rho_{ij}(A_{ij}|A^0_{ij})$. Consider now a specific measured data matrix $\mathbf{A}$. Any set of parameters produces a predicted $\mathbf{A}^0$ from which the probability $\rho(\mathbf{A}|\mathbf{A}^0)\ \Delta\mathbf{A}$ of the measured $\mathbf{A}$ may be computed (see Note 5). Conversely, one predicted data matrix may be regarded as somehow more "likely" than another if it predicts a higher probability for the available measured $\mathbf{A}$. The *maximum likelihood estimate* of the model parameter set (including both spectral amplitudes in $\mathbf{S}$ and possible nonlinear parameters from which the population matrix $\mathbf{M}$ is computed) is that set which produces a predicted $\mathbf{A}^0$ for which the probability $\rho(\mathbf{A}|\mathbf{A}^0)\Delta\mathbf{A}$ of the *measured* $\mathbf{A}$ is maximized. In the above-mentioned case in which probability densities may be defined individually for each matrix element, this requires that the expression

$$\begin{aligned}\prod \rho_{ij}(A_{ij}|A^0_{ij})\Delta\mathbf{A} &= (\prod \exp(\ln \rho_{ij}(A_{ij}|A^0_{ij})))\Delta\mathbf{A} \\ &= \exp(\ln \Delta\mathbf{A} + \sum \ln \rho_{ij}(A_{ij}|A^0_{ij}))\end{aligned} \tag{39}$$

be a maximum with respect to all parameters, or equivalently that the expression

$$\sum(-\ln \rho_{ij}(A_{ij}|A^0_{ij})) \tag{40}$$

be minimized.

For a given set of conditional probability density functions $\rho_{ij}(A_{ij}|A^0_{ij})$, the maximum likelihood estimate for some set of parameters (and therefore for the predicted data matrix $\mathbf{A}^0$) is produced by minimizing Eq. 40 with respect to all the parameters. The applicability of this procedure does not depend on special assumptions about the functional forms of the ρ_{ij}. Nevertheless, it is often useful to focus on the sub-class of problems for which the probability densities are only functions of the deviations between the predicted and the measured data matrices, i.e.,

$$\rho_{ij}(A_{ij}|A^0_{ij}) = \rho_{ij}(A_{ij} - A^0_{ij}) \tag{41}$$

In the event that these deviations are normally distributed—i.e., $\rho_{ij} \sim \exp(-((A_{ij} - A^0_{ij})/2\sigma_{ij})^2)$—maximum likelihood estimation reduces to the familiar problem of minimizing the functional

$$\chi^2 = \sum_{ij} \frac{1}{\sigma^2_{ij}} (A_{ij} - A^0_{ij})^2 \tag{42}$$

Practical aspects of maximum likelihood estimation in this special case, as well as for more general forms of the error probability densities ρ_{ij}, are discussed in Appendix C. In principle, the rigorously correct form of these probability densities may be deduced from a careful examination of the acquisition and analysis procedures by which a data matrix has been produced. However, such detailed information, while highly desirable, is not necessarily a prerequisite to a modeling analysis. Very often, a relatively simple probability function that captures the essential qualitative features of the true distribution may be "good enough" (see Note 6). For example, the model-fitting of a data set for which the measurement noise distribution is practically symmetric and "local," with negligible probability of outliers outside of some region, may often be conducted as effectively with a simple least-squares procedure as with a more costly treatment using the correct (non-normal) distribution. A simple illustration of this point appears in Fig. 4 *a*. Here, noise from a non-normal probability distribution has been added to a synthetic data set constructed using idealized species spectra $\mathbf{S}$ and populations $\mathbf{M}$. The chosen noise distribution is local (in the sense that the variance and all higher-order moments exist), albeit significantly less so than the normal distribution that underlies a least-squares analysis. In this case, recovery of the original species spectra from the noisy data set using a simple least-squares analysis is essentially as successful as the recovery by the full convergent maximum likelihood analysis using the iterative procedure described in Appendix C with the correct probability distribution. On the other hand, a least-squares or similarly "rigid" analysis will be sensitive to the presence of large and/or frequent outliers in a data set. Such a data set should be modeled assuming a probability distribution that better

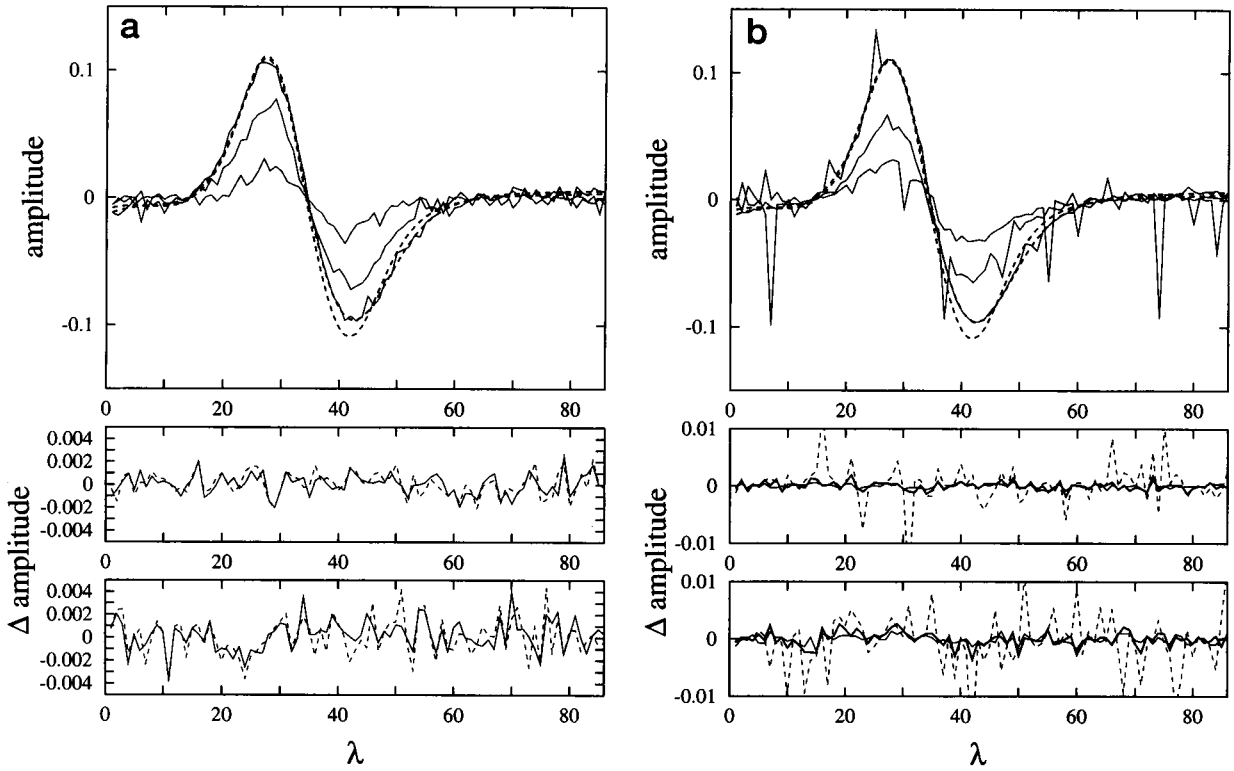

FIGURE 4 The recovery of species spectra from simulated data sets including noise from different probability distributions. Noise-free data matrices were constructed as in Fig. 3, using the same kinetic scheme and rates to compute the species populations, but with a distinct set of species spectra. Independently generated pseudo-random numbers from the chosen probability distribution were added to each of the elements of the noise-free data matrix to produce the simulated data. (*a*) Recovery of the species spectra from a data set constructed using a double exponential noise distribution $\rho(x) \sim \exp(-|x|)$. The maximum-likelihood problem for this distribution corresponds to minimizing the sum of the absolute deviations (rather than the sum of the squared deviations) between the measured and simulated data sets. The top panel shows the two difference spectra $\mathbf{S}_J - \mathbf{S}_I$ and $\mathbf{S}_K - \mathbf{S}_I$ used to synthesize the data matrix (*dashed curves*) and selected spectra from the simulated data set (*solid curves*). The lower panels show differences between each recovered spectrum and the corresponding original spectrum. The solid curve in each panel represents the recovery by solving the full maximum-likelihood problem using the actual noise distribution, and the dashed curve corresponds to the solution obtained by solving the simple least-squares problem. In this case, recovery by the two methods yields species spectra of comparable quality. (*b*) Same as (*a*), but for a probability distribution for the added noise having the basic form

$$\rho(x) \sim \frac{1}{(1 + ax^2)^2}$$

This distribution has finite variance, but infinite fourth and higher moments, and therefore has a tendency to produce outliers in the simulated data set. The obviously exaggerated outliers evident in the typical spectra shown in the upper panel produce large deviations between the spectra recovered using simple least-squares and the original spectra (*dashed curves* in *lower panels*). This effect is much reduced in the recoveries performed by solving the maximum-likelihood problem assuming either the correct probability distribution (*thick solid curve*) or the double exponential distribution (*thin solid curve*) used in (*a*).

reflects the observed appearance of large deviations. In Fig. 4 *b*, noise from a nonlocal probability distribution has been added to the same idealized data set. The species spectra extracted using simple least-squares (based on an assumed normal distribution of the measurement errors) are highly distorted by outlier points. These distortions are suppressed in spectra extracted by a maximum likelihood analysis assuming less localized error distributions, which exact a much smaller penalty for the appearance of outliers.

As a final illustration, consider a set of spectra that reflects some counting of random events (e.g., radioactive decay). For large counts, the measured amplitude in any one channel will be described by a normal distribution with variance (σ^2) equal to the mean. A maximum likelihood analysis of such a data set then requires minimization of χ^2 in Eq. 42, with each variance σ_{ij}^2 given by an estimate of the corresponding mean value A_{ij}^0. The results of such an analysis performed (as described in Appendix C) on a simulated data set are shown in Fig. 5. The extracted species spectra reproduce the original spectra extremely well, as expected. Moreover, the species spectra extracted using the matrix decomposition least-squares method described in Direct versus SVD-Based Analysis of the Data Matrix, which effectively assigns the same variance to each element of the data matrix, are of comparable quality. In this specific case, then, we would be justified in using the most efficient method to solve the corollary linear least-squares problem, with some confidence that the resulting spectra will differ

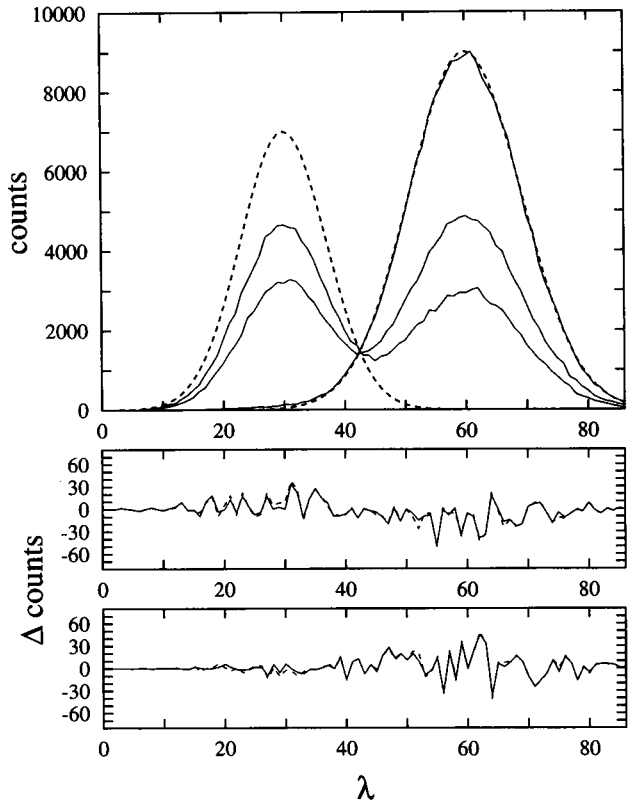

FIGURE 5 Recovery of species spectra from a simulated data set constructed for a counting process (with the variance of each measurement equal to the expected measured value). The noise-free data matrix was constructed as in Fig. 3 from two species spectra and a set of species populations derived from a simple kinetic scheme. To each element of the resulting data matrix was added a pseudo-random number taken from a normal distribution with variance equal to the noise-free value of that element. The top panel shows the two species spectra used to synthesize the data matrix (*dashed curves*) and selected spectra from the simulated data set (*solid curves*). The lower panels show differences between each recovered spectrum and the corresponding original spectrum. The solid curves represent the spectra recovered by solving the weighted least-squares problem using the correct variance at each point, and the dashed curves show the recovered spectra from the solution of the simple unweighted least-squares problem.

only negligibly from those obtained from the more complete analysis.

A GENERALIZED SPECTROSCOPIC MODELING PROBLEM

One final topic we would like to touch upon is a possible generalization of the spectroscopic modeling problem beyond that introduced in Definition and Features of the Spectroscopic Modeling Problem. Up to this point we have assumed that the species spectra are independent of experimental conditions; rather, all such dependences may be isolated in a population matrix function $\mathbf{M}(\{\xi_k\})$, which multiples the matrix of species spectra in the least-squares problem in Eq. 7. One can imagine situations in which one or more of the spectra characterizing a system may vary with experimental conditions (e.g., a dependence on pH or temperature). We begin simply by rewriting Eq. 5 to reflect this added flexibility:

$$A(\lambda, \{x_i\}) \approx \sum_j S_j(\lambda, \{x_i\}, \{\zeta_l\}) M_j(\{x_i\}, \{\xi_k\}) \qquad (43)$$

We have also included possible dependences of species spectra on a set of adjustable parameters $\{\zeta_l\}$. In this general form, the condition-dependence of the spectra does not allow us to write the least-squares problem in a matrix-product form $\mathbf{A} \approx \mathbf{SM}$. However, if we assume (in the spirit of our earlier discussion and Appendix A) that variations of the species spectra with conditions may be incorporated into condition-dependent linear operators (i.e., matrices) operating on a set of condition-independent "prototype" spectra, then we can recover a representation of the least-squares problem in Eq. 43 in a matrix form. Specifically, we write the species spectra as column vectors:

$$S_j(\lambda, \{x_i\}, \{\zeta_l\}) = \mathbf{T}_j(\{x_i\}, \{\zeta_l\}) S_j^0(\lambda) \qquad (44)$$

where all condition- and parameter-dependences have been absorbed into the $n_\lambda \times n_\lambda$ matrix functions $\mathbf{T}_j$, and the "prototype" spectra assume the "0" superscript. Each spectrum A_i ($\equiv A(\lambda, \{x_i\})$) may be written as a column vector

$$A_i \approx \sum_j M_j(\{x_i\}, \{\xi_k\}) \mathbf{T}_j(\{x_i\}, \{\zeta_l\}) S_j^0(\lambda)$$

$$= \sum_j \mathbf{W}_j(\{x_i\}, \{\xi_k\}, \{\zeta_l\}) S_j^0(\lambda) \qquad (45)$$

where we have absorbed all condition- and parameter-dependences into the matrix functions $\mathbf{W}_j$. This expression is clearly equivalent to Eq. 7 in the event that all species spectra are condition-independent, in which case each matrix $\mathbf{W}_j(\{x_i\}, \{\xi_k\}, \dots)$ reduces to a single scalar $M_j(\{x_i\}, \{\xi_k\})$ (or this scalar multiplied by the $n_\lambda \times n_\lambda$ identity matrix). These quantities are then arranged in a population matrix $\mathbf{M}$, which post-multiplies a spectral matrix $\mathbf{S}^0$ as before. In the general case, however, each prototype spectrum is multiplied by a unique condition-dependent matrix to construct each spectrum of the simulated data set. If we now array the n_p column vectors A_i in a single-column "supervector," and likewise for the n_s prototype spectra S_j^0, then the least-squares problem in Eq. 43 becomes

$$\begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_{n_p} \end{pmatrix} \approx \begin{pmatrix} \mathbf{W}_1(\{x_1\}) & \mathbf{W}_2(\{x_1\}) & \dots & \mathbf{W}_{n_s}(\{x_1\}) \\ \mathbf{W}_1(\{x_2\}) & \mathbf{W}_2(\{x_2\}) & \dots & \mathbf{W}_{n_s}(\{x_2\}) \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{W}_1(\{x_{n_p}\}) & \mathbf{W}_2(\{x_{n_p}\}) & \dots & \mathbf{W}_{n_s}(\{x_{n_p}\}) \end{pmatrix} \begin{pmatrix} S_1^0 \\ S_2^0 \\ \vdots \\ S_{n_s}^0 \end{pmatrix} \qquad (46)$$

$$\text{vec}(\mathbf{A}) \approx \mathbf{W}\,\text{vec}(\mathbf{S})$$

where for notational simplicity we have suppressed the parameter dependences of the matrix blocks $\mathbf{W}_j(\{x_i\})$ that make up the "supermatrix" $\mathbf{W}$. The *vec*(. . .) operation is described in Appendix B, and creates a column vector $\mathbf{a}$ from the matrix $\mathbf{A}$ and a column vector $\mathbf{s}$ from the matrix $\mathbf{S}$. Having now formally reduced the problem to the simple matrix-vector form $\mathbf{a} \approx \mathbf{W}(\{\xi_k\}, \{\zeta_l\})\mathbf{s}$, we can proceed in much the same way as has been outlined earlier in this article: During each iteration of some nonlinear least-squares optimization procedure, the matrix $\mathbf{W}$ is first constructed using the current values of the model parameters $\{\xi_k\}$ and spectral transformation parameters $\{\zeta_l\}$; this will involve computing transformation matrices $\mathbf{T}_j$ and/or species populations M_j for each set of conditions. This matrix is then used to solve the linear least-squares problem $\mathbf{a} \approx \mathbf{W}\mathbf{s}$ for a current set of consistent prototype spectra in $\mathbf{s}$. All adjustable parameters are varied until the residual between $\mathbf{a}$ and the product of the current $\mathbf{W}$ and the corresponding $\mathbf{s}$ is minimized.

A very simple example of such a problem is shown in Fig. 6. A synthetic data matrix $\mathbf{A}$ is created consisting of analytical Lorentzian spectra (for a single "species"), which vary with condition variable x (which may be interpreted as time, for concreteness) only in their peak position. A fitting operation is performed to recover a single "prototype" spectrum from which all the spectra in the data set are generated by application of the condition-dependent shift operator $\exp(-\Delta(x, \{\zeta_l\})\mathbf{Z})$ constructed as described in Appendix A. The fit is performed by varying only the parameters $\{\zeta_l\}$ associated with the presumed functional form of the condition-dependent shift Δ. For each such set of parameters, the "supermatrix" $\mathbf{W}$ is constructed as:

$$\mathbf{W} = \begin{pmatrix} \exp(-\Delta(x_1, \{\zeta_l\})\mathbf{Z}) \\ \exp(-\Delta(x_2, \{\zeta_l\})\mathbf{Z}) \\ \vdots \\ \exp(-\Delta(x_{n_p}, \{\zeta_l\})\mathbf{Z}) \end{pmatrix} \qquad (47)$$

$\mathbf{W}$ is then used to solve the linear least-squares problem $vec(\mathbf{A}) \approx \mathbf{W}\mathbf{s}$ for the single prototype spectrum $\mathbf{s}$. The parameters $\{\zeta_l\}$ are varied until this procedure produces an optimal approximation of the data matrix. The fitting procedure succeeds in producing an excellent approximation to the original Lorentzian prototype spectrum—*with no prior assumptions about the actual shape of the spectrum*—even in the presence of significant added noise in the synthetic data set.

CONCLUDING REMARKS

The goal of this article has been to present a general framework for the model-based analysis of large sets of spectroscopic data that exploits to the greatest possible extent the power and economies offered by matrix analysis. This framework relies on our ability to write a simulated data set as the product of a matrix $\mathbf{S}$ of species spectra and a matrix

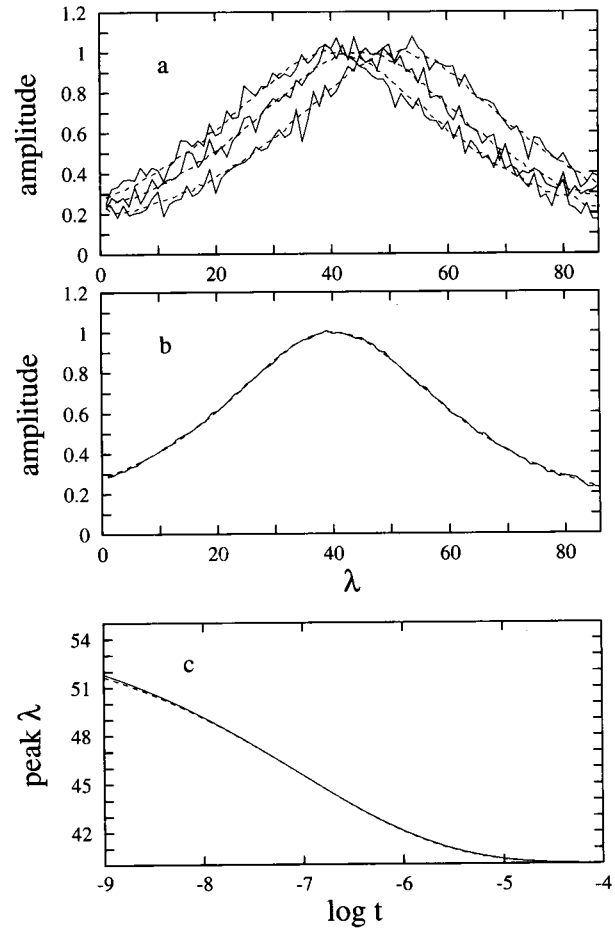

FIGURE 6 A simple illustration of the generalized spectroscopic modeling problem, in which species spectra are themselves allowed to vary with experimental conditions. A noise-free data set was constructed from a single prototype spectrum, which is a Lorentzian spectrum with a constant amplitude. Individual spectra were computed for various times by allowing the peak position of the prototype spectrum to evolve according to the "stretched exponential" decay $\exp(-(kt)^\beta)$:

$$(t_i) = \frac{1}{1 + ((\lambda - \lambda_m(t_i))/\Gamma)^2}$$

$$\lambda_m(t) = \lambda_m(\infty) + (\lambda_m(0) - \lambda_m(\infty))\exp(-(kt)^\beta)$$

Here $\Gamma = 25$, $\lambda_m(0) = 55$ and $\lambda_m(\infty) = 40$; $k = 10^7\ s^{-1}$ and $\beta = 0.3$. Thirty such spectra were computed for times uniformly distributed on a logarithmic scale. Normally distributed pseudo-random numbers (with variance 0.04) were added to these spectra to produce the final simulated data set. The prototype spectrum S_0, and the parameters $\Delta(0)$, k, and β, were recovered by performing a least-squares fit of all the spectra S_i to time-dependent shifts of the prototype spectrum given by $\exp(-\Delta(t_i)\mathbf{Z})S_0$, where $\Delta(t_i) \sim \Delta(0)\exp(-(kt_i)^\beta)$. Extraction of the prototype spectrum corresponding to specific values of the parameters $\Delta(0)$, k, and β was performed as described in the text. (*a*) Selected spectra from the simulated data set (*solid curves*) with the corresponding spectra from the best-fit data set (*dashed curves*). (*b*) Original prototype Lorentzian spectrum (*dashed curve*) and the final prototype spectrum extracted from the model fit (*solid curve*). (*c*) Original time course of the peak position in the noise-free data set (*dashed curve*) and the time course derived from the least-squares fit (*solid curve*).

M of model species populations, and thereby decompose the model-fitting process into separate optimizations in the space of (possibly nonlinear) model parameters from which **M** is constructed, and in the space of linear parameters that make up the spectral matrix **S**. From this perspective we have dramatically reduced the dimensionality of the modeling problem, with some necessary costs in terms of flexibility in prescribing the detailed features of the modeling process. (Both the advantages and the possible compromises associated with this reduction are further increased when the analysis is performed within a rank-reduced—e.g., SVD-based—representation of the spectral amplitudes.) However, we have demonstrated here that many desirable mechanisms for "guiding" the modeling process in useful ways, including the ability to incorporate knowledge or assumptions about species spectra to varying levels of detail, are accessible within a matrix framework. Moreover, certain analytical prescriptions arise naturally in this framework that would be very difficult to implement outside of it (specifically, the encoding of global relationships between species spectra as linear operators.

The advantages of such an approach are most clear when the measurement errors are normally distributed with a uniform variance, in which case simple least-squares, and in particular the very efficient matrix-based linear least-squares used to obtain **S**, is appropriate. Still, we have shown that the analysis for the cases of normal error distributions with nonuniform variances and general non-normal distributions may also be cast into matrix form to some apparent advantage (see Appendix C). We have taken the position, illustrated using simple examples in Spectroscopic Modeling in the Presence of Measurement Errors, that assuming the simplest possible error distribution that reflects the qualitative features of the actual distribution can greatly facilitate the *search* for optimal model parameters and species spectra. Thus, data sets for which the distribution of measurement errors is known to be local may be treated for most purposes using simple least-squares, with more detailed information about the error distribution (if available) only required in the final stages of the modeling process or in the statistical characterization of optimal parameter sets. On the other hand, the analysis of data sets for which a local error distribution is clearly not appropriate may require at least the use of some generic nonlocal error distribution to extract the best maximum-likelihood estimates of model parameters and species spectra.

A few remarks are in order concerning the implementation of the various procedures described here. (Interested readers should contact the author for more detailed information.) We have programmed and tested all of these algorithms as user-defined procedures within a general-purpose interactive data-analysis program of our own design, which runs on various UNIX platforms. Most, if not all, of the algorithms may also be coded to run within other environments (including commercial analysis programs), which offer a core set of matrix analysis capabilities. The required capabilities (including singular value and other matrix decompositions and efficient solution of the linear least-squares problem) may be found in freely available subroutine libraries; we recommend the LAPACK library (Anderson et al., 1992), available from the NETLIB repository (e-mail:netlib@netlib.org; anonymous ftp: ftp.netlib.org; World-Wide Web: http://www.netlib.org). Procedures for solving nonlinear least-squares and other optimization problems are described and coded, for example, by Press et al. (1993). Another source for research-quality implementations of these and other numerical algorithms is the TOMS archive (of the Association for Computing Machinery's Transactions on Mathematical Software), also available from NETLIB. (For example, a version of the algorithm LSEI, which incorporates the solution of the linear least-squares problem in the presence of linear equality and inequality constraints is available as TOMS algorithm #587 (Hanson and Haskell, 1982).)

APPENDIX A: A MATRIX OPERATOR FOR PRODUCING SHIFTED SPECTRA

As discussed in Incorporating Known Species Spectra, we may wish to describe certain species spectra (defined as vectors) as parametrized linear transforms of a minimal set of prototype spectra. For example, in some instances it is useful to incorporate into an analysis two or more species spectra that differ only by a wavelength shift. To accomplish this we therefore require a matrix operator which, when applied to a spectrum expressed as a column vector, has the effect of shifting the set of wavelengths on which the spectrum is sampled by an adjustable amount.

Consider a spectrum S that is sampled on a uniform set of N spectroscopic indices $\{\lambda_i\}$ to produce a vector **S** such that $S_i = S(\lambda_i)$. We now wish to derive a matrix function $\mathbf{L}(\Delta)$, which effectively shifts the spectrum **S** by an amount Δ; for integer Δ we require

$$(\mathbf{L}(\Delta)\mathbf{S})_i \simeq S_{i-\Delta} \tag{A1}$$

We would like to extend this property to all real values of Δ (over some reasonable range of values by which the spectrum may be indexed). In this case $\mathbf{S}_{i-\Delta}$ is taken to signify some appropriate interpolation between $\mathbf{S}_{i-floor(\Delta)}$ and $\mathbf{S}_{i-ceil(\Delta)}$, where *floor*($x$) and *ceil*($x$) are the nearest integers $\leq x$ and $\geq x$, respectively. We would like $\mathbf{L}(\Delta)$ to have the properties:

$$\mathbf{L}(0) = \mathbf{I}_N \tag{A2}$$

$$\mathbf{L}(\Delta_1 + \Delta_2) = \mathbf{L}(\Delta_1)\mathbf{L}(\Delta_2)$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix.

For any displacement Δ, Eq. A1 requires that each element of the shifted spectrum (the right-hand side) be constructed as some linear combination of the elements of the unshifted spectrum, with coefficients (elements of $\mathbf{L}(\Delta)$), which are determined by Δ. We proceed by approximating elements S_i of the unshifted spectrum locally in the vicinity of each i using a polynomial of degree n which is fit to the set of points $(i - m, S_{i-m}), \ldots, (i, S_i), \ldots, (i + m, S_{i+m})$, for some chosen interval radius m. The approximating polynomial for each i may be written

$$f_i(x) = a_{i,0} + a_{i,1}(x - i) + \cdots + a_{i,n}(x - i)^n \tag{A3}$$

For specified values of m and n, the coefficients $\{a_{i,j}\}$ may be expressed as linear combinations of the values $\{S_{i-m}, \ldots, S_i, \ldots, S_{i+m}\}$. That is,

$$\begin{aligned} a_{i,0} &= \alpha^0_{i,-m}S_{i-m} + \alpha^0_{i,-m+1}S_{i-m+1} + \cdots + \alpha^0_{i,m}S_{i+m} \\ a_{i,1} &= \alpha^1_{i,-m}S_{i-m} + \alpha^1_{i,-m+1}S_{i-m+1} + \cdots + \alpha^1_{i,m}S_{i+m} \\ &\vdots \\ a_{i,n} &= \alpha^n_{i,-m}S_{i-m} + \alpha^n_{i,-m+1}S_{i-m+1} + \cdots + \alpha^n_{i,m}S_{i+m} \end{aligned} \tag{A4}$$

The coefficients $\alpha^k_{i,j}$ are functions only of m and n, and methods for computing them lie in the province of smoothing algorithms (e.g., Savitsky-Golay smoothing) that rely on local polynomial fitting (see, for example, Press et al. (1993)). The zero-order coefficients $a_{i,0}$ determined independently at each point i are in fact the values of the smoothed spectrum derived from **S** by a $(2m + 1)$-point, degree-n polynomial smoothing algorithm. The first-order coefficients $a_{i,1}$ give the linear correction to the spectrum value S_i at the central point i for displacements from that point. For small displacements δ, we can write

$$S_{i-\delta} \approx S_i - (\alpha^1_{i,-m}S_{i-m} + \cdots + \alpha^1_{i,0}S_i + \cdots + \alpha^1_{i,m}S_{i+m})\delta \tag{A5}$$

at each point i.

As noted above, if we shift the spectrum **S** by δ, the new value at index i would be the old value at "index" $i - \delta$. For small δ, the shifted spectrum is therefore given by

$$\mathbf{S}(\delta) \simeq \mathbf{S} - \delta \begin{pmatrix} \alpha^1_{1,0} & \cdots & \alpha^1_{1,m} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \alpha^1_{i,-m} & \cdots & \alpha^1_{i,m} & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \alpha^1_{N,-m} & \cdots & \alpha^1_{N,0} \end{pmatrix} \begin{pmatrix} S_1 \\ \vdots \\ S_{m+1} \\ \vdots \\ S_{i-m} \\ \vdots \\ S_{i+m} \\ \vdots \\ S_{N-m} \\ \vdots \\ S_N \end{pmatrix} = (\mathbf{1} - \delta\mathbf{Z})\mathbf{S} \tag{A66}$$

This equation becomes exact in the limit of small shifts δ. The corresponding expression for a finite shift of amplitude Δ may be constructed by building up the product of a large number of "infinitesimal" shifts—that is

$$\begin{aligned} \mathbf{S}(\Delta) &= \lim_{p\to\infty}\left(1 - \frac{\Delta}{p}\mathbf{Z}\right)^p \mathbf{S} \\ &= \exp(-\Delta\mathbf{Z})\mathbf{S} \end{aligned} \tag{A7}$$

This identifies the matrix exponential $\exp(-\Delta\mathbf{Z})$ as the shift operator $\mathbf{L}(\Delta)$. An operator of this form clearly satisfies the requirements listed in Eq. A2, for a fixed choice of matrix **Z**.

Some comments about the construction and use of this operator are in order. The matrix **Z** is a $N \times N$ matrix, each row of which contains coefficients characteristic of a polynomial fit of a specified degree to an interval of specified width surrounding a single point of the spectrum. The reliability of the shift operator depends upon the validity of these fits as descriptions of the local variations of the spectrum in the vicinity of each point. The first corollary of this fact is that the presence of large amounts of noise in the spectrum will tend to adversely affect the performance of the shift operator. Secondly, even in the absence of noise, if the polynomial fits in some region of the spectrum are of insufficiently high degree to track rapid variations in the spectrum, then distortions of the shifted spectrum may result. (Fortunately, there is complete freedom in constructing the matrix **Z** to adjust the degree of the fitting polynomial to be higher for those rows that correspond to wavelength regions of a typical spectrum where rapid variations are expected.) Finally, some care may be required in constructing the rows of **Z** corresponding to the ends of the spectrum. The shift operator is designed to produce a full N-point spectrum for shifts of any size. Therefore, a nonzero shift will "push" some of the spectrum off one end and create new information at the other end. This new information is generated by an extrapolation of the spectrum into new wavelength regions based upon the polynomial approximations to the spectrum close to the affected end. The intervals over which these polynomial fits are performed are necessarily smaller as the endpoints are approached and are not symmetric about the point of interest; the extrapolated portions of the shifted spectrum may exhibit nonphysical distortions as a result. The effect of these distortions on an analysis may be minimized by ensuring that the significant spectral information lies well within the "interior" of the chosen wavelength region, and also by careful choices of the polynomial degree and interval size close to the endpoints. In even the most carefully constructed shift operator, the inevitable breakdown of the extrapolation from the endpoints places an effective upper limit on the size of the shift for which the operator is usable.

As noted above, in the process of constructing the matrix **Z** there is considerable freedom in choosing polynomial degrees and fitting-interval sizes to produce acceptable behavior of the shift operator for the types of spectra being studied. Once this matrix is available, then computation of the shift operator requires evaluating the matrix exponential $\exp(-\Delta\mathbf{Z})$ for the specified shift Δ. The calculation of the matrix exponential is discussed by, among others, Golub and Van Loan (1989). The method described by them based on Padé approximants is straightforward to implement and of more than adequate quality for this application.

Typical performance of the shift operator is demonstrated in Fig. 7. Two competing effects of the specific choices of interval size and polynomial degree for the polynomial approximations are shown. The solid curve is for a matrix **Z** computed from a seventh-degree polynomial approximation applied over 17-point intervals centered on each point, and the dashed curve is for **Z** computed from a third-degree approximation applied over 9-point intervals. The former high-order approximation provides a more faithful representation of the shifted spectrum in the vicinity of the peak absorbance than the latter approximation, but is more sensitive to extrapolation artifacts toward the left endpoint and "ringing" due to over-defined polynomial fitting of the spectrum toward the right endpoint. A matrix **Z** computed using low polynomial degree and smaller intervals toward the endpoints and higher degree and wider intervals in the vicinity of the peak reproduces the best features of both (*long-dashed curve*).

It is instructive also to consider briefly how other useful transformation operators might be constructed. First, it should be noted that, if we regard the index i for the spectral elements S_i as a continuous parameter, then the matrix **Z** defined by Eq. A6 is essentially an approximate matrix representation of a first-derivative operator with respect to i, and the shift matrix $\mathbf{L}(\Delta)$ is an approximate matrix representation for the operator

$$\mathbf{L}_\Delta = \exp\left(-\Delta\frac{d}{di}\right) \tag{A8}$$

It is in fact easily shown that the exponential of the derivative operator with respect to some argument produces a constant shift in the value of that argument when applied to a function. That is, if the function $f(x)$ is infinitely differentiable over some interval, then

$$f(x + \Delta) = \exp\left(\Delta\frac{d}{dx}\right)f(x) \tag{A9}$$

Assume now that a function f has a Fourier representation

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\tilde{f}(k)dk \tag{A10}$$

We consider the effect of applying the exponential of the *second*-derivative

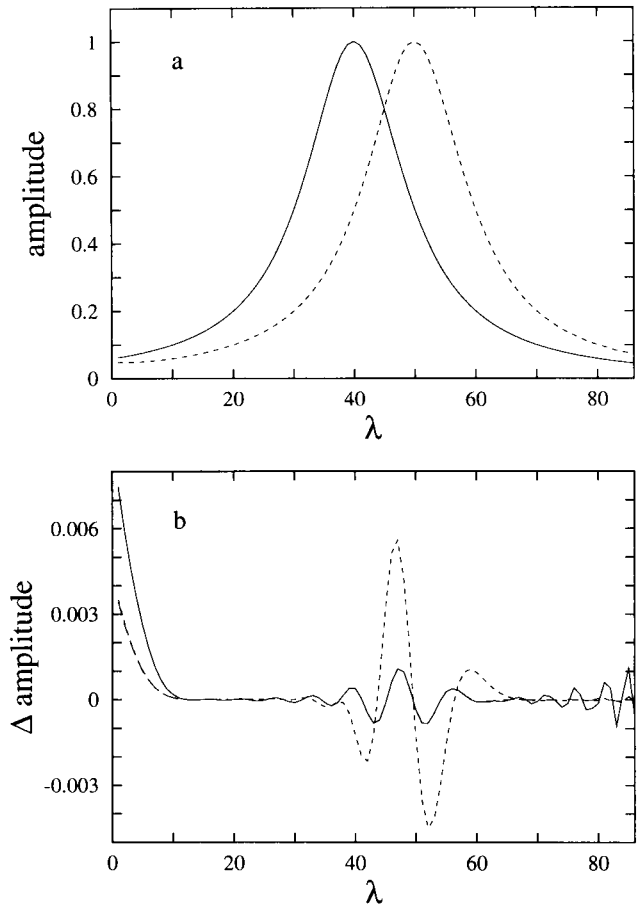

FIGURE 7 Typical performance of the shift operator developed in Appendix A. (*a*) An analytical Lorentzian spectrum, with a maximum at $\lambda = 40$, sampled at 86 points (*solid curve*) and a shifted vector computed by multiplying this vector by $\exp(-\Delta\mathbf{Z})$ for $\Delta = 10$ (channels) and a typical choice of $\mathbf{Z}$ (*dashed curve*). (*b*) The differences between three shifted vectors (produced using three distinct matrices $\mathbf{Z}$) and the "correct" shifted vector sampled from the Lorentzian analytical spectrum with the peak position displaced by 10 channels. The solid curve is for a matrix $\mathbf{Z}$ computed from a seventh-degree polynomial approximation (smoothly reduced to a second-degree approximation as the endpoints are approached) applied over 17-point intervals centered on each point; the short-dashed curve is for $\mathbf{Z}$ computed from a third-degree approximation applied over 9-point intervals. The long-dashed curve is for $\mathbf{Z}$ computed using a seventh-degree approximation over 17-point intervals at central channels (in the vicinity of the peak of the Lorentzian) with a smooth transition to a third-degree approximation over 9-point intervals toward the lower and higher channels.

operator to this function:

$$\exp\left(\gamma\frac{d^2}{dx^2}\right)f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(1 - \gamma k^2 + \frac{\gamma^2 k^4}{2!} - \cdots\right)e^{ikx}\tilde{f}(k)\,dk$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}e^{-\gamma k^2}\tilde{f}(k)\,dk \qquad \text{(A11)}$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\tilde{g}(k)\tilde{f}(k)\,dk$$

where we have assumed that $\gamma > 0$, and $\tilde{g}(k)$ is the Fourier transform of the normalized Gaussian function

$$g(x) = \left(\frac{1}{2\pi\gamma}\right)^{1/2} e^{-x^2/2\gamma} \qquad \text{(A12)}$$

From the convolution theorem for Fourier transforms, the last expression in Eq. A11 is just the convolution of g with f, which has the effect of broadening f (i.e., increasing its second moment, if it exists) by an amount proportional to γ. Thus the operator

$$\mathbf{G}(\gamma) = \exp\left(\gamma\frac{d^2}{dx^2}\right) \qquad \text{(A13)}$$

has attributes of a broadening operator, for $\gamma > 0$. Moreover, this form for the convolution operation may be formally extended to the inverse, or deconvolution operation, because $\mathbf{G}^{-1}(\gamma) = \mathbf{G}(-\gamma)$, allowing us to define a general "width-adjustment" operator for both positive and negative values of γ. (In practice, the useful range of negative values of γ will be limited by the extreme sensitivity of the deconvolution operation to small variations in the input data.) We may derive an approximate matrix representation of the second-derivative operator using the same procedure that we used for the first-derivative operator leading to Eq. A6, except that we use the second-order coefficients $\alpha^2_{i,j}$ from the local polynomial approximation, instead of the linear coefficients $\alpha^1_{i,j}$. Alternatively, because the second-derivative operator is simply the square (i.e., two consecutive applications) of the first-derivative operator, we can use the square of the matrix $\mathbf{Z}$ of the first-derivative operator as an estimate of the second-derivative operator matrix; this approximation is, in fact, more stable for many applications. The exponential of a scalar multiple of this matrix yields a matrix representation of the operator $\mathbf{G}$ applicable to spectra. An example of the performance of this operator is shown in Fig. 8. The remarks made above concerning the implementation and limitations of the shift operator apply in this case as well. It should also be noted that

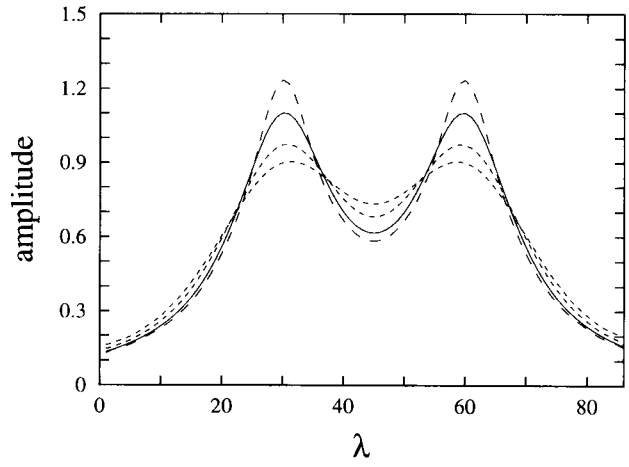

FIGURE 8 Demonstration of the effect of the "width-adjustment" operator described in Appendix A. This matrix operator is computed as $\exp(\gamma\mathbf{Y})$, where $\mathbf{Y}$ is the matrix representation of the second-derivative operator computed using a seventh-degree polynomial approximation (smoothly reduced to a second-degree approximation as the endpoints are approached) applied over 17-point intervals centered on each point. The solid curve is the sum of two analytical Lorentzian curves. The short-dashed and long-dashed curves are produced by multiplying this vector by the matrix operator computed for $\gamma = 10$ and 20 (broadening) and $\gamma = -5$ (narrowing), respectively. The useful range of negative values of γ is limited—in this case to values greater than ~ -8—by undesirable distortions (i.e., "ripples") of the resulting spectral shapes created by the deconvolution process.

convolution of a function with the normalized Gaussian $g(x)$, and therefore also to a good approximation the application of our synthetic operator **G**, leave the integral of the function (or the integrated area under the transformed spectrum) unchanged. Therefore a broadening (narrowing) of a spectral shape performed in this way will be accompanied by a decrease (increase) in peak amplitude.

APPENDIX B: METHODS FOR SOLVING THE LINEAR LEAST-SQUARES PROBLEM $\mathbf{A} \approx \Sigma_k \mathbf{B}^{(k)}\mathbf{F}\mathbf{C}^{(k)}$

In Explicit Relationships between Species Spectra it was shown that expressing certain "free" spectra as linear transformations of others leads to the multi-term least-squares problem

$$\mathbf{A} \approx \sum_{k=0}^{p} \mathbf{B}^{(k)}\mathbf{F}\mathbf{C}^{(k)} \tag{B1}$$

where the matrices $\mathbf{B}^{(k)}$ and $\mathbf{C}^{(k)}$ are constant matrices incorporating current species populations and linear transformation matrices, and **A** is a data matrix, possibly with contributions of other "pre-set" species spectra and corresponding populations removed. The unknown matrix **F** consists of the minimal set of "prototype" free spectra. When the right-hand side of Eq. B1 is a single term—i.e., the problem takes the form $\mathbf{A} \approx \mathbf{BFC}$—the least-squares solution is simply $\mathbf{F} = \mathbf{B}^{+}\mathbf{A}\mathbf{C}^{+}$, where $\mathbf{B}^{+}$ is the *pseudoinverse* of **B** (Lawson and Hanson, 1974).

Strategies to solve Eq. B1 when more than one term is present involve somehow re-casting it into the simpler form

$$\mathbf{\Phi}(\{\mathbf{B}^{(i)}\}, \{\mathbf{C}^{(i)}\}, \mathbf{A}) \approx \mathbf{\Gamma}(\{\mathbf{B}^{(i)}\}), \{\mathbf{C}^{(i)}\})\mathbf{\Psi}(\mathbf{A}) \tag{B2}$$

where $\mathbf{\Gamma}(\{\mathbf{B}^{(i)}\}, \{\mathbf{C}^{(i)}\})$ is some "supermatrix" constructed from the elements of all of the $\mathbf{B}^{(i)}$ and $\mathbf{C}^{(i)}$, and the operators $\mathbf{\Phi}$ and $\mathbf{\Psi}$ produce "supervectors" from the elements of the matrices on which they operate. This least-squares problem is then solved by the usual methods to produce the vector $\mathbf{\Psi}(\mathbf{A})$, from which the original matrix **A** may be recovered. One such method makes use of the Kronecker product (Horn and Johnson, 1991) of the matrices involved. To see how this is accomplished, we first write out the ijth component of the right-hand side of Eq. B1 (with some re-labeling of indices):

$$\begin{aligned}\sum_{p}\sum_{kl} B_{ik}^{(p)}F_{kl}C_{lj}^{(p)} &= \sum_{kl}\left(\sum_{p} B_{ik}^{(p)}C_{lj}^{(p)}\right)F_{kl} \\ &= \sum_{kl}\left(\sum_{p} C_{jl}^{(p)T}B_{ik}^{(p)}\right)F_{kl}\end{aligned} \tag{B3}$$

The k sum is over all columns of **B**, and the l sum is over all rows of **C**. We can visualize the double summation over k and l as a single summation over a composite index $\{kl\}$ in which k varies most quickly and assumes all values between 1 and the number of columns of **B**, and l varies most slowly and assumes all values between 1 and the number of rows of **C**; the composite indices are therefore ordered as $\{11\}, \{21\}, \ldots, \{12\}, \{22\}, \ldots$. Indexing the elements of the matrix **F** in this manner converts it into the column vector produced by successively concatenating all the columns of **F**. We represent this mapping of a matrix onto a column vector as *vec*(**F**). The final expression in Eq. B3 may then be rewritten:

$$\sum_{\{kl\}}\left(\sum_{p} C_{jl}^{(p)T}B_{ik}^{(p)}\right)[\mathrm{vec}(\mathbf{F})]_{\{kl\}} = \sum_{\{kl\}}\left(\sum_{p} K_{\{ij\};\{kl\}}^{(p)}\right)[\mathrm{vec}(\mathbf{F})]_{\{kl\}} \tag{B4}$$

where the matrices $\mathbf{K}^{(p)}$ are the so-called Kronecker products of $\mathbf{C}^{(p)T}$ and $\mathbf{B}^{(p)}$, which we indicate by $\mathbf{C}^{(p)T} \otimes \mathbf{B}^{(p)}$. The Kronecker or tensor product of two matrices, the first of which is $m \times n$ and the second of which is $q \times r$, is a $(mq) \times (nr)$ matrix made up of all possible products of elements of the two matrices. We can number the rows of the product using the composite index $\{ij\}$, where i varies most quickly and takes its values from the row indices of the second matrix, and j takes its values from the row indices of the first matrix. The columns of the product are similarly indexed using $\{kl\}$, where k varies most quickly and takes its values from the column indices of the second matrix, and l takes its values from the column indices of the first matrix. With this indexing scheme, the Kronecker product of two matrices **S** and **T** may be compactly defined as

$$(\mathbf{S} \otimes \mathbf{T})_{\{ij\};\{kl\}} = S_{jl}T_{ik} \tag{B5}$$

The least-squares problem in Eq. B4 may now be written as

$$\mathrm{vec}(\mathbf{A}) \approx \left(\sum_{p}(\mathbf{C}^{(p)T} \otimes \mathbf{B}^{(p)})\right)\mathrm{vec}(\mathbf{F}) \tag{B6}$$

which is in the form of a simple linear least-squares problem for the vector vec(**F**) and may be solved by the usual methods. The matrix **F** that solves the original problem is then produced by "un*vec*ing" this solution, which simply involves re-distributing the elements of the vector solution into the appropriate columns of the properly dimensioned matrix.

This method of solution is formally straightforward, but applying it in real analytical situations is often complicated by the large size of the Kronecker-product matrices required. Specifically, in treating spectral data sets consisting of n_λ spectral indices λ, the matrices $\mathbf{B}^{(p)}$ appearing in Eq. B1 have dimensions $n_\lambda \times n_\lambda$; if the data set consists of n_p experimental conditions, and the modeling treats n_s spectrally distinct species, then the matrices $\mathbf{C}^{(p)}$ are $n_s \times n_p$. The Kronecker product $(\mathbf{C}^{(p)} \otimes \mathbf{B}^{(p)})$ is then $(n_\lambda n_p) \times (n_\lambda n_s)$, which for typical spectral data sets may require many thousands of rows and many hundreds of columns.

We introduce here an alternative method that presents somewhat less severe size requirements for problems of this type. The least-squares problem of Eq. B1 requires minimization of the sum of squared deviations between the matrices on both sides. Using component notation, this summed deviation takes the form

$$\delta^2 = \sum_{ij}\left(A_{ij} - \sum_{p}\sum_{kl} B_{ik}^{(p)}F_{kl}C_{lj}^{(p)}\right)^2 \tag{B7}$$

and this expression must be minimized with respect to all of the elements of **F**. This requires that the derivative with respect to each component be zero, i.e.,

$$\frac{\partial}{\partial F_{mn}}\delta^2 = 2\sum_{ij}\left(A_{ij} - \sum_{p}\sum_{kl} B_{ik}^{(p)}F_{kl}C_{lj}^{(p)}\right)\left(-\sum_{q} B_{im}^{(q)}C_{nj}^{(q)}\right) = 0 \tag{B8}$$

After regrouping of factors this becomes:

$$\begin{aligned}\sum_{kl} F_{kl}\left(\sum_{pq}\left(\sum_{i} B_{ik}^{(p)}B_{im}^{(q)}\right)\left(\sum_{j} C_{ij}^{(p)}C_{nj}^{(q)}\right)\right) &= \sum_{ij} A_{ij}\sum_{q} B_{im}^{(q)}C_{nj}^{(q)} \\ \sum_{kl} F_{kl}\left(\sum_{pq}(\mathbf{B}^{(p)T}\mathbf{B}^{(q)})_{km}(\mathbf{C}^{(p)}\mathbf{C}^{(q)T})_{ln}\right) &= \sum_{q}(\mathbf{B}^{(q)T}\mathbf{A}\mathbf{C}^{(q)T})_{mn}\end{aligned} \tag{B9}$$

We now define a matrix **Q** and a vector **N**, which are both indexed using the type of composite indices described above:

$$\begin{aligned}Q_{\{mn\};\{kl\}} &= \sum_{pq}(\mathbf{B}^{(p)T}\mathbf{B}^{(q)})_{km}(\mathbf{C}^{(p)}\mathbf{C}^{(q)T})_{ln} \\ N_{\{mn\}} &= \sum_{q}(\mathbf{B}^{(q)T}\mathbf{A}\mathbf{C}^{(q)T})_{mn}\end{aligned} \tag{B10}$$

Equation B9 may then be written in the matrix form

$$\mathbf{Q}\,\mathrm{vec}(\mathbf{F}) = \mathbf{N} \tag{B11}$$

which may be solved for vec(**F**) using standard techniques, thereby also yielding the solution matrix **F**.

Using the same matrix sizes as were used in discussing the Kronecker product above, the square matrix **Q** has dimensions $(n_\lambda n_s) \times (n_\lambda n_s)$—i.e., the number of experimental conditions n_p does not enter into the size. Sizes of several hundred rows and columns are therefore expected, which makes this method often preferable to the Kronecker product method for applications of the type being considered here.

APPENDIX C: MAXIMUM-LIKELIHOOD ANALYSIS OF THE SPECTROSCOPIC MODELING PROBLEM

In the context of Spectroscopic Modeling in the Presence of Measurement Errors, we consider an experimental data set A_{ij} for which measurement errors (i.e., deviations from the "true" values A_{ij}^0 predicted by some model) have probability distributions which are functions $\rho_{ij}(A_{ij} - A_{ij}^0)$ only of the deviations $A_{ij} - A_{ij}^0$ themselves. For the spectroscopic modeling problem $\mathbf{A} \approx \mathbf{SM}(\{\xi_k\})$, a maximum likelihood estimate of the model parameters $\{\xi_k\}$ and species spectra **S** (giving rise to the predicted values A_{ij}^0) is produced by minimizing the functional

$$\begin{aligned} F(\{S\},\{\xi_k\}) &= \sum_{ij}(-\ln \rho_{ij}(A_{ij} - A_{ij}^0(\{\xi_k\}))) \\ &= \sum_{ij}\left(-\ln \rho_{ij}\left(A_{ij} - \sum_l S_{il}M_{lj}(\{\xi_k\})\right)\right) \end{aligned} \tag{C1}$$

with respect to all parameters and spectral amplitudes. For convenience, we write the probability densities as functions of the *squared* deviations and define the functions

$$\begin{aligned} f_{ij}((\delta A_{ij}(\{S\},\{\xi_k\}))^2) &\equiv f_{ij}\left(\left(A_{ij} - \sum_l S_{il}M_{lj}(\{\xi_k\})\right)^2\right) \\ &= -\ln \rho_{ij}\left(A_{ij} - \sum_l S_{il}M_{lj}(\{\xi_k\})\right) \end{aligned} \tag{C2}$$

The usual approach to minimizing $F = \Sigma f_{ij}$ with respect to all model and spectral parameters would use a general-purpose minimization algorithm (e.g., a simplex algorithm) while treating all parameters on an equal footing. In the spirit of Direct versus SVD-Based Analysis of the Data Matrix, we seek to perform the minimization of F with respect to the parameters $\{\xi_k\}$ and $\{S\}$ explicitly in the space of the $\{\xi_k\}$ only, with a set of optimal spectral amplitudes corresponding to a specific set $\{\xi_k\}$ determined by solving a corollary minimization problem using the appropriate fixed matrix **M**. This separation of the problem into linear parameters and nonlinear parameters, while still meaningful, does not offer obvious advantages, because no straightforward matrix-decomposition procedure is available to solve the linear subproblem in the general case analogous to that available for the simple least-squares problem. However, a matrix-based approach does appear to offer enhanced performance in solving this subproblem in at least some cases.

To develop the method, we assume that the matrix **M** is fixed, and we seek a minimum of the functional F with respect to the elements of the spectral matrix **S**. We require that the derivative of F with respect to each such element be zero:

$$\begin{aligned} \frac{\partial F}{\partial S_{pq}} &= \sum_{ij}\frac{\partial}{\partial S_{pq}} f_{ij}\left(\left(A_{ij} - \sum_l S_{il}M_{lj}\right)^2\right) \\ &= \sum_{ij} f'_{ij}\frac{\partial}{\partial S_{pq}}\left(A_{ij} - \sum_l S_{il}M_{lj}\right)^2 \\ &= -2\sum_j f'_{pj}\left(A_{pj} - \sum_l S_{pl}M_{lj}\right)M_{qj} = 0 \end{aligned} \tag{C3}$$

where f'_{ij} is the derivative of f_{ij} with respect to its argument (the squared residual $(\Delta A_{ij})^2$) evaluated at the values of the spectral amplitudes S_{ij} at the minimum. This dependence of f_{ij} on the amplitudes themselves complicates the solution of this system of equations. The simplest case, with $f_{ij}(x) = x$, is just the simple least-squares problem, for which Eq. C3 reduces to

$$\sum_j A_{pj}M_{qj} = \sum_{jl} S_{pl}M_{lj}M_{qj} \tag{C4}$$

In matrix form, this may be written as $\mathbf{AM}^T = \mathbf{SMM}^T$. The solution of this equation is $\mathbf{S}_m = \mathbf{AM}^+$, where $\mathbf{M}^+$ is the pseudoinverse of **M** (Lawson and Hanson, 1974).

The next simplest case for Eq. C3 is that with $f_{ij}(x) = x/\sigma_{ij}^2$, for a set of (generally distinct) constants σ_{ij}^2. The functional F is then equivalent to χ^2 as given in Eq. 42. This corresponds to each element A_{ij} having a measurement error which is normally distributed with variance σ_{ij}^2. We will discuss this case in some detail here, both because it is the immediate generalization of the simple least-squares case and because certain aspects of its treatment will be useful in approaching problems in which other functions f_{ij} are required. The system of equations C3 takes the form

$$\sum_j \frac{1}{\sigma_{pj}^2} A_{pj}\,M_{qj} = \sum_l S_{pl}\sum_j \frac{1}{\sigma_{pj}^2} M_{lj}\,M_{qj} \tag{C5}$$

Before discussing this system in general, we consider one useful special case for which the solution is easily constructed. Note that the inverse variances $1/\sigma_{ij}^2$ depend on the spectroscopic variable λ through the first index i and the remainder of the experimental conditions through the second index j. Suppose that these dependences may be factored out separately in the simplest way, i.e., a "rank-one" approximation of the inverse variances may be constructed:

$$\frac{1}{\sigma_{ij}^2} = \omega_i\eta_j \tag{C6}$$

Here the ω_i are elements of a vector $\boldsymbol{\omega}$, which depend only on the spectroscopic variable, and the η_j are elements of a vector $\boldsymbol{\eta}$, which depend only on the other experimental conditions. Then Eq. C5 may be written in a convenient form:

$$\begin{aligned} \omega_p\sum_j A_{pj}\eta_j M_{qj} &= \omega_p\sum_l S_{pl}\sum_j M_{lj}\eta_j M_{qj} \\ \sum_j \tilde{A}_{pj}\tilde{M}_{qj} &= \sum_l S_{pl}\sum_j \tilde{M}_{lj}\tilde{M}_{qj} \\ \tilde{A}_{ij} &= A_{ij}\eta_j^{1/2} \\ \tilde{M}_{ij} &= M_{ij}\eta_j^{1/2} \end{aligned} \tag{C7}$$

By analogy with the noise-free treatment this has the formal solution in terms of the new matrices given by

$$\mathbf{S} = \tilde{\mathbf{A}}\tilde{\mathbf{M}}^+ \tag{C8}$$

In many situations for which the experimental errors have been sufficiently well characterized to merit their explicit incorporation into the analysis, a decomposition of the form shown in Eq. C6 will likely be an adequate approximation. Should this not be the case, a direct solution of Eq. C5 is possible as follows: In the most general case the set of conditions in Eq. C5 may be rewritten as a set of matrix-vector equations, one for each row vector $\mathbf{s}_l$ of the spectral matrix $\mathbf{S}$:

$$\mathbf{a}_l = \mathbf{s}_l \mathbf{N}_l \tag{C9}$$

where we define the vector $\mathbf{a}_l$ and the symmetric matrix $\mathbf{N}_l$ in component form as

$$(\mathbf{a}_l)_m = \sum_j \frac{1}{\sigma_{lj}^2} A_{lj} M_{mj}$$
$$(\mathbf{N}_l)_{km} = \sum_j \frac{1}{\sigma_{lj}^2} M_{kj} M_{mj} \tag{C10}$$

As before, each of these equations has the formal solution

$$\mathbf{s}_l = \mathbf{a}_l \mathbf{N}_l^+ \tag{C11}$$

The construction of the vector $\mathbf{a}_l$ and matrix $\mathbf{N}_l$, and the solution of Eq. C11, for each of the n_λ rows of the spectral matrices may be expensive operations to perform each time one of the model parameters ξ (and therefore the matrix $\mathbf{M}$) is changed during an iterative model-fitting procedure. It may then be helpful (and as discussed in Spectroscopic Modeling may not necessarily be a severe compromise) to revert to the much more efficient solution of the "noise-free" linear least-squares problem (Eq. 10) during most of the calculation, or even make use of an approximate decomposition of the type shown in Eq. C6 to solve the problem shown in Eq. C5. The full solution method using Eqs. C9–C11 should only be strictly required in the vicinity of the final minimum.

We now return to the general system of equations in C3, which we write as

$$\sum_j f'_{pj} A_{pj} M_{qj} = \sum_l S_{pl} \sum_j f'_{pj} M_{lj} M_{qj} \tag{C12}$$

This system of equations would be formally equivalent to that in Eq. C5 (and therefore solvable by the same methods) if the factors f'_{pj} were in fact constants (analogous to $1/\sigma_{ij}^2$). This suggests a possible matrix-based iterative solution procedure: We first produce an initial estimate $\mathbf{S}_0$ of the matrix of spectral amplitudes S_{ij}, most conveniently by solving the system with all $f'_{pj} = 1$ using matrix decomposition methods. The squared residuals $(\Delta A_{ij})^2$ are then computed using this estimate, and new coefficients f'_{pj} are computed from these residuals. A new estimate $\mathbf{S}_1$ of the spectral amplitude matrix is produced by solving the new system, treating the coefficients f'_{pj} as constants, using for example the method outlined above, and formally replacing the constants $1/\sigma_{ij}^2$ with the current coefficients f'_{ij}. The new spectral amplitudes are used to compute new squared residuals, from which new coefficients f'_{ij} are obtained, and so forth. The procedure is iterated to self-consistency—i.e., until $\|\mathbf{S}_{n+1} - \mathbf{S}_n\| < \delta$ for some tolerance δ.

In using this procedure, there is little prior assurance that simple iteration will converge to a self-consistent set of spectral amplitudes. However, it may be shown that this specific form of iteration is approximately equivalent to minimizing the functional F using a Newton-Raphson procedure (Press et al., 1993) re-cast in the form of an iteratively reweighted least-squares problem (Green, 1984). Therefore, the present procedure will tend to be useful in those cases for which a Newton-Raphson analysis has acceptable convergence properties. A simple example is the weighted least-squares problem, for which the coefficients f'_{pj} are, in fact, constant and the measurement errors are, by assumption, normally distributed. In this case both the Newton-Raphson procedure and the present procedure by design produce a solution in a single iteration. For a given form of the functions f_{ij} the convergence properties of the sequence $\{\mathbf{S}_n\}$ are most easily determined by numerical simulation using data matrices and model population matrices representative of the type of system being studied. We have found that with simulated data sets using a double exponential error distribution ($\rho(\Delta A) \sim \exp(-|\Delta A|) \Rightarrow f((\Delta A)^2) \sim ((\Delta A)^2)^{1/2}$) or a Lorentzian error distribution ($\rho(\Delta A) \sim (1 + (\Delta A/\sigma)^2)^{-1} \Rightarrow f((\Delta A)^2) \sim \ln(1 + (\Delta A)^2/\sigma^2)$), the iterative procedure converges quite satisfactorily within a few iterations. Of course, when this method of solving the system of equations C12 for the spectral amplitudes is being used in conjunction with a general-purpose minimization algorithm in the space of model parameters $\{\xi_k\}$, some economies may be achieved by relaxing the convergence requirements on the sequences $\{\mathbf{S}_n\}$ during the early stages of the minimization and making them more stringent only as the final minimum in the space of model parameters is being approached.

The above-mentioned correspondence between this method of producing maximum-likelihood spectra and the method of iteratively re-weighted least-squares suggests an outline of a procedure for solving the maximum likelihood problem in the complete space of model parameters $\{\xi_k\}$ and species spectra $\mathbf{S}$. Each step of the procedure consists of the following: First, a few cycles of least-squares optimization are performed in the space of parameters $\{\xi_k\}$ using the current set of weights f'_{ij} assigned to the individual data points. (At the very beginning of the procedure, these weights may be set, for example, uniformly to 1.) During each cycle of this operation, the populations $\mathbf{M}$ are first computed from the current values of the model parameters, after which the corresponding spectral amplitudes $\mathbf{S}$ are determined by solving the system of equations C12 using the current weights. The least-squares optimization phase is followed by an iterative solution (which need not be carried to complete convergence) of the maximum-likelihood problem in the space of spectral amplitudes $\mathbf{S}$ alone using the above procedure and the current "best" set of populations $\mathbf{M}$. The final weights f'_{ij} from the latter phase then become the new set of weights to be used in the least-squares optimization phase of the next step. The steps (weighted least-squares $\rightarrow$ "best" $\mathbf{M} \rightarrow$ maximum-likelihood determination of $\mathbf{S} \rightarrow$ new weights $\rightarrow \ldots$) are repeated until the model parameters $\{\xi_k\}$ and spectra $\mathbf{S}$—as well as the weighting factors f'_{ij}—at the end of a step are consistent with those at the beginning of the step, to within some tolerance.

NOTES

1. Self-contained procedures for the complete orthogonal decomposition are provided, for example, in the LAPACK linear algebra subroutine library (Anderson et al., 1992). Also, the solution of the least-squares problem $\mathbf{B} \approx \mathbf{XA}$ for the matrix $\mathbf{X}$ is provided in the form of an elementary binary operation involving the matrices $\mathbf{A}$ and $\mathbf{B}$ in (among others) the commercial matrix analysis program MATLAB (The MathWorks, Inc., Natick, MA).

2. Strictly speaking, the least-squares problem $\mathbf{V'}^T \approx \mathbf{CM}$ is only equivalent to the problem $\mathbf{U'V'}^T \approx \mathbf{U'CM}$ if the columns of $\mathbf{U'}$ form an orthonormal set of vectors—i.e., $\mathbf{U'}^T\mathbf{U'} = \mathbf{I}$. For a general full-rank matrix $\mathbf{U'}$, the equivalent least-squares problem is $\mathbf{NV'}^T \approx \mathbf{NCM}$, where $\mathbf{N}$ is a nonsingular square matrix which relates $\mathbf{U'}$ to such a matrix $\mathbf{O}$ of orthonormal vectors through $\mathbf{U'} = \mathbf{ON}$. In light of this, it is a sensible practice to design the post-processing of the SVD output to ensure that the column vectors of $\mathbf{U'}$ form an orthonormal set so that the simpler equivalent problem (Eq. 15) may be treated and the need for mixing and/or differential weighting of the rows of $\mathbf{V'}^T$ in the least-squares fitting is avoided.

3. In many situations the spectrum of a species may consist of a superposition of several more elementary forms (e.g., distinct peaks centered at different values of λ). These are treated simply by constructing the $f_i(\lambda, \{\zeta_l\})$ as appropriately parametrized sums of the simpler "natural" lineshapes.

4. This is an especially simple case to treat, because the function g and its first derivative are both continuous at $x = 0$, independent of the choices of a and b. In general, relations among the various parameters would arise from the continuity conditions we choose to impose on functions defined in such a piecewise fashion, and some care may be required to ensure that a useful degree of parametric freedom remains.

5. $\Delta\mathbf{A}$ may be regarded as the fixed hypervolume of a region of the mn-dimensional space of $m \times n$ matrices which is small enough so that the probability density ρ is effectively constant over the region, and also small enough to contain a single measured $\mathbf{A}$. More careful consideration of this constant factor is not necessary, because it does not appear in the final expression (Eq. 40).

6. The use of such a simple probability function is intended to facilitate the search process, rather than to serve as a statistical characteristic of a final minimum. We focus here especially on the value of such simplicity in accelerating the production of optimal spectra consistent with a specified set of model parameters $\{\xi_k\}$. Statistical assessments of results of the modeling should always be based on the best available knowledge of the error distributions.

The author thanks James Hofrichter and Attila Szabo for critical reading of the manuscript.

REFERENCES

Anderson, E., Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen. 1992. LAPACK Users' Guide. SIAM, Philadelphia.

Beechem, J. 1992. Global analysis of biochemical and biophysical data. *Methods Enzymol.* 210:37–54.

Brent, R. P. 1973. Algorithms for Minimization Without Derivatives. Prentice-Hall, Englewood Cliffs, New Jersey.

Brown, K. M., and J. E. Dennis. 1972. Derivative free analogues of the Levenberg-Marquardt and Gauss algorithms for nonlinear least squares approximation. *Numer. Math.* 18:289–297.

Golub, G. H., and V. Pereyra. 1973. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM J. Numer. Anal.* 10:413–432.

Golub, G. H., and C. Van Loan. 1989. Matrix Computations, 2nd ed. Johns Hopkins University Press, Baltimore.

Green, P. J. 1984. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *J. R. Stat. Soc. B.* 46:149–192.

Hanson, R. J., and K. H. Haskell. 1982. LSEI and WNNLS: FORTRAN subprograms for solving least squares problems with linear equality and/or inequality constraints. *ACM Trans. Math. Software* 8:323–333.

Henry, E. R., and J. Hofrichter. 1992. Singular value decomposition: application to analysis of experimental data. *Methods Enzymol.* 210: 129–192.

Henry, E. R., C. M. Jones, J. Hofrichter, and W. A. Eaton. 1997. Can a two-state MWC allosteric model explain hemoglobin kinetics? *Biochemistry*, in press.

Hoff, W. D., I. H. M. van Stokkum, H. J. van Ramesdonk, M. E. van Brederode, A. M. Brouwer, J. C. Fitch, T. E. Meyer, R. van Grondelle, and K. J. Hellingwerf. 1994. Measurement and global analysis of the absorbance changes in the photocycle of the photoactive yellow protein from *Ectothiorhodospira halophila*. *Biophys. J.* 67:1691–1705.

Horn, R. A., and C. R. Johnson. 1991. Topics in Matrix Analysis. Cambridge University Press.

Jones, C. M., E. R. Henry, Y. Hu, C.-K. Chan, S. D. Luck, A. Bhuyan, H. Roder, J. Hofrichter, and W. A. Eaton. 1993. Fast events in protein folding initiated by nanosecond laser photolysis. *Proc. Natl. Acad. Sci. U.S.A.* 90:11860–11864.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. Optimization by simulated annealing. *Science* 220:671–680.

Lawson, C. L., and R. J. Hanson. 1974. Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, New Jersey.

Marquardt, D. W. 1963. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Industr. Appl. Math.* 11:431–441.

Nelder, J. A., and R. Mead. 1965. A simplex method for function minimization. *Comp. J.* 7:308–313.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1993. Numerical Recipes: The Art of Scientific Computing, 2nd ed. Cambridge University Press.

Shrager, R. I. 1970. Nonlinear regression with linear constraints: an extension of the magnified diagonal method. *J. Assoc. Comput. Mach.* 17:446–452.

Szu, H., and R. Hartley. 1987. Fast simulated annealing. *Phys. Lett. A* 122:157–162.