Lesson 23

- Some final comments on structure solution
- Non-linear least squares
- SHELXL

When Solution Fails

- Check space group

- Check formula

- Copy original data file into xl.hkl and rerun XPREP

- Make sure that you aren't eliminating a valid solution due to bias

- Try a super solver such as SHELXD (XM)

- Re-integrate data perhaps using different software.

Why structure solution fails

- Bad Data usually from a bad crystal. Sometimes re-collecting on a different crystal will help. Usually not.

- Twinned crystal—two types

  - Multiple crystal with some overlapping spots.

  - Real twin where crystal Laue group is not as observed for example an orthorhombic group where a and b are about the same length and disorder to appear as tetragonal.

- In some cases the crystal will never solve!

Linear Least Squares

- Linear refers to a functions where y=ax+bz+... i.e. All terms are additive.

- Lets begin with the simple straight line fit.

- y=mx+b where m is the slope and b is the intercept.

- The two parameters are the slope and the intercept, y is the dependent variable and x is the independent variable.

- Called least squares because the sum of $(d_c - d_o)^2$ is minimized where d is the data.

Fitting a Line

- There is an equation that when solved gives a value for the slope and intercept and their standard uncertainties (su).

- There is also the correlation coefficient which is a number between -1 and 1. This is a measure of the accuracy of treating the data as a straight line as anything can be fit by this method. The closer the absolute value of this is to one the more the data has a linear relationship.

Data to Parameter Ratio

- To solve by linear least squares there must be more observed data than parameters.

- To get good values there should be considerably more data than parameters.

- An irony—as the number of data goes up the precession goes down (su) but the accuracy goes up.

- NOTE—all measurements count even if they have a value of zero.

Non-linear Least Squares

- Biggest difference is the method is iterative—each determination provides shifts which when applied to the parameters yield better values.

- Instead of the correlation coefficient it provides the goodness of fit (goof)

    - This can range from a large number to zero

    - The ideal value is one

- Run until the refinement converges; that is the maximum shift is much less than the error for that parameter

Shift/error

The shift is the absolute value of difference between the original parameter before the least squares cycle and the new calculated value.

For a completed structure the maximum value should be less than 0.1
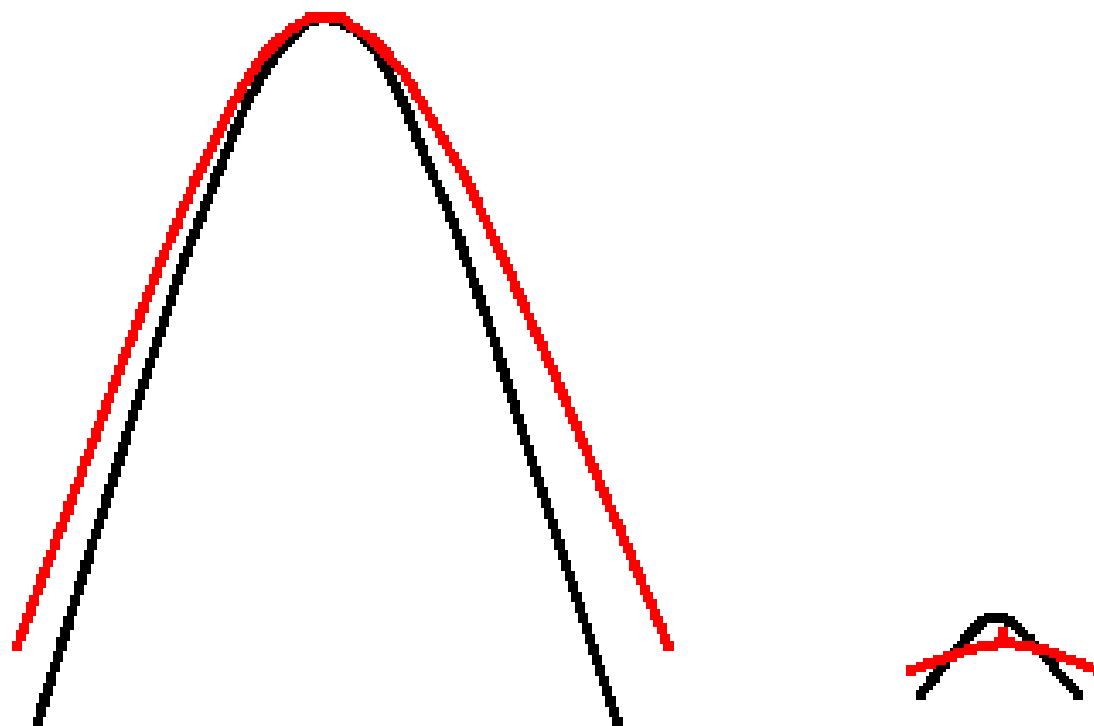
Error   shift
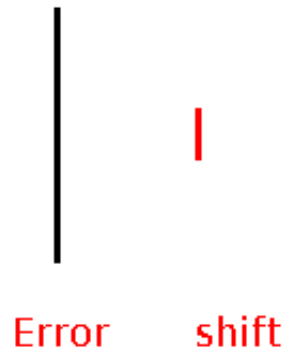
Weighting

- In general linear least squares covers a relatively small range of values.

- Non-linear least squares needs to deal with a much wider range of values.

- Remember the goal is to minimize the sum of $(d_c-d_o)^2$. Thus errors in very large values of data will increase the sum more than errors in weak data.

  - A 1% error in 10,000 is 100 while in 10 is 1.

An rough idea

Weighting

- Need to find a way to normalize the values so that magnitude of the value effects the sum the same way.

- This is done by weighting. The function minimized is $\Sigma w(d_c - d_o)^2$ where w is a weighting function.

- The weight must also maintain the accuracy of the statistics.

- If each measurement has an error then $w = 1/\sigma^2$ is a first approximation

Weighting in SHELX

- $w = 1/(\sigma^2(F_o^2) + (a*P)^2 + b*P)$ where $P = (F_o^2 + 2F_c^2)/3$

The first P term downweights the intense data even more

- The second term downweights the weak data

- Until the structure is well refined the default is a=0.1 and b=0.0

- Program suggests new values for a and b

- Since SHELX refines on F^2 the squares are the data type
- Weights must maintain statistical neutrality
- Typical values
  - 0.0<a<0.1
  - 0.0<b<1.0 (however it appears that for heavy atom data collected with copper the value of b may be much higher. (You will get a checkcif error if it is--ignore)

Observed vs Calculated Data

- The observed data is the reduced measured data

- The calculated data is given by

$$\mathbf{F_{hkl}} = \sum_{j=1} \mathit{f_j'} \, \mathbf{exp[+2\pi \mathit{i}(\mathit{hx_j+ky_j+lz_j})]}$$

Where f_j is the atomic scattering from the jth atom corrected for vibration.

f'=fexp[-

$2\pi^2(U_{11}h^2a^{*2}+U_{22}k^2b^{*2}+U_{33}l^2c^{*2}+2U_{23}klb^*c^*+2U_{13}hla^*c^*+2U_{12}hka^*b^*)]$

In this anisotropic case the U's are variables just like xyz.

Data to Parameter Ratio

- This is much more important in non-linear least squares than the linear version.

- Ideally there should be at least 10 data for each parameter.

- Having fewer makes the errors lower but also less precise. At some low point the errors become meaningless.

Some Math (sort of)

- To perform non-linear least squares the derivative of each parameter with respect to each data point needs to be calculated.

- This is stored in an array called the Jacobean which is number of parameters by the number of data.

- The transpose of this matrix is multiplied into a diagonal matrix containing the weights and then into the matrix to obtain a square matrix which is number of parameters squared. This matrix is symmetric about the diagonal

- This matrix is then inverted to obtain the shifts.

- In order to invert a matrix no column or row can be the same as another. This means all the parameters must truly be independent.

- The diagonal elements of this matrix are related to the shift while the off diagonal elements are related to the correlation of the parameters.

- It is not unusual in crystallography to have 10,000 data and 1000 parameters.

Correlation

If two parameters are actually only one then they will be totally correlated
Correlation is measured by the correlation coefficient which has a value of 1 for a fully correlated parameter and 0 for independent. The most likely cause of correlation is making symmetry related atoms into independent atoms.

Crystallographic Least Squares

- The data will be the measured data in one of three forms

  - I the measured intensity (rarely used)

  - F^2 the corrected measured intensity

  - F the amplitude calculated from $\sqrt{F^2}$

- Today F^2 is preferred. Solves the problem of what to do with negative intensity and also the error is more accurate

Parameters

- Generally each atom to be refined has three fractional coordinates which are parameters.

- Obviously atoms on special positions may have as few as no coordinate parameters.

- Each atom also has atom displacement parameters (adp's)

  – For isotropic atoms there is 1 parameter

  – For anisotropic atoms there are up to 6 parameters

- One overall scale factor

SHELXL

- The crystallographic refinement program to be used is SHELXL

- This is started by xl name where name is an optional name for the files which are of the type name.*xxx*. If name is omitted than the files are assumed to be named xl.*xxx*. At Purdue it is assumed that each structure is in its own directory and the SHELX files are named xl.*xxx*.

SHELXL Files

- Two input files are need for SHELXL

  - .ins contains the instructions and atomic data (in free format)

  - .hkl contains the data

- Two output files are created by SHELXL

  - .lst a list file containing the detailed output

  - .res a file just like .ins which contains the new values for refinement.

- If refinement is ok then the .res file is renamed to .ins to continue refinement (DANGER WILL ROBINSON)

A Note on .res Files

- These are created near the beginning of a refinement cycle.

- If things go wrong they can be empty or be missing the atom data

- Be careful when renaming as you can lose the entire structure.

- The REFINE gui tries its best to prevent renaming the file if it is no good but sometimes even it fails.

- After a good Shelxl run our xl will create a file xl.rescp if things worked OK.

Refinement Instructions

L.S. 5   Do 5 cycles of least squares refinement
ACTA Do a difference Fourier and calculate bonds and angles
BOND $H Include hydrogen atoms in the bond/angle tables
CONF Calculate torsional angles
PLAN -20 Use 20 peaks from the difference Fourier map in the bond/angle calculations
WGHT   0.027500   19.763300 Weighing parameters
FVAR   0.47033 Overall scale factor

The weighting scheme used by SHELX in the final cycles is $w=1/[\sigma^2(F^2) + (aP)^2 + bP]$ where $P = (1/3F_o^2 + 2/3\ F_c^2)$ where a and b are copied from suggested values given in the output. In the

Atom Input

```
CU1  7   0.412237   0.301190   0.167650
11.00000   0.03016   0.04105 =
      0.03941  -0.01083   0.00615   0.00142
CU2  7   0.204678   0.212250   0.097133
11.00000   0.02107   0.04247 =
      0.03061  -0.00129   0.00053   0.00489
F11  6   0.625395   0.481403   0.138240
11.00000   0.04820   0.05657 =
      0.09118  -0.01013   0.02931   0.00620
F12  6   0.678954   0.437258   0.051441
11.00000   0.05631   0.05078 =
      0.05403  -0.00345   0.00362   0.00741
...
```

The first four characters are the atom name. IUCr requires each atom be named by its element and at least one number.
The second number is the element type. The number refers to the order in the SFAC   card. Note the

Fixing Values in SHELX

- At times the program needs to know that a value that could be refined is not to be refined but kept fixed.

- SHELX considers any value on the atom card greater than 5 to be a fixed parameter with a value of entered value-10.0.

- Thus an occupancy value of 11.0000 is actually a fixed value of 1.00000

Occupancy Factor

- This is the number of atoms to be placed at this position.

- It can be refined but for most atoms has a fixed value of 1.000 (11.000) in SHELX.

- For atoms on special positions if will be fractional where the value is
  number of special position in cell / number of general positions

- This occurs because the program applies symmetry which translates the atom into itself.

The End of .ins

HKLF   4   This tells the program there are no more data or instructions and tells the type of data. Type 4 is F^2 data. Type 5 is data for twins.

REM xl in P2(1)/n   Lines that begin with REM are comments
REM R1 = 0.0768 for   7640 Fo > 4sig(Fo) and 0.1089 for all 10745 data
REM   662 parameters refined using   3 restraints

END The last instruction everything below is ignored

WGHT   0.0278   19.7599
Recommended weight for next crycle
REM Highest difference peak 0.589, deepest hole -0.825, 1-sigma level 0.087
Q1   1   0.1227 0.1801 -0.0757 11.00000 0.05   0.59

SHELXL Runtime OUTPUT

- As SHELXL runs if puts out data to give you an idea what is happening.

- You can stop the program by cntrl-C.

- Remember if SHELXL has completed even one cycle there will be a .res file that you probably do not want.

Before any cycle.

```
++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++
++++++++++++++
 + XL    -    CRYSTAL STRUCTURE
REFINEMENT   -   SHELXTL Version
2008/1  +
 + Copyright(c) 2008 Bruker Analytical X-
ray Solutions  All Rights Reserved  +
 + xl                          started at
11:32:54 on 03-Mar-2008  +
 ++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++
++++++++++++++

 Read instructions and data
 Data:  10745 unique,     0 suppressed
R(int) = 0.0000   R(sigma) = 0.0570
 Systematic absence violations:  29   Bad
equivalents:   0
   ** Cell contents from UNIT instruction
and atom list do not agree **
```

Output from Each Refinement Cycle

wR2 = 0.1641 before cycle   1 for 10745 data and   662 /   662 parameters
 GooF = S =   1.086;
Restrained GooF =   1.085 for 3 restraints
 Mean shift/esd =   0.001
Maximum =   0.014 for U13 Cu2 at 11:33:16
 Max. shift = 0.000 A for H22E
Max. dU = 0.000 for H1W1

GooF is the goodness of fit parameter.

wR2 is defined as $\Sigma\ w(Fo^2-Fc^2)^2/\Sigma wFo^2$

End Output

 GooF = S =   1.086;   Restrained GooF
=   1.085 for   3 restraints
 R1 = 0.0768 for   7640 Fo > 4sig(Fo) and
0.1089 for all 10745 data
 wR2 = 0.1641, GooF = S =   1.086,
Restrained GooF =   1.085 for all data
 R1 = 0.1053 for 10745 unique reflections
after merging for Fourier
 Highest peak   0.59 at 0.8773 0.8199
0.0757 [1.29 A from F21]
 Deepest hole   -0.83 at 0.4294 0.2762
0.1429 [0.81 A from CU1]

 +++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++
+++++++++++++
 + xl   finished at 11:34:51   Total
CPU time: 33.4 secs +
 +++++++++++++++++++++++++++++++++
++++++++++++++++++++++++++++++++++
+++++++++++++