

The Analysis of Vibrational Spectrum of a Linear Molecule ©

Richard W. Schwenz, Department of Chemistry and Biochemistry
University of Northern Colorado, Greeley, CO 80639
rwschwe@bentley.unco.edu

William F. Polik, Department of Chemistry
Hope College, Holland, MI 49423
polik@hope.edu

Sidney Young, Department of Chemistry
University of South Alabama, Mobile, AL 36688
syoung@jaguar1.usouthal.edu

© Copyright 2000 by the Division of Chemical Education, Inc., American Chemical Society. All rights reserved. For classroom use by teachers, one copy per student in the class may be made free of charge. Write to JCE Online, jceonline@chem.wisc.edu, for permission to place a document, free of charge, on a class Intranet.

Objectives: after using this document you will be able to:

1. analyze the infrared spectrum of a linear molecule, for example HCl or CO.
2. perform a regression using more than one independent variable.
3. explain, orally or in written form, the relationship between energy levels and spectroscopy.

Assumptions/ Prerequisites: You should

1. have read and understood the appropriate sections of your lecture text on the rigid rotor and harmonic oscillator models and their spectroscopy, and on the infrared spectra of diatomic and polyatomic molecules
2. be able to read, interpret, and edit a Mathcad document
3. have read the experimental introduction, and collected the data on the infrared spectrum of a diatomic molecule. Many of the equations you will need are contained in the laboratory notes provided by your instructor
4. have read the assigned sections in the article by Schwenz and Polik (reference 1) to see how a number of different analyses of FT-IR data are performed; most important is the description of the use of the energy expression (equation 3 in reference 1), the subtraction of two similar energy expressions (equations 4a and 4b in reference 1), and a mechanism for finding the molecular constants from a fit of the data to the energy expression difference.
5. be familiar with the concept of non-linear regression and how to perform a linear regression (see reference 5).

Introduction:

This document will lead you through the analysis of the infrared spectrum of a linear molecule. Let's begin by writing expressions for the energy levels of the rigid rotor and harmonic oscillator in the space below and labeling those as equation 1. Units of wavenumbers (cm^{-1}) are particularly useful. Use your text or laboratory manual to refresh your memory.

Place your equation 1 here: (1)

To obtain the vibration/rotation energy levels of a linear diatomic molecule, the energy expressions for the harmonic oscillator and rigid rotor models are combined into one expression. That expression does not adequately describe most molecules so it is extended by expansion to a five term energy expression with rigid rotor and harmonic oscillator level modified to second order in $(v+0.5)$ and $J(J+1)$ where v is the vibrational quantum number and J is the quantum number for rotation. The expansion to second order (power 2) accounts for the anharmonicity ($\omega_e x_e$), vibration - rotation interaction (α_e), and centrifugal distortion (D_e) terms mentioned in laboratory texts. Equation (2) is

$$E_{v,J} = \omega_e(v+1/2) - \omega_e x_e(v+1/2)^2 + B_e J(J+1) - \alpha_e(v+1/2)J(J+1) - D_e J^2(J+1)^2 \quad (2)$$

A spectrum is produced by the system changing from one energy level to another. The difference between two energy levels is obtained by subtracting the expressions for the two energies using equation (2) where the proper quantum numbers for each energy level are used in the energy expression for each energy level.

Remember that our task is to take the two energy expressions, one for each energy level, compute the difference between these energy levels, and then use this difference and the non-linear equation fitting process to obtain the five spectroscopic constants characteristic of the molecule. In our fitting process the observed spectrum is the dependent (y) variable. The five combinations of quantum numbers from equation (2) are the independent variables. The five spectroscopic constants are the fitting parameters. The fitting parameters are chosen by the algorithm used in this document to give a curve that minimizes the squared deviation of each experimental point, observed frequency, from the fitted data, a computed frequency.

The calculation in this document begins by reading the file that contains the experimental line positions and quantum numbers of the spectral lines for the vibration of a linear molecule. You create this file from the data you obtained in the laboratory. The file of data can be created using your favorite word processing program or a spreadsheet. After you enter the data into your software you should save it as an ascii file or export the spreadsheet as a comma separated file. It helps to give this file either a prn or txt extension. The data in your comma or space delimited file should contain the data as a list with the line number (1,2,...), spectral line frequency, lower state vibrational quantum number, upper state vibrational quantum number, lower state rotational quantum number, and upper state rotational quantum number in that order from left to right. Remember you should give the file a name that Mathcad can use. Check the Mathcad manual for more information about data file names. For example, to load a different data file into this document, save it with a name like mydata.prn and edit the following line to read **Data:= READPRN(mydata)**. Remember to check your Mathcad documentation to see how to read data files into a Mathcad document for your version of Mathcad.

Data := READPRN("hcl.txt")

For Mathcad 6, 7, and 8 use the correct READPRN command

The hcl.txt file that accompanies this document should be in the same directory as this Mathcad template. The test data in hcl.txt illustrates the structure of the data file, and allows you to practice with this document without worrying about the using your own experimental data at first. Obviously, to find the molecular constants associated with your molecule from your experiment you need to create a data file similar to this one. Open a word processing document such as WordPad or NotePad to look at the sample data file now. Notice how hcl.txt is a list of line numbers and quantum numbers associated with each spectral line where the spectral line line is recorded as the frequency in cm^{-1} . Close this file without changing it.

The first thing we need to do when analyzing data is to get the data into the Mathcad software memory. Remember, we read in an array of data, i.e. a set of data in six columns, see the READPRN command above. This data is now in a matrix called **Data**. Next we will assign each column of **Data** to a specific variable name. Remember that the number of rows in **Data** is equal to the number of spectral lines.

$N := \text{length}(\text{Data}^{<0>})$

Here we determine the number of rows in **Data**. Recall that Mathcad starts numbering rows and columns in arrays with the integer zero.

$N = 50$

The value of N shown here should equal the number of lines you obtained in your spectrum or the number of rows in hcl.prn.

$\text{lineno} := \text{Data}^{<0>}$

Here we assign the data in column zero to the variable lineno (line number) and the data in column five to ExpFreq (experimental frequency). Notice how the variable name helps us to remember what is stored in that variable.

$\text{ExpFreq} := \text{Data}^{<5>}$

You should associate the vibrational quantum numbers from columns 1 and 2 of the **Data** matrix and rotational quantum numbers from columns 3 and 4 in the same way to vlow, vup, Jlow, and Jup respectively.

$\text{vlow} := \text{Data}^{<1>}$ $\text{vup} := \text{Data}^{<2>}$

$\text{Jlow} := \text{Data}^{<3>}$ $\text{Jup} := \text{Data}^{<4>}$

These variables should not be included in the final Mathcad worksheet distributed to students. When students enter this themselves they learn more about the data analysis and software.

This next equation (equation 3) presents the energy expressions for the lines in a diatomic molecule. The molecular constants to be determined by the non-linear curve fitting process will be located in the vector called **MolecularConstants**. This vector contains ω_e , $\omega_e x_e$, B_e , α_e , and D_e in turn from top to bottom. This means that the zero element of the vector will contain ω_e . Remember that the molecular constants are the fitting parameters that we're trying to determine.

Note that the structure of *this document requires that data on multiple vibrational states be included* in the array **Data**. This means that a fundamental and first overtone spectrum should be obtained experimentally. Ideally the data has equal uncertainties (error bars) in the frequency as one expects from an FTIR instrument.

Compare the following equation to the first energy expression given in this worksheet. Which element in the vector **MolecularConstants** is associated with the spectroscopic constant B_e ?

$$\begin{aligned}
 E(\text{MolecularConstants}, v, J) := & \text{MolecularConstants}_0 \cdot (v + 0.5) \dots \\
 & + - \text{MolecularConstants}_1 \cdot (v + 0.5)^2 \dots \\
 & + \text{MolecularConstants}_2 \cdot (J \cdot (J + 1)) \dots \\
 & + - \text{MolecularConstants}_3 \cdot (v + 0.5) \cdot J \cdot (J + 1) \dots \\
 & + - \text{MolecularConstants}_4 \cdot J^2 \cdot (J + 1)^2
 \end{aligned} \tag{3}$$

We recognize that the energies of the lines in a spectrum are due to the difference in energy between the upper and lower states. We use equation 3 to obtain the energy expression for both the upper and lower state of a transition. This leads to equation 4.

$$\begin{aligned}
 \text{FitFreq}(v_{\text{low}}, v_{\text{up}}, J_{\text{low}}, J_{\text{up}}, \text{MolecularConstants}) := & E(\text{MolecularConstants}, v_{\text{up}}, J_{\text{up}}) \dots \\
 & + - E(\text{MolecularConstants}, v_{\text{low}}, J_{\text{low}})
 \end{aligned} \tag{4}$$

Note the Mathcad treatment of terms in equations that are wider than one page (the + -), e.g. in equations 3 and 4.

Notice how the FitFreq is given by the difference between two energy levels. We will collect the quantum numbers into columns of the matrix called **QNComb**. The combinations of quantum numbers will be the coefficients of the MolecularConstants in the energy expressions, equation (3), that will be used in the non-linear curve fitting process to reproduce the observed line frequencies. Remember that the curve fitting process will minimize the deviations squared between the experimental frequencies and the computed frequencies for the spectrum of the diatomic molecule. The columns in the **QNComb** matrix correspond to the independent variables in the regression (and in the energy expressions 3 and 4), and the observed line frequencies are the dependent variables.

This Mathcad document uses a matrix formulation to determine the molecular constants for the non-linear regression. For further information you should consult references 4 and 5. In order to use this method you need to define the independent variables in a matrix with the number of columns equal to the number of independent variables and the numbers of rows equal to the number of data points. In this document, the matrix **QNComb** (Quantum Number Combination) is that matrix

$$\text{QNComb}^{<0>} := \overrightarrow{((\text{vup} + 0.5)) - ((\text{vlow} + 0.5))}$$

You should now be able to enter in the expressions for columns 1 through 4 of the QNComb matrix from the energy expressions 3 and 4. Use the space provided below.

$$\text{QNComb}^{<1>} := - \overrightarrow{((\text{vup} + 0.5))^2 - (\text{vlow} + 0.5)^2}$$

These equations should not appear in the student version. Students will enter these expressions here as an exercise.

$$\text{QNComb}^{<2>} := \overrightarrow{(\text{Jup} \cdot (\text{Jup} + 1)) - (\text{Jlow} \cdot (\text{Jlow} + 1))}$$

$$\text{QNComb}^{<3>} := - \overrightarrow{((\text{vup} + 0.5) \cdot \text{Jup} \cdot (\text{Jup} + 1)) - ((\text{vlow} + 0.5) \cdot \text{Jlow} \cdot (\text{Jlow} + 1))}$$

$$\text{QNComb}^{<4>} := - \overrightarrow{(\text{Jup} \cdot (\text{Jup} + 1))^2 - (\text{Jlow} \cdot (\text{Jlow} + 1))^2}$$

The following line represents a standard algorithm for determining the molecular constants vector (the values of the constants associated with your molecule) by multiple regression with the **QNComb** matrix of independent variables and the experimental frequencies, the dependent variable. Further information on the formulas used in this regression technique are available in references 4 and 5.

$$\text{MolecularConstants} := \left(\text{QNComb}^T \cdot \text{QNComb} \right)^{-1} \cdot \left(\text{QNComb}^T \cdot \text{ExpFreq} \right)$$

We just calculated the numerical values of each of the molecular constants; let's see what the values are! Write down the numerical values of the molecular constants ω_e, \dots , given here in your notebook.

$$\text{MolecularConstants} = \begin{bmatrix} 2.989 \cdot 10^3 \\ 51.796 \\ 10.589 \\ 0.302 \\ 5.206 \cdot 10^{-4} \end{bmatrix}$$

Because we have five independent variables (the five combinations of quantum numbers) and one dependent variable (the line frequencies), we would have to look at a six dimensional graph to display the results (something we're not very good at). Two ways to check for correctness of your results are: (1) to look at the standard deviations of the resulting molecular constants (they should be small with respect to the value itself) and, (2) to look at the residual for each point (the difference between the experimental and fitted line energy). To do this we start by calculating the sum of squares of the deviations over the full range of data; this should be small as we minimized the least square deviation in the curve fitting process we just completed.

The fitting parameters we obtained above permit us to calculate the spectral frequencies for our molecule. The calculated frequencies can then be compared to the experimental frequencies. This will give us some idea about how good our fitting parameters are. In fact this is indeed the way we can get at the standard deviations of the fitting parameters and also a way to see if we have any bad data points. We should now calculate the deviations between the experimental and fitted line positions.

$$i := 1..(N - 1)$$

$$\text{Residual}_i := \text{ExpFreq}_i - \text{FitFreq}(\text{vlow}_i, \text{vup}_i, \text{Jlow}_i, \text{Jup}_i, \text{MolecularConstants})$$

$$\text{SumSquares} := \sum_{i=0}^{N-1} \frac{(\text{Residual}_i)^2}{N-5}$$

Why do we call this variable "SumSquares"?

Why is N-5 in the denominator?

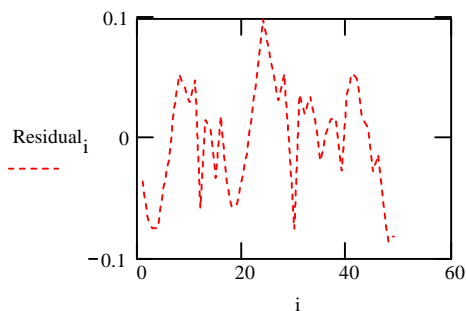
$$\text{SumSquares} = 2.313 \cdot 10^{-3}$$

Is the sum of squares small as we expected?

Record the experimental resolution and wavenumber accuracy here. Compare the magnitude of the largest residual to the experimental resolution, and also to the accuracy of the wavenumbers from your instrument. Comments?

You should determine if any of the residuals are larger than the experimental resolution now. If they are you probably have a data input error on the line position. Correct the error in your HCl.txt file where you typed your experimental data. Save the file as before and then refit by recalculating this worksheet.

Construct a plot of the residual vs. i . Are there any observable systematic trends in the residuals? Are there any outlying points? If so, what might they be due to?



This plot should not be included in the document distributed to students. Students should be able to construct this plot and answer the questions on their own.

Questions about the residuals:

1. Are the residuals distributed symmetrically around zero?
2. Would the average value of the residuals tell us anything about the significance of our data?
3. Should we consider dropping some points to improve the fit and the distribution of the residuals?
4. How would one decide how to determine when to drop a point from the data set?

Printout^{<0>} := Data^{<0>}

Printout^{<1>} := Data^{<1>}

Printout^{<2>} := Data^{<2>}

Printout^{<3>} := Data^{<3>}

Printout^{<4>} := Data^{<4>}

Printout^{<5>} := Data^{<5>}

Printout^{<6>} := Residual

Here we accumulate all the data and the vector of residuals into the matrix **Printout**. We can now export the matrix to a file for viewing with other software and printing. We can also look at the data here in Mathcad. Print all the data back out for examination of the deviations (residuals) for data examination for errors.

Instructors are encouraged to leave most of these commands out of student versions of the document so that they can assume greater responsibility for what is happening in this document.

WRITEPRN("output.prn") := Printout

Here we create a file of our results to view with other software such as a word processor or spreadsheet program. You may also view the contents of the Printout matrix here in this worksheet by typing Printout= in the space to the right. Be sure you use the correct WRITEPRN command, i.e. the one that matches your version of Mathcad.

Now let's determine the errors in each of the molecular constants by looking at the diagonal elements of the variance matrix which we will create below. Remember that the error associated with a molecular constant should be small with respect to the value of that molecular constant. The variance matrix is the fancy term for the matrix which describes the correlation of the sum of squares changing as a result of changing two of the variables simultaneously. The diagonal elements are the square of the standard deviation in the molecular constants. The off-diagonal elements measure the extent of any relationship between variables. The correlation coefficient that your calculator displays when you do a linear least squares is a measure of the ratio of the covariances to the variances in the linear case. Here we have more than one independent variable, so we have multiple covariances (which we don't look at here) and multiple variances. We want to calculate the standard deviation of a molecular parameter from the variance; we'll do the first one; you do the others!

$$\text{SigmaMolecularConstants}_0 := \sqrt{\text{SumSquares}\left[\left(\text{QNComb}^T \cdot \text{QNComb}\right)^{-1}\right]_{0,0}}$$

$$\text{SigmaMolecularConstants}_1 := \sqrt{\text{SumSquares}\left[\left(\text{QNComb}^T \cdot \text{QNComb}\right)^{-1}\right]_{1,1}}$$

$$\text{SigmaMolecularConstants}_2 := \sqrt{\text{SumSquares}\left[\left(\text{QNComb}^T \cdot \text{QNComb}\right)^{-1}\right]_{2,2}}$$

$$\text{SigmaMolecularConstants}_3 := \sqrt{\text{SumSquares}\left[\left(\text{QNComb}^T \cdot \text{QNComb}\right)^{-1}\right]_{3,3}}$$

$$\text{SigmaMolecularConstants}_4 := \sqrt{\text{SumSquares}\left[\left(\text{QNComb}^T \cdot \text{QNComb}\right)^{-1}\right]_{4,4}}$$

The last four of these expressions should not be included in the document distributed to students.

The molecular constants are given in the order ω_e , $\omega_e x_e$, B_e , α_e , and D_e from top to bottom in the vector **MolecularConstants**. Care needs to be taken in comparing the computed molecular constants with the literature values shown below because the number of expansion terms may differ between the method used in this worksheet compared to the method used by the original investigator as described in the literature.

$$\text{MolecularConstants} = \begin{bmatrix} 2.98928 \cdot 10^3 \\ 51.79551 \\ 10.58919 \\ 0.30167 \\ 5.20621 \cdot 10^{-4} \end{bmatrix} \quad \text{SigmaMolecularConstants} = \begin{bmatrix} 0.034 \\ 0.011 \\ 8.824 \cdot 10^{-4} \\ 9.76 \cdot 10^{-5} \\ 2.876 \cdot 10^{-6} \end{bmatrix}$$

Examine the SigmaMolecularConstants vector. Are the errors small with respect to the molecular constants? If they aren't, why? This is not a trick question.

$$\text{Lit} := \begin{bmatrix} 2989.74 \\ 52.05 \\ 10.5909 \\ 0.3019 \\ 0.00053 \end{bmatrix}$$

The literature values quoted in this document are for $^1\text{H}^{35}\text{Cl}$ (reference 6.)

Do all the molecular constants agree with the literature values for the analysis you are performing? Should they?

Well now you have this wonderful data on a molecule! So what can you do with it? How about calculating characteristics of the molecule, like the bond length, force constant for the bond and dissociation energy?

Rearrange equations (1) which you wrote in this worksheet to calculate the bond length, force constant, and dissociation energy. The energy expansion we use here is characteristic of the Morse potential. For the Morse potential it is known that the dissociation energy, $D_e = \omega_e^2 / 4\omega_e x_e$. Be careful of the units you use! Do that calculation here, along with the propagation of errors for the values. How does the error from your experimental data in the molecular constant compare with the size of the constant?

You may find the Mathcad document (reference 7) entitled "Exploring the Morse Potential: MorsePotential.mcd" by T. J. Zielinski useful.

If you have data on multiple isotopic variants of the same molecule $^1\text{H}^{35}\text{Cl}$, $^1\text{H}^{37}\text{Cl}$, $^2\text{H}^{35}\text{Cl}$, ... , then you should compare the bond lengths, force constants, and dissociation energies for all variants here. The Born-Oppenheimer approximation states that these quantities should be independent of isotopic variant. Are they? Is the approximation valid for the molecule under examination?

Extensions / Questions

1. What will you do if data on only the harmonic band (0 to 1) is available?
2. At what frequency would the 0 to 3 absorption band occur for the anharmonic oscillator we are using? Is this observable in the instrument you are using?
3. What modifications would be necessary to analyze the infrared spectrum of the asymmetric stretching mode of carbon dioxide or acetylene?

Mastery Exercise

4. What modifications would be necessary to analyze the infrared spectrum of the bending mode of carbon dioxide? To do this you need to look up the energy expression for the bending of carbon dioxide and separate that into (molecular constants)*(combinations of quantum numbers).
5. In the 1960's, noble gas compounds were new and exciting. Xenon difluoride was first synthesized in 1963, was quickly discovered to have a significant vapor pressure, and to react with almost everything (draw the Lewis structure of xenon difluoride). However, its structure was unknown because it was so reactive. The infrared spectrum was taken (Reichman & Schreiner, J. Chem. Phys., *vol. 51*, p. 2355, 1969) and discovered to be quite simple. Analyze the spectrum, show that the molecule must be linear, and determine the Xe - F bond distance. (You should note that this discovery helped prompt the discussion of non-octet rule compounds you had in general chemistry.)
6. One interesting alternative to the multiple linear regression would use Mathcad's MINERR function to minimize the sum of square of the residuals.

References

1. Schwenz and Polik, " Analysis of the Infrared Spectra of Diatomic Molecules", *J. Chem. Educ.* in press **1999**.
2. Shoemaker, Garland, Nibler, *Experiments in Physical Chemistry*, 6th ed., McGraw-Hill, New York, 1996.
3. Sime, *Physical Chemistry: Methods, Techniques, and Experiments*, Saunders, Philadelphia, 1990.
4. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1969.
5. Young, S. H.; Wierzbicki, A. "Linear Least Squares Regression" and "Non-Linear Least-Squares Regression", *J. Chem. Educ.* **2000**, *77*, 669.
6. Huber and Herzberg, *Molecular Spectra and Molecular Structure IV. Constants of Diatomic Molecules*, Van Nostrand, New York, 1979.
7. Zielinski, T. J. "Exploring the Morse Potential: MorsePotential.mcd", *J. Chem. Educ.* **1999**, *75*, 1191.

Acknowledgment: The authors acknowledge the National Science Foundation for support of the 1998 NSF-UFE Workshop on "Numerical Methods in the Undergraduate Chemistry Curriculum Using the Mathcad Software" and the organizers (Jeff Madura, Andrzej Wierzbicki and Sidney Young, University of South Alabama). RWS acknowledges partial support from the National Science Foundation through grant NSF CHE-8851651. WFP acknowledges partial support from the National Science Foundation through grant NSF CHE-9157713.

