JOURNAL OF MOLECULAR SPECTROSCOPY 165, 255–264 (1994)

On the Least-Squares Fitting of Correlated Data: Removing the Correlation

JOEL TELLINGHUISEN

Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37235

In the most general case of linear least-squares fitting, the input data may be mutually correlated. This correlation is accommodated through a weight matrix **W** which contains off-diagonal elements. The existence of strong correlation in the data makes it difficult to interpret the fit residuals in terms of "good" and "bad" data. However, it is possible to transform the original correlated problem into an equivalent uncorrelated, unweighted fit through a simple linear transformation. This transformation is readily obtained from the orthogonal transformation which diagonalizes **W**.

INTRODUCTION

Doubtless the method of least squares is the most important data analysis tool in the spectroscopist's bag of tricks. The typical spectrum analysis project probably entails dozens if not hundreds of least-squares (LS) fits by the time the final tables are attached to the manuscript in submission for publication. Frequently the data may be analyzed and fitted in subsets for convenience, and the subsets are then merged in a global fit in the final stages of the analysis. For example, in the analysis of electronic spectra, it is customary to process different vibronic bands individually, at least initially, while the assignments are being bootstrapped into place. The raw data are usually completely uncorrelated; but the results of these initial fits are no longer "raw," which means that their employment in subsequent fits should properly utilize methods of correlated fitting.

While these points have been widely appreciated by spectroscopists for a number of years now (*1, 2*), there are aspects of correlated fitting which warrant further clarification. In the present work I offer comments on the simple question: How does one deal with residuals and recognize outliers in a fit of correlated data? My suggestion: Through a simple orthogonal transformation it is possible to convert the initially correlated data into an equivalent set of unweighted, uncorrelated data. The fits of these transformed data can then be treated and interpreted by all the standard methods of unweighted fitting. In particular, for normally distributed errors, the residuals should yield a Gaussian histogram, from which outliers may be identified in the usual fashion—by the number of σ's they differ from their calculated values.

THEORETICAL BACKGROUND

Correlated fitting methods can be used to treat both linear and nonlinear (i.e., iterative) problems. However, typical applications in the spectroscopic literature have been linear, so I will confine my attention to the linear case in the present work. In any linear fit the algebraic equations to be solved can be expressed in matrix notation as (*1–3*)

0022-2852/94 $6.00

$$\mathbf{X}^T\mathbf{W}\mathbf{X}\beta = \mathbf{X}^T\mathbf{W}\mathbf{y}, \tag{1}$$

where the design matrix $\mathbf{X}$ relates the n measured values in the column vector $\mathbf{y}$ to the p adjustable parameters β through

$$\mathbf{y} = \mathbf{X}\beta + \delta. \tag{2}$$

The design matrix may contain functions of the independent variables (e.g., the quantum numbers for the various assigned spectroscopic lines); the vector δ contains the n residuals (defined as observed-calculated here). The LS fit minimizes the quantity

$$S = \delta^T\mathbf{W}\delta. \tag{3}$$

Operationally Eqs. (1) are solved for the parameters β, which permits evaluation of the residuals from Eqs. (2) and thence the quantity S in (3).

In the simplest application of these relations, the $n \times n$ weight matrix $\mathbf{W}$ is diagonal, with every element identical, $w_{ii} = c$. This is the case of unweighted fitting. In typical situations the inherent statistical quality of the data is not well known in advance, so with no loss in generality the constant c may be taken as unity. Then the results of the fit are used to assess the quality of the data, by taking the estimated fit variance s^2 as S/f, where f is the degrees of freedom, $n - p$. The variance–covariance matrix $\mathbf{V}$ is then obtained as $s^2\mathbf{A}^{-1}$, where $\mathbf{A} = \mathbf{X}^T\mathbf{W}\mathbf{X}$. The p diagonal elements of $\mathbf{V}$ represent the estimated variances in the parameters, while the off-diagonal elements are the estimated covariances. These are referred to as *external* estimates in this procedure (*4*).[1] Both the parameters and their statistical errors can be seen to be independent of c by virtue of its cancellation in Eqs. (1) and in the two factors, s^2 and $\mathbf{A}^{-1}$. If all goes "according to Hoyle"—normally distributed errors and a correct fit model—a histogram of the residuals should approximate a Gaussian of form $\exp(-(\delta^2/2s^2))$.

If the data are known to come from more than one parent distribution of differing inherent quality, then $\mathbf{W}$ remains diagonal, but the elements w_{ii} are no longer all identical. Again the results of the fit are independent of an arbitrary scale factor in the weights, but one must know the *relative* precisions of the different data subsets. In this case it is customary to designate one subset as reference and assign $w_{ii} = 1$ for these data, with all other weights taken as s^2_{ref}/s_i^2 (hence the need for knowledge of the relative precisions). Then the quantity s^2 ($=s^2_{\text{ref}}$) obtained from the fit is more properly referred to as the "estimated variance for data of unit weight," and the estimated variance for a general point in the data set is s^2/w_{ii}. A histogram of the reduced residuals, $\rho_i \equiv \delta_i/s_i$, should now approximate the normal distribution for unit variance, proportional to $\exp(-\rho^2/2)$.

In favorable circumstances one may have good *a priori* information about the quality of the data and may be able to assess the experimental variance in advance of carrying out the fit to obtain the spectroscopic parameters. For example, if each data point is the result N_j of a counting measurement, it is usually reliable to estimate the variance as $\sigma_j^2 \approx N_j$, in accord with the statistics of counting. Line frequencies may perhaps have been extracted from a number of such points near a peak. From the statistics of many such determinations, one may have more-or-less reliable estimates of the inherent precision in the data. In that case it is appropriate to take the weights as $w_{ii} = \sigma_i^{-2}$ (where for present purposes I will leave alone the fact that for experimental data, the

[1] This designation apparently goes back to Birge (*5*). However, in some works (*6*) the meanings of "external" and "internal" in this context have been reversed.

true variance σ^2 must forever remain known only unto God). Then the sum designated above as S becomes an estimate of χ^2 for the fit, and the quantity s^2 becomes an estimated reduced chi-square, χ_f^2 for f degrees of freedom. The statistical expectation value for χ^2 is f (or equivalently, for χ_f^2 is unity). The experimental value obtained for χ^2 can thus be used to judge the "goodness of fit." If the fit is deemed "acceptable," then it is also appropriate to take for the variance simply $\mathbf{V} = \mathbf{A}^{-1}$, i.e., to *omit* the "quality factor" s^2, which is the same as using the theoretical expectation value for χ_f^2. In Deming's notation (*4*), this procedure yields internal estimates of the parameter variances and covariances.[1] Again, a histogram of the normalized residuals δ_i/σ_i should approximate the normal distribution.[2]

From a formal standpoint the procedure just described can be seen to be equivalent to rewriting Eqs. (1)–(3) as

$$\mathbf{Y}^{\mathrm{T}}\mathbf{Y}\boldsymbol{\beta} = \mathbf{Y}^{\mathrm{T}}\mathbf{z}, \tag{4}$$

$$\mathbf{z} = \mathbf{Y}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

and

$$S = \boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{\varepsilon}, \tag{6}$$

where $\mathbf{Y} \equiv \mathbf{W}^{1/2}\mathbf{X}$, $\mathbf{z} \equiv \mathbf{W}^{1/2}\mathbf{y}$, and $\boldsymbol{\varepsilon} \equiv \mathbf{W}^{1/2}\boldsymbol{\delta}$. Here $\mathbf{W}^{1/2}$ remains diagonal, with elements which are the square roots of the corresponding elements of $\mathbf{W}$; and Eqs. (4)–(6) are simply the equivalent of incorporating a factor of $w_{ii}^{1/2}$ $(=1/\sigma_i)$ in every measurement y_i. Thus this trivial transformation of the original data and design matrix has converted the original *weighted* fit into an equivalent *unweighted* fit.

In the most general case the weight matrix $\mathbf{W}$ contains off-diagonal elements. As was noted above, this occurs when the raw data have become "partially cooked" through some preliminary mathematical processing (e.g., conversion to combination differences or term values (*7*)), or are themselves the output of an LS fit to a subset of the raw data. The latter case is the one of importance in "merged fitting" methods (*1, 2*). In typical applications of merged fitting, $\mathbf{W}$ is block-diagonal, with the kth block coming from a preliminary fit of the kth subset of the data to a set of intermediate parameters $\boldsymbol{\beta}_k$. The merged fit of these intermediate parameters then yields the global parameters for the entire data set. The block $\mathbf{W}_k$ of $\mathbf{W}$ is proportional to the inverse of the variance–covariance matrix from the preliminary fit, $\mathbf{W}_k \propto \mathbf{V}_k^{-1}$. Although there are questions about the correct way for assessing the proportionality constant in this relation,[3] I will assume for present purposes that the "true" variances and covariances are somehow known, and I will take the proportionality constant as unity. The computational processing of the data via Eqs. (1)–(3) remains straightforward

[2] The randomness of the statistical sampling process may cause this histogram to be somewhat broader or narrower than the reference Gaussian, $\exp(-(\rho^2/2))$, in accordance with deviations of the experimental χ_f^2 from its expectation value of 1. However, the variance in χ_f^2 is $2/f$, so for reasonably large data sets which produce acceptable fits, this effect should be nominal.

[3] Coxon (*2*) has noted that in stepwise merging, exact agreement with a one-step merge is obtained only if the weight blocks are taken as proportional to the inverse dispersion matrices rather than to the inverse estimated variance–covariance matrices. These two matrices are related by the factor I have called s^2, designated as $\hat{\sigma}_k^2$ in (*1*). Since this quantity is proportional to the estimated reduced chi-square for the preliminary fit, it can be seen that Coxon's recommendation corresponds to using *internal* estimates of the errors in the data, while the method of Albritton *et al.* (*1*) employs *external* estimates. For reasons similar to those advanced by Coxon, it can be shown (*8*) that internal estimates are the more appropriate choice in one-step merges as well as in the stepwise merges investigated by Coxon.

in this case. However, because of the correlation in the intermediate parameters that are input to this merge, it is no longer reliable to interpret large residuals in the usual way—as representing outliers in the data. In fact it can happen that some of the major offenders in a merged fit seem innocuous from the standpoint of their residuals, and *vice versa*. This reality is a source of frustration in the final "data massaging" stages of a comprehensive spectrum analysis project.

Coxon (*2*) suggested the treatment of large data sets by a stepwise merging procedure. In this scheme, one could hope to recognize a questionable subset of the data by its deleterious effect on the statistics of the fit when it is added to an "agglomerating" merge. In this sense the stepwise merge becomes analogous to a direct global fit in which data are added subset by subset.

There are some drawbacks to the stepwise merge. First, in many cases one must start with a significant subset of the data just to get the fit underway. For example, if one seeks the conventional polynomial expansion parameters for the vibrational energy and rotational constant of a diatomic molecule, one must include enough bands to adequately sample the range of v encompassed by the data. With no easy way to "read" the residuals from a correlated fit, this may entail some trial-and-error fitting to arrive at a satisfactory "core" fit of acceptable data. Then, too, the stepwise merge itself can be tedious if many data subsets (e.g., vibronic bands) are involved. Of course, one may have to alter the fit model as more data are included, for example, increasing the polynomial order to accommodate a wider range of v levels in the case of a diatomic molecule.

These inconveniences could be alleviated if it were possible to interpret the residuals from correlated fits in a straightforward way. In fact this can be accomplished through a transformation analogous to that behind Eqs. (4)–(6). Because $\mathbf{W}$ is no longer diagonal, we cannot obtain $\mathbf{W}^{1/2}$ in the same simple way. However, since variance-covariance matrices (and hence $\mathbf{W}$) are symmetric and positive definite, we *can* find an orthogonal transformation that will diagonalize $\mathbf{W}$, such that for the kth block (*3*),

$$\mathbf{O}_k^{\mathrm{T}}\mathbf{W}_k\mathbf{O}_k = \mathbf{D}_k, \tag{7}$$

in which $\mathbf{D}_k$ is diagonal with all elements positive. Then Eqs. (4)–(6) hold once again, provided we define

$$\mathbf{Y} \equiv \mathbf{D}^{1/2}\mathbf{O}^{\mathrm{T}}\mathbf{X}; \tag{8a}$$

$$\mathbf{z} \equiv \mathbf{D}^{1/2}\mathbf{O}^{\mathrm{T}}\mathbf{y}; \tag{8b}$$

$$\boldsymbol{\varepsilon} \equiv \mathbf{D}^{1/2}\mathbf{O}^{\mathrm{T}}\boldsymbol{\delta}; \tag{8c}$$

with each block of $\mathbf{O}$ being defined by (7). The elements of $\mathbf{D}$ are the eigenvalues of $\mathbf{W}$, obtained via the diagonalizing orthogonal transformation. The mechanics of obtaining this transformation will already be familiar to most spectroscopists, from the importance of this problem in quantum mechanics. As a result of this transformation, *the original correlated fit has been converted into an equivalent simple unweighted fit.* Accordingly its residuals $\boldsymbol{\varepsilon}$ can now be interpreted in the usual way.

EXAMPLES

To illustrate these points I have chosen some examples from a recent analysis of the $D' - A'$ transition in IBr (*9*). The primary data in that study came from high-resolution data for 71 $v' - v''$ bands recorded in absorption and 19 in emission. The

nature of the absorption experiments made it particularly expedient to record and analyze the bands individually, and so the emission data were treated likewise, yielding preliminary results in the form of a band origin, upper-, and lower-state rotational constant and their associated 3×3 variance–covariance matrix for each of the 90 bands. Earlier work with similar data had shown that the correlation between the band origin and the rotational constants is of no real importance in the analysis of the vibrational parameters (*10, 11*), and so the vibrational fitting was done separately from the rotational fitting. A large number of low-precision band origins were incorporated in the final vibrational analysis, which became thereby a simple weighted fit. On the other hand, the high correlation between the two rotational constants, B' and B'', obtained for a given band made correlated fitting essential in the rotational analysis.

Continuing work on the IBr $D' \rightarrow A'$ emission spectrum (*12*) has included some global fitting of assigned rotational lines in 23 bands together with the high- and low-resolution band origins from Ref. (*9*), suitably weighted. Figure 1 illustrates the fit residuals for the low-resolution origins and a representative subset of the emission data containing a comparable number of data points. The abscissae of the two histograms differ by a scale factor of 10, reflecting the factor of ~ 7 greater precision in the rotational data. In a simple illustration of the point behind Eqs. (4)–(6), Fig. 2 shows both sets of residuals in Fig. 1, now rescaled by dividing by their respective σ's so that they can be histogrammed together and compared with the normal distribution.

In the correlated fits of the rotational constants from the band-by-band analyses to the usual expressions for B'_v and B''_v, the weight matrix contains 90 2×2 blocks along its diagonal. For the sake of illustration, let us suppose that B'_v is to be expressed as $B'_v = B'_e - \alpha'_e u' + \gamma'_e u'^2$, while B''_v will be fitted to $B''_v = B''_e - \alpha''_e u''$, with $u \equiv v +$

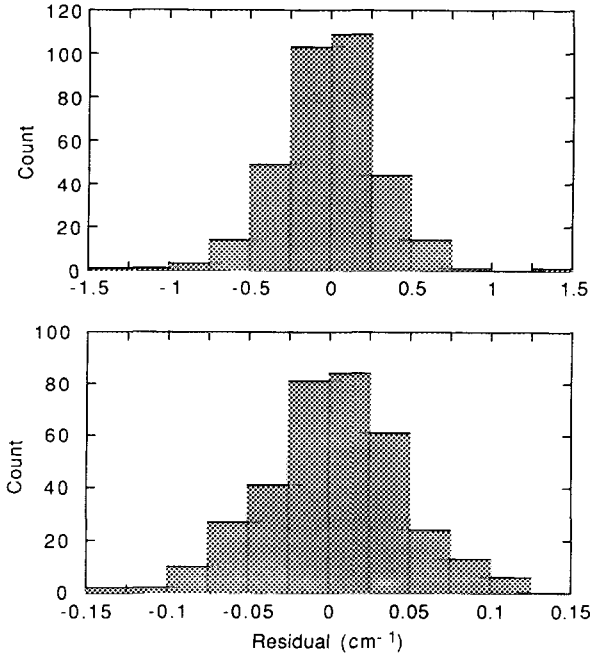

FIG. 1. Histograms of residuals from a weighted least-squares fit of IBr D'–A' data including low-resolution estimates of band origins (top, 340 points, $\sigma \approx 0.32$ cm^{-1}) and high-resolution rotationally assigned data (bottom, 351 points, $\sigma \approx 0.044$ cm^{-1}). Note different abscissae scales in the two plots.

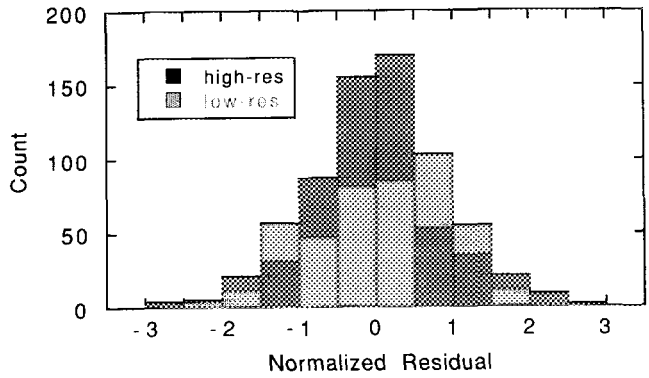

FIG. 2. Representation of data in Fig. 1 on a single histogram. Now all residuals have been converted to a common dimensionless scale through division by their respective estimated standard deviations.

$\frac{1}{2}$. Then the design matrix **X** and the vectors **y**, β, and δ take the form illustrated in Table I. Note that the odd-numbered data and their residuals represent B' values, while the even-numbered ones represent B'' values.

Figure 3 illustrates the residuals from this correlated fit in a manner which emphasizes the strong correlation between the B' and B'' values from each analyzed band. To put these on a common footing, I have scaled each entry by the square root of its respective diagonal element in **W**. If this were an uncorrelated fit with weights based on internal estimates of the precisions, this procedure would accomplish the same result as that

TABLE I

Illustration of the Matrix Equations for a Fit of 2 × 2 Correlated Rotational Constants from Band-by-Band Analyses of 90 v'–v'' Bands in an Electronic Transition in a Diatomic Molecule[a]

$$\begin{bmatrix} B'(1) \\ B''(1) \\ B'(2) \\ B''(2) \\ B'(3) \\ B''(3) \\ B'(4) \\ \vdots \\ B''(90) \end{bmatrix} = \begin{bmatrix} 1 & -u'(1) & u'(1)^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & -u''(1) \\ 1 & -u'(2) & u'(2)^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & -u''(2) \\ 1 & -u'(3) & u'(3)^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & -u''(3) \\ 1 & -u'(4) & u'(4)^2 & 0 & 0 \\ & & \vdots & & \\ 0 & 0 & 0 & 1 & -u''(90) \end{bmatrix} \cdot \begin{bmatrix} B_e' \\ \alpha_e' \\ \gamma_e' \\ B_e'' \\ \alpha_e'' \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \vdots \\ \delta_{180} \end{bmatrix}$$

[a]Numbers in parentheses designate the band number; $u \equiv v + 1/2$.

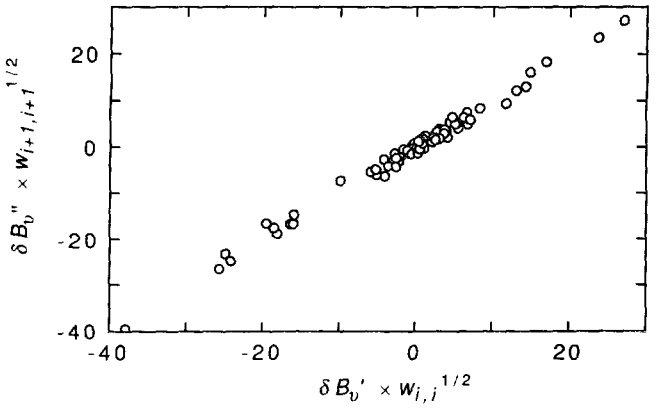

FIG. 3. Scatterplot of residuals from correlated fit of rotational constants from 90 v'–v'' bands to a quartic polynomial for $B_{v'}$ and the same for $B_{v''}$ (see Ref. (9)). Each point represents the residuals in both $B_{v'}$ and $B_{v''}$ for a given band, after scaling by the square root of the respective diagonal elements of **W**.

of Fig. 2. On this basis a number of points appear to deviate from their calculated values by more than 20σ! However, as Fig. 4 shows, this procedure is naive: Obviously the more appropriate normalization is division of each residual by its respective σ (the square root of the appropriate diagonal element of **V**).

Both Figs. 3 and 4 illustrate the strong correlation between B' and B'', and Fig. 4 shows a preponderance of points clustered within ± 1 units of the origin along both axes. However, it is improper to interpret these distributions this far and no further. Because of the high positive correlation between B' and B'' for each band, it is possible for a point right *on* the diagonal but far from the origin to make a smaller contribution to S than a point *off* the diagonal but near the origin. For example the point at (0.94, 0.48) contributes about 50% more to S than does the point at (1.69, 1.71).

Figure 5 shows the same scatterplot after the residuals have been "orthonormalized" via Eqs. (7) and (8c).[4] (Of course identical results are obtained by transforming the input data and the design matrix at the outset using (8a) and (8b) and then conducting a simple, unweighted fit.) Figure 5 now shows clearly which points exceed 2 or 3σ in one or both of the transformed variables. In particular, note that only the single point in the lower left corner exceeds 2σ in both of its coordinates, and in this case it also exceeds 3σ in both coordinates. Clearly this point makes the greatest single contribution to S, but whether it should have been deleted from the fit is still debatable. In Fig. 4 the corresponding point is the left-most point, and it falls both a long way from the origin and a good distance off the diagonal.

It is of interest to examine the orthogonalizing transformation in this case. It turns out that for every band the transformation comes close to a $\pi/4$ rotation, which means that it produces something close (but not identical) to the coefficients $(B' + B'')$ and $(B' - B'')$, which occur directly when P- and R-branch data are fitted to the "m-representation" of Herzberg (*13*, *14*). In the present case the differences are determined much more precisely, and so accordingly their weights are typically a factor of $\sim 10^2$–

[4] This language is suggested by similar operations in quantum mechanical calculations employing basis sets. The orthonormalization of a basis can be accomplished through a transformation analogous to that of Eqs. (7) and (8), in which the overlap matrix of the original basis fills the role played by **W** here.

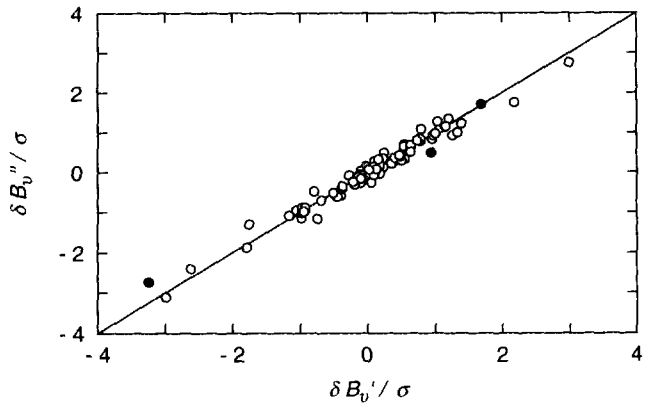

FIG. 4. Results of same correlated fit used to produce Fig. 3, but with each residual δ_i now divided by its respective σ_i from the band-by-band fits. Filled points are those which are given special attention in the text.

10^3 larger than those for the sums.[5] If just the 90 "differences" are used to obtain the rotational polynomial coefficients, they do about as well as the entire data set, with an exception: Because all of these are nearly simple differences between B' and B'', they can determine the *difference* $B_e' - B_e''$ precisely, but they yield only crude estimates of the two individual B_e values. On the other hand, inclusion of just two "sum" values—one for a low-v' band and one for a low-v'' band—serves also to pin down the two B_e values, although with precision substantially reduced from that obtained in the original correlated fit.

There is another quantity which can provide insight into the role of outliers in a data set, and that is the termwise accumulation of the sum of squared residuals S. This quantity is illustrated in Fig. 6 for the correlated fit under discussion here. Note

[5] However, the differences are also much smaller in magnitude, so that after multiplication by the factor $\mathbf{D}^{1/2}$ the resulting "orthonormalized" quantities are more nearly comparable in value.

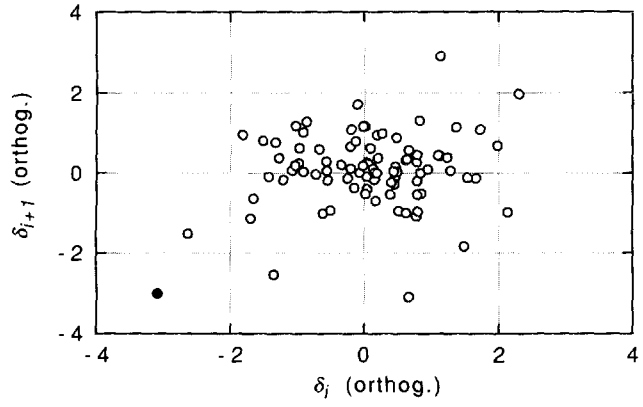

FIG. 5. Scatterplot of least-squares residuals from same correlated rotational fit, after transformation to an equivalent "orthonormal" set. The filled point at lower left represents the only band for which both residuals exceed 2σ.

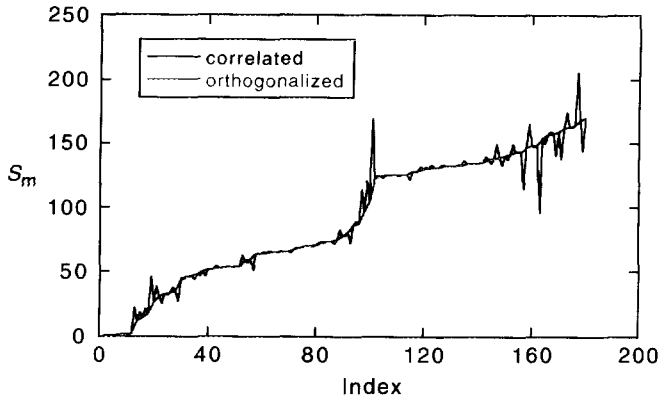

FIG. 6. Plot of accumulating sum of squared residuals, $S_m = \sum_{i=1}^{m} (\delta_i/\sigma_i)^2$, as a function of the running index m, for the correlated rotational fit of Figs. 3–5.

that because of the off-diagonal terms in **W** it is now possible for the individual terms *within* a block of correlated input to contribute *negatively* to S. Of course each block must make a net positive contribution, and so at the conclusion of each correlated block (every second point in this case) the results from Eq. (3) agree with those obtained using Eqs. (6) and (8c). It is easy to identify the big contributors to S in this plot from the sharp rises they produce. However, every proper data set must have some data which contribute much more appreciably to S than do the majority of points, and so some further work is still necessary to determine whether a particular large rise in such a plot can be justifiably attributed to a deletable outlier.

CONCLUSION

When the input data to a least-squares fit are correlated, this correlation must be taken into account in the fitting. Usually this is accomplished by defining the weight matrix **W** as the inverse of the variance–covariance matrix of the input data. The resulting off-diagonal elements in **W** make the interpretation of the residuals from a correlated fit less than straightforward. However, through a linear transformation which diagonalizes **W**, it is possible to convert a correlated fit into an equivalent simple unweighted fit, all the properties of which are interpretable in the usual ways. In particular, following such a transformation statistical outliers can be identified readily from the magnitude of their residuals.

Although the present work has focused on the residuals in correlated fits, there is at least one other valuable use for the orthogonalizing transformation. That is the situation where correlated data are to be incorporated in a very large merge, possibly one which contains many uncorrelated data. In such cases, the computational bookkeeping is greatly simplified if **W** is diagonal. Of course this does not come without a price, namely the need to obtain and store the transformation matrix **O**.

RECEIVED: November 30, 1993

REFERENCES

1. D. L. ALBRITTON, A. L. SCHMELTEKOPF, AND R. N. ZARE, *J. Mol. Spectrosc.* **67**, 132–156 (1977).
2. J. A. COXON, *J. Mol. Spectrosc.* **72**, 252–263 (1978).

3. W. C. HAMILTON, "Statistics in Physical Science," Ronald Press, New York, 1964.
4. W. E. DEMING, "Statistical Adjustment of Data," Dover, New York, 1964.
5. R. T. BIRGE, *Phys. Rev.* **40,** 207–227 (1932).
6. P. R. BEVINGTON AND D. K. ROBINSON, "Data Reduction and Error Analysis for the Physical Sciences," 2nd ed., McGraw-Hill, New York, 1992.
7. D. L. ALBRITTON, W. J. HARROP, A. L. SCHMELTEKOPF, R. N. ZARE, AND E. L. CROW, *J. Mol. Spectrosc.* **46,** 67–88 (1973).
8. J. TELLINGHUISEN, unpublished.
9. X. ZHENG, M. C. HEAVEN, AND J. TELLINGHUISEN, *J. Mol. Spectrosc.*, in press.
10. X. ZHENG, S. FEI, M. C. HEAVEN, AND J. TELLINGHUISEN, *J. Mol. Spectrosc.* **149,** 399–411 (1991).
11. X. ZHENG, S. FEI, M. C. HEAVEN, AND J. TELLINGHUISEN, *J. Chem. Phys.* **96,** 4877–4883 (1992).
12. D. T. RADZYKEWYCZ, C. D. LITTLEJOHN, M. B. CARTER, J. O. CLEVENGER, J. H. PURVIS, AND J. TELLINGHUISEN, submitted for publication.
13. G. HERZBERG, "Spectra of Diatomic Molecules," Van Nostrand, Princeton, NJ, 1950.
14. D. L. ALBRITTON, A. L. SCHMELTEKOPF, J. TELLINGHUISEN, AND R. N. ZARE, *J. Mol. Spectrosc.* **53,** 311–314 (1974).