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

John H. Shaw
Department of Physics

(NASA-CR-158716) INVESTIGATION OF INFRARED SPECTRA OF ATMOSPHERIC GASES TO SUPPORT STRATOSPHERIC SPECTROSCOPIC INVESTIGATIONS


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. NAG-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 O₃ at 750 Torr total pressure (O₂ + O₃) 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 H₂O and CO₂ in the optical path outside the absorption cell, and absorption by CO produced by reactions of ozone with carbon compounds in the cell and CH₄, an impurity in the oxygen.

Absorption by ozone occurs from 1850 to 2250 cm⁻¹. This is due to the wings of the ν1 + ν2 band near 1790 cm⁻¹, by the ν1 + ν3 band at 2110 cm⁻¹, and the 2ν₁ band near 2200 cm⁻¹. The ν₂ + 2ν₃ and ν₁ + ν₂ + ν₃ bands are seen between 2700 and 2800 cm⁻¹. There is also absorption by O₃ 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 experi-
ments.

Dr. E. Niple has investigated the types of information contained in
spectra of Lorentz shaped lines obtained with instruments of finite reso-
lution 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 strato-
sphere. 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 in-
formation 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 ob-
tained by a line by line analysis of ozone spectra obtained with an instru-
ment 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 O₃. 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 N₂O and CO₂, 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 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 widespread 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 \( \eta \) is a known function

\[
\eta = f(\xi_1, \xi_2, \ldots, \xi_k, \theta_1, \theta_2, \ldots, \theta_p) = f(\xi, \theta)
\]  \hspace{1cm} (1)

of \( k \) variables whose levels are denoted by the elements \( \xi_1, \ldots, \xi_j, \ldots, \xi_k \) of the vector \( \xi \) and of \( p \) parameters \( \theta_1, \ldots, \theta_j, \ldots, \theta_p \) elements of the vector \( \theta \). 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 \( N \times k \) matrix \( D = \{ \xi_{ju} \} \) called the design matrix. The \( j \)th row \( \xi_{ju} \) of this matrix with elements \( \xi_{u1}, \ldots, \xi_{uj}, \ldots, \xi_{uk} \) provides the levels of the \( k \) variables at which the response is to be observed in the \( j \)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 $y_\text{u}$, the observed response for the $u^{th}$ set of experimental conditions $\xi_\text{u}$, is such that:

$$E(y_\text{u}) = \eta_\text{u} = f(\xi_\text{u}; \theta), \quad (2)$$

and

$$E(y_\text{u} - \eta_\text{u})(y_\text{v} - \eta_\text{v}) = \begin{cases} \sigma^2, & u = v, \\ 0, & u \neq v \end{cases}, \quad u, v = 1, \ldots, N, \quad (3)$$

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

$$\sum_{u=1}^{N} (y_\text{u} - f(\xi_\text{u}; \hat{\theta}))^2$$

and has an approximate variance-covariance matrix $\Sigma$ given by

$$\Sigma = \sigma^2 \{F'F\}^{-1}, \quad (4)$$

where

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

and

$$f_{\text{ru}} = \left[ \frac{\partial f(\xi; \theta)}{\partial \theta} \right]_{\theta = \hat{\theta}}.$$

The Fisher information matrix $W$, used below, is related to $\Sigma$ according to

$$W = \Sigma^{-1} = 1/\sigma^2 \{F'F\}. \quad (5)$$

The Box and Lucas technique requires that the design matrix $D$ be chosen so that the determinant $|F'F|$ (or equivalently $|W|$) 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 \( |W| \) corresponds to minimizing \( \sigma^2_\theta \), the variance of the parameter estimate. This is an obvious choice in this instance; however, it is by no means obvious that \( |W| \) should be used when \( p \geq 2 \).

**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 \( \eta \) as a function of the observation time \( \xi_1 \) of an intermediate species \( B \) in the consecutive reaction

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

where each \( \xrightarrow[\theta_i]{\theta_i} \) represents a first-order, irreversible decay with rate \( \theta_i \). It may be shown that

\[
\eta(\xi_1) = \left\{ \theta_1 / (\theta_1 - \theta_2) \right\} \left\{ \exp(-\theta_2 \xi_1) - \exp(-\theta_1 \xi_1) \right\}. \tag{7}
\]

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 \( |F| \) are shown as functions of the two observation times \( \xi_{11} \) and \( \xi_{12} \). The design criterion is simply

\[
|F'F| = \{|F|\}^2
\]

when \( k = p \). In order to estimate the optimum design, the values of \( \theta_1 \) and \( \theta_2 \) must be estimated. We have used the same values as Box and
Lucas:
\[ \theta_1 = 0.7 \text{ and } \theta_2 = 0.2. \]

Figure 1 is similar to Fig. 3, p. 83 of their article. The D-optimal design occurs at
\[ \xi_{11} = 1.23, \quad \xi_{12} = 6.86 \] (8)
with \[ |F'F| = 0.66. \]

For this simple case of two observations and two unknowns, the values of \( \theta_1 \) and \( \theta_2 \) are estimated by solving the simultaneous equations
\[
\begin{align*}
y_1 &= \left( \frac{\theta_1}{\theta_1 - \theta_2} \right) \left\{ \exp(-\theta_2 \xi_{11}) - \exp(-\theta_1 \xi_{11}) \right\}, \\
y_2 &= \left( \frac{\theta_1}{\theta_1 - \theta_2} \right) \left\{ \exp(-\theta_2 \xi_{12}) - \exp(-\theta_1 \xi_{12}) \right\},
\end{align*}
\]
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 \( \theta_1 \) is
\[ \sigma^2_{\theta_1} = (\frac{\partial \theta_1}{\partial y_1})^2 \sigma^2_{y_1} + (\frac{\partial \theta_1}{\partial y_2})^2 \sigma^2_{y_2}. \] (9)
Under the assumptions in (2) and (3), this gives \( \sigma^2_{y_1} = \sigma^2_{y_2} = \sigma^2 \), and some straightforward 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
\[ \xi_{11} = 7.8, \quad \xi_{12} = 1.15, \quad \text{(Fig. 2)} \] (10a)
and
\[ \xi_{11} = 5.9, \quad \xi_{12} = 3.45. \quad \text{(Fig. 3)} \] (10b)

Clearly, if only \( \theta_2 \) 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 \( \theta_1 \) 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 \( \theta_1 \) and \( \theta_2 \) 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 $W$ is modified according to

$$W^* = CWC,$$

where the diagonal matrix $C$ is given by

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

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

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

$W^*$ is then diagonalized and the value of the smallest eigenvalue examined. According to Lees (1970, p.127) "...a small $\lambda_i^*$ [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 $W^*$ 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 $W$ (similar to Fig. 5) and the trace of $W$ (shown in Fig. 6), which may be better than $|W|$. 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 $|W|$, but rather in determining whether any 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 $(X, L, \mu_i)$ 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 $\mu_i$ over the elements of $L$. Assume that $\mu_i$ and
μ₂ 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 α ⊥ β, when there exists no set E ∈ L such that α(E) = 0 and β(E) ≠ 0 or β(E) = 0 and α(E) ≠ 0.)

By the Radon-Nikodym theorem, each hypothesis H₁ has associated with it a generalized probability density f₁(x) such that

\[ \mu_1(E) = \int_E f_1(x) \, d\lambda(x) \]

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

If each element x represents a possible set of observed responses y₁, ..., yₙ for some experiment, then f₁(x) is the probability density of observing the particular set x given that H₁ is true. When f₁(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

\[ I(1:2;x) = \log \left( \frac{f_1(x)}{f_2(x)} \right) \]

as the information in x for discrimination in favor of H₁ against H₂.

From Bayes' theorem (Box and Tiao (1973, p.10)) (11) is equivalent to

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

where P(H₁) and P(H₁ | x) are the prior and posterior probabilities of H₁. This quantity measures the relative change in the probabilities of H₁ and H₂ 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

\[
I(1:2;X) = I(1:2) = \int f_1(x)I(1:2;x)d\lambda(x)
\]

\[
= \int_{X} \log \left( \frac{P(H_1 | x)}{P(H_2 | x)} \right) \text{d}\mu_1(x) - \log \left( \frac{P(H_1)}{P(H_2)} \right).
\]  

(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 \( P(H_1) \) in communication problems (Jamison (1970, p.29); Box and Tiao (1973, pp.12-18)). One is in terms of frequencies of occurrences for \( H_1 \) 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. \( H_1, H_2 \in H \) implies \( \mu_1 = \mu_2 \). 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 0 in some open convex set in a p-dimensional Euclidean space and that each \( f_i(x) = f(x; 0) \) is of the same functional form. Under certain regularity conditions, \( I(\theta; \theta + \Delta \theta) \) can be approximated by a linear operator \( L_{op} \) for \( \Delta \theta \) sufficiently small (Kullback (1959, pp. 26-28)), giving

\[
I(\theta; \theta + \Delta \theta) \simeq L_{op}(\theta, \theta + \Delta \theta).
\] (14)

Each coordinate representation for \( H \) produces a matrix representation of \( L_{op} \), and Kullback shows that this matrix is simply \( \frac{1}{2} \) times \( W \), the Fisher information matrix. \( I(1:2) \) under these conditions is therefore similar to the metric of a curved space or the inertia of an extended object, and all the scalar measures introduced above (\( I, C, \) etc.) merely represent various aspects of this nonscalar measure. (Figure 7 shows the contours of \( I(\theta; 0) = \theta^T F F \theta \), 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 \( W \), on the other hand, changes for different coordinate representations.

**Information "Components"**

Another set of useful scalar measures are the quantities

\[
I_r \equiv \frac{\hat{\theta}_r^2}{\sigma_{\theta r}^2}, \quad r = 1, \ldots, p
\] (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 hypoth­
eses whose associated probabilities are normal distributions with
variances $\sigma_{\theta_r}^2$ and means $\hat{\theta}_r$ and 0, respectively (Miple (1978)). Con­
tour plots of these quantities for the two decay rates $\theta_1$ and $\theta_2$ are
similar to Figs. 2 and 3 (when $\sigma^2$ is independent of $\xi_{11}$ and $\xi_{12}$)
except the scales are multiplied by constants. The positions of the
maximum information components about $\theta_1$ and $\theta_2$ are thus given in (10).
Maximum information component corresponds to minimum percent uncertainty
$\%\theta_r$ where

$$%\theta_r = \left[ \frac{\sigma_{\theta_r}}{\hat{\theta}_r} \right] \times 100\%.$$  

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 experi­
menters 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 \( \theta \) as well as the design matrix \( D \). Because of this, estimation of the information depends also on estimation of \( \theta \).

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


Figures

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

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 $\xi_{11} \geq \xi_{12}$.

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

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 $\xi_{11} \geq \xi_{12}$. The lines with $C=0.0$ indicate uncorrelated designs.

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

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

Figure 7  Contour of $I(t_0;0)$ with $\theta_1 = 0.7$, $\theta_2 = 0.2$, $\sigma^2 = 1$ as a function of the observation times $\xi_{11} \geq \xi_{12}$. 
FIGURE 1.
FIGURE 2.
FIGURE 3.
FIGURE 4.
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 $\alpha$, the instrument 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 $H/\alpha \approx 1.5$ 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, and is Lorentzian. The signals $y_i$ are recorded at $N$ equally spaced frequencies $\nu_i$ from some lower frequency $\nu_L$ to an upper frequency

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

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

$$y_i = \xi_i + B(\nu_i) \int R(\nu', \nu_i; H) \exp \left[ -\frac{1}{2} \frac{(\nu' - \nu_i)^2}{\Delta \nu^2} \right] d\nu' + \mathcal{F}(\nu),$$

where $i = 1, \ldots, N$.

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

$$B(\nu_i) = a + b(\nu_i - \langle \nu \rangle)$$

with $b \ll 0$ 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

$$R(\nu', \nu_i; H) = \begin{cases} 1 - |\nu' - \nu_i|/H, & |\nu' - \nu_i| \leq H, \\ 0, & |\nu' - \nu_i| > H. \end{cases}$$

This form is triangular with full width $H$ at half height. $\xi_i$ 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 \( \kappa_o P_T \). The line parameters \( S \) and \( \kappa_o \) are the line intensity and the pressure broadening coefficients respectively, \( \nu_o \) 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: \( \nu_o \), \( S \), and \( \kappa_o \), and nine physical parameters: \( \nu_L \), \( \nu_U \), \( N \), \( a \), \( b \), \( H \), \( \sigma^2 \), \( u \) 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 \( \kappa_o \) and \( S \) are known and the physical parameters \( u \) and \( P_T \) are retrieved from the spectrum as well as \( a \), \( b \), \( \nu_o \), \( H \), and \( \sigma^2 \). In the other, \( P_T \) and \( u \) are known and the line parameters are retrieved instead. In both cases \( \nu_L \), \( \nu_U \) 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

\[
y_i = \epsilon_i + f(x_1^{(i)}, \ldots, x_k^{(i)}; \theta_1, \ldots, \theta_p), \quad i = 1, \ldots, N, \tag{5}
\]

where \( f \) is some function of \( k \) known variables \( x_1, \ldots, x_k \) and \( p \) unknown parameters \( \theta_1, \ldots, \theta_p \), and \( \epsilon_i \) is a normally distributed random variable with variance \( \sigma^2 \) and mean zero, the asymptotic variance-covariance matrix of the Least Squares estimates \( \hat{\theta}_1, \ldots, \hat{\theta}_p \) for the parameters is given by

\[
\hat{\Sigma} = \sigma^2 (J^T J)^{-1}. \quad \tag{6}
\]

In this, \( \sigma^2 \) is the estimated noise variance and \( J \) is the Jacobian matrix of \( f \) with

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

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

\[
\% \hat{\theta}_j = 100(\hat{\sigma}_j/\hat{\theta}_j). \tag{8}
\]

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

\[
Q(\theta) = \sum_{i=1}^{N} \left[ y_i - f(x_1^{(i)}, \ldots, x_k^{(i)}; \theta_1, \ldots, \theta_p) \right]^2.
\]

In NS-1 the quantity

\[
I_j = \left( \hat{\theta}_j / \hat{\sigma}_j \right)^2 \tag{9}
\]
was recommended as a measure of the component or "piece" of information in a set of data about the \( \hat{\theta}_j \) 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 \( \hat{\theta}_j \) parameter, as reflected in a large \( \% \hat{\theta}_j \), then that data set does not contain very much information about \( \hat{\theta}_j \). (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 \( \left\{ \xi_\alpha \right\} \) and \( \sigma^2 \) 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 \( S, \alpha_0, \gamma_0, a, H \) (or equivalently \( u, \pi, \gamma_0, a, H \)) can be written as

\[
I_j \approx (\text{SNR})^2 N F_j(E, C, \alpha/\kappa, H/\kappa),
\]

where \( F_j \) is some complex function of the four variables \( E, C, \alpha/\kappa \) and \( H/\kappa \).

These are defined as:

\[
E \equiv (\gamma_U - \gamma_L)/W_{\text{obs}},
\]

\[
\text{SNR} \equiv a/\nu,
\]
\[ C \equiv (\nu_0 - \nu_L)/(\nu_U - \nu_L), \]

\[ \Omega \equiv \sqrt{S_u \alpha / T}, \]

and \( \alpha \equiv \alpha P_T \).

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 \( \Omega / \alpha \) and \( H / \alpha \) 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 \( \Omega / \alpha \) and \( H / \alpha \) are shown in Fig. 2. Also shown are the contours of \( \tau_p \) 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 \( \Omega / \alpha \) and \( H / \alpha \). 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 \( W_{\text{obs}} \) over the same range of values with \( \alpha = 0.07084 \text{cm}^{-1} \).
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 $\nu_0$, $H$, $\alpha$, $\Omega = \sqrt{S\alpha}$ and $Su$ varies as a function of $E$. The calculations were done for $10$ points / observed width, and the quantity

$$I_{\nu_0} = \left(\frac{W_{\text{obs}}}{\sigma_{\nu_0}}\right)^2$$

was used for the line center component. Each $I_j$ was normalized to one at $E = 10$. Notice that, except for $I_{\nu_0}$, 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 $E$, $C$, $H/\alpha$, $\Omega/\alpha$ for instruments where $(\text{SNR})^2 \cdot N$ 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 $\sqrt{Su/H/\alpha} \approx 1.3$ and $H/\alpha \leq 2$.

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, $\alpha$ and a line center transmittance of $\approx 0.25$. Figure 6 also shows that little additional information is gained by improving the resolution beyond $H/\alpha \approx 1.5$. 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 $I_{\alpha}$, again labelled with $\% \alpha$. These contours are similar to those of $I_{Su}$ except that, in each case, $I_{\alpha}$ is strictly less than $I_{Su}$. We attribute this to the different types of information about the spectrum described by the parameters $S$ and $\alpha$. 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 $\alpha$ 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 \( \alpha \) than \( Su \), as would be expected for any line shape parameter. This is seen in the steeper upward slope along the optimum ridge as \( H/\alpha \) 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 (\( H/\alpha \) large) contains little information about \( \alpha \). Similarly a line whose shape is determined by collision broadening effects (\( H/\alpha \) small) contains little information about \( H \).

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

Figure 10 shows the contours of \( I_{\phi_0} \). (These are labelled with \( W_{obs}/\sigma_{\phi_0} \) in percent.) As with \( I_{\alpha} \), the \( t_p \) influence is the most important. In many cases, the quantity of interest is \( \sigma_{\phi_0}^2 \), 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 (\( H/\alpha \) small) and intermediate strength (\( \sqrt{Su/\mu_{\alpha}} \simeq 1.5 \)). Consequently the lines on the ridge of Fig. 6 for \( I_{Su} \) are nearly optimum for \( I_{Su}, I_{\alpha}, \) and \( \sigma_{\phi_0}^2 \).
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\( ^\text{(9)} \), away from the diffraction limit,

\[ \text{SNR} \propto W_S^2, \]

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 \( H_{\text{min}} \), SNR goes to zero.

The noise dependence for interferometers is discussed by Flemming.\( ^\text{(10)} \)

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 \( H, N, E, C, H/\kappa, \) and \( \alpha/\kappa, \) 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. \( S_u, \omega, \omega, \) 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

$$SNR \propto \left[\frac{H/\alpha}{d}\right]^d,$$

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 $I_\alpha$ in Fig. 7, for the case where $d=1/2$. The optimum (labelled +) occurs at $H/\alpha \approx 1.5$, $\Omega/\alpha \approx 1.3$, 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 $\circ$ 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 tradeoff curves of Backus and Gilbert$^{(11)}$, except the importance of $\Omega/\alpha$ (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 $Su$, $\alpha$ 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 \( \Omega \) or \( \gamma_0 \) 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 \( I_\Omega \)), however, improving the resolution beyond \( \frac{H}{\alpha} \propto 1 \) 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 \( \frac{H}{\alpha} \), a unique optimum spectrum exists for measuring \( \mu_\Omega \) or \( \alpha \) which has a larger \( \frac{H}{\alpha} \) value for higher powers of \( \frac{H}{\alpha} \).

Figure 4 shows the importance of including data points in the far wings of the line in order to measure \( \mu_\Omega, \alpha, \mu \) or \( \Omega \), but not \( \gamma_0 \). 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 \( \mu_\Omega, \alpha, \mu \) and \( \gamma_0 \) (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-I. 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 $I_{Su}$, $I_{v}$ and $I_{\nu_{0}}$) 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


Figure Captions

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

Fig. 2 - The range of values for $\omega / \alpha$ and $H / \alpha$ 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 $W_{\text{obs}}$ (in cm$^{-1}$ assuming $\omega = 0.0708$ cm$^{-1}$).

Fig. 4 - The dependence of the Information components on the extent of the spectrum $E$, for the components associated with $\gamma_0$ (line center), $H$ (resolution), $\alpha$ (Lorentz width), $Su$ and $\omega$.

Fig. 5 - The dependence of the information components on the number of data points $N$, for the components associated with $\gamma$ (line center), $H$ (resolution), $\alpha$ (Lorentz width), $Su$ and $\omega$.

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 $\alpha$ (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 $\omega$ (labelled with $\% \omega$) for $N = 300$, SNR = 100, $E = 10$ and $C = 0.5006$.

Fig. 10 - Contours of information component associated with $\gamma_0$ (labelled with $\% \gamma_0 / W_{\text{obs}} \times 100\%$) for $N = 300$, SNR = 100, $E = 10$ and $C = 0.5006$.

Fig. 11 - Contours of information component associated with $\alpha$ for $N = 300$, $E = 10$ and $C = 0.5006$ when $\text{SNR} \approx \frac{H}{\alpha}$.

+ indicates maximum. 0 indicates maximum when $\text{SNR} \approx \frac{H}{\alpha}$.

X indicates maximum when $\text{SNR} \approx \frac{H}{\alpha}^{3/2}$ (spacing corresponds to 2db change).
TRANSMITTANCE

FREQUENCY

CENTERING \equiv \left( v_o - v_L \right) / \left( v_U - v_L \right)

EXTENT \equiv \left( v_U - v_L \right) / W_{OBS}

\Omega \equiv \sqrt{S_{\text{obs}}} / \pi
Figure 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 \(^{35}\)Cl and \(^{37}\)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

\[
\nu_i = E_i' - E_i'',
\]

(1)

where \(i = 1\) identifies a model parameter of \(^{13}\)C\(^{16}\)O or \(^{37}\)Cl, and \(i = 2\) identifies the corresponding parameter of \(^{12}\)C\(^{16}\)O or \(^{35}\)Cl.
For these molecules
\[ E'_{i} = v_{bi} + B_{1}^i J(J+1) + D_{1}^i J^2(J+1) + H_{1}^i J^3(J+1) \],

and \[ E''_{i} = B''_{1} J(J+1) + D''_{1} J^2(J+1) + H''_{1} J^3(J+1) \],

(2)

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

The individual line intensities \( S_{oi}^{1}(m) \) (cm\(^{-1}\)/molecule) can be represented \((12)\) by
\[
S_{oi}^{1}(m) = \frac{8\pi^3}{3hc} \nu_{i}(m) \left| R_{oi}^{1}(m) \right|^2 \left| 1 + m \right| |Q_i| \exp \left[ \frac{-E''_i(m)}{kT} \right],
\]

(3)

where \( \left| R_{oi}^{1}(m) \right|^2 \) is the square of the dipole moment matrix element,
\[
Q_i \approx kT/hcB''_i,
\]

(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,
\[
\left| R_{oi}^{1}(m) \right|^2 = \left| R_{oi}^{1}(0) \right|^2 \left( 1 + \beta_1 m + \beta_2 m^2 \right),
\]

(5)

where \( \beta_1 \) and \( \beta_2 \) are assumed the same for both isotopes. Their values are small for both HCl and CO. Eq. (3) may be written as
\[
S_{oi}^{1}(m) = \omega_i \left( 1 + \beta_1 m + \beta_2 m^2 \right) \nu_{i}(m) \left| m \right| \exp \left[ \frac{-E''_i(m)}{kT} \right],
\]

(6)

where
\[
\omega_i = \frac{8\pi^3}{3hc} \left| R_{oi}^{1}(0) \right|^2 / 3hcQ_i,
\]

(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

\[ k_{mi}(\nu) = \left[ \frac{S_{oi}(m)}{m^2} \right] \left\{ \frac{\nu - \nu_{oi}(m)}{\Delta \nu_{oi}(m)} \right\}^{-1} \left\{ \frac{1}{\nu - \nu_{oi}(m)} + \frac{\alpha^2(m)}{\nu - \nu_{oi}(m)} \right\}^{-1}, \]  

(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

\[ \alpha_s(|m|) = \alpha_0 + \alpha_1 |m| + \alpha_2 m^2 + \alpha_3 |m|^3 + \alpha_4 m^4, \]  

(9)

where the coefficients \( \alpha_j \) 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

\[ \alpha(|m|) = \frac{\alpha_s(|m|)}{P_s}. \]  

(10)

Initial values for the coefficients \( \alpha_j \) 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

\[ k_b(\nu) = \sum_{i=1,2} \sum_{i=1,2} k_{bi}(\nu). \]  

(11)

The corresponding monochromatic band transmittance

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

(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 \( T \)

\[ N_i = \int L \rho_a T L \rho_s T, \]  

(13)
where $f_i$ is the fractional abundance (assumed known) of an isotope, $T_0 = 273K$, 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

$$V_i = \frac{W_i N_i}{T_i}, \quad (14)$$

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

The monochromatic transmittance can also be written as

$$T_{bt}(\nu) = \frac{I_t(\nu)}{I_0(\nu)}, \quad (15)$$

or

$$I_t(\nu) = I_0(\nu) T_{bt}(\nu), \quad (16)$$

where $I_0(\nu)$ 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

$$I(\nu) = I_0(\nu) T(\nu)
= \int I_0(\nu') T_{bt}(\nu') \sigma(\nu, \nu') d\nu' \int \sigma(\nu, \nu') d\nu', \quad (17)$$

where $\sigma(\nu, \nu')$ describes the shape of the spectral response function of the instrument. Previous analyses 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,

$$I_0(\nu) = a + b(\nu - \overline{\nu}) + c(\nu - \overline{\nu})^2, \quad (18)$$

where the values of the parameters $a$, $b$, and $c$ are to be estimated and $\overline{\nu}$ 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: $\beta_1$ and $\beta_2$, 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.
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 5 cm\(^{-1}\) 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 $5.2736(9) = 5.2736 \pm 0.0009$. 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

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

$$(\text{SNR})_{est} = I_o \sigma \left[ \frac{\gamma}{\sigma'} \right]_{16}, \hspace{1cm} \text{(SNR)}_{16} = 100.20. \hspace{1cm} (20)$$

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 \( \nu_0 \), \( B \), \( D \), 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 HC1 broadened to 50 Torr with N 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 $H_1' - H_1''$ 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 $\alpha_2$ 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 $\beta_2$ 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 $\alpha_j$ were found by fitting the data of Benedict et al. (14) since the $\nu$ dependence of the widths is much less than that of HC1.

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 HC1 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 $^{12}\text{C}^{16}\text{O}$ 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 $^{13}\text{C}^{16}\text{O}$ and $^{12}\text{C}^{16}\text{O}$. 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 $\left|R_0^4(0)\right|^2$ estimated. These values could then be used as initial guesses to analyze more highly resolved spectra.
REFERENCES


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.
TABLE 1. Sample conditions and spectral characteristics

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absorber</td>
<td>HCl</td>
<td>HCl</td>
<td>HCl</td>
<td>CO</td>
</tr>
<tr>
<td>$f'_2/f'_1$ Ratio</td>
<td>3.065</td>
<td>3.065</td>
<td>3.065</td>
<td>89.25</td>
</tr>
<tr>
<td>Absorber amt. atm. cm$^{-1}$(296K)</td>
<td>0.421</td>
<td>0.421</td>
<td>0.421</td>
<td>0.350</td>
</tr>
<tr>
<td>Partial Pressure of N$_2$ (Torr)</td>
<td>700</td>
<td>700</td>
<td>50</td>
<td>700</td>
</tr>
<tr>
<td>Sample Temp (K)</td>
<td>296</td>
<td>296</td>
<td>296</td>
<td>296</td>
</tr>
<tr>
<td>Path Length (m)</td>
<td>171</td>
<td>171</td>
<td>171</td>
<td>171</td>
</tr>
<tr>
<td>Spectral Limits Upper (cm$^{-1}$)</td>
<td>3050</td>
<td>3050</td>
<td>3050</td>
<td>2235</td>
</tr>
<tr>
<td>Lower (cm$^{-1}$)</td>
<td>2671</td>
<td>2671</td>
<td>2671</td>
<td>2025</td>
</tr>
<tr>
<td>Spacing of data points (cm$^{-1}$)</td>
<td>0.0603</td>
<td>0.0603</td>
<td>0.0603</td>
<td>0.0537</td>
</tr>
<tr>
<td>No. of points analyzed</td>
<td>3611</td>
<td>3611</td>
<td>3611</td>
<td>3920</td>
</tr>
<tr>
<td>$\nu$ (cm$^{-1}$)</td>
<td>2877</td>
<td>2877</td>
<td>2877</td>
<td>2130</td>
</tr>
</tbody>
</table>
TABLE 2. HCl Spectra: estimated parameter values

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>Parameter</th>
<th>Input Values</th>
<th>Retrieved</th>
<th>Retrieved</th>
<th>Retrieved</th>
<th>Retrieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( a )</td>
<td>5.2740</td>
<td>5.276</td>
<td>44.09</td>
<td>44.09</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( 10^6 b ) cm</td>
<td>0</td>
<td>1</td>
<td>-48</td>
<td>-2.3 \times 10^3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( 10^8 c ) cm^2</td>
<td>0</td>
<td>-2</td>
<td>80</td>
<td>4 \times 10^2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( 10^5 R ) cm^-1</td>
<td>1.04</td>
<td>1.15</td>
<td>1.19</td>
<td>1.19</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>( 10^5 v_1 )</td>
<td>7.604</td>
<td>6.68</td>
<td>53</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>( 10^4 v_2 )</td>
<td>2.349</td>
<td>2.00</td>
<td>17</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>( 10^2 \delta_1 )</td>
<td>-2.60</td>
<td>-3.16</td>
<td>-6</td>
<td>-6</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>( 10^4 \theta_2 )</td>
<td>4.50</td>
<td>-16</td>
<td>-28</td>
<td>-28</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>( \nu_{b1} ) cm^-1</td>
<td>2883.8705</td>
<td>2883.870</td>
<td>2883.870</td>
<td>2883.870</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>( B_1 ) cm^-1</td>
<td>10.12090</td>
<td>10.12115</td>
<td>10.1234</td>
<td>10.1234</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>( 10^4 \theta_1 ) cm^-1</td>
<td>5.200</td>
<td>5.21</td>
<td>5.242</td>
<td>5.242</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>( 10^8 H_1 ) cm^-1</td>
<td>1.63</td>
<td>2</td>
<td>3.9</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>( B_1^- ) cm^-1</td>
<td>10.42430</td>
<td>10.42417</td>
<td>10.42458</td>
<td>10.42458</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>( 10^4 \theta_1^- ) cm^-1</td>
<td>5.270</td>
<td>5.25</td>
<td>5.350</td>
<td>5.350</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>( 10^8 H_1^- ) cm^-1</td>
<td>1.63</td>
<td>1</td>
<td>3.9</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>( \nu_{b2} ) cm^-1</td>
<td>2885.9775</td>
<td>2885.979</td>
<td>2885.980</td>
<td>2885.980</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>( B_2^- ) cm^-1</td>
<td>10.136223</td>
<td>10.13615</td>
<td>10.13614</td>
<td>10.13614</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>( 10^4 \theta_2^- ) cm^-1</td>
<td>5.2167</td>
<td>5.20</td>
<td>5.203</td>
<td>5.203</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>( 10^8 H_2^- ) cm^-1</td>
<td>1.65</td>
<td>0</td>
<td>0.2</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>( B_2^- ) cm^-1</td>
<td>10.440254</td>
<td>10.44032</td>
<td>10.44064</td>
<td>10.44064</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>( 10^4 \theta_2^- ) cm^-1</td>
<td>5.2835</td>
<td>5.290</td>
<td>5.325</td>
<td>5.325</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>( 10^8 H_2^- ) cm^-1</td>
<td>1.65</td>
<td>1.8</td>
<td>0.2</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>( 10^8 a_0 ) cm^-1</td>
<td>1.0497</td>
<td>1.049</td>
<td>1.06</td>
<td>1.06</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>( 10^4 a_1 ) cm^-1</td>
<td>3.84</td>
<td>2</td>
<td>5.55</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>( 10^3 a_2 ) cm^-1</td>
<td>-1.568</td>
<td>-1.58</td>
<td>-1.748</td>
<td>-2 \times 10^{-3} (bound)</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>( 10^6 a_3 ) cm^-1</td>
<td>-5.76</td>
<td>-4</td>
<td>-3.00</td>
<td>-18</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>( 10^8 a_4 ) cm^-1</td>
<td>+6.64</td>
<td>6.8</td>
<td>9.65</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3. Some intermediate parameter values obtained during the analysis of a synthetic HCl spectrum

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Res. Sum. Squares</th>
<th>a</th>
<th>R(cm⁻¹)</th>
<th>10²B₁</th>
<th>10⁴D₁ (cm⁻¹)</th>
<th>B₂(cm⁻¹)</th>
<th>10α₀(cm⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Guess</td>
<td></td>
<td>5.2740</td>
<td>0.10400</td>
<td>-2.60</td>
<td>5.270</td>
<td>10.13600</td>
<td>1.050</td>
</tr>
<tr>
<td>Iteration Number</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>22.827</td>
<td>5.2740</td>
<td>0.10363</td>
<td>-2.62</td>
<td>5.272</td>
<td>10.13597</td>
<td>1.162</td>
</tr>
<tr>
<td>5</td>
<td>13.330</td>
<td>5.2739</td>
<td>0.10458</td>
<td>-2.60</td>
<td>5.274</td>
<td>10.13621</td>
<td>1.059</td>
</tr>
<tr>
<td>10</td>
<td>10.064</td>
<td>5.2746</td>
<td>0.10206</td>
<td>-2.52</td>
<td>5.247</td>
<td>10.13621</td>
<td>1.053</td>
</tr>
<tr>
<td>15</td>
<td>9.931</td>
<td>5.2737</td>
<td>0.10388</td>
<td>-2.46</td>
<td>5.249</td>
<td>10.13614</td>
<td>1.050</td>
</tr>
<tr>
<td>16</td>
<td>9.929</td>
<td>5.2736</td>
<td>0.10391</td>
<td>-2.46</td>
<td>5.252</td>
<td>10.13615</td>
<td>1.049</td>
</tr>
</tbody>
</table>
TABLE 4. Estimate of the asymptotic correlation matrix obtained from the analysis of a synthetic HCl spectrum (all values multiplied by 1000)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>R</th>
<th>v₁</th>
<th>v₂</th>
<th>β₁</th>
<th>β₂</th>
<th>vb₁</th>
<th>B₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>435</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>-529</td>
<td>-334</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>-322</td>
<td>-153</td>
<td>166</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v₁</td>
<td>-474</td>
<td>-437</td>
<td>321</td>
<td>472</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v₂</td>
<td>-520</td>
<td>-423</td>
<td>342</td>
<td>602</td>
<td>866</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₁</td>
<td>-31</td>
<td>-15</td>
<td>16</td>
<td>341</td>
<td>284</td>
<td>286</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β₂</td>
<td>-184</td>
<td>158</td>
<td>430</td>
<td>148</td>
<td>-73</td>
<td>100</td>
<td>-56</td>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vb₁</td>
<td>513</td>
<td>-441</td>
<td>-172</td>
<td>23</td>
<td>112</td>
<td>59</td>
<td>63</td>
<td>-354</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>B₁</td>
<td>223</td>
<td>260</td>
<td>397</td>
<td>-34</td>
<td>74</td>
<td>48</td>
<td>70</td>
<td>298</td>
<td>-36</td>
<td>1000</td>
</tr>
<tr>
<td>D₁</td>
<td>447</td>
<td>430</td>
<td>281</td>
<td>-109</td>
<td>14</td>
<td>-17</td>
<td>56</td>
<td>326</td>
<td>105</td>
<td>846</td>
</tr>
<tr>
<td>H₁</td>
<td>582</td>
<td>504</td>
<td>79</td>
<td>-154</td>
<td>-81</td>
<td>-108</td>
<td>74</td>
<td>239</td>
<td>201</td>
<td>600</td>
</tr>
<tr>
<td>B₁</td>
<td>77</td>
<td>105</td>
<td>680</td>
<td>34</td>
<td>122</td>
<td>103</td>
<td>66</td>
<td>285</td>
<td>144</td>
<td>842</td>
</tr>
<tr>
<td>D₂</td>
<td>-33</td>
<td>-170</td>
<td>857</td>
<td>47</td>
<td>163</td>
<td>158</td>
<td>16</td>
<td>398</td>
<td>166</td>
<td>584</td>
</tr>
<tr>
<td>H₂</td>
<td>-66</td>
<td>-212</td>
<td>806</td>
<td>20</td>
<td>94</td>
<td>97</td>
<td>40</td>
<td>384</td>
<td>142</td>
<td>334</td>
</tr>
<tr>
<td>vb₂</td>
<td>-232</td>
<td>6</td>
<td>834</td>
<td>132</td>
<td>61</td>
<td>93</td>
<td>47</td>
<td>525</td>
<td>-181</td>
<td>458</td>
</tr>
<tr>
<td>B₂</td>
<td>-127</td>
<td>-553</td>
<td>202</td>
<td>-13</td>
<td>378</td>
<td>332</td>
<td>-29</td>
<td>-250</td>
<td>376</td>
<td>27</td>
</tr>
<tr>
<td>H₂</td>
<td>-109</td>
<td>-740</td>
<td>208</td>
<td>75</td>
<td>433</td>
<td>378</td>
<td>62</td>
<td>-223</td>
<td>677</td>
<td>-93</td>
</tr>
<tr>
<td>B₂</td>
<td>-337</td>
<td>-26</td>
<td>467</td>
<td>19</td>
<td>184</td>
<td>196</td>
<td>-50</td>
<td>219</td>
<td>-366</td>
<td>282</td>
</tr>
<tr>
<td>D₂</td>
<td>-492</td>
<td>373</td>
<td>504</td>
<td>79</td>
<td>-50</td>
<td>7</td>
<td>-54</td>
<td>535</td>
<td>-884</td>
<td>292</td>
</tr>
<tr>
<td>H₂</td>
<td>-294</td>
<td>457</td>
<td>236</td>
<td>13</td>
<td>-215</td>
<td>-163</td>
<td>-50</td>
<td>462</td>
<td>-793</td>
<td>166</td>
</tr>
<tr>
<td>a₀</td>
<td>463</td>
<td>465</td>
<td>-134</td>
<td>-671</td>
<td>-770</td>
<td>-900</td>
<td>-240</td>
<td>-17</td>
<td>-142</td>
<td>129</td>
</tr>
<tr>
<td>a₁</td>
<td>31</td>
<td>14</td>
<td>-15</td>
<td>-299</td>
<td>-229</td>
<td>-250</td>
<td>-859</td>
<td>-35</td>
<td>-52</td>
<td>-64</td>
</tr>
<tr>
<td>a₂</td>
<td>-247</td>
<td>-630</td>
<td>-221</td>
<td>196</td>
<td>453</td>
<td>402</td>
<td>66</td>
<td>-566</td>
<td>430</td>
<td>-335</td>
</tr>
<tr>
<td>a₃</td>
<td>-11</td>
<td>-63</td>
<td>-57</td>
<td>228</td>
<td>158</td>
<td>184</td>
<td>595</td>
<td>-70</td>
<td>101</td>
<td>-19</td>
</tr>
<tr>
<td>a₄</td>
<td>86</td>
<td>497</td>
<td>351</td>
<td>-72</td>
<td>-110</td>
<td>-99</td>
<td>96</td>
<td>449</td>
<td>-416</td>
<td>371</td>
</tr>
</tbody>
</table>
TABLE 4 (cont)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$D_1^*$</th>
<th>$H_1^*$</th>
<th>$B_1^*$</th>
<th>$D_1^*$</th>
<th>$H_1^*$</th>
<th>$\nu_{b2}$</th>
<th>$B_2^*$</th>
<th>$D_2^*$</th>
<th>$H_2^*$</th>
<th>$B_2^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1^*$</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_1^*$</td>
<td>912</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_1^*$</td>
<td>649</td>
<td>410</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_1^*$</td>
<td>610</td>
<td>463</td>
<td>834</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_1^*$</td>
<td>444</td>
<td>398</td>
<td>614</td>
<td>919</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_{b2}$</td>
<td>394</td>
<td>233</td>
<td>682</td>
<td>827</td>
<td>785</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_2^*$</td>
<td>64</td>
<td>46</td>
<td>138</td>
<td>200</td>
<td>163</td>
<td>-313</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D_2^*$</td>
<td>-86</td>
<td>-110</td>
<td>183</td>
<td>249</td>
<td>-225</td>
<td>-78</td>
<td>677</td>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_2^*$</td>
<td>-127</td>
<td>-151</td>
<td>176</td>
<td>231</td>
<td>217</td>
<td>65</td>
<td>366</td>
<td>916</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>$B_2^*$</td>
<td>253</td>
<td>144</td>
<td>274</td>
<td>342</td>
<td>287</td>
<td>101</td>
<td>547</td>
<td>-94</td>
<td>-343</td>
<td>1000</td>
</tr>
<tr>
<td>$D_2^*$</td>
<td>134</td>
<td>-35</td>
<td>228</td>
<td>249</td>
<td>218</td>
<td>552</td>
<td>-379</td>
<td>-642</td>
<td>-609</td>
<td>466</td>
</tr>
<tr>
<td>$H_2^*$</td>
<td>30</td>
<td>-85</td>
<td>72</td>
<td>47</td>
<td>58</td>
<td>473</td>
<td>-670</td>
<td>-695</td>
<td>-525</td>
<td>-7</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>219</td>
<td>264</td>
<td>73</td>
<td>67</td>
<td>108</td>
<td>134</td>
<td>-361</td>
<td>-449</td>
<td>-392</td>
<td>-83</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>-61</td>
<td>-81</td>
<td>-49</td>
<td>-11</td>
<td>-32</td>
<td>-38</td>
<td>20</td>
<td>-28</td>
<td>-59</td>
<td>44</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-447</td>
<td>-437</td>
<td>-269</td>
<td>-370</td>
<td>-395</td>
<td>-579</td>
<td>601</td>
<td>738</td>
<td>616</td>
<td>-99</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-22</td>
<td>14</td>
<td>-32</td>
<td>-73</td>
<td>-54</td>
<td>-80</td>
<td>61</td>
<td>93</td>
<td>90</td>
<td>-55</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>440</td>
<td>381</td>
<td>345</td>
<td>450</td>
<td>451</td>
<td>639</td>
<td>-510</td>
<td>-587</td>
<td>-463</td>
<td>143</td>
</tr>
</tbody>
</table>
TABLE 4 (cont)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$D_2^-$</th>
<th>$H_2^-$</th>
<th>$\alpha_0$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_2^-$</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_2^-$</td>
<td>836</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>165</td>
<td>285</td>
<td>1000$\cdot$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>52</td>
<td>49</td>
<td>215</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>-615</td>
<td>-647</td>
<td>-578</td>
<td>-52</td>
<td>1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>-150</td>
<td>-147</td>
<td>-204</td>
<td>-877</td>
<td>158</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>642</td>
<td>632</td>
<td>336</td>
<td>-29</td>
<td>-867</td>
<td>-177</td>
<td>1000</td>
</tr>
</tbody>
</table>
### TABLE 5. CO Spectrum. Estimated Parameter Values

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Retrieved</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4 Calculated</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>53.07 (2)</td>
<td>53.11 (2)</td>
<td>53.11 (1)</td>
<td>53.095</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>( 10^8 \beta ) cm</td>
<td>0.0f</td>
<td>-1</td>
<td>0.0f</td>
<td>0.0f</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 10^9 ) cm²</td>
<td>0.0f</td>
<td>-1</td>
<td>0.0f</td>
<td>0.0f</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 10^6 ) ( \nu_1 )</td>
<td>1.040f</td>
<td>0.69 (1)</td>
<td>0.69 (1)</td>
<td>0.689</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 10^6 \nu_2 )</td>
<td>1.30 (2)</td>
<td>1.22 (4)</td>
<td>1.22 (4)</td>
<td>1.215</td>
<td>1.227</td>
<td></td>
</tr>
<tr>
<td>( 10^5 \beta_1 )</td>
<td>-2.80f</td>
<td>-3</td>
<td>-2.8f</td>
<td>-2.8f</td>
<td>-2.8</td>
<td>12</td>
</tr>
<tr>
<td>( 10^6 \beta_2 )</td>
<td>7.50f</td>
<td>7.5f</td>
<td>7.5f</td>
<td>7.5</td>
<td>7.5</td>
<td>12</td>
</tr>
<tr>
<td>( \nu_1 ) cm⁻¹</td>
<td>2096.0686</td>
<td>2096.06</td>
<td>2096.06</td>
<td>2096.058</td>
<td>2096.0670</td>
<td>11</td>
</tr>
<tr>
<td>( B_1 ) cm⁻¹</td>
<td>1.8216 (1)</td>
<td>1.822 (3)</td>
<td>1.822 (3)</td>
<td>1.8217</td>
<td>1.82161</td>
<td>11</td>
</tr>
<tr>
<td>( 10^6 D_1 ) cm⁻¹</td>
<td>6.2 (3)</td>
<td>6.05 (4)</td>
<td>6.04 (6)</td>
<td>5.922</td>
<td>5.594</td>
<td></td>
</tr>
<tr>
<td>( 10^5 H )</td>
<td>5.0f</td>
<td>5.0f</td>
<td>5.0f</td>
<td>5.0</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>( 10^6 D_1 ) cm⁻¹</td>
<td>1.83809 (9)</td>
<td>1.838 (3)</td>
<td>1.838 (3)</td>
<td>1.8380</td>
<td>1.83797</td>
<td>11</td>
</tr>
<tr>
<td>( 10^5 H )</td>
<td>6.5 (2)</td>
<td>5.58 (4)</td>
<td>5.58 (7)</td>
<td>5.478</td>
<td>5.596</td>
<td></td>
</tr>
<tr>
<td>( 10^5 D_2 ) cm⁻¹</td>
<td>6.103 (1)</td>
<td>6.039 (3)</td>
<td>6.039 (3)</td>
<td>6.0400</td>
<td>6.1195</td>
<td>10</td>
</tr>
<tr>
<td>( 10^6 D_2 ) cm⁻¹</td>
<td>5.44f</td>
<td>6</td>
<td>5.44f</td>
<td>5.44f</td>
<td>5.44</td>
<td>10</td>
</tr>
<tr>
<td>( B_2 ) cm⁻¹</td>
<td>1.922575 (9)</td>
<td>1.9225 (3)</td>
<td>1.922504 (5)</td>
<td>1.922510</td>
<td>1.922529</td>
<td>10</td>
</tr>
<tr>
<td>( 10^5 D_2 ) cm⁻¹</td>
<td>6.142 (3)</td>
<td>6.066 (3)</td>
<td>6.066 (3)</td>
<td>6.0673</td>
<td>6.1202</td>
<td>10</td>
</tr>
<tr>
<td>( 10^5 H )</td>
<td>5.60f</td>
<td>6</td>
<td>5.60f</td>
<td>5.60f</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>( 10^5 D_2 ) cm⁻¹</td>
<td>8.1 (2)</td>
<td>8.7 (2)</td>
<td>8.7 (2)</td>
<td>8.73</td>
<td>8.18</td>
<td>14</td>
</tr>
<tr>
<td>( 10^5 D_2 ) cm⁻¹</td>
<td>-4.7 (2)</td>
<td>-4.8 (1)</td>
<td>-4.8 (1)</td>
<td>-4.75</td>
<td>-5.06</td>
<td>14</td>
</tr>
<tr>
<td>( 10^5 D_2 ) cm⁻¹</td>
<td>2.6 (1)</td>
<td>3.0 (1)</td>
<td>3.0 (1)</td>
<td>2.91</td>
<td>2.91</td>
<td>14</td>
</tr>
<tr>
<td>( 10^5 D_2 ) cm⁻¹</td>
<td>-5.9 (3)</td>
<td>-7.2 (3)</td>
<td>-7.2 (3)</td>
<td>-7.01</td>
<td>-5.84</td>
<td>14</td>
</tr>
</tbody>
</table>
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 \cite{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 $k(v)$ at frequency $v$, the transmitted signal $I_m(v)$ is related to the signal $I_0(v)$, obtained with the absorber removed, by

$$I_m(v) = I_0(v) \exp[-k(v) \cdot u],$$

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

$$u = \frac{P_{\text{abs}}}{P_0} \cdot \frac{T_0}{T_{\text{abs}}} \cdot \ell(\text{atm-cm},P_0,T_0),$$

where the sample temperature is $T_{\text{abs}}$, the partial pressure of the absorber is $P_{\text{abs}}$, the absorbing path length is $\ell(\text{cm})$, and $T_0$ and $P_0$ 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

\[ I(\nu) = \frac{\int I_m(\nu') \sigma(\nu, \nu') \, d\nu'}{\int \sigma(\nu, \nu') \, d\nu'} \]

(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. \( I_0(\nu) = A_1 + A_2(\nu - \nu_0) \), where \( A_1 \) and \( A_2 \) are constants to be determined, and \( \nu_0 \) is chosen to lie near the middle of the spectral region analyzed. In the analysis of synthetic spectra described here \( \nu_0 = 0.7320 \) 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

\[ k(x, y) = \frac{k(\nu)}{k_0} = \frac{y}{\pi} \int_{-\infty}^{\infty} \frac{\exp(-t^2)}{y^2 + (x - t)^2} \, dt \]

(4)

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

\[ k_0 = \frac{S}{\alpha_D} \left( \frac{\ln 2}{\pi} \right)^{\nu} \]

\[ y = \frac{\alpha_D}{\ln 2} (\ln 2)^{\nu} \]

\[ x = \frac{\nu - \nu_0}{\alpha_D} (\ln 2)^{\nu} \]

\[ S = \int_{-\infty}^{\infty} k(\nu) \, d\nu \], is the line intensity.
\( \alpha_L = \) Lorentz half width due to radiation damping or collisional broadening,

\[ \alpha_D = \frac{\nu_0}{c} \left( \frac{2kT \omega r}{m} \right)^{1/2}, \]

is the Doppler half-width,

and

\( \nu_0 = \) line center position.

The Voigt function cannot be evaluated exactly but a number of techniques for calculating its approximate value have been described. A fast calculational method for obtaining its value to about one part in \( 10^4 \) has been described by Pierluissi et al. 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 \( 0 \leq x \leq 1.8, 0 \leq y \leq 3.0 \) 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 \( A_1, A_2, H, S, u, \nu_0, \alpha_D \) and \( \alpha_L \). 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 \( \alpha_D \) 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., 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 al9 and this is used together with
Eqs. (1-4) to calculate the transmitted signal I(v). The subroutine for
use with BMDP3R also gives analytical expressions for the partial deriva­
tives 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 pro­
posed by Pierluissi et al is given by the real part of w(z), where w(z)
is provided by Abramowitz and Stegun10 as the following;
in region I (0 ≤ |x| < 3, 0 ≤ y < 1.8)
\[ w(z) = e^{-z^2}[1 - \text{erf}(-iz)] , \]
(5)
where z = x + iy,

\[ \text{erf}(-iz) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n! \left(2n + 1\right)} (-iz)^{2n+1} , \]
(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., in region II \( (3 \leq |x| < 5, 1.8 \leq y < 5) \),

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

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

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

The derivatives of \( w(z) \) are,

in region I,

\[
\frac{dw(z)}{dz} = (-2z) \cdot w(z) + \frac{2}{\sqrt{\pi}} i,
\]

in region II,

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

and in region III,

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

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 \( \mu \) 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 \( \mu \) 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. 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. 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 $\pm 0.001$ 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* 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 N2 added to give a total pressure of 700 Torr in a 171 m path at 296 K. In a previous analysis 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. 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. 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. 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


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


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

<table>
<thead>
<tr>
<th>Voigt Region</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>I, II, and III Combined</th>
<th>I, II, and III Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_D$ (cm$^{-1}$)</td>
<td>0.027506</td>
<td>0.0027506</td>
<td>0.0027506</td>
<td>0.0027506</td>
<td>0.0027506</td>
</tr>
<tr>
<td>$c_L$ (cm$^{-1}$)</td>
<td>0.000132</td>
<td>0.0070</td>
<td>0.03304</td>
<td>0.005616</td>
<td>0.002632</td>
</tr>
<tr>
<td>$SU/\pi$ (cm$^{-1}$)</td>
<td>0.012046</td>
<td>0.004015</td>
<td>0.037477</td>
<td>0.008031</td>
<td>0.010708</td>
</tr>
<tr>
<td>$v_0$ (cm$^{-1}$)</td>
<td>0.7320</td>
<td>0.7320</td>
<td>0.7320</td>
<td>0.7320</td>
<td>0.7320</td>
</tr>
<tr>
<td>$H$ (cm$^{-1}$)</td>
<td>0.00060</td>
<td>0.00080</td>
<td>0.0500</td>
<td>0.008000</td>
<td>0.00200</td>
</tr>
<tr>
<td>$A_1$</td>
<td>5.274</td>
<td>5.274</td>
<td>5.274</td>
<td>5.274</td>
<td>5.274</td>
</tr>
<tr>
<td>$A_2$ (cm)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\Delta v$ (cm$^{-1}$)</td>
<td>0.00050</td>
<td>0.00020</td>
<td>0.015</td>
<td>0.0010847</td>
<td>0.001241</td>
</tr>
<tr>
<td>Total number of points</td>
<td>240</td>
<td>356</td>
<td>90</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>Number of points for $T &lt; 0.95$</td>
<td>200</td>
<td>273</td>
<td>21</td>
<td>100</td>
<td>93</td>
</tr>
<tr>
<td>Number of iterations</td>
<td>32</td>
<td>25</td>
<td>11</td>
<td>45</td>
<td>15</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Exp.</th>
<th>Data</th>
<th>Initial Value</th>
<th>Retrieved 1σ</th>
<th>Fixed 1σ</th>
<th>No. of Pts.</th>
<th>No. of Ex.</th>
<th>No. of Trans &lt; 0.95</th>
<th>1σ 10²σ (cm⁻²)</th>
<th>1σ 10²Ω (cm⁻²)</th>
<th>1σ 10²Ω'/Ω (cm⁻²)</th>
<th>1σ Ω (cm⁻²)</th>
<th>1σ 10²H (cm⁻²)</th>
<th>Total No. Pts.</th>
<th>No. of Pts.</th>
<th>No. of Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial Value</td>
<td>2.75</td>
<td></td>
<td>5.50</td>
<td>6.33</td>
<td>0.56</td>
<td>0.11</td>
<td>0.015</td>
<td>0.7330</td>
<td>5.274</td>
<td>0</td>
<td>5.295</td>
<td>100</td>
<td>20</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td>Retrieved</td>
<td>6.53</td>
<td></td>
<td>8.031</td>
<td>7.12</td>
<td>1.2</td>
<td>0.001</td>
<td>-</td>
<td>0.7330</td>
<td>5.274</td>
<td>0</td>
<td>-0.001</td>
<td>16.5</td>
<td>50</td>
<td>2.20</td>
</tr>
<tr>
<td>2a</td>
<td>Initial Value</td>
<td>2.75</td>
<td></td>
<td>5.50</td>
<td>6.48</td>
<td>0.71</td>
<td>0.12</td>
<td>0.003</td>
<td>0.7330</td>
<td>5.274</td>
<td>0</td>
<td>0</td>
<td>16.5</td>
<td>50</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td>Retrieved</td>
<td>6.72</td>
<td></td>
<td>8.00</td>
<td>7.28</td>
<td>1.0</td>
<td>0.001</td>
<td>-</td>
<td>0.7321</td>
<td>5.271</td>
<td>0</td>
<td>-0.001</td>
<td>16.7</td>
<td>100</td>
<td>2.20</td>
</tr>
<tr>
<td>3</td>
<td>Initial Value</td>
<td>4.00</td>
<td></td>
<td>8.00</td>
<td>8.10</td>
<td>0.49</td>
<td>0.10</td>
<td>0.001</td>
<td>0.7330</td>
<td>5.274</td>
<td>0</td>
<td>0</td>
<td>12.0</td>
<td>100</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>Retrieved</td>
<td>8.00</td>
<td></td>
<td>8.00</td>
<td>8.13</td>
<td>1.3</td>
<td>0.12</td>
<td>0.003</td>
<td>0.7321</td>
<td>5.271</td>
<td>0</td>
<td>0.001</td>
<td>12.3</td>
<td>100</td>
<td>5.00</td>
</tr>
<tr>
<td>4</td>
<td>Initial Value</td>
<td>4.00</td>
<td></td>
<td>12.0</td>
<td>12.15</td>
<td>0.9</td>
<td>0.16</td>
<td>0.0002</td>
<td>0.7320</td>
<td>5.274</td>
<td>0</td>
<td>5.268</td>
<td>20.0</td>
<td>100</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>Retrieved</td>
<td>12.15</td>
<td></td>
<td>12.0</td>
<td>12.0</td>
<td>1.9</td>
<td>0.10</td>
<td>0.0002</td>
<td>0.7321</td>
<td>5.271</td>
<td>0</td>
<td>0.001</td>
<td>23.0</td>
<td>100</td>
<td>5.00</td>
</tr>
<tr>
<td>5</td>
<td>Initial Value</td>
<td>3.00</td>
<td></td>
<td>12.0</td>
<td>12.05</td>
<td>0.9</td>
<td>0.2</td>
<td>0.0002</td>
<td>0.7320</td>
<td>5.274</td>
<td>0</td>
<td>5.270</td>
<td>28.0</td>
<td>200</td>
<td>6.00</td>
</tr>
<tr>
<td></td>
<td>Retrieved</td>
<td>12.1</td>
<td></td>
<td>12.0</td>
<td>12.0</td>
<td>0.9</td>
<td>0.2</td>
<td>0.0002</td>
<td>0.7321</td>
<td>5.271</td>
<td>0</td>
<td>0.001</td>
<td>28.3</td>
<td>200</td>
<td>6.00</td>
</tr>
<tr>
<td>6</td>
<td>Initial Value</td>
<td>27.5</td>
<td></td>
<td>12.0</td>
<td>12.05</td>
<td>0.132</td>
<td>0.12</td>
<td>0.0002</td>
<td>0.7320</td>
<td>5.274</td>
<td>0</td>
<td>5.271</td>
<td>26.0</td>
<td>210</td>
<td>7.60</td>
</tr>
<tr>
<td></td>
<td>Retrieved</td>
<td>12.0</td>
<td></td>
<td>12.0</td>
<td>12.0</td>
<td>0.01</td>
<td>0.10</td>
<td>0.0002</td>
<td>0.7321</td>
<td>5.271</td>
<td>0</td>
<td>0.001</td>
<td>28.0</td>
<td>210</td>
<td>7.60</td>
</tr>
</tbody>
</table>
Table 3. Comparison of HCL F(1) line parameters retrieved by using the single and double precision versions of HUPFR for Lorentz and Voigt shapes

<table>
<thead>
<tr>
<th>Exp</th>
<th>Iterations</th>
<th>Precision</th>
<th>Line Shape</th>
<th>Constraints</th>
<th>Residual</th>
<th>$A_1$ (cm)</th>
<th>$A_2$ (cm)</th>
<th>$S_1U/\pi$ (cm$^{-1}$)</th>
<th>$S_2U/\pi$ (cm$^{-1}$)</th>
<th>$Q_{\theta}$ (cm$^{-1}$)</th>
<th>$v_1$ (cm$^{-1}$)</th>
<th>$v_2$ (cm$^{-1}$)</th>
<th>$\Pi$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9</td>
<td>Double</td>
<td>Voigt</td>
<td>$S_2/S_1$ = 3.0186</td>
<td>$H = 0.104$</td>
<td>2.84655</td>
<td>5.275 ±0.005</td>
<td>0.0006 ±0.0007</td>
<td>0.127 ±0.003</td>
<td>0.577 ±0.010</td>
<td>0.091 ±0.002</td>
<td>2863.014 ±0.002</td>
<td>2865.093 ±0.002</td>
</tr>
<tr>
<td>B</td>
<td>12</td>
<td>Double</td>
<td>Voigt</td>
<td>None</td>
<td>2.8350</td>
<td>5.276 ±0.005</td>
<td>0.0006 ±0.0007</td>
<td>0.125 ±0.003</td>
<td>0.559 ±0.002</td>
<td>0.095 ±0.006</td>
<td>2863.014 ±0.002</td>
<td>2865.093 ±0.002</td>
<td>0.089 ±0.003</td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>Double</td>
<td>Lorentz</td>
<td>$S_2/S_1$ = 3.0186</td>
<td>$H = 0.104$</td>
<td>2.8424</td>
<td>5.277 ±0.005</td>
<td>0.0007 ±0.0007</td>
<td>0.137 ±0.003</td>
<td>0.577 ±0.002</td>
<td>0.091 ±0.002</td>
<td>2863.014 ±0.002</td>
<td>2865.093 ±0.002</td>
</tr>
<tr>
<td>D</td>
<td>12</td>
<td>Double</td>
<td>Lorentz</td>
<td>None</td>
<td>2.8350</td>
<td>5.276 ±0.005</td>
<td>0.0006 ±0.0007</td>
<td>0.125 ±0.003</td>
<td>0.559 ±0.027</td>
<td>0.095 ±0.006</td>
<td>2863.014 ±0.002</td>
<td>2865.093 ±0.002</td>
<td>0.089 ±0.003</td>
</tr>
<tr>
<td>E</td>
<td>26</td>
<td>Single</td>
<td>Voigt</td>
<td>$S_2/S_1$ = 3.0186</td>
<td>$H = 0.104$</td>
<td>2.8425</td>
<td>5.275 ±0.005</td>
<td>0.0007 ±0.0007</td>
<td>0.137 ±0.003</td>
<td>0.577 ±0.002</td>
<td>0.091 ±0.002</td>
<td>2863.014 ±0.002</td>
<td>2865.093 ±0.002</td>
</tr>
<tr>
<td>F</td>
<td>10</td>
<td>Single</td>
<td>Voigt</td>
<td>None</td>
<td>2.8392</td>
<td>5.276 ±0.005</td>
<td>0.0006 ±0.0007</td>
<td>0.125 ±0.004</td>
<td>0.559 ±0.002</td>
<td>0.095 ±0.006</td>
<td>2863.014 ±0.002</td>
<td>2865.093 ±0.002</td>
<td>0.089 ±0.003</td>
</tr>
<tr>
<td>G</td>
<td>26</td>
<td>Single</td>
<td>Lorentz</td>
<td>$S_2/S_1$ = 3.0186</td>
<td>$H = 0.104$</td>
<td>2.8423</td>
<td>5.275 ±0.005</td>
<td>0.0006 ±0.0007</td>
<td>0.137 ±0.003</td>
<td>0.577 ±0.002</td>
<td>0.091 ±0.002</td>
<td>2863.014 ±0.002</td>
<td>2865.093 ±0.002</td>
</tr>
<tr>
<td>H</td>
<td>17</td>
<td>Single</td>
<td>Lorentz</td>
<td>None</td>
<td>2.8399</td>
<td>5.276 ±0.005</td>
<td>0.0007 ±0.0007</td>
<td>0.125 ±0.006</td>
<td>0.559 ±0.027</td>
<td>0.095 ±0.006</td>
<td>2863.014 ±0.002</td>
<td>2865.093 ±0.002</td>
<td>0.089 ±0.003</td>
</tr>
</tbody>
</table>

* $S_2/S_1$ = ratio of the band intensities of H$^{16}$Cl and H$^{37}$Cl

$^7$ 700 Torr H$_2$ pressure-broadened width
| X-ray Temperature | 300°C | 500°C | 700°C | 900°C | 1100°C | 1300°C | 1500°C | 1700°C | 1900°C | 2100°C | 2300°C | 2500°C | 2700°C | 2900°C | 3100°C | 3300°C | 3500°C | 3700°C | 3900°C | 4100°C | 4300°C | 4500°C | 4700°C | 4900°C | 5100°C | 5300°C | 5500°C | 5700°C | 5900°C | 6100°C | 6300°C | 6500°C | 6700°C | 6900°C | 7100°C | 7300°C | 7500°C | 7700°C | 7900°C |
|------------------|-------|-------|-------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
Table 5. Comparison of NCL F(1) line parameters retrieved by using BMDPAR and BMDP3R in double precision for Voigt shape

<table>
<thead>
<tr>
<th>Program</th>
<th>Constraints</th>
<th>Iterations</th>
<th>Residual</th>
<th>$h_1$</th>
<th>$A_0$ (cm)</th>
<th>$S_0 U_0$ (cm$^{-1}$)</th>
<th>$S_1 U_1$ (cm$^{-1}$)</th>
<th>$Q_0^k$ (cm$^{-1}$)</th>
<th>$v_1$ (cm$^{-1}$)</th>
<th>$v_2$ (cm$^{-1}$)</th>
<th>$H$ (cm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMDP3R</td>
<td>$S_n/S_b = 3.0016$ $H = 0.604$</td>
<td>9</td>
<td>2.8427</td>
<td>5.275</td>
<td>0.0006</td>
<td>0.187</td>
<td>0.577</td>
<td>0.191</td>
<td>2863.019</td>
<td>10.002</td>
<td>10.002</td>
</tr>
<tr>
<td>BMDP3R</td>
<td>$S_n/S_b = 3.0016$ $H = 0.604$</td>
<td>39</td>
<td>2.847</td>
<td>5.277</td>
<td>-0.0006</td>
<td>0.188</td>
<td>0.570</td>
<td>0.192</td>
<td>2863.019</td>
<td>10.002</td>
<td>10.002</td>
</tr>
<tr>
<td>BMDP3R</td>
<td>None</td>
<td>12</td>
<td>2.8390</td>
<td>5.276</td>
<td>0.0006</td>
<td>0.186</td>
<td>0.559</td>
<td>0.195</td>
<td>2863.019</td>
<td>10.002</td>
<td>10.002</td>
</tr>
<tr>
<td>BMDP3R</td>
<td>None</td>
<td>37</td>
<td>2.8393</td>
<td>5.276</td>
<td>0.0006</td>
<td>0.186</td>
<td>0.559</td>
<td>0.195</td>
<td>2863.019</td>
<td>10.002</td>
<td>10.002</td>
</tr>
</tbody>
</table>

*700 Torr $A_0$ 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 0.42l atm-cm (296 K) of HCl with N₂ 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. 2300 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.
Figure 1
APPENDIX I

SUBROUTINE FOR USE WITH BMDP3R TO RETRIEVE
THE PARAMETERS OF TEN VOIGT SHAPED LINES
SUBROUTINE FUN(FDF, PXNKASENVARNPAR, IPASSXLOSS)
REAL*8 OF(NPAR), P(NPAR), X(NVAR)
REAL*3 A(32), ST(32), SM(32), X0, DOP(2), COVST(2), PC1, PC2, PCC1, PCC2,
* SACK, YEU(2), V2(2), S(2), T(2), USER(2), VSR(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), G(2), H(2), AKU(32),
* TM(32), RX, DFF(g)
COMPLEX*16 ZA(2), ZB(2), ZAZ(2)

REAL*8 A(30)/1.0000000,-.33333333,.O0000000,-.23095296E-3,
* 4.62902963E-3,7.57575757E-4,1.068376068E-4,-1.32751323E-5,
* 1.4391690E-5,1.65038522E-7,1.31223296E-9,-1.08922104E-6,
* 1.350702795E-11,-5.947794016E-12,3.95429516E-13,
* -3.46327011E-14,1.448326464E-15,-.03735012E-17,
* -4.221407289E-18,-2.107854191E-19,
* -1.06922104E-19,-4.551846750E-22,1.977046756E-23,
* -5.329149299E-25,-3.29260345E-26,-1.264107699E-27,-.678483522E-29,
* -1.659761793E-30,-.754191644E-32,-1.916942862E-33/,
* A5/0.02893894/, A6/5.32534370/, A7/5.12422424/, A8/2.7525510/,
* A9/0.0516536/, A10/2.72474500/, PISQ/1.128379167/, PISQRT/1.7725538091/
C FIND LINE PARAMETERS AND BACKGROUND USING VOIGT PROFILE AND
C TRIANGULAR SLIT FUNCTION
C F=BACKGROUND*CONVOLVED TRANSMITTANCE
C BACK=BACKGROUND+(AI+AZ*X11-X0))
C P(1)=AL
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(LV(2))/DOP(2)
X0=76.343000
DOP(1)=0.26754999-2
DOP(2)=0.2750570-2
CONST(1)=3.111791000
CONST(2)=3.026840000
PC1=P(3)*CONST(1)
PC2=P(4)*CONST(2)
PCC1=PC1*CONST(1)
PCC2=PC2*CONST(2)
SACK=P(11)+P(12)*(X11-X0)
F=0.00
DO 1 I=1, 8
DFF(I)=0.00
1 CONTINUE
DO 2 K=1, 32
B(K)=(33.00-2*K)/32.00
CONTINUE
DO 3 I=1,31,2
N=(I+1)/2
ST(N)=I/32.DO
ST(N+1)=(32.DO-I)/32.DO
CONTINUE
DO 4 K=1,32
CM(K)=(X(I)+B(K)*P(K))
YEU(1)=DABS(CM(K)-P(6))
YEU(2)=DABS(CM(K)-P(7))
DO 5 J=1,2
U(J)=YEU(J)*CONJG(J)
V(J)=P(J)*CONJG(J)
S(J)=U(J)*U(J)-V(J)*V(J)
T(J)=2.DO*U(J)*V(J)
Z2(J)=DCMPLX(S(J),T(J))
IF(V(J).GE.5.DO) GO TO 6
IF(V(J).GE.5.DO) GO TO 7
CONTINUE
CONTINUE
DO 4 K=1,32
CM(K)=(X(I)+B(K)*P(K))
YEU(1)=DABS(CM(K)-P(6))
YEU(2)=DABS(CM(K)-P(7))
DO 5 J=1,2
U(J)=YEU(J)*CONJG(J)
V(J)=P(J)*CONJG(J)
S(J)=U(J)*U(J)-V(J)*V(J)
T(J)=2.DO*U(J)*V(J)
Z2(J)=DCMPLX(S(J),T(J))
IF(V(J).GE.5.DO) GO TO 6
IF(V(J).GE.5.DO) GO TO 7
CONTINUE
CONTINUE
DO 4 K=1,32
CM(K)=(X(I)+B(K)*P(K))
YEU(1)=DABS(CM(K)-P(6))
YEU(2)=DABS(CM(K)-P(7))
DO 5 J=1,2
U(J)=YEU(J)*CONJG(J)
V(J)=P(J)*CONJG(J)
S(J)=U(J)*U(J)-V(J)*V(J)
T(J)=2.DO*U(J)*V(J)
Z2(J)=DCMPLX(S(J),T(J))
IF(V(J).GE.5.DO) GO TO 6
IF(V(J).GE.5.DO) GO TO 7
CONTINUE
CONTINUE
DO 4 K=1,32
CM(K)=(X(I)+B(K)*P(K))
YEU(1)=DABS(CM(K)-P(6))
YEU(2)=DABS(CM(K)-P(7))
DO 5 J=1,2
U(J)=YEU(J)*CONJG(J)
V(J)=P(J)*CONJG(J)
S(J)=U(J)*U(J)-V(J)*V(J)
T(J)=2.DO*U(J)*V(J)
Z2(J)=DCMPLX(S(J),T(J))
IF(V(J).GE.5.DO) GO TO 6
IF(V(J).GE.5.DO) GO TO 7
CONTINUE
CONTINUE
DO 4 K=1,32
CM(K)=(X(I)+B(K)*P(K))
YEU(1)=DABS(CM(K)-P(6))
YEU(2)=DABS(CM(K)-P(7))
DO 5 J=1,2
U(J)=YEU(J)*CONJG(J)
V(J)=P(J)*CONJG(J)
S(J)=U(J)*U(J)-V(J)*V(J)
T(J)=2.DO*U(J)*V(J)
Z2(J)=DCMPLX(S(J),T(J))
IF(V(J).GE.5.DO) GO TO 6
IF(V(J).GE.5.DO) GO TO 7
CONTINUE
CONTINUE
DO 4 K=1,32
CM(K)=(X(I)+B(K)*P(K))
YEU(1)=DABS(CM(K)-P(6))
YEU(2)=DABS(CM(K)-P(7))
DO 5 J=1,2
U(J)=YEU(J)*CONJG(J)
V(J)=P(J)*CONJG(J)
S(J)=U(J)*U(J)-V(J)*V(J)
T(J)=2.DO*U(J)*V(J)
Z2(J)=DCMPLX(S(J),T(J))
IF(V(J).GE.5.DO) GO TO 6
IF(V(J).GE.5.DO) GO TO 7
CONTINUE
* 2S(t1+G(J)Y(J))/G(J)*G(J)+R(J))} +
  2S(t1-E(J)Y(J))/E(J)*E(J)+R(J))}
ROER(J)=-DIMAG(ZDER(J))
UDER(J)=DREAL(ZDER(J))
GO TO 5
7  R(J)=T(J)*T(J)
   T(J)=T(J)*U(J)
   ES(J)=S(J)-B2
   GCJ=S(J)-B4
   ZBI(J)=Z2(J)-B2
   ZBJ(J)=Z2(J)-B4
   ZDER3(J)=(B1/ZBI(J)+B3/ZB3(J))-
   *(2.00*Z2(J))*(B1/ZBI(J)**2+B3/ZB3(J)**2)
   VOIGT(J)=B1*{(t1-E(J)Y(J))/(E(J)*E(J)+R(J))} +
   *(2.00*Z2(J))*(B1/ZBI(J)**2+B3/ZB3(J)**2)
   ROER(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))
RR(K)=BAC*ST(K)*TM(K)/16.00
OFF(3)=VOIGT(1)*CONST(1)*PISQRT
OFF(4)=VOIGT(2)*CONST(2)*PISQRT
OFF(5)=PC1*UDER(1)+PC2*UDER(2)*PISQRT
IF(CM(K).EQ.P(6)) GO TO 14
IF(CM(K).LT.P(6)) GO TO 11
OFF(6)=PC1*ROER(1)*PISQRT
GO TO 12
11 OFF(6)=O.000
12 CONTINUE
IF(CM(K).EQ.P(7)) GO TO 18
IF(CM(K).LT.P(7)) GO TO 15
OFF(7)=PC2*ROER(2)*PISQRT
GO TO 16
15 OFF(7)=O.000
16 CONTINUE
OFF(8)=-(OFF(6)+OFF(7))*B(K)
F=F+RR
DF(1)=DF(1)+ST(K)*TM(K)/16.00
DF(2)=DF(2)+ST(K)*TM(K)/16.00*(X(1)-XO)
DF(3)=DF(3)+RR*OFF(3)
DF(4)=DF(4)+RR*OFF(4)
DF(5)=DF(5)+RR*OFF(5)
DF(6)=DF(6)+RR*OFF(6)
DF(7)=DF(7)+RR*OFF(7)
DF(8)=DF(8)+RR*OFF(8)
4 CONTINUE
RETURN
END

//GO EXEC PGM=LOADER,PARM='MAP,EP=MAIN',COND=(8,LT),TIME=(11,30)
//GO,FT01FOO1 DO UNIT=SYSDA,SPACE=(CYL,(1,1)),
//DCB=(RECFM=VBS,LRECL=324,8LKSIZ=2298)
//FT02FOO1 DO UNIT=SYSDA,SPACE=(CYL,(1,1)),
//FT05FOO1 DO DONAME=SYSIN
//FT06FOO1 DO SYSGUT=AO,DCB=(RECFM=F8A,LRECL=133,PLKSIZ=931,9UFNO=1)
//SYSLIB DO ONSNAME=SYSLI.FORTLIB,DISP=SHR
// GC.SYSLIN DO DSN='LL08J',DISP=(OLD,DELETE)
// DD DSNNAME='SYSL4.9MPLIB(8MCP9R)',DISP=SHR
//SYSOUT DD SYSOUT=A,DCB=BUFNO=1
// GC.SYIN DD *
PROB TITLE IS 'TEST OF VOIGT LINE'.
INPUT VARIABLES ARE 2.
FORMAT IS '(F10.4,F10.3)'.
UNIT=10.
CASE ARE 299./
VARIABLE NAMES ARE X,P.
REGRESSION TITLE IS 'LINE PARAMETERS OF DOUBLE VOIGT LINE'.
DEPEND=F.
TOL=0.0000001.
PARM=8.
CONST=2.
ITER=45.
HALF=23./
PARM INIT=5.24,0.3,0.019,0.0617,0.0275,95,77.7,0.103
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=(5),1, K=0.104.
NAMES=A1,A2,SUOVPI1,SUOVPI2,LORENTZ,NEU1,NEU2,H.
END/
FINISH/
//GO.FTIOFOOI DO DSN='TS0744.HCL.DATALINOZ',DISP=SHR
//
APPENDIX II

SUBROUTINE FOR USE WITH BMDPAR TO RETRIEVE THE PARAMETERS OF A SINGLE VOIGT LINE
SUBROUTINE FUN(FPXNKASE,NVARNPARIPASSvXLOSS)
REAL*8 P(NPAR),X(NVAR)
REAL48 B(32),ST(32),CM(32),XODOP,CONST,PC1PCC1,BACKtYEU,UVS,T,
* USER,VSERUNVNvU2,V2,UNEWVNEWDEG1,G2,VOIGTRIt,
* REH,GAKU(32),TM(32)tRRI
REAL*t
AN(301/1.0000000-.33333333,1000000,-.238095238E-1,
* 4.6Z962963E-3,-7.57575757E-4,1.0683760660E-4,-1.322751323E-5,
* 1.45891690E-6,-1.4503052222E-7,-1.3122532636E-8,-1.089222104E-9,
* 8.35702795E-11,-5.947794014E-12,3.955429516E-13,
* -2.46682701E-14,1.4483264664E-15,-8.032735012E-17,
* 4.221407269E-16,-2.107855191E-19,
* 1.002516494E-20,-.551846759E-22,1.977604754E-23,
* 8.230149299E-25,3.289260349E-26,-1.264107899E-27,4.67848352E-29,
* -1.0409761793E-30,3.754191644E-32,1.916942862E-33,/
* B4/.0272474500/, PISQ/1.128379167/, PISQRT/1.7724535091/

C FIND LINE PARAMETERS AND BACKGROUND USING VOIGT PROFILE AND
C TRIANGULAR SLIT FUNCTION
C F=BACKGROUND*CONVOLVED TRANSMITTANCE
C BACK=BACKGROUND=(At A2*(X(I)-XO))
C P(1)=AI
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
X0=0.7320DO
DOP=0.275D-2
CONST=3.027471313D02
PC1=I(3)*CONST
PCC1=PC1*CONST
BACK=(P(1)+P2)*(X(I)-XO))
F=0.00
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/32.00
  ST(N+16)=(32.00-I)/32.00
3 CONTINUE
DO 4 K=1,32
  CM(K)=(X(I)+B(K)*P(6))
  YEU=OABS(CM(K)-P(5))
  U=YEU*CONST
  V=P(4)*CONST
  S=U*V
  T=2.0D0*U*V
IF(V.EQ.0.000) GO TO 6
IF(V.GE.5.000.00.OR.U.GE.5.000.00)) GO TO 7

IF(V.GE.1.8D0 OR U.GE.3.0D0) GO TO 8
6 CONTINUE
USER=V
VSE=-U
UN=V
VN=-U
U2=S
V2=T
M=6.842D0*U+8.0D0
IF(M.GT.29) M=29
IF(U.EQ.0.0D0) M=15
DO 9 I9=1,M
UNEW=UN*U2-VN*V2
VNEW=V2*UN+VN*U2
USER=USER+UNEW*AN(I9+1)
VSE=VSE+VNEW*AN(I9+1)
UN=UNEW
VN=VNEW
9 CONTINUE
DE=DEXP(-S)
G1=(1.0D0-PISQ*USER)
G2=PISQ*VSE
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)=PC*VOIGT*PISQRT
TM(K)=DEXP(-AKU(K))
RR1=BACK*ST(K)*TM(K)/16.0D0
F=F+RR1
4 CONTINUE
RETURN
END
//GO EXEC PGM=LOADER,PARM=MAIN%,COND=(8,LT),TIME=(5,30)
//FT01F001 DD UNIT=SYSDA,SPACE=(CYL,1,1)
//FT02F001 DD UNIT=SYSDA,SPACE=(CYL,1,1)
//FT03F001 DD DNAME=SYSIN
//SYSLIB DD DNAME=SYS1.FORTLIB,DISP=SHR
//SYSLIN DD DNAME=SYS1.FORLIB,DISP=OLD,DELETE
//DD DNAME=SYS1.BMDLIB(BMDPAR),DISP=SHR
//SYSOUT DD DNAME=A,D,G3=BUFNO=1
//GO,SYSIN DD *
//PROBLEM TITLE IS TEST
//INPUT VARIABLES ARE Z.
//FORMAT IS *(FIO.4,5X,FIO.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 .0,0.1,0.120,0.0990,0.9320,0.9
  MIN ARE 4.0,-0.1,1.E-7,0.000001,0.3320,0.00001
  NAME=AI, AZ, SUOVRPI, LDRENTZ, NEUD, H.
/END
/FINISH
//GO.FT10F001 DD DSN=TS0744+HCL+DATA(ANN),DISP=SHR
//
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.

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 nonlinear 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.


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.

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.

SPECTRAL ANALYSIS V. WHOLE-BAND ANALYSIS OF ROTATION-VIBRATION BANDS OF LINEAR MOLECULES - N$_2$O

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

The whole-band analysis technique described by Shaw and Lin (1) is being applied to N$_2$O absorption bands. Progress in applying this technique to N$_2$O spectra obtained with a Fourier Transform Spectrometer is described.

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