Global Least-Squares Analysis of the IR Rotation–Vibration Spectrum of HCl

W

Joel Tellinghuisen
Department of Chemistry, Vanderbilt University, Nashville, TN 37235; *joel.tellinghuisen@vanderbilt.edu*

Many data-analysis problems can be addressed in multiple ways, ranging from a series of related "local" fitting problems up to a single comprehensive "global" analysis. The local approach typically deals with subsets of the data treated in similar ways, with the results often subjected to subsequent analysis to arrive at a set of final parameters. While the stepwise approach can be useful in assessing the data and in obtaining preliminary estimates of the parameters, the global, or simultaneous analysis of all the data, is normally the statistically optimum method for obtaining the final parameters. As an example, consider the analysis of the IR fundamental and first overtone of HCl (or $H^{35}Cl$ and $H^{37}Cl$) *(1–7).* The data for each of the two bands for each isotopomer might be analyzed by the method of combination differences *(8)* to obtain estimates of the rotational and centrifugal distortion constants (B_υ and D_υ) for each of the two involved vibrational levels (υ = 0 and 1 for the fundamental, 0 and 2 for the overtone). This entails at least three fits for each band—one for $B_{\upsilon'}$ and $D_{\upsilon'}$ (for the upper level υ'), another for $B_{\upsilon''}$ and $D_{\upsilon''}$ (for lower level υ''), and a third to assess the band origin $\tilde{\nu}_0$. A statistically better treatment is a single fit of all the lines in each band to yield the five parameters at once. But the separate analyses of the fundamental and overtone yield duplicate estimates of B_0 and D_0. Since these constants must be identical for the two bands, one should properly fit both bands simultaneously to yield single estimates of these two parameters, along with B_1, B_2, D_1, D_2, $\tilde{\nu}_{10}$, and $\tilde{\nu}_{20}$. Alternatively, the same data might be fitted to the "equilibrium" parameters B_e, α_e, D_e, β_e, ω_e, and $\omega_e x_e$ *(8).* Then it also becomes possible to take advantage of the isotopic relations among these parameters to fit both bands for both isotopomers simultaneously.

These kinds of global fits have been widely used by researchers in spectroscopy since the 1970s, and spreadsheet methods for carrying them out have been described recently in this *Journal (2–5).* Normally one needs more than one independent variable (e.g., υ', υ'', J', and J'') to distinguish the different subsets of the data in the global analysis. That would appear to rule out the use of "two-dimensional" programs like KaleidaGraph,[1] which are geared toward display of the dependent variable as a function of a single independent variable. That is unfortunate, since such programs make it particularly easy to fit data, through the user-defined fitting option *(9),* with the added advantage of yielding directly the estimated standard errors in the fit parameters.

Actually, it *is* possible to use programs like KaleidaGraph for global least-squares analysis, and the tricks for doing so are a significant component of the present article. My methods involve altering the values of the independent variable for sorting purposes and then compensating appropriately in the definition of the fit relation. The use of library functions makes this approach a powerful one for fitting data to moderately complex models. It also makes it easy for the user to alter the fit model in trial-and-error fashion to check the significance of different terms, very much as one would handle similar problems on the research level. These methods are illustrated here for the analysis of HCl IR absorption spectra.

Spectroscopic Theory and Notation

All spectroscopic transitions represent energy differences in the substance under study,

$$\Delta E_{\text{substance}} = E_{\text{photon}} = h\nu = hc\tilde{\nu} \qquad (1)$$

where h, c, ν, and $\tilde{\nu}$ are Planck's constant, the speed of light, the frequency, and the wavenumber of the transition, respectively. For a transition in a diatomic molecule,

$$hc\tilde{\nu}(\upsilon', J', \upsilon'', J'') = E'(\upsilon', J') - E''(\upsilon'', J'') \qquad (2)$$

where single primes represent upper levels and double primes lower. The energy of a bound (υ, J) level of a diatomic can be represented as a sum of electronic, vibrational, and rotational energies, given by T_e, G_υ, and $F_\upsilon(J)$ in Herzberg's notation. The rotational energy includes contributions from centrifugal distortion, giving, for the simplest kind of molecular states ($^1\Sigma$),

$$\frac{E(\upsilon, J)}{hc} = T_e + G_\upsilon + B_\upsilon\kappa - D_\upsilon\kappa^2 + H_\upsilon\kappa^3 + \ldots \qquad (3)$$

where $\kappa \equiv J(J+1)$, B_υ is the rotational constant, D_υ and H_υ are the first and second centrifugal distortion constants, and all quantities on the right side are in the spectroscopist's "energy" units of cm^{-1}. Often T_e is combined with G_υ to give the vibronic energy T_υ. In the customary theoretical treatment of the diatomic molecule, involving expansion of the potential energy function for the electronic state in question about the minimum at internuclear distance $R = R_e$ and energy (cm^{-1}) $E/hc = T_e$, each of the quantities T_υ, B_υ, D_υ, H_υ, ... is itself expressed as a polynomial in $z \equiv (\upsilon + {}^1/_2)$,

$$T_\upsilon = T_e + \omega_e z - \omega_e x_e z^2 + \omega_e y_e z^3 - \ldots \qquad (4a)$$

$$B_\upsilon = B_e - \alpha_e z + \gamma_e z^2 - \ldots \qquad (4b)$$

$$D_\upsilon = D_e + \beta_e z + \ldots \qquad (4c)$$

$$H_\upsilon = H_e + \ldots \qquad (4d)$$

where the notation is that of Herzberg *(8).* The negative signs are included in eqs 3, 4a, and 4b, because the corresponding terms are almost always negative (meaning D_υ, $\omega_e x_e$, and α_e are almost always positive).[2] The first few "e" (for equilibrium) coefficients in each expression in eqs 4a–d have theoretical significance and carry special names: equilibrium

vibrational frequency (ω_e), first anharmonicity constant ($\omega_e x_e$), equilibrium rotational constant (B_e), vibration–rotation interaction constant (α_e). However, in dealing with data spanning large ranges of υ and J, these expressions are customarily treated in ad hoc fashion, with the number of included terms adjusted to do statistical justice to the data. The "e" constants are thus considered a means to an end in representing the more fundamental quantities on the left side of eqs 4a–d. There is a similar element of the ad hoc in the use of eq 3, where the range of J spanned by the data determines how many centrifugal distortion constants are needed.

If it is assumed that all the energy levels for the electronic state in question are determined by a single potential energy function $U(R)$ for that state, the state is said to follow the *rotating-oscillator model*. Then the centrifugal distortion constants D_υ, H_υ, L_υ, ... are not independent quantities but are, in effect, determined by the G_υ and B_υ values for the state. In most recent work on diatomic molecules, the data are analyzed with the centrifugal distortion constants constrained to be consistent with the G_υ and B_υ values, in accord with this model *(11)*. As examples of this interdependence, the centrifugal distortion parameters in eqs 4c and 4d are given by *(8, 10)*:

$$D_e = \frac{4B_e^3}{\omega_e^2} \tag{5a}$$

$$\beta_e = D_e\left(\frac{8\omega_e x_e}{\omega_e} - \frac{5\alpha_e}{B_e} - \frac{\alpha_e^2\omega_e}{24B_e^3}\right) \tag{5b}$$

$$H_e = \frac{2D_e}{3\omega_e^2}\left(12B_e^2 - \alpha_e\omega_e\right) \tag{5c}$$

In the Born–Oppenheimer approximation, all isotopic forms (isotopomers) of a molecule have identical electronic energy states and potential energy functions $U(R)$ for each such state. Then there are simple relations connecting the "e" coefficients of the different isotopomers. If we choose one isotopomer having reduced mass μ as reference species, the "isotopic rho factor" for a different isotopomer i is defined as *(8)*,

$$\rho_i = \left(\frac{\mu}{\mu_i}\right)^{1/2} \tag{6}$$

and for the coefficients given explicitly in eqs 4a–d,

$$\begin{gathered} T_{e,i} = T_e; \quad \omega_{e,i} = \rho_i\omega_e; \quad \omega_e x_{e,i} = \rho_i^2\omega_e x_e; \\ \omega_e y_{e,i} = \rho_i^3\omega_e y_e; \quad B_{e,i} = \rho_i^2 B_e; \quad \alpha_{e,i} = \rho_i^3\alpha_e; \\ \gamma_{e,i} = \rho_i^4\gamma_e; \quad D_{e,i} = \rho_i^4 D_e; \quad \beta_{e,i} = \rho_i^5\beta_e; \\ H_{e,i} = \rho_i^6 H_e \end{gathered} \tag{7}$$

For bookkeeping purposes, it is useful to note that the energies for species i can be obtained from the spectroscopic parameters for the reference species by using eq 3 and including a factor of ρ_i with each power of $(\upsilon + {}^1/_2)$ and a factor of ρ_i^2 with each power of κ. (This usage is clearer when Dunham's notation is used.[2]) For HCl, if $H^{35}Cl$ is the reference, $\rho = 0.99924302$ for $H^{37}Cl$.

Methods of Analysis

Direct and Difference Fitting Methods

Authors of recent contributions to this *Journal* on this problem have discussed and used several different methods to estimate the spectroscopic constants for HCl from data obtained by recording the IR fundamental and first overtone *(2–7)*. The spectrum contains R and P lines, for which $J' = J'' + 1$ and $J' = J'' - 1$, respectively. Lines of both branches in a given band follow the "m-representation" *(8)*,

$$\begin{aligned} \tilde{\nu}(m) = \tilde{\nu}_0 &+ \left(B_{\upsilon'} + B_{\upsilon''}\right)m \\ &+ \left(B_{\upsilon'} - B_{\upsilon''} - D_{\upsilon'} + D_{\upsilon''}\right)m^2 \\ &- 2\left(D_{\upsilon'} + D_{\upsilon''}\right)m^3 - \left(D_{\upsilon'} - D_{\upsilon''}\right)m^4 \end{aligned} \tag{8}$$

in which $m = J'' + 1$ for R lines and $-J''$ for P, and centrifugal distortion terms beyond the first in eq 3 are neglected. The band origin $\tilde{\nu}_0$ is $T_{\upsilon'} - T_{\upsilon''}$, which is $G_{\upsilon'} - G_{\upsilon''}$ for an IR absorption band. A direct fit of all the lines in a band to eq 8 will yield statistically best estimates of all five parameters and their standard errors *(9, 12)*.

While fitting data to eq 8 is routine today, in the precomputer era it was not. Accordingly, spectroscopists resorted to various difference techniques to simplify the analysis and to render the data into straight-line forms suitable for graphical analysis. Chief among these are the methods of combination differences and successive differences. In the first of these, one computes the differences $R(J) - P(J)$ of the R and P lines for a given J, and also $R(J-1) - P(J+1)$, where the lines are numbered by their J'' values. The first such difference contains only the rotational and centrifugal distortion constants for the upper υ level, while the second depends on only the lower level; thus the two analyses serve to determine $B_{\upsilon'}$, $D_{\upsilon'}$, ... and $B_{\upsilon''}$, $D_{\upsilon''}$, ..., respectively. For example, again neglecting centrifugal distortion constants beyond D_υ,

$$\begin{aligned} \Delta_2F'(J) &\equiv R(J) - P(J) \\ &= (4J+2)\left[B_{\upsilon'} - 2D_{\upsilon'}\left(J^2 + J + 1\right)\right] \end{aligned} \tag{9}$$

It is usually reasonable to assume that the wavenumbers of all the R and P lines of a given band are determined with equal precision, in which case the differences $\Delta_2F'(J)$ also have constant precision (larger by $\sqrt{2}$), and an unweighted fit to eq 9 is appropriate. On the other hand, if this equation is linearized by dividing through by $(4J + 2)$, the fitted quantities $[\Delta_2F'(J)/(4J + 2)]$ no longer possess constant uncertainty and should be weighted proportional to their inverse variances, or $w_J \propto (4J + 2)^2$. The unweighted fit to eq 9 and the weighted straight-line fit yield identical results for the adjustable parameters ($B_{\upsilon'}$ and $D_{\upsilon'}$) and their standard errors *(9, 12, 13)*.

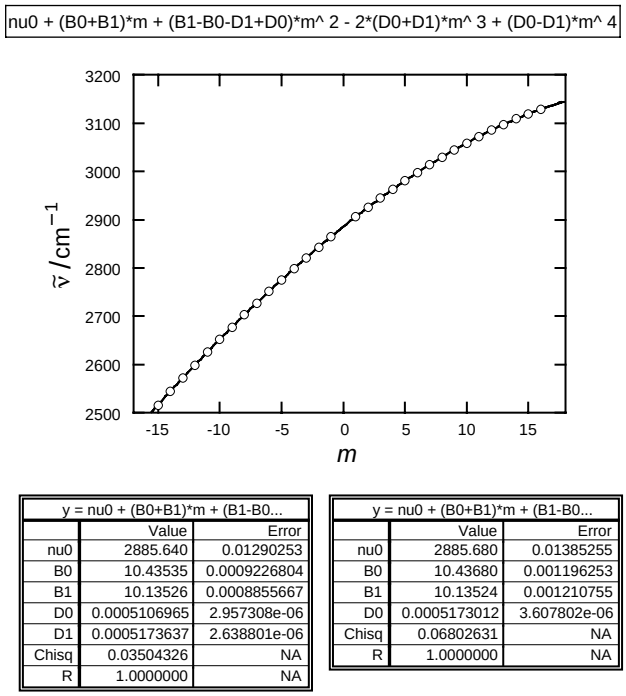
In the method of successive differences, the quantities $\tilde{\nu}(m+1) - \tilde{\nu}(m)$ are computed and fitted. This procedure reduces the dimensionality of the fit by removing the band origin $\tilde{\nu}_0$; and with certain approximations, it also yields a straight-line plot *(3, 5)*.

The method of combination differences is a sound statistical procedure, capable of giving results identical to those from a direct fit to eq 8 under most favorable conditions, as I have shown in recent work *(12).* However, this method has obvious drawbacks: (i) two separate analyses are required for each band; (ii) in its visually appealing straight-line form, a weighted fit is required; (iii) the band origin $\tilde{\nu}_0$ must still be determined; and (iv) there is no simple, statistically proper way to do this. In the method of successive differences, the subtraction process produces correlated data, requiring analysis by the complex method of correlated least squares; even then, it yields results of much poorer precision *(12).* Here, too, the band origin must still be determined. The method of successive differences is statistically deficient and really should be retired from the teaching literature. The method of combination differences is useful in understanding the principles but should not be the final word in quantitative analysis.

In the rest of this article, I will use only direct fits of all data to eq 8 or appropriate variants thereof, since such a direct fit of all the data is statistically optimal.

User-Defined Fitting with KaleidaGraph

The user-defined fitting routine is one of the most powerful data-analysis tools in KaleidaGraph and similar programs (DeltaGraph, IGOR Pro). In KaleidaGraph (KG) this is the "general" option under the curve-fit menu, and it may be used to tackle both linear and nonlinear fitting problems, with and without weighting of the data. In this regard it is the nature of the fitting problem that determines whether the fit is linear or nonlinear, not the computational algorithm.[3] In all of the fitting of the HCl spectral data discussed in refs *1–7,* and in most discussed here, the fit models are linear in the adjustable parameters, qualifying these cases as "linear least squares". Thus all the usual properties of linear least squares (LS) apply to the results *(13).* In particular, convergence to a minimum in the sum of weighted squared residuals is assured, and such convergence will occur in a single iteration in some nonlinear algorithms.[3] Conversely, failure to achieve convergence indicates a flaw in the algorithm or a pathology in the data.



y = nu0 + (B0+B1)*m + (B1-B0...		
	Value	Error
nu0	2885.640	0.01290253
B0	10.43535	0.0009226804
B1	10.13526	0.0008855667
D0	0.0005106965	2.957308e-06
D1	0.0005173637	2.638801e-06
Chisq	0.03504326	NA
R	1.0000000	NA

y = nu0 + (B0+B1)*m + (B1-B0...		
	Value	Error
nu0	2885.680	0.01385255
B0	10.43680	0.001196253
B1	10.13524	0.001210755
D0	0.0005173012	3.607802e-06
Chisq	0.06802631	NA
R	1.0000000	NA

Figure 1. Data for fundamental IR band of $H^{35}Cl$ from ref *3,* displayed as a function of the index *m* and fitted to eq 8 with separate values for D_0 and D_1 and with a common value for both (results box on right). The untruncated fit function for the five-parameter fit is shown exactly as entered, in the box at top. The displayed values of "Chisq" give 0.037 and 0.050 cm^{-1} for s_y.

I have described the use of KG on several common data-analysis problems in the physical chemistry laboratory curriculum *(9),* but it is useful to review some key properties of the program here. The first step in an analysis is the creation of the data sheet, with the independent and dependent variables appearing in different columns. (The data may be imported from text or spreadsheet files.) The data must then be displayed as a "line" or "scatter" plot before the fit can be invoked from the curve-fit menu. After one of the "general" fits is selected from the latter, clicking on "define" opens the fit-definition box, into which the right side [$f(x)$ of $y = f(x)$] of the fit relation is entered. "Weight data" is checked for a weighted fit, and nonzero initial values are entered for each adjustable parameter.[4] When the open boxes are closed, the dependent variable is selected and the fit iterations proceed.

KG's default variables are `m0` for the independent variable and `m1–m9` for the adjustable parameters. The first five of these are redefined in the default Macro Library, as `x=m0`, `a=m1`, `b=m2`, `c=m3`, and `d=m4`. This mimics algebraic notation and simplifies equation entry. These quantities can be redefined by the user in the Macro Library, if desired, as can also the remaining five adjustable parameters.

Importantly, the "general" fit output includes estimates of the standard errors of the parameters, and of χ^2 and Pearson's correlation coefficient *R* for the fit. The quantity χ^2 ("Chisq" in KG) is defined as

$$\chi^2 = \sum w_i \delta_i^2 = \sum \left(\frac{\delta_i}{\sigma_i} \right)^2 \qquad (10)$$

where δ_i is the (calculated – observed) residual for the *i*th point and σ_i is the standard deviation of y_i. The σ_i values must appear in a third data column for a weighted fit; for unweighted, KG takes $\sigma_i = 1$ for all *i*. In the latter case "Chisq" is not really χ^2, rather it is the sum *S* of squared residuals, from which the estimated variance in *y* as $s_y^2 = S/(N - p)$ can be computed, where *N* is the number of data points and *p* the number of adjustable parameters. ($N - p$ is also known as the number of degrees of freedom, *f*.) For a weighted fit (including one of constant σ_i) in which the correct absolute σ_i values are used, "Chisq" *is* an estimate of χ^2 for the fit. Accordingly its value should approximate *f*, its theoretical average value *(13, 14).* Significant deviation from this value (or of S/f from unity) indicates flaws in the data, weights, or the fit model.

Illustrations

Single-Band Fits

It is instructive to work with the spectra for $H^{35}Cl$ of Schwenz and Polik *(3),* since the data have been tabulated and subjected to global analysis by these authors. Figure 1 illustrates the results obtained when just the fundamental band

is fitted to eq 8 and to an altered form with the two D_υ values taken to be identical. Two different general fits must be defined to obtain both sets of results on a single plot; however, the second fit can be obtained by simply copying the first expression into the definition box for the second fit and then replacing each "D1" by "D0". The results show that separate values for D_0 and D_1 are warranted, since (i) the values so obtained differ by more than their combined standard errors and (ii) s_y is significantly larger for the four-parameter fit than for the five-parameter fit.

Note that to achieve the fit results illustrated in Figure 1, the fit function is entered *exactly* as illustrated in the boxed text in the figure (after inserting the definitions m=m0, nu0=m1, B0=m2, B1=m3, D0=m4, and D1=m5 in the Macro Library). It is easier to use this direct transcription of eq 8 than to rewrite it factorized into single terms for each adjustable parameter, but the results are identical, including those for the parameter standard errors. On the other hand it would be a big mistake to simply fit to a quartic polynomial in *m*. That would necessitate a subsequent algebraic solution to extract the desired parameters from the polynomial coefficients. Even worse, correlated error propagation would be required to obtain the correct parameter standard errors *(15).* Such methods are not feasible in most instructional settings. By contrast, once the data have been entered and plotted, the fitting illustrated in Figure 1 can be accomplished as fast as the parameters can be defined in the Macro Library and the boxed expression entered correctly, which takes only minutes if transcription errors are avoided.[5]

The fit results in Figure 1 are shown exactly as returned from the KG program. While the user has some control over formatting, it is generally not possible to output all the parameters and their associated standard errors with the appropriate numbers of significant figures. The truncation of the fit definition expression in the first line is also unavoidable for long fit definitions. On the other hand, all quantities in the results table can be edited for display using the KG text editor.

It is trivially easy to extend the LS analyses of Figure 1 to the other bands—the overtone of $H^{35}Cl$ and both bands for $H^{37}Cl$—the data for which can be in the same or in separate data sheets. The user picks "template" on the "gallery" menu and selects the appropriate data columns. When "new plot" is clicked, the fits proceed automatically and are displayed on a new plot formatted exactly the same as Figure 1. Of course, for the overtone bands the results for the upper level ($\upsilon = 2$) will be incorrectly labeled with "1," subject to correction by redefinition or by editing the text.

The results for the $H^{35}Cl$ overtone band are shown in Figure 2. The case for separate D_υ values is less convincing here, as the two values from the first fit now agree within their (larger) combined errors, and s_y rises by only 10% on reducing the number of adjustable parameters by one. It is interesting that this s_y value is a factor of two smaller than for the fundamental; in our experience using a 10-cm absorption cell, the much weaker overtone spectra are noisier, hence give larger s_y values (even using ~1 atm of HCl for the overtone versus ~0.1 atm for the fundamental).

Suppose we repeat the computations behind Figures 1 and 2 using the weighted option and entering as the σ values in the "weights" column the exact values for s_y computed from

A

y = nu0 + (B0+B1)*m + (B1-B0-D1+D0)*m^ 2 - 2*(D0+D1)*m^ 3 + (D0-D1)*m^ 4		
	Value	Error
nu0	5667.832	0.008
B0	10.4394	0.0010
B1	9.8339	0.0009
D0	5.11 e-4	9 e-6
D1	4.98 e-4	7 e-6
Chisq	0.004526	NA
R	1.000	NA

B

y = nu0 + (B0+B1)*m + (B1-B0-D0+D0)*m^ 2 - 2*(D0+D0)*m^ 3 + (D0-D0)*m^ 4		
	Value	Error
nu0	5667.820	0.007
B0	10.4383	0.0010
B1	9.8340	0.0010
D0	4.98 e-4	8 e-6
Chisq	0.006230	NA
R	1.000	NA

Figure 2. Similar results for the first overtone (2–0 band) of $H^{35}Cl$ from ref *3*: (A) separate D_0 and D_1 values and (B) a common value (D_0) for both υ levels. Here the values have been edited for display. The estimated s_y values are 0.018 and 0.020 cm^{-1}, respectively.

the KG Chisq values (e.g., 0.03671 for the first fit in Figure 1). The fit will return identical values for all parameters and their errors, and the exact expected value for χ^2 (31 lines − 5 parameters = 26 in the first fit in Figure 1). If the entered σ values are changed, the parameters will be unaffected, but the standard errors will scale with σ while the Chisq output will scale with σ^{-2}. This illustrates the different ways KG handles weighted and unweighted fits and will be useful below *(9, 13).*

An approximate calculation of H_e from eq 5c yields 1.7×10^{-8} cm^{-1}, for which the term that is cubic in κ in eq 3 yields contributions > 0.2 cm^{-1} for the largest *J* values in the data set. This is both systematic and much larger than the estimated s_y for the fits, so it is possible that this term should be included in the fit model. Rather than modify eq 8 for inclusion of the H_υ terms, which requires some tedious algebra, let us use eq 3 directly. To expedite entry of the fit function, we will also use the Macro Library to define several key quantities. First we check the new approach on the fit to eq 8, by entering:[6]

```
k(m) = (m*(m+1));
F0(m) = (B0*k(m)-D0*k(m)^2);
F1(m) = (B1*k(m)-D1*k(m)^2);                    (11)
fun(m) = ((m>0)?(nu0+F1(m)-F0(m-1)):
                (nu0+F1(-m-1)-F0(-m)));
```

Users familiar with the C programming language will recognize the branching test in fun(m), used to distinguish the R ($m = J + 1$) and P lines ($m = -J$) for different treatment. "fun(m)" is entered in the fit definition box, yielding results that agree exactly with those displayed in Figure 1. Next the significance of H_υ is checked by defining H0=m6 and H1=m7 in the library and adding "+H0*k(m)^3" to F0(m) and "+H1*k(m)^3" to F1(m). This fit gives a 6% reduction in Chisq (Figure 3) but a slight increase in s_y (since *f* is reduced from 26 to 24). H_0 and H_1 are only marginally significant, but their inclusion in the fit model has produced systematic changes in the D_υ values and sizable increases in their standard errors. Use of a common H_υ value for the two

υ levels yields a value almost identical to that for H_1. All of these estimates are statistically consistent with the theoretical values—1.67 and 1.62 × 10^{-8} cm^{-1} for H_0 and H_1, respectively *(16)*—and one is justified in fixing these parameters at these values in further fitting. This change leads to the second set of results in Figure 3.

Global Fits

To expand the analysis to include both bands simultaneously, we must put all the data in the same two columns and alter the *m* values for one band to distinguish it from the other.[7] I have done this by just adding 30 to all the *m* values for the overtone. The Macro Library is again used to facilitate function entry. First, define `fn(m)` as the function given in Figure 1 and set `nu1=m6, B2=m7,` and `D2=m8`. Then define the overtone band similarly, but with `nu1, B2,` and `D2` substituted for their counterparts in the fundamental:

```
ov(m) = (nu1 + (B0+B2)*m
      + (B2-B0-D2+D0)*m^2                (12)
      - 2*(D0+D2)*m^3
      + (D0-D2)*m^4);
```

Both bands are thus represented at the level of eq 8. The fit definition function now contains a conditional test to distinguish the data for the overtone (m > 18) from those for the fundamental, yielding the results shown in Figure 4.

If the fit is repeated with weights, in recognition of the apparent factor of two greater precision for the overtone, many of the parameters change by more than their apparent standard errors, and χ^2 = 69.6. This is significantly larger than its expected value of 42 (50 lines, 8 parameters) and suggests some systematic errors in the data or limitations of the model *(17)*.[8] In this regard, the small value for D_2 seems anomalous. (Correction for the H_υ term does not significantly alter this situation, reducing χ^2 only to 67.5.) In light of these questions, the further fits described below were done with equal weights for all lines (i.e., unweighted).

To fit to the "e" parameters in eqs 4a–d, we must again define the key parameters in the Macro Library, for example, `we=m1; wx=m2; ....` The B_υ, D_υ, and $\tilde{\nu}_0$ values are then defined explicitly in terms of these, for example, for the overtone band origin, `nu1=(2*we-6*wx);`.[W] The fit function defined in Figure 4 can then be applied directly, yielding results shown in Figure 5A, from which it can be seen that (i) `Chisq` is significantly greater than in Figure 4 and (ii) the

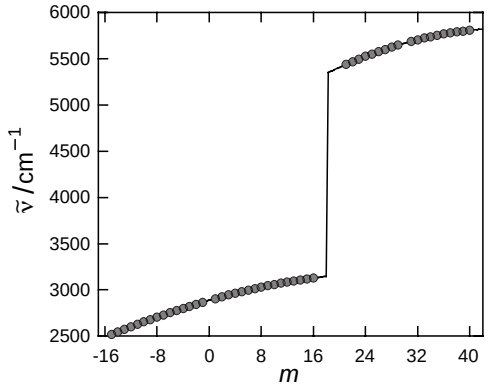
A

y = fun(m)		
	Value	Error
nu0	2885.643	0.016
B0	10.4370	0.0017
B1	10.1367	0.0016
D0	5.27 e-4	1.4 e-5
D1	5.31 e-4	1.2 e-5
H0	4.2 e-8	3.5 e-8
H1	3.1 e-8	2.8 e-8
Chisq	0.03301	NA
R	1.000	NA

B

y = fun(m)		
	Value	Error
nu0	2885.640	0.013
B0	10.4361	0.0009
B1	10.1360	0.0009
D0	5.176 e-4	2.9 e-6
D1	5.243 e-4	2.6 e-6
Chisq	0.03369	NA
R	1.000	NA

Figure 3. Results for fundamental band with inclusion of H_υ centrifugal distortion term in the model, with both H_υ values freely fitted (A) or fixed at their calculated values from ref *16* (B).



y = ((m>18)?(ov(m-30)):(fn(m)))		
	Value	Error
nu0	2885.645	0.01324095
B0	10.43755	0.0007416346
B1	10.13729	0.0007295697
D0	0.0005165513	2.584545e-06
D1	0.0005224596	2.334812e-06
nu1	5667.835	0.01662221
B2	9.831383	0.001176677
D2	0.0004923015	9.293977e-06
Chisq	0.06024322	NA
R	1	NA

Figure 4. Simultaneous fit of the IR fundamental and first overtone bands to B_υ and D_υ values for υ = 0, 1, and 2, and two band origins. The fit results are presented in "raw" form to facilitate consistency checks and detailed comparisons.

A

y = ((m>18)?(ov(m-30)):(fn(m)))		
	Value	Error
we	2989.281	0.03420965
wx	51.79673	0.01173693
Be	10.58892	0.0009809508
ae	0.301529	0.0002436824
De	0.0005193061	3.547938e-06
bt	7.824433e-07	1.210022e-06
Chisq	0.1042565	NA
R	1	NA

C

y = ((m>18)?(ov(m-30)):(fn(m)))		
	Value	Error
we	2989.106	0.043
wx	51.73	0.02
Be	10.5856	0.0010
ae	0.29465	0.00129
De	5.00 e-4	5 e-6
bt	4.2 e-5	1.0 e-5
ge	-0.0028	0.0005
p	-1.8 e-05	5 e-06
Chisq	0.060362	NA
R	1	NA

B

y = ((m>18)?(ov(m-30)):(fn(m)))		
	Value	Error
we	2989.101	0.04293362
wx	51.72791	0.01557496
Be	10.58555	0.0009892759
ae	0.2945961	0.00129171
De	0.0005000723	4.992371e-06
bt	4.19747e-05	9.693309e-06
ge	-0.00282844	0.0005419427
p	-1.80332e-05	4.607272e-06
Chisq	0.06024322	NA
R	1	NA

Figure 5. Same global fit, in terms of the "e" parameters of eqs 4, for linear representations of B_υ and D_υ (A) and quadratic (B). The results of a systematic rounding-refitting procedure are given in (C).

parameter bt (β_e) is statistically insignificant. By setting this parameter to zero in the Library, we reproduce exactly the multiple regression results given by Schwenz and Polik in their Table 2 *(3)*.[9] To regain the statistical quality of the eight-parameter fit in Figure 4, it is necessary to add the quadratic terms to the definitions of the B_v and D_v values [e.g., B2=(Be-2.5*ae+6.25*ge);].[W] This yields the results in Figure 5B, which are statistically equivalent to those in Figure 4 (i.e., identical values of Chisq and, by back calculation, of all fitted parameters in Figure 4).

Reporting the Results

The critical reader might have noticed that several of the parameters given in Figures 2 and 3 were stated to more significant figures than seem necessary from their estimated standard errors. For example, in Figure 2A, why not report the first three parameters with one less digit each? The goals of an LS fit often include the compact representation of the data in a way that permits reliable recalculation of observed data and prediction of unobserved, and to that end one must exercise care in rounding the LS fit results. The reason is that the fit parameters are often highly correlated, and naive rounding of correlated parameters can lead to unacceptable loss in reliability of the fit results. In the case of the aforementioned results in Figure 2, removing one significant digit in each of the quantities results in increases of Chisq by 43% in (A) and 73% in (B). A safe and straightforward way to handle this problem is to round the parameters one at a time, refitting the remaining parameters each time *(18)*.[10] This procedure is easy to apply using KG,[W] and it yields the results given in Figure 5C for the fit summarized in Figure 5B. The resulting rise in Chisq is a modest 0.2%. By contrast, rounding we, Be, and ae to 2989.10, 10.586, and 0.295—values that seem reasonable from Figure 5B—results in a nearly threefold increase in Chisq.

Extension to Multiple Isotopomers

The "e" representation of Figure 5B, while exactly equivalent to the fit in Figure 4, can be used to expand the global analysis to more than one isotopomer, by taking advantage of the isotopic relations in eqs 7. This entails defining expressions for the new parameters that include their dependence on ρ_i [e.g., r=0.999243017; B0r=((Be-ae*r/2+ge*r^2/4)*r^2);].[W] It is also necessary to define new fundamental and overtone functions containing these "r" parameters; and additional branches are required in the fit-definition function and must be incorporated in the *m* values for the second isotopomer (e.g., by adding 60 to all the *m* values for the fundamental of $H^{37}Cl$ and 90 to the overtone). This kind of fit is especially sensitive to absolute errors in the wavenumbers, since now there is redundancy in the vibrational parameters. It is illustrated in the Supplemental Material[W] for very accurate and precise literature data *(19–21)*.[11]

Comments

The implementation of the computations described here is truly as easy as entering the data in the data sheet and the various functions in the fit-definition box and library. The spreadsheet and MathCad-based approaches are much more demanding, especially with respect to the kinds of trial-and-error modifications I have illustrated here, like adding or removing parameters, or setting them to constant values. Having said this, I must still acknowledge that KaleidaGraph is not very helpful when mistakes are made. For example, an error as simple as unbalanced parentheses in a library function will produce the ubiquitous "syntax error" warning from the fit-definition routine. Thus, it is wise to start small and build up to the more complex models. To that end it is useful to know that the function definitions can be saved easily, either in the Macro Library itself, or by copying and pasting into the "posted note" of the data sheet. Errors in library functions will not register on entry, but these functions can be tested individually using the "formula entry" window. Of course no instructor would want to make this kind of project the students' first encounter with KaleidaGraph or, for that matter, with spreadsheet computations, either.

Equations 5a–c for the key centrifugal distortion parameters can be incorporated easily in the "e"-representation fits, like those summarized in Figure 5 and employed for multiple isotopomers.[W] Their inclusion reduces the number of adjustable parameters but renders the fit nonlinear. Because of the small systematic errors in the data analyzed here,[11] these constraints produce large increases in Chisq—for example, a two-fold increase when eq 5a is used for D_e in Figure 5B. Very precise data fully justify their use and even demonstrate a need to correct for deviations from the Born–Oppenheimer model behind eqs 7 *(16)*.

Students groan under the data-analysis burden placed upon them by the HCl experiment, especially when data are analyzed for more than one isotopomer. No analysis method can fully remove that burden, but some can reduce the "pain-to-gain" ratio. In most settings, it would not be reasonable to expect students to perform all the different calculations I have described here, so instructors will need to preview typical spectra, decide upon a suitable computational strategy, and guide the students accordingly. Students tend to think in terms of "right" and "wrong" when it comes to their computational work, so the notion of testing parameters for significance will likely require special nurturing.

Neither KaleidaGraph nor the global-fitting techniques described here are in any way limited to spectroscopy applications. For example, the multiple sets of Charles's law data used to estimate absolute zero in a recent contribution to this *Journal (22)* can be readily analyzed using KG and the kind of data sorting techniques applied here in Figures 4 and 5.

[W]Supplemental Material

Global-fitting examples are available in this issue of *JCE Online*.

Notes

1. This relatively inexpensive program (from Synergy Software, Reading, PA) is available for both Macs and PCs, with files interchangeable across platforms.

2. Another frequently used notation is that of Dunham *(10)*, which explicitly recognizes the double-polynomial nature of E/hc by expressing it as $\sum_{ij} Y_{ij} z^i \kappa^j$. For key parameters, the correspondence is as follows: $Y_{10} = \omega_e$, $Y_{20} = -\omega_e x_e$, $Y_{01} = B_e$, $Y_{11} = -\alpha_e$, and $Y_{02} = -D_e$.

3. There is much misunderstanding of this point in the literature. In the "inverse Hessian" formulation of nonlinear least squares (LS) *(14),* one can show that for a linear problem the equations for the corrections to the parameters, starting with initial values of zero, are identical to those obtained directly via a linear approach *(13).* Thus, for example, using Excel's Solver routine on a linear LS problem does not make it nonlinear.

4. In the earlier versions of KG, the default initial values of 0 yielded "singular matrix" failure, even for linear fit models; in the latest version (3.6) it appears that initial values are not needed for linear fits. An option for entering partial derivatives seems rarely to be needed, which is fortunate, because having to enter this information would greatly reduce the utility of the program. Without this option checked, the needed derivatives are estimated numerically.

5. These fit expressions are entered once in the appropriate fit definition boxes; then the boxes are clicked shut, and the fit proceeds. There is no copying of formulas from cell to cell, as in Excel. Also, there are just two data columns—the m values and their associated $\tilde{\nu}(m)$.

6. It is wise to enclose all functions defined in the Library in parentheses, to ensure their proper interpretation when parsed out by the fitting program. The semicolon terminates each definition and can be used to "inactivate" definitions, since all following material is taken as comments.

7. The latest version of the KG program (3.6) does permit use of multiple columns in the data sheet as independent variables; see the Supplemental Material[W] for illustrations.

8. From Table C.4 in ref *17*, a value of χ^2 this large occurs only ~1% of the time.

9. In a linear fit, if any parameter has a standard error larger than its magnitude, setting that parameter to zero is guaranteed to decrease s_y for the fit *(17).* Hence removal of the parameter from the fit model is statistically justified, unless its inclusion is warranted on other than ad hoc statistical grounds.

10. The remaining parameters become progressively more precise in this sequential rounding, refitting procedure. As a guideline one is usually safe in rounding by $\sim\sigma/4$, and it is the current σ that is relevant here. However, the proper parameter error to report is the one from the original all-parameter fit.

11. Most of the results obtained here for the spectra of ref *3* are not in statistical agreement with the current best values *(16).* A comparison of these spectra with the tabulated wavenumbers in ref *19* shows discrepancies ranging from 0.0 cm^{-1} to 0.35 cm^{-1} for the fundamental, with a systematic dependence on J. For the overtone, the difference is a more nearly constant 0.15 cm^{-1}.

Literature Cited

1. Shoemaker, D. P.; Garland, C. W.; Nibler, J. W. *Experiments in Physical Chemistry,* 6th ed.; McGraw-Hill: New York, 1996; pp 397–404.
2. Iannone, M. *J. Chem. Educ.* **1998,** *75,* 1188–1189.
3. Schwenz, R. W.; Polik, W. F. *J. Chem. Educ.* **1999,** *76,* 1302–1307.
4. Zielinski, T. J. *J. Chem. Educ.* **2000,** *77,* 668–670.
5. Ogren, P.; Davis, B.; Guy, N. *J. Chem. Educ.* **2001,** *78,* 827–836.
6. Feller, S. E.; Blaich, C. F. *J. Chem. Educ.* **2001,** *78,* 409–412.
7. Glendening, E. D.; Kansanaho, J. M. *J. Chem. Educ.* **2001,** *78,* 824–826.
8. Herzberg, G. *Spectra of Diatomic Molecules;* D. Van Nostrand: Princeton, NJ, 1950.
9. Tellinghuisen, J. *J. Chem. Educ.* **2000,** *77,* 1233–1239.
10. Dunham, J. L. *Phys. Rev.* **1932,** *41,* 721–731.
11. Tellinghuisen, J.; McKeever, M. R.; Sur, A. *J. Mol. Spectrosc.* **1980,** *82,* 225–245.
12. Tellinghuisen, J. *J. Mol. Spectrosc.* **2003,** *221,* 244–249.
13. Tellinghuisen, J. *J. Phys. Chem. A* **2000,** *104,* 2834–2844.
14. Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes;* Cambridge University Press: Cambridge, United Kingdom, 1986.
15. Tellinghuisen, J. *J. Phys. Chem. A* **2001,** *105,* 3917–3921.
16. Coxon, J. A.; Hajigeorgiou, P. G. *J. Mol. Spectrosc.* **2000,** *203,* 49–64.
17. Bevington, P. R.; Robinson, D. K. *Data Reduction and Error Analysis for the Physical Sciences,* 2nd ed.; McGraw-Hill, New York, 1992.
18. Tellinghuisen, J. *J. Mol. Spectrosc.* **1989,** *137,* 248–250.
19. Rank, D. H.; Rao, B. S.; Wiggins, T. A. *J. Mol. Spectrosc.* **1965,** *17,* 122–130.
20. Webb, D. U.; Rao, K. N. *Appl. Opt.* **1966,** *5,* 1461–1463.
21. Webb, D. U.; Rao, K. N. *J. Mol. Spectrosc.* **1968,** *28,* 121–124.
22. Salter, C. *J. Chem. Educ.* **2003,** *80,* 1033–1035.