RF Project 761420/711626
Semi-Annual Report

the
ohio
state
university

research foundation
1314 kinnear road
columbus, ohio
43212

INVESTIGATION OF INFRARED SPECTRA OF ATMOSPHERIC GASES TO SUPPORT STRATOSPHERIC SPECTROSCOPIC INVESTIGATIONS

John H. Shaw
Department of Physics

For the Period
1 November 1978 - 30 April 1979

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D. C. 20546

Grant No. NSG-7479

June, 1979

INVESTIGATION OF INFRARED SPECTRA OF ATMOSPHERIC GASES TO SUPPORT STRATOSPHERIC SPECTROSCOPIC INVESTIGATIONS

by

John H. Shaw
Department of Physics

1 November 1978 - 30 April 1979

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D. C. 20546

Grant No. NSG-7479

June, 1979

Ozone Spectra

A survey spectrum of ozone between 1850 and 4500 cm^{-1} is shown in Fig. 1. The sample contained approximately 10 atm cm O_3 at 750 Torr total pressure (O_2 + O_3) and the spectral resolution is about 0.5 cm^{-1}. At this pressure and resolution few of the details in the bands can be identified. In addition to bands of ozone, there is also absorption by H_2O and CO_2 in the optical path outside the absorption cell, and absorption by CO produced by reactions of ozone with carbon compounds in the cell and CH_4, an impurity in the oxygen.

Absorption by ozone occurs from 1850 to 2250 cm^{-1}. This is due to the wings of the $\nu_1 + \nu_2$ band near 1790 cm^{-1}, by the $\nu_1 + \nu_3$ band at 2110 cm^{-1}, and the $2\nu_1$ band near 2200 cm^{-1}. The $\nu_2 + 2\nu_3$ and $\nu_1 + \nu_2 + \nu_3$ bands are seen between 2700 and 2800 cm^{-1}. There is also absorption by O_3 between 2950 and 3050 cm^{-1} and weaker features near 3200 and 4030 cm^{-1}. High resolution spectra of these bands will be obtained in the next several months and an atlas showing their appearance for different sample conditions will be prepared together with a listing of the line positions. Estimates of the line intensities and widths will be obtained in future work.

The spectra in Fig. 1 and other studies have shown that bands of ozone consist of densely packed series of lines which are incompletely resolved even with a spectrometer of 0.04 cm^{-1} resolution. The blending of the lines makes it difficult to measure the line positions and even more difficult to estimate the line intensities and widths. The conventional methods of estimating these latter parameters from measurements of the equivalent widths fail and other methods of analysis are desirable.

A very powerful method of extracting line parameter information from spectra has recently been developed by our group. This consists of non-linear, least-squares, fitting of the observed spectral signals and the results obtained are usually superior to those obtained by other techniques. Even this method cannot be successfully applied to single lines when the number of signal values in the vicinity of the line is small. Thus further improvements in the curve fitting method are required. Some of these have been investigated during the past few months. Several approaches have been considered and a review of our recent work is given below.

Laboratory spectra are usually obtained from samples containing arbitrary amounts of the absorbing gas and at arbitrary total pressures. Very low pressures are usually used if the highest resolution of the spectral details is required. This improvement is due to the decrease in the line widths with pressure. As the pressure is decreased, the line shape changes from the Lorentz shape to the Voigt shape. When the line widths are smaller than the spectral resolution, it becomes increasingly difficult to retrieve the line intensities and widths.

These considerations suggest that, for a given spectral line, observed with an instrument whose performance characteristics are known, there are some optimum sample conditions which allow the best estimates of the line

parameters to be obtained. The estimation of these optimum conditions is a problem of experimental design. Although the principles of experiment design have been applied to many types of investigations, there have been few applications of these concepts to the design of spectroscopic experiments.

Dr. E. Niple has investigated the types of information contained in spectra of Lorentz shaped lines obtained with instruments of finite resolution by using the statistical methods developed by, for example, Fisher and Kullback.

As a result of this work, two articles have been prepared and copies are enclosed with this report. The first of these, "Information Measures in Nonlinear Experimental Design," describes the general method used to identify the optimum experimental conditions required to obtain the best estimates of a desired parameter. The second "Information in Spectra of Collision Broadened Absorption Lines" describes the application to Lorentz shaped lines. The approaches described in these papers can be applied to many other spectroscopic problems.

More recently, Dr. Niple has begun to use these methods to investigate the types of information which can be expected from spectra of the stratosphere. The experiment chosen for analysis was that of occultation solar spectra obtained with an instrument of moderately high resolution. Some preliminary results are described in an interim report currently being prepared.

These approaches allow optimum experimental designs to be evaluated and also the nature of the laboratory spectroscopic data required for the analyses to be quantitatively described.

Concurrently with these investigations, analyses of the type of information which can be retrieved from spectra of Voigt shaped lines have been undertaken. The attached article "Least Squares Analysis of Voigt Shaped Lines" has been accepted by JQSRT. This article shows that a number of parameters of Voigt lines can be estimated provided a sufficient number of signal values are available for analysis.

These investigations have shown that it is possible to define optimum configurations for spectroscopic measurements and that all the important line parameters can be estimated from a single spectrum of an absorption line. They also indicate that it is unlikely that all the line information required for spectroscopic investigations of the stratosphere can be obtained by a line by line analysis of ozone spectra obtained with an instrument of 0.04 cm^{-1} resolution.

However, the aim of many analyses of high resolution spectra of gases is to describe the individual line characteristics by models containing far fewer adjustable parameters than the number of lines observed. Although the values of these adjustable parameters are usually obtained by fitting the models to the measurements of individual lines, it is clear that a

considerable reduction in the time and effort required can occur if the parameter values are obtained by simultaneous analyses of all the lines in the band. From this viewpoint it is also clear that, even though there may not be sufficient information in a given series of spectra to allow estimates of the individual line parameters to be obtained separately, estimates of the important band parameters can still be obtained by collecting all the information concerning them from all of the lines simultaneously.

This approach has been tested by analyzing low resolution (0.1 cm^{-1}), rather noisy spectra of HCl and CO. Some initial results are described in the attached article "Band Analysis by Spectral Curve Fitting." The results are very encouraging.

These, and other investigations by our group were described in a series of presentations at the Thirty-Fourth Symposium on Molecular Spectroscopy, The Ohio State University, June 1979. Abstracts of these talks are attached.

The results obtained thus far indicate that a substantial increase in the precision with which band parameter values can be estimated can be achieved by analyzing higher-resolution, less-noisy spectra. Bands of HCl and CO consist of well separated lines, in marked contrast to the densely packed lines in bands of O_3. Thus the extension of these whole band analysis techniques to ozone spectra is a formidable undertaking. We intend to proceed in this direction by analyzing bands of N_2O and CO_2, before tackling the more complicated problems of ozone. The band models for these molecules are considerably less complicated than those of ozone. However, the presence of overlapping bands of isotopic molecules and other hot bands will serve as a crucial test of this method.

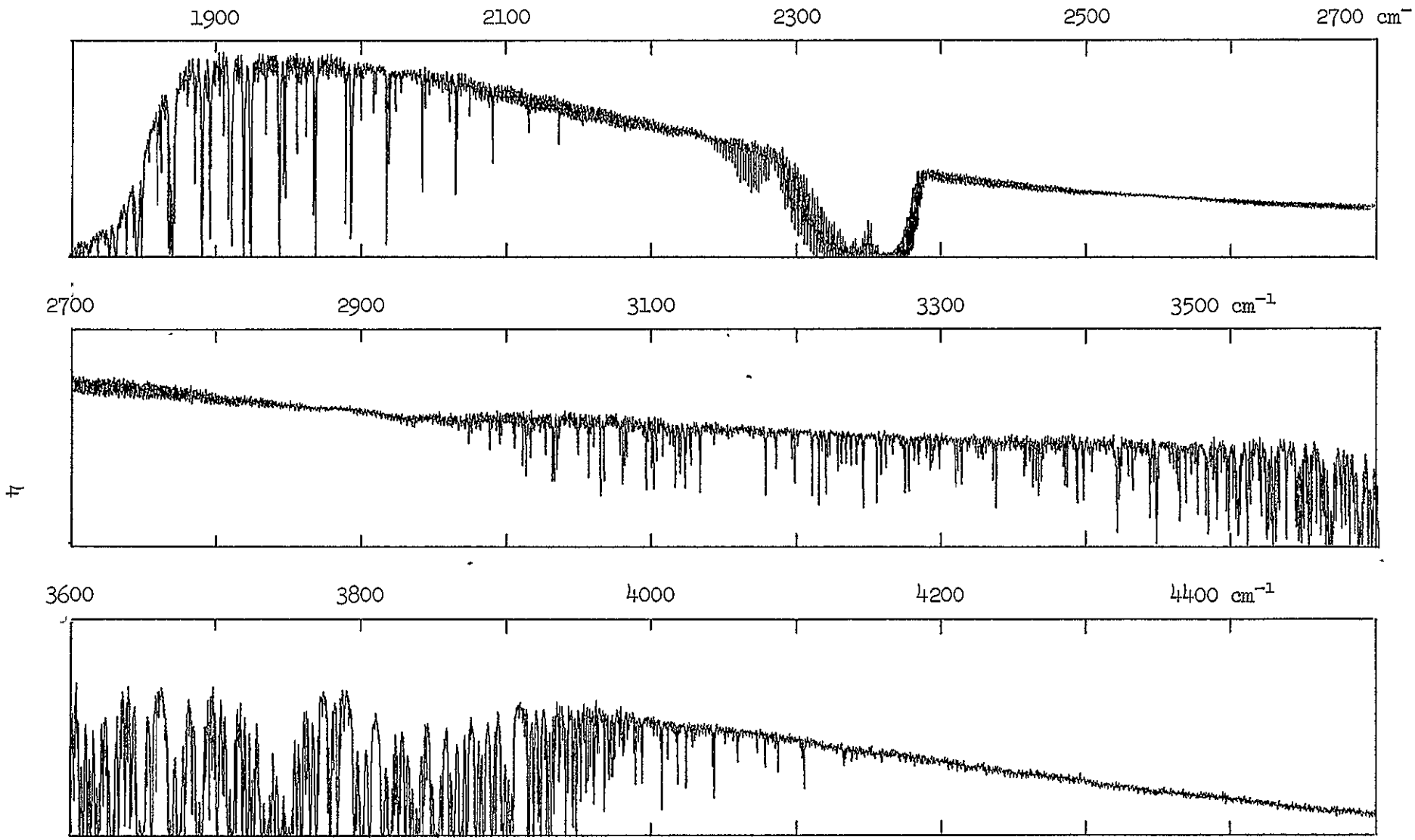


Fig. 1a - Background spectrum; 0.5 cm^{-1} resolution

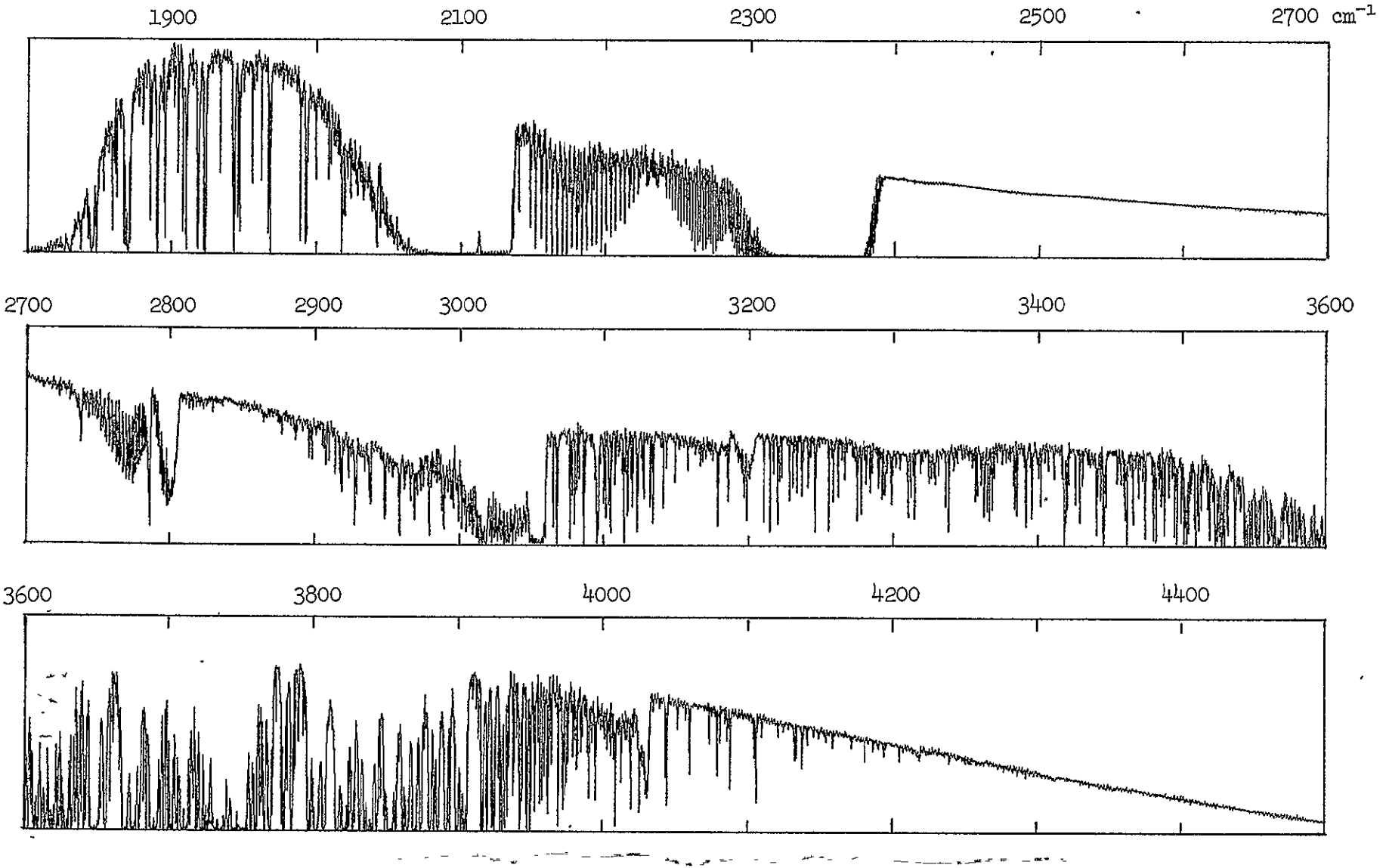


Fig. 1b - Spectrum of approximately 10^{-2} atm cm O_3; 0.5 cm^{-1} resolution

INFORMATION MEASURES IN NONLINEAR EXPERIMENTAL DESIGN*

E. Niple
J. H. Shaw

Department of Physics
The Ohio State University
Columbus, Ohio 43210

*This work supported in part by NASA Grant No. NS9-7479.

ABSTRACT

Some different approaches to the problem of designing experiments which estimate the parameters of nonlinear models are discussed. The assumption in these approaches that the information in a set of data can be represented by a scalar is criticized, and the nonscalar discrimination information is proposed as the proper measure to use. The two-step decay example in Box and Lucas (1959) is used to illustrate the main points of the discussion.

Introduction

Many experiments involve estimating the parameters of nonlinear models. The purpose of this paper is to collect together and organize some different approaches to the design problem for such experiments. The basis for this organization is what Rosenkrantz (1970, p.58) has called "the obvious analogy" between experimentation, or measurement processes, and communication processes. It will be shown that this analogy allows the design problem to be treated as an Information Theory (IT) problem. Since IT is interdisciplinary in nature, the results can be applied to many different fields, in our case, experimental physics and chemistry. Most of the examples discussed below are from these areas.

There have been several previous attempts to pursue this analogy.* We shall not review these here because they have not found wide spread use among experimental designers. We feel that this is due, in part, to several inappropriate fundamental assumptions carried over from Communication Theory without sufficient justification. (By Communication Theory we mean that impressive body of work which followed from the papers of Shannon and Wiener.) Principal among these assumptions is that the formal quantity which represents the information in a set of data has to be scalar. Equally unfortunate is the neglect in these approaches of what may be called (with apologies to Einstein) the Relativity Principle of Belief, which states that the information in a set of data is independent of the beliefs and prior expectations of the experimenter.

*Perhaps the most well known is that of Brillouin (1956). Many more can be found in the journal Measurement Techniques.

We do not need to propose a new IT approach to the nonlinear design problem. Instead, some of the non-IT approaches currently in use are reviewed and it is shown that these can be restated in terms of an existing IT formulation whose roots lie in statistics instead of Communication Theory.

The Box Lucas Approach

Perhaps the most theoretically developed of the non-IT approaches is that introduced in the well known article by Box and Lucas (1959) and extended by Draper and Hunter (1966). (See St. John and Draper (1975) for a recent review of the use of this technique.) The general outline of the problem they discussed is (Box and Lucas (1959; p.77))

"...some response η is a known function

$$\eta = f(\xi_1,\xi_2,\ldots,\xi_k,\theta_1,\theta_2,\ldots,\theta_p) = f(\underline{\xi},\underline{\theta}) \qquad (1)$$

of k variables whose levels are denoted by the elements $\xi_1,\ldots,\xi_i,\ldots,\xi_k$ of the vector $\underline{\xi}$ and of p parameters $\theta_1,\ldots,\theta_r,\ldots,\theta_p$ elements of the vector $\underline{\theta}$... The problem here considered is that of selecting a programme of trials [i.e., a set of experiments] such that they may be expected to provide results from which the p parameters can be estimated with high accuracy. In general the experimental programme may be defined by an Nxk matrix $D = \{\xi_{iu}\}$ called the design matrix. The u^{th} row $\underline{\xi}_u'$ of this matrix with elements $\xi_u,\ldots,\xi_{iu},\ldots,\xi_{ku}$ provides the levels of the k variables at which the response is to be observed in the u^{th} trial."

We assume that this problem may be reformulated as: how can the most information regarding the parameters be obtained? Box and Lucas appear to share this assumption because, when they consider the inclusion of experimental costs, they write (p.80), "One could then attempt to find the design giving the *most information* [italics added] for a given cost." In a later article (Box (1970, p.585)) the Box and Lucas design criterion (defined below) which is to be maximized for

the optimum design, is even referred to as the "information" of a set of data. It should be noted, however, that this criterion is not a valid information measure, as such are usually defined in IT, since it is not additive for independent sets of data.

Assume now that y_u, the observed response for the u^{th} set of experimental conditions $\underline{\xi}_u$, is such that

$$E(y_u) = \eta_u = f(\underline{\xi}u;\underline{\theta}), \tag{2}$$

and

$$E(y_u-\eta_u)(y_v-\eta_v) = \begin{cases} \sigma^2, & u=v \\ 0, & u\neq v \end{cases} \qquad u,v=1,\ldots,N, \tag{3}$$

where σ^2, the variance of the observed responses, is unknown in general. One estimate for $\underline{\theta}_T$, the true values of the parameters, is the least squares estimate, $\underline{\hat{\theta}}$. This minimizes the sum of squares

$$\sum_{u=1}^{N} \{y_u - f(\underline{\xi}u;\underline{\theta})\}^2$$

and has an approximate variance-covariance matrix $\underline{S}$ given by

$$\underline{S} = \sigma^2\{\underline{F}'\underline{F}\}^{-1}, \tag{4}$$

where

$$\underline{F} = \{f_{ru}\},$$

and

$$f_{ru} \equiv \left[\frac{\partial f}{\partial \theta_r}(\xi_u;\underline{\theta})\right]_{\underline{\theta}=\underline{\hat{\theta}}}$$

The Fisher information matrix $\underline{W}$, used below, is related to $\underline{S}$ according to

$$\underline{W} = \underline{S}^{-1} = 1/\sigma^2\ \{\underline{F}'\underline{F}\}. \tag{5}$$

The Box and Lucas technique requires that the design matrix $\underline{D}$ be chosen so that the determinant $|\underline{F}'\underline{F}|$ (or equivalently $|\underline{W}|$) is maximized.

Thus the determinant of the Fisher information matrix is used as the information measure in this approach. Such a design is referred to as a D-optimal design (Kiefer and Wolfowitz (1959).

When $p = 1$, i.e. there is just one unknown parameter, maximizing $|\underline{W}|$ corresponds to minimizing σ^2_θ, the variance of the parameter estimate. This is an obvious choice in this instance; however, it is by no means obvious that $|\underline{W}|$ should be used when $p \geqq 2$.

A Specific Example

To illustrate the difficulties encountered by using this approach, we return to the example discussed by Box and Lucas. This involves observing the yield η as a function of the observation time ξ_1 of an intermediate species B in the consecutive reaction

$$A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C, \tag{6}$$

where each $\xrightarrow{\theta_i}$ represents a first-order, irreversible decay with rate θ_i. It may be shown that

$$\eta(\xi_1) = \{\theta_1/(\theta_1-\theta_2)\}\{\exp(-\theta_2\xi_1) - \exp(-\theta_1\xi_1)\}. \tag{7}$$

If just two values of the yield are to be observed ($k = 2$), the Box and Lucas optimum design is obtained from Figure 1. Here the contours of $|\underline{F}|$ are shown as functions of the two observation times ξ_{11} and ξ_{12}. The design criterion is simply

$$|\underline{F}'\underline{F}| = \{|\underline{F}|\}^2$$

when $k = p$. In order to estimate the optimum design, the values of θ_1 and θ_2 must be estimated. We have used the same values as Box and

Lucas:

$$\theta_1 = 0.7 \quad \text{and} \quad \theta_2 = 0.2.$$

Figure 1 is similar to Fig. 3, p.83 of their article. The D-optimal design occurs at

$$\xi_{11} = 1.23, \qquad \xi_{12} = 6.86 \tag{8}$$

with $\quad |\underline{F}'\underline{F}| = 0.66.$

For this simple case of two observations and two unknowns, the values of θ_1 and θ_2 are estimated by solving the simultaneous equations

$$y_1 = \{\theta_1/(\theta_1-\theta_2)\}\{\exp(-\theta_2\xi_{11}) - \exp(-\theta_1\xi_{11})\},$$

$$y_2 = \{\theta_1/(\theta_1-\theta_2)\}\{\exp(-\theta_2\xi_{12}) - \exp(-\theta_1\xi_{12})\},$$

where y_1 and y_2 are the observed yield values. The variances of these estimates can be obtained from (4) or from the standard propagation of errors formulae (Bevington (1969, p.59)). For example, the variance of θ_1 is

$$\sigma^2_{\theta_1} \approx (\partial\theta_1/\partial y_1)^2 \, \sigma^2_{y_1} + (\partial\theta_1/\partial y_2)^2 \, \sigma^2_{y_2}. \tag{9}$$

Under the assumptions in (2) and (3), this gives $\sigma^2_{y_1} = \sigma^2_{y_2} = \sigma^2$, and some straight forward algebra verifies that (9) yields the same result as (4).

Another approach to nonlinear design considers these variances as the relevant design parameters (see e.g., Hernandez (1978), Kovanic (1970), Passi (1977)). Figures (2) and (3) show the contours of $\sigma^2/\sigma^2_{\theta_1}$, and $\sigma^2/\sigma^2_{\theta_2}$, respectively over the same range of times as Fig. 1.

The maxima (corresponding to minimum variances) occur at

$$\xi_{11} = 7.8, \quad \xi_{12} = 1.15, \quad \text{(Fig.2)} \tag{10a}$$

and

$$\xi_{11} = 5.9, \quad \xi_{12} = 3.45. \quad \text{(Fig.3)} \tag{10b}$$

Clearly, if only θ_2 is to be estimated, the design in (10b) is better than that in (8) selected by the Box and Lucas approach. Similarly, the design in (10a) is better if θ_1 alone is to be found. Even if both parameters are to be estimated, one may be more important than the other. No allowance for this possibility is made in the Box and Lucas approach. What, then, justifies the "optimal" nature of (8)?

Some Other Criteria

According to Day (1969, p.110) the usual procedure in linear design is to choose a design in which the various parameter estimates are uncorrelated.

The correlation coefficient for the two rates θ_1 and θ_2 is given by

$$C = S_{12} / \sqrt{S_{11} S_{12}},$$

and the contours of this quantity are shown in Fig. 4. It is seen that there exists a range of different uncorrelated designs ($C = 0$) for this example, although nonlinear problems do not always have them (Day (1969,p.11)). Whatever the advantages of such designs may be, in this case, Figs. 2, 3, and 4 show that these do not include minimum parameter variances.

Lees (1970) discusses the use of a different quantity for non-linear design. The matrix $\underline{W}$ is modified according to

$$\underline{W}^* = \underline{C}\,\underline{W}\,\underline{C},$$

where the diagonal matrix $\underline{C}$ is given by

$$\underline{C} = \{C_{rs}\},$$

$$C_{rr} = 1/\sqrt{W_{rr}}\ ,$$

and $C_{rs} = 0\ ,\quad r \neq s.$

$\underline{W}^*$ is then diagonalized and the value of the smallest eigenvalue examined. According to Lees (1970,p.127) "...a small λ_i^* [the eigenvalue] definitely indicates a near linear dependence [among the unknown parameter estimates]." Such a dependence indicates a poor design due to the difficulties in numerically inverting the data to retrieve the parameter estimates.

Figure 5 shows the contours of the smallest eigenvalue of $\underline{W}^*$ for the decay example. While this quantity may be useful in deciding which designs are worst, it is unclear how it could be used to decide which one is best.

Day (1969,p.110) has suggested several other quantities, among them the spectral condition number of $\underline{W}$ (similar to Fig. 5) and the trace of $\underline{W}$ (shown in Fig. 6), which may be better than $|\underline{W}|$.

It seems to us, however, that the problem lies not in deciding which of the scalar measures in Figs. 2-6 to use in place of $|\underline{W}|$, but rather in determining whether <u>any</u> scalar measure is adequate. Figures 2 and 3 offer convincing evidence that no scalar is.

As Rivkin (1968,p.230) has noted, this same problem arises in approaches which use scalar Communication Theory measures such as the entropy.

The crucial question which needs to be answered is: what type of mathematical object (scalar, vector, etc.) should be used to represent the information in data? Most approaches to nonlinear design have assumed that a scalar should be used, but this is unacceptable when multiple parameters are involved.

Exploring the entire range of possible types would be a formidable task, fortunately, however, there already exists an alternate approach which uses a nonscalar measure whose properties have long been known. The basics of this were outlined by Fisher (1939) prior to the work of Box and Lucas or even the pioneering work of Shannon and Wiener in Communication Theory. As extended by Good, Kullback, and others, this approach defines an information measure which, for a given set of data, is a mathematical operator that takes as input two hypotheses H_1 and H_2 about the nature of the physical system being studied and produces as output a number. This number measures the amount of information in the data for discriminating between H_1 and H_2.

Details of the Discrimination Information

Most of this section is taken from Kullback (1959).

Let each hypothesis H_i, $i = 1,2$ be such that it defines a probability space (X, L, μ_i) composed of a set of elements $x \in X$ (called the sample space), a collection L of all sets of elements from X, and a probability measure μ_i over the elements of L. Assume that μ_1 and

μ_2 are absolutely continuous with each other and with a third measure λ. (Two measures α and β are absolutely continuous with each other, written as $\alpha \equiv \beta$, when there exists no set $E \in L$ such that $\alpha(E) = 0$ and $\beta(E) \neq 0$ or $\beta(E) = 0$ and $\alpha(E) \neq 0$.)

By the Radon-Nikodyn theorem, each hypothesis H_i has associated with it a generalized probability density $f_i(x)$ such that

$$\mu_i(E) = \int_E f_i(x) d\lambda(x)$$

for all $E \in L$ except for sets of measure zero in λ.

If each element x represents a possible set of observed responses $\eta_1,\ldots,\eta_n$ for some experiment, then $f_i(x)$ is the probability density of observing the particular set x given that H_i is true. When $f_i(x)$ is considered as a probability over hypotheses with x fixed, it is referred to as a likelihood (Box and Tiao (1973,p.10)).

Kullback (1959,pp.4-5) defines the quantity

$$I(1:2;x) \equiv \log \left\{ \frac{f_1(x)}{f_2(x)} \right\} \tag{11}$$

as the information in x for discrimination in favor of H_1 against H_2. From Bayes' theorem (Box and Tiao (1973,p.10)) (11) is equivalent to

$$I(1:2;x) = \log \left\{ \frac{P(H_1|x)}{P(H_2|x)} \right\} - \log \left\{ \frac{P(H_1)}{P(H_2)} \right\} , \tag{12}$$

where $P(H_i)$ and $P(H_i|x)$ are the prior and posterior probabilities of H_i. This quantity measures the relative change in the probabilities of H_1 and H_2 as a result of observing x. (See Osteyee and Good (1974, pp.1-17) for a discussion of the relationship between I(1:2;x) and the information measures of Communication Theory.) The expected value

of $I(1:2;x)$, given that H_1 is true, is

$$I(1:2;X) = I(1:2) = \int_X f_1(x) I(1:2;x) d\lambda(x)$$

$$= \int_X \log\left\{\frac{P(H_1|x)}{P(H_2|x)}\right\} \quad d\mu_1(x) - \log\left\{\frac{P(H_1)}{P(H_2)}\right\} . \tag{13}$$

Although $I(1:2;x)$ and $I(1:2)$ can be expressed in terms of the prior and posterior probabilities, they are actually invariant with respect to changes in these probabilities since they only depend on the likelihoods. This is not true for the entropy or the semantic measures discussed by Hilpinen (1970). Commonly there are two different interpretations given to $P(H_i)$ in communication problems (Jamison (1970,p.29); Box and Tiao (1973,pp.12-18)). One is in terms of frequencies of occurrences for H_i among all the different hypotheses. These frequencies are usually much more difficult to determine than the likelihoods for most nonlinear experiments. A second interpretation is in terms of degree of belief (Box and Tiao (1973,p.14)), in which case the entropy of a set of probabilities becomes a function of the observer's belief system rather than of the experimental situation. While in some instances such a dependence may be desirable, our conception of the present problem requires that the information be strictly a property of the experiment and the data it produces. This is the Relativity Principle of Belief.

Connection with Fisher Information Matrix

Let H be a homogeneous set (or space) of hypotheses, i.e. H_i, $H_j \epsilon H$ implies $\mu_i \equiv \mu_j$. Then the measure $I(1:2)$ is called (Kullback (1959,p.7)) a directed divergence in this space and performs much

the same function as a metric or distance. Suppose that *H* is also parametric with each member represented by a vector $\underline{\theta}$ in some open convex set in a p-dimensional Euclidean space and that each $f_i(x) = f(x;\underline{\theta})$ is of the same functional form. Under certain regularity conditions, $I(\underline{\theta}:\underline{\theta}+\underline{\Delta\theta})$ can be approximated by a linear operator L_{op} for $\underline{\Delta\theta}$ sufficiently small (Kullback (1959,pp.26-28)), giving

$$I(\underline{\theta}:\underline{\theta}+\underline{\Delta\theta}) \backsimeq L_{op}(\underline{\theta},\underline{\theta}+\underline{\Delta\theta}). \tag{14}$$

Each coordinate representation for *H* produces a matrix representation of L_{op}, and Kullback shows that this matrix is simply ½ times $\underline{W}$, the Fisher information matrix. I(1:2) under these conditions is therefore similar to the metric of a curved space or the nertia of an extended object, and all the scalar measures introduced above ($|\underline{W}|$, C, etc.) merely represent various aspects of this nonscalar measure. (Figure 7 shows the contours of $I(\underline{\theta}:\underline{0}) = \underline{\theta}'\underline{F}'\underline{F}\,\underline{\theta}$, with $\sigma^2 = 1$ for the decay example.)

It should be noted, however, that the information is an operator whose expected value I(1:2) for two hypotheses is invariant with respect to coordinate transformations in the space used to specify the hypotheses. The Fisher information matrix $\underline{W}$, on the other hand, changes for different coordinate representations.

Information "Components"

Another set of useful scalar measures are the quantities

$$I_r \equiv \hat{\theta}_r^2 \,/\, \sigma_{\theta r}^2 \qquad r = 1,\ldots,p \tag{15}$$

which measure the components* of information in the data associated with the p parameters. Each one is a valid information measure which gives the amount of information for discriminating between two hypotheses whose associated probabilities are normal distributions with variances $\sigma^2_{\theta_r}$ and means $\hat{\theta}_r$ and 0, respectively (Niple (1978)). Contour plots of these quantities for the two decay rates θ_1 and θ_2 are similar to Figs. 2 and 3 (when σ^2 is independent of ξ_{11} and ξ_{12}) except the scales are multiplied by constants. The positions of the maximum information components about θ_1 and θ_2 are thus given in (10). Maximum information component corresponds to minimum percent uncertainty $\%\theta_r$, where

$$\%\theta_r \equiv \left\{\sigma_{\theta_r} / \theta_r\right\} \times 100\%.$$

Discussion

Many approaches to the problem of nonlinear design suffer from overly simplified assumptions about the mathematical complexity of the quantity used to represent the information in experimental data. The example of a two step decay process has illustrated this. No one optimum design suffices for this experiment because different experimenters performing the same experiment may be interested in different aspects of the decay. For this reason, the discrimination information measure, which is nonscalar, is a more realistic choice. It incorporates many different aspects of the information, and individual experimenters can then select those aspects most relevant to their purposes when

*The word "components" is perhaps misleading since it seems to imply that the information is a vector. The word "pieces", introduced by dePristo and Robitz (1978) in a somewhat different context may be more appropriate.

designing their experiments. Thus the question: which design yields the most information? is ill-posed since the word "most" implies a scalar measure. A better question is: what design yields the *best* information? where now the word "best" allows for the multifaceted nature of the information in a set of data.

The discrimination information provides a rigorous and general formalization of the nonlinear design problem which includes as specific cases the other approaches discussed above.

The insights gained by this approach to design should allow a number of previously intractable problems to be analyzed. The design problem for isolated absorption lines has already been analyzed (Niple (1978)) and solutions to additional spectroscopy problems will be given elsewhere.

The expected information for a particular design can be a function of the unknown parameters $\underline{\theta}$ as well as the design matrix $\underline{D}$. Because of this, estimation of the information depends also on estimation of $\underline{\theta}$. A complete treatment of the resulting determination of an optimum design will involve prior probabilities and utilities in the usual way (Rosenkrantz (1970)). At an informal level, a knowledge of the coarse variation of the information components together with some of the other scalar measures should be sufficient for most applications, although situations in which the cost of the designs is a rapidly varying function of the design parameters may require a more involved treatment.

List of References

1. Bevington, P.R. (1969), Data Reduction and Error Analysis for the Physical Sciences, New York: McGraw-Hill.
2. Box, G.E.P., and Lucas, H.L. (1959), "Design of Experiments in Nonlinear Situations", Biometrika 46, 77-90.
3. Box, G.E.P., and Tiao, G.C. (1973), Bayesian Inference in Statistical Analysis, Reading: Addison-Wesley.
4. Box, M.J. (1970), "Some Experiences with a Nonlinear Experimental Design Criterion", Technometrics 13, 19-31.
5. Brillouin, L. (1956), Science and Information Theory, New York: Academic Press.
6. Day, N.E. (1969), "Statistical Considerations for Nonlinear Least Squares Problems", in Least Squares Methods in Data Analysis, R.S. Anderson and M.R. Osborne (editors), 1969, St. Lucia: University of Queensland Press.
7. dePristo, A.E. and Rabitz, H. (1978), "On the Use of Various Scaling Theories in the Deconvolution of Rotational Relaxation Data: Application to Pressure-Broadened Linewidth Measurements", J. Chem. Phys. 69, 902.
8. Draper, N.R. and Hunter, W.G. (1966), "Design of Experiments for Parameter Estimation in Multi-response Situations", Biometrika 53, 525-533.
9. Fisher, R.A. (1939), The Design of Experiments, London: Oliver and Boyd.
10. Hermandez, G. (1978), "Analytical Description of a Fabry-Perot Spectrometer 4: Signal Noise Limitations in Data Retrieval: Winds, Temperature and Emission Rate", Appl. Opt. 17, 2967-2972.
11. Hilpinen, R. (1970), "On the Information Provided by Observations", in Information and Inference, J. Hintikka and P. Suppes (editors), 1970, Dordrecht-Holland: D. Reidel Publishing Co.
12. Jamison, D. (1970). "Bayesian Information Usage", in Information and Inference, J. Hintikka and P. Suppes (editors). 1970, Dordrecht-Holland: D. Reidel Publishing Co.

13. Kiefer, J. and Wolfowitz, J. (1959), "Optimum Designs in Regression Problems", Ann. Math. Statist. 30, 271-294.
14. Kovanic, P. (1970), "Accuracy Limitations in an Unbiased Optimum Data Treatment", J. Comp. Phys. 6, 473-482.
15. Kullback, S. (1959), Information Theory and Statistics, New York: John Wiley and Sons.
16. Lees, R.M. (1970), "Linear Dependence in Least Squares Analysis, with an Application to Internal Rotation", J. Mol. Spec. 33, 124-136.
17. Niple, E.R. (1978), "On the Information Content of a Spectrum", Ph.D. thesis, The Ohio State University.
18. Osteyee, D.B. and Good, I.J. (1974), Information, Weight of Evidence, the Singularity between Probability Measures and Signal Detection, no. 376 in Lecture Notes in Mathematics, A. Dold and B. Eckmann (editors), 1974, Berlin: Springer-Verlag.
19. Passi, R.M. (1977), "Use of Nonlinear Least Squares in Meteorological Applications", J. of Applied Meteorology 16, 827-832.
20. Rivkin, A.S. (1968), "Application of the Information Theory Concept to Measurement-Technique Problems", Measurement Techniques no. 2, 229-232, translated from Izmeritel'naya Tekhnika, no. 2, Feb. 1978, 63-65.
21. Rosenkrantz, R. (1970), "Experimentation as Communication with Nature", in Information and Inference, J. Hintikka and P. Suppes (editors), 1970, Dordrecht-Holland: D. Reidel Publishing Co.
22. St. John, R.C. and Draper, N.R. (1975), "D-Optimality for Regression Designs: A Review", Technometrics 17, 15-23.

Figures

Figure 1 Contours of $|\underline{F}'\underline{F}|$ with $\theta_1 = 0.7$, $\theta_2 = 0.2$ as a function of the observation times $\xi_{11} \geqq \xi_{12}$.

Figure 2 Contours of $\sigma^2/\sigma^2_{\theta_1}$ with $\theta_1 = 0.7$, $\theta_2 = 0.2$ as a function of the observation times $\xi_{11} \geqq \xi_{12}$.

Figure 3 Contours of $\sigma^2/\sigma^2_{\theta_2}$ with $\theta_1 = 0.7$, $\theta_2 = 0.2$ as a function of the observation times $\xi_{11} \geqq \xi_{12}$.

Figure 4 Contours of C, the correlation coefficient for the rate constants with $\theta_1 = 0.7$, $\theta_2 = 0.2$ as a function of the observation times $\xi_{11} \geqq \xi_{12}$. The lines with $C = 0.0$ indicate uncorrelated designs.

Figure 5 Contours of $\log_{10}(\lambda^*_{min})$ with $\theta_1 = 0.7$, $\theta_2 = 0.2$ as a function of the observation times $\xi_{11} \geqq \xi_{12}$.

Figure 6. Contour of the trace of $\underline{F}'\underline{F}$ with $\theta_1 = 0.7$, $\theta_2 = 0.2$ as a function of the observation times $\xi_{11} \geqq \xi_{12}$.

Figure 7. Contour of $I(\underline{\theta}:\underline{0})$ with $\theta_1 = 0.7$, $\theta_2 = 0.2$, $\sigma^2 = 1$ as a function of the observation times $\xi_{11} \geqq \xi_{12}$.

FIGURE 1.

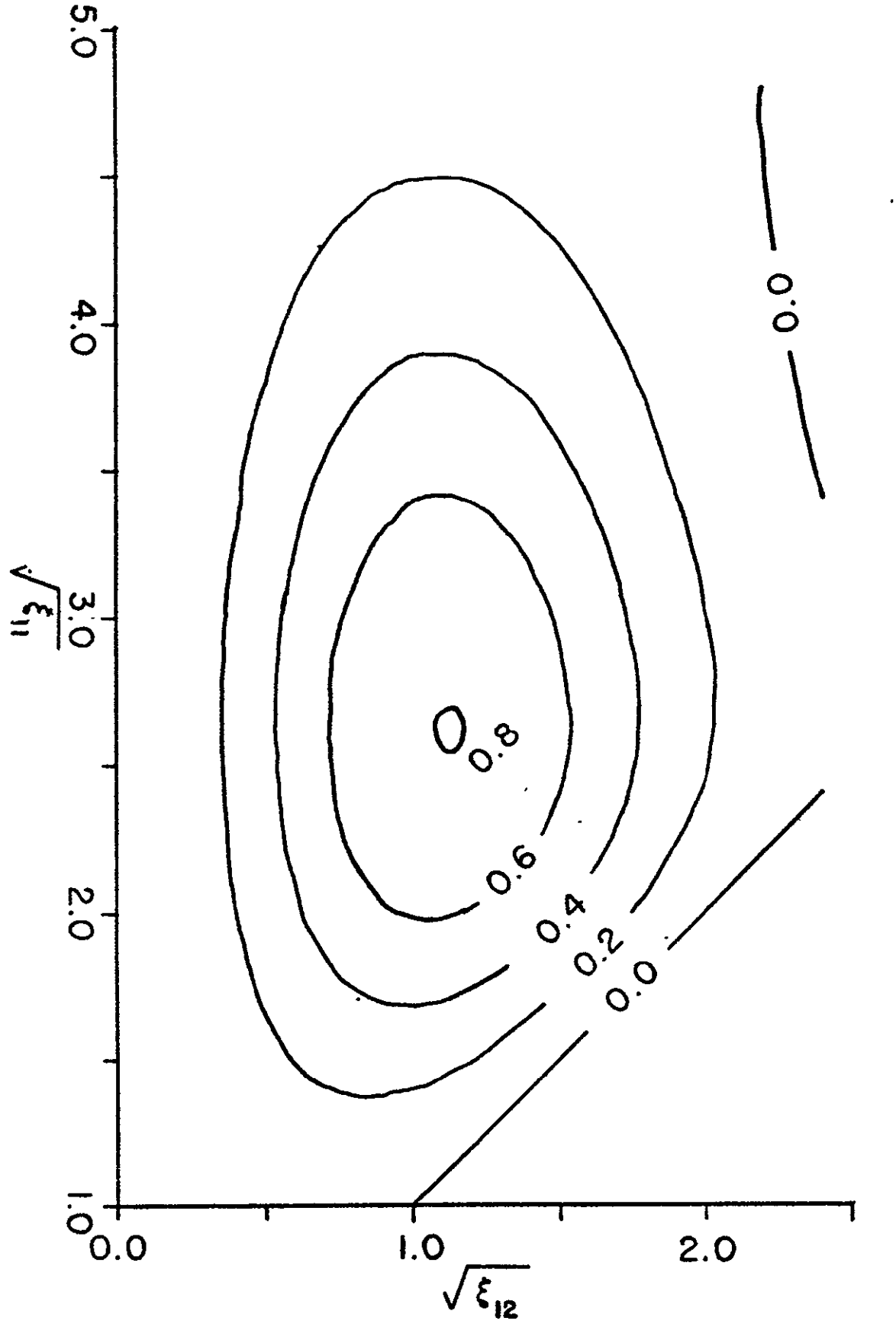


FIGURE 2.

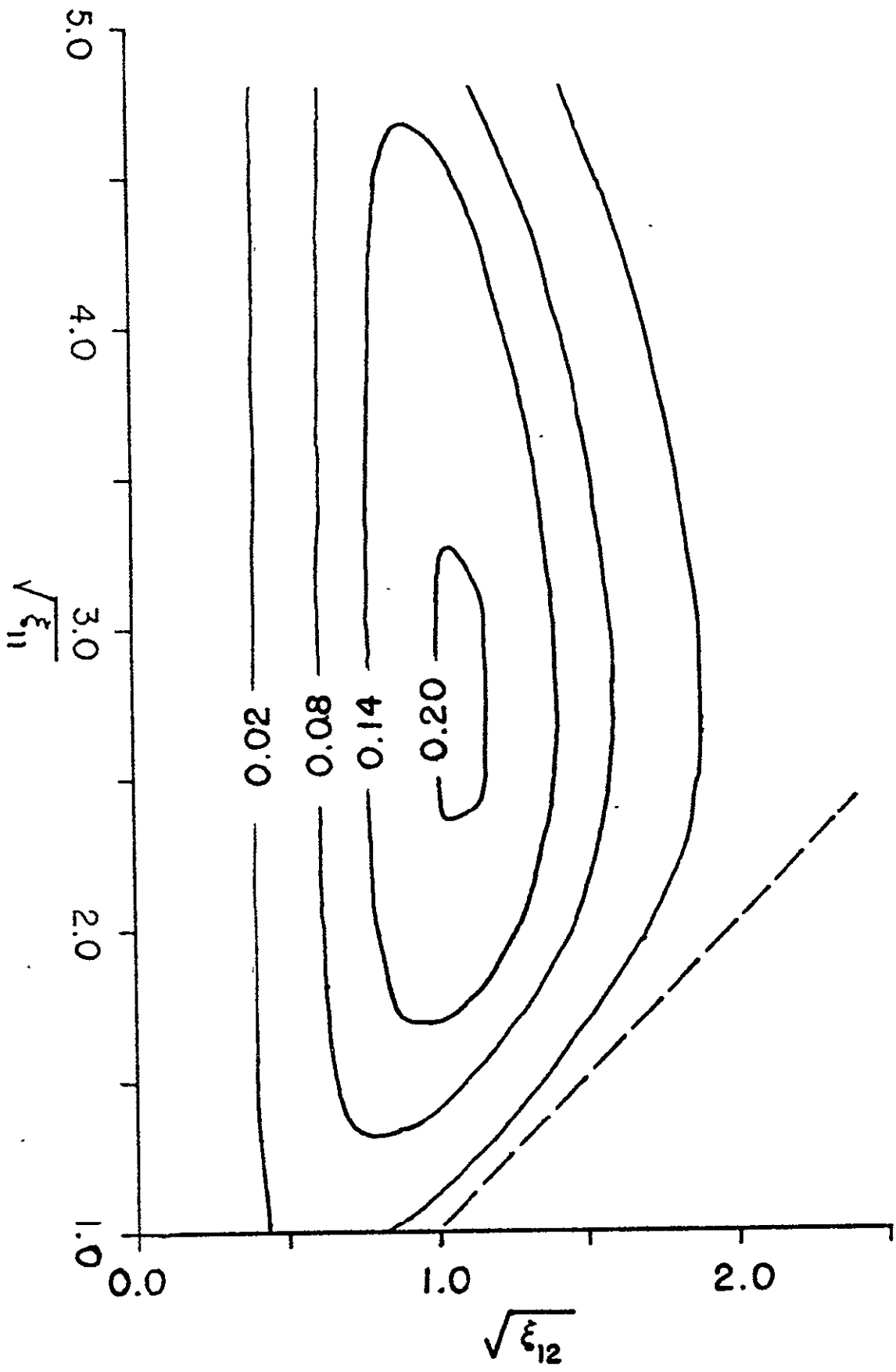


FIGURE 3.

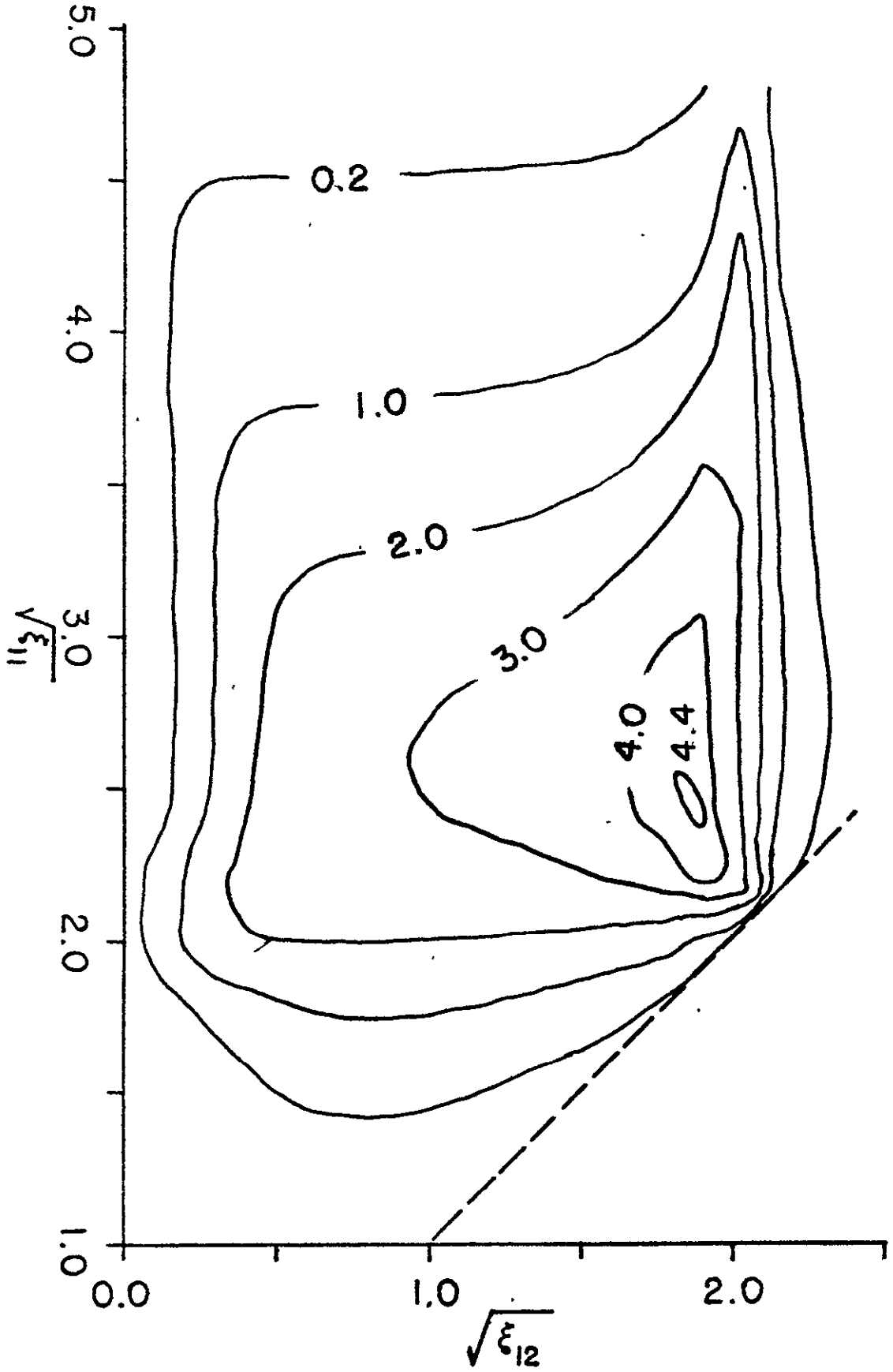

5.0
4.0
3.0
2.0
1.0
$\frac{\eta_3}{\sqrt{\xi_{11}}}$
0.2
1.0
2.0
3.0
4.0
4.4
0.0
1.0
2.0
$\sqrt{\xi_{12}}$

FIGURE 4.

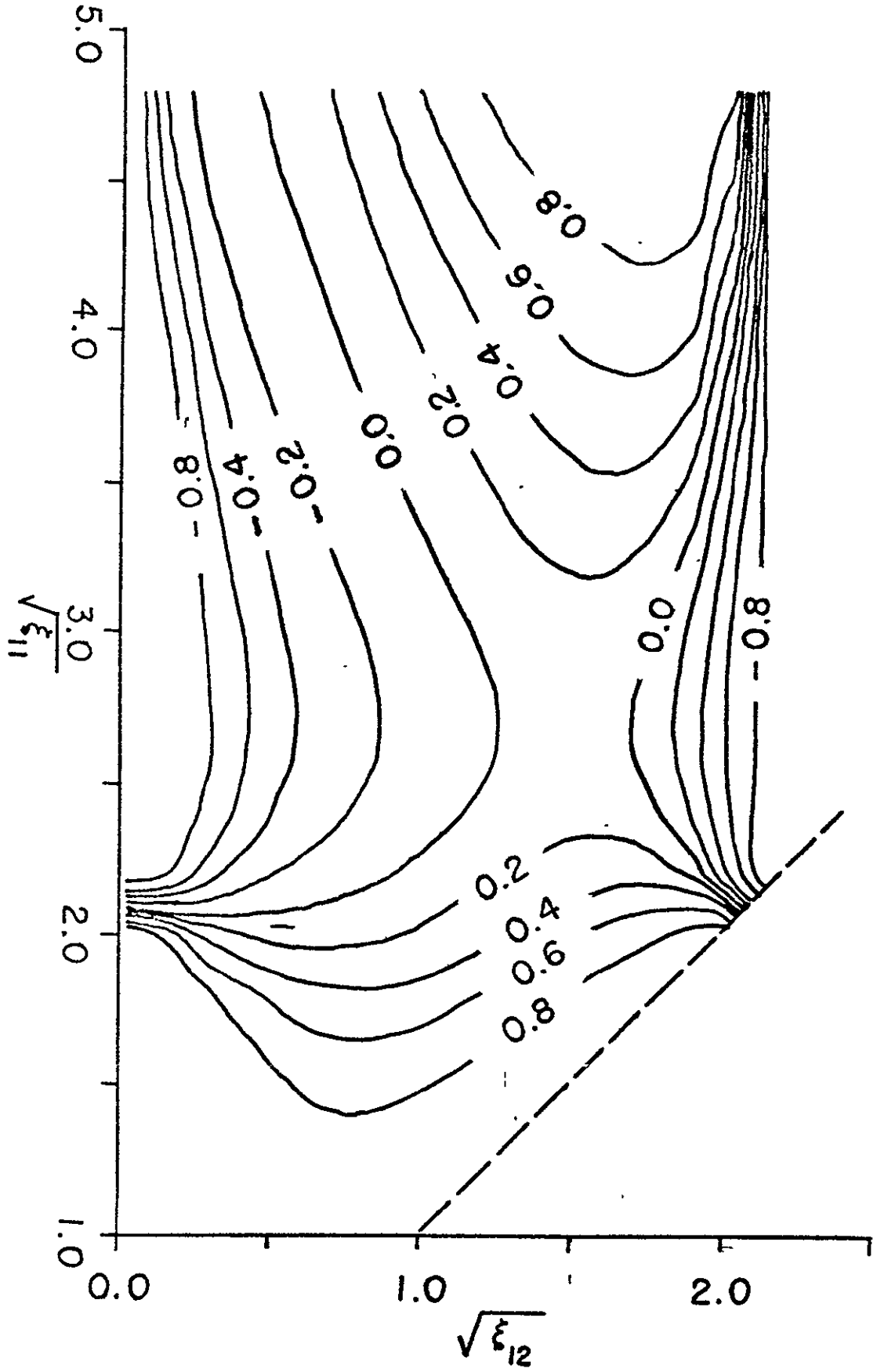

5.0
4.0
3.0
2.0
1.0
0.0
1.0
2.0
$\sqrt{\xi_{11}}$
$\sqrt{\xi_{12}}$
8.0
9.0
4.0
2.0
0.0
-0.4
-0.8
0.0
-0.8
0.2
0.4
0.6
0.8

FIGURE 5.

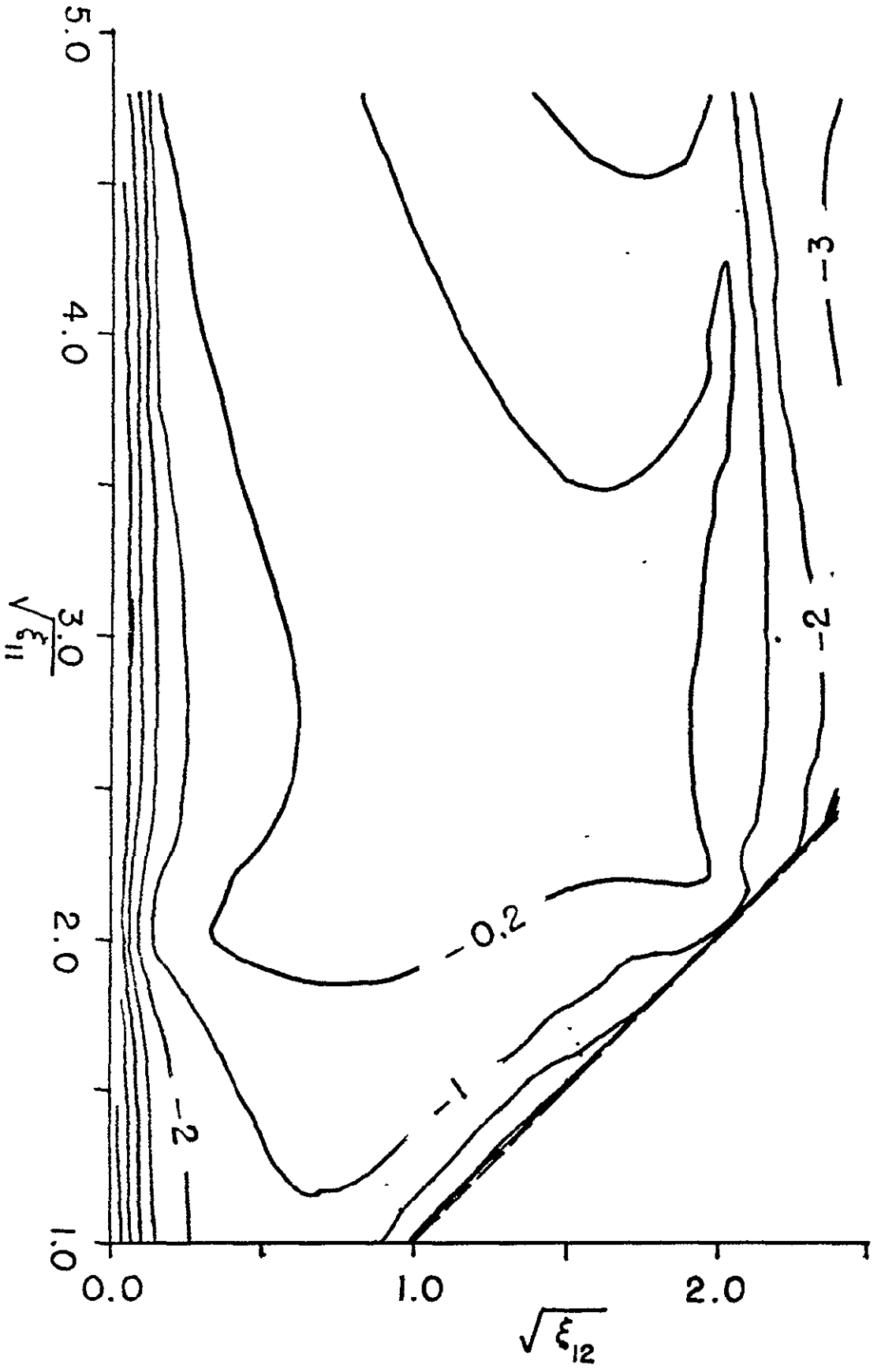


FIGURE 6.

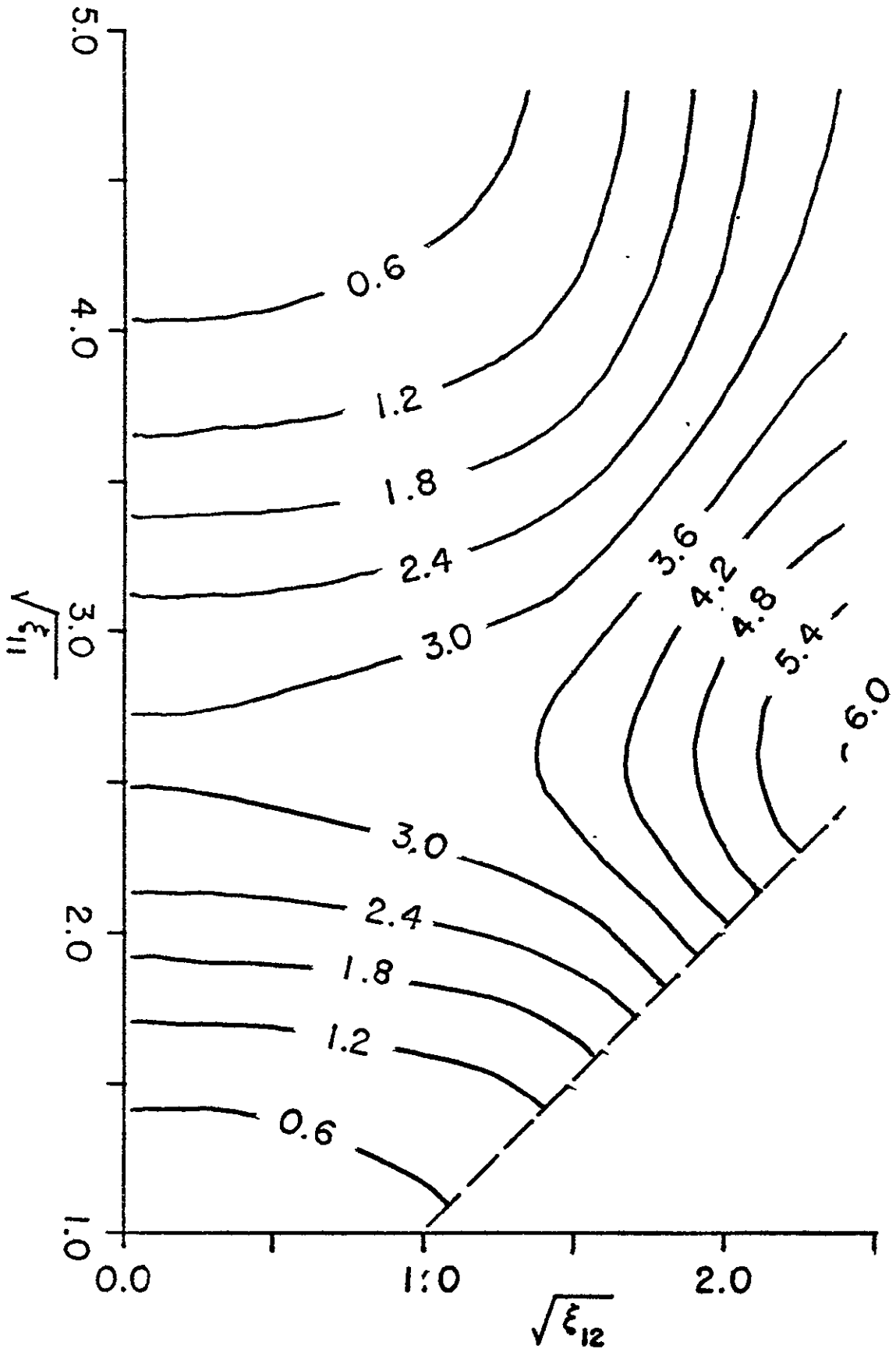
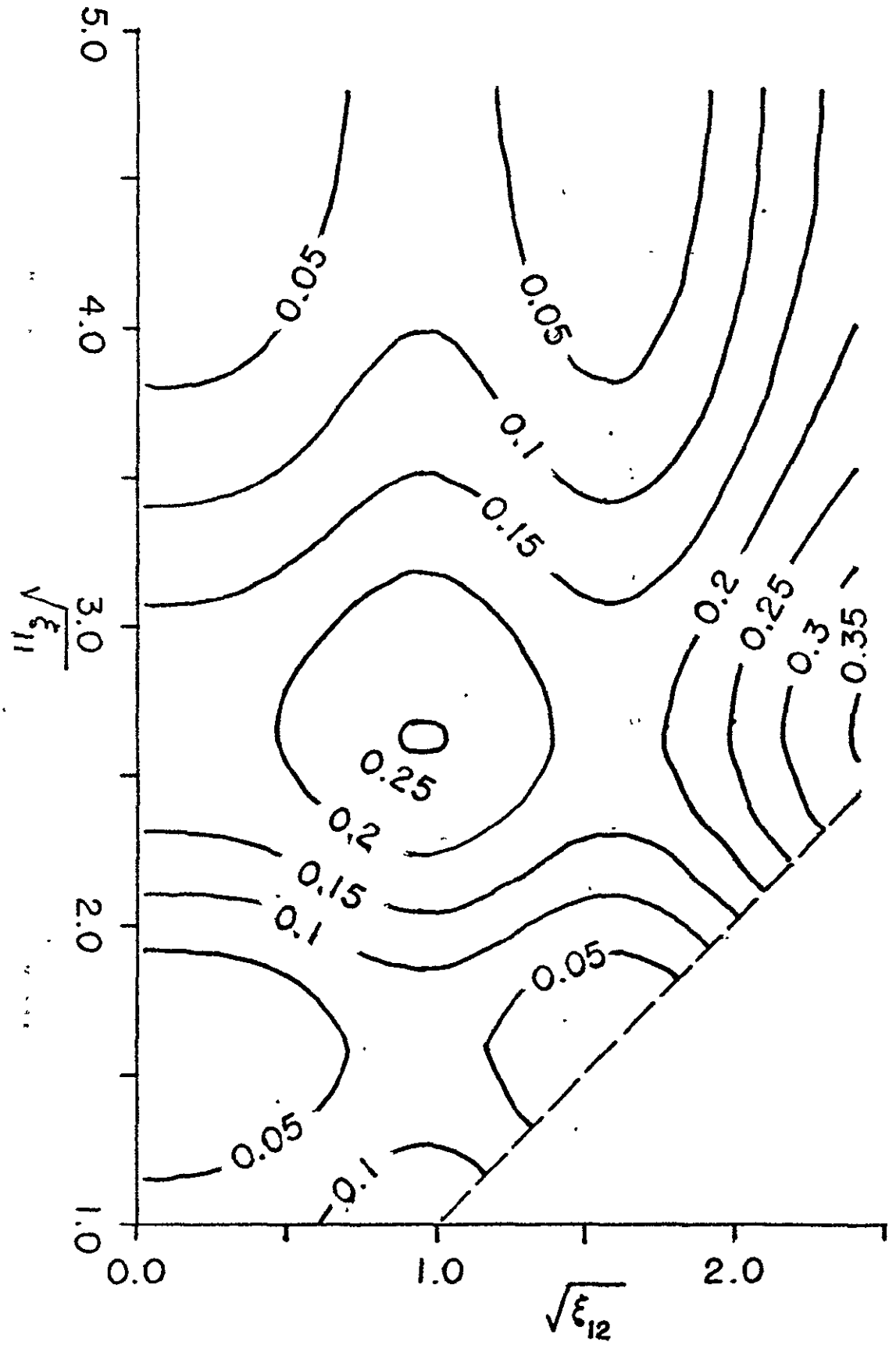

5.0
4.0
3.0
2.0
1.0
0.0
1:0
2.0
$\sqrt{\xi_{12}}$
$\sqrt{\xi_{11}}$
0.6
1.2
1.8
2.4
3.0
3.6
4.2
4.8
5.4
6.0
3.0
2.4
1.8
1.2
0.6

FIGURE 7.


5.0
4.0
3.0
2.0
1.0
0.0
1.0
2.0
$\sqrt{\xi_{11}}$
$\sqrt{\xi_{12}}$
0.05
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0
0.25
0.2
0.15
0.1
0.05
0.05
0.1

INFORMATION IN SPECTRA OF COLLISION BROADENED ABSORPTION LINES*

E. Niple
J. H. Shaw

Department of Physics
The Ohio State University
Columbus, Ohio 43210

*This work supported in part by NASA Grant No. NS9-7479.

ABSTRACT

A mathematical model is introduced for experiments which yield spectra of isolated, collision broadened absorption lines. A nonlinear design analysis is then performed on the model to identify the optimum experimental conditions. The information in the spectra is split into components associated with each of the unknown parameters to be estimated from the spectra. By exploring the variation of these components it is shown that spectra with line center transmittances of 0.25 are nearly optimum for measuring the line position, the Lorentz width α, the instument resolution H, and the product of line intensity and absorber amount. The trade off between signal to noise ratio (SNR) and resolution is investigated for several different resolution dependent noise levels. The analysis shows that when SNR is proportional to the square root of the ratio of resolution and Lorentz width, a single optimum resolution and line center transmittance combination of $H/\alpha \cong 1.5$ and 0.28, respectively, exists.

Introduction

Qualitatively there are two kinds of data: good data and bad data. The goal of Experimental Design is to determine, ahead of time, which kind a particular experiment would be likely to yield. The influence of the signal-to-noise ratio (SNR) on the quality of experimental data has long been recognized, yet other effects may also be significant. In spectroscopic experiments the spectral resolution is equally as important as SNR, and the data quality is often determined by a trade-off between these two. For the purposes of experimental design, therefore, it is important to have a quantitative criterion which describes all the factors which influence the outcome.

Such a multi-faceted criterion was discussed in a previous paper[1] (referred to as NS-1). It is the discrimination information measure developed by Fisher[2], Good[3], Kullback[4] and others. In the present paper, the information components, introduced in NS-1, are used to analyze experiments in which spectra of an isolated, pressure-broadened, (Lorentzian), absorption line obtained by instruments of finite resolution are collected. Such spectra have been analyzed previously e.g. Chang et al[5]. Although qualitative descriptions of the type of spectral data most suitable for analysis were given, no quantitative investigation of the experimental design was made.

It is hoped that the results presented below will lead to a better understanding of the power of the discrimination information measure and, more generally, to improved techniques of experimental design in Spectroscopy.

Pressure Broadened Lines

Consider an experiment which measures the intensity of radiation transmitted by an absorbing gas sample at a uniform temperature and total pressure P_T. When P_T is sufficiently high, the line shape is determined by collision broadening,[6] and is Lorentzian. The signals y_i are recorded at N equally spaced frequencies ν_i from some lower frequency ν_L to an upper frequency

$$\nu_U = \nu_L + (N-1)\Delta\nu, \tag{1}$$

where $\Delta\nu$ is the constant frequency spacing. If the noise in the signal has a normal distribution with rms value σ, and is independent of frequency, the form of y in the vicinity of an isolated absorption line is described by

$$y_i = \epsilon_i + B(\nu_i)\int_0^\infty R(\nu', \nu_i; H)\exp -Su\alpha_o P_T/\pi\left\{(\nu' - \nu_o)^2 + \alpha_o^2 P_T^2\right\} d\nu' \tag{2}$$
$$i = 1, \ldots, N.$$

In this equation $B(\nu_i)$ represents the background signal with the absorber removed and is assumed to be of the form

$$B(\nu_i) = a + b(\nu_i - \langle\nu\rangle) \tag{3}$$

with $b \simeq 0$ and $\langle\nu\rangle$ the mean frequency. This gives a background nearly independent of frequency. $R(\nu, \nu_i; H)$ is the instrumental spectral response function and is assumed to be of the form

$$R(\nu', \nu_i; H) = (1 - |\nu' - \nu_i|/H)/H, \qquad |\nu' - \nu_i| \leq H,$$
$$= 0, \qquad |\nu' - \nu_i| > H. \tag{4}$$

This form is triangular with full width H at half height. ϵ_i represents the random noise.

Equation (2), which is called the model for the spectrum, assumes that the combined effects of foreign and self broadening can be represented by the single term $\alpha_o P_T$. The line parameters S and α_o are the line intensity and the pressure broadening coefficients respectively, ν_o is the line center frequency, and u, the absorber amount, is the product of radiation path length and absorber partial pressure.

A typical spectrum calculated from the model is shown in Fig. 1. Also shown in this figure are two parameters W_{obs} and t_p that are used to describe absorption features in general. The width of the line W_{obs} at the half peak absorption points is a complex function of the instrumental response function width H and the line parameters. The transmittance at line center is defined as t_p.

There are twelve parameters involved in the model: the three line parameters: ν_o, S, and α_o, and nine physical parameters : ν_L, ν_U, N, a, b, H, σ^2, u and P_T. There are two types of experiments which require the collection and analysis of spectra described by this model. In one, the line parameters α_o and S are known and the physical parameters u and P_T are retrieved from the spectrum as well as a, b, ν_o, H, and σ^2. In the other, P_T and u are known and the line parameters are retrieved instead. In both cases ν_L, ν_U and N are also known, and the goal is to retrieve the remaining seven unknowns. One way to effect this retrieval is by nonlinear Least Squares Regression Analysis, which is summarized in the next section.

The Least Squares Variance-Covariance Matrix and Information Components

When a set of observations $\{y_i\}$ is described by the model

$$y_i = \epsilon_i + f(\xi_1^{(i)}, \ldots, \xi_k^{(i)}; \theta_1, \ldots, \theta_p) \quad i = 1, \ldots, N, \tag{5}$$

where f is some function of k known variables $\xi_1, \ldots, \xi_k$ and p unknown parameters $\theta_1, \ldots, \theta_p$, and ϵ_i is a normally distributed random variable with variance σ^2 and mean zero, the asymptotic variance-covariance matrix of the Least Squares estimates $\hat{\theta}_i, \ldots, \hat{\theta}_p$ for the parameters is given by [7]

$$\underline{\hat{S}} = \hat{\sigma}^2 (\underline{J}' \underline{J})^{-1}. \tag{6}$$

In this, $\hat{\sigma}^2$ is the estimated noise variance and $\underline{J}$ is the Jacobian matrix of f with

$$J_{ij} = \left. \frac{\partial f}{\partial \theta_j} \left(\xi_1^{(i)}, \ldots, \xi_k^{(i)}; \theta_1, \ldots, \theta_p \right) \right|_{\hat{\theta}_1, \ldots \hat{\theta}_p} \tag{7}$$

The diagonal element $\hat{S}_{jj}$ of $\underline{\hat{S}}$ is the asymptotic variance $\hat{\sigma}_j^2$ of $\hat{\theta}_j$ and $\hat{S}_{ij} / \hat{\sigma}_j \hat{\sigma}_i$ is the asymptotic correlation coefficient between $\hat{\theta}_j$ and $\hat{\theta}_i$. The asymptotic percent uncertainty $\% \hat{\theta}_j$ is therefore given by

$$\% \hat{\theta}_j \equiv 100 (\hat{\sigma}_j / \hat{\theta}_j). \tag{8}$$

The values of $\hat{\theta}_1, \ldots, \hat{\theta}_p$ are those which minimize the sum of square deviations

$$Q(\theta) = \sum_{i=1}^{N} \left[y_i - f(\xi_1^{(i)}, \ldots, \xi_k^{(i)}; \theta_1, \ldots, \theta_p) \right]^2 .$$

In NS-1 the quantity

$$I_j \equiv (\hat{\theta}_j / \hat{\sigma}_j)^2 \tag{9}$$

was recommended as a measure of the component or "piece" of information in a set of data about the j^{th} unknown parameter. From Eqs.(8) and (9) it is seen that I_j is large when the per cent uncertainty $\%\hat{\theta}_j$ is small and vice versa. This corresponds to the intuitive sense that, if a set of data leads to a parameter estimate with a large uncertainty for the j^{th} parameter, as reflected in a large $\%\hat{\theta}_j$, then that data set does not contain very much information about θ_j. (See NS-1 and its references for a more formal justification of the term "information" above). .

The Information Components for Equation (2)

From Eqs. (6), (7) and (10) it is seen that each I_j depends upon the $\left\{\xi_\alpha^{(\iota)}\right\}$ and σ^2 as well as the $\left\{\hat{\theta}_\beta\right\}$. For the model in Eq. (2) this means that twelve different variables can influence the information content of a spectrum. To simplify the analysis, the value of b was fixed at zero in all spectra analyzed. This is not the same as assuming that b is known before analyzing the spectrum. It can be shown[8] that the remaining eleven variables contain several redundancies (related to symmetry properties of the information) so that each of the information components for S, α_o, ν_o, a, H (or equivalently u, P_T, ν_o, a, H) can be written as

$$I_j \simeq (SNR)^2 \; N \; F_j(E, C, \Omega/\alpha, H/\alpha), \qquad (10)$$

where F_j is some complex function of the four variables E, C, Ω/α and H/α. These are defined as:

$$E \equiv (\nu_U - \nu_L)/W_{obs} \,,$$

$$SNR \equiv a/\sigma,$$

$$C \equiv (\nu_o - \nu_L)/(\nu_U - \nu_L),$$

$$\Omega \equiv \sqrt{Su\alpha/\pi},$$

and $\alpha \equiv \alpha_o P_T$.

The extent of the spectrum E is measured in observed widths. C specifies the centering of the line in the observed frequency range, and a value of 0.5 corresponds to an exactly centered line. The approximate nature of Eq. (10) is due to the N dependence. (The validity of this approximation is discussed below.)

In order to present the variation of the information components in a meaningful and easily understood manner, the values of E and C were fixed at 10 observed widths and 0.5006 respectively, and a range of values for Ω/α and H/α was selected. Representative values for SNR and N of 100 and 300 respectively were then used to generate three dimensional surfaces for each I_j. These can be shown as either two dimensional projections or contour plots.

The ranges of values for Ω/α and H/α are shown in Fig. 2. Also shown are the contours of t_p over these ranges and several examples of spectra generated from Eq. (2). These spectra were calculated for fixed values of N, E and C and five pairs of values for Ω/α and H/α . For the purpose of reference, spectra corresponding to points in the lower part of Fig. 2 correspond to weak lines (as indicated by a line center transmittance > 0.85) and those at the top to strong lines. Similarly, spectra corresponding to points at the left correspond to high resolution and those on the right to low resolution. Figure 3 shows the contours of the observed width W_{obs} over the same range of values with $\alpha = 0.07084 cm^{-1}$.

Before investigating the variation of the I_j with E and N fixed, the effects of changing each of these, one at a time, on the information content associated with spectrum C in Fig. 2 were investigated. Figure 4 shows how each of the I_j for ν_o, H, α, $\Omega = \sqrt{S_u \alpha}$ and S_u varies as a function of E. The calculations were done for 10 points / observed width, and the quantity

$$I_{\nu_o} = (W_{obs}/\sigma_{\nu_o})^2$$

was used for the line center component. Each I_j was normalized to one at E = 10 . Notice that, except for I_{ν_o}, each of the I_j continues to increase with increasing extent even though there is no significant absorption beyond about five observed widths from line center. We attribute this increase to the correlations between the background parameters a and b and the other parameters. Data points far from line center make the greatest contribution to the precision with which the background can be retrieved and thus they also indirectly contribute to the precision of parameters correlated with the background.

Figure 5 explores the N- dependence in Eq. (10). The value of each I_j/N (normalized to unity at N = 600) is shown as a function of N, again for spectrum C in Fig.2, with E = 10 . Provided N is larger than 70 , the approximation is quite good. This result also allows the definition of an average information rate which only depends on E, C, H/α , Ω/α for instruments where $(SNR)^2 \cdot N$ is proportional to the time required to collect the data. Often this is approximately true for both grating and interferometer type spectrometers.

Figure 6 shows the contours of I_{Su} over the range of spectra in Fig.2. (Recall that SNR = 100). This will be the component associated with S when u is known, and the component associated with u when S is known. These contours are labelled with the value of the percent uncertainty %Su rather than with the value of I_{Su}, since the former is more familiar. The distance between adjacent contour lines corresponds to a 1db change in % Su and hence to a 2db change in I_{Su}. This figure shows a ridge of high information content about Su for spectra with $\sqrt{Su/\pi\alpha} \simeq 1.3$ and $H/\alpha \leqslant 2$. From Fig. 2 it is seen that this ridge lies approximately along the $t_p = 0.25$ contour. Consequently the best spectra for measuring Su have a resolution on the order of, or less than, α and a line center transmittance of $\simeq 0.25$. Figure 6 also shows that little additional information is gained by improving the resolution beyond $H/\alpha \simeq 1.5$. This assumes that SNR is independent of resolution. More is said about the case where they are not independent in the discussion below.

Figure 7 shows the contours of I_α, again labelled with $\%\alpha$. These contours are similar to those of I_{Su} except that, in each case, I_α is strictly less than I_{Su}. We attribute this to the different types of information about the spectrum described by the parameters S and α. Although the effects of these parameters on the spectrum are by no means independent, nevertheless, in a crude sense, $S = \int k(\nu)\,d\nu$ is estimated from the area of the line whereas α is a shape factor. The results in Figs. 5 and 6 show that, in a given spectrum, there is more information about the equivalent width of the line than the

shape. Also, the effect of improving resolution is more marked for α than Su, as would be expected for any line shape parameter. This is seen in the steeper upward slope along the optimum ridge as H/α decreases.

Figure 8 shows the variation of the information component %H associated with the resolution parameter H. This figure is complementary to Fig.7. This is as expected since a line whose shape is determined mainly by the instrument spectral response function (H/α large) contains little information about α. Similarly a line whose shape is determined by collision broadening effects (H/α small) contains little information about H.

Figure 9 shows the variation of I_Ω. The similarity between this figure and Fig. 2 indicates that the dominant factor in determining the information about Ω is the line center transmittance t_p. This figure was obtained from the values of σ_{Su} and σ_α in the previous figures along with the correlation coefficient between Su and α according to the standard propagation-of-errors formulae.

Figure 10 shows the contours of I_{ν_o}. (These are labelled with W_{obs}/σ_{ν_o} in percent.) As with I_Ω, the t_p influence is the most important. In many cases, the quantity of interest is $\sigma^2_{\nu_o}$, the absolute uncertainty in the estimated line center. By comparing Fig. 10 with Fig. 3 of W_{obs}, it is seen that the lowest absolute line center uncertainty occurs for spectra with high resolution (H/α small) and intermediate strength ($\sqrt{Su/\pi\alpha} \simeq 1.5$). Consequently the lines on the ridge of Fig. 6 for I_{Su} are nearly optimum for I_{Su}, I_α, and $\sigma^2_{\nu_o}$.

The Case Where SNR Depends on Resolution

As stated above, the percent uncertainty values used to label the contours in Figs. 6 through 10 are based on a uniform SNR of 100 and $N = 300$ points. Eq. (10) can be used to compute the values of the I_j for other values of these quantities.

It is interesting to explore the situation where SNR is a function of the resolution H. Many different choices for the functional dependence are possible, depending upon the operating characteristics of the spectrometer used to collect the spectrum. For grating spectrometers it is often found that[9], away from the diffraction limit,

$$SNR \propto W_S^2 ,$$

where W_S is the physical slit width. The relationship between W_S and H is complex and depends on the alignment of the spectrometer, but, for narrow slit widths, H is nearly independent of W_S and, for large slit widths, H becomes linearly dependent on W_S. For large slit widths, therefore, the SNR can be expected to be proportional to H^2 and, as H decreases toward some limit H_{min}, SNR goes to zero. The noise dependence for interferometers is discussed by Flemming.[10]

It is a simple matter to substitute an appropriate function of H into Eq. (10) for SNR and calculate values for any of the information components, given values for H, N, E, C, H/α, and Ω/α, where the values of the F_j are taken directly from Figs. 6 through 10. The resulting functions of six variables could be optimized on a computer, or by trial and error, to obtain the best spectral parameters for measuring specific unknown parameters i.e. Su, α, ω, etc. However, because of the many variables involved, the results of such an

optimization are very difficult to present graphically, We have therefore chosen to assume that the interdependence of SNR and H (and N, where appropriate) can be expressed in the form

$$SNR \propto \left[H/\alpha\right]^{d}, \qquad (11)$$

where d is some positive constant. The resulting information components can now be presented in contour plots by fixing E and C and including the N- dependence as a scale factor, as was done before. Figure 11 shows the effects of a resolution dependent SNR on the contours of I_{α} in Fig. 7, for the case where d=1/2. The optimum (labelled +) occurs at $H/\alpha \simeq 1.5$, $\Omega/\alpha \simeq 1.3$, with the information falling off at 2db per contour line going away from this optimum. The corresponding locations of the optima for d = 1 and d = 3/2 are shown by the O and X, respectively. For d = 2 the optimum lies to the right of the figure. The track of the optimum, as d increases, is indicated by the dotted line, and is seen to follow the ridge of Fig. 7. Similar results were obtained for I_{Su}. Such figures are similar to the well known resolution - SNR trade-off curves of Backus and Gilbert[11], except the importance of Ω/α (or equivalently the line center transmittance) has now been included as well.

Discussion

One of the conclusions to be drawn from the contour plots in Figs. 6 through 8 is that neither spectra of very weak lines nor spectra of very strong lines contain very much information about Su, α or H. This is seen in the high percent uncertainties at the tops and bottoms of these figures. Spectra of very weak lines, especially

low resolution spectra, do not contain much information about Ω or ν_o either, but high resolution spectra of strong lines do contain considerable information about these two parameters. All the figures indicate increased information with improving resolution (except I_H), however, improving the resolution beyond $H/\alpha \simeq 1$ leads to little additional information, provided the signal to noise ratio, SNR, is independent of the resolution. When the SNR varies as some power of the relative resolution H/α, a unique optimum spectrum exists for measuring Su or α which has a larger H/α value for higher powers of H/α.

Figure 4 shows the importance of including data points in the far wings of the line in order to measure Su, α, H or Ω, but not ν_o. A similar analysis of the centering dependence (see Ref. (8)) showed that the line should be centered in the middle of the observed frequency range.

Also apparent in the figures is the importance of the line center transmittance value, t_p for determining the information content. For Su, α, H and ν_o (assuming the absolute value of the variance is used) a t_p value of 25% transmittance at highest resolution yields approximately the best information for all these parameters.

We feel that the results given above are best viewed within the Information Theory framework discussed in NS-1. In this framework, Figs. 6 through 11 are interpreted as displaying various aspects of a multidimensional mathematical operator which formally represents the information in a set of data; in the present case, a spectrum. The overall implications of the figures agree with one's intuitive feel

for the problem, as , indeed, they should. For example, the conclusion that spectra of very weak lines contain little information about any of the line parameters (as indicated by small I_{Su}, I_{α} and $I \cdot \nu_o$) is well known to anyone who has ever tried to analyze such spectra. Yet the value of a quantitative measure that expresses what one can expect to learn from a set of data should not be underestimated, especially for the case where cost analyses are being performed on several competing experimental proposals. A reliable quantitative measure greatly simplifies the difficult decision of choosing between repeated performances of inexpensive but less informative experiments, and a single trial of a more expensive but very informative one. Hopefully the techniques presented in the present work will prove useful in such instances.

REFERENCES

1. E. Niple and J. H. Shaw, J. Phys., E, (submitted for publication).

2. R. A. Fisher, The Design of Experiments, Oliver and Boyd, London (1937).

3. I. J. Good, Probability and the Weighing of Evidence, Charles Griffin, London (1950).

4. S. Kullback, Information Theory and Statistics, John Wiley and Sons, New York (1959).

5. Y. S. Chang and J. H. Shaw, Appl. Spectrosc., 31, 213 (1977).

6. W. S. Benedict, R. Herman. G. Moore and S. Silverman, Can. J. Phys., 34, 830 (1956).

7. N. Day, in Least Squares Methods in Data Analysis, R. Anderssen and M. Osborne Ed., University of Queensland Press, St. Lucia (1969) p. 105.

8. E. Niple, On the Information Content of a Spectrum, Ph. D. dissertation, The Ohio State University (1978).

9. Y. S. Chang, Line Intensities and Half Widths of the H_2O ν_2 Band near 2000cm^{-1} Obtained by Using a Least Squares Fit Method, Ph.D. dissertation, The Ohio State University (1976).

10. J. Flemming, Infrared Physics, 17, 263 (1977).

11. G. Backus and F. Gilbert, Phil. Trans. Roy. Soc., A1173, 226, 123 (1970).

Figure Captions

Fig. 1 - A typical absorption line of type described by Eq. (2).

Fig. 2 - The range of values for Ω/α and H/α used, including contours of line center transmittance t_p, and five representative spectra corresponding to points in the range.

Fig. 3 - Contours of the observed width W_{obs} (in cm^{-1} assuming $\alpha = 0.0708\ cm^{-1}$).

Fig. 4 - The dependence of the Information components on the extent of the spectrum E, for the components associated with ν_o (line center), H (resolution), α (Lorentz width), Su and Ω.

Fig. 5 - The dependence of the information components on the number of data points N, for the components associated with ν_o (line center), H (resolution), α (Lorentz width), Su and Ω.

Fig. 6 - Contours of Information component associated with Su (labelled with %Su) for N = 300, SNR = 100, E = 10 and C = 0.5006.

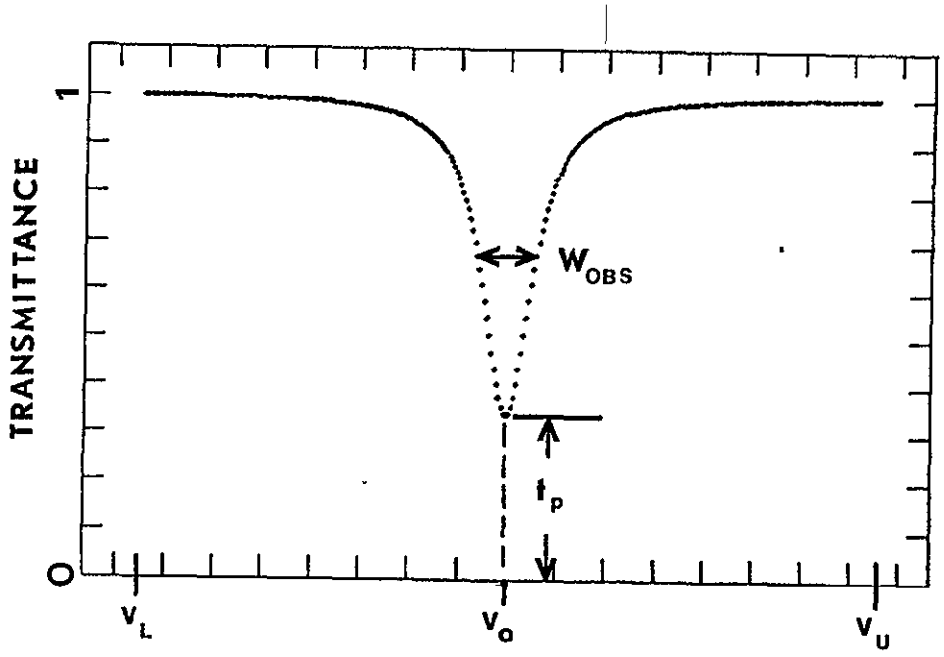
Fig. 7 - Contours of Information component associated with α (labelled with $\%\alpha$) for N = 300, SNR = 100, E = 10 and C = 0.5006.

Fig. 8 - Contours of information component associated with H (labelled with %H) for N = 300, SNR = 100, E = 10 and C = 0.5006.

Fig. 9 - Contours of information component associated with Ω (labelled with $\%\Omega$) for N = 300, SNR = 100, E = 10 and C = 0.5006.

Fig. 10 - Contours of information component associated with ν_o (labelled with σ_{ν_o}/W_{obs} x 100%) for N = 300, SNR = 100, E = 10 and C = 0.5006.

Fig. 11 - Contours of information component associated with α for N = 300, E = 10 and C = 0.5006 when SNR $\simeq \sqrt{H/\alpha}$.
+ indicates maximum. O indicates maximum when SNR $\simeq H/\alpha$.
X indicates maximum when SNR $\simeq (H/\alpha)^{3/2}$ (spacing corresponds to 2db change).

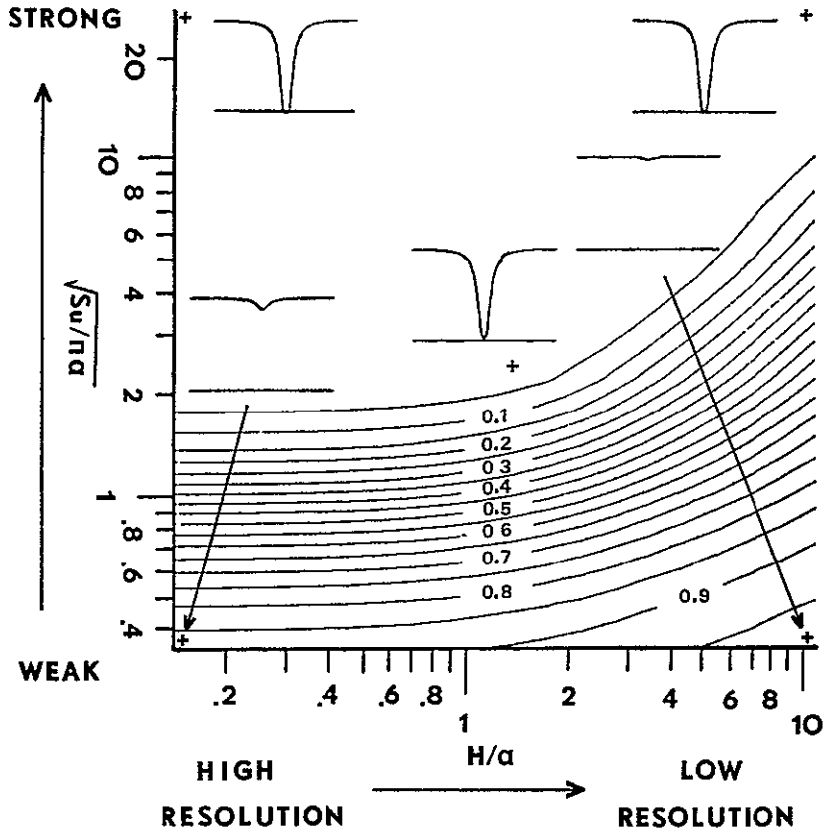


$$\text{CENTERING} \equiv [V_o - V_L]/[V_U - V_L]$$

$$\text{EXTENT} \equiv [V_U - V_L]/W_{OBS}$$

$$\Omega \equiv \sqrt{Sua/n}$$

Fig 1


STRONG
WEAK
√Su/πa
20
10
8
6
4
2
1
.8
.6
.4
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
.2
.4
.6
.8
1
2
4
6
8
10
H/a
HIGH
RESOLUTION
LOW
RESOLUTION

11

Fig 2

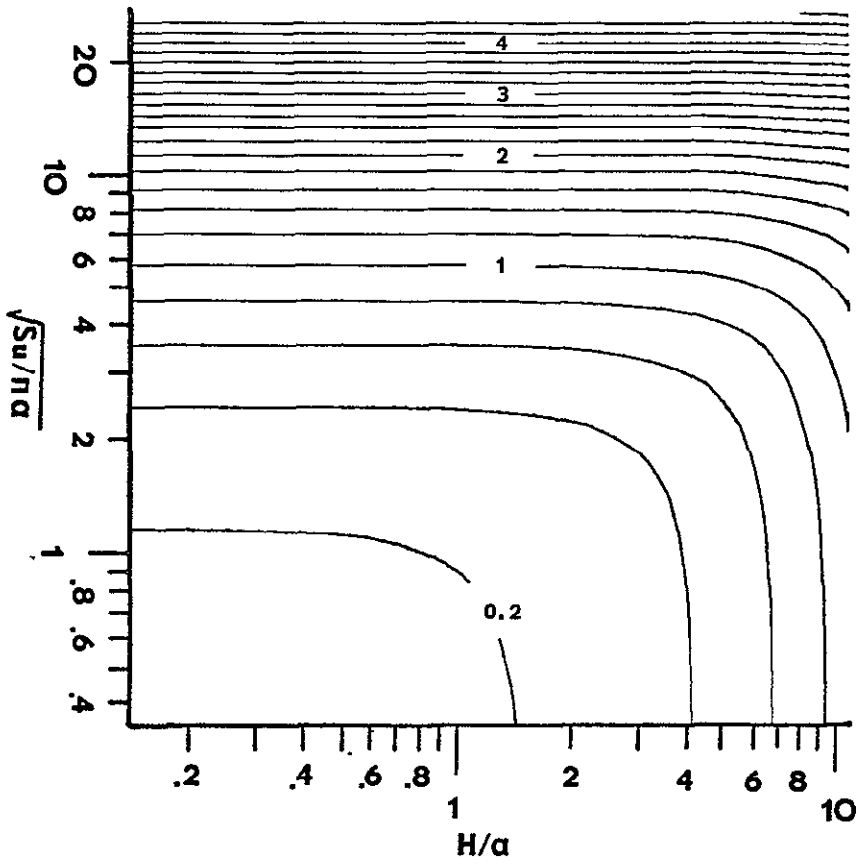


Fig 3

Fig 4

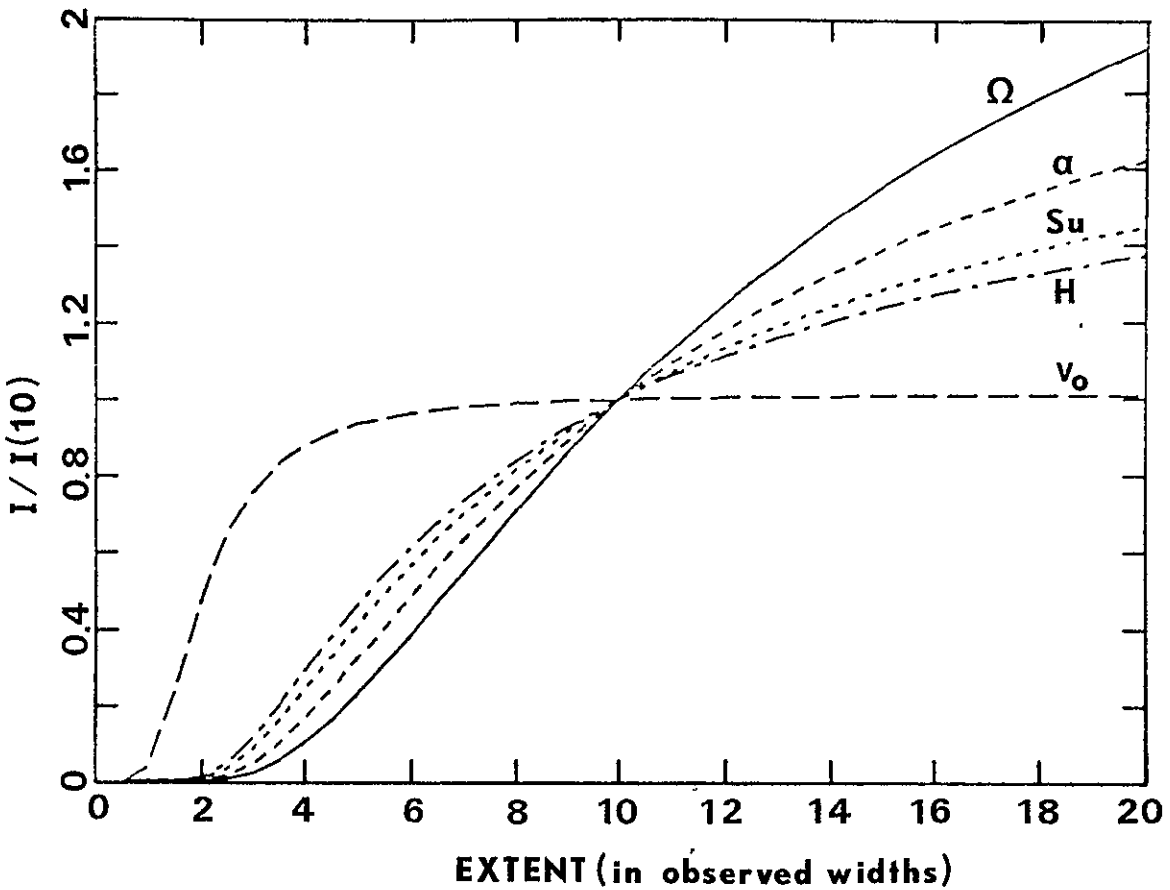
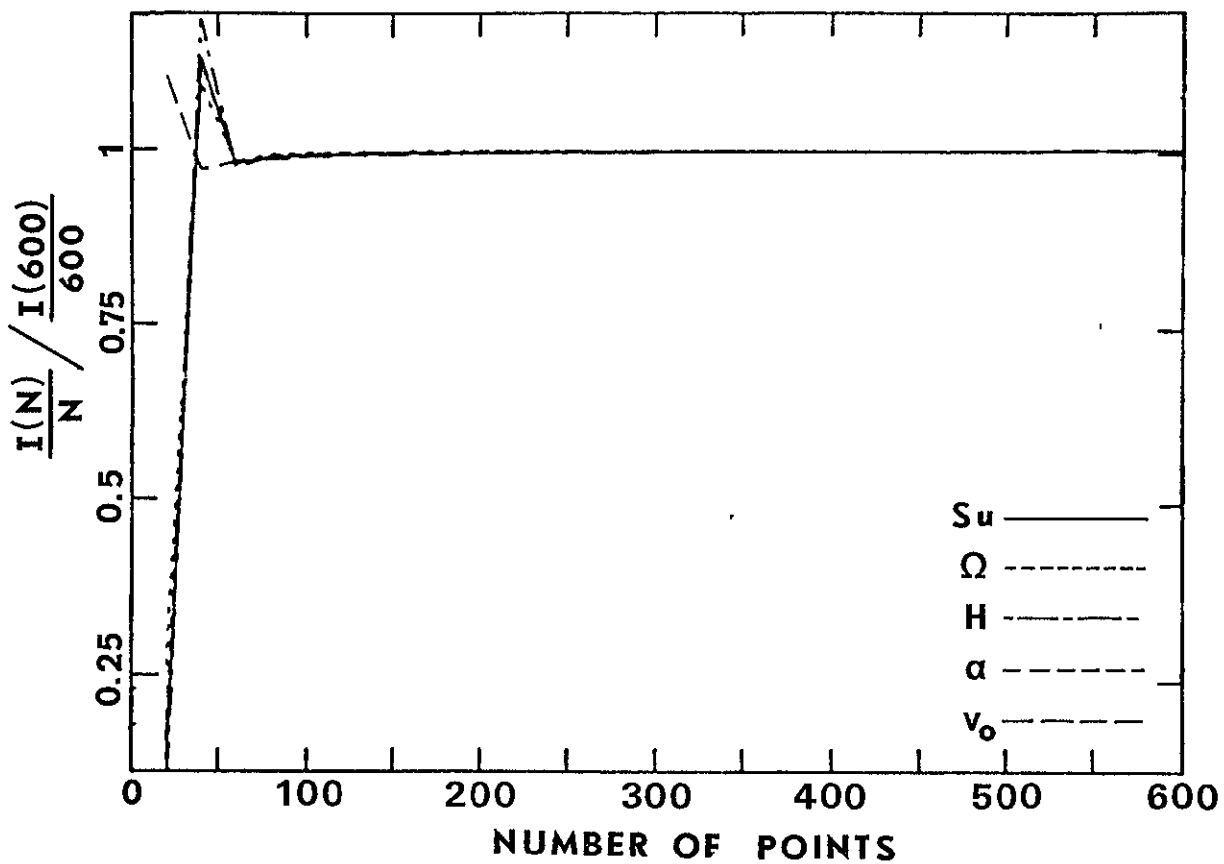

Ω
α
Su
H
V_o
I / I(10)
0
0.4
0.8
1.2
1.6
2
0
2
4
6
8
10
12
14
16
18
20
EXTENT (in observed widths)

Fig 5


$\frac{I(N)}{N} / \frac{I(600)}{600}$
0.25
0.5
0.75
1
0
100
200
300
400
500
600
NUMBER OF POINTS
Su
Ω
H
α
v_o

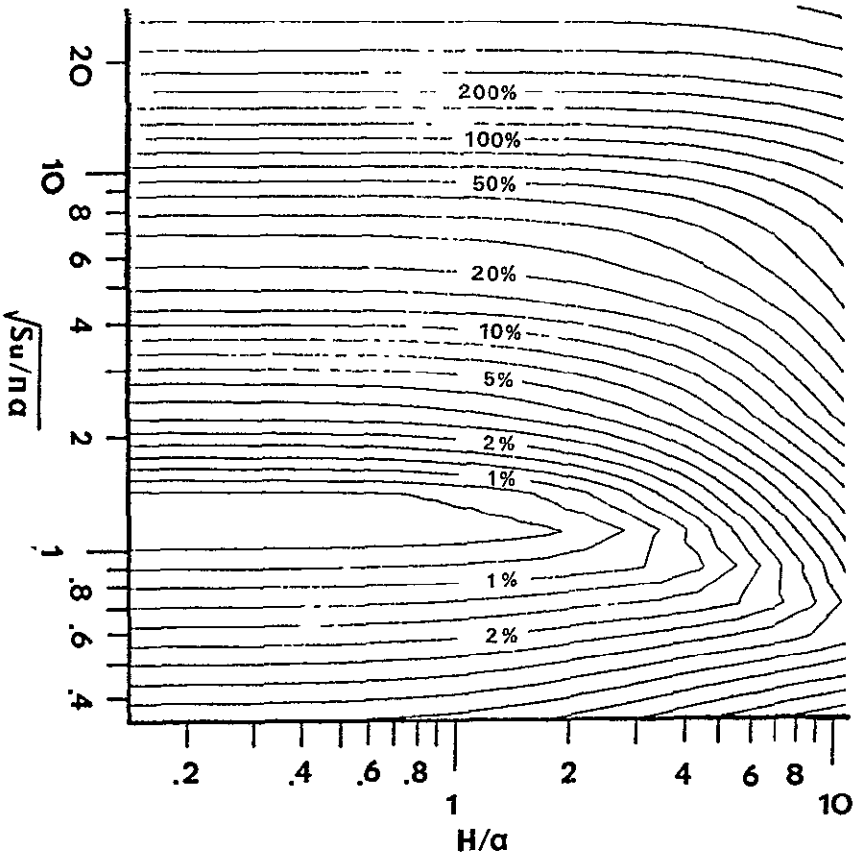


Fig 6

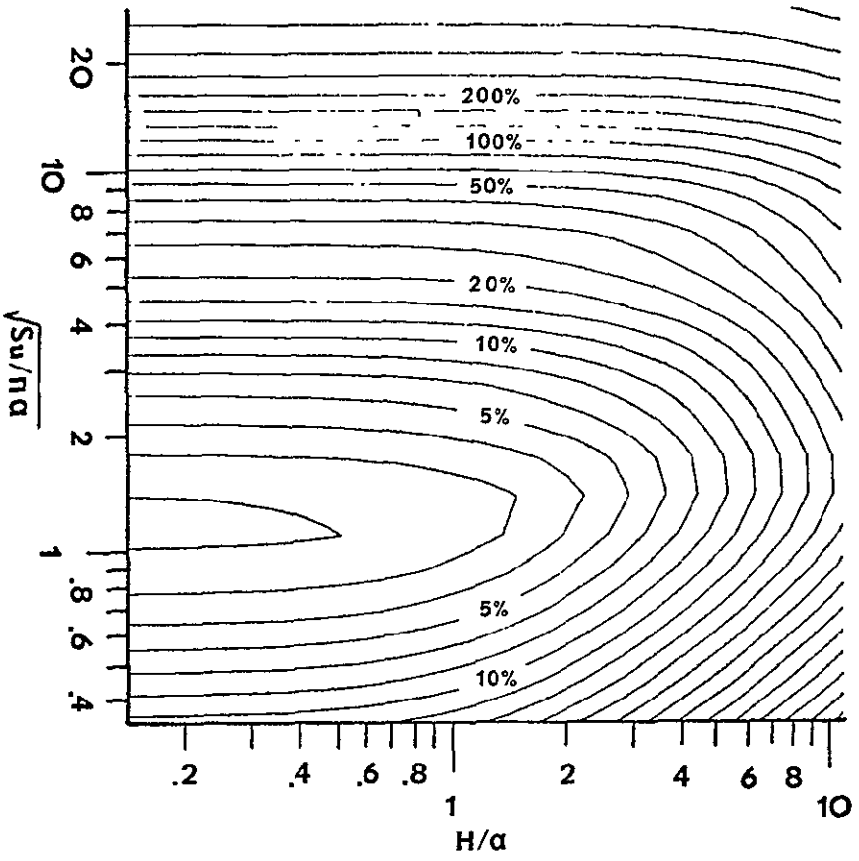


Fig 7

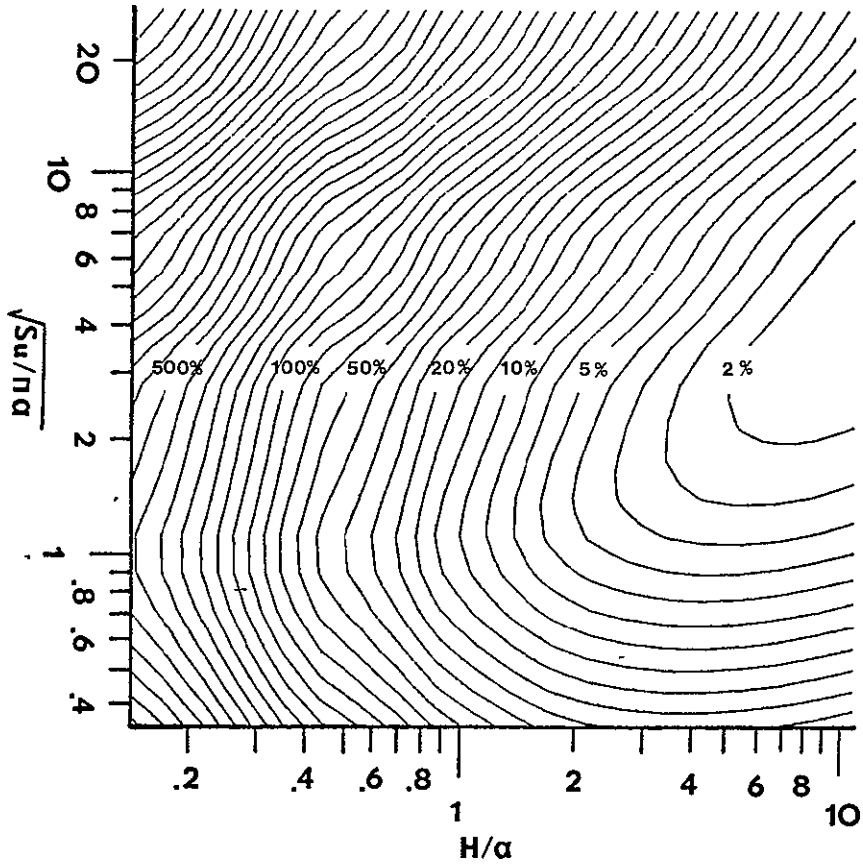


Fig 8

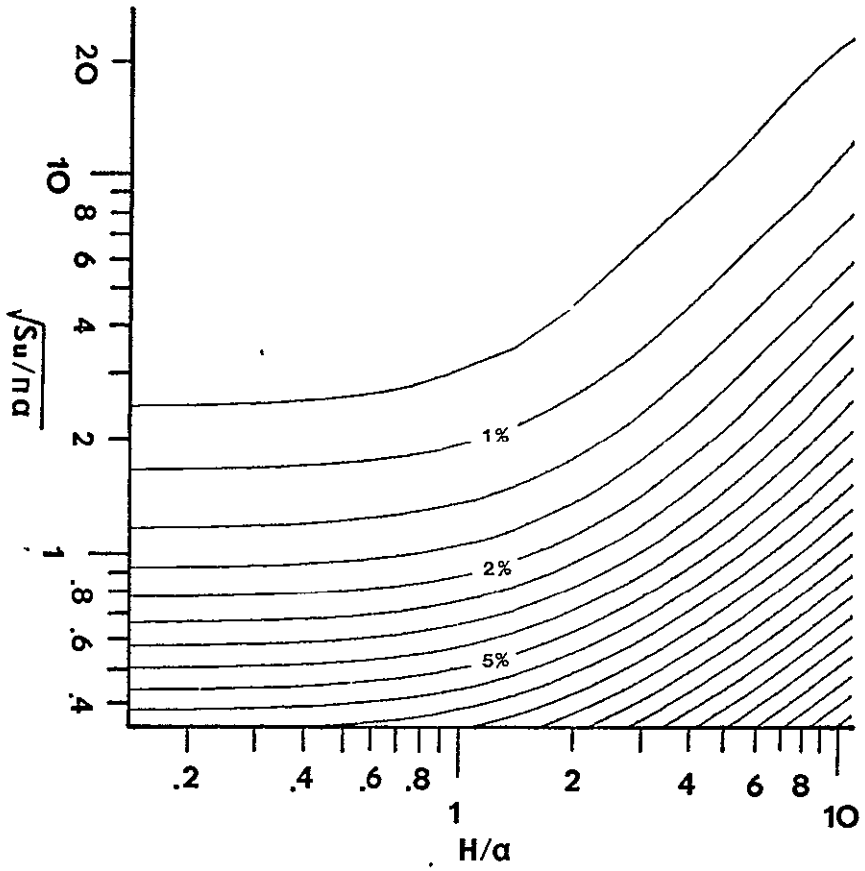


Fig 9

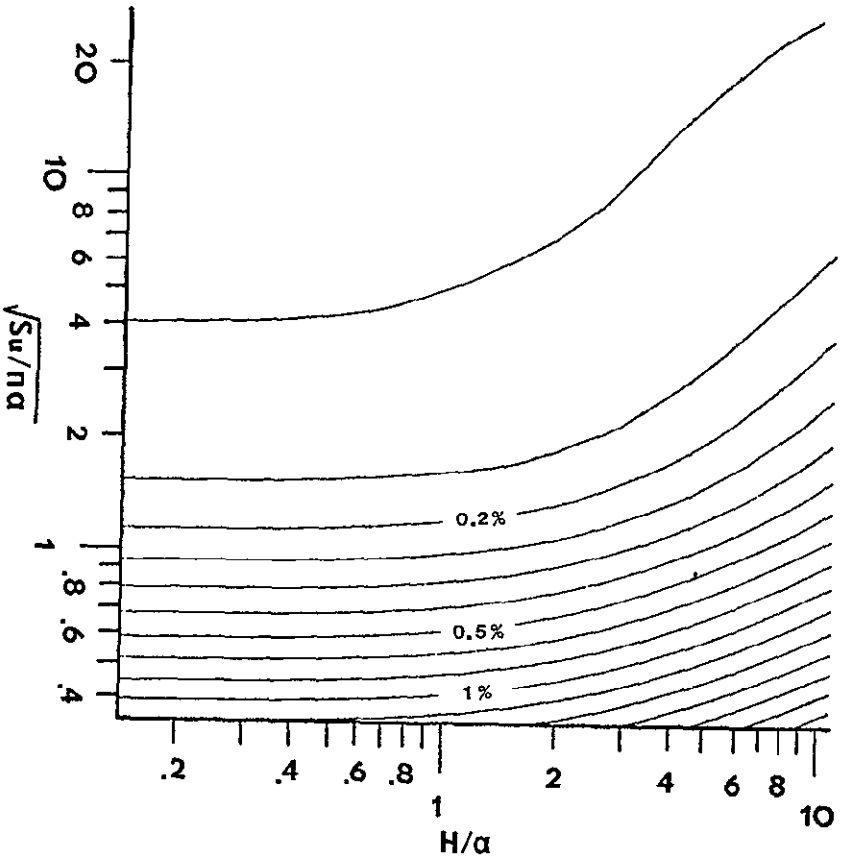


Fig 10

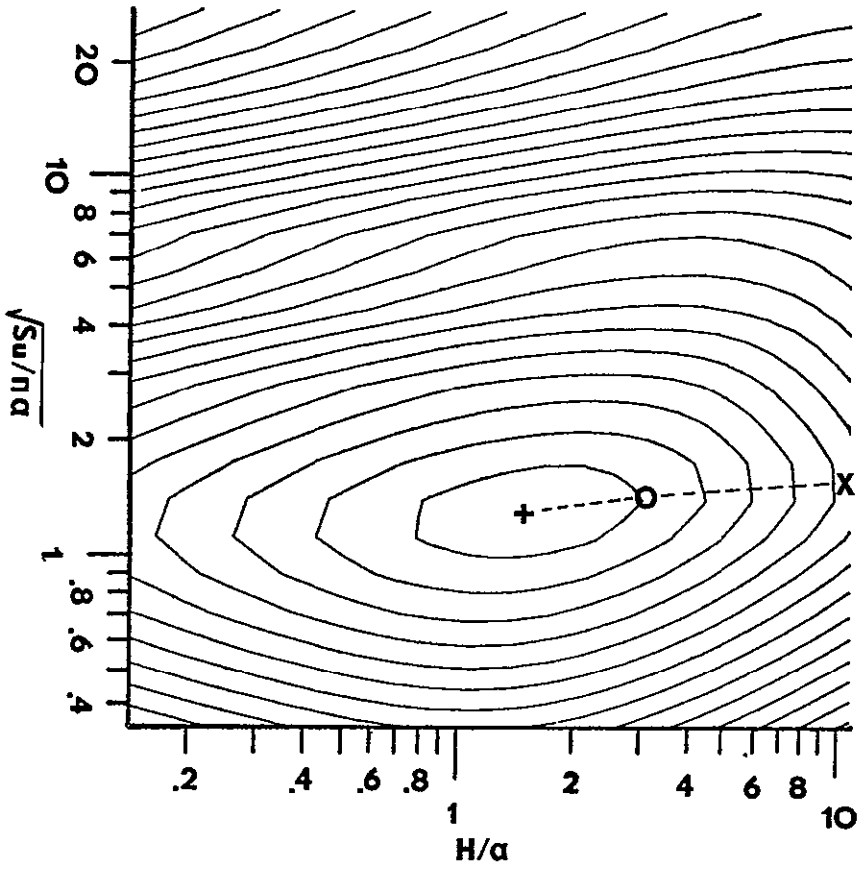


Fig 11

BAND ANALYSIS BY SPECTRAL CURVE FITTING*

C. L. Lin
J. H. Shaw
Department of Physics

and

J. G. Calvert
Department of Chemistry
The Ohio State University
Columbus, Ohio 43210

*This work was supported in part by EPA Grant R803868
and NASA Grant NSG 7469

ABSTRACT

A method of estimating the values of the parameters in the models describing the positions, widths, and intensities of the lines in rotation-vibration bands of gases without the need for line by line analysis is described. To illustrate the technique, portions of the 1-0 bands of HCl and CO have been analyzed. The values of up to 27 parameters, their standard deviations, and the correlations between the parameters required to describe the spectra have been obtained.

Introduction

Experimental data (signals) are often analyzed by assuming a model and determining the values of its adjustable parameters. The models may be based on theoretical considerations or empirically derived. In either case, an often-used criterion for estimating the best parameter values is that the sum of the squares of the residuals - the squares of the differences between the signals predicted by the model and the experimental signals - should be a minimum. Indeed, as Albritton, Schmeltekopf and Zare[1] have noted, if the appropriate model is linear in the unknown parameters "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." Corresponding assumptions are often made concerning non-linear models. In all cases, the objectives of the analysis include a desire to obtain the parameter values as precisely as possible, to minimize observer bias, and to show the model used is an appropriate one.

In some methods of analysis an intermediate set of parameters is first obtained from the experimental signals and, from this set, other characteristics of the system studied are inferred. Thus, for example, the individual lines in rotation-vibration spectra of gases can be described by a position, shape, and intensity. The values of these line parameters, e.g. the line positions, may be described by other relations containing fewer parameters. These band parameters are usually obtained by analysis of the line position values. If these line positions are obtained by the usual line-by-line analysis of the spectrum it must then be assumed that there are no correlations

between these line position values. Again quoting from Albritton et al[1] "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)." In addition, if the line positions are estimated by the observer, the errors associated with these estimates are based on the observer's judgment. Although these errors typically depend on the appearance of each line in the spectrum it is not usually feasible to weight the line position estimates individually. Indeed either the same weight[1] is given to all the line positions obtained or the weight is related in some ad hoc manner to, for example, the peak absorptance of the lines. In many spectra some fraction of the lines are blended and the information from these lines is included or discarded based on the judgment of the observer.

Thus the final estimates of the band parameters and their precision are usually based on a series of steps, each step requiring the judgment of the observer, and the end results depend on the skill of the observer.

In the present method these intermediate steps are omitted, and the desired band parameters are obtained directly from the experimental spectral data. Equal weights are assigned to all the signals in the observed spectrum and the band constants are retrieved directly by a non-linear, least-squares, spectral-curve-fitting method. This is the equivalent of weighting each line in a consistent and statistically valid way and the band constants are unaffected by observer bias.

Since this method eliminates the intermediate step of measuring individual lines and analyses the complete data set, more of the information present in the spectrum can be extracted than in conventional approaches.

In this paper, it is implicitly assumed that the noise in the spectrum is independent of position and signal strength, that it is randomly distributed with mean zero, and that the model is sufficiently accurate to represent the data. It is useful if the data are available in digital form and it is important that the experimental conditions do not change significantly during data collection. Spectra obtained by Fourier Transform Spectrometers (FTS) are particularly suited for this method of analysis since information about the entire spectral region is obtained simultaneously. Indeed the ability of FTS to acquire data rapidly requires a similar speed in the analysis if the full potential of this type of instrumentation is to be realized.

In Fourier Transform Spectroscopy the experimental signals consist of interferograms. Our interferograms are transformed into conventional spectra by programs supplied by the FTS manufacturer. It is not our purpose to evaluate the relative merits of different transformation techniques and it has been assumed that the transformed spectra are the experimental data sets to be analyzed. As in many statistical analyses it is assumed that there is no error in the independent variable (here the frequencies at which signal information is obtained). These positions are based on the wavelength of a helium-neon laser associated with the interferometer and they also show a small dependence on the optical alignment of the system. We

have detected small shifts in the positions of spectral features in similar spectra taken on different occasions. These may introduce small systematic errors in the values of the estimated parameters reported here, since we are unable to model them at this time.

The usefulness of this technique has been explored by analyzing spectra of the 1-0 bands of HCl and CO. These were chosen because of the simplicity of the band models and because the molecular constants have been previously determined with high precision. The analyses were performed with the non-linear, derivative-free, computer-program BMDPAR described by Brown et al[2] by using an Amdahl computer system. This and similar programs have been successfully used to analyze other types of spectra[3].

The method requires that the line intensities, shapes, widths, and spectral positions be appropriately modelled and that the effects of finite spectral resolution on the appearance of the spectra be included. Previous experience[4] has shown that good values for the line intensities and widths are obtained provided there are a sufficient number of data points in the vicinity of each line and the line widths are comparable to the spectral resolution. It has also been found that the line positions are not strongly correlated with the other parameters. Since most gases have pressure-broadened half-widths of from 0.03 to 0.1 cm^{-1}/atm at room temperature, these previous results suggest that, if the spectral resolution is about 0.1 cm^{-1}, then the sample pressure should be near one atmosphere if the line widths are to be determined. The line widths of HCl are strongly J dependent[5] and become very narrow for high J values. There is also a pressure

shift of the line positions[6]. Thus, in addition to reanalyzing previously studied[5] spectra of HCl, a spectrum of the 1-0 bands of CO has been investigated. The J dependence of the line widths[7] and pressure shifts[8] are considerably smaller for CO than for HCl.

Theory

The theoretical bases for the models describing the rotation-vibration bands of gases are described in an extensive body of literature. As the experimental spectral resolution increases the number of adjustable parameters required by the models increases. Only those parameters which have significant influences on the observed spectra should be included in the analysis.

The spectra analyzed were obtained with a Digilab FTS having a nominal resolution of 0.1 cm^{-1}. Analyses [5,9] of these and similar spectra have shown that the individual line positions can be estimated to about 0.002 cm^{-1}, and the intensities and widths to a few percent. Spectra of the 1-0 bands of HCl and CO analyzed here show two sets of lines. These belong to the isotopes of $H^{35}Cl$ and $H^{37}Cl$ and to $^{12}C^{16}O$ and $^{13}C^{16}O$ respectively. Lines of other isotopes and of hot bands have been ignored in this analysis.

Provided the pressure of the gas sample is not too large, the positions of the lines are given by

$$\nu_i = E_i' - E_i'' , \qquad (1)$$

where $i = 1$ identifies a model parameter of $^{13}C^{16}O$ or $H^{37}Cl$, and $i = 2$ identifies the corresponding parameter of $^{12}C^{16}O$ or $H^{35}Cl$.

For these molecules

$$E_i' = \nu_{bi} + B_i' J(J+1) + D_i' J^2(J+1)^2 + H_i' J^3(J+1)^3 ,$$

$$\text{and } E_i'' = B_i'' J(J+1) + D_i'' J^2(J+1)^2 + H_i'' J^3(J+1)^3, \qquad (2)$$

where the symbols have their usual meanings and allowed transitions occur for $\Delta J = \pm 1$. The values [10,11] of the H_i terms are small and are not expected to be retrieved with any precision from the spectra analyzed here.

The individual line intensities $S^1_{oi}(m)$ (cm^{-1}/molecule) can be represented [12] by

$$S^1_{oi}(m) = (8\pi^3/3hc)\nu_i(m)\left|R^1_{oi}(m)\right|^2(|m|/Q_i)\exp\left[-E_i''(m)/kT\right] , \quad (3)$$

where $\left|R^1_{oi}(m)\right|^2$ is the square of the dipole moment matrix element,

$$Q_i \simeq kT/hcB_i'' , \qquad (4)$$

is the rotational partition function, $m = -J$ in the P branch ($\Delta J = -1$), and $m = J + 1$ in the R branch ($\Delta J = 1$).

The square of the dipole moment matrix element can be expressed [13] as the product of a rotationless factor and a rotational or Herman-Wallis factor,

$$\left|R^1_{oi}(m)\right|^2 = \left|R^1_{oi}(0)\right|^2(1 + \beta_1 m + \beta_2 m^2) , \qquad (5)$$

where β_1 and β_2 are assumed the same for both isotopes. Their values are small for both HCl and CO. Eq. (3) may be written as

$$S^1_{oi}(m) = W_i(1 + \beta_1 m + \beta_2 m^2)\,\nu_i(m)\,|m|\,\exp\left[-E_i''(m)/kT\right] , \quad (6)$$

where

$$W_i = 8\pi^3\left|R^1_{oi}(0)\right|^2/3hcQ_i , \qquad (7)$$

is constant for all the lines of a given isotope.

The pressures of the samples were such that the lines have an essentially Lorentzian shape with absorption coefficients given by

$$k_{mi}(\nu) = \left[S^{1}_{oi}(m)\ (|m|)/\pi\right]\left\{\left[\nu-\nu_{oi}(m)\right]^{2} + \alpha^{2}(|m|)\right\}^{-1}, \qquad (8)$$

where $\alpha(|m|)$ are the line widths. Lines of both isotopes with the same $|m|$ value have been assigned the same widths, and deviations from the Lorentz shape in the far wings have been ignored. The dependence of $\alpha(|m|)$ on $|m|$ has been modelled by

$$\alpha_{s}(|m|) = \alpha_{o} + \alpha_{1}|m| + \alpha_{2}m^{2} + \alpha_{3}|m|^{3} + \alpha_{4}m^{4}, \qquad (9)$$

where the coefficients α_{j} are to be determined and $\alpha_{s}(|m|)$ is the half width at some standard pressure (760 Torr). The corresponding widths at pressure P are

$$\alpha(|m|) = P\alpha_{s}(|m|)/P_{s}. \qquad (10)$$

Initial values for the coefficients α_{j} were found by fitting Eqs. (9) and (10) to previously reported values for the line widths[5,14]

Eqs. (1) - (10) can be used to give the combined absorption coefficient due to all the lines

$$k_{b}(\nu) = \sum_{i=1,2} k_{bi}(\nu). \qquad (11)$$

The corresponding monochromatic band transmittance

$$T_{bt}(\nu) = \exp\left[-\sum_{i=1,2} k_{bi}(\nu)N_{i}\right], \qquad (12)$$

where N_{i} are the numbers of absorbing molecules of each isotope in the path. If the absorbing path length is l, the absorber partial pressure is P_{a}, and the sample temperature is T

$$N_{i} = f_{i}\, l\, P_{a}T_{o}L/P_{s}T, \qquad (13)$$

where f_i is the fractional abundance (assumed known) of an isotope, T_o = 273K, and L is Loschmidt's number. It is then found that each of the terms on the right hand side of Eq. (12) contains a term of the form

$$V_i = W_i N_i / \pi \,, \tag{14}$$

which consists of the unknown parameter $\left|R^1_{oi}(0)\right|^2$ and other, known parameters. In the analyses, the estimated values of V_i were obtained.

The monochromatic transmittance can also be written as

$$T_{bt}(\nu) = I_t(\nu) / I_o(\nu) \,, \tag{15}$$

$$\text{or} \quad I_t(\nu) = I_o(\nu)\, T_{bt}(\nu), \tag{16}$$

where $I_o(\nu)$ is the signal observed in the absence of the sample and $I_t(\nu)$ is the signal with the sample present. The corresponding signal observed with a spectrometer of finite resolution

$$\begin{aligned} I(\nu) &= I_o(\nu)T(\nu) \\ &= \int I_o(\nu')T_{bt}(\nu')\,\sigma(\nu, \nu')\,d\nu' \Big/ \int \sigma(\nu, \nu')\,d\nu', \end{aligned} \tag{17}$$

where $\sigma(\nu, \nu')$ describes the shape of the spectral response function of the instrument. Previous analyses [5] of our FTS spectra have shown that $\sigma(\nu, \nu')$ can be well represented by a triangular function with full width at half height R, and that the signal, in the absence of the sample,

$$I_o(\nu) = a + b(\nu - \bar{\nu}) + c(\nu - \bar{\nu})^2, \tag{18}$$

where the values of the parameters a, b, and c are to be estimated and $\bar{\nu}$ is chosen to lie near the center of the spectral region analyzed. Typically, the background $I_{os}(\nu)$ of the sample spectrum

shows a marked dependence on frequency because of variations in the spectral radiance of the source and of the spectral detectivity of the detector. Most of these variations can be removed by ratioing the signals in the sample spectrum with the signals of a spectrum obtained with the sample removed. In such cases the background $I_o(\nu)$ is nearly independent of frequency and the last two terms in Eq. (12) are essentially zero.

The above expressions enable the experimental spectra to be modelled by a function containing 27 adjustable parameters. Eight of these (two background parameters: b and c, two Herman-Wallis factors: β_1 and β_2, and four H terms)are expected to have only a small influence on the spectra. Estimates of each of the parameter values are obtained by analyzing the entire set of signal values in the spectral region of interest. This includes signal values in the vicinity of blended lines as well as those where there is little line absorption. Estimates of the asymptotic standard deviation of each parameter and the asymptotic correlation matrix are also produced. The reliability of the retrieved values depends both on the quality of the spectra analyzed and on the accuracy of the models used. The model accuracy can be estimated by comparing the spectrum calculated from the retrieved parameter values with the observed spectrum. The differences should be of the same order as the noise level in the experimental spectrum.

Results

The method was tested by analyzing a synthetic spectrum of HCl calculated from the parameter values given in Tables 1 and 2, under the heading Spectrum 1. These values were chosen to produce a spectrum similar in appearance to an actual HCl spectrum analyzed previously[5] by the conventional line by line approach. The sample conditions and half width parameter values were obtained from this previous work, the line position parameters from the data of Rao[10], and the parameters defined by Eq. (5) from Toth et al[12]. A signal to noise ratio of 100 was simulated by adding random numbers with a Gaussian distribution to the signal values. To reduce computation time only the spectral region containing the P(8) to R(9) lines was analyzed and all signal values more than $5cm^{-1}$ from a line center were removed.

The analysis can be performed by constraining any number of the variable parameters to fixed values, or by fixing the ratios or differences of pairs of values. The data may be weighted, and the number of iterations and the tolerance for convergence specified. In these results, all the data were equally weighted, and a stringent convergence condition imposed. A large number of iterations may be required before this latter condition is satisfied and many of the results described here were obtained after the specified number of iterations but before the convergence criterion was satisfied.

The results obtained by analyzing the synthetic spectrum with no constraints are shown in Table 2 under the heading Spectrum 1.

These results were obtained after 16 iterations. Some representative parameter values obtained for intermediate iterations are shown in Table 3, together with the sum of the squares of the residuals. This sum was still decreasing after 16 iterations, and the convergence criterion had not been satisfied. The results in Table 3 show oscillations about the expected values and a slow convergence to the final results as the number of iterations increases. The estimated parameter values in Table 2 also include the asymptotic standard deviations as $5.2736(9) \equiv 5.2736 \pm 0.0009$. In most cases the final estimated parameter values agree with the values used to calculate the spectrum within one or two standard deviations. The estimated values for the band centers were only printed to three decimal places.

The root mean square error of the measurements can be estimated from the relation

$$\sigma = \left[\sum(\text{residual})^2 / (N-m)\right]^{1/2}, \qquad (19)$$

where N (3611) is the number of signal values analyzed and m(27) is the number of variable parameters. The estimated signal-to-noise ratio obtained by using the value of $\sum(\text{residual})^2$ obtained after 16 iterations

$$\begin{aligned} (SNR)_{est} &= I_o(\nu)/\sigma_{16}, \\ &= 100.20. \end{aligned} \qquad (20)$$

This is close to the value of 100 used to compute the spectrum and indicates that a near optimum solution has been obtained. Additional iterations are expected to improve the estimates only slightly. This

increase in precision must be balanced against computer costs. A new spectrum must be calculated for each iteration and for the increment halvings required in any iteration. Most of the results described here were obtained by using inefficient methods of calculating the spectra and each analysis required about one hour of computer time. New programs now allow the same computations to be made four times more rapidly.

The correlation coefficients between the variable parameters for this analysis of a synthetic spectrum are shown in Table 4. In addition to the expected high correlations between parameters such as ν_0, B, D, and H, other strong correlations also occur. In many types of spectral analysis only a subset of the available information is desired. If implicit assumptions are made concerning the influence of the unmeasured parameters on the spectrum, then the data in Table 4 show that most empirical weighting schemes become subjective in their application.

A spectrum of a sample of HCl at 700Torr was then analyzed. This spectrum, shown in Fig. 1 of reference (5), has a signal-to-noise ratio of about 50. Information about the sample and spectrum is given in Column 2 of Table 1. The parameter values obtained after 12 iterations are shown under the heading Spectrum 2 in Table 2. Because of degeneracies encountered during the iterations, correlation coefficients and standard deviations could not be estimated. These degeneracies may have been caused by the poor quality of the data, the small number of lines analyzed, or the pressure shifts of the lines which were not modelled. They could have been avoided by fixing more of the parameter

values. The results are included because they show that most of the parameter estimates are reasonable. In particular the important line position parameters, half width parameters and band strength parameter values are in good agreement with those anticipated.

The results obtained by analyzing the spectrum of a sample of HCl broadened to 50 Torr with N_2 are shown under the heading Spectrum 3 in Table 2. This spectrum is shown in Fig. 1 of reference (5) and the sample conditions are given in Column 3 of Table 1. In this retrieval, the differences $H_i' - H_i''$ were fixed to be less than 10^{-12}. The half widths of the lines in this spectrum are much smaller than the spectral resolution and could not be retrieved accurately, indeed, the estimate of α_2 reached the lower bound, and its standard deviation was not obtained. The large uncertainties in the half width parameters are reflected in a corresponding loss of precision of the spectral resolution and band strength parameters, as expected from Table 4. However, the parameters which determine the line positions are not so strongly correlated with the half width parameters and many have been retrieved with the highest precision of all the analyses in Table 2. The intensities of corresponding lines of $H^{35}Cl$ and $H^{37}Cl$ are approximately the same and hence the precision with which the corresponding band parameters are estimated is similar. The estimates of the other parameters could have been improved by using the values of the half width parameters or the band intensity parameters obtained from the spectrum of the sample at 700 Torr in the analysis. We note that the estimated values of β_2 obtained from both the experimental spectra are negative. A similar result has previously been obtained[5] although

the precision of all the measurements is low.

The experimental spectrum of CO shown in Fig. 1 has also been analyzed. The sample conditions are given in Table 1, Column 4. Initial values for the line position parameters were obtained from Rao[10] and Chen et al[11], and the values for the constants in Eq. (5) from Toth et al[12]. Estimates of only the first four half width parameters α_j were found by fitting the data of Benedict et al[14] since the m dependence of the widths is much less than that of HCl. A synthetic spectrum calculated with the parameter values shown in the last column of Table 5 showed good agreement with the observed spectrum after a uniform shift of 0.0384 cm^{-1} was applied to the observed data. This shift is included in the results shown in Table 5, all of which are based on the analysis of a single spectrum.

The first set of results was obtained by fixing nine of the 26 parameter values and making twelve iterations. The spectral resolution was fixed to the value obtained from previous analyses of an HCl spectrum[5] although Chang et al[9] have found that a value of 0.075 cm^{-1} was more representative of the spectral resolution of a CO spectrum taken on the same day as that shown in Fig. 1. These authors estimated the spectrum analyzed was shifted by 0.0537 cm^{-1} with respect to the line positions in the Air Force Geophysical Laboratories' listing of the parameters of atmospheric lines[15]. Two of the background parameters, the Herman-Wallis parameters, and the H terms were also fixed because they are expected to have little influence on the spectrum and are unlikely to be retrieved successfully.

In the next retrieval only three parameters were fixed and one

iteration performed to obtain estimates of the standard deviations of the other parameters. This confirmed that the parameters previously fixed have estimates with very low precision. Eight parameter values were fixed in the last two retrievals. The third set of results, obtained after four iterations, gave a considerable improvement in the precision of many of the line position parameters of $^{12}C^{16}O$ although the remaining parameter values were little changed. The last set of values was obtained after twelve iterations. No standard deviations were estimated, although the residual sum of squares was about 1% lower than the previous two analyses. The results of four analyses are shown to indicate the stability of the solutions for differing conditions of analysis.

Only a few comments on these analyses of CO are given... The fixing of the spectral resolution (retrieval 1) to an incorrect value substantially increased the residual sum of squares compared to the subsequent retrievals. This incorrect value affected the half width parameter and band intensity parameter values, a result previously observed in the analysis of HCl spectra. Some of the line position parameters were also affected by the poor choice of the spectral resolution. This may be caused by the blending of some lines in the experimental spectrum.

The marked difference in the precision of the line position parameters of the two CO isotopes is probably associated with large differences between the intensities of corresponding lines of $^{13}C^{16}O$ and $^{12}C^{16}O$. However, when the relatively small numbers of lines of both HCl and CO contained in the spectral regions analyzed is considered,

the precision with which the line position parameters have been estimated is satisfactorily high. The power of this method of entire band analysis is also indicated by the observation that previous attempts to use the spectral curve fitting method to analyze single lines of CO were unsuccessful unless several parameter values were fixed.

The spectrum calculated from the parameter estimates of retrieval 4, Table 5 is shown in Fig. 1 together with the differences between the observed and calculated spectra. The agreement is excellent. The dependence of the half widths on $|m|$ is shown as the continuous curve in Fig. 2. The open circles are the experimental values of Benedict et al[14] and the other values were obtained from the AFGL line listing[15].

Discussion

A new method of obtaining unbiased estimates, standard deviations, and correlations between the values of the adjustable parameters in models of vibration-rotation bands, from experimental spectra, has been tested. In addition to allowing the relative influence of each parameter on the spectrum to be assessed, the method can be used to test the merits of different models and to allow local anomalies caused, for example, by accidental resonances between close-lying states, to be detected.

Although considerable effort is required to write the programs, a single spectrum can be analyzed by this method in far less time than is required by conventional methods. Thus, in the line by line method of analysis, the intensity, width, and position of each line is

first obtained. These must be suitably weighted and combined to retrieve the appropriate band parameters. It is often necessary to analyze several spectra to obtain the desired line information. The present technique removes the necessity for measuring single lines and has a greater potential for extracting band parameter information from the spectrum.

The precision with which each parameter is estimated depends on the appearance of the spectrum. This has been discussed by Niple[16] for the case of single lines. It is clearly desirable to maximize the spectral resolution and signal to noise ratio. Many methods of improving the appearance of spectra, degraded by noise or finite resolution, by smoothing or partial deconvolution have been proposed. These may have value if no model of the "true" spectrum is available, and they are often considered useful to apply before the results are analyzed by conventional methods. However, if parameter values are to be estimated by fitting the data to a model, Kullback[17] and others have shown that manipulation of the experimental data before analysis cannot add to its information content and tends to remove information. In our method no prior manipulation of the data is required, and the non linear least squares method of analysis is generally accepted as giving the best estimates of the desired parameters when a non linear model must be used.

The appearance of the spectrum is also determined by the physical conditions (temperature, pressure and absorber amount) of the sample. We have not explored these relationships although the results obtained suggest there is no unique set of sample conditions which will allow all the parameters to be estimated with optimum precision. Thus the

analysis of several spectra may be required. The present method can be modified to allow several spectra to be analyzed simultaneously, or to allow weighted parameter values from other sources to be included in the analysis to obtain single best estimates of the parameters from the entire data set.

A single spectrum may contain several bands of a molecule. If the band models contain common parameter values - for example if the bands have common upper or lower states - improved estimates can be obtained by simultaneous analysis of all of these bands appearing in different spectra. The ability to retrieve information is thus limited only by the quality of the experimental data and computer costs.

We have not explored the strategies to be used to analyze an unknown spectrum. The spectral examples discussed here are of molecules whose properties are well known and good initial guesses of the parameter values can be made. In other cases, the appropriate band model may not be known, or the spectrum may consist of incompletely resolved lines due either to insufficient spectral resolution or to the close spacing of the lines.

Thus, for example, if only very low resolution spectra of the 1-0 bands of CO showing unresolved P and R branches were available for analysis, only a few of the adjustable parameters could be obtained. In this case, the contributions by the weaker isotopes would be neglected, the lines would be assumed to have some width independent of m, and only the background parameter a, the band center, the average B value, and $\left|R_o^1(0)\right|^2$ estimated. These values could then be used as initial guesses to analyze more highly resolved spectrum.

REFERENCES

1. D. L. Albritton, A. L. Schmeltekopf, and R. N. Zare, An Introduction to the Least Squares Fitting of Spectral Data, in Molecular Spectra - Modern Research, Vol 2, K. N. Rao Ed., Academic Press (1976).

2. BMDP-77, Biomedical Computer Programs, P Series, M. B. Brown, Editor, University of California Press, Berkeley (1977).

3. Y. S. Chang, J. H. Shaw, J. G. Calvert, and W. M. Uselman, JQSRT, 19, 599 (1978).

4. Y. S. Chang and J. H. Shaw, Appl. Spectrosc., 31, 213 (1977).

5. C. L. Lin, E. Niple, J.H.Shaw, and J.G.Calvert, JQSRT, 20, 581 (1978).

6. G. Guelachvili and M. A. H. Smith, JQSRT, 20, 35 (1978).

7. R. H. Hunt, R. A. Toth, and E. K. Plyler, J. Chem. Phys., 49, 3909 (1968).

8. R. B. Nerf, Jr. and M. A. Sonnenberg, J. Mol. Spec., 58, 474 (1975).

9. Y.S.Chang, J. H. Shaw, W. M. Uselman, and J. G. Calvert, App. Opt., 16, 2116 (1977).

10. K. N. Rao, High Resolution Infrared Spectroscopy: Aspects of Modern Research, in Physical Chemistry Series Two, Vol. 3, D. A. Ramsay, Ed., Butterworths (1976).

11. D. Chen, K. N. Rao, and R. S. McDowell, J. Mol. Spec., 61, 71 (1976)

12. R. A. Toth, R. H. Hunt, and E. K. Plyler, J. Mol. Spec/, 32, 85 (1969) and 35, 110 (1970).

13. R. A. Toth, R. H. Hunt, and E. K. Plyler, J. Mol. Spec., 32, 74 (1969).

14. W. S. Benedict, R. Herman, and S. Silverman, Astrophys. J., 135, 277 (1962).

15. L. S. Rothman and R. A. McClatchey, Appl. Opt., 15, 2616 (1976).

16. E. Niple, Ph.D. dissertation, The Ohio State University (1978).

17. S. Kullback, Information Theory and Statistics, John Wiley and Sons, New York (1956).

Figure Captions

Fig. 1. - The experimental CO spectrum analyzed is shown in the top curve, beneath it is the spectrum calculated from the retrieved values in column 4, Table 5. The other curves show the ratios and differences of these two spectra.

Fig. 2. - The continuous curve represents the polynomial dependence of the CO line widths on m calculated from the retrieved values shown in column 4 , Table 5. The open circles represent the average half widths obtained from Benedict, Herman and Silverman and the asterisks show the half widths in the latest version of the AFGL atmospheric line listing.

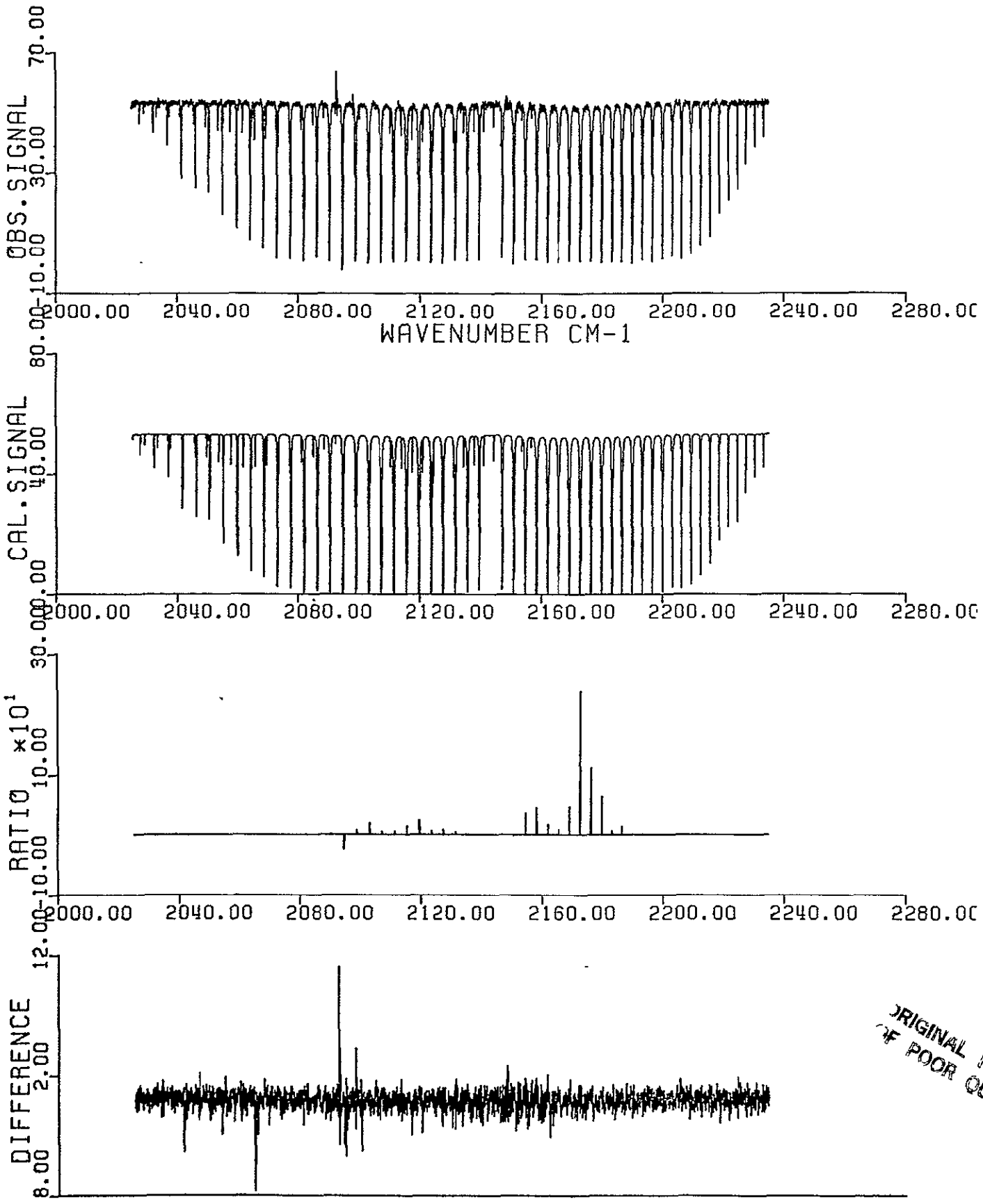


Fig. 1-1

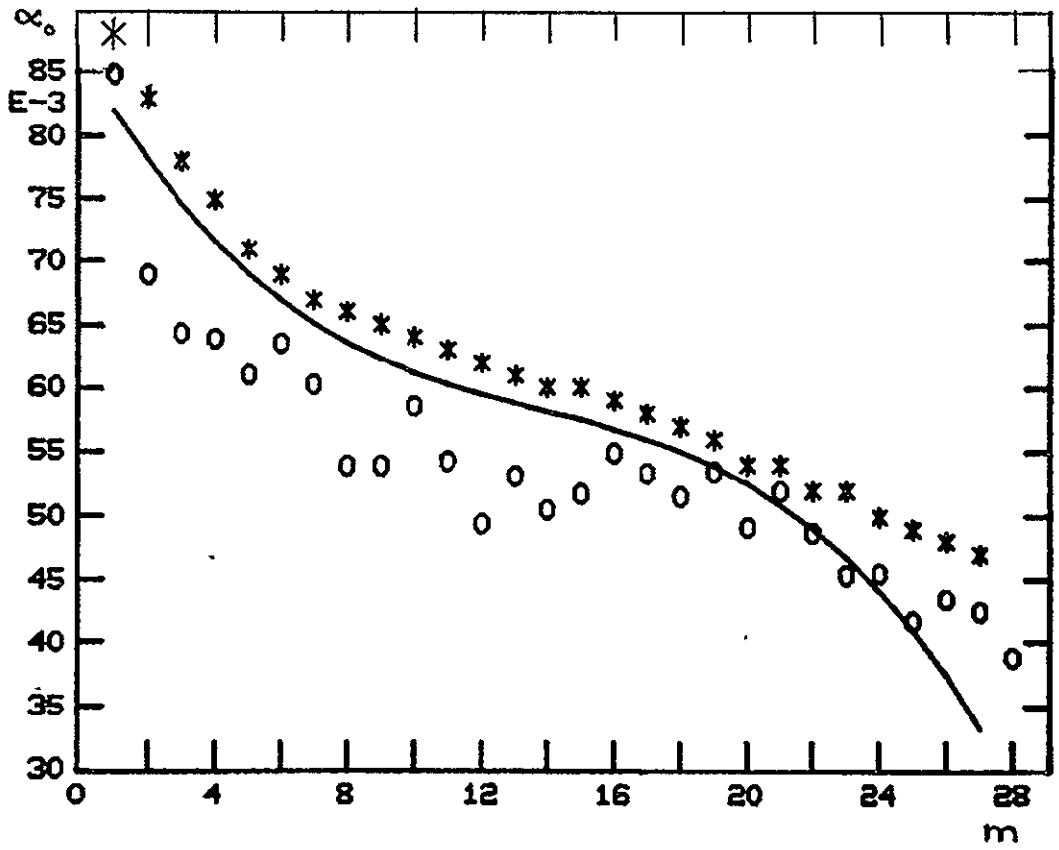


Fig. 2

TABLE 1. Sample conditions and spectral characteristics

Spectrum	1	2	3	4
Absorber	HCl	HCl	HCl	CO
f_2/f_1 Ratio	3.065	3.065	3.065	89.25
Absorber amt. atm. cm-(296K)	0.421	0.421	0.421	0.350
Partial Pressure of N_2 (Torr)	700	700	50	700
Sample Temp (K)	296	296	296	296
Path Length (m)	171	171	171	171
Spectral Limits Upper (cm^{-1})	3050	3050	3050	2235
Lower (cm^{-1})	2671	2671	2671	2025
Spacing of data points (cm^{-1})	0.0603	0.0603	0.0603	0.0537
No. of points analyzed	3611	3611	3611	3920
$\bar{\nu}$ (cm^{-1})	2877	2877	2877	2130

TABLE 2. HCl Spectra: estimated parameter values

Spectrum		1			2	3	
#	Parameter	Input Values	Retrieved		Retrieved	Retrieved	
1	a	5.2740	5.2736	(9)	5.276	44.09	(3)
2	10^6b cm	0	1	(2)	-48	-2.3×10^3	(2)
3	10^8c cm^2	0	-2	(3)	80	4×10^2	(2)
4	10R cm^{-1}	1.04	1.04	(2)	1.15	1.19	(2)
5	10^5V_1	7.604	7.56	(5)	6.68	53	(5)
6	10^4V_2	2.349	2.35	(2)	2.00	17	(2)
7	$10^2\beta_1$	-2.60	-2.5	(1)	-3.16	-6	(3)
8	$10^4\beta_2$	4.50	5	(3)	-16	-28	(23)
9	ν_{b1} cm^{-1}	2883.8705	2883.870 ±.0003		2883.870	2883.870 ±0.00004	
10	B_1' cm^{-1}	10.12090	10.12090	(9)	10.12115	10.12134	(7)
11	$10^4D_1'$ cm^{-1}	5.200	5.21	(2)	5.242	5.241	(3)
12	$10^8H_1'$ cm^{-1}	1.63	2	(1)	3.9	3.9008	(3)
13	B_1''	10.42430	10.42417	(9)	10.42485	10.42458	(7)
14	$10^4D_1''$ cm^{-1}	5.270	5.25	(2)	5.350	5.328	(3)
15	$10^8H_1''$ cm^{-1}	1.63	1	(1)	3.9	3.9009	(3)
16	ν_{b2} cm^{-1}	2885.9775	2885.979 ±.0003		2885.980	2885.980 ±.00004	
17	B_2' cm^{-1}	10.136223	10.13615	(7)	10.13601	10.13614	(4)
18	$10^4D_2'$ cm^{-1}	5.2167	5.20	(1)	5.203	5.203	(3)
19	$10^8H_2'$ cm^{-1}	1.65	0	(1)	0.2	0.6	(6)
20	B_2'' cm^{-1}	10.440254	10.44032	(5)	10.44064	10.44032	(2)
21	$10^4D_2''$ cm^{-1}	5.2835	5.290	(8)	5.325	5.305	(4)
22	$10^8H_2''$ cm^{-1}	1.65	1.8	(7)	0.2	0.6	(6)
23	$10\alpha_o$ cm^{-1}	1.0497	1.049	(1)	1.06	1.4	(1)
24	$10^4\alpha_1$ cm^{-1}	3.84	2	(1)	5.55	50	(30)
25	$10^3\alpha_2$ cm^{-1}	-1.568	-1.58	(3)	-1.748	-2×10^{-3} (bound)	
26	$10^6\alpha_3$ cm^{-1}	-5.76	-4	(2)	-3.00	-18	(19)
27	$10^8\alpha_4$ cm^{-1}	+6.64	6.8	(3)	9.65	11	(1)

TABLE 3. Some intermediate parameter values obtained during the analysis of a synthetic HCl spectrum

Parameter	Res. Sum. Squares	a	R(cm^{-1})	$10^2\beta_1$	$10^4D_1''$ (cm^{-1})	B_2' (cm^{-1})	$10\alpha_0$ (cm^{-1})
Initial Guess		5.2740	0.10400	-2.60	5.270	10.13600	1.050
Iteration Number							
1	22.827	5.2740	0.10363	-2.62	5.272	10.13597	1.162
5	13.330	5.2739	0.10458	-2.60	5.274	10.13621	1.059
10	10.064	5.2746	0.10206	-2.52	5.247	10.13621	1.053
15	9.931	5.2737	0.10388	-2.46	5.249	10.13614	1.050
16	9.929	5.2736	0.10391	-2.46	5.252	10.13615	1.049

TABLE 4. Estimate of the asymptotic correlation matrix obtained from the analysis of a synthetic HCl spectrum (all values multiplied by 1000)

Parameter	a	b	c	R	v_1	v_2	β_1	β_2	v_{b1}	B_1'
a	1000									
b	435	1000								
c	-529	-334	1000							
R	-322	-153	166	1000						
V_1	-474	-437	321	472	1000					
V_2	-520	-423	342	602	866	1000				
β_1	-31	-15	16	341	284	286	1000			
β_2	-184	158	430	148	-73	100	-56	1000		
v_{b1}	513	-441	-172	23	112	59	63	-354	1000	
B_1'	223	260	397	-34	74	48	70	298	-36	1000
D_1'	447	430	281	-109	14	-17	56	326	105	846
H_1'	582	504	79	-154	-81	-108	74	239	201	600
B_1''	77	-105	680	34	122	103	66	285	144	842
D_1''	-33	-170	857	47	163	158	16	398	166	584
H_1''	-66	-212	806	20	94	97	40	384	142	334
v_{b2}	-232	6	834	132	61	93	47	525	-181	458
B_2'	-127	-553	202	-13	378	332	-29	-250	376	27
D_2'	-131	-799	234	53	477	416	25	-295	687	-73
H_2'	-109	-740	208	75	433	378	62	-223	677	-93
B_2''	-337	-26	467	19	184	196	-50	219	-366	282
D_2''	-492	373	504	79	-50	7	-54	535	-884	292
H_2''	-294	457	236	13	-215	-163	-50	462	-793	166
α_o	463	465	-134	-671	-770	-900	-240	-17	-142	129
α_1	31	14	-15	-299	-229	-250	-859	-35	-52	-64
α_2	-247	-630	-221	196	453	402	66	-566	430	-335
α_3	-11	-63	-57	228	158	184	595	-70	101	-19
α_4	86	497	351	-72	-110	-99	96	449	-416	371

TABLE 4 (cont)

Parameter	D_1'	H_1'	B_1''	D_1''	H_1''	ν_{b2}	B_2'	D_2'	H_2'	B_2''
D_1'	1000									
H_1'	912	1000								
B_1''	649	410	1000							
D_1''	610	463	834	1000						
H_1''	444	398	614	919	1000					
ν_{b2}	394	233	682	827	785	1000				
B_2'	64	46	138	200	163	-313	1000			
D_2'	-86	-110	183	249	225	-78	677	1000		
H_2'	-127	-151	176	231	217	65	366	916	1000	
B_2''	253	144	274	342	287	101	547	-94	-343	1000
D_2''	134	-35	228	249	218	552	-379	-642	-609	466
H_2''	30	-85	72	47	58	473	-670	-695	-525	-7
α_o	219	264	73	67	108	134	-361	-449	-392	-83
α_1	-61	-81	-49	-11	-32	-38	20	-28	-59	44
α_2	-447	-437	-269	-370	-395	-579	601	738	616	-99
α_3	-22	14	-32	-73	-54	-80	61	93	90	-55
α_4	440	381	345	450	451	639	-510	-587	-463	143

TABLE 4 (cont)

Parameter	D_2''	H_2''	α_o	α_1	α_2	α_3	α_4
D_2''	1000						
H_2''	836	1000					
α_o	165	285	1000				
α_1	52	49	215	1000			
α_2	-615	-647	-578	-52	1000		
α_3	-150	-147	-204	-877	158	1000	
α_4	642	632	336	-29	-867	-177	1000

TABLE 5. CO Spectrum. Estimated Parameter Values

Retrieved Parameter		1		2		3		4	Calculated	Ref.
1	a	53.07	(2)	53.11	(2)	53.11	(1)	53.095		
2	10^8b cm	0.0f		-1	(2)	0.0f		0.0f		
3	10^9c cm^2	0.0f		-1	(4)	0.0f		0.0f		
4	10R cm^{-1}	1.040f		0.69	(1)	0.69	(1)	0.689		
5	10^6V_1	1.30	(2)	1.22	(4)	1.22	(4)	1.215	1.227	
6	10^4V_2	1.25	(3)	1.14	(3)	1.14	(3)	1.121	1.143	
7	$10^4\beta_1$	-2.80f		-3	(2)	-2.8f		-2.8f	-2.8	12
8	$10^6\beta_2$	7.50f		7.5f		7.5f		7.5f	7.5	12
9	ν_1 cm^{-1}	2096.0686	(8)	2096.06	(1)	2096.06	(1)	2096.058	2096.0670	11
10	B_1' cm^{-1}	1.8216	(1)	1.822	(3)	1.822	(3)	1.8217	1.82161	11
11	$10^6D_1'$ cm^{-1}	6.2	(3)	6.05	(4)	6.04	(6)	5.922	5.594	11
12	$10^{12}H_1'$	5.0f		5.0f		5.0f		5.0f	5.0	11
13	B_1'' cm^{-1}	1.83809	(9)	1.838	(3)	1.838	(3)	1.8380	1.83797	11
14	$10^6D_1''$	6.5	(2)	5.58	(4)	5.58	(7)	5.478	5.596	11
15	$10^{12}H_1''$ cm^{-1}	5.0f		5.0f		5.0f		5.0f	5.0	11
16	ν_2 cm^{-1}	2143.2696	(4)	2143.269	(1)	2143.2686	(1)	2143.26848	2143.2715	10
17	B_2' cm^{-1}	1.905045	(8)	1.9050	(3)	1.904984	(5)	1.904990	1.905026	10
18	$10^6D_2'$	6.103	(1)	6.039	(3)	6.039	(3)	6.0400	6.1195	10
19	$10^{12}H_2'$ cm^{-1}	5.44f		6	(9)	5.44f		5.44f	5.44	10
20	B_2'' cm^{-1}	1.922575	(9)	1.9225	(3)	1.922504	(5)	1.922510	1.922529	10
21	$10^6D_2''$ cm^{-1}	6.142	(3)	6.066	(3)	6.066	(3)	6.0673	6.1202	10
22	$10^{12}H_2''$ cm^{-1}	5.60f		6	(9)	5.60f		5.60f	5.60	10
23	$10^2\alpha_o$ cm^{-1}	8.1	(2)	8.7	(2)	8.7	(2)	8.73	8.18	14
24	$10^3\alpha_1$ cm^{-1}	-4.7	(2)	-4.8	(1)	-4.8	(1)	-4.75	-5.06	14
25	$10^4\alpha_2$ cm^{-1}	2.6	(1)	3.0	(1)	3.0	(1)	2.91	2.91	14
26	$10^6\alpha_3$ cm^{-1}	-5.9	(3)	-7.2	(3)	-7.2	(3)	-7.01	-5.84	14

LEAST SQUARES ANALYSIS OF VOIGT SHAPED LINES*

C. L. Lin
J. H. Shaw
Department of Physics

and

J. G. Calvert
Department of Chemistry
The Ohio State University
Columbus, Ohio 43210

*This work was supported by EPA Grant No. R803868030

ABSTRACT

The retrieval of the Lorentzian width, intensity, and position of Voigt shaped lines obtained with an instrument of finite spectral resolution by spectral curve fitting and the use of non linear least squares techniques is described. The accuracy with which the values of these parameters can be retrieved was tested by analyzing synthetic spectra to which random noise was added. Some of the requirements of the experimental data to obtain good retrievals are identified.

Introduction

A knowledge of the positions, shapes, and intensities of spectral lines is a prerequisite for the quantitative analysis of line spectra. Line position measurements are often made independently from estimates of the line intensity and halfwidth. These latter quantities are usually determined from curve of growth measurements of the equivalent widths of lines obtained from several spectra.

In a recent series of papers Chang et al[1,2,3] have discussed the retrieval of information from spectra by curve fitting using non linear least squares techniques. It has been shown that all the line parameters can be obtained from the analysis of a single spectrum provided the physical conditions of the sample-pressure, temperature, absorber concentration, and path-length are known, and that suitable models for the shapes of the lines and the instrument spectral response function are used. Conversely, if the line parameters are known, the physical conditions of the sample can be estimated.

In this previous work a Lorentzian line shape was assumed. This describes the line shape adequately for samples near ambient temperature and atmospheric pressure but it becomes necessary to use the mixed Doppler-Lorentz (Voigt) profile as the sample pressure is reduced. Thus, although the Lorentzian shape can be used to predict the infrared transmittance in the troposphere, the Voigt shape is required for stratospheric or mesospheric calculations.

In this paper it is shown that the curve fitting technique can be extended to allow Voigt shaped lines to be analyzed. The retrieval of the line parameters from an absorption spectrum, consisting of either a single

isolated line or double lines, obtained with an instrument of finite resolution is discussed. By analyzing synthetic spectra to which random noise has been added it is shown that, provided the Doppler width of the line and the spectral response function shape are known, the line position and intensity, the Lorentz half width, the spectral resolution, and the position of the background curve in the absence of line absorption can be retrieved from spectra taken with a wide variety of experimental conditions. In addition, experimental data of the P(1) and P(9) pairs of isotopic lines in the 1-0 band of HCl have been analyzed by assuming them to have both a Lorentz shape and a Voigt shape.

The principal computer programs used in this work were developed by the Health Sciences Computing Facility at the University of California and are described in Ref. (4). Both single and double precision versions have been used to verify the accuracy of the computations.

Method

If a homogeneous absorbing gas sample has a monochromatic absorption coefficient $k(\nu)$ at frequency ν, the transmitted signal $I_m(\nu)$ is related to the signal $I_o(\nu)$, obtained with the absorber removed, by

$$I_m(\nu) = I_o(\nu) \exp[-k(\nu)\, u] , \tag{1}$$

where u is a measure of the amount of absorber present,

$$u = \frac{P_{abs}}{P_o} \cdot \frac{T_o}{T_{abs}} \cdot \ell(\text{atm-cm}, P_o, T_o) , \tag{2}$$

where the sample temperature is T_{abs}, the partial pressure of the absorber is P_{abs}, the absorbing path length is ℓ(cm), and T_o and P_o refer to some standard temperature and pressure, respectively. All of the quantities defining u are assumed known.

If the spectrum is obtained with an instrument of finite resolving power then the observed signal can be written as

$$I(\nu) = \frac{\int I_m(\nu')\ \sigma(\nu,\nu')\ d\nu'}{\int \sigma(\nu,\nu')\ d\nu'} , \tag{3}$$

where $\sigma(\nu,\nu')$ describes the instrument spectral response function. A triangular shape has been used in this and previous work.

We have assumed:

1. $\sigma(\nu,\nu')$ is triangular whose full width at half-height H is to be retrieved in the analysis.
2. $I_O(\nu) = A_1 + A_2(\nu - \nu\bar{o})$, where A_1 and A_2 are constants to be determined, and $\nu\bar{o}$ is chosen to lie near the middle of the spectral region analyzed. In the analysis of synthetic spectra described here $\nu\bar{o} = 0.7320\ cm^{-1}$.
3. Values of $I(\nu)$ are obtained at frequency intervals $\Delta\nu$.
4. $k(\nu)$ is described by the Voigt line shape.

The Voigt line shape can be described in terms of the Voigt function

$$k(x,y) = \frac{k(\nu)}{k_O} = \frac{y}{\pi}\int_{-\infty}^{\infty} \frac{\exp(-t^2)}{y^2 + (x - t)^2}\ dt , \tag{4}$$

where $k(\nu)$ is the absorption coefficient,

$$k_O = \frac{S}{\alpha_D}\left(\frac{\ell n\ 2}{\pi}\right)^{1/2} ,$$

$$y = \frac{\alpha_L}{\alpha_D}(\ell n\ 2)^{1/2} ,$$

$$x = \frac{\nu - \nu_O}{\alpha_D}(\ell n\ 2)^{1/2} ,$$

$S = \int_{-\infty}^{\infty} k(\nu)\ d\nu$, is the line intensity,

α_L = Lorentz half width due to radiation damping or collisional broadening,

$\alpha_D = \frac{\nu_o}{c}\left(\frac{2kT\ \ell n\ 2}{m}\right)^{1/2}$, is the Doppler half-width,

and

ν_o = line center position.

The Voigt function cannot be evaluated exactly but a number of techniques for calculating its approximate value have been described.[5,6,7,8] A fast calculational method for obtaining its value to about one part in 10^4 has been described by Pierluissi et al.[9] Three different approximations are used which depend on the values of x and y. This program has been modified to include the case of $y = 0$ in the region $I(0 \leq x \leq 1.8,\ 0 \leq y \leq 3.0)$ and to reach higher precision for the case of pure Doppler lines.

In order to calculate a spectrum of a Voigt line it is necessary to know the values of the quantities A_1, A_2, H, S, u, ν_o, α_D and α_L. Since the Doppler width is not strongly dependent on temperature and since the molecular mass of the absorbing gas is usually known we have assumed the α_D can be calculated. Thus, up to six other parameter values must be determined by spectral curve fitting.

In the technique described by Chang et al,[1,2,3] a calculated spectrum, based on initial guesses of the unknown parameters, is compared with the observed spectrum. The parameter values are then varied until the best fit (defined in the usual manner, that the sum of the squares of the differences between the observed and calculated spectrum is a minimum) is

obtained. There are a number of computational methods for arriving at this best fit, we have used the programs BMDP3R and BMDPAR described in Ref. (4).

The subroutines for analyzing Voigt shaped lines for use with these programs are given in Appendices I and II. The subroutine in Appendix I is designed to be used with BMDP3R for the simultaneous retrieval of the parameters of two absorption lines. Appendix II gives the corresponding subroutine for a single line for use with BMDPAR. Based on the initial parameter guesses, values for x and y are calculated for each position in the spectrum. The corresponding value of the Voigt function is evaluated by the method of Pierluissi et al[9] and this is used together with Eqs. (1-4) to calculate the transmitted signal $I(\nu)$. The subroutine for use with BMDP3R also gives analytical expressions for the partial derivatives of Eq. (3) with respect to each of the unknown parameters required for the curve fitting program.

The Voigt function k(x,y) used in the three different regions proposed by Pierluissi et al is given by the real part of w(z), where w(z) is provided by Abramowitz and Stegun[10] as the following;

in region I ($0 \leq |x| < 3$, $0 \leq y < 1.8$)

$$w(z) = e^{-z^2}[1 - \mathrm{erf}(-iz)] , \tag{5}$$

where $z = x + iy$,

$$\mathrm{erf}(-iz) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\,(2n+1)} (-iz)^{2n+1} , \tag{6}$$

and $n = 15$ if $x = 0$, otherwise $n = 6.842x + 8.0$. The number of terms to be included was determined through an empirical relation in terms of x

and y proposed by Pierluissi et al,[9]

in region II ($3 \leq |x| < 5$, $1.8 \leq y < 5$),

$$w(z) = iz\left(\frac{0.4613}{z^2 - 0.1901} + \frac{0.09999}{z^2 - 1.7844} + \frac{0.00288}{z^2 - 5.5253}\right), \tag{7}$$

and in region III ($|x| \geq 5$, $y \geq 5$),

$$w(z) = iz\left(\frac{0.5124}{z^2 - 0.2752} + \frac{0.05176}{z^2 - 2.7247}\right). \tag{8}$$

The derivatives of w(z) are,

in region I,

$$\frac{dw(z)}{dz} = (-2z) \cdot w(z) + \frac{2}{\sqrt{\pi}} i , \tag{9}$$

in region II,

$$\frac{dw(z)}{dz} = i\left(\frac{0.4613}{z^2 - 0.1901} + \frac{0.09999}{z^2 - 1.7844} + \frac{0.00288}{z^2 - 5.5253}\right)$$

$$- i(2z^2)\left(\frac{0.4613}{(z^2 - 0.1901)^2} + \frac{0.09999}{(z^2 - 1.7844)^2} + \frac{0.00288}{(z^2 - 5.5253)^2}\right), \tag{10}$$

and in region III,

$$\frac{dw(z)}{dz} = i\left(\frac{0.5124}{z^2 - 0.2752} + \frac{0.05176}{z^2 - 2.7247}\right)$$

$$- i(2z^2)\left(\frac{0.5124}{(z^2 - 0.2752)^2} + \frac{0.05176}{(z^2 - 2.7247)^2}\right). \tag{11}$$

Equations (5-11) are used to obtain the derivatives of Eq. (3) with respect to each unknown parameter. It should be noted that the value of x defined in Eq. (4) may take either positive or negative values but the values used in the expressions for the Voigt function given by Pierluissi et al are always positive or zero. Thus, although the absolute values of x are used to obtain the absorption coefficient k(x,y), the correct sign must be assigned to the partial derivatives required in the analytical program shown in Appendix I.

Results

The accuracy of the modified method described by Pierluissi et al for calculating k(x,y) was tested by comparing values for this quantity for wide ranges of x and y values with the tabulation of k(x,y) given by Young.[5] The agreement was well within the accuracy of one part in 10^4.

The analytical program of BMDP3R requires partial derivatives of Eq. (3) for each unknown parameter for each of the three approximations to the Voigt function. Tests of the accuracy of these derivative expressions were performed by calculating the transmittances (to six significant figures) of a single Voigt shaped line, for the parameter values in Table 1. This table also shows the total number of data points calculated in each spectrum, their spacing $\Delta\nu$, and the number of data points for which the transmittance was less than 95%. These parameter values describe lines in each of the three regions defined by Pierluissi et al.[9] Values of the line parameters and other quantities were then retrieved by using program BMDP3R in single precision by considering these calculated transmittances to be "experimental" data. The initial guesses for the line parameters were usually within 30% of the true values. In all cases the values retrieved by the program agreed with the initial values in Table 1 to better than one part in 10^4 after the number of iterations shown.

These tests confirm the accuracy of the programs but they do not establish their ability to retrieve information from actual spectra. Since our present spectrometers have insufficient resolution to obtain suitable spectra of lines in the regions I and II defined by Pierluissi et al, it was necessary to use synthetic data. A series of line

profiles was calculated for several sets of values of the adjustable parameters and random Gaussian noise was added to these profiles before analysis to give a signal-to-noise ratio of 100. Some of the results are shown in Table 2.

This table shows the parameter values used to generate each spectrum, the total number of data points analyzed, the number of data points for which the transmittance was less than 0.95, the spacing between data points, and the number of iterations carried out by the program. The retrieved values together with the standard deviations obtained by the program are also shown, where, in all cases it was assumed that the Doppler width was known. In some of the retrievals the background parameter A_2 was constrained to lie between ± 0.001 cm; in other cases the upper and lower bounds were increased.

It is seen that essentially all the retrieved values agree with the initial values within two standard deviations. In some cases the standard deviations are large, either because the number of data points containing useful information was small or because the ratios of the spectral resolution or Lorentz width to the Doppler width of the line were too small.

These results were obtained by using the single precision version of the BMDP3R source program. It is obvious that this precision is adequate for the present work. New versions of the BMDP programs[4] which use double precision are now available. In addition to the double precision version of BMDP3R there is a new program BMDPAR which does not require the partial derivatives of the function with respect to each parameter. This greatly reduces the subroutine programming. We have compared results obtained by using the program BMDPAR and also the single precision and

double precision versions of BMDP3R both for Lorentz lines and Voigt lines. In analyzing synthetic spectra of Voigt lines, we have found no significant differences in the values of the parameters retrieved by the single and double precision versions of BMDP3R. In some cases the number of iterations required by the double precision version was less than that for the single precision version and there was a small decrease in the computer time required.

Experimental spectra of the P(1) and P(9) pairs of isotopic lines of the 1-0 band of HCl have also been analyzed. The spectra were obtained from a sample containing 0.421 atm cm (296 K) of HCl with N_2 added to give a total pressure of 700 Torr in a 171 m path at 296 K. In a previous analysis[11] of these spectra the lines were assumed to have a Lorentz shape and the single precision version of BMDP3R was used to retrieve the line parameters. Analyses of many pairs of lines in the band suggested the best value of the spectral resolution was 0.104 cm^{-1}. The Voigt expression approaches the Lorentz expression for large x and y values and the spectra have been reanalyzed by using the double precision versions of BMDP3R and BMDPAR. The results of several analyses of the P(1) and P(9) spectra are shown in Table 3 and Table 4 respectively. In some of the retrievals, the spectral resolution and the line intensity ratio were constrained as described in our previous analysis of these spectra.[11] The different analyses in each of these tables give essentially the same results and confirm that all the lines analyzed can be described adequately by the Lorentz shape.

The spectra of the P(1) lines have also been analyzed by using the derivative free BMDPAR program. Both Voigt and Lorentz line shapes have

been used. A comparison of the results obtained by using both BMDP3R and BMDPAR is given in Table 5. The results are essentially identical. It has been found that the BMDPAR program is the more expensive and that the initial values should be as close as possible to the best values for rapid convergence.

Examples of some of the data analyzed are shown in Figs. 1-4. The three lines in Fig. 1 have a pure Doppler shape, mixed Doppler-Lorentz shape, and a pure Lorentz shape, and their corresponding k(x,y) values lie entirely in the regions I, II, and III defined by Pierluissi et al. The parameters used to obtain these lines are given in the first three columns of Table 1.

Figures 2 and 3 show synthetic spectra with simulated 1% Gaussian random noise, the spectra calculated from the retrieved parameters, and the ratio and difference of the spectra which correspond to experiments 2a and 3 respectively in Table 2.

Figure 4 shows the experimental spectrum of the P(9) lines of HCl, the spectrum calculated from the constants obtained by assuming a Voigt profile and by using BMDPAR to analyze the data together with the ratio and difference of the observed and calculated spectra. Figure 5 shows the ranges of the values of the Voigt parameters x and y which correspond to the line parameters given in Tables 1-5.

Discussion

Absorption spectra of Voigt shaped lines obtained with an instrument of finite resolution must be modeled as complex convolutions of several analytical functions. These functions contain a small number of adjustable parameters. It has been shown that the values of these parameters

can often be retrieved with reasonable accuracy by spectral curve fitting of a limited amount of data from a single spectrum provided:

1. the frequency spacing between the experimental data points is comparable to, or smaller than, the widths of the lines,
2. the spectral resolution is comparable to the line width,
3. if the Lorentz width is to be determined, it should not be smaller than the Doppler width.

The synthetic spectra analyzed in this work correspond to the types of spectra obtained, for example, by high resolution Fourier Transform Spectrometers and the analysis was restricted to the retrieval of line parameters. Spectra with Voigt shaped lines can be collected by other types of instruments either for the purpose of obtaining information about line parameters or of the physical characteristics of the absorbing medium. Chang et al[1,2,3] have explored applications of spectral curve fitting to retrieving information from spectrum consisting of Lorentz shaped lines. The present work shows that these same techniques can now be used to analyze spectra with Voigt shaped lines.

In the process of exploring these techniques, we have attempted to determine, within the limitations imposed either by the available instrumentation or the nature of the absorbing sample, the amount of information contained in the spectrum and methods of optimizing the desired information content in the data collected. Thus, for example, if it is desired to measure the Lorentz width and intensity of an isolated spectral line, it is desirable to maximize the spectral resolution, the number of data points collected, and the signal-to-noise ratio in the spectrum, and to minimize the spacing between data points. Also the sample pressure should

be increased to make the Lorentz width large compared with the Doppler width. If these conditions are met, then the line will have a Lorentz shape and can be analyzed by the method of Chang et al.[1,2,3] However, the infrared spectra of many molecules consist of closely spaced lines and the sample pressure must be reduced to narrow the lines and minimize the amount of overlapping. Under these circumstances, the line may have a Voigt shape. Although spectra of finite resolution have been discussed, it is now possible to obtain spectra with essentially infinite resolution. In these cases, the true line profile is obtained and the spectral curve fitting technique will often allow the simultaneous determination of both the Lorentz and Doppler line widths, if desired.

The ability to analyze Voigt shaped lines will also allow, for example, solar spectra observed above the tropopause to be analyzed. The telluric lines in these spectra due to gases in the stratosphere and mesosphere have Voigt shapes and contain information about the vertical distributions of the absorbing gases and the pressure and temperature profiles of the atmosphere. These least squares techniques will allow sets of spectra to be analyzed separately or simultaneously, to retrieve the physical characteristics of the atmosphere by assuming the line parameters are known, and an appropriate model atmosphere.

REFERENCES

1. Y. S. Chang and J. H. Shaw, Appl. Spectros., 31, 213 (1977).

2. Y. S. Chang, J. H. Shaw, J. G. Calvert, and W. M. Uselman, JQSRT, 18, 589 (1977).

3. Y. S. Chang, J. H. Shaw, J. G. Calvert, and W. M. Uselman, JQSRT, 19, 599 (1977).

4. BMDP Biomedical Computer Programs P-Series 1977, Health Sciences Computing Facility, Department of Biomathematics, School of Medicine, UCLA.

5. C. Young, Tables for Calculating the Voigt Profile, Technical Report, ORA Project 05863, High Altitude Engineering Lab., Department of Aerospace Engineering, College of Engineering, The University of Michigan (1973).

6. B. H. Armstrong, JQSRT, 7, 61 (1967).

7. J. F. Kielkopf, J. Opt. Soc. Am., 63, 987 (1973).

8. S. R. Drayson, JQSRT, 16, 611 (1976).

9. J. H. Pierluissi, P. C. Vanderwood, and R. B. Gomez, JQSRT, 18, 555 (1977).

10. M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, p. 328 (1964) National Bureau of Standards.

11. C. L. Lin, E. Niple, J. H. Shaw, W. M. Uselman, and J. G. Calvert, Line Parameters of HCl Obtained by Simultaneous Analysis of Spectra, JQSRT, (to be published).

Table 1. Values of the parameters used to generate synthetic spectra to test the accuracy of the computer programs

Voigt Region	I	II	III	I, II, and III Combined	I, II, and III Combined
α_D (cm^{-1})	0.027506	0.0027506	0.0027506	0.0027506	0.0027506
α_L (cm^{-1})	0.000132	0.0070	0.03304	0.005616	0.002632
SU/π (cm^{-1})	0.012046	0.004015	0.037477	0.008031	0.010708
ν_o (cm^{-1})	0.7320	0.7320	0.7320	0.7320	0.7320
H (cm^{-1})	0.00060	0.00080	0.0500	0.008001	0.00200
A_1	5.274	5.274	5.274	5.274	5.274
A_2 (cm)	0.000	0.000	0.000	0.000	0.000
$\Delta\nu$ (cm^{-1})	0.00050	0.00020	0.015	0.0010847	0.001241
Total number of points	240	356	90	100	200
Number of points for $T < 0.95$	200	273	21	100	93
Number of iterations	32	25	11	45	15

Table 2. Initial values of the parameters used to calculate absorption spectra of single Voigt shaped lines. The parameter values and their standard deviations retrieved by using the single precision version of the program BMDP3R are also given.

Exp.	Data	$10^3\alpha_D$ (cm^{-1})	$10^3\alpha_L$ (cm^{-1})	$10^3SU/\pi$ (cm^{-1})	ν_o (cm^{-1})	A_1	A_2 (cm)	10^3H (cm^{-1})	Total No. Pts.	No. of Pts. Trans < 0.95	$10^3\Delta\nu$ (cm^{-1})	No of Iterations
1	Initial Value Retrieved 1σ	2.75 Fixed -	5.50 6.53 0.56	4.015 4.25 0 11	0.7320 0.7318 0.0001	5.274 5.295 0.015	0 -0.001 -	5.50 2.6 4.6	50	20	2.20	9
2a	Initial Value Retrieved 1σ	2.75 Fixed -	5.50 6.68 0.71	8.031 8.32 0.12	0.7320 0.7318 0.0001	5.274 5.317 0.023	0 -0.00 0.22	16.5 15.5 1.2	50	28	2.20	7
2b	Initial Value Retrieved 1σ	2.75 Fixed -	5.50 5.32 0.49	8 031 8.09 0.10	0.7320 0.7321 0.0001	5.274 5.267 0.009	0 +0.001 -	16.5 16.7 0.9	100	28	2.20	11
3	Initial Value Retrieved 1σ	4.00 Fixed -	8.00 8.10 0.64	8.031 8.13 0.13	0.7320 0.7324 0.0002	5.274 5.269 0.007	0 0.00 0.03	12.0 11.2 2.3	100	14	5.00	5
4	Initial Value Retrieved 1σ	4.00 Fixed -	12.0 11.9 0.9	12.05 12.15 0.16	0.7320 0.7324 0.0002	5.274 5.268 0.008	0 0.02 0.03	20.0 20.0 0.2	100	21	5.00	7
5	Initial Value Retrieved 1σ	3.00 Fixed -	12.0 11.4 0.9	12.05 12.1 0.2	0.7320 0.7315 0.0002	5.274 5.270 0.004	0 0.00 0.01	24.0 24.3 0.2	200	16	6.00	8
6	Initial Value Retrieved 1σ	27.5 Fixed -	0.132 0.01 0.50	12.05 12.02 0.10	0.7320 0.7320 0.0000	5.274 5.271 0.008	0 0.00 0.03	0.606 3.6 10.0	240	210	7.60	24

Table 3. Comparison of HCl P(1) line parameters retrieved by using the single and double precision versions of BMDP3R for Lorentz and Voigt shapes

Exp.	Iterations	Precision	Line Shape	Constraints	Residual	A_1	A_2 (cm)	S_1U/π (cm^{-1})	S_2U/π (cm^{-1})	α_L§ (cm^{-1})	ν_1 (cm^{-1})	ν_2 (cm^{-1})	H (cm^{-1})
A	9	Double	Voigt	S_2/S_1* = 3.0186 H = 0.104	2.8265	5.275 ±0.005	0.0006 ±0.0007	0.187 ±0.003	0.577 ±0.010	0.091 ±0.002	2863.014 ±0.002	2865.093 ±0.002	0.104 -
B	12	Double	Voigt	None	2.8390	5.276 ±0.005	0.0006 ±0.0007	0.185 ±0.003	0.559 ±0.027	0.095 ±0.006	2863.014 ±0.002	2865.093 ±0.002	0.089 ±0.003
C	9	Double	Lorentz	S_2/S_1 = 3.0186 H = 0.104	2.8424	5.275 ±0.005	0.0006 ±0.0007	0.187 ±0.003	0.577 ±0.010	0.091 ±0.002	2863.014 ±0.002	2865.093 ±0.002	0.104 -
D	12	Double	Lorentz	None	2.8390	5.276 ±0.005	0.0006 ±0.0007	0.185 ±0.004	0.559 ±0.027	0.095 ±0.006	2863.014 ±0.002	2865.093 ±0.002	0.089 ±0.003
E	26	Single	Voigt	S_2/S_1 = 3.0186 H = 0.104	2.8425	5.275 ±0.005	0.0006 ±0.0007	0.187 ±0.003	0.577 ±0.010	0.091 ±0.002	2863.014 ±0.002	2865.093 ±0.002	0.104 -
F	18	Single	Voigt	None	2.8392	5.276 ±0.005	0.0006 ±0.0007	0.185 ±0.004	0.559 ±0.027	0.095 ±0.006	2863.014 ±0.002	2865.093 ±0.002	0.089 ±0.003
G	26	Single	Lorentz	S_2/S_1 = 3.0186 H = 0.104	2.8423	5.275 ±0.005	0.0006 ±0.0007	0.187 ±0.003	0.577 ±0.010	0.091 ±0.002	2863.014 ±0.002	2865.093 ±0.002	0.104 -
H	17	Single	Lorentz	None	2.8389	5.276 ±0.005	0.0006 ±0.0007	0.185 ±0.004	0.559 ±0.027	0.095 ±0.006	2863.014 ±0.002	2865.093 ±0.002	0.089 ±0.003

*S_2/S_1 = ratio of the band intensities of $H^{35}Cl$ and $H^{37}Cl$

§700 Torr H_2 pressure-broadened width

Table 4. Comparison of HCl P(9) line parameters retrieved by using the single and double precision versions of BMDP3R for Lorentz and Voigt shapes

Exp	Iterations	Precision	Line Shape	Constraints	Residual	A_1	A_2 (cm)	S_1U/π (cm^{-1})	S_2U/π (cm^{-1})	α_L§ (cm^{-1})	ν_1 (cm^{-1})	ν_2 (cm^{-1})	H (cm^{-1})
I	7	Double	Voigt	S_2/S_1 x = 3.0186 H = 0.104	1.8448	5.311 ±0.005	0.001 ±0.001	0.0247 ±0.0017	0.076 ±0.005	0.014 ±0.002	2675.952 ±0.003	2677.728 ±0.002	0.104 -
J	9	Double	Voigt	None	1.7753	5.312 ±0.005	0.002 ±0.001	0.0213 ±0.0010	0.058 ±0.005	0.023 ±0.005	2675.952 ±0.003	2677.728 ±0.002	0.084 ±0.008
K	7	Double	Lorentz	S_2/S_1 = 3.0186 H = 0.104	1.8453	5.311 ±0.005	0.001 ±0.001	0.0248 ±0.0018	0.077 ±0.006	0.014 ±0.002	2675.952 ±0.003	2677.728 ±0.002	0.104 -
L	9	Double	Lorentz	None	1.7751	5.312 ±0.005	0.002 ±0.001	0.0213 ±0.0010	0.058 ±0.005	0.023 ±0.005	2675.952 ±0.003	2677.728 ±0.002	0.084 ±0.008
M	9	Single	Voigt	S_2/S_1 = 3.0186 H = 0.104	1.8447	5.311 ±0.005	0.001 ±0.001	0.0247 ±0.0017	0.076 ±0.005	0.014 ±0.002	2675.952 ±0.003	2677.728 ±0.002	0.104 -
N	10	Single	Voigt	None	1.7752	5.312 ±0.005	0.002 ±0.001	0.0213 ±0.0010	0.058 ±0.005	0.023 ±0.005	2675.952 ±0.003	2677.728 ±0.002	0.084 ±0.008
O	9	Single	Lorentz	S_2/S_1 = 3.0186 H = 0.104	1.8452	5.311 ±0.005	0.001 ±0.001	0.0248 ±0.0018	0.077 ±0.006	0.014 ±0.002	2675.952 ±0.003	2677.728 ±0.002	0.104 -
P	10	Single	Lorentz	None	1.7750	5.312 ±0.005	0.002 ±0.001	0.0213 ±0.0010	0.058 ±0.005	0.023 ±0.005	2675.952 ±0.003	2677.728 ±0.002	0.084 ±0.008

x S_2/S_1 = ratio of band intensities of $H^{35}Cl$ and $H^{37}Cl$

§ 700 Torr N_2 pressure-broadened width

Table 5. Comparison of HCl P(1) line parameters retrieved by using BMDPAR and BMDP3R in double precision for Voigt shape

Program	Constraints	Iterations	Residual	A_1	A_2 (cm)	$S_1 U/\pi$ (cm^{-1})	$S_2 U/\pi$ (cm^{-1})	α_L* (cm^{-1})	ν_1 (cm^{-1})	ν_2 (cm^{-1})	H (cm^{-1})
BMDP3R	S_2/S_1 = 3.0816 H = 0.104	9	2.8427	5.275 ±0.005	0.0006 ±0.0007	0.187 ±0.003	0.577 ±0.010	0.091 ±0.002	2863.014 ±0.002	2865.093 ±0.002	0.104 -
BMDPAR	S_2/S_1 = 3.0816 H = 0.104	39	2.8647	5.277 ±0.005	-0.0006 ±0.0001	0.188 ±0.004	0.570 ±0.016	0.092 ±0.003	2863.014 ±0.002	2865.093 ±0.001	0 104 -
BMDP3R	None	12	2.8390	5.276 ±0.005	0.0006 ±0.0007	0.186 ±0.003	0.559 ±0.027	0.095 ±0.006	2863.014 ±0.002	2865.093 ±0.002	0.089 ±0 003
BMDPAR	None	37	2.8393	5.276 ±0.005	0.0006 ±0.0004	0.186 ±0.003	0.559 ±0.022	0.095 ±0.005	2863.014 ±0.002	2865.093 ±0.002	0 089 ±0.002

*700 Torr N_2 pressure-broadened width

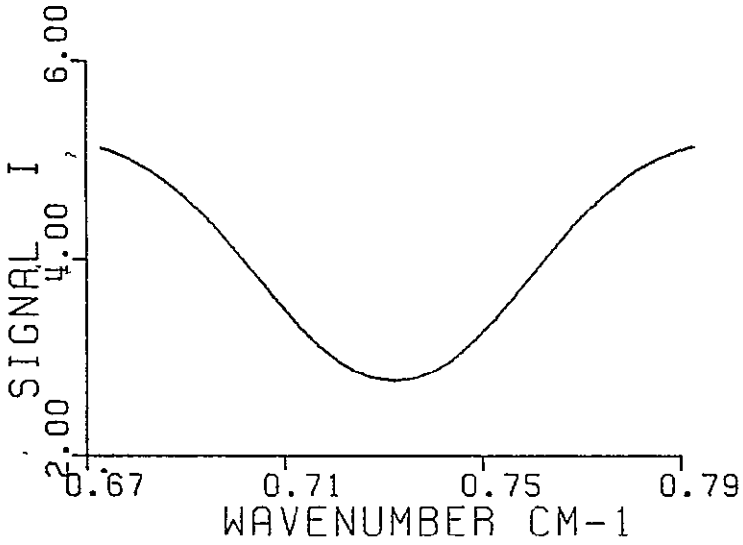
Fig. 1 - Noise-free spectra and absorption coefficients of Voigt shaped lines calculated for the parameter values in Table 1. Top curve: Region I line; middle curve: Region II line; bottom curve: Region III line.

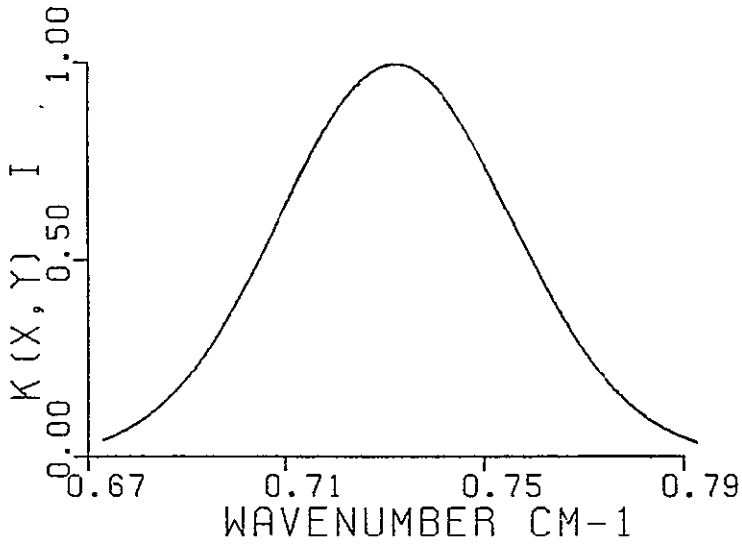
Fig. 2 - The top curve shows the spectrum with 1% added noise calculated from the parameter values of Experiment 2a in Table 2, the next curve is the spectrum calculated from the retrieved values. The bottom curves show the ratio and differences of these spectra.

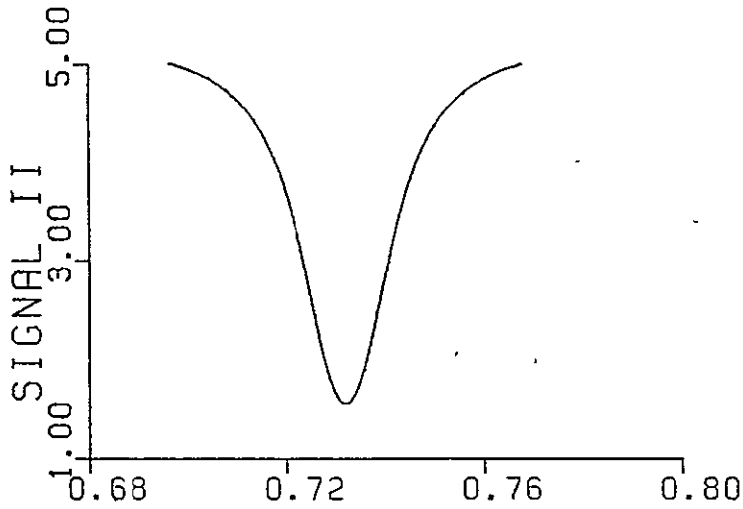
Fig. 3 - The top curve shows the spectrum with 1% added noise calculated from the parameter values of Experiment 3 in Table 2, the next curve is the spectrum calculated from the retrieved values. The bottom curves show the ratio and difference of these spectra.

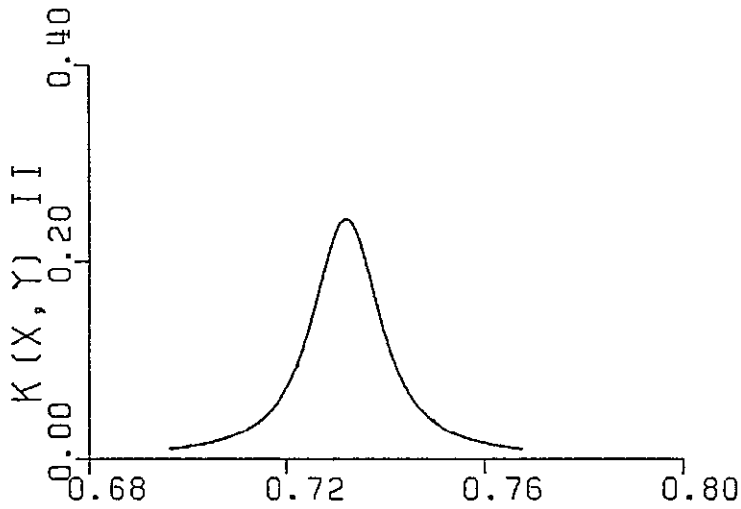
Fig. 4 - The top curve shows part of an experimental spectrum near the P(9) lines of the 1-0 band of HCl for a sample containing 0.421 atm-cm (296 K) of HCl with N_2 added to give a total pressure of 700 Torr. Beneath it is a spectrum calculated from the retrieved values of the parameters given in Table 4. The bottom curve shows the ratio and difference of these spectra. 2800 cm^{-1} should be added to the abscissa values.

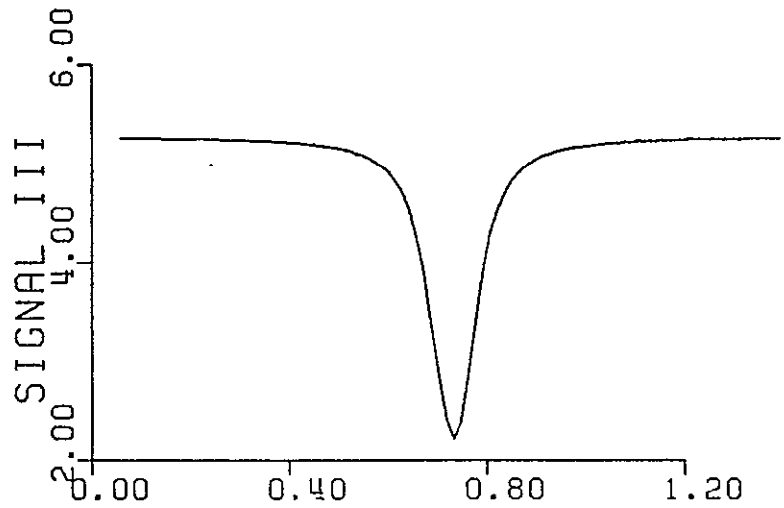
Fig. 5 - The ranges of the values of the Voigt parameters x and y corresponding to the line parameters in Tables 1-5. The boundaries of the three regions defined by Pierluissi et al, are also shown.


SIGNAL I
6.00
4.00
2.00
0.67
0.71
0.75
0.79
WAVENUMBER CM-1


K(X,Y) I
1.00
0.50
0.00
0.67
0.71
0.75
0.79
WAVENUMBER CM-1


SIGNAL II
5.00
3.00
1.00
0.68
0.72
0.76
0.80


K(X,Y) II
0.40
0.20
0.00
0.68
0.72
0.76
0.80


SIGNAL III
6.00
4.00
2.00
0.00
0.40
0.80
1.20

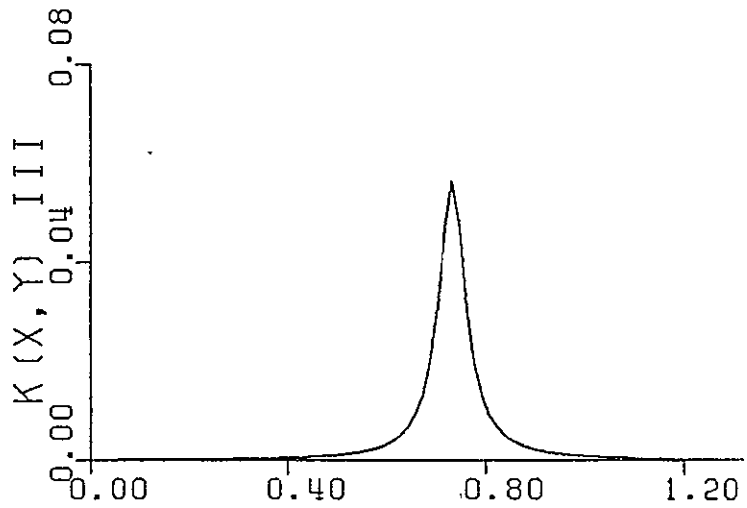

K(X,Y) III
0.08
0.04
0.00
0.00
0.40
0.80
1.20

Fig 1

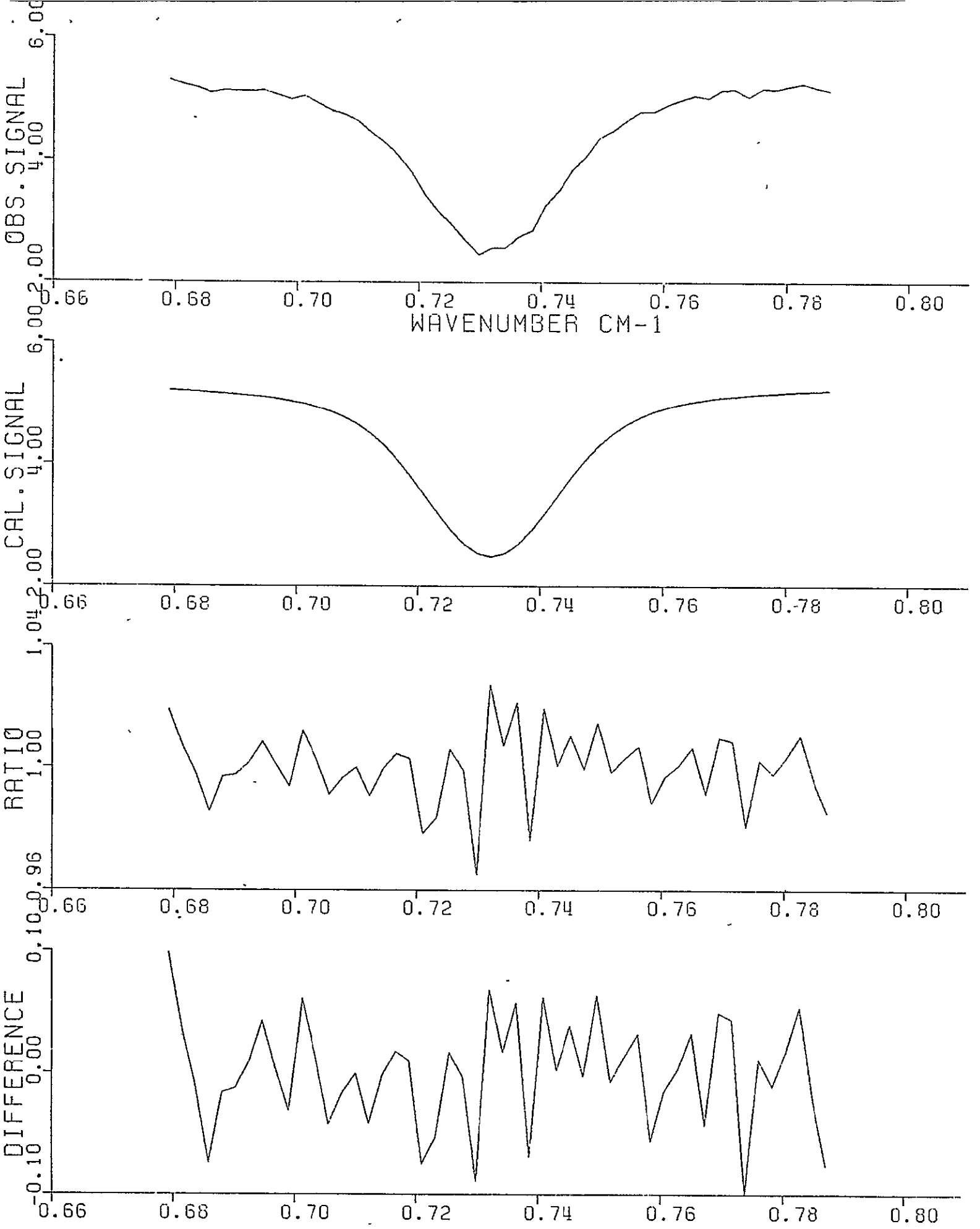

OBS. SIGNAL
6.00
4.00
2.00
0.66
0.68
0.70
0.72
0.74
0.76
0.78
0.80
WAVENUMBER CM-1
CAL. SIGNAL
6.00
4.00
2.00
0.66
0.68
0.70
0.72
0.74
0.76
0.78
0.80
RATIO
1.04
1.00
0.96
0.66
0.68
0.70
0.72
0.74
0.76
0.78
0.80
DIFFERENCE
0.10
0.00
-0.10
0.66
0.68
0.70
0.72
0.74
0.76
0.78
0.80

Fig. 2.

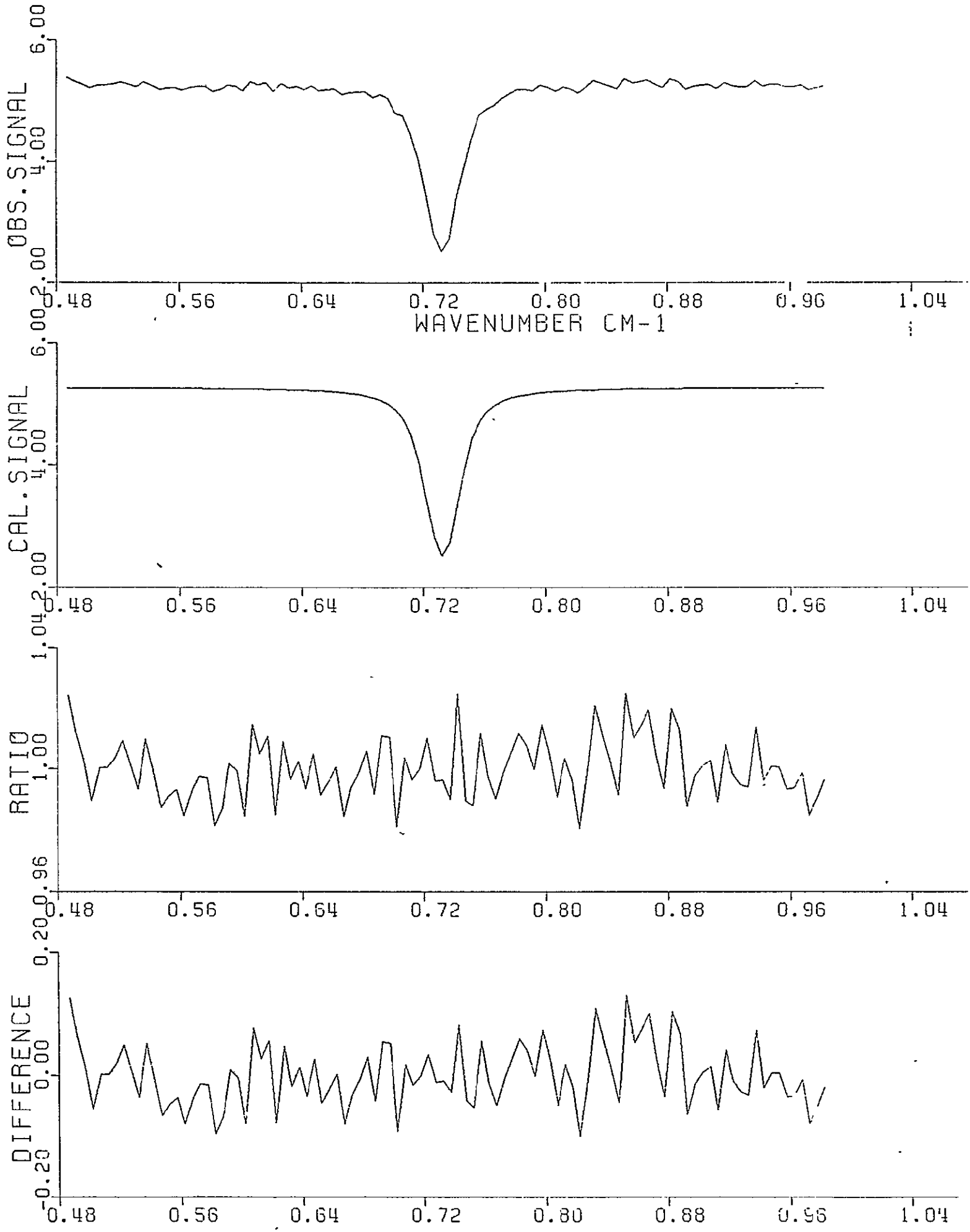

OBS. SIGNAL
6.00
4.00
2.00
0.48
0.56
0.64
0.72
0.80
0.88
0.96
1.04
WAVENUMBER CM-1
CAL. SIGNAL
6.00
4.00
2.00
0.48
0.56
0.64
0.72
0.80
0.88
0.96
1.04
RATIO
1.04
1.00
0.96
0.48
0.56
0.64
0.72
0.80
0.88
0.96
1.04
DIFFERENCE
0.20
0.00
-0.20
0.48
0.56
0.64
0.72
0.80
0.88
0.96
1.04

Fig 3

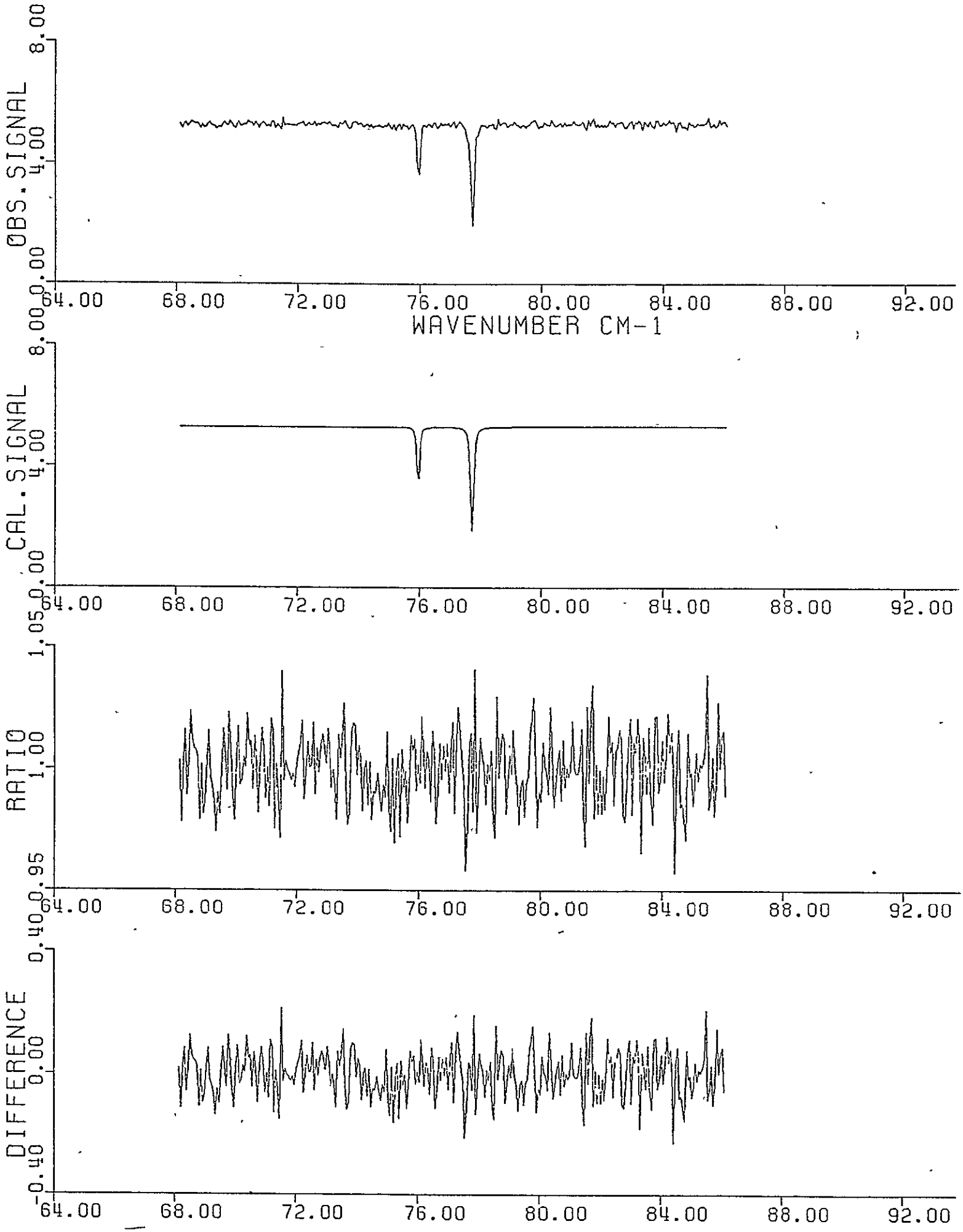

OBS. SIGNAL
8.00
4.00
0.00
64.00
68.00
72.00
76.00
80.00
84.00
88.00
92.00
WAVENUMBER CM-1
CAL. SIGNAL
8.00
4.00
0.00
64.00
68.00
72.00
76.00
80.00
84.00
88.00
92.00
RATIO
1.05
1.00
0.95
64.00
68.00
72.00
76.00
80.00
84.00
88.00
92.00
DIFFERENCE
0.40
0.00
-0.40
64.00
68.00
72.00
76.00
80.00
84.00
88.00
92.00

Fig 4

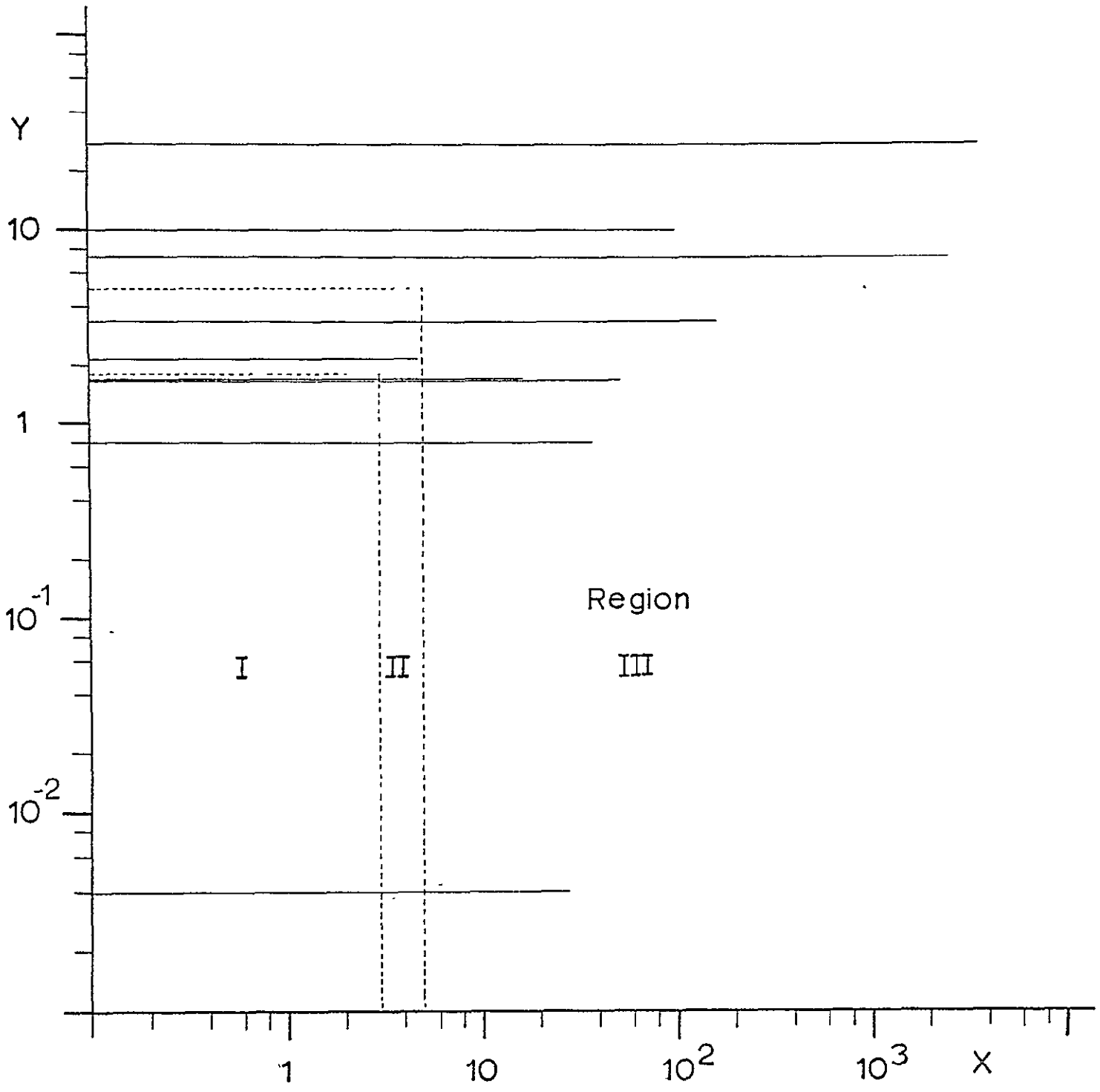


Fig. 5

APPENDIX I

SUBROUTINE FOR USE WITH BMDP3R TO RETRIEVE THE PARAMETERS OF TEN VOIGT SHAPED LINES

```
// TIME=12,REGION=500K
//CMP    EXEC    PGM=IFEAAB,PARM=ID,TIME=(0,30)
//SYSLIN  DD  UNIT=SYSDA,DSNAME=&&OBJ,SPACE=(CYL,(1,1)),
//            DISP=(MOD,PASS),DCB=(RECFM=FB,LRECL=80,BLKSIZE=3120)
//SYSPRINT DD SYSOUT=A,DCB=BUFNO=1
//CMP.SYSIN  DD  *
      SUBROUTINE FUN(F,DF,P,X,N,KASE,NVAR,NPAR,IPASS,XLOSS)
      REAL*8 DF(NPAR),P(NPAR),X(NVAR)
      REAL*8 B(32),ST(32),CM(32),XO,DOP(2),CONST(2),PC1,PC2,PCC1,PCC2,
     *BACK,YEU(2),U(2),V(2),S(2),T(2),USER(2),VSER(2),UN(2),VN(2),U2(2),
     *V2(2),UNEW(2),VNEW(2),DE(2),Q1(2),Q2(2),G1(2),G2(2),VOIGT(2),
     *R1(2),UR1(2),UR2(2),RDER(2),UDER(2),R(2),E(2),G(2),H(2),AKU(32),
     *TM(32),RR1,DFF(8)
      COMPLEX*16 Z2(2),ZA1(2),ZA3(2),ZA5(2),
     *ZDER2(2),ZDER3(2),ZB1(2),ZB3(2)
      REAL*8 AN(30)/1.00000000,-.33333333,.10000000,-.238095238E-1,
     *  4.629629629E-3,-7.575757576E-4,1.068376068E-4,-1.522751323E-5,
     *  1.45891690E-6,-1.450385222E-7,1.312253296E-8,-1.089222104E-9,
     *  8.350702795E-11,-5.947794014E-12,3.955429516E-13,
     *  -2.406827016E-14,1.448326464E-15,-8.032735012E-17,
     *  4.221407289E-18,-2.107855191E-19,
     *  1.002516404E-20,-4.551846759E-22,1.977064754E-23,
     *-8.230149299E-25,3.289260349E-26,-1.264107899E-27,4.678483526E-29,
     *  -1.669761793E-30,5.754191644E-32,-1.916942862E-33/,
     *  A1/.46131350/,
     *  A2/.19016350/,A3/.09999216/,A4/1.78449270/,A5/.002893894/,
     *  A6/5.52534370/,B1/.51242424/,B2/.27525510/,B3/.05176536/,
     *  B4/2.72474500/,PISQ/1.128379167/,PISQRT/1.77245385091/
C     FIND LINE PARAMETERS AND BACKGROUND USING VOIGT PROFILE AND
C     TRIANGULAR SLIT FUNCTION
C     F=BACKGROUND*CONVOLVED TRANSMITTANCE
C     BACK=BACKGROUND=(A1+A2*(X(1)-XO))
C     P(1)=A1
C     P(2)=A2
C     P(3)=(LINE#1 INTENSITY)*(CONCENTRATION*PATH LENGTH)/PI
C     P(4)=(LINE#2 INTENSITY)*(CONCENTRATION*PATH LENGTH)/PI
C     P(5)=(LINE#1 LORENTZIAN HALF WIDTH)=(LINE#2 LORENTZIAN HALF WIDTH)
C     P(6)=WAVENUMBER OF LINE CENTER OF LINE#1
C     P(7)=WAVENUMBER OF LINE CENTER OF LINE#2
C     P(8)=SLIT FUNCTION FULL HALF WIDTH
C     DOP(1)=DOPPLER HALF WIDTH OF LINE#1
C     DOP(2)=DOPPLER HALF WIDTH OF LINE#2
C     CONST(1)=SQRT(LN(2))/DOP(1)
C     CONST(2)=SQRT(LN(2))/DOP(2)
      XO=76.343000
      DOP(1)=0.2675490-2
      DOP(2)=0.2750570-2
      CONST(1)=3.111781002
      CONST(2)=3.026840002
      PC1=P(3)*CONST(1)
      PC2=P(4)*CONST(2)
      PCC1=PC1*CONST(1)
      PCC2=PC2*CONST(2)
      BACK=(P(1)+P(2)*(X(1)-XO))
      F=0.00
      DO 1 I=1,8
      DF(I)=0.00
    1 CONTINUE
      DO 2 K=1,32
      B(K)=(33.00-2*K)/32.00
```

```
2    CONTINUE
     DO 3 I=1,31,2
     N=(I+1)/2
     ST(N)=I/32.D0
     ST(N+16)=(32.D0-I)/32.D0
3    CONTINUE
     DO 4 K=1,32
     CM(K)=(X(1)+B(K)*P(8))
     YEU(1)=DABS(CM(K)-P(6))
     YEU(2)=DABS(CM(K)-P(7))
     DO 5 J=1,2
     U(J)=YEU(J)*CONST(J)
     V(J)=P(5)*CONST(J)
     S(J)=U(J)*U(J)-V(J)*V(J)
     T(J)=2.D0*U(J)*V(J)
     Z2(J)=DCMPLX(S(J),T(J))
     IF(V(J).EQ.0.0D0) GO TO 6
     IF(V(J).GE.5.0D0.OR.U(J).GE.5.0D0) GO TO 7
     IF(V(J).GE.1.8D0.OR.U(J).GE.3.0D0) GO TO 8
6    CONTINUE
     USER(J)=V(J)
     VSER(J)=-U(J)
     UN(J)=V(J)
     VN(J)=-U(J)
     U2(J)=-S(J)
     V2(J)=-T(J)
     M=6.842D0*U(J)+8.0D0
     IF(M.GT.29) M=29
     IF(U(J).EQ.0.0D0) M=15
     DO 9 I9=1,M
     UNEW(J)=UN(J)*U2(J)-VN(J)*V2(J)
     VNEW(J)=V2(J)*UN(J)+VN(J)*U2(J)
     USER(J)=USER(J)+UNEW(J)*AN(I9+1)
     VSER(J)=VSER(J)+VNEW(J)*AN(I9+1)
     UN(J)=UNEW(J)
     VN(J)=VNEW(J)
9    CONTINUE
     DE(J)=DEXP(-S(J))
     Q1(J)=(U(J)*DCOS(-T(J))-V(J)*DSIN(-T(J)))
     Q2(J)=(V(J)*DCOS(-T(J))+U(J)*DSIN(-T(J)))
     G1(J)=(1.D0-PISQ*USER(J))
     G2(J)=PISQ*VSER(J)
     VOIGT(J)=DE(J)*(DCOS(-T(J))*G1(J)+DSIN(-T(J))*G2(J))
     R1(J)=(-2.D0)*DE(J)*(Q1(J)*G1(J)+Q2(J)*G2(J))
     UR1(J)=(-2.D0)*DE(J)*(Q1(J)*(-G2(J))+Q2(J)*G1(J))
     UR2(J)=PISQ
     RDER(J)=R1(J)
     UDER(J)=UR1(J)+UR2(J)
     GO TO 5
8    R(J)=T(J)*T(J)
     T(J)=T(J)*U(J)
     E(J)=S(J)-A6
     G(J)=S(J)-A4
     H(J)=S(J)-A2
     ZA1(J)=Z2(J)-A2
     ZA3(J)=Z2(J)-A4
     ZA5(J)=Z2(J)-A6
     ZDER2(J)=(A1/ZA1(J)+A3/ZA3(J)+A5/ZA5(J))-
    *       (2.D0*Z2(J))*(A1/ZA1(J)**2+A3/ZA3(J)**2+A5/ZA5(J)**2)
     VOIGT(J)=A1*((T(J)-H(J)*V(J))/(H(J)*H(J)+R(J)))+
```

```
     *       A3*((T(J)-G(J)*V(J))/(G(J)*G(J)+R(J)))+
     *       A5*((T(J)-E(J)*V(J))/(E(J)*E(J)+R(J)))
      RDER(J)=-DIMAG(ZDER2(J))
      UDER(J)=DREAL(ZDER2(J))
      GO TO 5
  7   R(J)=T(J)*T(J)
      T(J)=T(J)*U(J)
      E(J)=S(J)-B2
      G(J)=S(J)-B4
      ZB1(J)=Z2(J)-B2
      ZB3(J)=Z2(J)-B4
      ZDER3(J)=(B1/ZB1(J)+B3/ZB3(J))-
     *       (2.D0*Z2(J))*(B1/ZB1(J)**2+B3/ZB3(J)**2)
      VOIGT(J)=B1*((T(J)-E(J)*V(J))/(E(J)*E(J)+R(J)))+
     *       B3*((T(J)-G(J)*V(J))/(G(J)*G(J)+R(J)))
      RDER(J)=-DIMAG(ZDER3(J))
      UDER(J)=DREAL(ZDER3(J))
  5   CONTINUE
      AKU(K)=(PC1*VOIGT(1)+PC2*VOIGT(2))*PISQRT
      TM(K)=DEXP(-AKU(K))
      RR1=BACK*ST(K)*TM(K)/16.D0
      DFF(3)=-VOIGT(1)*CONST(1)*PISQRT
      DFF(4)=-VOIGT(2)*CONST(2)*PISQRT
      DFF(5)=(PCC1*UDER(1)+PCC2*UDER(2))*PISQRT
      IF(CM(K).EQ.P(6)) GO TO 14
      IF(CM(K).LT.P(6)) GO TO 11
      DFF(6)=PCC1*RDER(1)*PISQRT
      GO TO 12
 11   DFF(6)=-PCC1*RDER(1)*PISQRT
      GO TO 12
 14   DFF(6)=0.0D0
 12   CONTINUE
      IF(CM(K).EQ.P(7)) GO TO 18
      IF(CM(K).LT.P(7)) GO TO 15
      DFF(7)=PCC2*RDER(2)*PISQRT
      GO TO 16
 15   DFF(7)=-PCC2*RDER(2)*PISQRT
      GO TO 16
 18   DFF(7)=0.0D0
 16   CONTINUE
      DFF(8)=-(DFF(6)+DFF(7))*B(K)
      F=F+RR1
      DF(1)=DF(1)+ST(K)*TM(K)/16.D0
      DF(2)=DF(2)+ST(K)*TM(K)/16.D0*(X(1)-X0)
      DF(3)=DF(3)+RR1*DFF(3)
      DF(4)=DF(4)+RR1*DFF(4)
      DF(5)=DF(5)+RR1*DFF(5)
      DF(6)=DF(6)+RR1*DFF(6)
      DF(7)=DF(7)+RR1*DFF(7)
      DF(8)=DF(8)+RR1*DFF(8)
  4   CONTINUE
      RETURN
      END
//GO      EXEC   PGM=LOADER,PARM='MAP,EP=MAIN',COND=(8,LT),TIME=(11,30)
//GO.FT01F001 DD UNIT=SYSDA,SPACE=(CYL,(1,1)),
//             DCB=(RECFM=VBS,LRECL=324,BLKSIZE=2298)
//FT02F001  DD  UNIT=SYSDA,SPACE=(CYL,(1,1))
//FT05F001   DD DDNAME=SYSIN
//FT06F001   DD SYSOUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=931,BUFNO=1)
//SYSLIB    DD  DSNAME=SYS1.FORTLIB,DISP=SHR
```

```
//GO.SYSLIN DD DSN=&&OBJ,DISP=(OLD,DELETE)
//    DD         DSNAME=SYS4.BMDPLIB(BMDP3R),DISP=SHR
//SYSLOUT  DD   SYSOUT=A,DCB=BUFNO=1
//GO.SYSIN  DD  *
PROB TITLE IS 'TEST OF VOIGT LINE'./
INPUT VARIABLES ARE 2.
      FORMAT IS '(F10.4,5X,F10.3)'.
      UNIT=10.
      CASE ARE 299./
VARIABLE NAMES ARE X,F./
REGRESSION TITLE IS 'LINE PARAMETERS OF DOUBLE VOIGT LINE '.
      DEPEND=F.
      TOL=0.00000001.
      PARM=8.
      CONST=2.
      ITER=45.
      HALF=20./
PARM  INIT=5.24+,0.0,0.019,0.0617,0.02,75.95,77.7,0.103
     MAXIMUMS ARE 6.0,0.1,0.3,0.6,0.14,77.0,79.0,0.9
     MINIMUMS ARE 4.0,-0.1,1.E-7,0.00001,0.0001,71.0,73.0,0.0001
      CONST=(4)1,(3)-3.0866. K=0.
      CONST=(8)1. K=0.104.
     NAMES=A1,A2,SUOVPI1,SUOVPI2,LORENTZ,NEU1,NEU2,H./
END/
FINISH/
//GO.FT10F001  DD  DSN=TSO744.HCL.DATA(LIN02),DISP=SHR
//
```

APPENDIX II

SUBROUTINE FOR USE WITH BMDPAR TO RETRIEVE THE PARAMETERS OF A SINGLE VOIGT LINE

```
// TIME=6,REGION=500K
//CMP    EXEC    PGM=IFEAAB,PARM=ID,TIME=(0,30)
//SYSLIN  DD  UNIT=SYSDA,DSNAME=&&OBJ,SPACE=(CYL,(1,1)),
//            DISP=(MOD,PASS),DCB=(RECFM=FB,LRECL=80,BLKSIZE=3120)
//SYSPRINT DD SYSOUT=A,DCB=BUFNO=1
//CMP.SYSIN  DD  *
      SUBROUTINE FUN(F,P,X,N,KASE,NVAR,NPAR,IPASS,XLOSS)
      REAL*8 P(NPAR),X(NVAR)
      REAL*8 B(32),ST(32),CM(32),XO,DOP,CONST,PC1,PCC1,BACK,YEU,U,V,S,T,
     *       USER,VSER,UN,VN,U2,V2,UNEW,VNEW,DE,G1,G2,VOIGT,R1,
     *       R,E,H,G,AKU(32),TM(32),RR1
      REAL*8 AN(30)/1.00000000,-.33333333,.10000000,-.238095238E-1,
     *  4.62962963E-3,-7.57575757E-4,1.068376068E-4,-1.322751323E-5,
     *  1.45891690E-6,-1.450385222E-7,1.312253296E-8,-1.089222104E-9,
     *  8.350702795E-11,-5.947794014E-12,3.955429516E-13,
     *  -2.46682701E-14,1.448326464E-15,-8.032735012E-17,
     *  4.221407289E-18,-2.107855191E-19,
     *  1.002516494E-20,-4.551846759E-22,1.977064754E-23,
     *-8.230149299E-25,3.289260349E-26,-1.264107899E-27,4.67848352E-29,
     *  -1.669761793E-30,5.754191644E-32,-1.916942862E-33/,
     *  A1/.46131350/,
     *  A2/.19016350/,A3/.09999216/,A4/1.78449270/,A5/.002883894/,
     *  A6/5.52534370/,B1/.51242424/,B2/.27525510/,B3/.05176536/,
     *  B4/2.72474500/,PISQ/1.128379167/,PISQRT/1.77245385091/
C     FIND LINE PARAMETERS AND BACKGROUND USING VOIGT PROFILE AND
C     TRIANGULAR SLIT FUNCTION
C     F=BACKGROUND*CONVOLVED TRANSMITTANCE
C     BACK=BACKGROUND=(A1+A2*(X(1)-XO))
C     P(1)=A1
C     P(2)=A2
C     P(3)=(LINE INTENSITY)*(CONCENTRATION*PATH LENGTH)/PI
C     P(4)=LORENTZIAN LINE HALF WIDTH
C     P(5)=WAVENUMBER OF LINE CENTER
C     P(6)=SLIT FUNCTION FULL HALF WIDTH
C     DOP=DOPPLER HALF WIDTH OF THE LINE
C     CONST=SQRT(LN(2))/DOP
      XO=0.732D0
      DOP=0.275D-2
      CONST=3.027471313D02
      PC1=P(3)*CONST
      PCC1=PC1*CONST
      BACK=(P(1)+P(2)*(X(1)-XO))
      F=0.D0
      DO 2 K=1,32
      B(K)=(33.D0-2*K)/32.D0
  2   CONTINUE
      DO 3 I=1,31,2
      N=(I+1)/2
      ST(N)=I/32.D0
      ST(N+16)=(32.D0-I)/32.D0
  3   CONTINUE
      DO 4 K=1,32
      CM(K)=(X(1)+B(K)*P(6))
      YEU=DABS(CM(K)-P(5))
      U=YEU*CONST
      V=P(4)*CONST
      S=U*U-V*V
      T=2.D0*U*V
      IF(V.EQ.0.0D0) GO TO 6
      IF(V.GE.5.0D0.OR.U.GE.5.0D0) GO TO 7
```

```
      IF(V.GE.1.800.OR.U.GE.3.000) GO TO 8
  6   CONTINUE
      USER=V
      VSER=-U
      UN=V
      VN=-U
      U2=-S
      V2=-T
      M=6.84200*U+8.000
      IF(M.GT.29) M=29
      IF(U.EQ.0.000) M=15
      DO 9 I9=1,M
      UNEW=UN*U2-VN*V2
      VNEW=V2*UN+VN*U2
      USER=USER+UNEW*AN(I9+1)
      VSER=VSER+VNEW*AN(I9+1)
      UN=UNEW
      VN=VNEW
  9   CONTINUE
      DE=DEXP(-S)
      G1=(1.D0-PISQ*USER)
      G2=PISQ*VSER
      VOIGT=DE*(DCOS(-T)*G1+DSIN(-T)*G2)
      GO TO 5
  8   R=T*T
      T=T*U
      E=S-A6
      G=S-A4
      H=S-A2
      VOIGT=A1*((T-H*V)/(H*H+R))+A3*((T-G*V)/(G*G+R))+
     *    A5*((T-E*V)/(E*E+R))
      GO TO 5
  7   R=T*T
      T=T*U
      E=S-B2
      G=S-B4
      VOIGT=B1*((T-E*V)/(E*E+R))+B3*((T-G*V)/(G*G+R))
  5   CONTINUE
      AKU(K)=PC1*VOIGT*PISQRT
      TM(K)=DEXP(-AKU(K))
      RR1=BACK*ST(K)*TM(K)/16.D0
      F=F+RR1
  4   CONTINUE
      RETURN
      END
//GO      EXEC   PGM=LOADER,PARM='MAP,EP=MAIN',COND=(8,LT),TIME=(5,30)
//FT01F001  DD  UNIT=SYSDA,SPACE=(CYL,(1,1))
//FT02F001  DD  UNIT=SYSDA,SPACE=(CYL,(1,1))
//FT05F001    DD DDNAME=SYSIN
//FT06F001    DD SYSOUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=931,BUFNO=1)
//SYSLIB    DD  DSNAME=SYS1.FORTLIB,DISP=SHR
//SYSLIN    DD   DSNAME=&&OBJ,DISP=(OLD,DELETE)
//    DD         DSNAME=SYS4.BMDPLIB(BMDPAR),DISP=SHR
//SYSLOUT  DD   SYSOUT=A,DCB=BUFNO=1
//GO.SYSIN  DD  *
/PROBLEM TITLE IS 'TEST'.
/INPUT VARIABLES ARE 2.
      FORMAT IS '(F10.4,5X,F10.7)'.
      UNIT=10.
      CASE ARE 50.
```

```
/VARIABLE NAMES ARE X,F.
/REGR TITLE IS 'LINE PARAMETER OF SINGLE VOIGT LINE'.
      DEPEND=F.
      TOL=0.00000001.
      PARM=6.
      ITER=45.
      HALF=20.
/PARAM     INIT=5.258,0.0,0.0075,0.0046,0.729,0.0152
           MAX ARE 6.0,0.1,0.120,0.0990,0.9320,0.9
           MIN ARE 4.0,-0.1,1.E-7,0.00001,0.3320,0.00001
      NAME=A1,A2,SUOVRPI,LORENTZ,NEUD,H.
/END
/FINISH
//GO.FT10F001  DD  DSN=TS0744.HCL.DATA(ANN),DISP=SHR
//
```

Reprinted from the Proceedings of the

Thirty-Fourth symposium on Molecular Spectroscopy
The Ohio State University
Columbus, Ohio
June 11-15, 1979

RA9.

SPECTRAL ANALYSIS I. EXPERIMENTAL DESIGN

E. NIPLE, J. H. SHAW

The intensities, positions and widths of spectral lines can be obtained from absorption spectra provided the physical conditions of the sample (temperature, path length, etc.) are known.

The accuracy with which these line parameters can be retrieved depends on many factors.

An Information Theory approach is adopted to make these dependences explicit for the case of an isolated, pressure-broadened line.

RA10.

SPECTRAL ANALYSIS II. SMOOTHING AND INTERPOLATION OF SPECTRAL DATA

M. HOKE, R. HAWKINS, J. H. SHAW

The parameters of individual spectral lines have been successfully retrieved by non-linear regression analysis of experimental data. (1)

The effect on these parameter values due to incorrect modeling (e.g. improper choice of the number of lines present) or of prior smoothing or interpolation of the raw data have been investigated by analyzing synthetic spectra with added noise.

Examples of the effects studied will be presented.

1. Y. Chang, J. Shaw, Appl. Spectros. 31, 213 (1977).

RA11.

ANALYSIS OF SPECTRA III. VOIGT SHAPED LINES

C. L. LIN AND J. H. SHAW

Pressure broadening of spectral lines can cause fine details in the absorption spectra of gases to be lost. As the pressure is decreased the details are observed but the line shape typically changes from a Lorentz shape to the Voigt shape. It is shown that, provided the physical conditions of the sample are known, the line intensity, position and Lorentz width can be determined by spectral curve fitting for a wide variety of Voigt shaped lines. The technique of Pierluissi et al. (1) was used to calculate the absorption coefficients of the lines.

1. J. H. Pierluissi, P. C. Vanderwood, and R. B. Gomez, JQSRT,18,555 (1977).

RA12.

SPECTRAL ANALYSIS IV. WHOLE BAND ANALYSIS OF ROTATION-VIBRATION BANDS OF DIATOMIC GASES

J. H. SHAW, C. L. LIN

Absorption spectra recorded by Fourier Transform Spectrometers consist of the transmitted signals at discrete frequencies. Spectra of the 1 - 0 bands of $H^{35}Cl$ and $H^{37}Cl$ observed at a nominal resolution of 0.1 cm^{-1} contain more than 8000 signal values.

The individual line positions, intensities and (Lorentz) widths can be described by proper choice of 23 adjustable parameters in the expressions for the upper and lower state term values, the band intensities, Herman-Wallis factors, and the J dependence of the line widths. The observed spectrum can be fitted by synthetic spectra calculated from these expressions provided the shape (assumed triangular) and width of the spectral response function of the instrument and the shape (assumed to have a quadratic dependence on frequency) and position of the background curve can be determined.

Values for the 27 unknown parameters of these models have been simultaneously determined by least squares fitting more than 3500 signal values within ±5 cm^{-1} of the HCl line centers in a single spectrum. The standard deviations of the values and the correlations between the parameters were also determined. The results are compared with previously estimated values.

This method of band analysis requires models for all the factors which influence the spectrum and large amounts of computer time, but no experimenter manipulation of the raw data.

RA13.

SPECTRAL ANALYSIS V. WHOLE-BAND ANALYSIS OF ROTATION-VIBRATION BANDS OF LINEAR MOLECULES - N_2O

R. HAWKINS, M HOKE, J. H. SHAW

The whole-band analysis technique described by Shaw and Lin (1) is being applied to N_2O absorption bands. Progress in applying this technique to N_2O spectra obtained with a Fourier Transform Spectrometer is described.

1. J. H. Shaw, C. L. Lin (preceding paper in this session)