

Recent advances in small molecule refinement

Richard Cooper
University of Oxford

Contact details : richard.cooper@chem.ox.ac.uk

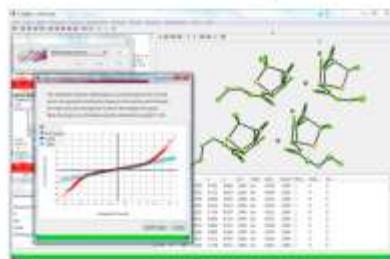
Web: <http://research.chem.ox.ac.uk/richard-cooper.aspx>

More than one single crystal refinement package?

The *de facto* standard software in the field is well established.

```
1070 ADAM.YN.PK  
XRAY 0.18172 0.191 3.193 0.1811 01.09.39  
LATT 1.  
SYMM 9; 2; 2  
1090 9; 2;  
1091 9; 2;  
1092 9; 2;  
1093 1,40.45,2.34.0  
LATT 1.0000000000000000  
1.0000000000000000  
APOL 1  
SCAL 1.1  
ANISI 0.000 0.000  
MWT 0.000 0.000  
AP 1.000 0.000 .000  
A 1.000 0.000 .000  
B 1.000 0.000 .000  
C 1.000 0.000 .000  
D 1.000 0.000 .000  
E 1.000 0.000 .000  
F 1.000 0.000 .000  
G 1.000 0.000 .000  
H 1.000 0.000 .000  
I 1.000 0.000 .000  
J 1.000 0.000 .000  
K 1.000 0.000 .000  
L 1.000 0.000 .000  
M 1.000 0.000 .000  
N 1.000 0.000 .000  
O 1.000 0.000 .000  
P 1.000 0.000 .000  
Q 1.000 0.000 .000  
R 1.000 0.000 .000  
S 1.000 0.000 .000  
T 1.000 0.000 .000  
U 1.000 0.000 .000  
V 1.000 0.000 .000  
W 1.000 0.000 .000  
X 1.000 0.000 .000  
Y 1.000 0.000 .000  
Z 1.000 0.000 .000  
HMLT 0
```

CRYSTALS structure analysis software



UI to encourage correct use of tools.
Guided mode.
Open source for testing new ideas.

'Standard' small molecule refinement has been around for decades.

Modern examples include, but may not be limited to: Olex2, Shelxle, WinGX, CRYSTALS, Oscail.

In our group we develop CRYSTALS – a package for analysis and refinement of crystal structures from X-ray or neutron diffraction data.

Where is the development headed now?

- Overview of the refining the ‘standard model’ of crystal structures (computation)
- Review some recent advances
- Specific examples: restraints, non-atomic scattering density, standard uncertainties

Talk in three parts

- (i) Recap the model commonly used to refine crystal structures. The model is usually fit to measured data using least-squares.
A simple non-linear least squares example is introduced. Addition of restraints to the fit is described and the example revisited.
- (ii) Examples of how the model can be made more complicated to better fit the data.
- (iii) Crystallographic examples of recent developments in CRYSTALS restraints, non-atomic scattering density, and remeasuring data to improve parameter estimates.

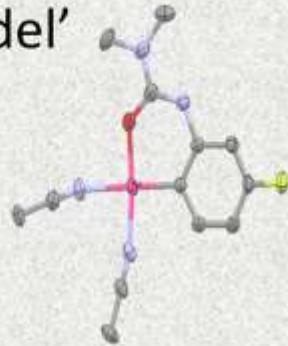
Standard model
Linear least squares
Non-linear least squares
Restraints example

REFINEMENT OVERVIEW

Standard model
Linear least squares
Non-linear least squares
Restraints example

The ‘standard model’

$$\sum f_j e^{-8\pi^2 s^2 U_j} e^{2\pi i (hx_j + ky_j + lz_j)}$$



$$\sum occ_j f_j e^{-8\pi^2 s^2 U_j} e^{2\pi i (hx_j + ky_j + lz_j)}$$

$$\sum_i x_i \left(G(g, I_i) \sum_j occ_j f_j e^{-8\pi^2 s^2 U_j} e^{2\pi i (h_i x_j + k_i y_j + l_i z_j)} \right)^2$$

Examples of simple mathematical models used for computing structure factors.

Top: the standard model computes structure factors from atom positions and displacement parameters in a unit cell.

f_j is the atomic scattering factor (or form factor) for the atom type.

Middle: Model allowing partial occupancy for disorder – this simple change enables modelling of significantly more complicated systems, including mixed site occupancies, partially occupied solvent molecules (for example), and disordered crystals.

Bottom: Model extended to extinction (G = function of extinction parameter and intensity) and twinning (sum over multiple domains with relative occupancy x_i).

Linear least-squares in one slide

$$y = c + ax$$

$$y = c + ax + bx^2$$



$$\begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} c \\ a \end{bmatrix} = \begin{bmatrix} y \end{bmatrix}$$

$$\begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} c \\ a \\ b \end{bmatrix} = \begin{bmatrix} y \end{bmatrix}$$

$$y_1 = c + ax_1$$

$$y_2 = c + ax_2$$

$$y_3 = c + ax_3$$



$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} c \\ a \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$\mathbf{A} \quad \mathbf{p} = \mathbf{Y}$$

The ‘best fit’ is given by parameters **p**

$$\mathbf{p} = [\mathbf{A}^\top \mathbf{A}]^{-1} \mathbf{A}^\top \mathbf{Y}$$

Top:

Linear equations can be written as matrices.

Middle:

Multiple equations with common parameters can be written as matrices too. These equations could correspond to experimental observations, y_1-y_3 , taken under conditions, x_1-x_3 respectively.

The parameters a and c are common to all of the equations and we would like to determine their values. In general when number of equations (here 3) is greater than number of parameters (2) there will not be an exact solution.

Least squares gives us the values of a and c which minimize the sum of the squares of the difference between the equation values and the measured values.

Bottom:

Matrices represented by bold letters.

A is sometimes called the design matrix (it tells us about the conditions used for the experiment, here x). A column for each parameter, one row for each equation. Contains the coefficients of the parameters in the equation (here 1 and x_n)

p is a vector of the parameters (a row for each parameter, one column)

Y is a vector of measurements $y_1 - y_3$. A row for each measurement / equation and one column.

Finally – the matrix operations required to find the least-squares fit for **p**.

When a function is non-linear *in the unknown parameters* we can't write the equations as matrices.

Linear:

$$y = c + ax$$

$$y = c + ax + bx^2$$

$$y = a \sin(x)$$

$$\mathbf{A} \quad \mathbf{p} = \mathbf{Y}$$

$$\begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} c \\ a \end{bmatrix} = \begin{bmatrix} y \end{bmatrix}$$

$$\begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} c \\ a \\ b \end{bmatrix} = \begin{bmatrix} y \end{bmatrix}$$

$$\begin{bmatrix} \sin(x) \end{bmatrix} \begin{bmatrix} a \end{bmatrix} = \begin{bmatrix} y \end{bmatrix}$$

Non-linear:

$$y = a e^b$$

$$y = a \sin(bx)$$



$$\begin{bmatrix} ? & ? \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y \end{bmatrix}$$

$$\begin{bmatrix} ? & ? \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y \end{bmatrix}$$

In reality one would use an optimised linear algebra library (e.g. OpenBLAS, MKL) to solve these equations. In fact, for most problems you would avoid inverting the matrix altogether for more efficient solution. In small molecule crystallography we make use of the inverted matrix to give us estimates of the parameter errors, so it is often computed by default.

Non-linear equations cannot be written in matrix form. It is not possible to separate the equation into a sum of parameters with linear coefficients (see examples above).

Non-linear least squares adjusts **an existing model** resulting in **iterative** convergence, and possible **multiple minima**

$$\begin{aligned} f(x_i, \mathbf{p}) &\approx f(x_i, \mathbf{p}^k) + \sum_j \frac{\partial f(x_i, \mathbf{p}^k)}{\partial p_j} (p_j - p_j^k) \\ &\approx f(x_i, \mathbf{p}^k) + \sum_j A_{ij} \Delta p_j \end{aligned}$$

$$\mathbf{A} \Delta \mathbf{p} = \Delta \mathbf{Y}$$

$$\Delta \mathbf{p} = [\mathbf{A}^\top \mathbf{A}]^{-1} \mathbf{A}^\top \Delta \mathbf{Y}$$

For non-linear problems we can find a best fit by writing the equations in terms of shifts to existing parameters.

Line 1. Approximate differences using the linear terms of a Taylor series expansion of the non-linear function, $f(x_i, \mathbf{p})$.

The current set of parameters is given by \mathbf{p}^k . The best solution is \mathbf{p}^k .

The first term of the Taylor expansion is given by the function evaluated with the current parameters. The second term is the first derivative of the function (evaluated at the current set of parameters) multiplied by the difference between these sets of parameters is $(p_j - p_j^k)$ on the right hand side.

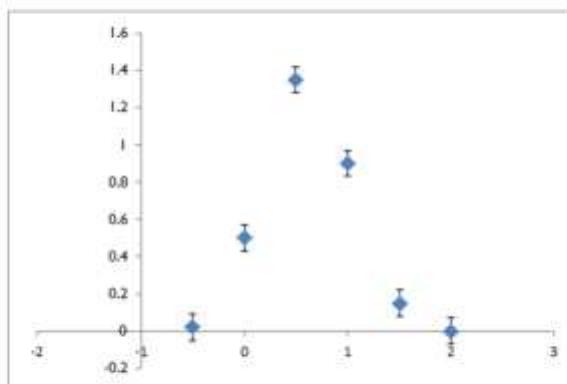
Line 2. Condense – the first derivatives (known) become our \mathbf{A} matrix and the parameters sought are the differences between the current model and best solution.

Line 3. Written as matrix algebra. We get $\Delta \mathbf{Y}$ by subtracting the first term in (2) from both sides (also note the sides are swapped for comparison to previous slides).

Line 4. The least squares solution for shifts to find the best fit (approximate) – therefore need to repeat until convergence.

Functions and matrices are a very dry way to get a feel for non-linear least squares – let's do an example

x	y
-0.5	0.02
0	0.5
0.5	1.35
1	0.9
1.5	0.15
2	0.0



$$y = a \exp(-(x-b)^2/2c^2)$$

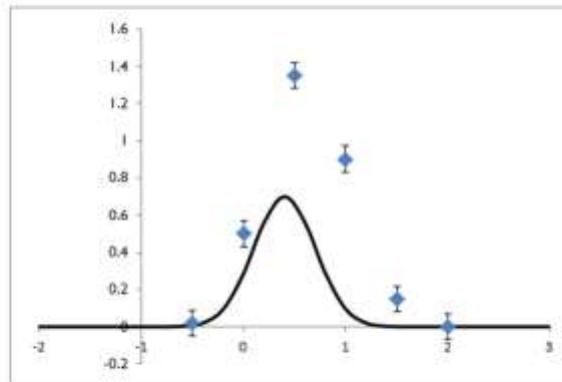
Six points which look like a gaussian curve. Gaussian model equation given below a is proportional to the height of the peak, b is the offset of the mean value from 0, c is the width of the peak.

Aside for pedants:

Although the equation above is non-linear, it can be transformed by taking logs of both sides to give a linear equation.

- (i) It is generally not advised to apply non-linear transformations to your data as the effect will be to change the distribution of errors, and least-squares is predicated on having a normal distribution of errors. However, in this case to improve stability you could use the transformation to find the required parameters and then optimize them a bit more with the non-linear least squares approach. This approach is used, for example, in finding peaks in electron density from Fourier maps.
- (ii) The transformation does not work if you need to fit the sum of two Gaussian shaped peaks (e.g. overlapped peaks), or even if you want to fit a background term along with the peak parameters.

$$a = 0.7; b = 0.4; c = 0.3$$



x	y	$f(x, a, b, c)$	Δy	$\frac{\partial f}{\partial a}$	$\frac{\partial f}{\partial b}$	$\frac{\partial f}{\partial c}$
-0.5	0.02	0.008	0.012	-0.011	0.078	-0.233
0	0.5	0.287	0.212	-0.411	1.279	-1.705
0.5	1.35	0.662	0.687	-0.946	-0.736	-0.245
1	0.9	0.094	0.805	-0.135	-0.631	-1.263
1.5	0.15	0.000	0.149	-0.001	-0.010	-0.037
2	0.0	0.000	0.000	0.000	-0.000	-0.000

$$\frac{\partial f_i}{\partial b} = \sigma \exp\left(-\frac{(x_i - b)^2}{2c^2}\right) \frac{(x - b)}{c^2}$$

$$\frac{\partial f_i}{\partial c} = \sigma \exp\left(-\frac{(x_i - b)^2}{2c^2}\right) \frac{(x - b)^2}{c^3}$$

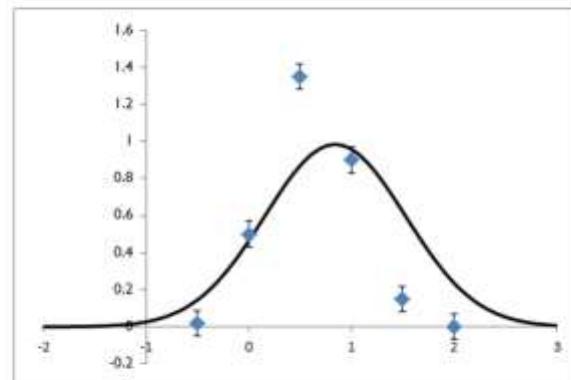
$$\Delta p = (A^T A)^{-1} A^T \Delta y$$

Starting guess Gaussian parameters a, b, c shown at top. First derivative equations for parameters b and c shown below (a is trivial).

For each equation / observation we compute the difference between the observed value and the value predicted by the model: Δy , and the first derivative evaluated for the current model with respect to each parameter a, b, c in turn.

Finally we apply the matrix algebra shown at the bottom right to obtain some shifts to the parameters (see next slide).

$$a = 0.98; b = 0.84; c = 0.69$$

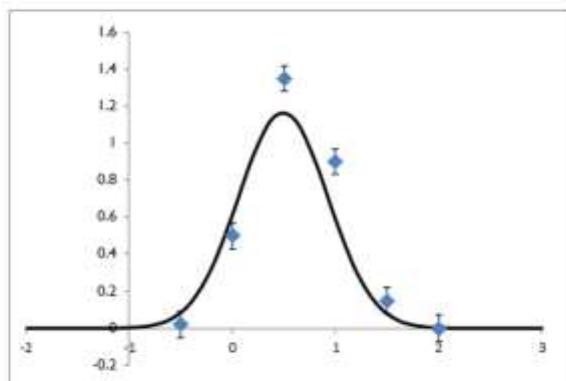


x	y	$f(x, a, b, c)$	Δy	$\frac{\partial f}{\partial a}$	$\frac{\partial f}{\partial b}$	$\frac{\partial f}{\partial c}$
-0.5	0.02	0.145	-0.125	-0.147	0.413	-0.810
0	0.5	0.463	0.037	-0.470	0.829	-1.019
0.5	1.35	0.869	0.481	-0.883	0.632	-0.316
1	0.9	0.959	-0.059	-0.974	-0.321	-0.074
1.5	0.15	0.621	-0.471	-0.631	-0.869	-0.834
2	0.0	0.236	-0.237	-0.240	-0.582	-0.984

1st iteration.

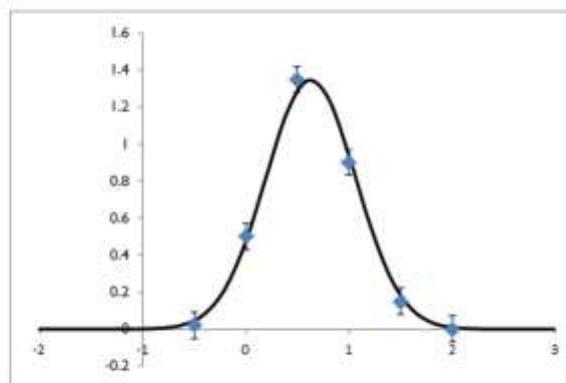
a has increased a bit, b has shifted a bit too far to the right and c has widened out.

$$a = 1.16; b = 0.49; c = 0.43$$



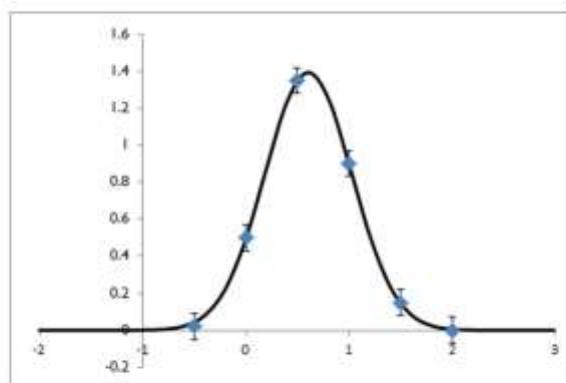
2nd iteration. Looking better!

$$a = 1.34; b = 0.62; c = 0.43$$



3rd iteration, nearly there.

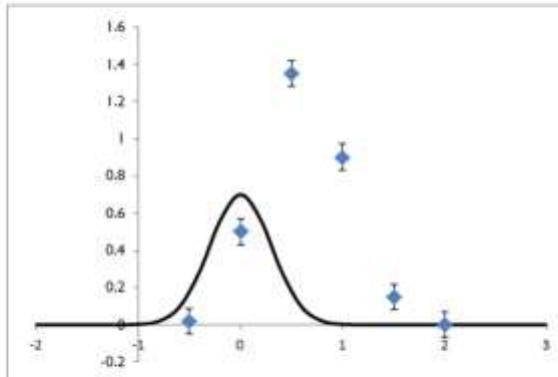
$$a = 1.39; b = 0.61; c = 0.42$$



4th iteration. Good fit and *very small shifts to parameters*. We are approaching convergence.

Looks good, try some different starting models

$$\begin{aligned} a &= 0.7 \\ b &= 0.0 \\ c &= 0.3 \end{aligned}$$



x	y	Δy	$df/d\alpha$	$df/d\beta$	$df/d\gamma$
-0.500	0.020	-0.155	-0.249	0.970	-1.616
0.000	0.500	-0.200	-1.000	0.000	0.000
0.500	1.350	1.175	-0.249	-0.970	-1.616
1.000	0.900	0.897	-0.004	-0.030	-0.100
1.500	0.150	0.150	0.000	0.000	0.000
2.000	0.000	0.000	0.000	0.000	0.000

1.12	0.00	0.81
0.00	1.88	0.00
0.81	0.00	5.23
-1.00	0.00	0.15
0.00	-0.53	0.00
0.15	0.00	-0.21

Normal
Matrix

Inverse:

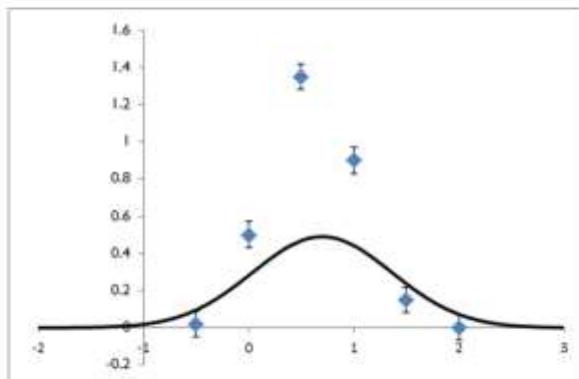
Non-linear least squares needs an approximate solution to start. Because the derivatives rely on the starting model, if the model is too far from the best fit then the derivatives will not help to find the best solution.

For this reason we can't solve the fit of Gaussian parameters by simply making up parameters and refining them to the right solution. The crystallographic equivalent would be throwing lots of atoms into a unit cell and refining to try to find the best model – it will rarely work.

Here, a and c are as before, but starting value of b is 0.0 instead of 0.4.

Looks good, try some different starting models

$$\begin{aligned}a &= 0.49 \\b &= 0.69 \\c &= 0.66\end{aligned}$$



x	y	Δy	df/dx	df/dy	df/dc
-0.500	0.020	-0.076	-0.196	0.261	-0.471
0.000	0.500	0.218	-0.575	0.446	-0.470
0.500	1.350	0.881	-0.956	0.212	-0.063
1.000	0.900	0.458	-0.903	-0.302	-0.137
1.500	0.150	-0.087	-0.484	-0.430	-0.519
2.000	0.000	-0.072	-0.147	-0.212	-0.416

2.35	0.00	0.86
0.00	0.63	0.01
0.86	0.01	0.91
-0.65	0.00	0.61
0.00	-1.58	0.02
0.61	0.02	-1.68

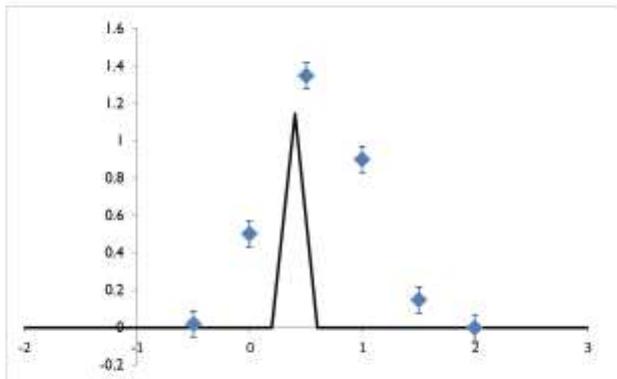
Normal
Matrix

Inverse:

Looks OK but a has shifted in the wrong direction.

Looks good, try some different starting models

$$\begin{aligned} a &= 1.27 \\ b &= 0.42 \\ c &= 0.04 \end{aligned}$$



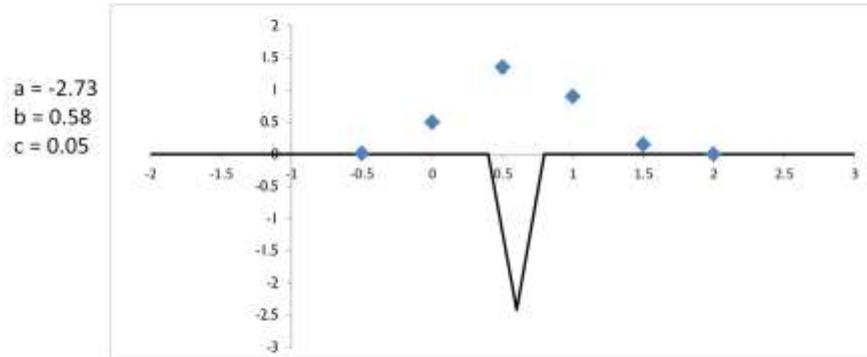
x	y	Δy	df/da	df/db	df/dc
-0.500	0.020	0.020	-2.4E-87	1.34E-84	-2.7E-83
0.000	0.500	0.5	-7.7E-19	1.95E-16	-1.8E-15
0.500	1.350	1.052597	-0.23342	-10.0973	-18.7593
1.000	0.900	0.9	-6.7E-35	-2.3E-32	-2.9E-31
1.500	0.150	0.15	-2E-119	-1E-116	-3E-115
2.000	0.000	-6E-255	5E-255	-5E-252	-2E-250

0.05	2.57	4.38	Normal Matrix
2.57	120.94	206.30	
4.38	206.30	351.91	Inverse

Now a has shifted back the other way and c has become far too narrow.

The shape is now too sharp to be represented by the sampling on my graph.

Looks good, try some different starting models



0.06	5.88	-9.89
5.88	584.35	-982.48
-9.89	-982.48	1651.86

Normal Matrix

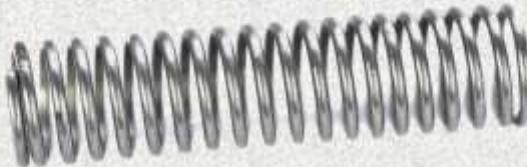
Determinant is zero

Uh-oh. a has now gone negative. c is still far too narrow. The normal matrix ($\mathbf{A}^T \mathbf{A}$) is shown. Its determinant is now zero so it can't be inverted. No further iteration is possible (even if it were desirable)...

Crystallographic restraints introduce external information as new 'data'

$$F_h = \sum_1^n f_n \exp(2\pi i(hx_n + ky_n + lz_n))$$

$$d^2 = (x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2$$



Restraints are extra equations that include some of the parameters of the model. Think of them as extra data.

The equations above for crystallographic structure factor and interatomic distance (in a cartesian coordinate system) show their equivalence. In each case the left hand side value is known (or measured), and the right hand side is a function (non-linear in this case) of some atom parameters (x_i , y_i and z_i) in a model.

For the pedants:

The coordinate systems are different for each equation, the purpose of the equations is to demonstrate as simply as possible the equivalence of these types of equation.

Here is a simple restraint that can improve convergence of non-linear least squares:

$$\begin{bmatrix} \frac{\partial f}{\partial p_1} & \frac{\partial f}{\partial p_2} & \frac{\partial f}{\partial p_3} \\ \vdots & & \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta a \\ \Delta b \\ \Delta c \end{bmatrix} = \begin{bmatrix} y^1 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

This restraint can be read as $\Delta a = 0$, $\Delta b = 0$, $\Delta c = 0$ (i.e. don't change!)

When forming the normal matrix $[A^T A]$, this effectively adds a 1 to every element on the diagonal.

Remember that restraints don't enforce conditions, they are just extra data.

In non-linear least squares the derivatives of the restraints are added into the design matrix.

This is probably the simplest restraint possible, but can be tricky to imagine as it doesn't map onto a real world observation. *We are restraining a to be equal to the current value of a.*

In this case the first restraint equation would be $y=a$.

The first partial derivative $dy/da = 1$

The difference (observed y – model y) = 0.

And so on for b and c .

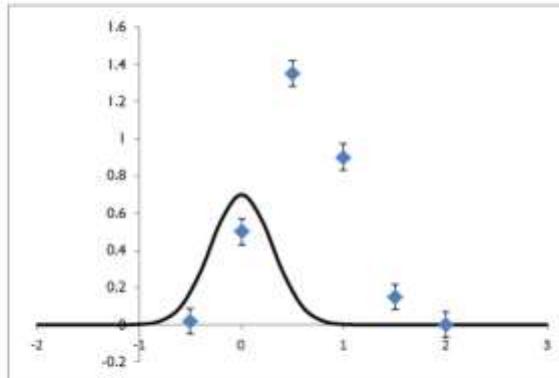
The equations take their place at the bottom of the design matrix (one extra row for each new restraint equation).

The number of parameters is unchanged.

The Δy vector is extended by one row for each restraint equation.

Add this restraint to our bad starting model for the peak

$$\begin{aligned}a &= 0.7 \\b &= 0.0 \\c &= 0.3\end{aligned}$$

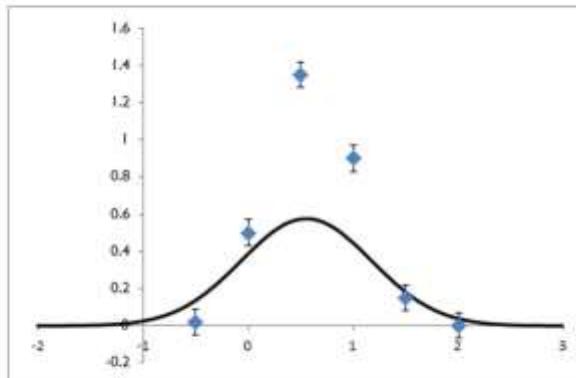


Why should this restraint improve our fit?

It won't affect the end result (since $a=a$ introduces no new information), however it will tend to reduce the shift in parameters between cycles, thus it can stabilize our unstable example from earlier.

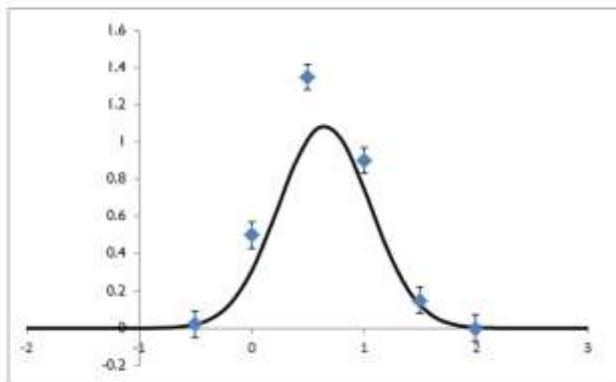
The first six equations provide information which will drive the parameters towards better values, the last three will tend to counteract this and make the shifts smaller – prevents catastrophic overshifts which end in divergence from the best fit.

$a = 0.58$
 $b = 0.55$
 $c = 0.62$



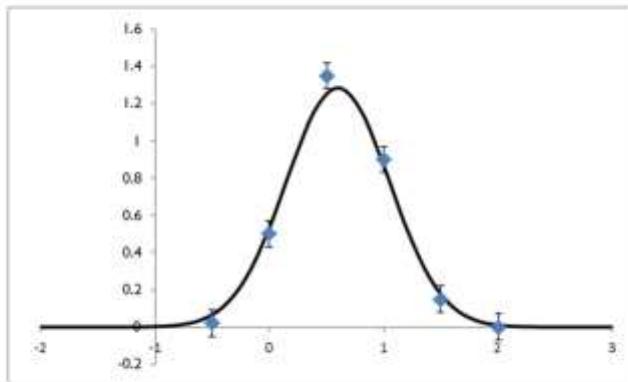
Much better!

$$\begin{aligned}a &= 1.08 \\b &= 0.65 \\c &= 0.41\end{aligned}$$



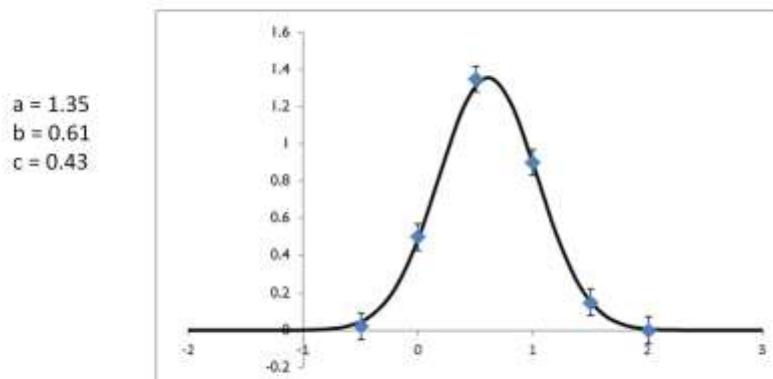
Hooray!

$a = 1.28$
 $b = 0.60$
 $c = 0.45$



Almost there

Damping or 'shift-limiting' restraints slow down convergence, but can prevent calamitous divergence of the parameters.



An alternative approach to avoid divergence is to apply a shift-multiplier to the solution: $p^{k+1} = p^k + c \Delta p$ [0.0 < c < 1.0]

Two effects:

- (i) Convergence is slowed down. These restraints can be removed once the refinement is stable.
- (ii) Augmenting the normal matrix with these restraints means that the parameters uncertainties will be smaller than those estimated by the data alone. It is a philosophical choice and/or what is expected in your area of science to decide whether this is OK. In crystallography it is usually not OK and a final cycle without restraints is recommended to obtain 'true' parameter error estimates.

The purpose of restraints

Class 1: Poor starting model (solution) – need to ‘keep things reasonable’ while structure is improved.

Class 2: Insufficient data for stability in full final refinement (model becomes physically unreasonable). Need extra data in the form of *justifiable* restraints.

The restraints above fall into class 1, where they are needed to stabilize the mathematical solutions due to poor starting models.

Note:

The use of weights in least squares has so far been omitted. They can easily be introduced – just multiply every equation in the list by its own weight which reflects the accuracy of the measurement or restraint information.

$$\begin{aligned} W A \Delta p &= W \Delta y \\ \Delta p &= (A^T W A)^{-1} A^T W \Delta y \end{aligned}$$

W is a square matrix (one row/col for every equation, but is diagonal, so mostly zeroes. Each diagonal term, w_{ii} , is the weight for the i^{th} equation.

More complicated models
Automation

ADVANCES IN STRUCTURE REFINEMENT

More complicated models
Automation

More detail in the model

Two approaches:

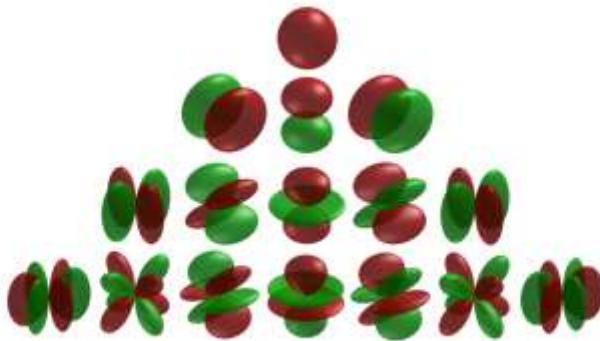
- i. collect better data and refine aspherical atomic density distributions
- ii. use independent data / theory to build a better model of atomic scattering density

These approaches used in charge density analysis and invariom methods respectively are related. The models for the latter either come from high-quality refinement or *ab-initio* DFT calculations.

In fact, using quantum mechanical calculations to obtain atomic scattering factors is the basis of the standard model. The Cromer-Mann parameters used to compute the non-resonant part of the spherical atomic scattering factor f_j are a fit to relativistic Hartree-Fock wave functions for the elements. See references in
http://reference.iucr.org/dictionary/Cromer%20%93Mann_coefficients

Beyond spherical atom scattering: multipoles

$$\rho_i(\mathbf{r}) = P_{i,\text{core}} \rho_{i,\text{core}}(\mathbf{r}) + P_{i,\text{valence}} \kappa^3_i \rho_{i,\text{valence}}(\mathbf{r}) + \sum_{l=0,l_{\max}} [\kappa'^3_i R_l(k' l) \sum_{m=0,l} (P_{lm} + y_{lm} + (r/r))]$$

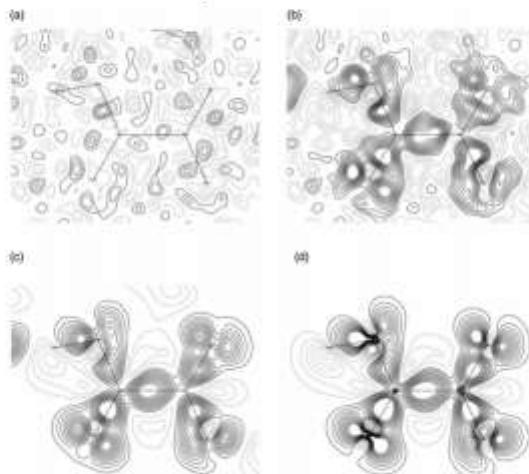


Modern charge density studies: the entanglement of experiment and theory *Macchi, Piero 2013 Crystallography Reviews*

“The most adopted method to reconstruct the electron density is the multipolar model, where $\rho(\mathbf{r})$ is expanded into atomic - or better pseudo-atomic - multipolar functions, based on a radial function centered at the nuclear site and an angular function (spherical harmonics, usually truncated at hexadecapolar level).”

Standard small molecule refinement obtains R-factors of 2%-5%. Each atom is an independent and spherical scattering centre that oscillates harmonically about its nucleus, i.e. according to a restoring force $F = -kx$. To fit more complicated atomic models, we must consider that as well as the non-spherical atomic density, this 2-5% discrepancy includes statistical errors of the diffraction measurements and systematic errors (such as crystal sample absorption).

Charge density maps reveal density not in the independent atom model.



Modern charge density studies: the entanglement of experiment and theory
Macchi, Piero. 2013 Crystallography Reviews

Charge density maps highlight the deformation from the spherical atom model, hence show up bonding (valence) electron density.

The residual density map after multipole refinement (a),
the X-X deformation density map (b),
the model dynamic deformation density map (c),
the static deformation density map (d)

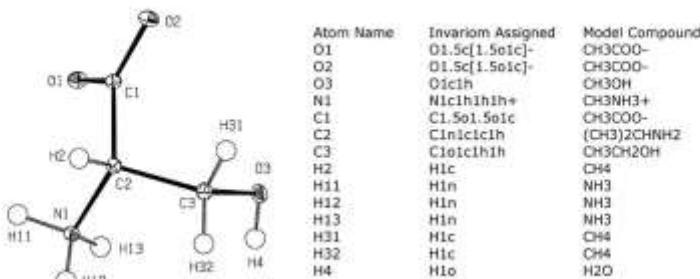
of the hydrogen oxalate ion in L-histidinium hydrogen oxalate [14], from experimental electron density modelling using data at 100 K, and at 0.45 Å diffraction resolution.

All data are included in these Fourier summations. Contours are drawn at $\pm 0.05 \text{ e}\text{\AA}^{-3}$, with positive electron density contours shown as solid lines, negative electron density contours shown as dotted lines.

Invarioms – transferable multipole models for local atom environments

- Local coordinate system for each atom, database of ‘ideal’ multipole parameters.
- Systematic difference in H-position removed.

Acta Cryst. (2005). A61, 314–320



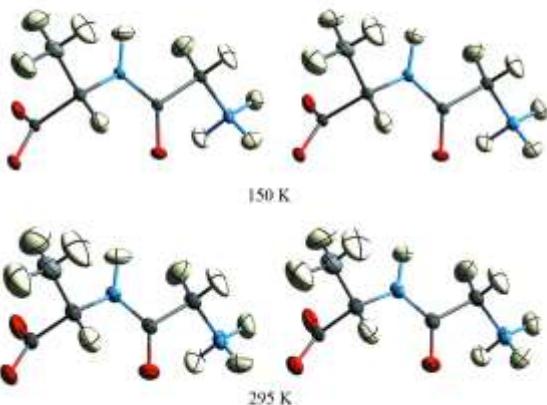
Acta Cryst. (2005). A61, 314–320

Ideally we could compute or find examples of the electron density deformation around an atom and apply it to other structures where we don’t have good enough data to refine these parameters.

This is the invariom model.

Can be applied to refinement of protein data (e.g. with MoPro) by transferring data from high-res amino acids to protein structure.

Iterative hirshfeld atom refinement uses scattering density for each atom from quantum mechanics



Projections of the Gly-L-Ala molecule as obtained from X-ray data after HAR at the BLYP/cc-pVTZ level (left column) and from neutron data (right column). ADPs are shown at the 50% probability level.

Silvia C. Capelli, Hans-Beat Bürgi, Birger Dittrich, Simon Grabowsky and Dylan Jayatilaka

IUCrJ 1(5) 2014 361-379

"Hirshfeld atom refinement (HAR) is a method which determines structural parameters from single-crystal X-ray diffraction data by using an aspherical atom partitioning of tailor-made ab initio quantum mechanical molecular electron densities without any further approximation. Here the original HAR method is extended by implementing an iterative procedure of successive cycles of electron density calculations, Hirshfeld atom scattering factor calculations and structural least-squares refinements, repeated until convergence. The importance of this iterative procedure is illustrated via the example of crystalline ammonia."

Silvia C. Capelli Hans-Beat Bürgi Birger Dittrich Simon
Grabowsky and Dylan Jayatilaka
IUCrJ 1(5) 2014 361-379

Automation of structure solution has been very successful over several decades



Above: Bruker X2S automated small molecule crystallography instrument.

Can work with or without a chemical formula, using multiple solution methods and multiple space groups to try to solve routine structures.

Atoms are modeled anisotropically where the data supports it and hydrogen atoms are included in calculated geometric positions. A final structure report is generated.

Structure solution has been automated to such an extent that we rarely even consider interacting with a solution program these days.

The attempts at structure refinement below show that it is a harder problem – but some inroads are being made, especially for ‘routine’ samples.

Automated solution

ShelxT

Superflip/EDMA

Automated solution and refinement

System-S (Platon)

X2S (Bruker)

Autochem (Olex2/Rigaku)

Displacement parameter restraints
Non-atomic scattering density
Targetted parameter precision

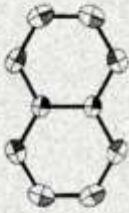
DEVELOPMENTS IN REFINEMENT

Displacement parameter restraints
Non-atomic scattering density
Targetted parameter precision

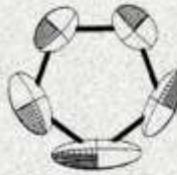
Displacement parameter restraints



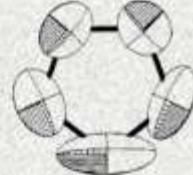
92K



239K



173K



293K

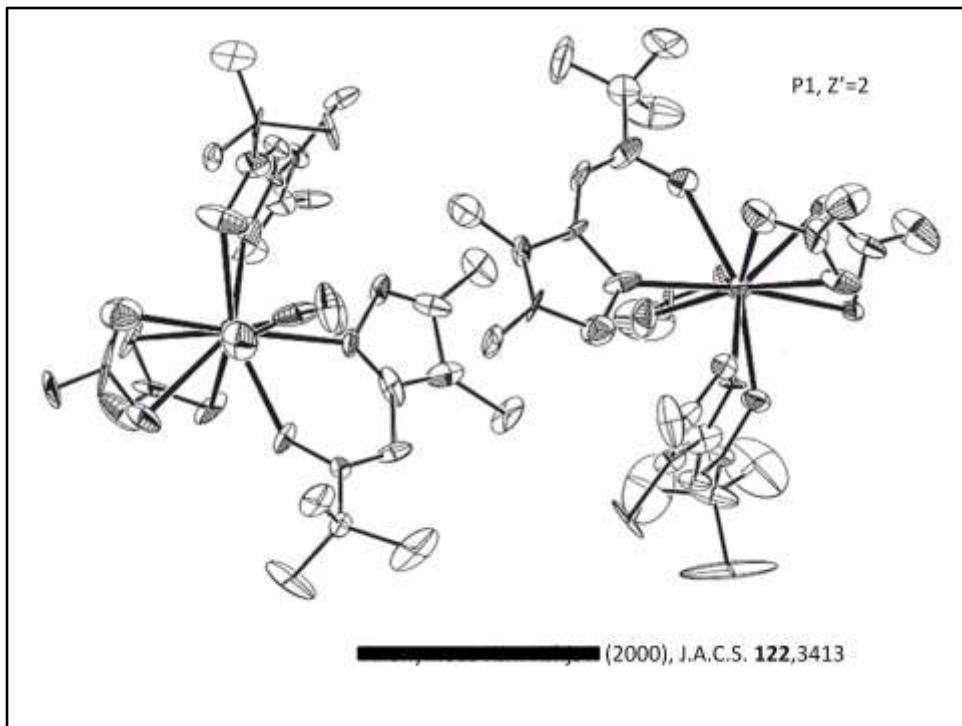
Figure 5. Equiprobability ellipsoids for the carbon atoms of naphthalene²³ at 92 and 239 K. Each ellipsoid axis is 1.54 times the rms displacement along the axis, corresponding to 50% probability.

Figure 6. Equiprobability ellipsoids for the carbon atoms of monoclinic ferrocene²⁴ at 173 and 293 K. Each ellipsoid axis is 1.54 times the rms displacement along the axis.

Dunitz, Schomaker, Trueblood (1988) *J. Phys. Chem.* **92**, 856-867

Displacement parameters have an intuitive relationship with each other. For a given fragment of a molecule in a crystal they are typically about the same size.

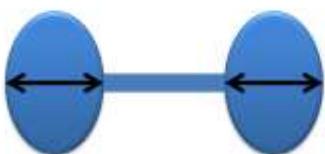
The Hirshfeld ‘rigid-bond’ test states that for equal atoms, the components of the displacement along a bond should be the same.



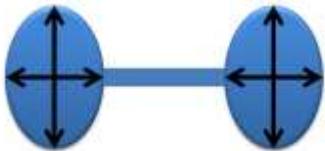
When a model or data is wrong the displacement parameters are often the first place that errors show up.

Above – examples of cigar shaped and oblate (flat) adps. Many adps connected by bonds show entirely different behaviour. The problem is explained away here by a heavy atom (Pr) which dominates the scattering and make the other parameters difficult to determine. However, in fact, the space group should be P-1, not P1. Hence 2x too many parameters are introduced into the refinement, giving a near singular matrix inversion and resulting in distorted bonds, angles and adps.

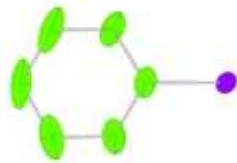
Traditional displacement parameter restraints



Rigid bond (VIBR/DELU)



ADP (SIMU/UIJS)

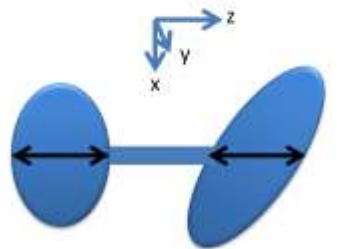


For the cases of poor or incomplete data, additional information about the relationship between displacement parameters may be introduced in the form of restraints.

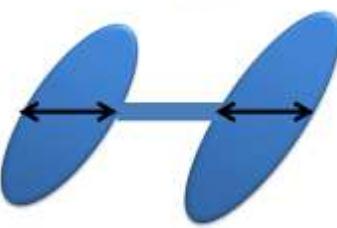
Above – the Hirshfeld rigid bond test criteria (see previous slide) as VIBR (CRYSTALS) or DELU (SHELXL) and a straightforward equivalence between all ADP parameters, SIMU (SHELXL) and UIJs (CRYSTALS).

The former restraint is physically meaningful (atoms connected and moving in a symmetric potential), the latter is less meaningful (see example picture top right), but may be required in the event of very disordered regions or poor data:parameter ratio.

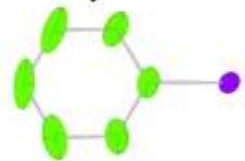
Restrained alignment (RIGU)



$$\Delta U_{zz} = 0$$



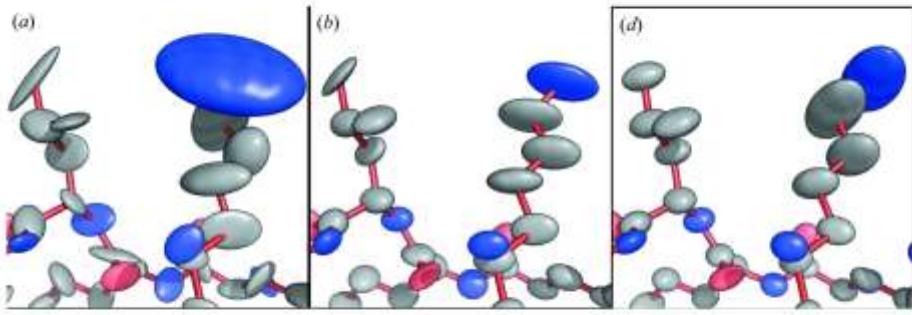
$$\begin{aligned}\Delta U_{zz} &= 0 \\ \Delta U_{yz} &= 0; \Delta U_{xz} = 0\end{aligned}$$



Thorn, Dittrich, Sheldrick (2012) *Acta Cryst. A* **68**, 448-551

RIGU introduced into SHELXL by Thorn, allows restraining direction (but not magnitude) of displacements of atoms related by a bond.

Influence on restricted resolution data

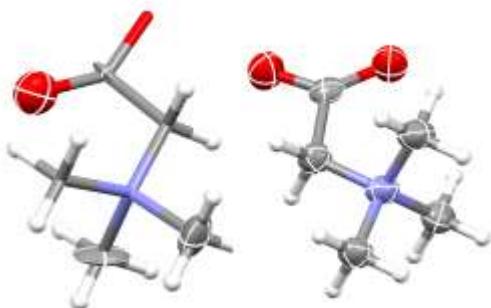


Thorn, Dittrich, Sheldrick (2012) *Acta Cryst. A*68, 448-551

RIGU introduced into SHELXL by Thorn, allows restraining direction (but not magnitude) of displacements of atoms related by a bond.

Effect on protein side chain shown. Paper describes validation of restraint against high quality examples (i.e. molecules really do tend to behave like this – as expected).

Applications include the refinement of high-pressure data (left) where an isotropic model does not give a satisfactory fit but unrestrained refinement of all anisotropic displacement parameters is not possible.



Left: Unrestrained ($R_1=0.079$)
Right: 'UBAR' restrained ($R_1=0.068$ + physical sense)

Refinement of the high pressure structure of betaine monohydrate reported by R. D. L. Johnstone, A. R. Lennie, S. Parsons, E. Pidcock, J. E. Warren, 2009 Acta Cryst. B65, 731-748.

The small molecule equivalent of low-res protein data sets could be high pressure data sets where the opening angle of the pressure cell physically limits the resolution of the experiment, and reciprocal space may not be fully accessible due to the fixed relationship between the sample and the path of the beam.

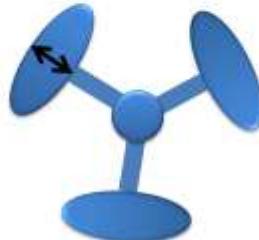
Above – RIGU-type restraint in CRYSTALS allows better model than either isotropic model (published) or unrestrained anisotropic model (left).

Additional adp alignment restraints

$$\Delta U_{zx} = 0; U_{yz} = 0; U_{xz} = 0$$

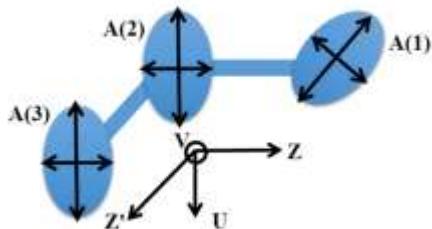
Force principal axes

perpendicular to a given bond or axis direction.



$$U_{xy} = 0$$

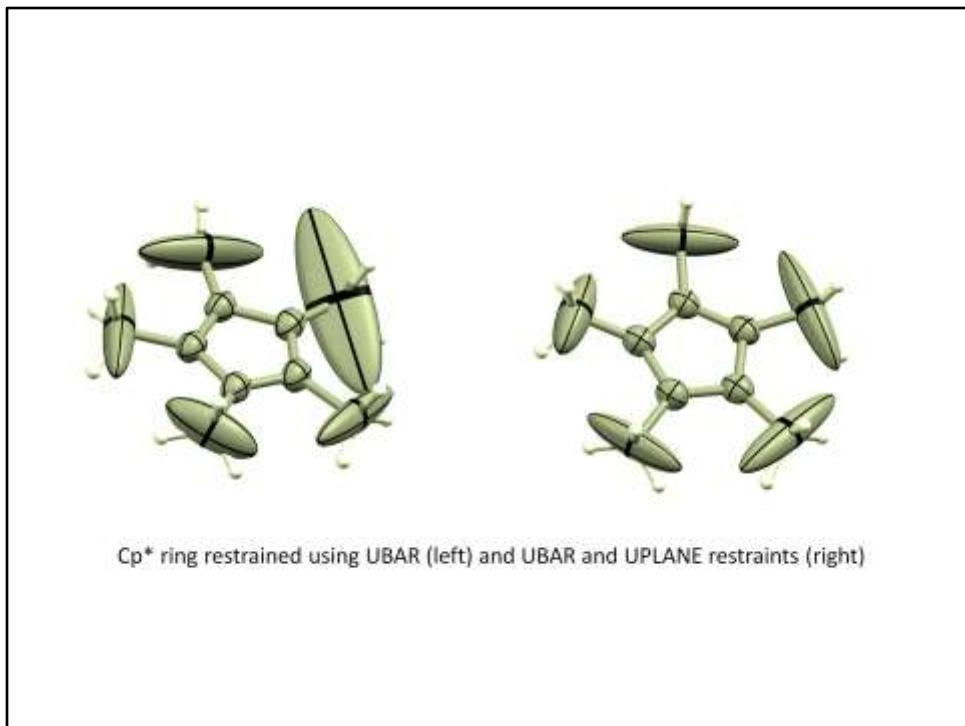
Force principal axes to a lie **in a plane** given three atoms defining the plane.



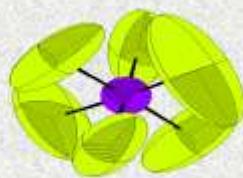
Both ideal for CF_3 , Cp-ring, t-butyl where the principal motion is a rotation.

More adp restraints are possible, where applicable:

Rotating groups usually have adps aligned perpendicular to a bond (top), and the adps tend to be aligned in the plane of rotation (bottom).



Example of refinement against truncated data set.
A Ru-Cp* ring with poor data:parameter ratio (left)
is fixed by perpendicular and in-plane restraints.

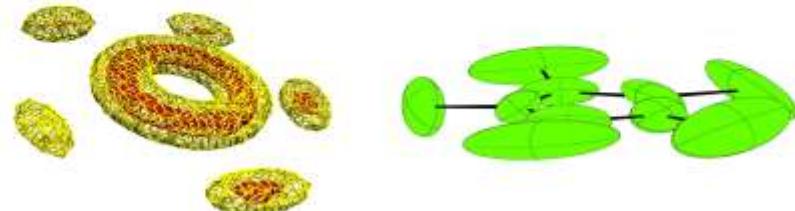


Non-atomic electron density distribution by convolution of a shapes with an atom position.

SPECIAL SHAPES

These are similar mathematically to the invarioms seen earlier - in that we are modifying the shape of the scattering density of an atom. However the purpose here is to fit non-atomic average density in the crystal structure (e.g. due to disorder due to molecular rotations, or randomly filled solvent channels), rather than to attempt a high-resolution fit of the data.

Convolutions of an atom scattering factor with a ring, line or shell are possible.



Refinable parameters

All shapes have an isotropic displacement term

Spherical shell also has a radius

Lines have an orientation and length

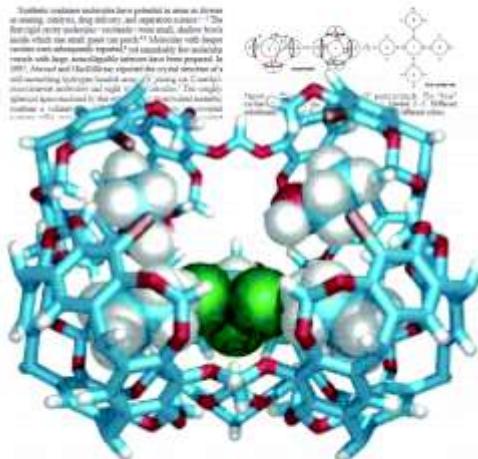
Rings have an orientation and radius

These shapes are convolutions of scattering density in real space. Remember that convolutions in real space are just multiplications in reciprocal space, so if we can describe the shape by a function, its analytical FT can be substituted into the structure factor equation and parameters fit.

Supercage Conformer Molecules

Elizabeth D. Sherburn,¹ Jason L. Irwin,¹ Alison J. Edwards,¹ and Michael S. Sherburn^{2*}
¹ School of Chemistry, Australian National University, Canberra ACT 2601, Australia; and
² School of Chemistry, University of Sydney, Sydney NSW 2006, Australia

Received September 15, 2004; E-mail: m.sherburn@sydney.edu.au

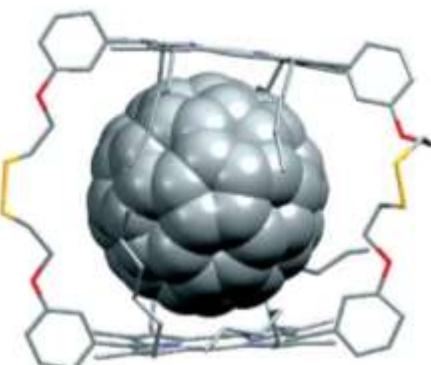


"A ring of electron density¹ is employed to account for the diffuse electron density associated with the partial disorder of two of the three Cl sites of the chloroform occupying the basal cavity and bowl of the structure."²

¹L.Schroder, D. J. Watkin, A. Cousson, R.I. Cooper, and W. Paulus (2004) *J. Appl Cryst.* **37**, 545-550

²E. S. Barrett, J. L. Irwin, A. J. Edwards, and M. S. Sherburn (2004) *J. Am. Chem. Soc.*, **126**, 16747-16749

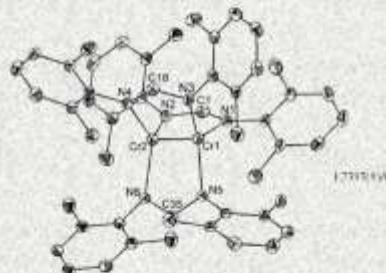
Using CRYSTALS, a structure reported by A. L. Kieran *et al.* (*Chem. Commun.* 2005, 1276–1278) used just five parameters to model a disordered C₆₀ in a porphyrin cage.



Experiment Strategy: find the *best experiment* to answer specific questions.

We might want to know:

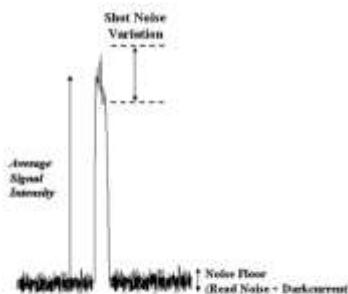
- the distance of this interaction.
- The s.u. of this bond length.
- Is it significantly different from others?



Angew. Chem. Int. Ed. (2008) **47**, 7250–7253
Remarkably Short Metal–Metal Bonds: A
Lantern-Type Quintuply Bonded Dichromium(1)
Complex
Yi-Chou Tsai, Chia-Wei Hsu, Jen-Shiang K. Yu,
Gene-Hsiang Lee, Yu Wang, and Ting-Shen Kuo

The question is not “how can I make the structure more accurate?”, but “how can I make the Cr-Cr bond more accurate?”.

How to improve signal-to-noise by a factor of 2?
Remeasure every reflection $\sim 4x$ longer to double the precision.



$$I = 9; \sigma \propto \sqrt{I} = 3$$

$$I = 18; \sigma \propto \sqrt{I} = 4.24$$

$$I = 36; \sigma \propto \sqrt{I} = 6$$

Can we be more selective?

The standard uncertainty of the parameters is derived from the standard uncertainties of the data *via* the model.

The more accurate the data, the better the model parameters will be (on average).

Therefore a simple approach is to collect data for longer, or with more redundancy to improve the signal to noise.

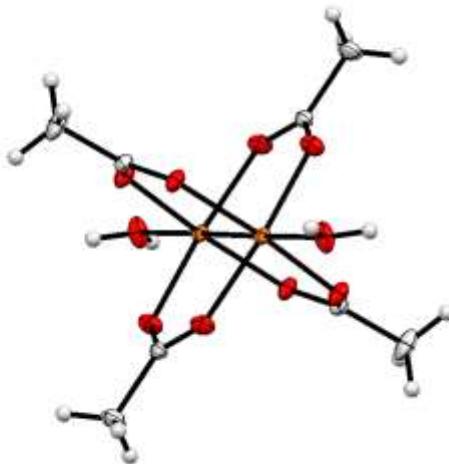
Unfortunately this gets boring quite quickly. Assuming Poisson counting stats we need to collect 4x longer to achieve 2x signal:noise improvement. Can we be more selective?

Standard uncertainties are derived from the least squares analysis

$\Delta \mathbf{p} = (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W} \Delta \mathbf{F}$	Solution of normal equations (shifts)
$V = \frac{w\Delta^2}{n-p} ([A^T W A]^{-1})$	Variance-covariance estimated from least-squares equations
$\sigma_i^2 = V_{ii}$	Variance of one parameter
$\sigma^2(d) = a^T V a$	Variance of a function of parameters including covariance terms, e.g. distance, angle
$a_i = \frac{\partial f(x)}{\partial x_i}$	

1. Solution of parameter shifts including weights (see earlier)
2. The Variance-Covariance matrix, V , is the inverse normal matrix scaled by the reduced chi-squared (called goodness-of-fit in crystallographic literature). When only the relative variance of the measurements is known this *a posteriori* correction puts the errors on the right scale. (For this reason anyone telling you that the goodness of fit should be one is wrong).
3. The variance of a single parameter is the corresponding diagonal element in this matrix. Take the square root for the standard uncertainty.
4. For functions of parameters (e.g. a distance) use a linearised version of the function with respect to each parameter (eqn 5) to extract both the variance and covariance in order to compute the variance of the function itself.

Example: how precise can we make the copper – water distance?



The following example shows a proof of concept that by some careful analysis of an initial model, data can be selected for remeasurement to target specific derived parameters (in this case, a distance).

We picked Cu-O because the Cu-Cu (which is a more obvious target of interest) is already incredibly well determined due to the high scattering power of Cu.

Try ‘add-one-in’ refinement and analysis for every possible reflection, to see whether remeasuring it would improve the precision of a specific parameter.

$$\frac{w\Delta^2}{n-p} [(A^T W A + z_H^T z_H)^{-1}]_{ii} \quad \text{Variance of one parameter}$$

$$a^T \frac{w\Delta^2}{n-p} (A^T W A + z_H^T z_H)^{-1} a \quad \text{Variance of a function}$$

Would require n matrix inversions.

A little thought experiment is shown here. We can surely do this analysis simply by adding one extra reflection at a time (a simulated remeasurement) into the least squares and seeing how it improves the variance of our selected parameter.

The computational time would be quite a waste.

Linear algebra shortcut (Prince 3rd Ed p122):

$$V_{new} = V_0 - \frac{V_0 z_H^T z_H V_0}{1 + z_H V_0 z_H^T} = V_0 - \Delta V$$

No inversion required.

Re-measurements that give ΔV values which most reduce the variance of a parameter can therefore be prioritized.

$$\Delta\sigma^2_i = -\Delta V_{ii}$$

$$\Delta\sigma^2(d) = -a^T [\Delta V] a$$

The shortcut is given by Prince's book Mathematical Techniques for Crystallography and Materials Science (3rd edition p122), and also his related article in International Tables for Crystallography.

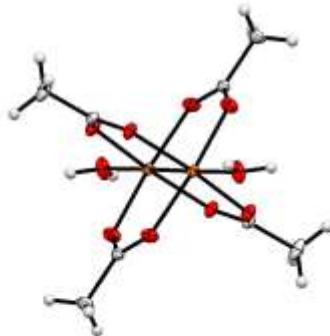
It shortcuts the inversion step to allow a quick calculation of the improvement in a parameter error estimate due to adding one more observation with a short bit of matrix multiplication (the z_H vector is one row of the design matrix with sqrt weight applied), this can be output by some crystallographic software. (e.g. CRYSTALS).

Targeted Precision

For Cu-OH₂ distance.
Measure a 60 minute
data set

Initial Cu-O distance is
 2.189 ± 0.0024

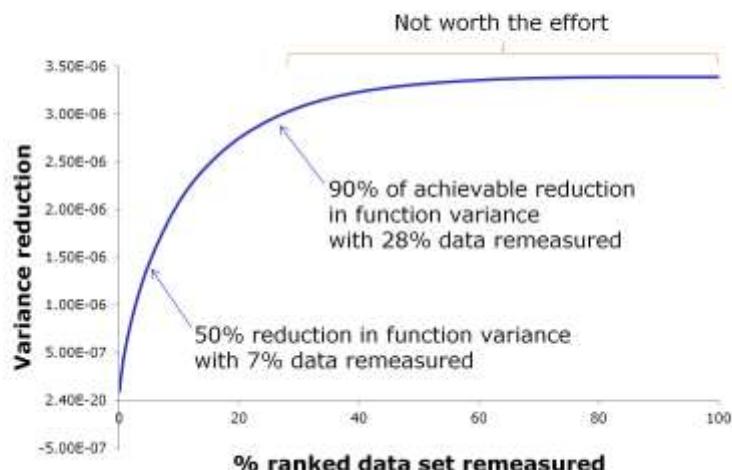
Sort reflections by the
amount they will reduce
 $\sigma(\text{Cu-O})$



Here are the initial values.

Note for pedants – we've quoted the error to more d.p's than the value. It is bad practice (and I could go back to the source and dig out the value), but note that it's the error that we are interested in improving – the value is sort of irrelevant until its precision is improved.

Some data are better than others



Here's the interesting graph. Each point is one reflection and how much it will improve the Cu-O error estimate if we remeasure it and add it into the refinement.

The data are sorted so the best improvements are at the left hand end of the horizontal axis.

Beyond 28% of the data set the rest of the improvements are negligible. We can get to a 50% reduction in Cu-O error (cumulative) with only 7% of the data.

NB. This improvement will probably not be additive, since after adding in the first extra observation, the model will change slightly, so the rest of the analysis is approximate. This is why its useful to try a real experiment...

Targeted Precision

Generate a strategy to re-measure
33 “most useful” reflections in
another 20 minutes [using
CrysAlisPro]

Predicted s.u. improvement:
 0.0009\AA

Initial Cu-OH₂ distance was
 $2.189\text{\AA} \pm 0.0024\text{\AA}$

New Cu-OH₂ distance is
 $2.188\text{\AA} \pm \mathbf{0.0018\text{\AA}}$

Actual improvement: 0.0006\AA



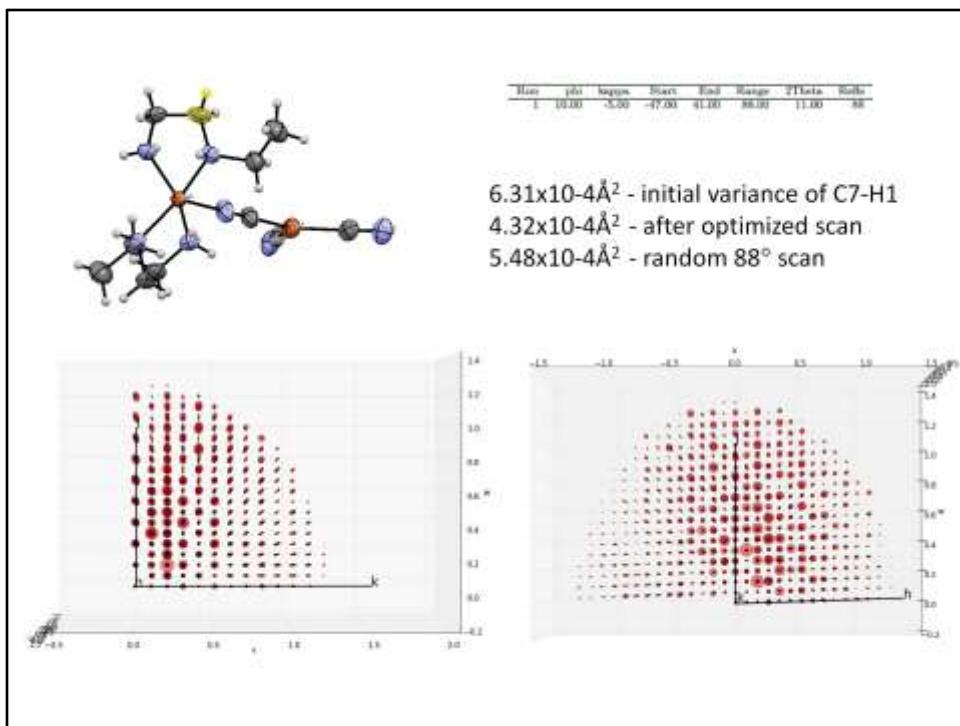
Here we define a strategy on an area detector instrument to remeasure 33 reflections.

Actual improvement in Cu-O error is 0.0024 to 0.0018 Angstrom (about 25% off).

This is about what we would expect if we remeasured the whole dataset for one hour, so there is some advantage to prioritising reflections.

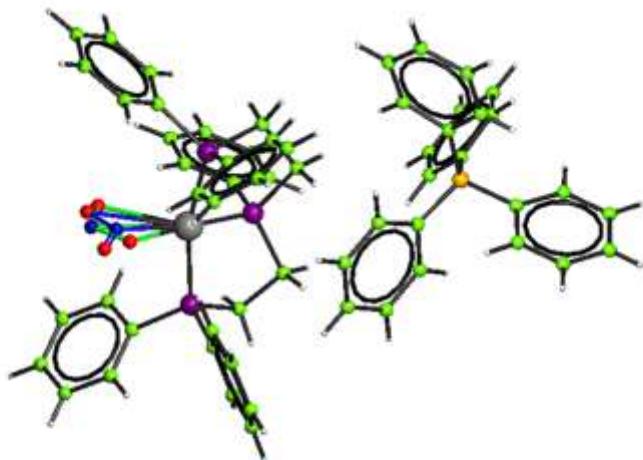


Output of this method is ideal for serial
diffractometers... but trickier for area detectors.
Which are the best scans?
Have to write own strategy software!



Patterns of reflections in reciprocal space are interesting. Different parameters have different reciprocal space distributions (according to the Fourier transform of the function studied).

E.g. points in reciprocal space parallel to a bond distance will have most influence (defining the distance between the atoms). For groups in a plane the most influential points defining geometry are found in a halo in the same plane in reciprocal space.



Linkage isomerism – metastable for seconds -> minutes depending on T.

Real applications: calculating optimal scans for excited state measurements. Differences may be small and collecting the right data quickly enables better time resolution.

Acknowledgements

David Watkin, Bob Carruthers, John Rollett (Oxford)

Pascal Parois, James Arnold (Oxford)

Andrea Thorn (MRC Cambridge)

Iain Oswald (Strathclyde)

Lauren Hatcher & Paul Raithby (Bath)

Mark Warren(Diamond Light Source)

Ton Spek (Utrecht)

Alison Edwards (ANSTO)

Ludger Schroder (CEA-CNRS, Saclay)

EPSRC EP/K013009/1

For those interested, CRYSTALS is available from <http://www.xtl.ox.ac.uk>