Structure Drawing and Molecular Mechanics Programs.

Jim Beatty

Programs to draw structures and manipulate them are many. This listing includes only those I have literature for. There are more on the market. Upgrades are constantly being made. Prices quoted are for single academic copies. Most suppliers will send you a demonstration disk for a small charge or no charge. Some suppliers will send you programs on an approval basis. Some include minimization programs such as MM2, while for others you must purchase the program separately.

I consider these programs as word processing programs for chemical structures and more. I believe that one or more should be available to undergraduate chemistry students.

I. Tripos Associates, 1699 South Hanley Road, Suite 303, St. Louis, MO 63144-2913. Phone 800-323-2960.

Alchemey III, a powerful molecular modeling program for IBM compatibles or Macintosh computers. Prices start at about $500. MM2 available for IBM machines for $200.

LabVision, an advanced molecular modeling program for computer workstations. $2500. 30 day free evaluation available.

II. Serena Software P.O. Box 3076, Bloomington, IN 47402-3076, Phone 812-333-08232.

PCMODEL Molecular modeling for all worlds. Prices start at $200 and go up to about $4000 for programs for the fastest workstations.

III. Trinity Software P.O. Box 960, Campton, NH 03223, Phone 800-352-1282.

CHEMWINDOW, CHEMINTOSH and CHEMDRAFT III are basically structure drawing progams. CHEMDRAFT in version III has grown into a 3-D molecular modeling and presentation graphics program. Prices start at $99.

IV. Oxford University Press, 200 Madison Avenue, NY, NY 10016, Phone 212-679-7300.

Desktop Molecular Modeller for the IBM PS/2 55 compatible or above. A powerful and interesting moleular structures program which has a number of interesting options including a symmetry operator program for inorganic structures. Prices start at $495.

V. Springer-Velag New York, Inc., Electronic Media Department, 175 Fifth Avenue, New York, NY 10010, Phone 212-460-1653.

Moby is an interesting molecular modeling program for the IBM PC or compatible. Some very interesting options. Prices start at $498.

THE USE OF LINEAR LEAST SQUARES IN UNDERGRADUATE CHEMISTRY LABORATORY COURSES

Donald Rosenthal, Department of Chemistry, Clarkson University, Potsdam NY 13676
ROSEN@CLVM.BITNET

INTRODUCTION

The availability of microcomputers and appropriate software make the use of numerical and statistical methods by undergraduates readily feasible. The important question is how useful and meaningful is the use of such software by undergraduates? There is a risk that either too little or too much information about a numerical or statistical method is provided to the student. If too little information is provided, the student goes through a process which provides some computer output, but obtains no real understanding of what the information obtained means and how it is to be used. If too much information is provided, the student may become confused and frustrated. The level of instruction must be appropriate, considering the abilities and needs of the students.

Linear least squares is one of the most useful statistical methods for beginning students. In the pre-computer days it was common to include an experiment in the junior or senior level physical chemistry or instrumental analysis laboratory course which involved a linear least squares calculation. This exposed the student to the technique. Because of the amount of time required to do the calculations either by hand or calculator, it wasn't really feasible to have the student carry out more than one or a few least squares fits.

At Clarkson where every student has a computer (currently an IBM PS/2) in his (or her) own room and where there are computers in the laboratory and elsewhere on campus, I have asked students to use linear least squares software in a second semester freshman laboratory course and a second semester junior combined analytical - physical chemistry laboratory course.

In this article I will describe how linear least squares can be introduced at an elementary level. In my view it is not essential that

the least squares equations are derived by the instructor. What the student needs to understand is how the method is to be used and how the results of linear least squares calculations are to be interpreted.

Linear Equations

There are TWO different linear equations which the software must be able to fit:

$$Y = AX + B \quad (1), \text{ and}$$
$$Y = AX \quad (2)$$

Some software will only fit data to equation (1). Equation (1) is really not appropriate in some instances. For example, the dependence of absorbance upon the concentration of a single absorbing constituent is given by the Beer-Lambert Law (A = abc) which has the form of equation (2). In chromatography it is usually expected that peak height or peak area (Y) is related to concentration or amount (X) by equation (2). These are only a few examples of where equation (2) rather than equation (1) is appropriate. Very often equation (1) should be used to fit experimental data. If a substance R reacts to produce a product P and the reaction is first order in R, the integrated form of the rate expression is:

$$\ln[Rt] = \ln[R0] - kt \quad (3)$$

where [Rt] is the concentration of R at time t, and [R0] is the concentration initially (at t = 0). This is of the same form as equation (1) where Y = ln[Rt] and X = t.

Many students will use equation (1) indiscriminately because it gives a "better" fit to the data. If the purpose of the experiment is to test the validity of a theoretical or an accepted relationship which is like equation (2) (e.g. Beer's Law or chromatography equations), then equation (2) and not equation (1) should be fitted.

THE THEORY AND ASSUMPTIONS OF LINEAR LEAST SQUARES

In the least squares fitting of experimental data to equation (1) or (2), values of A and B (for equation (1)) are found such that the sum of the squares of the deviations (SSD) is a minimum.

SSD = Sum (Y(experimental) - (Y(calculated))**2 (4)

Y(calculated) is the value of Y calculated fron equation (1) (or (2)). Indeed, least squares means least sum of squares of deviations (SSD).

While students may not realize it, they are already familiar with a least squares calculation. The average of a series of measurements is a least squares value.

Implicit in the assumption that the least squares calculation gives a "best fit" to the data are the assumptions that there is little or no error in X and that the error in Y is independent of the value of Y and is normally distributed. (Normally distributed errors give rise to the typical bell-shaped curve. Small errors in Y are more likely than large errors and positive and negative errors are equally likely.) In discussing errors in Y we are considering random errors. Where equation (1) is being fitted, residual errors in Y will affect the Y intercept, B, but not the least squares value of the slope, A.

The assumption that the magnitude of the error in Y is independent of the value of Y is implicit in the method of unweighted least squares. Where this assumption is not valid, it is possible to use weighted least squares methods. While such calculations are relatively simple to perform, the topic and related techniques are too confusing to introduce at a beginning level.

INFORMATION WHICH SHOULD BE OBTAINED FROM THE LEAST SQUARES CALCULATIONS AND HOW THIS INFORMATION CAN BE USED

In addition to obtaining the least squares slope, A, and intercept, B (for equation (1)), there are a number of other statistics which a satisfactory least squares program should provide. Typical output obtained with the program used at Clarkson is shown below:

This is simulated kinetic data least squares fittedusing equation (3) where EXPT X is the time in seconds and EXPT Y is ln[Rt] obtained from experimental measurements. Ideally, more experimental points would be obtained. Only a few points are used in this calculation to conserve space and make it easier to check some of the calculations.

(If the data had been fitted to equation (2), which is obviously inappropriate, the SLOPE would be -2.80E-03, S.D. SLOPE is 5.41E-04 and S.D. REG is 2.963.)

One of the most important statistics obtained is the standard deviation from regression (S.D. REG). The standard deviation from regression is the square root of the sum of the squares of the deviations

EXPT X	EXPT Y	CALCD Y	DIFFERENCE
1000	-6.990	-6.992	2.0E-03
2000	-7.710	-7.699	-1.1E-02
3000	-8.390	-8.406	1.6E-02
4000	-9.120	-9.113	-7.0E-03

LEAST SQUARES SUMMARY

SLOPE = -7.07E-04 S.D. SLOPE = 6.56E-06

INTERCEPT = -6.285 S.D. INTERCEPT = 0.0180

S.D. REG = 1.47E-02