

103

AN INTRODUCTION TO THE LEAST-SQUARES FITTING
OF SPECTROSCOPIC DATA

D. L. Albritton and A. L. Schmeltekopf

Aeronomy Laboratory
NOAA Environmental Research Laboratories
U. S. Department of Commerce
Boulder Colorado 80302

and

R. N. Zare

Department of Chemistry
Columbia University
New York, New York 10027

(to be published in Molecular Spectroscopy: Modern Research,
Volume II, Academic Press)

Contents

	<u>Page</u>
A. FITTING EQUATIONS TO DATA	1
1. Purposes	1
(a) <u>Curve fitting</u>	1
(b) <u>Model fitting</u>	1
2. Methods	4
B. NOTATION	9
C. ATTRIBUTES OF THE LEAST-SQUARES METHOD	12
1. Properties of Least-Squares Estimates	12
(a) <u>Unbiased</u>	13
(b) <u>Linear</u>	14
(c) <u>Minimum-variance</u>	14
2. Assumptions of the Least-Squares Method	15
(a) <u>Perfect model</u>	16
(b) <u>Linear model</u>	16
(c) <u>Zero-mean errors</u>	17
(d) <u>Finite-variance errors</u>	18
(e) <u>Known relative error variances</u>	19

	<u>Page</u>
C-2. (f) <u>Known relative error covariances</u>	19
(g) <u>Independent variable without error</u>	20
D. UNWEIGHTED LEAST-SQUARES FORMULATION	22
1. Statement of the Problem	22
2. Least-Squares Estimates	25
(a) <u>Estimates of the parameters</u>	25
(b) <u>Estimate of the variance of the random measurement errors</u>	29
(c) <u>Estimates of the variances and covariances of the estimated parameters</u>	32
3. Confidence Limits	47
(a) <u>Normal distribution</u>	47
(b) <u>Construction of the confidence limits</u>	51
E. WEIGHTED LEAST-SQUARES FORMULATION	56
1. Statement of the Problem	56
2. Variance-Covariance Matrix of the Measurement Errors	61
3. Estimated Molecular Parameters	66
4. Estimated Variance	67

	<u>Page</u>
E. 5. Estimated Variances and Covariances of the Molecular Parameters	68
F. CORRELATED LEAST-SQUARES FORMULATION	73
1. Statement of the Problem	73
2. Least-Squares Estimates	79
G. REPORTING LEAST-SQUARES RESULTS	84
1. Essential Information	84
(a) <u>Basic observations</u>	84
(b) <u>Weight matrix</u>	85
(c) <u>Model</u>	86
(d) <u>Estimated molecular constants</u>	86
(e) <u>Estimated variance of the measurement errors</u>	90
(f) <u>Degrees of freedom and standard errors</u>	91
2. Useful Information	95
(a) <u>Estimated variance-covariance matrix</u>	95
(b) <u>Normality</u>	98
H. CONCLUDING REMARKS	99

	<u>Page</u>
ACKNOWLEDGMENTS	101
APPENDIX: DERIVATION OF THE LEAST-SQUARES EQUATIONS . . .	102
1. Estimated Molecular Constants	102
2. Gauss-Markov Theorem	107
3. Estimated Variance of the Measurement Errors . . .	108
4. Estimated Variance-Covariance Matrix	111
5. Correlated Least-Squares Estimates	112
REFERENCES	118

A. FITTING EQUATIONS TO DATA

1. Purposes

Fitting equations to data is a household task in molecular spectroscopy and its purposes are usually twofold:

(a) Curve fitting. From a practical standpoint, it is simply convenient to summarize adequately a large set of spectroscopic data by some simple analytical function that has only a few adjustable parameters. An equation is a compact and easily remembered substitute for the data. This is data reduction in the pragmatic sense of "to diminish in number." Furthermore, the representation of the numerical data set by an appropriate equation is one useful aid to interpolation, differentiation, integration, or calibration. A familiar example is the fitting of a low-order polynomial to the standard lines on a photographic plate for the purpose of interpolating the wavelengths of the spectrum of interest. Since the somewhat arbitrarily chosen function used in curve fitting is generally not derived from "laws of nature," no physical significance is usually attributed to the values obtained for the adjustable parameters and the uncertainty of these values is seldom of interest.

(b) Model fitting. However, the fitting of equations to

spectroscopic data is usually approached from a more fundamental standpoint. Namely, the adjustable parameters represent molecular "constants" and the equations are algebraic representations of a physical model involving these constants as adjustable parameters. In fitting these particular equations to spectroscopic data, (i) the values obtained for the parameters are considered as estimates of the forever-unknown "true" values (if such can be said to exist!) of the molecular constants and (ii) the estimated uncertainties of these values are taken as measures of how well the estimated values could be expected to represent the "true" values. Thus, while the mathematical mechanics of model fitting do not differ too much from those of curve fitting, the functional form is now no longer arbitrary and the quality of parameter estimation emerges as a consideration.

Parameter estimation is clearly a more sophisticated and difficult task than simply curve fitting, but it is of considerably more applied and theoretical utility. Precise values of the parameters in an appropriate molecular model can perform a variety of tasks. For example, they can be used very successfully as predictors to enlarge a set of spectroscopic data by an incremental iterative technique of (i) fitting and (ii) interpolation to missing internal members of the set and, with care, extrapolation to new external members. The estimated uncertainties of the values of

the molecular parameters can be used to place limits within which it is realistic to search for the missing or new members. Furthermore, a molecular parameter estimated from one model can be used in another model that contains this parameter. For example, G_v and B_v values estimated by fitting a molecular Hamiltonian to observed line positions can be used to compute potential curves and hence to predict line intensities. Finally, an estimated "experimental" molecular parameter obtained from fitting a model to data can be compared to an ab initio value and the estimated uncertainty of the "experimental" value used as a measure of the agreement.

Spectroscopy is a mature scientific discipline in which the main structure of the models that are fitted to the data have a very sound theoretical basis. Furthermore, the number of measured line positions usually far exceeds the number of molecular parameters to be estimated. Consequently, these attributes are conducive to the meaningful fitting of equations to spectroscopic data.

2. Methods

There are, of course, a wide variety of techniques available for fitting equations to data. These range from straightforward graphical methods that use straight edges and french curves to sophisticated statistical methods that require digital computers. In spectroscopy, prior to the widespread availability of electronic computers, equations were fit to data by graphical methods (Herzberg, 1950, pp. 168-193). This method is inherently subjective, which on the one hand, can be an advantage since it can employ a spectroscopist's sound intuition and intimate familiarity with the data at each step of the analysis but which, on the other hand, can be a disadvantage since it can sometimes lead to unduly optimistic estimates of the precision of the resulting molecular constants (an example is given by Albritton, Harrop, Schmeltekopf, and Zare, 1973b). Furthermore, the graphical method is conceptually simple and visually advantageous, but it is difficult and tedious to apply to large quantities of spectral data to be fitted by complex (e.g., nonlinear) models. Thus, while graphical methods are indispensable in the early and intermediate stages of spectroscopic data reduction, increasingly, spectroscopists are using statistical methods to obtain their reported values of molecular constants.

There are many different statistical methods for fitting equations to data for the purpose of parameter estimation — least absolute deviation, least squared deviation, maximum likelihood, minimum chi-squared, — each method having a different "goodness of fit" criterion and all of the methods varying in sophistication and popularity. The oldest and by far the most widely used statistical estimation procedure is the method of least-squared deviation, or "least squares." In this method, the values that minimize the sum of the squared deviations of the calculated and observed quantities are taken as the estimates of the molecular constants of the model. When applied with an electronic computer, the method of least squares has the considerable advantage of being able, for a large amount of data and for complicated models, to provide quickly estimates of the molecular constants and their associated uncertainties, both with many well-defined and desirable attributes. However, the least-squares method (like all statistical methods) also has the distinct disadvantage of being able, even for those situations where the method is in varying degrees inapplicable, to nevertheless furnish numbers that superficially appear correct. This disadvantage is compounded when it is associated with the aura of infallibility often falsely attributed to numbers generated by an electronic computer.

In spite of this disadvantage, however, the method of least squares has played, and will continue to play, a valuable role in

the analysis of spectroscopic data. No doubt, other methods will see increasing application, but, as Tukey (1974, p. 5) points out, little information is in the hands of the physical scientist on general replacements for least squares. This dearth is in contrast to the wealth of statistical research papers, mathematical treatises, general textbooks, and "how-to-do-it" handbooks that are currently available on the method of least squares. Furthermore, many least-squares computer programs have been written for general use and are thus available for those who hate computers but want to use them. (In statistical texts, a least-squares fit is referred to frequently as a "regression," a term actually having a wider definition and curious historical origins, see Bennett and Franklin, 1954, p. 37.)

The goal of this Chapter is to provide an introduction to the least-squares method as it relates to common spectroscopic problems. The presentation here will be long on spectroscopic applications and short on statistical proofs, which are placed in the Appendix. The reasons for these choices are twofold. First, there are a number of excellent texts that consider the application of the method of least squares to broad scientific disciplines like physics or chemistry, but there appears to be no such text solely devoted to the introductory application of the least-squares method to molecular spectroscopy. Secondly,

too much rigor frequently leads to rigor mortis.

Unfortunately, space, time, and energy do not permit inclusion here of the interesting topic of the relative merits of the various types of least-squares fits used in spectroscopic data reduction: (i) one-state fits; i.e., the molecular constants of each state being determined separately from least-square fits to, e.g., combination differences or term values of that state; and (ii) two-state fits; i.e., the molecular constants of upper and lower states, or molecular constants of one state and term values of the other, being determined simultaneously from a direct fit to the measured line positions. This subject has received a lot of attention recently (Albritton, Harrop, Schmeltekopf, Zare, and Crow, 1973; Åslund, 1974; Plíva and Telfair, 1974; and Albritton, Schmeltekopf, Tellinghuisen, and Zare, 1974) and it may even be premature to review it here. However, the equally interesting and current topic of the relative merits of separate band-by-band and simultaneous multi-band fitting (Zare, Schmeltekopf, Harrop, and Albritton, 1973; Coxon, 1974; Plíva and Telfair, 1974; Albritton, Schmeltekopf, Zare, and Czarny, 1975; and Saenger, Zare, and Mathews, 1975) is considered in Sec. F, where it is shown that one often can achieve the advantages of both.

After a brief outline of the notation that will be used,

the format of this Chapter will be as follows. First, there is a general overview of the least-squares method, attempting to answer two necessary introductory questions, "What is so good about least squares?" and "When can least squares be applied?" Secondly, the following Sections will then concentrate on the application of the mathematical machinery of the least-squares method, beginning with the simple, most-restricted, and most-frequently applied least-squares formulation and ending with the general, least-restrictive, and least-frequently applied formulation. Lastly, the Chapter will conclude with comments on reporting least-squares results without loss of information.

B. NOTATION

The least-squares method is formulated here in vector and matrix notation, which has many advantages over the older algebraic notation involving explicit subscripts and summations. The equations are compact and easily manipulated. Furthermore, they apply to any least-squares fit, regardless of the number of measurements or parameters to be estimated. Finally, matrix manipulation is a natural form for computer programming, since many of the matrix operations, e.g., inversion, are generally available in standard subroutines at most computer centers. Lastly, almost all of the modern texts on the theory and practice of parameter estimation use matrix notation.

Throughout this Chapter, the matrix definitions and operations will be the standard ones; hence, only the symbols are defined here. A speaking acquaintance with matrix algebra is assumed herein. However, useful aids to dispelling possible rust on this subject are the sections on matrix algebra as it pertains to the method of least squares in Draper and Smith (1966, Chap. 2), Scheffé (1959, Apps. I and II), and Bard (1974, App. A), the latter two being more comprehensive but less elementary. The matrix notation of this Chapter will be:

(i) Boldface lower case letters, both Roman and Greek, will denote column vectors, e.g., \mathbf{y} , $\mathbf{\beta}$.

(ii) Boldface upper case letters, both Roman and Greek, will denote matrices, e.g., \mathbf{X} , $\mathbf{\Theta}$.

(iii) Roman letters denote matrices and vectors of known quantities, e.g., measurements, independent variables.

(iv) Greek letters denote matrices and vectors of unknown quantities to be estimated, e.g., molecular constants, measurement errors.

(v) The superscript T will denote the transpose of a vector or matrix, e.g., \mathbf{y}^T , \mathbf{X}^T .

(vi) The superscript -1 will denote the inverse of a (square) matrix, e.g., $(\mathbf{X}^T\mathbf{X})^{-1}$.

The statistical notation of this Chapter will be mercifully simple:

(i) The circumflex \wedge is the standard statistical notation for the estimated value of a parameter, e.g., $\hat{\mathbf{\beta}}$ is the column vector of numbers that are the estimates for the column vector of parameters $\mathbf{\beta}$. Here we reserve the circumflex to denote specifically the minimum-variance linear unbiased (MVLU) estimates defined in Sec. C-1.

(ii) The function $E(\)$ denotes the expectation value of the quantity in parentheses, i.e., the average value that

would be obtained by infinitely many repetitive samples, all made identically except for random measurement error.

C. ATTRIBUTES OF THE LEAST-SQUARES METHOD

1. Properties of Least-Squares Estimates

The popularity of the method of least squares is based on the desirable properties of the estimates that it furnishes. In the simplest format, if a set of n measurements

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad (C-1)$$

are to be fitted to an appropriate model that is linear in the m unknown parameters ($m < n$)

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix} \quad (C-2)$$

to determine estimates $\hat{\beta}$ of the unknown "true" values of β , then the method of least squares offers values that are the most precise (i.e., minimum-variance) unbiased estimates that are linear functions of the measurements. These properties are a consequence of the celebrated Gauss-Markov theorem, which is proved in the Appendix. It is easier to recognize that the adjectives "minimum-variance linear unbiased (MLU)", are indeed desirable properties of estimates after some elaboration of each one. The following comments are adapted from those of Daniel and Wood (1971, p. 7), except where noted.

(a) Unbiased. The "unbiased" property indicates that if the same appropriate equations could be fitted with the least-squares method to successive sets of data, each set taken under identical experimental circumstances and subject only to random measurement errors, then these varying sets of estimates will average out in the long run to the "true" values; i.e., in statistical notation, the expectation $E(\hat{\beta}) = \beta$. Thus, the linear least-squares estimation procedure itself does not introduce any bias, i.e., systematic error, into the estimates. This often-unfamiliar property of least-squares estimates is not a trivial one. There are other estimation procedures that do in fact introduce bias even for a perfect model. For example, the method of maximum likelihood does not necessarily give unbiased estimates for small

data sets (Bard, 1974, p. 61).

(b) Linear. The "linear" property indicates that the estimates $\hat{\beta}$ furnished by the least-squares method will be computed from expressions that are linear functions of the data. Naturally, this property is related to the fact that the model that was fitted to the data was linear, as discussed in Sec. 2-b.

(c) Minimum-variance. The "minimum-variance" property indicates that, again, if the same appropriate equations could be fitted with the least-squares method to successive sets of data, each set taken under identical experimental circumstances and subject only to random measurement errors, then these varying sets of estimates will "scatter" about the "true" values with the smallest possible variance, i.e. mean-squared deviations. The importance of this MV property is that the variance, or its square root - the standard deviation, is the almost universally accepted indicator of the degree of scatter of a set of values. The scatter in successive sets of estimates for the same parameter may appear slightly hypothetical, since it could be observed only if the measurement process were repeated many times under the same circumstances and each set of data fitted with the same equations. Therefore, the alternative and equivalent interpretation of minimum variance in terms of probability may be more appealing. Namely, the user of the least-squares method

can be assured that the one set of MVLU estimates $\hat{\beta}$ obtained from a fit to the one set of data y at hand has the "best chance" of being nearest the unknown "true" values; i.e., there is no estimation procedure that will give any other linear unbiased estimates with smaller variances.

Furthermore, in addition to estimated values $\hat{\beta}$ of the model parameters with the above MVLU estimates, the least-squares method can provide estimates of how far these $\hat{\beta}$ estimates may be expected to deviate from the unknown "true" values. Stated alternatively and equivalently, the estimates $\hat{\beta}$ and their standard errors can be used to construct, for each parameter, an interval (i.e., a confidence limit) within which it can be said that the unknown "true" value is likely to be found.

Examples of these properties will be given in Secs. D, E, and F.

2. Assumptions of the Least-Squares Method

The realization of the useful properties of least-squares estimates described above depends on how closely the given theoretical and experimental situation conforms to the assumptions of the least-squares method, which were only stated implicitly above. In this Section, the key, non-trivial assumptions of the least-squares method are examined in detail.

The first set of related assumptions pertain to the form of the equations, i.e., the model, used in the fit.

(a) Perfect model. It is assumed that the model perfectly describes the physical situation. As theorists know, a perfect model is a will-o'-the-wisp. When a model is incorrect, some of the least-squares estimates will naturally be biased; that is, regardless of the precision and extent of hypothetical successive measurements, the least-squares estimates will not average to the "true" values in the long run. Since this assumption of a perfect model is generally unattainable in practice, it is replaced by the more realistic assumption that the model is "adequate" for the case at hand. Because of this concession to reality, it is important to test critically the adequacy of the model during the data reduction. Thus model testing, which is discussed at an introductory level in Sec. D-3-b, is an important part of the application of the least-squares method, indicating that data analysis should be considered a continuation of the experiment.

(b) Linear model. The model must be linear in the parameters to be estimated. There are, however, many situations in spectroscopy where the model is nonlinear. Nonlinear least-squares fitting is now well-documented in several texts (Bard, 1974, and Daniel and Wood, 1971), but the details are beyond the scope of this introduction. It suffices here to say that most of the properties

of least-squares estimates discussed in this Chapter apply well to nonlinear least squares, since most spectroscopic models are approximately linear in a reasonable region about the estimated values of the parameters.

The second set of related assumptions pertain to the measurement errors. In particular, the unknown error ϵ_i in each measurement y_i is considered to be one member of an imagined population of errors whose frequency of occurrence would be described by $\varphi_i(\epsilon_i)$ if this measurement could be repeated identically many times. First, the only properties of each individual φ_i that need be specified are that each φ_i has a mean of zero and a finite variance, denoted by σ_i^2 . Secondly, the interrelations among the group of φ_i , $i = 1, 2, \dots, n$, must be specified; i.e., if the variances σ_i^2 and σ_j^2 are unequal for all $i \neq j$, then their ratios must be known, and if there is a nonzero interrelation between the φ_i and φ_j for all $i \neq j$, then the ratio of the "covariances" (defined below) must be known. These assumptions regarding the measurement errors are given here in a rather condensed form and merit elaboration.

(c) Zero-mean errors. The specification of zero mean for each φ_i , i.e. $E(\epsilon_i) = 0$, assumes that the systematic component of the error in each measurement is zero, a situation that, as experimentalists know, can only be approximated in varying degrees.

If they are not zero (or cannot be accounted for), then they can introduce bias into the least-squares estimates just as can be done by the model deficiencies described above. Again this idealized situation of no systematic errors is unattainable in practice. It is more realistically assumed that the systematic errors are much smaller than the variances of the φ_i (i.e., the "random errors"). Because of this concession to reality, it is important to test for the ratio of random to systematic errors. Ramsay (1974) has given numerous examples of how wavelength-dependent systematic errors (i.e., the nonzero means of the φ_i vary with wavelength) can creep into line position measurements and also points out that such systematic errors often exceed the average magnitude of the random errors. It is of crucial importance to realize that when this is indeed the case, any statistical meaning of the estimates and their uncertainties obtained by the usual application of the least-squares method (e.g., that are described in this Chapter) is virtually nonexistent! Fortunately, there are ways of testing for some types of systematic error, as will be noted in Sec. F-1.

(d) Finite error variances. The fact that the only stipulation that is required about the form of each φ_i is that it possess a finite variance, i.e. $E(\varphi_i^2) = \sigma_i^2$, seems too weak and

incomplete. Nevertheless, it is a remarkable feature of the method of least-squares that the only property that must be specified about the zero-mean error distributions φ_i is that they each possess a finite second moment (σ_i^2); other than this, φ_i can have any form! For example, contrary to what is often thought, the applicability of the method of least squares does not depend on the assumption that φ_i is a normal distribution, i.e. the familiar "bell-shaped" Gaussian curve. However, if the distribution is indeed normal, as it often nearly is, additional special advantages accrue to the least-squares method, as will be discussed in Sec. D-3-b.

(e) Known relative error variances. The specification that the variances must be known within a common factor permits the treatment of measurements of unequal precision by the more general "weighted" least-squares method (Sec. E). The common special case is, of course, equally precise measurements, i.e. $\sigma_i^2/\sigma_j^2 = 1$ for all i and j (Sec. D).

(f) Known relative error covariances. Finally, the specification that the covariances $E(\epsilon_i \epsilon_j)$ ($i \neq j$) be known within the same common factor in (e) permits the treatment of measurements with "correlated" errors by the more general correlated least-squares formulation (Sec. F). The common special case is, of course; zero correlation, $E(\epsilon_i \epsilon_j) = 0$ ($i \neq j$), that is, the magnitude and sign of ϵ_i do not depend on the magnitude and sign of ϵ_j . The correlations

of spectroscopic measurements themselves (e.g., "raw" line positions) apparently have always been implicitly assumed to be zero and indeed it would appear to be difficult to do otherwise (but, admittedly, this particular area is unexplored). On the other hand, the correlations of "calculated data" (e.g., $\Delta G_{v+1/2} = G_{v+1} - G_v$ or term values) can be computed straightforwardly and are generally nonzero.

We shall see in Sec. E-1 that (d), (e), and (f) are actually just different parts of one requirement, i.e., the elements of the variance-covariance matrix of the measurement errors $[\underline{\epsilon}]_{ij} = E(\epsilon_i \epsilon_j)$, must be finite and be known within a common factor. Furthermore, we shall see that this matrix must be considered as input, along with the measurements, in the most general form of the least-squares formulation.

(g) Independent variable without error. Although the above assumptions — the first group about the model and the second group about the errors of the data — are the only necessary assumptions of the least-squares method, there is an additional assumption that is made so frequently (often implicitly) that it should be stated here. Namely it is assumed that the independent variable is known without error; i.e., only the dependent variable has measurement error. Stated alternatively, the parameters to be estimated are assumed to be the only unknowns in the model; all numerical factors are known without error. A spectroscopic

example is the familiar assumption that the rotational quantum numbers J are known in the fitting of the molecular constants ν_0 , B' , D' , B'' , and D'' to a set of measured line positions $\nu_i(J_i', J_i'')$, $i = 1, 2, \dots, n$. Least-squares fits to data with errors in both the independent and dependent variables are rare and they are seldom examined in introductory texts. The application of the least-squares method when there are errors in both variables was first considered by Deming (1943, Chap. 4) and he gave an approximate method for computing the estimates. Powell and MacDonald (1972) and Britt and Luecke (1973) have recently developed techniques for computing the exact least-squares estimates and Tryon (1975) discusses the application of these methods to a spectroscopic problem. If gross errors exist in the independent variable (e.g., misassignments of the rotational quantum number J_i) and were ignored, then they could introduce bias in the least-squares estimates in much the same way as model deficiencies or systematic measurement errors.

D. UNWEIGHTED LEAST-SQUARES FORMULATION

In this section, the mechanics of applying the simplest least-squares formulation are examined in detail, and, rather than manipulate abstract quantities, a commonly occurring spectroscopic example is used.

1. Statement of the Problem

A set of measured line positions (i.e., wavenumbers) y_i , $i = 1, 2, \dots, n$, where $n \gg 5$, for a $^1\Sigma - ^1\Sigma$ band are to be used to obtain least-squares estimates of the band origin ν_0 and the rotational constants B' , D' , B'' , and D'' . The familiar model is expressed here as n equations:

$$y_i = \nu_0 + B'J_i'(J_i'+1) - D'[J_i'(J_i'+1)]^2 - B''J_i''(J_i''+1) + D''[J_i''(J_i''+1)]^2 + \epsilon_i$$

$$i = 1, 2, \dots, n \quad (D-1)$$

In accordance with assumption (a) of Sec. B-2 this model is taken to be completely adequate. Furthermore, it is clear from Eq. (D-1) that the model is linear in the unknown molecular constants, thereby satisfying assumption (b). The ϵ_i appearing in Eq. (D-1) are the unknown measurement errors associated with the line

positions y_i . They are assumed to be described by distribution functions of zero mean (assumption c) and common unknown variance σ^2 (assumptions d and e). In addition, the errors ϵ_i are assumed to be uncorrelated (assumption f). Lastly, the independent variables in Eq. (D-1) — 1 (for the coefficient of v_0), J_i' , and J_i'' — are assumed to be numerically known without error (assumption g).

That all of the measurement errors are described by one variance implies that the measurements are equally precise, i.e. weights of unity. This assumption and that of zero correlation between the errors are what make this example the special case of unweighted least-squares, which is more properly called unity-weighted, uncorrelated least-squares.

In matrix notation, Eq. (D-1) can be written quite succinctly as the vector equation

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon} , \quad (D-2)$$

where \underline{y} , $\underline{\beta}$, and $\underline{\epsilon}$ are the column vectors

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \nu_0 \\ B' \\ D' \\ B'' \\ D'' \end{bmatrix}, \quad \underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_n \end{bmatrix}, \quad (D-3)$$

the first of which is known. The known $n \times 5$ "coefficient" matrix \underline{X} has elements given by $1, J_i'(J_i'+1), -[J_i'(J_i'+1)]^2 - J_i''(J_i''+1)$, and $[J_i''(J_i''+1)]^2$. For a band with 50 R-lines and 50 P-lines, the $i = 1, 2, 50, 51, 52$, and 100 rows of the coefficient matrix \underline{X} [which correspond to the $R(J''=0), R(1), R(49), P(1), P(2)$, and $P(50)$] are

$$\underline{X} = \begin{bmatrix} 1 & 2 & -4 & 0 & 0 \\ 1 & 6 & -36 & -2 & 4 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 2550 & -6502500 & -2450 & 6002500 \\ 1 & 0 & 0 & -2 & 4 \\ 1 & 2 & -4 & -6 & 36 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & 2450 & -6002500 & -2550 & 6502500 \end{bmatrix}. \quad (D-4)$$

2. Least-Squares Estimates

The least-squares method cannot, of course, provide the "true" values of the molecular constants. However, it can provide prescriptions of how to calculate, from the measured line positions, values that are to be taken as estimates of the three key sets of unknowns: (a) the molecular constants, (b) the standard deviation of the measurement errors, and (c) the standard errors of the estimated molecular constants. In this Section, the prescriptions for obtaining these three sets of estimates are given and the properties of these estimates are examined.

(a) Estimates of the parameters. The values $\hat{\underline{\beta}}$ that minimize the sum of the squared deviations between the measured line positions \underline{y} and the calculated line positions $\hat{\underline{y}}$, which are given by

$$\hat{\underline{y}} = \underline{X}\hat{\underline{\beta}}, \quad (D-5)$$

are to be taken as the MVLU estimates of the unknown molecular constants $\underline{\beta}$. The prescription for calculating these $\hat{\underline{\beta}}$ values from the known coefficient matrix \underline{X} and the known measured line positions \underline{y} is the following expression

$$\hat{\underline{\beta}} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{Y} , \quad (D-6)$$

the derivation of which is given in the Appendix.

It is worthwhile to reemphasize the MVLU properties of the estimates $\hat{\underline{\beta}}$ by considering one of the five components of $\hat{\underline{\beta}}$, namely $\hat{\beta}_5$, the estimate of the lower-state centrifugal distortion constant D'' . First, since the model is assumed correct, $\hat{\beta}_5 = \hat{D}''$ is an unbiased estimate. Secondly, it is clear from Eq. (D-6) that \hat{D}'' is a linear function of the measured line positions, i.e., $\hat{D}'' = [(\underline{X}^T \underline{X})^{-1} \underline{X}^T]_{51} Y_1 + [(\underline{X}^T \underline{X})^{-1} \underline{X}^T]_{52} Y_2 + \dots + [(\underline{X}^T \underline{X})^{-1} \underline{X}^T]_{5n} Y_n$. Thus, \hat{D}'' is a linear estimate. Lastly, if the measurements could be repeated many times in an identical fashion, except for random measurement error, and each set of measurements fitted to yield $\hat{\underline{\beta}}$, then the hypothetical group of estimates \hat{D}'' (of which we happen to possess only one member) would scatter about the unknown "true" value of D'' with the smallest possible variance. Hence, \hat{D}'' is the minimum-variance estimate.

This last property merits some elaboration. It is, of course, impractical to make a large number of repeated measurements of the same set of lines under nearly identical experimental conditions to examine the minimum-variance scatter of the resulting set of \hat{D}'' values. However, such scatter can be simulated very realistically in the following example. From a given set of

"true" values $\nu_0 = 50\,000$, $B' = 1.5$, $D' = 4.0 \times 10^{-8}$, $B'' = 2.0$, and $D'' = 6.0 \times 10^{-8} \text{ cm}^{-1}$, a set of $n = 100$ "true" line positions, 50 R-lines $[R(J''=0), R(1), \dots, R(49)]$ and 50 P-lines, $[P(J''=1), P(2), \dots, P(50)]$, were calculated. A thousand randomly different sets of random errors, 100 in each set, were created from a normal distribution (see Sec. D-3-a) with a zero mean and a standard deviation of 0.05 cm^{-1} . These errors were combined with the "true" line positions to make a thousand sets of "synthetic" line positions, 100 per set, each of which therefore simulates exactly the conditions assumed in Sec. D-1 for the applicability of the unweighted least-squares formulation. The specification of a normal distribution is, of course, not necessary here, but one must select some type of distribution, and the particular choice of a normal distribution will make this example useful again in Sec. D-3-b. Applying Eq. (D-6) to each set of line positions yields a thousand different sets of $\hat{\beta}$, which scatter about the "true" values. The extent of this scatter is shown in the "frequency-of-occurrence histogram" of Fig. D-1, which gives for \hat{D} the number of values that lie in small intervals about the "true" value. The common measure of this scatter is, of course, the variance, which is computed from the sum of the squares of the residuals between the "true" value and the estimates:

$s_s^2 = \sum (D''_{\text{true}} - \hat{D}'')^2 / 1000$, where $D''_{\text{true}} = 6.0 \times 10^{-8} \text{ cm}^{-1}$. The standard deviation s_s of the scatter is the square root of s_s^2 . For the \hat{D}'' values in Fig. D-1, $s_s = 6.26 \times 10^{-8} \text{ cm}^{-1}$ and $D''_{\text{true}} \pm s_s$ are marked by the arrows. The powerful MV property of these 1000 estimated \hat{D}'' values can now be visually appreciated. Namely, there is no unbiased linear function of the measured line positions y that can give values that have smaller scatter, as measured by the variance, than this shown in Fig. D-1 (i.e., smaller scatter than that exhibited by the values calculated with Eq. D-6). Any other unbiased linear estimates must exhibit either the same or larger scatter.

Returning to real life, where, instead of hundreds of sets of line position measurements, we have only one set that has been fitted to yield one set of least-squares estimates $\hat{\theta}$ for the molecular constants, the advantage of the MV property of this set of values is a probabilistic one. Even though this \hat{D}'' value may lie anywhere in the distribution shown in Fig. (D-1) (e.g., far out in the left tail or near the center, close to the "true" value, etc.), we can nevertheless be assured that the chances are better that our MV estimate \hat{D}'' lies closer to the "true" value than would any non-MV estimate of D'' based on the same measurements.

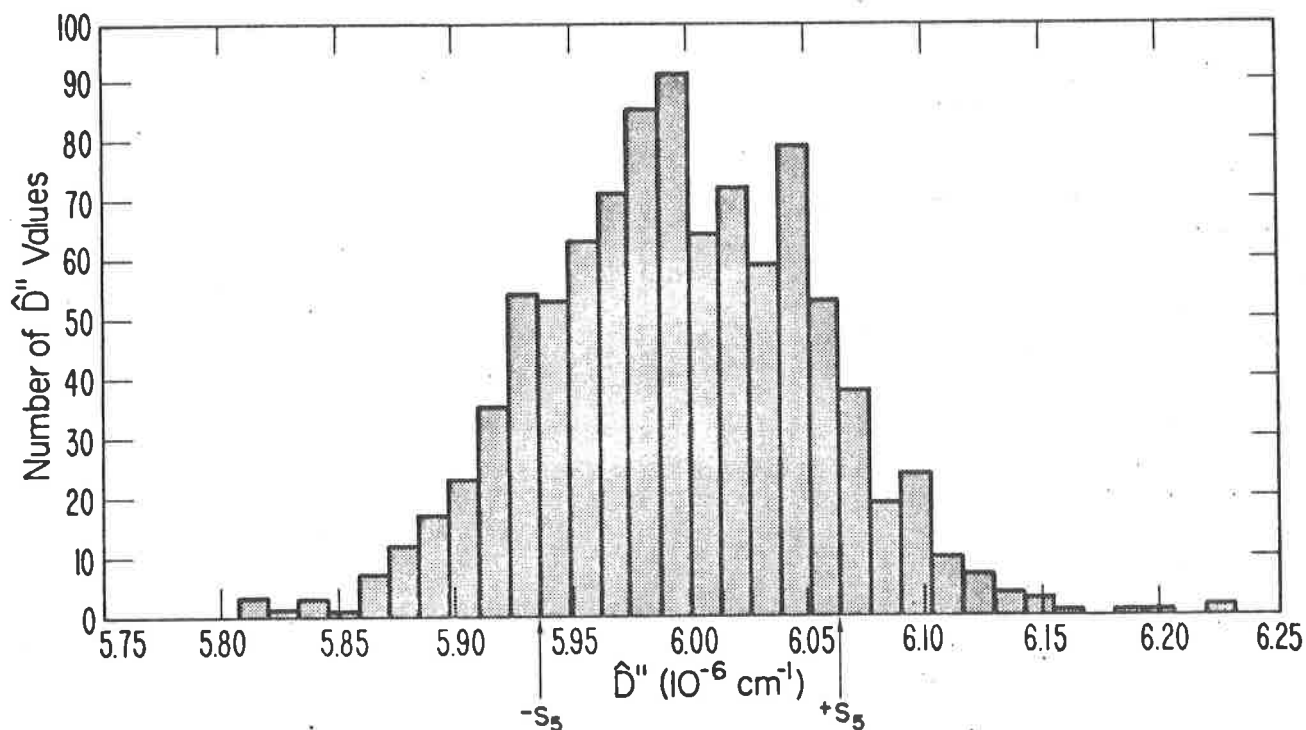


Figure D-1. Frequency-of-occurrence histogram of 1000 \hat{D}'' values. The values were obtained from least-squares fits to 1000 sets of synthetic $^1\Sigma - ^1\Sigma$ line positions, each set containing 100 lines: $R(J''=0), R(1), \dots, R(49)$ and $P(J''=1), P(2), \dots, P(50)$. The synthetic lines were constructed by adding random, normally-distributed "errors" with zero mean and $\sigma = 0.05 \text{ cm}^{-1}$ to line positions calculated from $\nu_0 = 50000$, $B' = 1.5$, $D' = 4.0 \times 10^{-6}$, $B'' = 2.0$, and $D'' = 6.0 \times 10^{-6} \text{ cm}^{-1}$, as indicated in Eq. (D-1). The arrows denote the locations of $D''_{\text{true}} \pm s_s$, where s_s is the standard deviation of the differences $(D''_{\text{true}} - \hat{D}'') = (6.0 \times 10^{-6} - \hat{D}'') \text{ cm}^{-1}$.

(b) Estimate of the variance of the random measurement errors. The variance of the measurement error distribution is generally unknown. The least-squares formulation prescribes (see the Appendix) that an estimate of this variance is to be taken as the value obtained by dividing the sum of the squares of the deviations between the measured and calculated line positions by the "degrees of freedom" of the least-squares fit, i.e.,

$$\hat{\sigma}^2 = \frac{1}{n-5} (\underline{y} - \underline{x}\hat{\underline{\beta}})^T (\underline{y} - \underline{x}\hat{\underline{\beta}}) , \quad (\text{D-7})$$

where the degrees of freedom are the number of independent measurements, n , minus the number of unknowns to be estimated, five in this example. The associated estimate of the standard deviation of the measurement error distribution is, of course, the square root of the variance, $\hat{\sigma} = (\hat{\sigma}^2)^{\frac{1}{2}}$.

Several points are noteworthy about the estimated variance. First, it is shown in the Appendix that $\hat{\sigma}^2$ is an unbiased estimate of the variance of the measurement errors. (However, this is not strictly true of $\hat{\sigma}$, since the square root is a nonlinear operation. Nevertheless, for the large degrees of freedom, the bias in $\hat{\sigma}$ is negligible. For an elaboration of this small point, which is ignored hereafter, see Bennett and Franklin [p. 164-166].) Being an estimate, $\hat{\sigma}^2$ is subject to variation like $\hat{\underline{\beta}}$ due to random

measurement errors. Figure D-2 shows the frequency-of-occurrence histogram for the 1000 estimates of $\hat{\sigma}$ obtained from the synthetic line position fits described in (a) above.

Secondly, the estimated standard deviation $\hat{\sigma}$ is not to be confused with the root-mean-square (rms) of the residuals, which has a similar, but not identical, form:

$$\text{rms} = \left[\frac{1}{n} (\underline{y} - \underline{x}\hat{\beta})^T (\underline{y} - \underline{x}\hat{\beta}) \right]^{\frac{1}{2}} \quad (\text{D-8})$$

The difference is that the standard deviation employs the degrees of freedom $n-5$ rather than simply the number of measurements n . When the degrees of freedom are large, the standard deviation and the rms deviation are nearly identical, but when $n \gtrsim m$, the distinction is important, particularly in model testing.

Thirdly, since $\hat{\sigma}$ is an estimate of the quality (i.e., the precision) of the set of assumed equally precise measurements, the value of $\hat{\sigma}$ changes very little when the quantity of data changes. This can be seen from the form of Eq. (D-7). As n is increased, the increase in the sum of the squared residuals $(\underline{y} - \underline{x}\hat{\beta})^T (\underline{y} - \underline{x}\hat{\beta})$ in Eq. (D-7) is proportional to n , on the average, and for $n \gg 5$, the increase in degrees of freedom is also approximately proportional to n . Thus, for substantially large degrees of freedom, there will be very little change in $\hat{\sigma}$ as the size of the data set

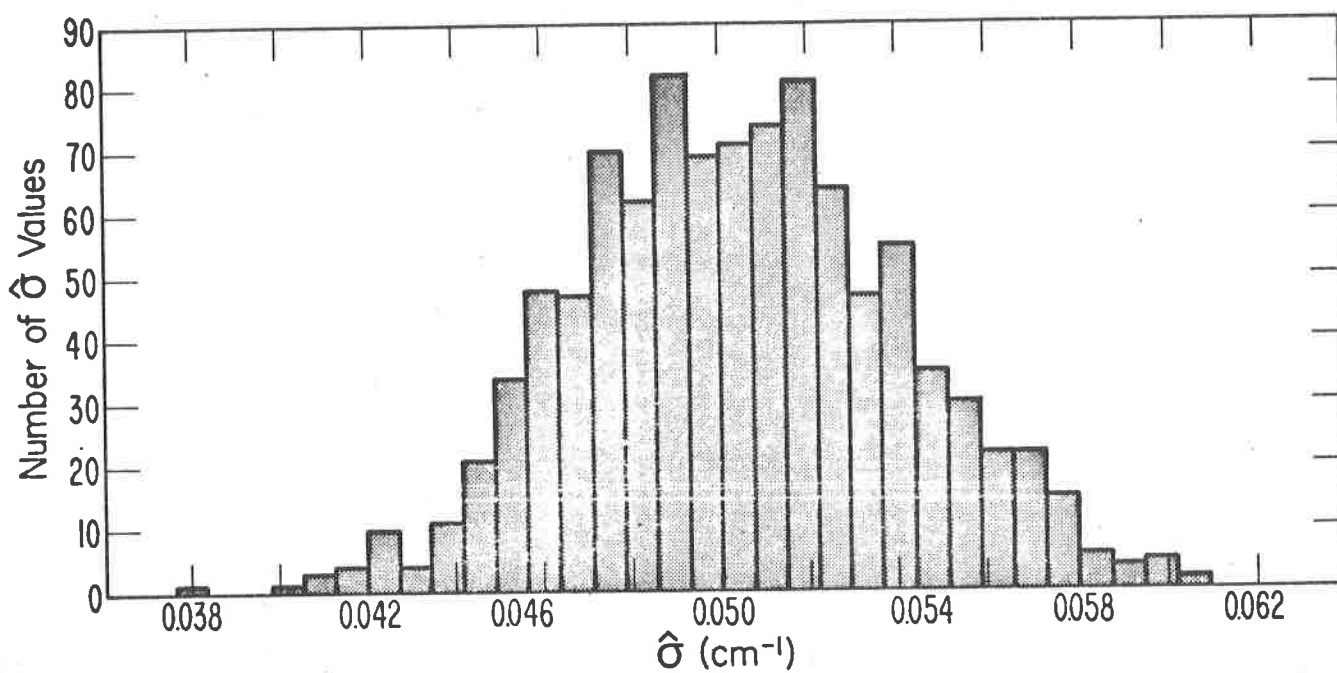


Figure D-2. Frequency-of-occurrence histogram of 1000 estimates $\hat{\sigma}$ of the standard deviation of the measurement error distribution. The values were obtained from 1000 least-squares fits described in Fig. D-1.

increases. This insensitivity also can be seen in Fig. D-3. For one of the sets of synthetic line positions described in (a) above, the equal number of R-lines and P-lines in the least-squares fit was varied from 3 to 100. There is little variation of $\hat{\sigma}$, except for the low J_{\max} values where there are only a few degrees of freedom and the least-squares fits are more sensitive to the particular values of the random errors. The insensitivity of $\hat{\sigma}$ to the quantity, i.e. extent, of nearly equally-precise data is a property that will be useful in (c) below.

Lastly, a check on the magnitude of the estimated standard deviation $\hat{\sigma}$ and an examination of the residuals between the measurements \underline{y} and the calculated values $\hat{\underline{y}}$ (Eq. D-5) are both valuable diagnostics that test whether the assumptions of the least-squares fit appear to be violated. Since $\hat{\sigma}$ should be an estimate of the precision of the measurements, this estimate is the value that is to be compared to the spectroscopist's own estimate of the random measurement error. If the estimated standard deviation is more than about 3 (?) times larger than it "should" be, then this is generally a good indication that something is wrong, usually with the selected model or with a few of the measurements. Hence, one or more of the assumptions necessary for the applicability of the least-squares method are invalid. Furthermore, an examination of the residuals can reveal similar

problems. A few "maverick" measurements (see Sec. D-3-a) in a set of data can be easily spotted. Plots of the residuals against the rotational quantum number J_i for each branch are an important method of testing the selected model. If the residuals show noticeable positive or negative (or both alternately) systematic trends with J_i , this is often a strong indication of the need for one or more additional terms in the model. A further type of systematic variation is when subdivisions of the data set have residuals of significantly different mean sizes. This would suggest that the measurements were not of equal precision and that a weighted least-squares fit (Sec. E) is needed. Finally, a plot of the frequency of occurrence of the magnitude and sign of the residuals can reveal the measurement error distribution function ϕ , a point discussed in detail in Sec. D-3-a. Plotting routines are commonly available at computing centers and, when appended to a least-squares program, remove the tedium from carrying out these informative tests. Additional discussions of the analysis of residuals are given by Draper and Smith (1966, pp. 86-99), Daniel and Wood (1971, pp. 27-32), and Bard (1974, pp. 198-201).

(c) Estimates of the variances and covariances of the estimated parameters. Figure D-1 shows how the estimates \hat{D} scatter about the "true" value when the measurements are repeated many times in an identical fashion, except for random measurement error, and each set of synthetic line positions fitted to yield \hat{g} . The

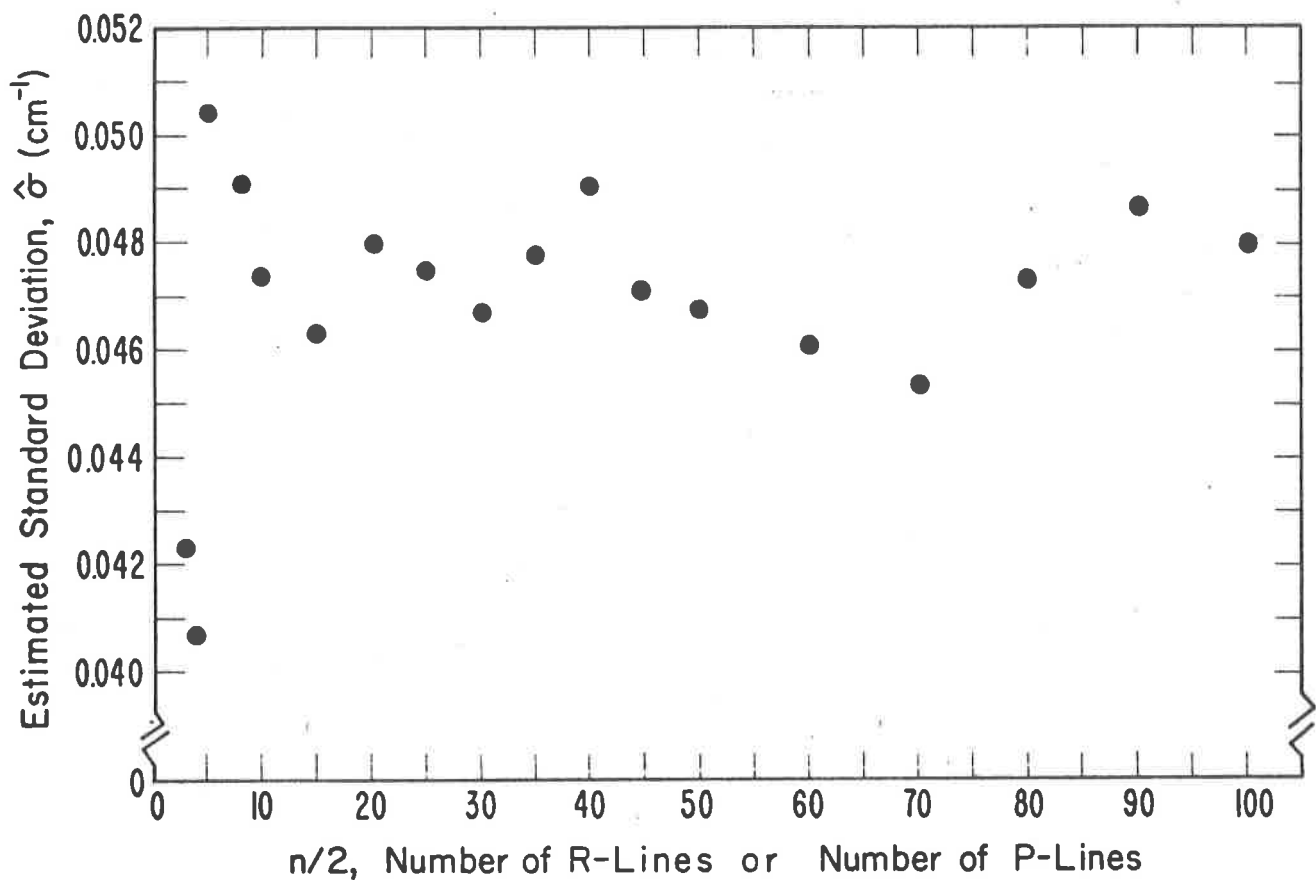


Figure D-3. Variation of estimated standard deviation $\hat{\sigma}$ of the error distribution as a function of the number of R-lines, $R(J''=0)$, $R(1)$, ..., $R(\frac{n}{2}-1)$, and the equal number of P-lines, $P(J''=1)$, $P(2)$, ..., $P(\frac{n}{2})$, in the fit. The line positions and model fitted was similar to one of the sets described in Fig. D-1.

standard deviation $s_{\hat{D}}$ of the scatter of the \hat{D} values is also indicated in the figure. In a probabilistic sense, this standard deviation for the 1000 \hat{D} estimates, for example, is the standard deviation of any one of the estimates. Namely, for a small $s_{\hat{D}}$, indicating a narrow distribution (i.e., less scatter), then the chances are better that any single estimate \hat{D} lies closer to the "true" value, than for a larger $s_{\hat{D}}$ indicating a wider distribution (i.e., more scatter). Thus, the variance of this distribution of \hat{D} values from many identically-repeated measurements is interpreted as the measure of the precision of any one of the single \hat{D} values.

However, it is clearly impractical to compute the variance of a molecular constant from the scatter exhibited by the values from many repeated sets of nearly identical experiments! We need an estimate of this variance that can be calculated from one set of measured line positions. The least-squares method prescribes (see the Appendix) that such estimates are to be taken as the values given by the diagonal elements of the square (5x5, in this case) symmetric "variance-covariance" matrix

$$\hat{\Theta} = \hat{\sigma}^2 \underline{V} \quad , \quad (D-9)$$

where the dispersion matrix \underline{V} is given by

$$\underline{V} = (\underline{X}^T \underline{X})^{-1} \quad (D-10)$$

Since $\hat{\underline{\Theta}}$ is clearly a linear function of $\hat{\sigma}^2$, it is an unbiased estimate of $\underline{\Theta}$, as described in (b) above. The estimated variance-covariance matrix contains a wealth of information and we shall examine its major attributes in detail.

Variances and standard errors The diagonal elements $\hat{\Theta}_{ii}$ are the estimated variances of the values of the molecular constants \hat{V}_0 , \hat{B}' , \hat{D}' , \hat{B}'' , and \hat{D}'' . The square roots of the $\hat{\Theta}_{ii}$ elements are the standard errors. (The term "error" is commonly used here, rather than "deviation" to distinguish standard errors of the molecular constants from the standard deviation of the measurement errors.) For one of the sets of synthetic lines considered above (the first 50 R-lines and 50 P-lines and assuming $\hat{\sigma} = 0.05 \text{ cm}^{-1}$), the diagonal and above-diagonal elements of the matrix $\hat{\underline{\Theta}}$ is given in Table D-1 and the parentheses on the diagonal contain the standard errors. (Since $\hat{\underline{\Theta}}$ is symmetric, there is no need to list the below-diagonal elements and the space is used to list the correlation coefficients, which will be defined below.) It is important to note that the standard error of \hat{D}'' , $\hat{\Theta}_{55}^{1/2} = 6.41 \times 10^{-8} \text{ cm}^{-1}$, is very nearly equal to the standard deviation $s_5 = 6.26 \times 10^{-8} \text{ cm}^{-1}$ computed from the deviations of the 1000 estimates in Fig. D-1 from the "true" value.

Thus, we see that the values of $\hat{\Theta}_{ii}$ computed from Eq. (D-9) and Eq. (D-7) based on one set of measured line positions is indeed the variance that would be computed from the scatter of many identically-repeated sets of measurements, if this were practical. Furthermore, $\hat{\Theta}_{ii}$ has the same probabilistic interpretations discussed above; namely, for smaller $\hat{\Theta}_{ii}$, it is more likely that the estimated value $\hat{\beta}_i$ is closer to the unknown "true" value than it is for larger $\hat{\Theta}_{ii}$. Because of this, the magnitudes of the estimated standard errors $\hat{\Theta}_{ii}^{1/2}$ are important least-squares indicators. For example, a \hat{D} value itself makes no statement about the width (i.e., standard error) of the scatter in Fig. D-1. The scatter may be as large as \hat{D} itself, as it often is for this molecular parameter, or it may be a small fraction of \hat{D} itself, as it is in the example. The point is that without the standard errors $\hat{\Theta}_{ii}^{1/2}$ there is no indication of the quality (i.e., precision) of the $\hat{\beta}_i$ values.

We are now in a position to see what the precision $\hat{\Theta}_{ii}^{1/2}$ of an estimated molecular constant (equivalently, e.g., the scatter in Fig. D-1) depends on. An analysis of this dependence is worthwhile because achieving smaller variances for the molecular constants is one clearly desirable goal of spectroscopic measurements. Toward this end, it is important to note in Eq. (D-9) that the $\hat{\Theta}_{ii}$ are the product of two distinctly different factors, $\hat{\sigma}^2$ and V_{ii} .

These two factors show how the uncertainty in an estimated molecular constant arises from separate sources. Specifically, the factor $\hat{\sigma}^2$ reflects the contribution from the precision of the measurements and the dispersion matrix element V_{ii} reflects the contribution from the structure of the measurements and the model. The first of these contributions appeals to intuition; namely, precise measurements (i.e., small $\hat{\sigma}^2$) must be a factor in obtaining precise estimates (i.e., small $\hat{\theta}_{ii}$) for molecular constants. However, the dependence of $\hat{\theta}_{ii}$ on V_{ii} shows that precise measurements alone are not sufficient; the data and the model must have a favorable structure. The following example sharpens our understanding of this last point.

The separate contributions of $\hat{\sigma}^2$ and V_{ii} can be examined by considering how the precision of the estimated molecular constants changes with the enlargement of the structure of the data set, namely, as the number of equally-precise lines n increases in going to higher J' and J'' levels. It was seen in Fig. D-3 that $\hat{\sigma}^2$ is almost insensitive to such a change. However, in contrast to this near-constancy of the $\hat{\sigma}^2$ component, V_{ii} is remarkably sensitive to the structure of the data set. Because there are few among us that can peer back through the matrix inversion, multiplication, and transposition to see how the size of a given element $V_{ii} = [(X^T X)^{-1}]_{ii}$ depends on the structure of X , this is best examined with a numerical example. Figure D-4 shows the

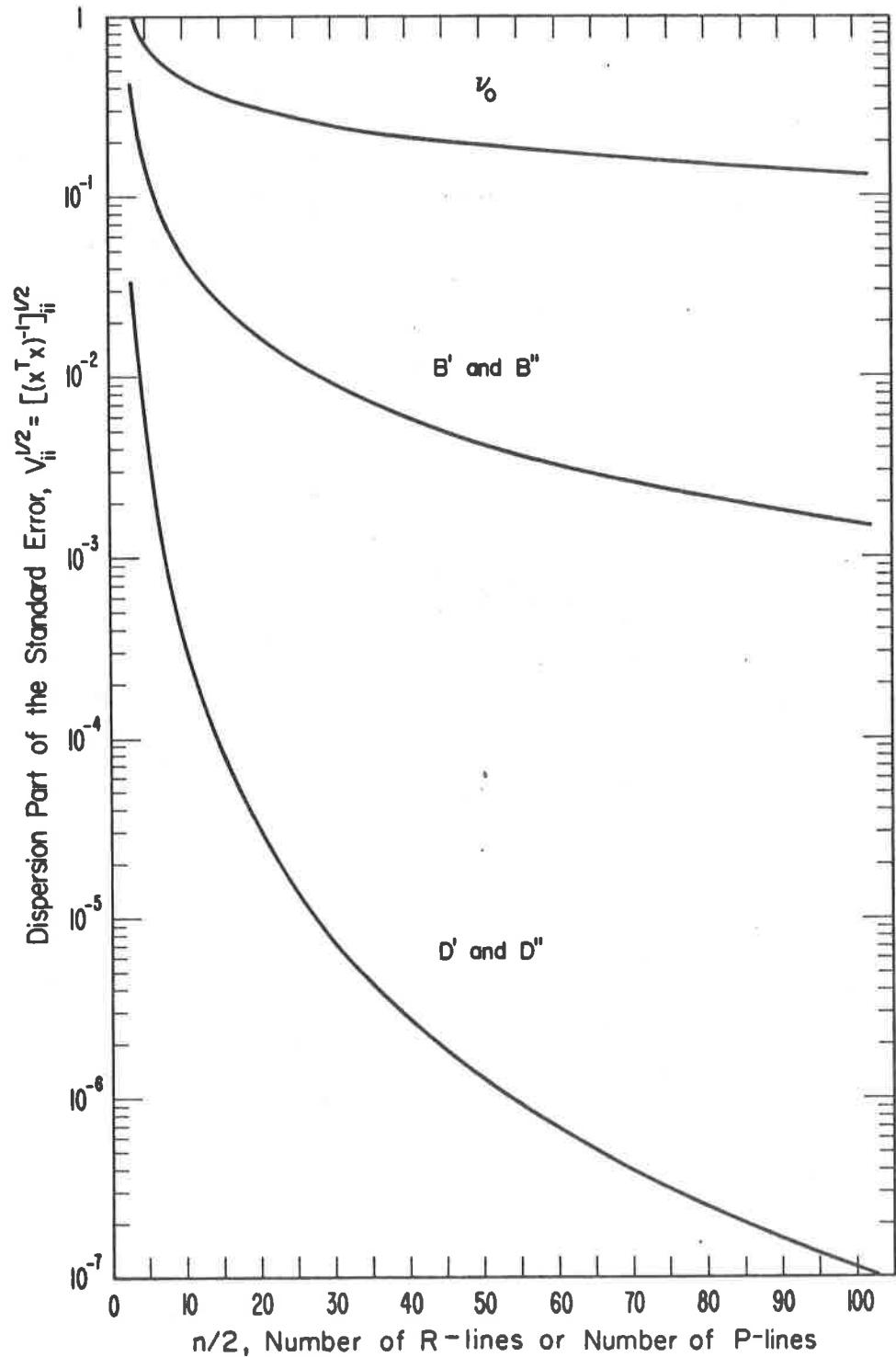


Figure D-4. Variation of the dispersion elements $[(\underline{x}^T \underline{x})^{-1}]_{ii}^{1/2}$ associated with v_0 , B , and D as a function of the number of R-lines $R(J''=0), R(1), \dots, R(\frac{n}{2}-1)$ and the equal number of P-lines, $P(J''=1), P(2), \dots, P(\frac{n}{2})$ in the coefficient matrix \underline{X} .

variation of the $V_{ii}^{\frac{1}{2}}$ elements that are associated with ν_0 , B' , B'' , D' , and D'' as a function of $n/2$, which is the number of R-lines $[R(J''=0), R(1), \dots, R(n/2 - 1)]$ and the equal number of P-lines $[P(J''=1), P(2), \dots, P(n/2)]$ composing the $n \times 5$ matrix \underline{X} . For this symmetric example, the variance associated with B' and B'' are equal, and similarly for D' and D'' (see Table D-1). These equalities are the result of the symmetries of \underline{X} (see Eq. D-4).

For the range covered, the $V_{ii}^{\frac{1}{2}}$ elements associated with ν_0 , B , and D decrease by about 1, 2, and 5 orders of magnitude, respectively. This relative behavior certainly agrees with the spectroscopist's expectations that $\hat{\nu}_0$ is least sensitive to the presence of high- J lines and \hat{D} is most sensitive.

As an example of a particular case, Fig. D-4 shows that if (i) each branch could be traced on for 10 more lines, i.e., the R-lines increased from $R(J''=0), R(1), \dots, R(49)$ to $R(J''=0), R(1), \dots, R(59)$ and similarly for the P-branch, and (ii) the measurements were of the same precision, then the standard errors of $\hat{\nu}_0$, \hat{B} , and \hat{D} would decrease by factors of 1.10, 1.31, and 1.88, respectively. This would mean that if the measurements could be repeated many times in an identical fashion, the scatter in the resulting sets of values of $\hat{\nu}_0$, \hat{B} , and \hat{D} would decrease by these factors. Using 1000 sets of synthetic lines constructed and fitted in the same manner described in Sec. D-2-a above except

that each of the branches have 60 members instead of 50, the frequency-of-occurrence histogram in Fig. D-5 was made for \hat{D}'' . The horizontal and vertical scales are the same as those in Fig. D-1, which was for the same conditions except for only 50 lines in each branch. Thus, comparison of these two histograms shows visually how the precision of an estimated molecular constant depends on the structure of the data, for a constant precision of the line position measurements themselves.

A few more short examples of this sensitivity of the variance of the molecular constants to the structure of the data will suffice. Frequently, in practice, the R-branch can be traced to higher J levels than the P-branch because of the relatively larger Hönl-London factors (Herzberg, 1950, pp. 126 and 204). It is interesting to note that even the extension of one branch improves the precision of the estimated molecular constants. The first three lines of Table D-2 show the $V_{ii}^{1/2}$ elements of three cases, two with symmetric branch structure and one with asymmetric structure. Note that the asymmetric 75 R-lines, 50 P-lines is better (smaller $V_{ii}^{1/2}$ elements) than 50 R-lines, 50 P-lines, but, as expected, is worse than 75 R-lines, 75 P-lines. Furthermore, the standard errors associated with B' and B'' will now be slightly different, as will be those associated with D' and D'' . The $\Delta J = \pm 1$ selection rule keeps them from being very different,

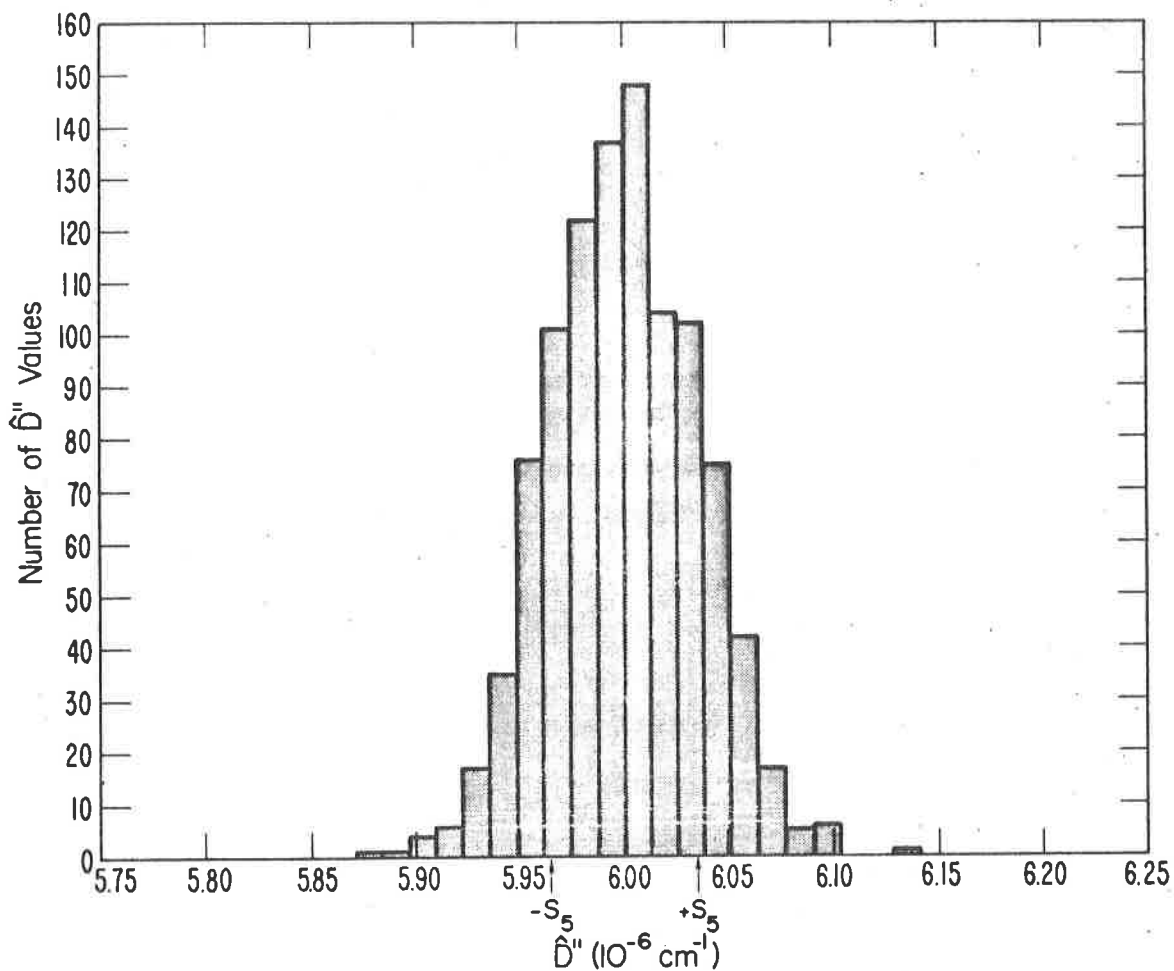


Figure D-5. Frequency-of-occurrence histogram of 1000 \hat{D}'' values. The values were obtained from least-squares fits to 1000 sets of synthetic $^1\Sigma - ^1\Sigma$ line positions, each set containing 120 lines: $R(J''=0), R(1), \dots, R(59)$ and $P(J''=1), P(2), \dots, P(60)$. The synthetic lines were constructed like the ones described in Fig. D-1. The arrows denote the locations of $D''_{\text{true}} \pm s_5$, where s_5 is the standard deviation of the differences $(D''_{\text{true}} - \hat{D}'') = 6.0 \times 10^{-6} - \hat{D}''$ cm⁻¹.

Table D-2. Comparison of the dispersion part of the standard errors, $V_{ii}^{\frac{1}{2}} = [(\tilde{X}^T X)^{-1}]_{ii}^{\frac{1}{2}}$, of the molecular constants ν_0 , B' , D' , B'' , and D'' for a ${}^1\Sigma - {}^1\Sigma$ band with varying numbers of R- and P-lines.

number ^a of R lines P lines		$V_{ii}^{\frac{1}{2}} = [(\tilde{X}^T X)^{-1}]_{ii}^{\frac{1}{2}}$				
f	g	ν_0 (i=1)	B' (i=2)	D' (i=3)	B'' (i=4)	D'' (i=5)
50	50	0.190	4.27×10^{-3}	1.28×10^{-6}	4.27×10^{-3}	1.28×10^{-6}
75	50	0.159	3.91×10^{-3}	1.07×10^{-6}	4.01×10^{-3}	1.13×10^{-6}
75	75	0.154	2.34×10^{-3}	3.13×10^{-7}	2.34×10^{-3}	3.13×10^{-7}
40 ^b	50	0.230	4.33×10^{-3}	1.30×10^{-6}	4.37×10^{-3}	1.31×10^{-6}
0	50	0.801	1.15×10^{-1}	1.25×10^{-4}	9.84×10^{-2}	1.20×10^{-4}

^a Except as noted in footnote b, the lines are distributed continuously, i.e., $R(J''=0)$, $R(1)$, ..., $R(f-1)$ and $P(J''=1)$, $P(2)$, ..., $P(g)$.

^b The R-lines are $R(J''=10)$, $R(11)$, ..., $R(49)$.

however. The fourth line in Table D-2 shows the effect on the molecular constants of ten missing low-J R-lines, which corresponds roughly to the commonly-occurring situation where the lines of a densely packed band head cannot be resolved. The effect of the missing low-J R-lines is pronounced only for ν_0 , again indicating the relative sensitivity of this molecular constant on low-J lines. The last line in Table D-2 is for the case that may occur in laser emission spectroscopy where only one branch can be measured. Note that the fact the $\underline{X}^T \underline{X}$ can be inverted demonstrates that molecular constants can, in fact, be determined, albeit poorly, from just one branch, a possibility that does not appear to be widely recognized.

Such tests are very valuable exercises, since \underline{y} is independent of the wavelength of the lines and the magnitudes of the molecular constants in a given model. Thus, with a good guess for $\hat{\sigma}$ (which most positional spectroscopists can quote from past experience, even before the measurements are made), the standard errors of each molecular constant can be predicted as a function of rotational development. For example, with the 50 P-lines described by the fifth line of Table D-2 and a measurement precision of $\hat{\sigma} \approx 0.05 \text{ cm}^{-1}$, B'' could be determined to a standard error of approximately 0.0049 cm^{-1} . Simple calculations like this can be very useful in planning the recording of the spectrum if, for example, an

improvement in the precision of existing molecular constants, or the determination of a particular small constant associated with some weak molecular interaction, is the goal. This utility has also been pointed out by Pliva and Telfair (1974), who give similar examples. Naturally, it also applies to molecular models other than the one used as the example here.

Covariances and correlation coefficients. As described above, the diagonal elements of the $\hat{\Theta}$ matrix are the estimated variances of the molecular constants. The off-diagonal elements $\hat{\Theta}_{ij}$ ($i \neq j$) are the estimated covariances. For the ${}^1\Sigma - {}^1\Sigma$ example at hand, the covariances of the molecular constants ν_0 , B' , D' , B'' , and D'' are the upper off-diagonal elements in Table D-1. Unlike a variance, which refers to one molecular constant, a covariance refers to a pair of molecular constants. For example, the element $\hat{\Theta}_{12} = -8.262588 \times 10^{-8} \text{ cm}^{-2}$ is the covariance of ν_0 and B' . Since $\hat{\Theta}$ is a symmetric matrix, the covariance of ν_0 and B' is always equal to the covariance of B' and ν_0 , i.e., $\hat{\Theta}_{ij} = \hat{\Theta}_{ji}$. Furthermore, because the particular example has branches of equal extent, the $\hat{\Theta}_{ij}$ elements in Table D-1 have the additional symmetries that $\hat{\Theta}_{12} = -\hat{\Theta}_{14}$, $\hat{\Theta}_{13} = -\hat{\Theta}_{15}$, $\hat{\Theta}_{23} = \hat{\Theta}_{45}$, and $\hat{\Theta}_{25} = \hat{\Theta}_{34}$. However, these exact symmetries do not exist for ${}^1\Sigma - {}^1\Sigma$ bands of general structure (although they often hold approximately).

The covariances, in contrast to the sought-after variances (or their square roots, the standard errors), are a generally neglected part of the variance-covariance matrix $\hat{\Theta}$. Seldom is the computer even requested to print them out. Nevertheless, the covariances carry useful information about the least-squares estimates $\hat{\Theta}$ of the molecular constants, and the remainder of this Section will consider some of the details of this information.

First, the concept of covariance is best examined in terms of what are called correlation coefficients, which are formed by normalizing the variance-covariance matrix to form a new matrix \underline{C} , whose diagonal elements are all 1 and whose off-diagonal elements lie between -1 and +1:

$$C_{ij} = \hat{\Theta}_{ij} / (\hat{\Theta}_{ii} \hat{\Theta}_{jj})^{1/2}. \quad (D-11)$$

Since the correlation matrix \underline{C} is also a symmetric matrix and the diagonal elements are always unity, it is sufficient to give only the upper or lower off-diagonal elements. In Table D-1, the correlation coefficients of ν_0 , B' , D' , B'' , and D'' are given in the lower off-diagonal positions. For the particular ${}^1\Sigma - {}^1\Sigma$ example in Table D-1, \underline{C} has the additional symmetries mentioned above.

As Table D-1 shows, correlation coefficients can be near zero, e.g., $C_{21} = -0.0407$ for B' and ν_0 , and can be near 1 or -1, e.g., $C_{42} = 0.9941$ for B' and B'' . The magnitude of the correlation coefficients gives a measure of the interdependence of the molecular constants β for a given data set. If C_{ij} ($i \neq j$ is understood when discussing correlation coefficients) is very near ± 1 , the interdependence is very high and if C_{ij} is very near zero, the interdependence is very low. For example, the large positive correlation coefficient between B' and B'' means that if a least-squares estimate \hat{B}' for B' differs from the "true" value by a positive (or negative) error, then the accompanying least-squares estimate \hat{B}'' for B'' will also differ from the "true" value by approximately the same positive (or negative) error. Thus, the errors in \hat{B}' and \hat{B}'' are "linked together." Nearly the same interdependence exists for the errors of \hat{D}' and \hat{D}'' , since their correlation coefficient of 0.9875 is also very close to unity.

One effect of the large correlation coefficient of D' and D'' is shown in Figure D-6, in which are plotted the pairs of \hat{D}' and \hat{D}'' values obtained from the least-squares fits to the first 50 sets of synthetic $^1\Sigma - ^1\Sigma$ line positions described in Sec. D-2-a above. Each point represents the pair of \hat{D}' and \hat{D}'' values obtained from one of the least-squares fits. The vertical and horizontal

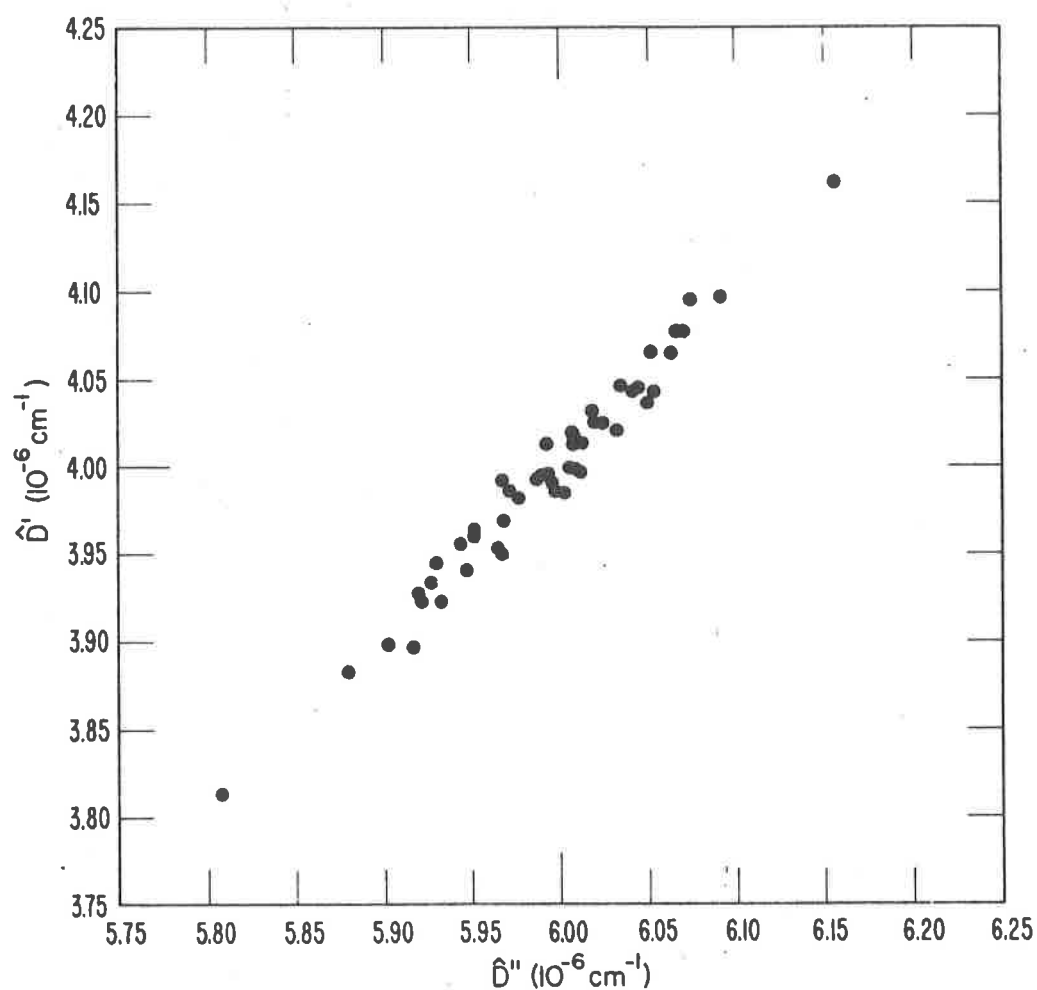


Figure D-6. Distribution of pairs of \hat{D}' and \hat{D}'' values. The (\hat{D}', \hat{D}'') pairs plotted here are the values obtained from the first 50 least-squares fits to the synthetic line positions described in Fig. D-1.

scales are the same as that used in Fig. D-1, namely, about ± 4 standard errors centered on the "true" values. The fact that the errors of \hat{D}' and \hat{D}'' are strongly "linked together," i.e., correlated, is clear from the pattern of the departures from the center of the figure, which represents the "true" values. A relatively large positive (or negative) error in \hat{D}' is probable only if it is accompanied by nearly the same positive (or negative) error in \hat{D}'' . Alternatively, a large positive (or negative) error in \hat{D}' and a large negative (or positive) error in \hat{D}'' is extremely unlikely. If the strong correlation coefficient of \hat{D}' and \hat{D}'' had been negative instead of positive, the points in Fig. D-6 would have a corresponding negative slope.

For smaller correlation coefficients, the small axis of the "ellipse" in Fig. D-6 would be larger and the points would show more "scatter" in all four quadrants. The ${}^1\Sigma-{}^1\Sigma$ band also affords an example of this. Table D-1 shows that the correlation coefficient of ν_0 and D'' is 0.0479. One effect of this nearly zero correlation is shown in Fig. D-7, which contains the pairs of $\hat{\nu}_0$ and \hat{D}'' values obtained from the least-squares fits to the first 50 sets of synthetic ${}^1\Sigma-{}^1\Sigma$ lines. The virtually random scatter shows that the errors of $\hat{\nu}_0$ and \hat{D}'' values have no relative constraints, i.e., essentially zero correlation.

Plots similar to Fig. D-6 could be made for the pairs (B', D'') ;

$(B', B''); (D', B''); (B', D')$; and (B'', D'') , since they have relatively large correlation coefficients. Similarly, plots like Fig. D-7 could be made for the pairs (ν_0, B') ; (ν_0, D') ; and (ν_0, B'') , since they have relatively small correlation coefficients.

Like the variances $\hat{\sigma}_{ii}$, the covariances $\hat{\sigma}_{ij}$ ($i \neq j$) of the molecular constants are estimated quantities made up of two distinctly different factors (see Eq. D-9): the estimated variance of the measurement error distribution $\hat{\sigma}^2$ (which is why the covariance is an estimate) and the dispersion elements $[(\underline{X}^T \underline{X})^{-1}]_{ij}$ ($i \neq j$). The first factor $\hat{\sigma}^2$ reflects the contribution of the precision of the measurements and the second factor, the dispersion elements, reflect the structure of the data and the model, as was discussed above for the variance.

On the other hand, a correlation coefficient C_{ij} is not an estimate, because the $\hat{\sigma}^2$ factor cancels in Eq. (D-11). Thus, C is determined only by the dispersion matrix \underline{V} . Hence, the correlation coefficients are known properties of only the structure of the data and the model, and not of the precision of the measurements. For a given model, varying degrees of rotational development of the branches alters the correlation coefficients, as it does the diagonal elements of the dispersion matrix, which was illustrated in Fig. D-4. The variation of the correlation coefficients for (ν_0, D'') and (B', B'') with the number of R-lines and the (equal)

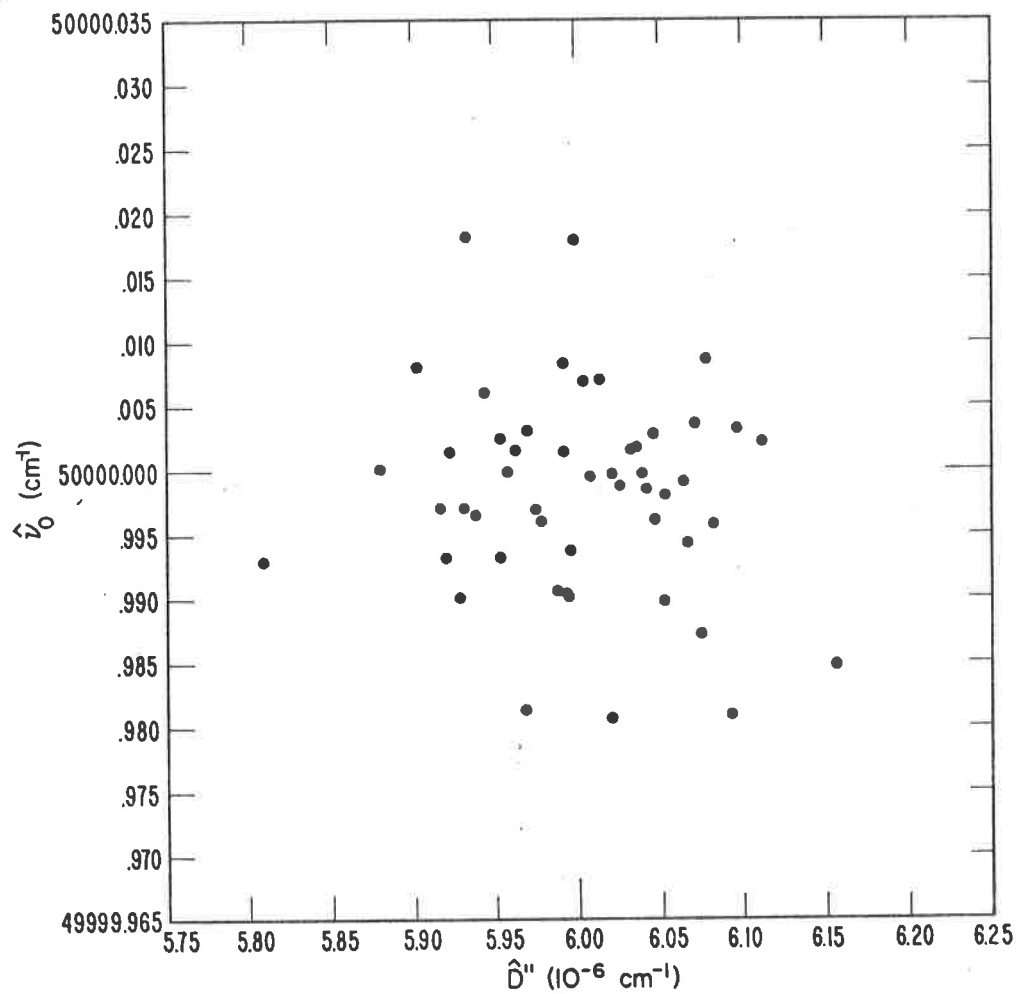


Figure D-7. Distribution of pairs of $\hat{\nu}_0$ and \hat{D}'' values. The $(\hat{\nu}_0, \hat{D}'')$ pairs plotted here are the values obtained from the first 50 least-squares fits to the synthetic line positions described in Fig. D-1.

number of P-lines is given in Fig. D-8. The variation is very sharp when the number of lines is small, but $C(\nu_0, D'')$ and $C(B', B'')$ approach zero and unity, respectively, very rapidly. Certainly, in the range of most spectroscopic measurements, these two coefficients are nearly constant and are approximately 0.1 and 0.9, respectively for a ${}^1\Sigma - {}^1\Sigma$ band. The other correlation coefficients involving ν_0 behave very much like $C(\nu_0, D'')$, and the remainder involving $B', D', B'',$ and D'' behave very much like $C(B', B'')$.

Examples of correlation coefficient matrices have been given for other spectroscopic applications: ${}^1\Delta - {}^1\Phi$ molecular constants and ${}^1\Sigma - {}^1\Sigma$ term values, Albritton, Harrop, Schmeltekopf, Zare, and Crow (1973); ${}^2\Pi - {}^2\Pi$ molecular constants, Albritton, Harrop, Schmeltekopf, and Zare (1973a); ${}^1\Sigma - {}^3\Sigma$ molecular constants, Albritton, Harrop, Schmeltekopf, and Zare (1973b); Dunham Y_{ij} coefficients, Roh and Rao (1974) and Kildal, Eng, and Ross (1974); and various ${}^1\Sigma - {}^1\Sigma$ configurations, Pliva and Telfair (1974). An example of a variance-covariance matrix for a ${}^1\Sigma - {}^1\Sigma$ band is given by Albritton, Schmeltekopf, Tellinghuisen, and Zare (1974).

It is very important to realize that the least-squares estimates $\hat{\beta}$ (or more strictly speaking, their errors) are always correlated in general; i.e., \underline{V} is generally not a diagonal matrix. The appearance of non-zero correlation is not, of course, limited just to spectroscopic least-squares fits. For example, in their

1973 least-squares adjustment of the fundamental constants, Cohen and Taylor (1973) list with their estimates of the constants and their standard errors the reminder that "the uncertainties of these constants are correlated" (see footnote a to their Table 33.1). Only when the parameters of the model are parameters of orthogonal functions is there zero correlation between the errors of the least-squares estimates (e.g., see Hunter, Khandekar, and Prokopenko, 1974).

The properties and implications of correlation coefficients have been examined in great detail; see for example, the discussion in Bennett and Franklin (1954, Sec. 6.4) on the correlations of raw data, and the comments in Dixon and Massey (1969, Chap. 11) on the relation between least-squares fits and correlation. Furthermore, "partial" correlation coefficients have been defined and have many useful properties, see the interesting example given by Hamilton (1964, pp. 184-188, particularly Example 5-9-2). However, our use of correlation coefficients per se here will be restricted just to their usefulness as qualitative indicators. For example, it will be shown (i) in Sec. G-2-a that nonzero correlation among the values of the molecular constants indicates that the general rules of propagation of error, which involve the covariances as well as the variances, must be used when computing the standard errors of quantities calculated from the molecular constants; (ii) in Sec. G-1-d that any large correlation coefficient

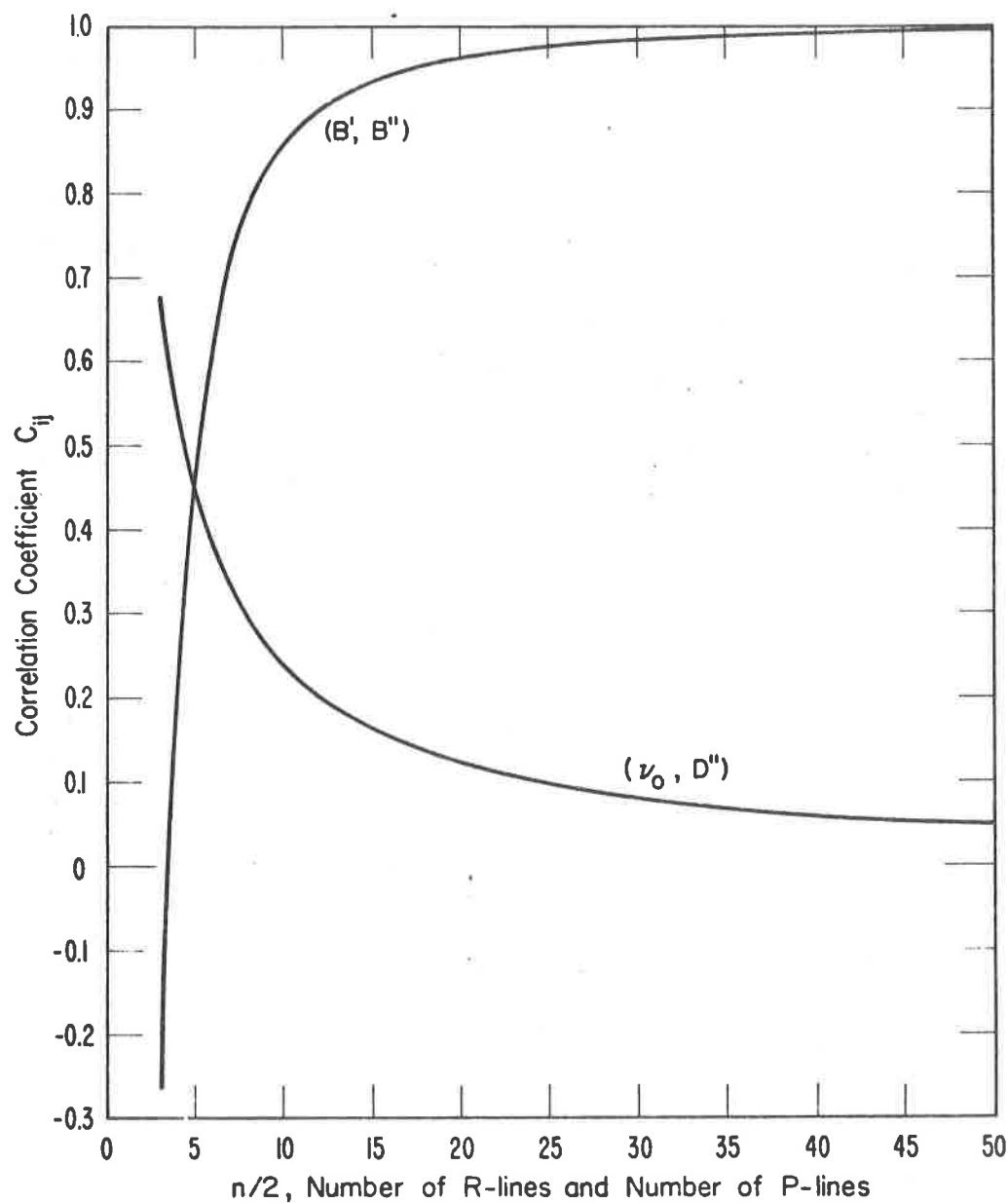


Figure D-8. Variation of the correlation coefficients $C(v_0, D'')$ and $C(B', B'')$ as a function of the number of R-lines, $R(J''=0), R(1), \dots, R(\frac{n}{2}-1)$, and the equal number of P-lines, $P(J''=1), P(2), \dots, P(\frac{n}{2})$, in the coefficient matrix \underline{X} .

among a set of molecular constants indicates that the least-squares values for these constants often cannot, without loss of information, be rounded to reflect one standard error; and (iii) in Sec. F-1 that large correlation coefficients among the input quantities of a least-squares fit indicates that the most general formulation of the least-squares method must be used.

3. Confidence Limits

It was stressed in Sec. D-2-a that the least-squares values $\hat{\beta}$ for the molecular constants β , obtained from a set of measured line positions, are only estimates of these constants and generally differ from the unknown "true" values. Furthermore, it was shown in Sec. D-2-b that the estimated standard errors $\hat{\sigma}_{ii}^{1/2}$ were measures (deliberately vague) of how much the estimates $\hat{\beta}$ could be expected to be different from the "true" values. In this Section, this measure is made quantitative in terms of confidence limits (often referred to as confidence intervals), which permit an expression of the probability (i.e., confidence) that the unknown "true" value lies within a specified range about an estimated value $\hat{\beta}_i$.

(a) Normal distribution. However, to construct confidence limits, which reflect probability, the distribution function

of the measurement errors must be specified. We use here the normal distribution, because of the tendency of the errors occurring in many types of measurements to be normally distributed. This pleasant situation is most likely due to the central-limit theorem, which states that if an error ϵ_i is composed of a sum of errors from several independent sources, each component having an arbitrary error distribution, then the sum will have a distribution that becomes more nearly normal as the number of individual components grows larger (Draper and Smith, 1966, p. 17). Thus, the assumption of a normal distribution is not unreasonable and, at any rate, the error distribution associated with a given experimental situation can be tested for normality, as will be described below. The literature of the normal distribution is enormous and the references cited here are obviously only a small fraction.

The form of the normal equation that is useful here is the one that yields a normal curve that has the same area as a given frequency-of-occurrence histogram (Dixon and Massey, 1969, p. 57)

$$\varphi(x) = \frac{n\ell}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad (D-12)$$

where φ is the height of the curve at the value x , n is the total number of occurrences, ℓ is the length of the class interval of the histogram (i.e., the width of each vertical "bar"), σ and μ

are the standard deviation and mean, respectively, of the distribution. The area under such a curve is $n\ell$.

Figure D-9 gives an example of how the errors in spectroscopic line position measurements frequently approach a normal distribution. The $n=272$ residuals between the observed lines and those calculated from the fitted molecular constants of the (3,1) band of the O_2^+ $b\ ^4\Sigma_g^- - a\ ^4\Pi_u$ First Negative system (Albritton, Schmeltekopf, Zare, and Czarny, 1975) were sorted into narrow intervals about zero. The width of each interval was $\ell = \hat{\sigma}/5$, where $\hat{\sigma} = 0.0306\text{ cm}^{-1}$. Thus, the histogram in Fig. D-9 shows the frequency of occurrence of the different sizes of errors about $\hat{\mu} = 0$. The solid curve is the normal curve computed from Eq. (D-12) with n , ℓ , $\hat{\mu}$, and $\hat{\sigma}$ as given. It is clear from Fig. D-9 that the shape is roughly normal.

However, a comparison that the eye can better judge is the simple and useful one shown in Fig. D-10, which was constructed in the following way. The number of occurrences in each class interval (i.e., "bar") of Fig. D-9 is divided by $n = 272$, to convert it to a percentage. Then, the resulting percentages are summed from left to right and the accumulating total is plotted in Fig. D-10 after each step of the summation. The resulting distribution as a function of the class interval bears the bulky full name of "cumulative percentage frequency-of-occurrence distribution." The grid of this figure is that of "normal probability graph paper," which is

commonly available and the vertical axis of which is constructed so that a cumulative normal distribution will be a straight line, as is represented by the dashed line in Fig. D-10. The solid line is the cumulative percentage distribution corresponding to the histogram of O_2^+ quartet residuals (Fig. D-9), which should, on the average, follow the dashed line, if the residuals are normally distributed.

The attractiveness of this type of comparison is that we all have an intuitive feeling of when a line is nearly straight. In Fig. D-10, the solid line is indeed nearly straight, except at the upper end. This departure is due, for the main part, to the two "outliers" on the far right side of the histogram in Fig. D-9, whose rejection could be considered using one of several criteria for removing such measurements from the fit (see Worthing and Geffner, 1943, p. 170; Dixon and Massey, 1969, p. 328; and Anscombe, 1960). Thus, our conclusion from the cumulative percentage distribution plot in Fig. D-10 is that the residuals are nearly normally distributed. If the model is believed to be adequate, as it is in this case, then the residuals are taken to be reliable estimates of the unknown measurement errors and the conclusion of near-normality is extended to the unknown measurement errors themselves.

It should be stressed in using such cumulative percentage

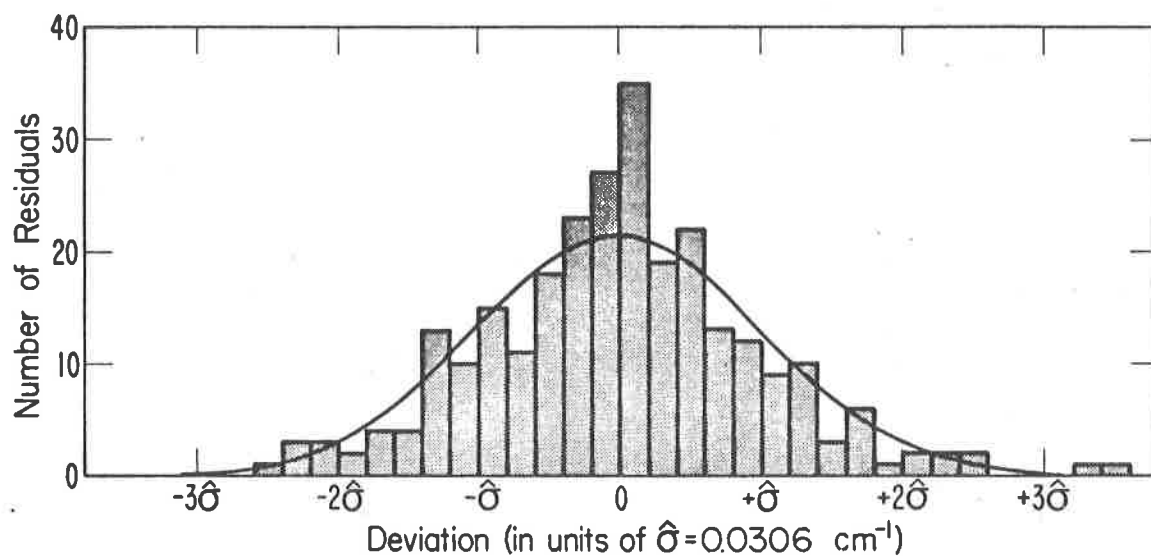


Figure D-9. Distribution of the residuals between observed and calculated lines of the $(3,1) \text{ O}_2^+ \text{ b } ^4\Sigma_g^- - \text{ a } ^4\Pi_u$ First-Negative band. The solid curve is a normal distribution (Eq. D-12) with $n = 272$, $\hat{\mu} = 0$, and $\hat{\sigma} = 0.0306 \text{ cm}^{-1}$.

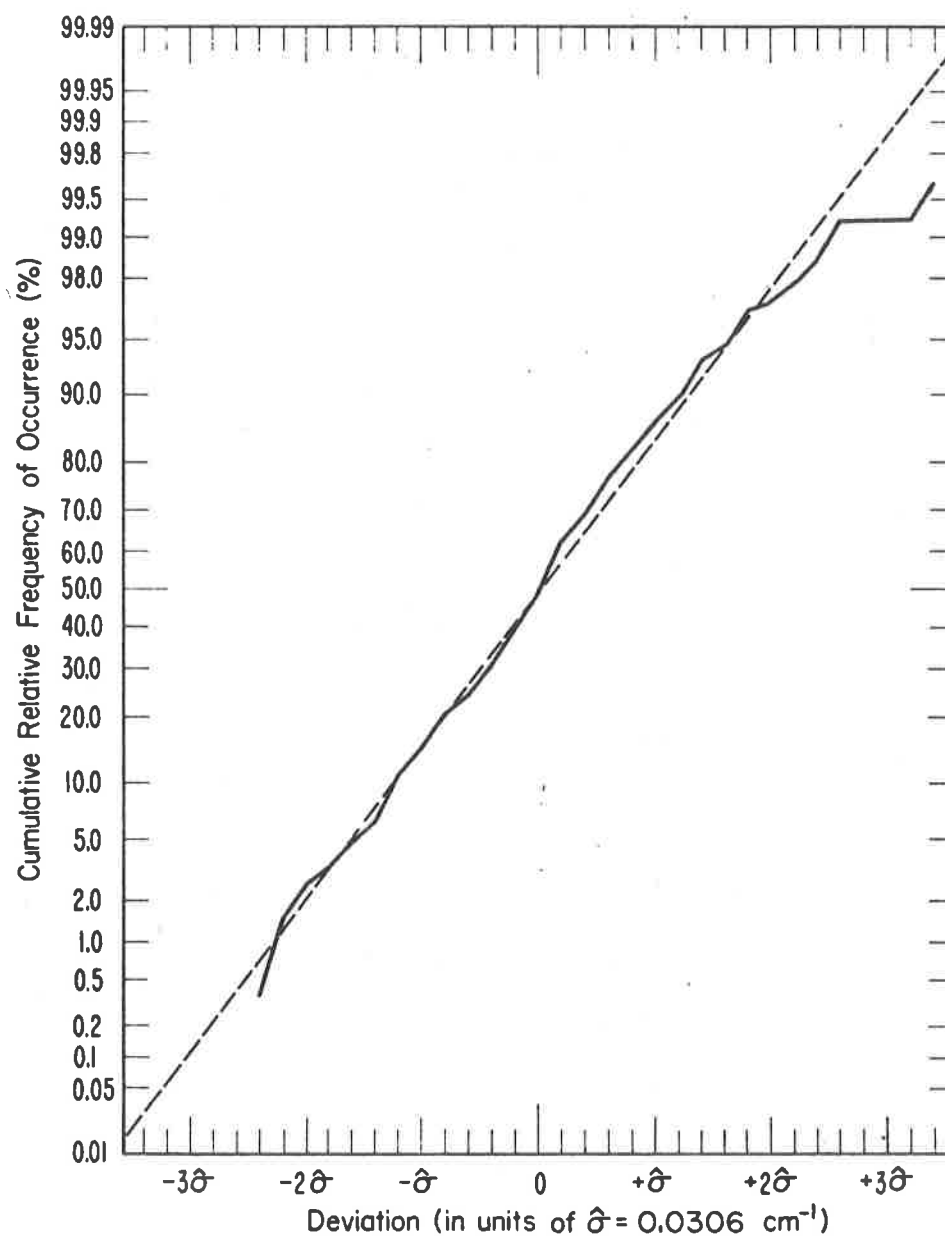


Figure D-10. Cumulative percentage distribution of the residuals between observed and calculated lines of the $(3,1) \text{ O}_2^+ \text{ b } ^4\Sigma_g^- - \text{a } ^4\Pi_u$ First-Negative band. The dashed line represents a cumulative normal distribution.

plots that even if the residuals were drawn randomly from a distribution that is known to be normal, the cumulative percentage distribution would still be a zig-zag line with varying degrees of "zig" and "zag," because of the randomness of the samples. The line will, however, get smoother as the sample size n increases and the class interval h decreases. In order to give a feeling of how cumulative percentage plots of truly normal distributions can look, Bennett and Franklin (1954, pp. 92-94) have given very useful plots of random samples of size 50, 100, and 200. Other examples, some normal and one non-normal, are given by Dixon and Massey (1969, pp. 61-66). Both of these references discuss this and other tests of normality in detail.

(b) Construction of the confidence limits. Provided the measurement error distribution is approximately normal, we are now in position to determine limits $\hat{\beta}_i \pm \hat{\omega}_i$ centered on the estimate $\hat{\beta}_i$ within which we may be "reasonably confident" that the unknown "true" value of β_i lies, provided we define what is meant by "reasonably confident." If $\hat{\omega}_i$ is of such magnitude that the chances are about 95 in 100 that the "true" value lies inside the limits $\hat{\beta}_i \pm \hat{\omega}_i$ and about 5 in 100 that it lies outside, then such limits are called 95% confidence limits; i.e., if we state that the "true" value of β_i is between the limits $\hat{\beta}_i - \hat{\omega}_i$ and $\hat{\beta}_i + \hat{\omega}_i$, then we shall be correct 95% of the time. Alternatively, the

risk is 5% that we will be making an incorrect statement. To the more cautiously inclined, a 5% risk of being wrong may seem too high. If so, 99% confidence limits, with a 1% risk of being wrong, may seem more attractive. Of course, for a given problem, the 99% limits will always be more widely spaced than will be the 95% limits. The point to be made here is that one is free to select the half width $\hat{\omega}_i$ of the confidence limits to reflect one's own taste in balancing the desire to be specific against the fear of being wrong. Confidence levels of 95 and 99% are popular choices (see Bennett and Franklin, 1954, pp. 148-149).

Specifically, within the assumptions of the least-squares method and the assumption of a normal distribution of measurement errors, one can be $100(1-\alpha)\%$ confident that the unknown "true" value of β_i lies within the interval

$$\hat{\beta}_i \pm t(n-m, 1-\alpha) \hat{\omega}_{ii}^{\frac{1}{2}} \quad (D-13)$$

Thus, confidence limits are given by some appropriate multiple of the standard error. The multiplicative factor is a function of the degrees of freedom $n-m$ (sometimes referred to as df or ν) and the selected degree of confidence $1-\alpha$, and is referred to as "Student's" t -factor [titled after its originator W. S. Gosset (1876-1937), who was a chemist employed by the Guinness Breweries and published pioneering statistical papers under the pseudonym

of "Student"]. The magnitude of the t-factor takes into account the possible variation of $\hat{\beta}_i$ from the true value in units of the estimated standard error $\hat{\sigma}_{ii}^{1/2}$ and the reliability of this estimated standard error as a unit (i.e., the possible variation of $\hat{\sigma}^2$ from the "true" variance). Student's t-factors are tabulated as functions of n-m and 1- α (or more frequently 1- $\frac{1}{2}\alpha$). Among the references cited here, Dixon and Massey (1969) have the most extensive and useful set of statistical tables. A large number of t(df, 1- $\frac{1}{2}\alpha$) factors are listed in their Table A-5.

For our use here, Fig. D-11 gives the t-factors for four levels of confidence: 90, 95, 99, and 99.9%. For the example of 50 R-lines and 50 P-lines and 5 molecular constants, the degrees of freedom are 95. Choosing a 95% level of confidence, Fig. D-11 yields a t-factor of 1.99, or conveniently rounded to 2.0. Thus, 95% confidence limits for, say, D'' would be $\hat{D}'' \pm 2.0 \hat{\sigma}_{55}^{1/2}$, or from the first line of Table D-2, $\hat{D}'' \pm 2.0 \hat{\sigma} (1.3 \times 10^{-6}) = \hat{D}'' \pm 2.6 \hat{\sigma} \times 10^{-6}$. From the \hat{D}'' and $\hat{\sigma}$ values of the 1000 least-squares fits described in Sec. D-2-a, this 95% confidence limit can be tested, since it predicts that the true value $6 \times 10^{-6} \text{ cm}^{-1}$ will lie inside of the limits $\hat{D}'' \pm 2.6 \hat{\sigma} \times 10^{-6}$ for 950 of the $(\hat{D}'', \hat{\sigma})$ values. In fact, 961 of the confidence limits included the true value, only 1% different from the prediction.

Figure D-11 shows that each t-factor converges very rapidly

to nearly a constant value as the degrees of freedom increase. For $n-m \gtrsim 30$, the t-factor for 95% confidence is approximately 2.0 and for $n-m \gtrsim 60$, the t-factor for 99% confidence is approximately 2.6. Since a 95% confidence limit has been a popular choice and since degrees of freedom are frequently greater than 30, it is clear why " \pm two standard errors" are often quoted as the "experimental uncertainty" in reported least-squares values. It should now be equally clear, however, that such a statement makes implicit assumptions about the degrees of freedom, confidence level, and approximate normality.

Confidence limits provide one of the simplest ways of model testing (Kirchhoff, 1972). If a particular confidence limit $\hat{\beta}_i \pm \omega_i$ includes zero, then to the chosen level of confidence, the estimate $\hat{\beta}_i$ is not significant; i.e., there is no reason to believe that the "true" value of β_i is different from zero. Thus this molecular constant may be omitted from the model. On the other hand, if the confidence limit does not include zero, then to the chosen level of confidence, the estimate $\hat{\beta}_i$ is significant. There are, of course, many more sophisticated and powerful ways of model testing. See, for example, Curl (1970); Draper and Smith, 1966, pp. 24-26, 71-72; Hamilton, 1964, pp.168-173; and Beaton and Tukey, (1974).

The confidence limits given by Eq. (D-13) describe the "limits of error," for a chosen degree of confidence, for any

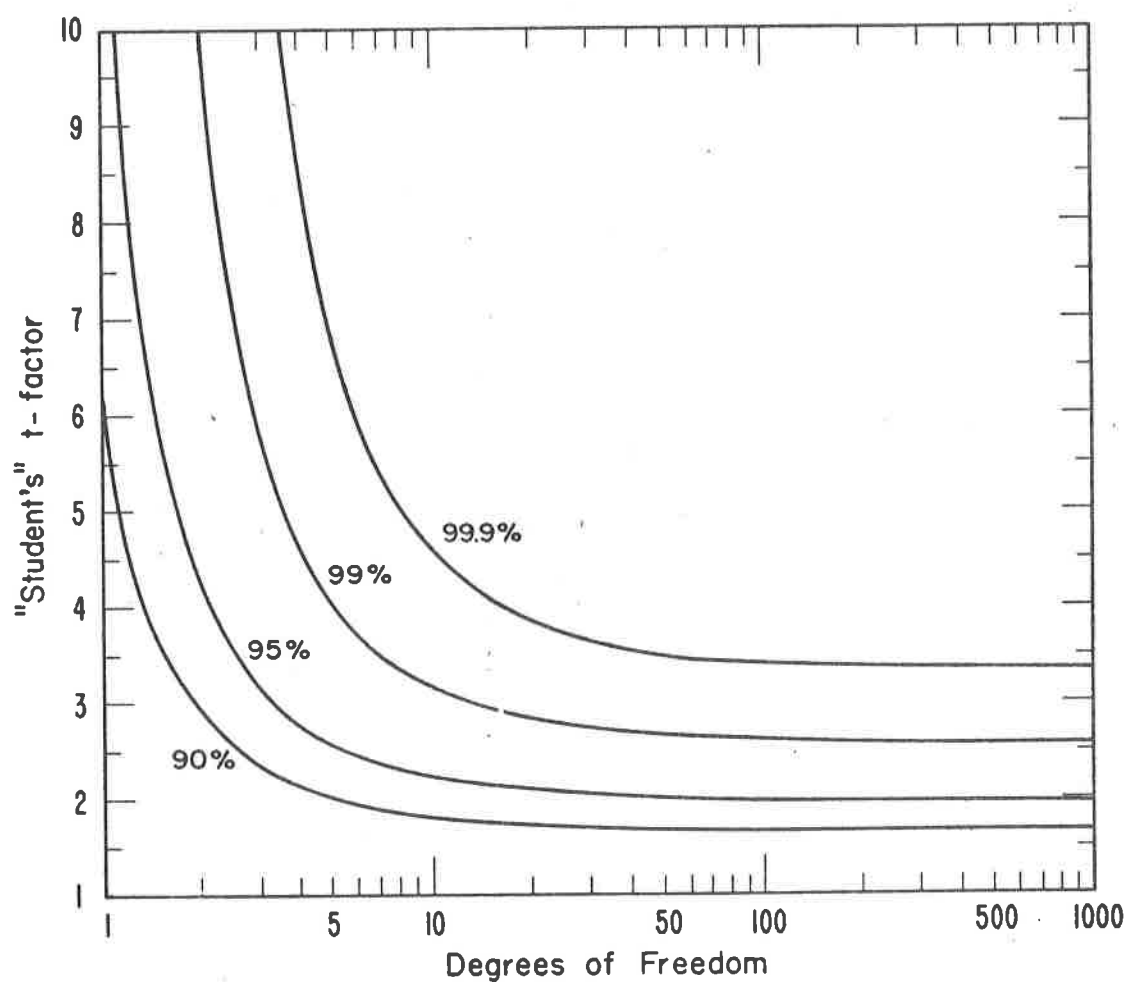


Figure D-11. "Student's" t-factor for 90, 95, 99, and 99.9% confidence limits as a function of the degrees of freedom.

one molecular constant. However, it was shown in Sec. D-2-c that when the correlation coefficients of the molecular constants are nonzero, the errors of the estimated values of the constants are not free to independently have any values over the separate confidence intervals. This shows the utility of considering a joint confidence region for two highly correlated constants. For a pair of constants, for example, the joint confidence region is an ellipse enclosed within a rectangle whose sides are approximately the separate confidence limits. The large correlation coefficient between D' and D'' implies that there are large regions of the rectangle, whose sides are given by the separate confidence intervals $\hat{D}' \pm t_{\hat{\theta}_{33}}^{\frac{1}{2}}$ and $\hat{D}'' \pm t_{\hat{\theta}_{55}}^{\frac{1}{2}}$, that are virtually inaccessible to values of \hat{D}' and \hat{D}'' simultaneously. The true region accessible to simultaneous \hat{D}' and \hat{D}'' is a thin ellipse inside the rectangle, the shape of which can be visualized by the scatter of the 50 pairs of estimates (\hat{D}', \hat{D}'') in Fig. D-6. The near-zero correlation coefficient of ν_0 and D'' , for example, implies that the joint confidence interval for these two will be essentially as large as the rectangle formed by their separate confidence limits. The scatter of the 50 pairs of estimates $(\hat{\nu}_0, \hat{D}'')$ in Fig. D-7 gives an indication of this. Further discussion and application of joint confidence regions are given by Draper and Smith (1966, pp. 64-67).

E. WEIGHTED LEAST-SQUARES FORMULATION

In this section, we examine the least-squares formulation that is appropriate when the data to be fitted are known to be of unequal precision, in contrast to the equally precise data considered in Sec. D. Again, a familiar spectroscopic example is used to illustrate this formulation.

1. Statement of the Problem.

A set of p measured line positions y_i , $i = 1, 2, \dots, p$ for an infrared vibration-rotation band and a set of q measured line positions y_i , $i = p+1, p+2, \dots, p+q \equiv n$ for microwave transitions between the rotational levels of the lower vibrational level of the infrared band are both used together to obtain least-squares estimates of the familiar molecular constants ν_0 , B' , D' , H' , B'' , D'' , and H'' . An example of this problem could be the determination of the molecular constants of the $v = 0$ and 2 vibrational levels of the $\text{CO } X^1\Sigma^+$ state from the R- and P-lines of the (2,0) infrared vibration-rotation band and the microwave transitions between the adjacent rotational levels, i.e., $J \rightarrow J+1$, of the $v = 0$ level. These data sets are clearly of significantly different precision; thus, this application requires the weighted least-squares

formulation.

The model is expressed by the n equations relating the molecular constants to the measured line positions, p equations for the infrared lines:

$$Y_i = \nu_0 + B'J_i'(J_i'+1) - D'[J_i'(J_i'+1)]^2 + H'[J_i'(J_i'+1)]^3 - B''J_i''(J_i''+1) + D''[J_i''(J_i''+1)]^2 - H''[J_i''(J_i''+1)]^3 + \epsilon_i \quad i = 1, 2, \dots, p, \quad (E-1a)$$

and q equations for the microwave lines:

$$Y_i = B''^2(J_i''+1) - D''^4(J_i''+1)^3 + H''^2(J_i''+1)^3 (3J_i''^2 + 6J_i'' + 4) + \epsilon_i \quad i = p+1, p+2, \dots, p+q = n \quad (E-1b)$$

The least-squares assumptions (Sec. C-2) of this problem are the same as those made for the unweighted problem in Sec. D-1, except that here the unknown measurement errors ϵ_i are assumed to be described by distributions of generally unequal variances σ_i^2 , but whose ratios σ_i^2/σ_j^2 , $i \neq j$, are adequately known.

In matrix notation, Eq. (E-1) can again be written quite succinctly as

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon} \quad (E-2)$$

The vector $\underline{\beta}$ of the seven unknown molecular constants is given by

$$\underline{\beta}^T = [\nu_0 \ B' \ D' \ H' \ B'' \ D'' \ H''] \quad (E-3)$$

and the vectors \underline{y} and $\underline{\epsilon}$ of the measurements and their errors have the same form given earlier in Eq. (D-3). The matrix \underline{X} is defined by the coefficients in Eq. (E-1). For the specific case (representative of CO X $^1\Sigma^+$) of the first 50 R-lines [R(J''=0), R(1), ..., R(49)] and the first 50 P-lines [P(J''=1), P(2), ..., P(50)] in a vibration-rotation band and the first 7 rotational transitions (J''=0) \rightarrow (J''=1), (J''=1) \rightarrow (J''=2), ..., (J''=6) \rightarrow (J''=7), the $i = 1, 50, 51, 100, 101$, and 107 lines of the coefficient matrix \underline{X} [which correspond to the R(J''=0), R(49), P(1), P(50), (J''=0) \rightarrow (J''=1), and (J''=6) \rightarrow (J''=7)] are

$$\underline{X} = \begin{bmatrix} 1 & 2 & -4 & 8 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2550 & -6502500 & 16581375000 & -2450 & 6002500 & -14706125000 \\ \\ 1 & 0 & 0 & 0 & -2 & 4 & -8 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 2450 & -6002500 & 14706125000 & -2550 & 6502500 & -16581375000 \\ \\ 0 & 0 & 0 & 0 & 2 & -4 & 8 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 14 & -1372 & 101528 \end{bmatrix}$$

(E-4)

In matrix notation, the assumed variances and covariances of the measurement errors \underline{e} take a particularly concise form; namely, the variance-covariance matrix $\underline{\Phi}$ of the measurement errors has the $n \times n$ diagonal form

$$\underline{\Phi} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{bmatrix}, \quad (E-5)$$

where the ratios σ_i^2/σ_j^2 , $i \neq j$, are assumed to be known. Alternatively, the assumption that the ratios σ_i^2/σ_j^2 are known can be

stated equivalently by the assumption that $\underline{\Phi}$ is known to within a common factor, or

$$\underline{\Phi} = \sigma_n^2 \underline{M}, \quad (\text{E-6})$$

where σ_n^2 is the unknown common factor and the measurement error dispersion matrix \underline{M} is a known diagonal matrix whose nonzero elements are ratios of variances; for example,

$$\underline{M} = \begin{bmatrix} \sigma_1^2/\sigma_n^2 & & & 0 \\ & \sigma_2^2/\sigma_n^2 & & \\ 0 & & \ddots & \\ & & & 1 \end{bmatrix}. \quad (\text{E-7})$$

The location of the unity element(s) is, of course, arbitrary, but we shall see in Sec. E-2 below that the location of the unity element(s) determines the interpretation of the common factor σ_n^2 .

In summary, the weighted least-squares formulation should be used when the variance-covariance matrix $\underline{\Phi}$ of the measurement errors has at least some unequal variance elements on-diagonal, zero covariance elements off-diagonal, and can be specified within a common unknown factor. Thus, a perhaps more accurate, but less

recognizable, label for this formulation would be the "unequally weighted, uncorrelated" least-squares. Finally, it should be clear that the unweighted formulation described in Sec. D is just a special case of the present formulation with $\underline{\Phi} = \sigma^2 \underline{I}$, where σ^2 is the common unknown variance of all of the measurements errors and \underline{I} is the identity matrix.

2. Variance-Covariance Matrix of the Measurement Errors

The matrix $\underline{\Phi}$, or \underline{M} , is obviously the key factor in the weighted least-squares formulation and it is worthwhile examining its properties in detail.

As will be shown below, it is actually the inverse of $\underline{\Phi}$ that appears in the expression for the estimated molecular parameters $\hat{\underline{\theta}}$. This inverse is given by

$$\underline{\Phi}^{-1} = \begin{bmatrix} 1/\sigma_1^2 & & & \\ & 1/\sigma_2^2 & & \\ & & \ddots & \\ 0 & & & 1/\sigma_n^2 \end{bmatrix} \quad (E-8)$$

The matrix

$$\underline{W} \equiv \underline{M}^{-1} = \begin{bmatrix} \sigma_n^2 / \sigma_1^2 & & & 0 \\ & \sigma_n^2 / \sigma_2^2 & & \\ 0 & & \ddots & \\ & & & 1 \end{bmatrix} \quad (\text{E-9})$$

is frequently referred to as the weight matrix, hence the name "weighted least-squares." It is clear from Eq. (E-9) that the weights, i.e., the diagonal elements of \underline{W} , are inversely proportional to the variances σ_i^2 . The common factor σ_n^2 , therefore, is the unknown variance of the measurement(s) with unit weight.

The terminology of "weights" can actually be a source of confusion unless it is accompanied by explicit definitions. For example, occasionally one sees where the so-called weight matrix in a weighted least-squares was constructed with diagonal elements inversely proportional to σ_i , rather than σ_i^2 . As we shall see below, this means that the $\hat{\underline{\beta}}$ values were not minimum-variance estimates.

Since $\underline{\Phi}$ must be known to within a common factor before $\hat{\underline{\beta}}$ can be calculated, its construction is the first step in applying the weighted least-squares formulation. This choice of the relative sizes of the diagonal elements of $\underline{\Phi}$ may appear, at first, to introduce an intolerable degree of arbitrariness into this least-squares formulation, since clearly by altering the relative

sizes of the elements of $\hat{\Phi}$, the values of $\hat{\Phi}$ can indeed be manipulated almost at will. However, in most cases, an honest assignment of the relative sizes of the elements $\hat{\Phi}_{ii}$ can be made without undue arbitrariness. Consider two extremes. First, for a group of line position measurements made under one set of experimental conditions (same wavelength range, resolving power, exposure time, etc.), it may be difficult to estimate with appreciable confidence that one line is significantly (i.e., $\sigma_a/\sigma_b \approx 1.5?$) more precise than others. In such cases, the unweighted, least-squares formulation of Sec. D would probably suffice, in view of the inability to segregate measurements of clearly different precision. On the other hand, the present example of combining infrared and microwave measurements in one least-squares fit represents nearly the other extreme. Taking the precision of an infrared line position measurement to be $\tilde{\sigma}_{IR} \approx 10^{-3} \text{ cm}^{-1}$ (Rao, 1972, p. 349) and the precision of a microwave transition to be $\tilde{\sigma}_M \approx 0.03 \text{ MHz} = 10^{-6} \text{ cm}^{-1}$ (Morino and Saito, 1972, p. 11), it is fairly clear that a ratio of $\tilde{\sigma}_{IR}^2/\tilde{\sigma}_M^2 = 10^8$ would be representative and it is absolutely clear that the weighted least-squares formulation must be used. (Here, we use the tilde \sim to denote an estimate that is only a "guess" or "hunch"). It is fortunate that only approximate relative variances prove to be generally adequate. For example, an analysis of the residuals of a preliminary unweighted least-squares fit often can provide adequate estimates of the weighting factors, should subdivisions of the data show residuals

of significantly different magnitudes. As Tukey (1974, p. 5) points out, the assignment of relative variances in $\underline{\Phi}$, if reasonably done, is almost never a crucial issue in least-squares applications.

To this we add that if the assignment of significantly different relative variances can indeed be reasonable done, then it always should be done.

Continuing with the infrared-microwave example, the matrix $\underline{\Phi}$ could be given by

$$\underline{\Phi} = \sigma_M^2 \begin{bmatrix} 10^8 & & & 0 \\ & \ddots & & \\ & & 10^8 & \\ 0 & & & 1 & \ddots & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix} \quad (E-10)$$

and in this case σ_M^2 represents the unknown variance of the microwave data. However, this construction of $\underline{\Phi}$ is not unique, since $\underline{\Phi}$ only need be specified within a common factor. One alternative is clearly

$$\underline{\Phi} = \sigma_{IR}^2 \begin{bmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ 0 & & & 10^{-8} & \ddots & \\ & & & & 10^{-8} \end{bmatrix} \quad (E-11)$$

and in this case σ_{IR}^2 represents the unknown variance of the infrared data. Another alternative is to use, not the ratios, but the crude estimates $\tilde{\sigma}_{\text{IR}}^2$ and $\tilde{\sigma}_{\text{M}}^2$ themselves,

$$\underline{\Phi} = \sigma^2 \begin{bmatrix} 10^{-8} & & & 0 \\ & \ddots & & \\ & & 10^{-8} & \\ 0 & & & 10^{-12} \\ & & & & \ddots \\ & & & & & 10^{-12} \end{bmatrix} \quad (\text{E-12})$$

In this case, σ^2 represents the unknown variance of neither the microwave nor the infrared data, since neither of these are represented by unity elements, but rather will have an estimated value $\hat{\sigma}^2$ near unity, provided $\tilde{\sigma}_{\text{IR}}$ and $\tilde{\sigma}_{\text{M}}$ are realistic, as will be discussed in the next part below. Since the three constructions of $\underline{\Phi}$ in Eqs. (E-10), (E-11), and (E-12) differ only by constant factors, we shall see that they will yield the same values for $\hat{\beta}$.

If $\underline{\Phi}$ is unwarrantedly set equal to $\sigma^2 \underline{I}$ when the data are of significantly unequal precision, the resulting estimated molecular constants $\underline{\beta}^*$ will still be unbiased linear estimates, but they will not be minimum-variance estimates (Hamilton, 1964, p. 146 and Draper and Smith, 1966, p.80). This means that if the measurements could be made repeatedly under identical experimental conditions,

except of course for random errors, the resulting sets of estimates $\underline{\theta}^*$ will scatter about the "true" values in a way that the average approaches the true values as the number of repetitions grows, but the scatter, as indicated by the variance, will be larger than that occurring when the proper $\underline{\theta}$ is used. This wider scatter is not a trivial matter, since it means that the chances are greater that the one set of estimates $\underline{\theta}^*$ calculated with $\underline{\theta} = \sigma^2 \underline{I}$ from the one set of available measured line positions will be further from the "true" values than the estimates $\hat{\underline{\theta}}$ calculated with the appropriate $\underline{\theta}$ from the same measurements. Of course, this does not prohibit a given set of estimates $\underline{\theta}^*$ from being nearly identical to the "true" values, but over a scientific lifetime, a scorn of proper weighting will surely take its toll.

3. Estimated Molecular Parameters

For the weighted least-squares formulation, the expression for the estimated molecular constants takes on a more general form than that given earlier in Sec. D. The appropriate new expression is stated here as a prescription; it is derived in the Appendix. In particular, the MVLU estimates $\hat{\underline{\theta}}$ of the molecular constants are calculated from the expression

$$\hat{\underline{\beta}} = (\underline{X}^T \underline{M}^{-1} \underline{X})^{-1} \underline{X}^T \underline{M}^{-1} \underline{y} , \quad (\text{E-13})$$

which, of course, necessarily involves only known quantities: namely, the coefficient matrix \underline{X} , the measurement error dispersion matrix \underline{M} , and the vector of measurements \underline{y} . One frequently sees Eq. (E-13) written explicitly in terms of the weight matrix \underline{W} (e.g., see Kildal, Eng, and Ross, 1974, Eq. 4). Furthermore, it may also be written in terms of $\underline{\Phi}$, rather than \underline{M} or \underline{W} , since the common factor σ^2 will cancel out:

$$\hat{\underline{\beta}} = (\underline{X}^T \underline{\Phi}^{-1} \underline{X})^{-1} \underline{X}^T \underline{\Phi}^{-1} \underline{y} . \quad (\text{E-14})$$

For the special case of $\underline{\Phi} = \sigma^2 \underline{I}$, Eqs. (E-13) and (E-14) are identical to Eq. (D-6) given earlier.

4. Estimated Variance

The common unknown factor σ^2 in the variance-covariance matrix $\underline{\Phi}$ of the measurement errors is calculated from the expression

$$\hat{\sigma}^2 = \frac{1}{n-m} (\underline{y} - \underline{X}\hat{\underline{\beta}})^T \underline{M}^{-1} (\underline{y} - \underline{X}\hat{\underline{\beta}}) , \quad (\text{E-15})$$

where $n-m$ are the degrees of freedom. For the special unity-weighted

case where $\underline{M} = \underline{I}$, Eq. (E-15) is the same as Eq. (D-7). As remarked earlier, for the special case of $\hat{\underline{\theta}}$ constructed with estimates of the measurement variances $\tilde{\sigma}_i^2$ (e.g., Eq. E-12), $\hat{\sigma}^2$ will be approximately unity if the estimates are reasonable. This can be seen from Eq. (E-15), which can be written in summation form as

$$\hat{\sigma}^2 = \frac{1}{n-m} \sum_{i=1}^n (y_i - \hat{y}_i)^2 / \tilde{\sigma}_i^2 . \quad (\text{E-16})$$

Thus, if, on the average, $\tilde{\sigma}_i^2$ is an adequate estimate of σ_i^2 and if the residual $y_i - \hat{y}_i$ is an adequate estimate of the measurement error ϵ_i , then $(y_i - \hat{y}_i)^2 / \tilde{\sigma}_i^2$ will fluctuate randomly about approximately unity. This implies that $\hat{\sigma}^2$ will be near unity. Therefore, this particular construction of $\hat{\underline{\theta}}$ provides a useful test of the reasonableness of the estimated variances $\tilde{\sigma}_i^2$.

5. Estimated Variances and Covariances of the Molecular Parameters

The $m \times m$ (for the present example, $m=7$) symmetric variance-covariance matrix $\hat{\underline{\Theta}}$ of the estimated molecular parameters $\hat{\underline{\theta}}$ is given by

$$\hat{\underline{\Theta}} = \hat{\sigma}^2 (\underline{X}^T \underline{M}^{-1} \underline{X})^{-1} . \quad (\text{E-17})$$

For the special unweighted case where $\underline{M} = \underline{I}$, Eq. (E-17) is the same as Eq. (D-9).

Like $\hat{\underline{\beta}}$, $\hat{\underline{\Theta}}$ is oblivious to a common factor in \underline{M} or \underline{W} . This can be seen by using Eq. (E-15) to eliminate $\hat{\sigma}^2$ from Eq. (E-17)

$$\hat{\underline{\Theta}} = \frac{1}{n-m} (\underline{y} - \underline{X}\hat{\underline{\beta}})^T \underline{M}^{-1} (\underline{y} - \underline{X}\hat{\underline{\beta}}) (\underline{X}^T \underline{M}^{-1} \underline{X})^{-1} \quad , \quad (\text{E-18})$$

which, despite its ungainliness, shows that any common factor always cancels. This is a point worth remembering: specifically, if \underline{M} is replaced by $\underline{M}' = c\underline{M}$, where c is any constant, then $\hat{\underline{\Theta}}$ given by Eq. (E-17) and $\hat{\underline{\beta}}$ given by Eq. (E-13) are unchanged. Only $\hat{\sigma}^2$ is changed, $(\hat{\sigma}^2)' = \hat{\sigma}^2/c$; i.e., c is simply absorbed into $\hat{\sigma}^2$. Since $\hat{\sigma}^2$ alone absorbs any constant factor in the assigned variance-covariance matrix $\underline{\Phi}$ of the measurement errors, then it is understandable why the values of $\hat{\sigma}^2$ from weighted least-squares fits can have such "curious" magnitudes.

Returning to the example of combining infrared vibration-rotation data with microwave data, it is very instructive to examine the elements of the resulting estimated variance-covariance matrix $\hat{\underline{\Theta}}$ of the seven molecular constants, since they are the indicators of how well the molecular constants have been determined and what gains have been made by combining the two types of data.

For the \underline{X} matrix in Eq. (E-4) and the $\underline{\Phi}$ matrix in Eq. (E-10), with $\hat{\sigma}_M$ taken to be 10^{-8} cm^{-1} , the $\hat{\underline{\Theta}}$ matrix is given as the diagonal and upper-right elements of Table E-1. The standard errors $\hat{\Theta}_{ii}^{1/2}$ are given in parentheses and the correlation coefficients are given as the lower-left elements. First, it is interesting to see the effect of the asymmetries introduced by the microwave rotational

lines. For example, the standard errors of B", D", and H" are factors of 11, 1.3, and 1.2 smaller than those of B', D', and H', respectively. Furthermore, the correlation coefficient B" with D' is 0.6656, but B' with D" is only 0.0417. All of these would have been equal with only the infrared vibration-rotation lines.

Secondly, it is even more interesting to examine what gains have been made by combining the two types of data. This is done in Table E-2, in terms of 95% confidence limits, assuming that the measurement errors are normally distributed. It will be recalled from Sec. D-3-b that such a limit is $\hat{\beta}_i \pm t(n-m, 0.95) \hat{\sigma}_{ii}^{1/2}$ and it represents the interval within which it may be said that the "true" value of β_i lies, with a 5% risk of being wrong. The first line gives the 95% confidence limits for the infrared vibration-rotation band alone. These data have two advantages with regard to determining a statistically significant value for the elusive molecular constant H. First, since they have high degrees of freedom, the t-factor, $t(93, 0.95) = 1.989$, is near its minimum value (see Fig. D-11). Secondly, since the data go to high rotational levels ($J_{\max} = 50$ here), the dispersion matrix element is small, $V_{\gamma\gamma}^{1/2} = 1.332 \times 10^{-9}$. As shown, the 95% confidence interval for H' (and H", since these lines are "symmetric") is $\hat{\beta}_H \pm t(93, 0.95) \hat{\sigma}_{\gamma\gamma}^{1/2} = \hat{\beta}_H \pm 1.989 \times 1.332 \times 10^{-12} = \hat{\beta}_H \pm 2.65 \times 10^{-12} \text{ cm}^{-1}$. This means that an estimated value $\hat{H} = 5 \times 10^{-12} \text{ cm}^{-1}$, which is fairly

Table E-1. Variances $\hat{\sigma}_{ii}$, standard errors $\hat{\sigma}_{ij}$ ($i \neq j$), covariances $\hat{\sigma}_{ij}^k$ ($i \neq j$), and correlation coefficients of the molecular constants ν_0 , B' , D' , H' , B'' , D'' , and H'' of an infrared vibrational-rotational band of a $^1\Sigma$ state with $R(J''=0)$, $R(1)$, ..., $R(49)$ and $P(J''=1)$, $P(2)$, ..., $P(50)$ and the microwave rotational lines $(J''=0) \rightarrow (J''=1)$, $(J''=1) \rightarrow (J''=2)$, ..., $(J''=6) \rightarrow (J''=7)$, with $\tilde{\sigma}_{IR}^2/\tilde{\sigma}_M^2 = 10^6$ and $\hat{\sigma}_M = 10^{-8}$ cm $^{-1}$.

	ν_0	B'	D'	H'	B''	D''	H''
ν_0	4.961813×10^{-8} (2.23×10^{-4})	$-1.746784 \times 10^{-10}$	$-1.465582 \times 10^{-13}$	$-3.400290 \times 10^{-17}$	2.777002×10^{-13}	4.305116×10^{-16}	2.600422×10^{-18}
B'	-0.7508	1.091055×10^{-12} (1.04×10^{-8})	1.165382×10^{-15}	2.972050×10^{-19}	5.549994×10^{-16}	5.609901×10^{-17}	4.742149×10^{-21}
D'	-0.3959	0.6714	2.761680×10^{-18} (1.66×10^{-9})	8.078619×10^{-22}	1.047655×10^{-18}	1.591628×10^{-18}	4.848787×10^{-22}
H'	-0.2948	0.5496	0.9390	2.680500×10^{-25} (5.18×10^{-13})	3.295717×10^{-20}	5.046215×10^{-22}	1.821949×10^{-26}
B''	0.0132	0.0561	0.6656	0.6721	8.969737×10^{-16} (9.47×10^{-8})	1.092377×10^{-18}	3.430104×10^{-20}
D''	0.0150	0.0417	0.7431	0.7562	0.8949	1.661323×10^{-18} (1.29×10^{-8})	5.256103×10^{-22}
H''	0.0264	0.0103	0.6607	0.7968	0.8201	0.9234	1.950330×10^{-26} (4.42×10^{-13})

^a The variances and standard errors (in parentheses) are on the diagonal. The covariances are above the diagonal and the correlation coefficients are below the diagonal. The units of the variances and covariances are cm $^{-2}$ and those of the standard errors are cm $^{-1}$. The correlation coefficients are dimensionless.

Table E-2. 95% confidence limits for the estimated $^1\Sigma$ -state molecular constants from separate fits to (a) infrared vibration-rotation lines $R(J''=0), R(1), \dots, R(49)$ and $P(J''=1), P(2), \dots, P(50)$; (b) microwave rotational lines $(J''=0) \rightarrow (J''=1), (J''=1) \rightarrow (J''=2), \dots, (J''=6) \rightarrow (J''=7)$; and (c) both the infrared and the microwave lines simultaneously.

number of lines		degrees of freedom	$\hat{\sigma}$ (cm^{-1})	$\hat{\omega}_i$, corresponding to 95% confidence limits, $\hat{B}_i \pm \hat{\omega}_i$ (cm^{-1})						
vib-rot	rot			ν_0	B'	D'	H'	B''	D''	H''
100	0	93	$\hat{\sigma}=10^{-3}$	$\pm 4.44 \times 10^{-4}$	$\pm 1.49 \times 10^{-6}$	$\pm 1.15 \times 10^{-8}$	$\pm 2.65 \times 10^{-12}$	$\pm 1.49 \times 10^{-6}$	$\pm 1.15 \times 10^{-8}$	$\pm 2.65 \times 10^{-12}$
0	7	4	$\hat{\sigma}=10^{-6}$	-	-	-	-	$\pm 5.36 \times 10^{-7}$	$\pm 1.97 \times 10^{-9}$	$\pm 2.11 \times 10^{-10}$
100	7	100	$\hat{\sigma}_{\text{IR}}^2 / \hat{\sigma}_{\text{M}}^2 = 10^3$ $\hat{\sigma}_{\text{M}} = 10^{-6}$	$\pm 4.43 \times 10^{-4}$	$\pm 2.08 \times 10^{-6}$	$\pm 3.30 \times 10^{-8}$	$\pm 1.03 \times 10^{-12}$	$\pm 1.88 \times 10^{-7}$	$\pm 2.56 \times 10^{-9}$	$\pm 8.78 \times 10^{-13}$

typical for this constant, would be statistically significant at the 95% level. In fact, a value of $5 \times 10^{-12} \text{ cm}^{-1}$ would be significant at the 99.95% level [i.e., $(5 \times 10^{-12}) / (1.332 \times 10^{-12}) = 3.754 = t(93, \approx 0.9995)$] (Federighi, 1959), which is very significant indeed!

In contrast to this ability of the infrared data to determine H , the second line of Table E-2, which gives the 95% confidence limits for the microwave rotational data alone, shows that these data cannot determine with 95% confidence a value for H'' that is smaller than $2.11 \times 10^{-10} \text{ cm}^{-1}$. Although the microwave data are extremely precise, they are handicapped here in two ways. First, because of the small degrees of freedom, they have a large t -factor, $t(4, 0.95) = 2.776$. Secondly, and more importantly for the molecular constant H'' , since the data only go to low rotational levels ($J_{\text{max}} = 7$ here), the dispersion element is large, $V_{33}^{\frac{1}{2}} = 7.61 \times 10^{-5}$. However, note that the microwave data can determine an estimate of B'' that is 28 times more precise than the estimate determined from the infrared data. For this molecular constant, the 10^3 factor of increased precision between the microwave and infrared measurements overcomes the 1.396 t -factor handicap and a 25.75 dispersion-element handicap.

Thus, the first two lines of Table E-2 show that the infrared data are much better at determining \hat{H}'' , the microwave data are

much better at determining \hat{B}'' , and the two sets of data can determine \hat{D}'' about equally well. Such a situation clearly suggests that gains in the precision of the molecular constants are available by combining the two sets of data into one weighted least-squares fit. The third line of Table E-2 shows that this is indeed true.

Comparison of the three lines of any of the last three columns shows that the precision of the estimated molecular constants \hat{B}'' , \hat{D}'' , and \hat{H}'' from the fit to the combined infrared and microwave data is in each case better than that from a fit to either the infrared or the microwave data alone. Furthermore, note that this improvement propagates into the upper state constants B' , D' , and H' , via the correlation between the upper and lower state constants arising from the infrared vibration-rotation lines (see Table D-1). Clearly the possibility for such a straightforwardly obtained improvement in the precision of estimated molecular constants should not be ignored! Only the band origin ν_0 , which is not involved with the microwave rotational data and which is only very weakly correlated with the rotational constants (see Table D-1), escapes significant improvement, as expected.

F. CORRELATED LEAST-SQUARES FORMULATION

We now examine the most general least-squares formulation, namely, the one that is appropriate when the data to be fitted are known to have both unequal weights and nonzero correlations. The problem that is used here as an example arises frequently in spectroscopy, but the present application of correlated least-squares toward its solution is relatively new (Albritton, Schmeltekopf, Zare, and Czarny, 1975; Saenger, Zare, and Mathews, 1975).

1. Statement of the Problem

The problem of combining infrared vibration-rotation and microwave rotation data to obtain an optimum set of molecular constants is considered again here. However, the present approach differs from the single weighted least-squares fit to both the infrared and microwave data simultaneously that was used in Sec. E, where only an educated "guess" was used for the ratio of the variances of the infrared and microwave line position measurements. Here, unweighted least-squares fits are first made separately to the infrared and to the microwave measurements to obtain separate infrared and microwave MVLU estimates for the molecular

constants, their variance-covariance matrix, and the variance of the measurement errors. Not only do these separate fits yield MVLU estimates of the infrared and microwave measurement-error variances to replace the "guesses," but they also provide the important opportunity to test for relative systematic errors between the infrared and microwave measurements, as described below. In the event such tests prove negative, the two sets of estimates are then merged in an optimum way. Namely, a correlated least-squares fit is made to both the infrared and the microwave estimates simultaneously to obtain an optimum single set of MVLU estimates for the molecular constants. Therefore, the problem is solved in two steps instead of one, and the second step serves here as an example of the application of the correlated least-squares formulation.

Specifically, the p infrared line position measurements $\underline{y}_{\text{IR}}$ are assumed to have a variance-covariance matrix $E(\underline{\epsilon}_{\text{IR}}\underline{\epsilon}_{\text{IR}}^T) = \hat{\sigma}_{\text{IR}}^2 \underline{I}_{\text{IR}}$. These measurements are fitted to the model $\underline{y}_{\text{IR}} = \underline{X}_{\text{IR}}\underline{\theta}_{\text{IR}} + \underline{\epsilon}_{\text{IR}}$ to obtain the estimates of the m_{IR} molecular constants $\hat{\underline{\theta}}_{\text{IR}}^T = [\hat{\nu}_{0\text{IR}} \hat{B}'_{\text{IR}} \hat{D}'_{\text{IR}} \hat{H}'_{\text{IR}} \hat{B}''_{\text{IR}} \hat{D}''_{\text{IR}} \hat{H}''_{\text{IR}}]$, the measurement error variance $\hat{\sigma}_{\text{IR}}^2$, and the $m_{\text{IR}} \times m_{\text{IR}}$ variance-covariance matrix of the molecular constants $\hat{\underline{\Theta}}_{\text{IR}} = \hat{\sigma}_{\text{IR}}^2 \underline{V}_{\text{IR}}$, where the dispersion matrix is given by $\underline{V}_{\text{IR}} = (\underline{X}_{\text{IR}}^T \underline{X}_{\text{IR}})^{-1}$. A specific example of the $p \times m_{\text{IR}}$ coefficient matrix $\underline{X}_{\text{IR}}$ used in this unweighted least-squares fit is given in

the first 100 rows of the matrix in Eq. (E-4). Similarly, the q microwave line position measurements \underline{y}_M are assumed to have a variance-covariance matrix $E(\underline{\epsilon}_M \underline{\epsilon}_M^T) = \underline{\Phi}_M = \sigma_M^2 \underline{I}_M$. These measurements are fitted to the model $\underline{y}_M = \underline{X}_M \underline{\beta}_M + \underline{\epsilon}_M$ to obtain the estimates of the m_M molecular constants $\hat{\underline{\beta}}_M = [\hat{B}_M'' \ \hat{D}_M'' \ \hat{H}_M'']$, $\hat{\sigma}_M^2$, and $\hat{\underline{\Theta}}_M = \hat{\sigma}_M^2 \underline{V}_M$, where $\underline{V}_M = (\underline{X}_M^T \underline{X}_M)^{-1}$. A specific example of the $q \times m_M$ matrix \underline{X}_M used in this unweighted least-squares fit is given in the last 7 rows and the last 3 columns of the matrix given in Eq. (E-4).

In general, the infrared estimates and the microwave estimates for the lower-state constants B'' , D'' , and H'' will differ. Before proceeding to the second step where optimum single values are determined for these constants, it is always very valuable to test whether such differences are consistent with the confidence limits of the infrared and microwave estimates. If the differences are much larger than could be expected on the basis of the confidence limits, then one suspects that non-negligible systematic errors may exist in either or both of the sets of measurements. If such systematic errors are ignored and the infrared and microwave results are combined, then the statistical meaning of the resulting combined values can be seriously impaired and misleading.

The simplest test for internal consistency is to see whether the, say, 95% confidence limits of the infrared results and 95% confidence limits of the microwave results overlap for most of

the common constants. If so, one is encouraged, since this suggests that gross relative systematic errors are probably absent. A more sophisticated test is to compute the confidence limits for the difference of each pair of corresponding estimates, e.g. $\hat{B}_{IR}'' - \hat{B}_M''$, and see if they include zero. The details of the computation of the confidence limits of the differences, which again involve Student's t-factor, are given by Bennett and Franklin (1954, pp. 176-180), Hamilton (1964, pp. 92-94), and Dixon and Massey (1969, pp. 114-119). If, say, the 95% confidence limits do not all include zero, then one becomes concerned that there are significant relative systematic errors, suggesting a reexamination of the experimental procedures or the models used in fitting the data. On the other hand, if most of the important confidence limits do include zero, then one is justified in combining the results, as is assumed to be the case for the example here.

The output quantities $\hat{\underline{\theta}}_{IR}$, $\hat{\underline{\theta}}_{IR}$, $\hat{\underline{\theta}}_M$, and $\hat{\underline{\theta}}_M$ from the first step are taken as the input quantities of the second step. Namely, a vector of $n = m_{IR} + m_M$ "observations"

$$\underline{y} = \begin{bmatrix} \hat{\underline{\theta}}_{IR} \\ \hat{\underline{\theta}}_M \end{bmatrix}, \quad (F-1)$$

is formed from the subvectors $\hat{\underline{\beta}}_{IR}$ and $\hat{\underline{\beta}}_M$. The $n \times n$ variance-covariance matrix $E(\underline{\epsilon}\underline{\epsilon}^T) = \underline{\Phi}$ associated with these observations is

$$\underline{\Phi} = \sigma^2 \underline{M} = \sigma^2 \begin{bmatrix} \hat{\sigma}_{IR}^2 V_{IR} & \underline{0} \\ \underline{0} & \hat{\sigma}_M^2 V_M \end{bmatrix}, \quad (F-2)$$

whose elements are formed from the $m_{IR} \times m_{IR}$ and $m_M \times m_M$ submatrices $\hat{\sigma}_{IR}^2 V_{IR} = \hat{\underline{\Theta}}_{IR}$ and $\hat{\sigma}_M^2 V_M = \hat{\underline{\Theta}}_M$, respectively.

The "observations" \underline{y} are fitted to the linear model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon}, \quad (F-3)$$

where the m molecular constants to be estimated are

$$\underline{\beta}^T = [\nu_0 \ B' \ D' \ H' \ B'' \ D'' \ H''] \quad (F-4)$$

The $n \times m$ coefficient matrix is given by

$$\tilde{X} = \begin{matrix} & \nu_0 & B' & D' & H' & B'' & D'' & H'' \\ \begin{matrix} \nu_0 \\ B'_{IR} \\ D'_{IR} \\ H'_{IR} \\ B''_{IR} \\ D''_{IR} \\ H''_{IR} \\ B''_M \\ D''_M \\ H''_M \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix} \quad (F-5)$$

Thus, the "model" assumed here is a particularly simple one, namely, that there is one value for each molecular constant.

This is a relatively rare example of a perfect model. Note that the model allows for changes (i.e., improvements) in the molecular constants ν_0 , B' , D' , and H' , even though the microwave measurements do not involve these constants, because of the correlation of these constants with the lower-state constants.

Since Eq. (F-2) shows that the errors of the "observations" \underline{y} have not only unequal variances but also some nonzero covariances, even the weighted least-squares formulation is inappropriate. This

application requires the correlated least-squares formulation (i.e., weighted and correlated least-squares), which is the most general formulation of the least-squares method for errors only in the dependent variable.

2. Least-Squares Estimates

In the Appendix, it is shown that for n observations \underline{y} whose errors $\underline{\epsilon}$ have a variance-covariance matrix $\underline{\Phi} = \sigma^2 \underline{M}$ with finite elements and for \underline{y} linearly related to m parameters $\underline{\beta}$ through the $n \times m$ coefficient matrix \underline{X} , $\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon}$, the MVLU estimates of the parameters $\hat{\underline{\beta}}$, the variance $\hat{\sigma}^2$, and the variance-covariance matrix of the constants $\hat{\underline{\Theta}}$ are given by

$$\hat{\underline{\beta}} = (\underline{X}^T \underline{M}^{-1} \underline{X})^{-1} \underline{X}^T \underline{M}^{-1} \underline{y}, \quad (\text{F-6})$$

$$\hat{\sigma}^2 = \frac{1}{n-m} (\underline{y} - \underline{X}\hat{\underline{\beta}})^T \underline{M}^{-1} (\underline{y} - \underline{X}\hat{\underline{\beta}}), \quad (\text{F-7})$$

$$\hat{\underline{\Theta}} = \hat{\sigma}^2 \underline{V} = \hat{\sigma}^2 (\underline{X}^T \underline{M}^{-1} \underline{X})^{-1} \quad (\text{F-8})$$

The unweighted and the weighted formulations of Secs. D and E are clearly just special cases of this general formulation with $\underline{M} = \underline{I}$ and \underline{M} equal to a diagonal matrix with unequal elements, respectively.

For the present example, the degrees of freedom for the correlated fit are small, $n-m=3$, which means that the variance

of the estimate $\hat{\sigma}^2$ given by Eq. (F-7) is relatively large. Hence, for the present example of sequential least-squares fits, it is best to compute $\hat{\sigma}^2$ from the alternative expression

$$\hat{\sigma}^2 = \frac{1}{p+q-m} \left\{ \begin{bmatrix} \underline{Y}_{IR} \\ \underline{Y}_M \end{bmatrix} - \begin{bmatrix} \underline{X}_{IR} & \dots & \dots & \dots \\ \underline{0} & \vdots & \vdots & \underline{X}_M \end{bmatrix} \hat{\underline{\beta}} \right\}^T \begin{bmatrix} \underline{I}_{IR}/\hat{\sigma}_{IR}^2 & \underline{0} \\ \underline{0} & \underline{I}_M/\hat{\sigma}_M^2 \end{bmatrix} \quad (F-9)$$

$$\left\{ \begin{bmatrix} \underline{Y}_{IR} \\ \underline{Y}_M \end{bmatrix} - \begin{bmatrix} \underline{X}_{IR} & \dots & \dots & \dots \\ \underline{0} & \vdots & \vdots & \underline{X}_M \end{bmatrix} \hat{\underline{\beta}} \right\} \cdot$$

Equation (F-9) uses the $p+q$ residuals constructed from the optimum estimates $\hat{\underline{\beta}}$ and the measured line positions \underline{Y}_{IR} and \underline{Y}_M to estimate $\hat{\sigma}^2$. In contrast, Eq. (F-7) uses the 10 residuals constructed from the optimum estimates $\hat{\underline{\beta}}$ and the "observations" $\hat{\underline{\beta}}_{IR}$ and $\hat{\underline{\beta}}_M$. Since $p+q \gg 10$ in general, the former turns out to be a preferable estimate for this example.

By analogy with Eq. (E-9) of the preceding Section, it is useful to define a "generalized" (i.e., nondiagonal) weight matrix, $\underline{W} = \underline{M}^{-1}$. For the present example,

$$\begin{aligned}
 \underline{W} \equiv \underline{M}^{-1} &= \begin{bmatrix} \underline{V}_{\text{IR}}^{-1} / \hat{\sigma}_{\text{IR}}^2 & \underline{0} \\ \underline{0} & \underline{V}_{\text{M}}^{-1} / \hat{\sigma}_{\text{M}}^2 \end{bmatrix} , \\
 &= \begin{bmatrix} \underline{X}_{\text{IR}}^T \underline{X}_{\text{IR}} / \hat{\sigma}_{\text{IR}}^2 & \underline{0} \\ \underline{0} & \underline{X}_{\text{M}}^T \underline{X}_{\text{M}} / \hat{\sigma}_{\text{M}}^2 \end{bmatrix} .
 \end{aligned} \tag{F-10}$$

The latter, more-explicit form of \underline{W} shows how the correlated least-squares fit utilizes simultaneously, as generalized "weights" (i.e., as reciprocals), both the differing precision estimates $\hat{\sigma}_{\text{IR}}^2$ and $\hat{\sigma}_{\text{M}}^2$ and the differing dispersions $(\underline{X}_{\text{IR}}^T \underline{X}_{\text{IR}})^{-1}$ and $(\underline{X}_{\text{M}}^T \underline{X}_{\text{M}})^{-1}$ of the separate infrared and microwave data in determining the optimum estimates of the molecular constants $\hat{\underline{\theta}}$ and their variances and covariances $\hat{\underline{\Theta}}$.

Albritton, Schmeltekopf, Zare, and Czarny (1975) have shown that these values for $\hat{\underline{\theta}}$ and $\hat{\underline{\Theta}}$ are identical with those that would be obtained from a single weighted least-squares fit to both the measured infrared and microwave lines $\underline{y}_{\text{IR}}$ and \underline{y}_{M} simultaneously using $\hat{\sigma}_{\text{IR}}^2$ and $\hat{\sigma}_{\text{M}}^2$ in the weight matrix. In other words, the present correlated least-squares fit gives results identical to the

weighted least-squares fit of Sec. E, provided the "guesses" $\tilde{\sigma}_{\text{IR}}^2$ and $\tilde{\sigma}_{\text{M}}^2$ are replaced by the MVLU estimates $\hat{\sigma}_{\text{IR}}^2$ and $\hat{\sigma}_{\text{M}}^2$ obtained from the individual fits to each type of data separately. Thus, all of the comments made in Sec. E-5 about the gains available by combining infrared and microwave data apply equally well here to the identical results of the present correlated fit. The key point here is that, in both of these equivalent fits, initial least-squares fits were made to separate subdivisions of the whole data set. In the present case, there are two subdivisions, the infrared data and the microwave data. These gave the individual subsets of estimates $(\hat{\beta}_{\text{IR}}, \hat{\sigma}_{\text{IR}}^2, \hat{\theta}_{\text{IR}})$ and $(\hat{\beta}_{\text{M}}, \hat{\sigma}_{\text{M}}^2, \hat{\theta}_{\text{M}})$. The weighted fit then uses $(\underline{y}_{\text{IR}}, \hat{\sigma}_{\text{IR}}^2, \underline{y}_{\text{M}}, \hat{\sigma}_{\text{M}}^2)$ as input; the correlated fit uses $(\hat{\beta}_{\text{IR}}, \hat{\theta}_{\text{IR}}, \hat{\beta}_{\text{M}}, \hat{\theta}_{\text{M}})$ as input. Both give identical final results. However, it is important to realize that an unweighted fit to $(\underline{y}_{\text{IR}}, \underline{y}_{\text{M}})$ simultaneously will not generally give the same results (these results will not be MV), since it would not utilize the information that $\hat{\sigma}_{\text{IR}}^2 \neq \hat{\sigma}_{\text{M}}^2$ gained from the key "interrogations" of the infrared and microwave subsets separately.

Clearly these techniques are not limited to just combining infrared and microwave data or even to combining only two subsets of data. For example, both the weighted and the correlated fits are very useful for reducing a multi-band system to an optimum set of molecular constants. The choice between the two is dictated

only by convenience. For applications to nonlinear models, the two-step correlated approach appears to be more practical than the two-step, weighted approach. In the former, the second step is a simple, linear fit and in the latter, the second step is an often-large, nonlinear fit, which can be very expensive.

In the present example, the correlation among the quantities to be fitted arose because they were output from a preceding least-squares fit. Correlation among quantities to be fitted can also arise because they are the output of a preceding linear transformation. One of the simplest types of transformations occurring in molecular spectroscopy is that of forming new quantities that are differences between other quantities. One example is the construction of $\Delta G_{v+\frac{1}{2}}$ values from band origins ν_0 . As noted by Lees (1973), another example is the calculation of the splittings between the measured frequencies of multiplet components. In these cases, errors in the $\Delta G_{v+\frac{1}{2}}$ values that involve the same ν_0 and the errors in the splittings that involve the same multiplet component will be correlated with correlation coefficients of $-\frac{1}{2}$, as can be straightforwardly shown using the propagation of variance and covariance (see Sec. G-2-a). Therefore, such $\Delta G_{v+\frac{1}{2}}$ values and splittings must be fitted with the correlated least-squares formulation in order to obtain MV estimates.

G. REPORTING LEAST-SQUARES RESULTS

The preceding Sections have dealt with the assumptions of the least-squares method, the properties of its estimates, and the mechanics of its application. The subject of this concluding Section is the next logical topic; namely, once the least-squares estimates have been obtained, what should be reported and how should it be reported in order to convey to the user a reasonably complete set of information. Here, we divide the least-squares input and results into two categories, first those that appear to be essential to report and secondly, those that are useful but sometimes impractical to report.

1. Essential Information

(a) Basic observations. The data to be fitted, y , clearly occupy a fundamental position in the whole endeavor of this Chapter. Therefore, their presentation is essential and may, in time, prove to be the most valuable part of a publication. We are all aware of examples where high-quality line positions that are decades old have "outlived" the accompanying estimates of the molecular constants, in that improved models were later devised that fit these data much better than those available initially.

To be sure, the large number of line positions frequently recorded in molecular spectroscopy creates a problem in their presentation in journals, but photoreduction of data tables, as well as permanent data depositories, contribute to its solution.

It is traditional to list all of the line positions that were measured in the investigation. For any of several reasons, some fraction (usually small) of the measured lines are often not satisfactorily represented by the model and hence were not included in the least-squares determination of the model parameters. For two reasons, it is useful to have these rejected but listed data identified. First, the particular measurements on which the estimates of the molecular constants were based are then clearly identified by "default." Secondly, it is possible that many of the rejected measurements may later prove to be a valuable source of new information. The classic spectroscopic example is that of initially-unfitted perturbed lines later yielding to analysis and providing the identity and location of new electronic states.

(b) Weight matrix. From the preceding Sections, it is clear that the variance-covariance matrix of the measurement errors $\underline{\Phi} = \sigma^2 \underline{M}$ must be considered as input that has equal standing with the observations \underline{y} themselves. Therefore, to completely define the input to a least-squares fit, the matrix \underline{M} must be adequately stated. For unweighted fits, this requirement is met with only

a few words. For weighted fits, the elements of the diagonal weight matrix $\underline{W} = \underline{M}^{-1}$ can be given. Only for correlated fits must a full generalized weight matrix be given or described.

(c) Model. Since the model basically defines the nature of the physical constants being estimated, it is clearly essential information to be reported. Without it, the estimates lose their usefulness as predictors. Some models are simple to specify, e.g., the vibrating-rotator used as examples here. On the other hand, the complex models of multiplet states require a more elaborate presentation, or careful reference, of Hamiltonian matrix elements. Part of model specification may also include the results of tests of significance of "borderline" constants like H_v that were not included in the model.

(d) Estimated molecular constants. Since it is needless to elaborate on why the estimates of the molecular constants $\hat{\beta}$ should be reported, we consider here the important point of how to report them without loss of information. These values appear first on the computer output, where they are printed with the number of digits determined by the extent of the chosen format statement, which is deliberately oversized. Clearly, not all of these digits are significant, and rounding up to the last significant digit before reporting these estimates is certainly called for. This apparently-trivial task is actually quite deceptive.

Since the standard error is a measure of the uncertainty of an estimate, the practice of rounding the estimate so that the last remaining digit reflects one standard error is a common one. For example, $\hat{B}' = 0.3224994\dots$ and $\hat{B}'' = 0.3320580\dots \text{ cm}^{-1}$, with standard errors of $0.0000075\dots$ and $0.0000076\dots \text{ cm}^{-1}$, respectively, could be presented as $B' = 0.322499(8)$ and $B'' = 0.332058(8) \text{ cm}^{-1}$ (where the number in parentheses reflects the standard error in the last digit). However, in this particular fit (Albritton, Schmeltekopf, Zare, and Tellinghuisen, 1974), $\Delta B \equiv B' - B'' = -0.0095586\dots \text{ cm}^{-1}$, with a standard error of $0.00000039\dots \text{ cm}^{-1}$. (The difference between two highly-correlated constants is almost always more precise than the two constants themselves. See Pliva and Telfair, 1974, and Albritton, Schmeltekopf, Tellinghuisen, and Zare, 1974.) Therefore, these two overly-rounded values cannot yield the ΔB value with full precision; one more digit is required, e.g., $B' = 0.3224994(76) \text{ cm}^{-1}$. Because of this, such overly-rounded values may yield calculated line positions that are significantly in error, thereby impairing, for example, the predictive usefulness of these reported values.

The possibility that rounding to reflect one standard error does indeed result in over-rounding increases with the correlation between the molecular constants. The reason for this is that, while the standard error is indeed a measure of the absolute precision of an estimated constant, the covariances describe the relative precision among the set of estimates. Thus, for large correlation

coefficients, using the standard errors alone as a guide for rounding can result in considerable loss of relative precision, which is important anytime the estimates are used as a set. An example of detrimental loss of relative precision by over-rounding follows.

Figure G-1 shows a computer output giving the results of an equally-weighted least-squares fit, $\underline{W} = \underline{I}/(0.30)^2 \text{ cm}^2$, to a set of 19 "measured" absorption band origins $\nu_0(v', 0)$ to determine estimates of the electronic term energy T_{0e} and the five vibrational Dunham coefficients Y_{i0} , $i = 1, 2, \dots, 5$, (commonly recognized as ω_e , $-\omega_e x_e$, $\omega_e y_e$, ...) based on the familiar model

$$\nu_0(v', 0) = T_{0e} + \sum_{i=1}^5 Y_{i0} (v' + \frac{1}{2})^i \quad (G-1)$$

The output lists the measured values \underline{y} and the calculated values, $\hat{\underline{y}} = \underline{X}\hat{\underline{\theta}}$, and compares the differences to the estimated experimental uncertainty. The estimated values $\hat{\underline{\theta}}$ and standard errors $\hat{\theta}_{ii}^{1/2}$ for the Dunham coefficients are given, along with the degrees of freedom $n-m$ and standard deviation of the residuals $\hat{\sigma}$. Lastly, the estimated variance-covariance matrix $\hat{\underline{\Theta}}$ and the correlation coefficients C_{ij} are displayed.

One immediately notes that the errors of the estimated coefficients are highly correlated. These magnitudes are a clear warning that the standard errors alone, which are here often only a factor of ten smaller than the estimated coefficient itself, should not be the sole guide for the rounding of digits.

CALCULATION OF DUNHAM EXPANSION COEFFICIENTS FROM BAND ORIGIN VALUES

V	MEASURED NU(V,0)	CALCULATED NU(V,0) AND STD ERR	MEAS - CALC RESIDUAL	EXPERIMENTAL UNCERTAINTY	DUNHAM EXPANSION COEFFICIENTS AND STD ERR
0	64748.48	64748.80(0.30)	-0.32	0.30	* TOE = 63996.0496(0.4847)
1	66230.15	66229.37(0.19)	0.78	0.30	* Y(1,0) = 1513.69728(0.5187)
2	67675.66	67676.08(0.19)	-0.42	0.30	* Y(2,0) = -16.2985963(0.1683)
3	69088.40	69088.44(0.17)	-0.04	0.30	* Y(3,0) = -1.84997169(0.2240) X 10 ⁻¹
4	70466.27	70466.18(0.15)	0.09	0.30	* Y(4,0) = 1.50107240(0.1298) X 10 ⁻²
5	71809.04	71809.24(0.15)	-0.20	0.30	* Y(5,0) = -4.10143806(0.2720) X 10 ⁻⁴
6	73117.55	73117.70(0.16)	-0.15	0.30	* DEGREES OF FREEDOM = 13
7	74391.73	74391.70(0.15)	0.03	0.30	* STANDARD DEVIATION = 1.08
8	75631.75	75631.45(0.15)	0.30	0.30	
9	76837.33	76837.14(0.14)	0.19	0.30	
10	78008.66	78008.90(0.15)	-0.24	0.30	
11	79146.91	79146.77(0.15)	0.14	0.30	
12	80250.51	80250.59(0.16)	-0.08	0.30	
13	81320.05	81320.05(0.15)	0.00	0.30	
14	82354.43	82354.55(0.15)	-0.12	0.30	
15	83352.97	83353.19(0.17)	-0.22	0.30	
16	84314.94	84314.73(0.19)	0.21	0.30	
17	85237.75	85237.51(0.19)	0.24	0.30	
18	86119.25	86119.42(0.30)	-0.17	0.30	
19		86957.87(0.76)			
20		87749.70(1.66)			

VARIANCE-COVARIANCE MATRIX AND CORRELATION COEFFICIENTS

	TOE	Y(1,0)	Y(2,0)	Y(3,0)	Y(4,0)	Y(5,0)
TOE	-1	-1	-2	-3	-4	-6
Y(1,0)	2.3493183 X 10 ⁻¹	-2.2281648 X 10 ⁻¹	6.3682009 X 10 ⁻²	-7.6439485 X 10 ⁻²	4.0684409 X 10 ⁻²	-7.9285833 X 10 ⁻²
Y(2,0)	-0.8863	2.6902173 X 10 ⁻¹	-8.4865649 X 10 ⁻²	1.0751016 X 10 ⁻¹	-5.9199829 X 10 ⁻²	1.1809675 X 10 ⁻¹
Y(3,0)	0.7806	-0.9721	2.8329825 X 10 ⁻²	-3.7216110 X 10 ⁻²	2.1016128 X 10 ⁻²	-4.2707581 X 10 ⁻²
Y(4,0)	-0.7042	0.9255	-0.9873	5.0156033 X 10 ⁻²	-2.8863167 X 10 ⁻²	5.9508795 X 10 ⁻²
Y(5,0)	0.6466	0.8792	0.9618	-0.9927	1.6853350 X 10 ⁻²	-3.5150273 X 10 ⁻²
Y(6,0)	-0.6013	0.8370	-0.9328	0.9768	-0.9953	7.4000574 X 10 ⁻²

-1
(ALL UNITS ARE CM⁻¹)

Figure G-1. Sample computer output of a least-squares fit to a set of absorption band origins to estimate the upper-state Dunham coefficients.

Specifically, the high correlation implies that if one reports these calculated band origins \hat{y} (which were generated by the computer using all of its digits, significant and insignificant) and also lists the coefficients rounded to reflect one standard error, then these reported calculated band origins cannot be reproduced by the reported (over-rounded) expansion coefficients! Table G-1 gives an example of this. The second column lists, first, the expansion coefficients rounded so that the last remaining digit reflects one standard error, and secondly, the differences that occur between the values calculated from these rounded values and those calculated from the unrounded values. The error introduced by over-rounding increases with increasing vibrational level and, at the upper extent of the data, is about four times larger than the experimental uncertainty. The third column shows that even when one more digit is preserved, the error is still intolerable, since it is simply due to excessive roundoff of digits that had relative significance. Finally, the last column shows that approximately three digits beyond the "one standard error digit" are required in this example to eliminate this annoying and unnecessary error.


The digits that possess relative significance vary, of course, with the degree of correlation, and hence are different for each example. Although "rules of thumb" can be developed for each

class of least-squares fit, (e.g., one additional digit beyond the "one standard error digit" seems to suffice for most electronic bands) it is nevertheless an excellent practice, and an easy task, to take one's final rounded set of values for the estimates $\hat{\beta}$ and simply verify that they do indeed give the original calculated \hat{y} values to some number of appropriate digits.

(e) Estimated variance of the measurement errors. The estimated variance $\hat{\sigma}^2$, or equivalently, the estimated standard deviation $\hat{\sigma}$, is the best single measure of the precision of the data and the success of the chosen model in fitting the data. As such, $\hat{\sigma}$ should definitely be reported.

Since $\hat{\sigma}^2$ is an estimate, it has some degree of uncertainty. While the "variance of the estimated variance" sounds like an elaboration of statistical minutia, it is nevertheless useful and worth a few lines here. The reason is that, in reporting $\hat{\sigma}$, it is convenient to know how many digits are significant, a small but practical matter. Unfortunately, as explained by Dixon and Massey (1969, p. 101-103), there is no simple, universal expression for the variance of $\hat{\sigma}^2$. For example, the scatter of the $\hat{\sigma}$ estimates in Fig. D-2 is slightly asymmetric; hence, there is no single " \pm " value that is strictly correct. However, an approximation

Table G-1. Differences between $v_0(v', 0)$ values calculated with unrounded and rounded Y_{i0} coefficients.

coefficient	rounded to one std. err.	rounded to one additional digit	"appropriately" rounded
T_{e0}	63996.05	63996.05	63996.050
Y_{10}	1513.7	1513.70	1513.6973
Y_{20}	-16.30	-16.299	-16.2986
Y_{30}	-1.85×10^{-1}	-1.850×10^{-1}	-1.84997×10^{-1}
Y_{40}	1.50×10^{-2}	1.501×10^{-2}	1.50107×10^{-2}
Y_{50}	-4.10×10^{-4}	-4.101×10^{-4}	-4.10144×10^{-4}
<u>v</u>	<u>difference</u>	<u>difference</u>	<u>difference</u>
0	0.00	0.00	0.00
1	0.00	0.00	
2	0.00	0.00	
3	0.01	0.00	
4	0.02	0.00	
5	0.04	0.00	
6	0.06	0.00	
7	0.09	0.00	
8	0.13	0.01	
9	0.18	0.02	
10	0.24	0.02	
11	0.32	0.03	
12	0.41	0.04	
13	0.52	0.05	
14	0.65	0.06	
15	0.80	0.07	
16	0.97	0.08	
17	1.17	0.09	
18	1.39	0.10	0.00

that is reasonable for moderately large degrees of freedom and approximately normally distributed measurement errors (Rossini and Deming, 1939) is that the standard deviation of the standard deviation is

$$\hat{\sigma}_{\hat{\sigma}} \cong \frac{\hat{\sigma}}{[2(n-m)]^{\frac{1}{2}}} \quad (G-2)$$

where $n-m$ are the degrees of freedom. This shows that seldom do more than two figures of $\hat{\sigma}$ have absolute precision. Thus, except for large degrees of freedom, $\hat{\sigma}$ need only be quoted to two nontrivial digits.

(f) Degrees of freedom and standard errors. To report only the estimated values $\hat{\beta}_i$ is incomplete, since no indication is given of how "good" or how "bad" such estimates are; that is, no information is given as to the width of the interval centered on the estimate and within which we may reasonably expect the unknown "true" value to lie. Without such information about the precision, the results of two experiments, or the results of experiment and theory, cannot be meaningfully compared. As described in Sec. D-3-b, both the standard errors and the degrees of freedom are required before such confidence limits can be computed. Thus, these two are discussed here together.

It was seen in Fig. D-11, that when the degrees of freedom of

a least-squares fit are large, the t-factor (i.e., the number of standard errors one must combine to be confident, at some level, that an interval has been constructed that includes the "true" value) is very nearly constant. In many spectroscopic investigations, even when the degrees of freedom are not explicitly stated, it will be obvious that they are, in fact, large. Therefore, the popular 95% confidence limits, for example, can be constructed by simply doubling the reported standard error. On the other hand, however, there are numerous investigations where the degrees of freedom cannot be easily estimated by inspection (e.g., perhaps the line lists were deposited). In instances like these, even with the standard errors given, the full measure of the precision of the estimated molecular constants cannot be obtained, or, perhaps at best, can be obtained only after tedious counting. Therefore, since they are available as output from the least-squares fit and are only simple integers that need only to be listed once for each fit, the reporting of the degrees of freedom seems to be very desirable.

Some of these remarks apply equally well to the standard errors themselves, but since they are in fact usually reported, in sharp contrast to the degrees of freedom, no elaboration on why they should be stated is required. However, it is useful to consider how they can be reported in order to convey their

information without possible confusion.

One can question what measure of precision should be reported: confidence limits or standard errors? As described in Sec. D-3-b, it is only the confidence limit that has the desired probabilistic interpretation and that corresponds to one's need for realistic limits of uncertainty due to random errors. Thus, two standard errors are often quoted, implying 95% confidence limits (and large degrees of freedom). On the other hand, one standard error is also often quoted. It is clear that such a quoted value should not be misinterpreted as a realistic confidence limit. If erroneously taken to be such, the risk of being wrong, due purely to random effects, is a rash 33 percent. The standard error is, however, in many ways a simpler and more appealing quantity to quote than a confidence limit. It is easily recognized and, to many, it lies at the heart of the meaning of precision. Furthermore, it permits the users to construct confidence limits with percentages corresponding to their own tastes, e.g., 95, 99, 99.9%, provided, of course, that the degrees of freedom are given. Actually, it does not appear to be critical whether either confidence limits or standard errors are quoted; however, whichever is quoted, it is critical that it be clearly identified as such!

One also sometimes sees, for example, five standard errors given as the "limits of error" (e.g., see Ramsay, 1974, p. 123). Usually, such

large multiples are given in order to account for some degree of "contamination" by systematic errors. While it is clear that this procedure can have no statistical basis and the use of the standard error as a "unit" for the multiple appears to be chiefly for convenience, it can be nevertheless a useful statement of the spectroscopist's judgement of the relative sizes of random and systematic error. For example, it could have been found that estimates of the same constant using several different bands required five standard errors to be consistent, and, to be on the safe side, such "limits of error" were adopted for all of the estimates. Again, possible misinterpretation can be avoided if the construction of the limits are clearly explained.

Since the precision of a standard error depends only on the precision of $\hat{\sigma}$, the general rule for the number of significant digits in a standard error is the same as that described above for $\hat{\sigma}$. Specifically, seldom do more than two nontrivial digits need to be reported in quoted standard errors, except for very large degrees of freedom. Since standard errors are rarely used as a set (in contrast, as explained below, to the variances and covariances), there appears to be no problem with over-rounding of digits with relative precision.

2. Useful Information

(a) Estimated Variance-Covariance Matrix. If it were not for the fact that its size often makes it impractical to publish, the usefulness of the estimated variance-covariance matrix $\hat{\Theta} = \hat{\sigma}^2 \underline{V}$ would place it in the category of essential information to report. While the reporting of the standard errors does mean that the diagonal elements of $\hat{\Theta}$ are known, the off-diagonal covariances are required for the determination of the errors of quantities whose calculation requires two or more of the estimated constants $\hat{\beta}$. It is informative to examine a few examples of commonly-occurring instances where the full variance-covariance matrix $\hat{\Theta}$ is useful.

First, for enlarging the data set \underline{y} by interpolation and short extrapolation to the location of missing or new elements $\underline{\psi}$, it is useful to know the standard errors of such predictions, since confidence limits based on these errors estimate the range over which it is practical to search for the new elements. Such estimates of the new elements are calculated from the estimated molecular constants

$$\hat{\underline{\psi}} = \underline{Z} \hat{\underline{\beta}} \quad , \quad (G-3)$$

where the matrix \underline{Z} is generally a simple enlargement of the original coefficient matrix \underline{X} . The standard errors associated with the estimates $\hat{\underline{\psi}}$ are the square roots of the diagonal elements of the estimated variance-covariance matrix of $\hat{\underline{\psi}}$,

$$\hat{\underline{\Omega}} = \underline{Z} \hat{\underline{\Theta}} \underline{Z}^T, \quad (\text{G-4})$$

whose calculation is based on the propagation rules of variance and covariance (the generalization of the more-familiar "propagation of error, i.e. variance," see Hamilton, 1964, p. 149). For example, Fig. G-1 shows the standard errors of the calculated $\nu_0(\nu', 0)$ values. Here the standard error "envelope" is symmetric about the central vibrational level of the observations because all of the weights were equal. With 95% confidence, the location of the origin of the predicted (19,0) band lies within the interval $86956.23 \leq \nu_0(19, 0) \leq 86959.51 \text{ cm}^{-1}$, provided the model of only five Dunham coefficients can be trusted for one additional vibrational level.

A further related example of the usefulness of $\hat{\underline{\Theta}}$ is the calculation of the zero-point vibrational energy

$$\begin{aligned} \hat{G}(\nu=0) &= \sum_{\nu=1}^5 \hat{Y}_{i_0} \left(\frac{1}{2}\right)^i \\ &= \underline{f}^T \hat{\underline{\beta}}, \end{aligned} \quad (\text{G-5})$$

where $\underline{f}^T = [\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}]$. For the example given in Fig. G-1, $\hat{G}(v=0) = 752.75 \text{ cm}^{-1}$. In comparison of this "experimental" value to ab initio calculations of this unobservable quantity, it is useful to know the standard error of this value, $[\underline{f}^T \hat{\Theta} \underline{f}]^{\frac{1}{2}} = 0.22 \text{ cm}^{-1}$. Further examples of the use of the full variance-covariance matrix have been given by Albritton, Schmeltekopf, Tellinghuisen, and Zare (1974).

Thus, in view of the valuable information content in the variance-covariance matrix $\hat{\Theta}$, it is a great pity that it is often so large. When it is impractical to publish $\hat{\Theta}$, it would be useful, almost essential, to at least briefly mention high correlation, when it does exist among some of the reported constants. Certainly, in those applications where $\hat{\Theta}$ is indeed a publishable size, it would be useful to make it available, perhaps in a form similar to the array in Fig. G-1 or to Table E-1 (without the standard errors, of course, since they are usually given separately in the tabulation of the $\hat{\beta}$ values). The variances and covariances are quantitatively useful and the correlation coefficients are qualitatively useful. Note that without $\hat{\sigma}$, reporting just correlation coefficients is not very quantitatively useful, since correlation coefficients alone cannot be used to reconstruct $\hat{\Theta}$. This and other reconstruction problems are the reasons why the "split" matrix containing both $\hat{\Theta}$ and C_{ij} is preferable to just

C_{ij} alone. It should also be noted that since the matrix multiplication in equations like Eq. (G-4) involves differences between large, almost-equal numbers, over-rounding of the elements of a reported $\hat{\underline{\theta}}$ is a possible pitfall.

Finally, it might be thought that, since $\hat{\underline{\theta}} = \hat{\sigma}^2 (\underline{X}^T \underline{X})^{-1}$, $\hat{\underline{\theta}}$ could always be reconstructed from the given $\hat{\sigma}^2$, structure of the data, and model. However, while this is indeed true, it would also mean that the least-squares fit would essentially have to be repeated, which is not generally attractive. For example, $\hat{\underline{\theta}}$ and $\hat{\sigma}^2$ could be the results of a nonlinear least-squares fit.

(b) Normality. The interpretation of a confidence limit in terms of probability is based on the assumption that the measurement errors are normally distributed (Sec. D-3-b). Fortunately, the reliability of the confidence limits in this regard is not unduly sensitive to moderate departures from normality (Hamilton, 1964, p. 173). Therefore, a brief statement summarizing the results of even the simplest normality test, like that described in Sec. D-3-a, would be useful information.

H. CONCLUDING REMARKS

Equations (F-6), (F-7), and (F-8) summarize the mathematical machinery of the method of least-squares. For a given model and set of input data this machinery will obediently grind out a set of resulting numbers. However, the overall goal of computing is insight, not just numbers. A necessary condition for attaining this goal in the campaigns of molecular constant estimation is that the numbers constitute well-defined estimates of the molecular constants. Clearly this hinges on how adequately the least-squares assumptions are met by the physical circumstances.

If there is fairly close correspondence, as often is the case, then the numbers are MVLU estimates, whose desirable statistical properties are well-defined. In such cases, it is particularly important to consider the estimated value of a molecular constant and its accompanying estimated standard error as inseparable quantities. The former can never be more than an estimate, and without the latter, it can never be more than an estimate of unknown quality. The estimated value and its estimated standard error permit the construction of a numerical range, within which the forever-unknown "true" value is expected, with a certain confidence, to lie. This confidence interval is the end product of the experimentation-estimation process. Its width, for a given

confidence, is the final measure of the investigator's skill.

On the other hand, if there is not fairly close correspondence between the least-squares assumptions and the physical situation, the resulting numbers will be impaired estimates, flawed to a degree that may not be superficially apparent. Perhaps the most common failings are systematic measurement errors and model deficiencies. Fortunately, estimation procedures need not be blind to such defects; they have the ability to test, to some degree, for the existence and extent of these failings through, for example, the reasonability of the estimated variance of the residuals, an analysis of the residuals for possible systematics, the normality of the residuals, the internal consistency of confidence limits, the confidence with which a new molecular constant can be added to the model, and other more sophisticated tests. Thus, even though the correction of these defects may be often a formidable or impossible task, just to be aware of their existence is nevertheless a gain. A devil known is at least better than a devil unknown, even though it cannot be exorcised.

ACKNOWLEDGMENTS

We warmly thank Professor John A. Coxon (Dalhousie University) who carefully proofread a first draft of this work and made many valuable suggestions. We are also indebted to Dr. E. L. Crow (Institute of Telecommunication Sciences), Professor Emeritus F. E. Croxton (Columbia University), and Dr. P. V. Tryon (National Bureau of Standards) for their useful comments on the statistics of this Chapter. R.N.Z. gratefully acknowledges support from the National Science Foundation.

APPENDIX: DERIVATION OF THE LEAST-SQUARES EQUATIONS

1. Estimated Molecular Constants

We start from the linear model equation

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon} , \quad (1)$$

which relates the n observations y_i , each with error ϵ_i , to the m parameters β_i ($m < n$) through a rectangular $n \times m$ matrix \underline{X} of known coefficients. For unweighted least squares the errors are assumed to arise from distributions (not necessarily normal) characterized by a mean value of zero

$$E(\underline{\epsilon}) = \underline{0} \quad (2)$$

and an unknown variance σ^2 , but zero covariance

$$E(\underline{\epsilon}\underline{\epsilon}^T) = \sigma^2 \underline{I} , \quad (3)$$

where \underline{I} is the $n \times n$ identity matrix. We form the sum of the squared residuals

$$\begin{aligned}
 S(\underline{\beta}) &= (\underline{y} - \underline{X}\underline{\beta})^T (\underline{y} - \underline{X}\underline{\beta}) \\
 &= \underline{y}^T \underline{y} - \underline{y}^T \underline{X}\underline{\beta} - \underline{\beta}^T \underline{X}^T \underline{y} + \underline{\beta}^T \underline{X}^T \underline{X}\underline{\beta} \quad , \quad (4)
 \end{aligned}$$

where we use the matrix algebra property $(\underline{AB})^T = \underline{B}^T \underline{A}^T$. We then minimize $S(\underline{\beta})$ with respect to the variation of $\underline{\beta}$.

We introduce the column vector operator

$$\underline{\nabla}_{\underline{\beta}} = \begin{bmatrix} \frac{\partial}{\partial \beta_1} \\ \frac{\partial}{\partial \beta_2} \\ \vdots \\ \frac{\partial}{\partial \beta_m} \end{bmatrix} \quad . \quad (5)$$

Because of its definition as a column vector, $\underline{\nabla}_{\underline{\beta}}$ can only be applied to the transpose of a column vector, i.e. a row vector. For example, if \underline{u} is an arbitrary 1×1 column vector, then $\underline{\nabla}_{\underline{\beta}} \underline{u}^T$ is a $m \times 1$ matrix

$$\underline{\nabla}_{\underline{\beta}} \underline{u}^T = \begin{bmatrix} \frac{\partial}{\partial \beta_1} \\ \frac{\partial}{\partial \beta_2} \\ \vdots \\ \frac{\partial}{\partial \beta_m} \end{bmatrix} \begin{bmatrix} u_1, u_2, \dots, u_\ell \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial u_1}{\partial \beta_1} & \frac{\partial u_2}{\partial \beta_1} & \dots & \frac{\partial u_\ell}{\partial \beta_1} \\ \frac{\partial u_1}{\partial \beta_2} & \frac{\partial u_2}{\partial \beta_2} & \dots & \frac{\partial u_\ell}{\partial \beta_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial u_1}{\partial \beta_m} & \frac{\partial u_2}{\partial \beta_m} & \dots & \frac{\partial u_\ell}{\partial \beta_m} \end{bmatrix} . \quad (6)$$

Unless $\underline{\nabla}_{\underline{\beta}}$ is of dimension 1×1 , the ordinary rules of differentiation, such as the rule for differentiating a product, are not valid. An exception occurs when \underline{a} and \underline{b} are two arbitrary $m \times 1$ column vectors. Then it is readily shown by direct evaluation that a modified product rule holds, namely,

$$\underline{\nabla}_{\underline{\beta}} (\underline{a}^T \underline{b}) = (\underline{\nabla}_{\underline{\beta}} \underline{a})^T \underline{b} + (\underline{\nabla}_{\underline{\beta}} \underline{b})^T \underline{a} . \quad (7)$$

It immediately follows from Eq. (7) that if \underline{c} is a $m \times 1$ column vector independent of β , then

$$\nabla_{\beta} (\underline{\beta}^T \underline{c}) = \nabla_{\beta} (\underline{c}^T \underline{\beta}) = \underline{c} \quad (8)$$

and if \underline{Z} is a $m \times m$ matrix independent of β , then

$$\nabla_{\beta} (\underline{\beta}^T \underline{Z} \underline{\beta}) = \underline{Z} \underline{\beta} + \underline{Z}^T \underline{\beta} \quad (9)$$

The condition for $S(\underline{\beta})$ to be a minimum is

$$\nabla_{\underline{\beta}} s(\underline{\beta})|_{\underline{\beta}=\hat{\underline{\beta}}} = \underline{0} \quad . \quad (10)$$

Applying Eq. (10) to Eq. (4) we obtain with the help of Eqs. (8) and (9)

$$\begin{aligned} \underline{0} &= \nabla_{\underline{\beta}} [\underline{y}^T \underline{y} - \underline{y}^T \underline{x} \underline{\beta} - \underline{\beta}^T \underline{x}^T \underline{y} + \underline{\beta}^T \underline{x}^T \underline{x} \underline{\beta}] |_{\underline{\beta}=\hat{\underline{\beta}}} \\ &= -2\underline{x}^T \underline{y} + 2\underline{x}^T \underline{x} \hat{\underline{\beta}} \end{aligned} \quad (11)$$

or

$$(\underline{x}^T \underline{x}) \hat{\underline{\beta}} = \underline{x}^T \underline{y} \quad . \quad (12)$$

Now $(\underline{x}^T \underline{x})$ is a $m \times m$ symmetric matrix having an inverse. Hence

$$\hat{\underline{\beta}} = (\underline{x}^T \underline{x})^{-1} \underline{x}^T \underline{y} \quad , \quad (13)$$

which is Eq. (D-6).

2. The Gauss-Markov Theorem

Our treatment follows that of Koerts and Abrahamse (1969, p. 11). Consider an arbitrary linear estimate $\underline{\theta}^* = \underline{A}\underline{y}$, which is also unbiased, i.e., $E(\underline{\theta}^*) = \underline{\theta}$. This property requires that

$$\begin{aligned}\underline{\theta} &= E(\underline{A}\underline{y}) \\ &= E[\underline{A}(\underline{X}\underline{\theta} + \underline{\epsilon})] \\ &= E(\underline{A}\underline{X}\underline{\theta}) ,\end{aligned}\tag{14}$$

since $E(\underline{\epsilon}) = 0$. Eq. (14) can only be satisfied if

$$\underline{A}\underline{X} = \underline{I} ,\tag{15}$$

where the identity matrix \underline{I} is of order m . Hence the matrix \underline{A} can be written as

$$\underline{A} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T + \underline{B}\tag{16}$$

where

$$\underline{B}\underline{X} = \underline{0} .\tag{17}$$

We now show that among all the unbiased linear estimates, $\underline{\beta} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}$ has minimum variance. We form the variance-covariance matrix of $\underline{\beta}^*$. First, we see that

$$\begin{aligned} \underline{\beta}^* - \underline{\beta} &= [(\underline{X}^T \underline{X})^{-1} \underline{X}^T + \underline{B}] \underline{y} - \underline{\beta} \\ &= [(\underline{X}^T \underline{X})^{-1} \underline{X}^T + \underline{B}] [\underline{X} \underline{\beta} + \underline{\epsilon}] - \underline{\beta} \\ &= [(\underline{X}^T \underline{X})^{-1} \underline{X}^T + \underline{B}] \underline{\epsilon} . \end{aligned} \quad (18)$$

It then follows that

$$\begin{aligned} E[(\underline{\beta}^* - \underline{\beta})(\underline{\beta}^* - \underline{\beta})^T] &= E\{[(\underline{X}^T \underline{X})^{-1} \underline{X}^T + \underline{B}] \underline{\epsilon} \underline{\epsilon}^T [\underline{X}(\underline{X}^T \underline{X})^{-1} + \underline{B}^T]\} \\ &= [(\underline{X}^T \underline{X})^{-1} \underline{X}^T + \underline{B}] E(\underline{\epsilon} \underline{\epsilon}^T) [\underline{X}(\underline{X}^T \underline{X})^{-1} + \underline{B}^T] \\ &= \sigma^2 [(\underline{X}^T \underline{X})^{-1} + \underline{B} \underline{B}^T] , \end{aligned} \quad (19)$$

where we have used the fact that $(\underline{X}^T \underline{X})$ and $(\underline{X}^T \underline{X})^{-1}$ are symmetric matrices. $\underline{B} \underline{B}^T$ is a symmetric matrix all of whose diagonal elements are greater than or equal to zero. Therefore the choice $\underline{B} = \underline{0}$, corresponding to $\underline{A} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T$ and thus $\underline{\beta}^* = \hat{\underline{\beta}}$, gives the minimum variance.

3. Estimated Variance of the Measurement Errors

Since the error vector $\underline{\epsilon}$ is equal to $\underline{y} - \underline{X} \underline{\beta}$, where $\underline{\beta}$ is unknown, $\underline{\epsilon}$ cannot be observed. Instead, we use the

least-squares estimate for the errors

$$\hat{\underline{e}} = \underline{y} - \underline{X}\hat{\underline{\beta}} = \underline{N}\underline{y} \quad (20)$$

where

$$\underline{N} = [\underline{I} - \underline{X}(\underline{X}^T \underline{X})^{-1} \underline{X}^T] \quad (21)$$

The $n \times n$ matrix \underline{N} has a number of special properties:

$$\underline{N} = \underline{N}^T \quad (22)$$

$$\underline{N}\underline{X} = \underline{0} \quad (23)$$

$$\underline{N} = \underline{N}\underline{N} = \underline{N}^T \underline{N} = \underline{N}\underline{N}^T \quad (24)$$

$$\text{trace } \underline{N} = n-m \quad (25)$$

These are all easily proven but perhaps Eq. (25) needs some elaboration. The trace of a square matrix is equal to the sum of the diagonal elements. The trace operation has the property $\text{trace}(\underline{A} + \underline{B}) = \text{trace } \underline{A} + \text{trace } \underline{B}$ and $\text{trace}(\underline{AB}) = \text{trace } (\underline{BA})$, \underline{A} and \underline{B} being square matrices of the same order. Hence

$$\begin{aligned}
\text{trace } \underline{N} &= \text{trace}[\underline{I} - \underline{X}(\underline{X}^T \underline{X})^{-1} \underline{X}^T] \\
&= \text{trace } \underline{I} - \text{trace}[\underline{X}(\underline{X}^T \underline{X})^{-1} \underline{X}^T] \\
&= \text{trace } \underline{I} - \text{trace}[(\underline{X}^T \underline{X})^{-1} (\underline{X}^T \underline{X})] \\
&= n - m \quad .
\end{aligned} \tag{26}$$

We now find an estimate for the unknown variance σ^2 of the errors. A well-known estimate of σ^2 is the mean square of the residuals

$$\begin{aligned}
s^2 &= \frac{1}{n} \hat{\underline{\epsilon}}^T \hat{\underline{\epsilon}} \\
&= \frac{1}{n} (\underline{N}\underline{y})^T \underline{N}\underline{y} \\
&= \frac{1}{n} \underline{y}^T \underline{N}^T \underline{N}\underline{y} \\
&= \frac{1}{n} \underline{y}^T \underline{N}\underline{y} \quad .
\end{aligned} \tag{27}$$

This estimate is not used much because it is biased, i.e., $E(s^2) \neq \sigma^2$. This bias can be removed, however, by noting that

$$\begin{aligned}
E(s^2) &= \frac{1}{n} E[\underline{y}^T \underline{N}\underline{y}] = \frac{1}{n} E\{[\underline{X}\underline{\beta} + \underline{\epsilon}]^T \underline{N}(\underline{X}\underline{\beta} + \underline{\epsilon})\} \\
&= \frac{1}{n} E[\underline{\epsilon}^T \underline{N}\underline{\epsilon}] \quad ,
\end{aligned} \tag{28}$$

where use is made of $\underline{NX} = \underline{0}$. Since $E(\underline{\epsilon}\underline{\epsilon}^T) = \sigma^2 \underline{I}$, and $\underline{\epsilon}^T \underline{N} \underline{\epsilon}$ is a scalar,

$$\begin{aligned} E(s^2) &= \frac{1}{n} E[(\underline{\epsilon}^T \underline{N} \underline{\epsilon})] \\ &= \frac{\sigma^2}{n} \text{trace } \underline{N} \\ &= \frac{n-m}{n} \sigma^2 . \end{aligned} \quad (29)$$

It is now clear that $\frac{n}{n-m} s^2$ is an unbiased estimate of σ^2 , i.e.,

$$\hat{\sigma}^2 = \frac{1}{n-m} [\underline{y} - \underline{X}\hat{\underline{\beta}}]^T [\underline{y} - \underline{X}\hat{\underline{\beta}}] , \quad (30)$$

which is (D-8) with $m=5$.

4. Estimated Variance-Covariance Matrix

The variance-covariance matrix of $\underline{\beta}$ is computed from

$$\begin{aligned} E\{(\underline{\beta} - \hat{\underline{\beta}})(\underline{\beta} - \hat{\underline{\beta}})^T\} &= E\{[\underline{\beta} - (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}][\underline{\beta} - (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y}]^T\} \\ &= E\{[\underline{\beta} - (\underline{X}^T \underline{X})^{-1} \underline{X}^T (\underline{X}\underline{\beta} + \underline{\epsilon})][\underline{\beta} - (\underline{X}^T \underline{X})^{-1} \underline{X}^T (\underline{X}\underline{\beta} + \underline{\epsilon})]^T\} \\ &= E\{[(\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\epsilon}][(\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{\epsilon}]^T\} \\ &= (\underline{X}^T \underline{X})^{-1} \underline{X}^T E(\underline{\epsilon}\underline{\epsilon}^T) \underline{X} (\underline{X}^T \underline{X})^{-1} \\ &= \sigma^2 (\underline{X}^T \underline{X})^{-1} , \end{aligned} \quad (31)$$

which is Eqs. (D-9) and (D-10).

5. Correlated Least-Squares Estimates

In correlated least-squares (of which unweighted and weighted least-squares are special cases), we begin once again with the linear model

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\epsilon} , \quad (32)$$

but here the errors $\underline{\epsilon}$, while having zero mean, also have in general unequal variances and non-zero covariances. The variance-covariance matrix of the errors is given by

$$E(\underline{\epsilon}\underline{\epsilon}^T) = \sigma^2 \underline{M} , \quad (33)$$

where \underline{M} is a known positive definite (i.e., all eigenvalues > 0) symmetric matrix and σ^2 is an unknown common factor. The unweighted least-squares formulation cannot be applied here because of the form of this variance-covariance matrix. However, this obstacle can be removed by finding a square matrix \underline{P}^T , independent of $\underline{\epsilon}$, that transforms the errors $\underline{\epsilon}$ into $\underline{\gamma}$

$$\underline{\gamma} = \underline{P}^T \underline{\epsilon} , \quad (34)$$

such that

$$E(\underline{Y}\underline{Y}^T) = \sigma^2 \underline{I} \quad . \quad (35)$$

Substituting Eq. (34) into Eq. (35) we find

$$\begin{aligned} \sigma^2 \underline{I} &= E[(\underline{P}^T \underline{\epsilon})(\underline{P}^T \underline{\epsilon})^T] \\ &= E[\underline{P}^T \underline{\epsilon} \underline{\epsilon}^T \underline{P}] \\ &= \sigma^2 \underline{P}^T \underline{M} \underline{P} \quad . \end{aligned} \quad (36)$$

Hence the transformation must have the property

$$\underline{P}^T \underline{M} \underline{P} = \underline{I} \quad . \quad (37)$$

We show next that it is always possible to find such a transformation.

Since \underline{M} is a real Hermitian matrix, there exists a real unitary matrix \underline{U} , i.e.

$$\underline{U}^T = \underline{U}^{-1} \quad , \quad (38)$$

that diagonalizes \underline{M}

$$\underline{U}^T \underline{M} \underline{U} = \underline{D} , \quad (39)$$

where \underline{D} is a diagonal matrix whose elements λ_i are the eigenvalues of \underline{M} . Since \underline{M} is a positive definite matrix, $\lambda_i > 0$. Thus we can define a matrix $\underline{D}^{\frac{1}{2}}$

$$D_{ii}^{\frac{1}{2}} = (\lambda_i)^{\frac{1}{2}} \quad (40)$$

so that

$$\underline{D} = \underline{D}^{\frac{1}{2}} \underline{D}^{\frac{1}{2}} . \quad (41)$$

The diagonal matrix $\underline{D}^{\frac{1}{2}}$ is nonsingular and its inverse is also a diagonal matrix $\underline{D}^{-\frac{1}{2}}$ where

$$D_{ii}^{-\frac{1}{2}} = (\lambda_i)^{-\frac{1}{2}} . \quad (42)$$

Then the matrix

$$\underline{P} = \underline{U} \underline{D}^{-\frac{1}{2}} \quad (43)$$

has the desired property we seek, namely,

$$\begin{aligned}
\underline{P}^T \underline{M} \underline{P} &= (\underline{D}^{-\frac{1}{2}})^T \underline{U}^T \underline{M} \underline{U} \underline{D}^{-\frac{1}{2}} \\
&= (\underline{D}^{-\frac{1}{2}})^T \underline{D} \underline{D}^{-\frac{1}{2}} \\
&= \underline{D}^{-\frac{1}{2}} \underline{D}^{\frac{1}{2}} \underline{D}^{\frac{1}{2}} \underline{D}^{-\frac{1}{2}} \\
&= \underline{I} \quad , \quad (44)
\end{aligned}$$

where we make use of the fact $(\underline{D}^{-\frac{1}{2}})^T = \underline{D}^{-\frac{1}{2}}$ since a diagonal matrix is always symmetric. Moreover, the matrix \underline{P} is nonsingular and has an inverse given by

$$\begin{aligned}
\underline{P}^{-1} &= (\underline{U} \underline{D}^{-\frac{1}{2}})^{-1} \\
&= \underline{D}^{\frac{1}{2}} \underline{U}^{-1} \\
&= \underline{D}^{\frac{1}{2}} \underline{U}^T \quad . \quad (45)
\end{aligned}$$

Hence we find from Eq. (37) that

$$\underline{M} = (\underline{P}^T)^{-1} \underline{P}^{-1} \quad (46)$$

and

$$\begin{aligned}
\underline{M}^{-1} &= [(\underline{P}^T)^{-1} \underline{P}^{-1}]^{-1} \\
&= \underline{P} \underline{P}^T \quad . \quad (47)
\end{aligned}$$

We now transform Eq. (30) by premultiplying by \underline{P}^T

$$\underline{P}^T \underline{Y} = \underline{P}^T \underline{X} \underline{\beta} + \underline{P}^T \underline{\epsilon} \quad (48)$$

Equation (48) can be recognized as the unweighted least-squares model equation with $\underline{P}^T \underline{Y}$ as the new observation vector with errors $\underline{P}^T \underline{\epsilon}$ and a new known coefficient matrix $\underline{P}^T \underline{X}$. Hence, we use the results derived previously for the unweighted least-squares to obtain the MVLU estimates in the general case

$$\begin{aligned} \hat{\underline{\beta}} &= [(\underline{P}^T \underline{X})^T (\underline{P}^T \underline{X})]^{-1} (\underline{P}^T \underline{X})^T \underline{P}^T \underline{Y} \\ &= [\underline{X}^T \underline{P} \underline{P}^T \underline{X}]^{-1} \underline{X}^T \underline{P} \underline{P}^T \underline{Y} \\ &= [\underline{X}^T \underline{M}^{-1} \underline{X}]^{-1} \underline{X}^T \underline{M}^{-1} \underline{Y} \quad (49) \end{aligned}$$

The variance-covariance matrix of $\underline{\beta}$ is computed from

$$\begin{aligned} \hat{\underline{\Theta}} &= \hat{\sigma}^2 [(\underline{P}^T \underline{X})^T (\underline{P}^T \underline{X})]^{-1} \\ &= \hat{\sigma}^2 [\underline{X}^T \underline{P} \underline{P}^T \underline{X}]^{-1} \\ &= \hat{\sigma}^2 [\underline{X}^T \underline{M}^{-1} \underline{X}]^{-1} \quad (50) \end{aligned}$$

where the best estimate of σ^2 is given by

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-m} (\underline{P}^T \underline{\epsilon})^T (\underline{P}^T \underline{\epsilon}) \\
&= \frac{1}{n-m} \underline{\epsilon}^T \underline{P} \underline{P}^T \underline{\epsilon} \\
&= \frac{1}{n-m} [\underline{y} - \underline{X}\hat{\beta}]^T \underline{M}^{-1} [\underline{y} - \underline{X}\hat{\beta}]
\end{aligned} \tag{51}$$

Note that $\hat{\beta}$, $\hat{\theta}$, and $\hat{\sigma}^2$ can be calculated without having to know the transformation matrix \underline{P}^T explicitly. Eqs. (49-51) are the general correlated least squares Eqs. (F-6), (F-7), and (F-8).

REFERENCES

Albritton, D. L., W. J. Harrop, A. L. Schmeltekopf, and R. N.

Zare (1973a) Rotational analysis of the O_2^+ $A^2\Pi_u - X^2\Pi_g$ second negative band system, J. Mol. Spectrosc. 46, 89-102.

Albritton, D. L., W. J. Harrop, A. L. Schmeltekopf, and R. N.

Zare (1973b) Resolution of the discrepancies concerning the optical and microwave values for B_0 and D_0 of the $X^3\Sigma_g^-$ state of O_2 , J. Mol. Spectrosc. 46, 103-118.

Albritton, D. L., W. J. Harrop, A. L. Schmeltekopf, R. N. Zare,

and E. L. Crow (1973) A critique of the term value approach to determining molecular constants from the spectra of diatomic molecules, J. Mol. Spectrosc. 46, 67-88.

Albritton, D. L., A. L. Schmeltekopf, J. Tellinghuisen, and R. N.

Zare (1974) Least-squares equivalence of different representations of rotational constants, J. Mol. Spectrosc. 53, 311-314.

Albritton, D. L., A. L. Schmeltekopf, R. N. Zare, and J. Czarny

(1975) Rotational analysis of the O_2^+ $b^4\Sigma_g^- - a^4\Pi_u$ first negative band system, to be published.

Anscombe, F. J. (1960) Rejection of outliers, Technometrics 2,

123-147.

- Åslund, N. (1974) A numerical method for the simultaneous determination of term values and molecular constants, J. Mol. Spectrosc. 50, 424-434.
- Bard, Y. (1974) "Nonlinear Parameter Estimation," Academic Press, New York.
- Beaton, A. E. and J. W. Tukey (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, Technometrics 16, 147-185.
- Bennett, C. A. And N. L. Franklin (1954) "Statistical Analysis in Chemistry and the Chemical Industry," John Wiley and Sons, Inc., New York.
- Britt, H. I. and R. H. Luecke (1972) The estimation of parameters in nonlinear implicit models, Technometrics 15, 233-247.
- Cohen, E. R. and B. N. Taylor (1973) The 1973 least-squares adjustment of the fundamental constants, J. Phys. Chem. Ref. Data 2, 663-734.
- Coxon, J. A. (1974) Rotational vibrational analysis of the $B^3\Pi(0^+) - X^1\Sigma^+$ band system of $^{79}\text{Br}^{35}\text{Cl}$ and $^{81}\text{Br}^{35}\text{Cl}$, J. Mol. Spectrosc. 50, 142-165.
- Curl, R. F. (1970) Diagnostic least squares, J. Comp. Phys. 6, 367-377.
- Daniel, C. and F. S. Wood (1971) "Fitting Equations to Data," Wiley-Interscience, New York.

- Deming, W. E. (1943) "Statistical Adjustment of Data," John Wiley and Sons, New York.
- Dixon, W. J. and F. J. Massey, Jr. (1969) "Introduction to Statistical Analysis," Third Edition, McGraw-Hill Book Co., New York.
- Draper, N. R. and H. Smith (1966) "Applied Regression Analysis," John Wiley and Sons, Inc., New York.
- Federighi, E. T. (1959) Extended tables of the percentage points of Student's t-distribution, J. Amer. Statist. Assoc. 54, 683-688.
- Hamilton, W. C. (1964) "Statistics in Physical Science," The Ronald Press Company, New York.
- Herzberg, G. (1950) "Spectra of Diatomic Molecules," D. Van Nostrand Company, Inc., Princeton, New Jersey.
- Hunter, G., S. M. Khandekar, and S. M. L. Prokopenko, A minimum covariance basis for diatomic vibrational constants, J. Mol. Spectrosc. 53, 393-401.
- Kildal, H., R. S. Eng, and A. H. M. Ross (1974) Heterodyne measurements of $^{12}\text{C}^{18}$ laser frequencies and improved Dunham coefficients, J. Mol. Spectrosc. 53, 479-488.
- Kirchhoff, W. H. (1972) On the calculation and interpretation of centrifugal distortion constants: a statistical basis for model testing: the calculation of the force field, J. Mol. Spectrosc. 41, 333-380.

- Koerts, J. and A. P. Abrahamse (1969) "On the Theory and Applications of the General Linear Model," Rotterdam University Press, Rotterdam.
- Lees, R. M. (1973) Least-squares analysis: correlated data in treatment of frequency splittings, *J. Mol. Spectrosc.* 47, 286-296.
- Morino, Y. and S. Saito (1972) "Microwave Spectra" in "Molecular Spectroscopy: Modern Research," K. Narahari Rao and C. Weldon Mathews, Eds., Academic Press, New York.
- Pliva, J. and W. B. Telfair (1974) Correlations and accuracy of estimation of spectroscopic constants and term values, *J. Mol. Spectrosc.* 53, 221.
- Powell, D. R. and J. R. MacDonald (1972) A rapidly convergent iterative method for solution of the generalised nonlinear least-squares problem, *Computer J.* 15, 148-155.
- Ramsay, D. A. (1974) "Critical Evaluation of Molecular Constants from Optical Spectra," in "Critical Evaluation of Chemical and Physical Structural Information," Proceedings of a Conference at Dartmouth College, June 24-29, 1973, editors D. R. Lide, Jr. and M. A. Paul, National Academy of Sciences, Washington, D. C.

- Rao, K. Narahari (1972) "Large Plane Gratings for High-Resolution Infrared Spectrographs" in "Molecular Spectroscopy: Modern Research," K. Narahari Rao and C. Weldon Mathews, Eds., Academic Press, New York.
- Roh, W. B. and K. Narahari Rao (1974) CO laser spectra, J. Mol. Spectrosc. 49, 317-321.
- Rossini, F. D. and W. E. Deming (1939) The assignment of uncertainties to the data of chemistry and physics, with specific recommendations for thermochemistry, J. Wash. Acad. Sci. 29, 416-441.
- Saenger, K. L., R. N. Zare, and C. W. Mathews (1975) Reanalysis of the HCl^+ A-X band system, J. Mol. Spectrosc. (to be published).
- Scheffé, H. (1959) "The Analysis of Variance," John Wiley and Sons, New York.
- Tryon, P. V. (1975) Generalized nonlinear least-squares, a perspective (to be published).
- Tukey, J. W. (1974) "Introduction to Today's Data Analysis," in "Critical Evaluation of Chemical and Physical Structural Information," Proceedings of a Conference at Dartmouth College, June 24-29, 1973, editors D. R. Lide, Jr. and M. A. Paul, National Academy of Sciences, Washington, D. C.

- Worthing, A. C. and J. Geffner (1943) "Treatment of Experimental Data," John Wiley and Sons, Inc., New York.
- Zare, R. N., A. L. Schmeltekopf, W. J. Harrop, and D. L. Albritton (1973) A direct approach for the reduction of diatomic spectra to molecular constants for the construction of RKR potentials, J. Mol. Spectrosc. 46, 37-66.

