Crystallography & Least Squares

Richard Cooper
Original course material by David Watkin
Chemical Crystallography
Chemistry, University of Oxford
The Components of a Structure
The Components of a Structure
An Overview

Three Basic Equations
Integration and summation limits
The atomic model
The form factor
Positional coordinates,
Atomic displacement parameters
Solvent
Disorder
Occupation factors
Extinction
Flack Parameter
Modulated & Incommensurate Structures
The Components of a Structure

During a structure analysis, there are four places where your skill can make a difference

1. Growing and choosing a crystal
2. Setting the data collection parameters
3. Refining the model
4. Interpreting the result

Growing a nice crystal may be the most difficult
Refining the structure will probably consume the most of your own time.
Structure Analysis Throughput

- Optical examination and initial frames
- Data Collection
- Image Processing, Solution and Refinement

Time (days)
The Components of a Structure

The interaction of an X-ray beam with a crystal is a 3-dimensional interference phenomenon, giving rise to diffracted beams of radiation.
The Components of a Structure

The geometric arrangement of these beams (also called reflections) depends upon the geometry of the unit cell, and of the instrument,

The intensity of the beams depends upon the electron distribution inside the unit cell.
The 3 key equations

Reverse Fourier transform

\[ \rho_{\text{xyz}} = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} \left| F_{hkl} \right| e^{\alpha_{hkl}} e^{-2\pi i (hx+ky+lz)} \] [1]

Fourier transform

\[ F_{hkl} = \iiint \rho_{\text{xyz}} e^{2\pi i (hx+ky+lz)} \, dx \, dy \, dz \] [2]

Atomic model approximation

\[ F_{hkl} \approx \sum_{j} f_{j} e^{-\left(8\pi^{2} U_{j} \sin^{2} \theta \right)/\lambda^{2}} \, e^{2\pi i (hx_{j}+ky_{j}+lz_{j})} \] [3]
Bragg’s Law is only concerned with the geometry of the diffraction pattern.

The intensity of a reflection \( hkl \) can be represented as the result of an integration over the continuous periodic electron density in the crystal.

\[
F_{hkl} = \iiint \rho_{xyz} e^{2\pi i (hx+ky+lz)} \, dx\,dy\,dz \quad \text{[2]}
\]

Note the \( i \) in the exponent. This means that \( F \) has both magnitude and phase.

We can measure the magnitude simply by counting photons. \textit{It is very difficult to measure the phase angles.}
The Components of a Structure

Basic Equations - 2

$$F_{hkl} = \iiint \rho_{xyz} e^{2\pi i (hx + ky + lz)} \, dx \, dy \, dz \quad [2]$$

The structure factors, $F$, are Fourier Transforms of the electron density. Note that the integration is over the total contents of the unit cell.

If we know the magnitude $|F|$ and phase angle $\alpha$ of every reflection, we can perform the reverse transformation and compute the electron density at any point in the cell

$$\rho_{xyz} = \frac{1}{V} \sum_h \sum_k \sum_l |F|_{hkl} e^{\alpha_{hkl}} e^{-2\pi i (hx + ky + lz)} \quad [1]$$

Note that the summation is over all possible reflections.
The reconstruction

\[
\rho_{xyz} = \frac{l}{V} \sum_{h} \sum_{k} \sum_{l} |F_{hkl}| e^{\alpha_{hkl}} e^{-2\pi i (hx + ky + lz)} \tag{1}
\]

Given the magnitude of the structure factors, \(|F_{hkl}|\), and the phases, \(\alpha\), we can reconstruct the electron density.

The task of structure solution programs is to determine good estimates of the unobservable phases.

Electron density map of potassium benzyl penicillin, derived by Dorothy Hodgkin.
The Components of a Structure

Basic Equations - 3

The integration over the electron density:

\[ F_{hkl} = \iiint \rho_{xyz} e^{2\pi i (hx + ky + lz)} \, dx \, dy \, dz \] [2]

can be replaced by an approximation (a *model*) which involves a summation over the atomic contents of the unit cell

\[ F_{hkl} \approx \sum_j f_j e^{-\left(8\pi^2 U_j \sin^2 \theta \right)/\lambda^2} e^{2\pi i (hx_j + ky_j + lz_j)} \] [3]

The summation is over the *j* atoms situated at \(x, y, z\) with properties \(f\).
The nature of the model

\[ F_{hkl} \approx \sum_j f_j e^{-\left(\frac{8\pi^2 U_j \sin^2 \theta}{\lambda^2}\right)} e^{2\pi i (hx_j + ky_j + lz_j)} \]  

The centre of each atom is located at \(x,y,z\). The atomic properties consist of two parts:

1. The form factor, \(f_j\). This quantifies the interaction of the electrons with X-rays as a function of Bragg angle, is structure independent, and depends only on the element type.

2. The atomic displacement parameters. ADPs (atomic displacement parameters). These represent the probable mean displacements of atoms from their equilibrium positions. They used to be called ‘temperature factors’ or ‘Debye Factors’.
Form (Scattering) Factors

Scattering factor vs resolution

- (Ca2+)
- (Fe2+)

Electrons

$\sin(\theta)/\lambda$
Form Factors

These are element-specific, and consist of 3 components:

$f$ the *normal part*, is wavelength independent.

$f'$ is the *real part* of the *anomalous dispersion*, is wavelength dependent, and changes rapidly near an absorption edge.

$f''$ is the *imaginary part* of the *anomalous dispersion*, and for non-centrosymmetric structures leads to information about the absolute structure.
We have a model representing the geometry of the molecule itself.

We know the positions of all the atoms, so we can calculate geometrical properties such as bond lengths.
Derived Parameters

Atomic parameters are almost always reported as *fractions of the unit cell axes*, that is, they are *dimensionless quantities*.

This means that they cannot be used directly to compute physical parameters which have dimensions, e.g. distances or angles.

Some older books (e.g. Stout & Jensen) explain how to do the calculation by hand, but it is best to use a computer program.
Crystallographic Coordinates

To calculate a distance by hand, use 3D Pythagoras in a non-orthogonal space

\[ d^2 = u^2 + v^2 + w^2 - 2vw\cos\alpha - 2uw\cos\beta - 2uv\cos\gamma \]

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.2</td>
<td>.3</td>
</tr>
<tr>
<td>2</td>
<td>.8</td>
<td>.3</td>
</tr>
<tr>
<td>3</td>
<td>.2</td>
<td>.7</td>
</tr>
<tr>
<td>4</td>
<td>.8</td>
<td>.7</td>
</tr>
</tbody>
</table>

distance 1→2 = 3→4 = (0.8-0.2)a = u, v = 0

distance 1→3 = 2→4 = (0.7-0.3)b = v, u = 0

distance 1→4 = ?

distance 2→3 = ?
In a general computer program, the fractional coordinates are converted to orthogonal coordinates using an *orthogonalisation matrix*, $L$. The elements of this 3x3 matrix are built up from the cell parameters.

$$x_{\text{orthogonal}} = L \cdot x_{\text{fractional}}$$
The model also shows the atomic displacements

\[ F_{hkl} = f_j e^{i\left(8^{-1}u_j \sin^2 \theta \right) j^2 e^{i(hx_j+ky_j+lz_j)}} \]
Atomic Displacement Parameters

$\textit{adps}$ are now measured in Angstrom$^2$. Small molecule studies are usually reported in terms of $U_{\text{iso}}$, $U_{\text{aniso}}$, $U_{ij}$ or $U_{\text{equiv}}$. Macromolecular studies are reported in terms of $B_{\text{iso}}$, $B_{\text{aniso}}$, $B_{ij}$ or $B_{\text{equiv}}$.

$$B = 8 \pi^2 U$$

Older papers use $b$ or $\beta$, which are dimensionless and make comparisons between different axial directions difficult.
Anisotropic Atomic Displacement Parameters

The tensor is symmetric, so only the 6 unique terms need to be determined. *The size and orientation are determined by all 6 terms together.*

For an atom on a special position, the terms may not be independent because the ellipsoid must conform to the site symmetry.

\[
\begin{array}{ccc}
U_{11} & U_{12} & U_{13} \\
U_{21} & U_{22} & U_{23} \\
U_{31} & U_{32} & U_{33}
\end{array}
\]
Understanding ADPs

The 3D maths is messy, and best left to a computer. For a 2x2 tensor, the interpretation can be visualised:

Angle related to $U_{11}$, $U_{22}$ and $U_{12}$ taken together
The Model

An atom will fit into a 3x3x3Å box. It can be represented by a 3D grid of **700 electron density values** if plotted at 3 points per Angstrom.

\[
F_{hkl} \approx \sum f_j \cdot e^{2\pi i (hx_j + ky_j + lz_j)}
\]

An model normally consists of atoms with:

- 3 positional parameter \((x, y, z)\)

and

- either 1 isotropic adp

- or 6 anisotropic components.

There will be cases where this simple model will not accurately represent what is in the crystal.
Final Result

<table>
<thead>
<tr>
<th>type ID</th>
<th>Position</th>
<th>Displacement Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATOM O 1</td>
<td>0.523</td>
<td>0.392</td>
</tr>
<tr>
<td>ATOM C 2</td>
<td>0.764</td>
<td>0.383</td>
</tr>
</tbody>
</table>

The three positional parameters $x, y, z$ must be taken together to define where the atom is.

The six displacement parameters must be taken together to define its vibration.

The actual values of these parameters are optimised by the **Method of Least Squares**.
Components of a Structure

Solvent molecules

When crystals of blue copper sulphate are heated to just below 100 C, they decompose to a grey-white powder. If water is added to the powder, it turns blue again.
Components of a Structure

Copper Sulphate

The crystals were grown by slowly cooling a hot saturated solution of copper sulphate in water, and then removing the supernatant liquid.

The upper image shows fresh crystals.

The lower image is after allowing the crystals to stand for several days in a warm dry atmosphere.
Components of a Structure

Solvent molecules

The X-ray structure analysis shows that the crystals contain two kinds of water:

That coordinated directly to the copper

That just held in the lattice by hydrogen bonds

Mild heating just dehydrates the lattice, causing it to collapse to a powder. Stronger heating releases the coordinated water.
Components of a Structure

Solvent molecules

Note the channels containing dichloromethane.

If this material were removed from the mother liquor for any length of time, it is certain that the crystals would collapse to powder as a result of loss of solvent.

Chloro-methyl-(N-((2-(diphenylphosphino)phenyl)methylene)-6,6-dimethylbicyclo[3.1.1]2-methanamine-N,P)-palladium(ii) dichloromethane solvate
Disorder

- The observation of each reflection takes a few seconds (which is a long time compared with atomic vibrations) so that the data corresponds to a *time* average.

- In the absence of any other effects, this vibration makes itself evident in the Debye factor, $B$ or $U$.

- $B$ or $U$ were once called the *temperature factors*, but are now called *atomic displacement parameters (adps)* because they can include other displacements from mean positions.
Disorder

• The data collection process also requires a macroscopic crystal, so that the observed data corresponds to a *space average* of many million unit cells.

• It is possible that a given *atom* in a molecule may occupy different spatial positions in different unit cells. This is called *positional disorder*.

• It is possible that a given *position* in different unit cells may contain different atoms. This is called *substitutional disorder*. 
Disorder: Positional vs. Substitutional

- Space average of different conformations gives a 25% unhappy face

- Space average of different elements gives a 25% small, 75% big face = ~87% big face.

- If unsure of the composition it can be difficult to work out the atomic species of a substitutionally disordered site. Charge balancing can help
Components of a Structure
Substitutional Disorder

This is most commonly seen in inorganic extended lattice materials where an atom may be replaced by an atom of a different element with the same valence and size.

The alums, with general formula $M^{+}T^{3+}(SO_4).12H_2O$ are cubic, space group $Pa-3$.

$M = Na, K, Rb, Cs, NH_2$
$T = Al, Cr, Ti, V, Mo$

\begin{align*}
K,Al & \quad a = 12.16 \\
K,Cr & \quad a = 12.20 \\
NH_4,Al & \quad a = 12.24 \\
Rb,Al & \quad a = 12.24 \\
Cs,Ti & \quad a = 12.40 \\
Cs,Mo & \quad a = 12.44 \\
K,Al_{0.95}Cr_{0.05} & \quad a = 12.165
\end{align*}
Alums

Potassium aluminium alum
Potassium chrome alum
Mixed aluminium/chrome alum
Common alum over-growing chrome alum
Components of a Structure

Occupation Factor

Static disorder means that an atom is in one place (I) in one cell, and a different place (II) in another.

If there are $n_{\text{total}}$ molecules in the crystal,

$$n_{\text{total}} = n_I + n_{\text{II}}$$

$$1 = \frac{n_I}{n_{\text{total}}} + \frac{n_{\text{II}}}{n_{\text{total}}}$$

$n_i/n_{\text{total}}$ is called the *fractional occupancy*, and lies in the interval 0 to 1, and the sum of the *occupancies* is usually unity.

Perovskite is a mineral composed of calcium titanate, with the formula CaTiO$_3$. The term is now applied to any material with the general formula ($^{XII}A^{2+}$ $^{VII}B^{4+} X^{2-} 3$) where A and B can be replaced by a variety of ions, e.g. (Mg,Ca,Sr,Ba)(Fe,Nb,Ti)O$_3$
Components of a Structure
Disorder

If the space average disorder involves sites which are very close together, it can give symptoms which are very similar to time average disorder, and appears as unusual adps.

**Substitutional disorder** can appear as atomic displacement parameters which are too large (if the contaminant is heavy) or too small (if the contaminant is light) atomic displacement parameters.

**Positional disorder** over sites which are close together can appear as very anisotropic displacement parameters.
Components of a Structure
Static and Dynamic Disorder

Dynamic disorder of groups of atoms in molecular compounds implies either the whole molecule moves, or there is *torsional* freedom. This enables part of a molecule to move with respect to the rest.

If the disorder has been modelled as an ellipsoid, performing an analysis at two or more temperatures may resolve static and dynamic disorder.

*Dynamic disorder usually reduces significantly as the temperature is dropped*, so that the long axis of the ellipsoid should decrease in length.
The usual description of a macroscopic crystal is that, excepting for faults, it consists of identical unit cells replicated in every direction by simple translation.

Static disorder can modify this description by only requiring that cells are identical on average.
There are cases where adjacent cells are not exactly the same, but differ in a repeating way. In this example, a new unit cell is chosen to incorporate two of the original cells - a Z' = 2 structure.

The perturbation can extend in all three dimensions, and may include more than just the first neighbour.
Modulated Structures

If the original unit cell contained \( n \) atoms, the new cell will contain \( 2n \) atoms, to be located and refined.

However, to a first approximation, most of the atoms in the second motif may be related to those in the first by a simple, exact, translation.

It may be more practical to refine just the atoms of one average motif in the smaller cell, and include additional parameters to indicate that there is a periodic (modulated) change from cell to cell.

This strategy is particularly useful if the slight changes include more than just the first neighbour.
Components of a Structure

Other refineable parameters

So far, our model consists of:

- Positional parameters, $x, y, z$
- Atomic displacement parameters, $U_{iso}$ or $U_{ij}$
- Site occupation parameters, $Occ$

Other parameters include:

- Secondary Extinction
- The Flack parameter
The basic expression for the intensity of the diffracted beam from crystallites which are large enough to act as diffracting agents assumes that every plane in the crystallite is exposed to the same intensity.
Components of a Structure

Primary Extinction

![Diagram of multiple reflections from a family of lattice planes.]

The diffracted beam from the second layer falls at exactly the right angle to be reflected back in the direction of the incident beam from the first layer.

However, there is a phase shift of $\pi/2$ on reflection, so the twice reflected beam is out of phase with the incident beam, leading to destructive interference.

In fact, every incident and emergent beam can be multiply diffracted, leading to a reduction in the intensity of the diffracted beam.

This is most serious for very strong reflections and for very perfect crystals.

It is called **primary extinction**, and leads to $F$ being proportional to the measured intensity, $I$.

**It is not a common effect with ordinary crystals.**
Components of a Structure

Primary Extinction

In most macroscopic samples, the crystal consists of many small volumes in which unit cells are accurately aligned by translation, but these volumes are inclined at small angles to each other.

The mis-matches mean that diffracted beams from one volume have no special phase relationship with those from other volumes, so that interference cannot occur.

The little volumes are called mosaic blocks. A sample consisting of optimally sized blocks is said to be ideally imperfect.
The basic expression for the intensity of the diffracted beam assumes that every plane in the crystal is exposed to the same intensity.

If a particular set of planes is very strongly diffracting, a substantial amount of intensity is diffracted out before the incident beam reaches the back of the crystal. This reduction in the incident intensity creates a reduction in the diffracted intensity for very strong reflections.

This effect is called secondary extinction.

The refinement of an extinction coefficient is a good approximate correction for this effect.
Extinction

Plot of Fo (vertical) vs Fc.

Extinction parameter excluded. The strongest observed intensities are weaker than expected.

Extinction parameter included
Components of a Structure
Secondary Extinction

Typically, the average mis-orientation is of the order of one half a degree. If the mosaic spread becomes larger, this leads to degradation of the diffraction pattern. Soft or mis-handled crystals can show mosaic spreads of up to 4-5 degrees.
The appearance of modulated diffraction often occurs after a phase transition, usually as a result of cooling the sample.

Above the transition temperature the diffraction corresponds to the smaller cell.

Below the transition the cell increase 6-fold in each direction (many more reflections closer together). However, most of the new reciprocal lattice is unobservably weak.

The principal reflections define the average structure, the satellites describe the modulation.

Modulated Structures

The essence of a modulated structure is that a basic motif is modified in a periodic way from adjacent cell to cell.

If, after a finite number of unit cells, the perturbation gets back in step before getting out of step again, the modulation is called \textit{commensurate}.

If it never gets back in step, the modulation is \textit{incommensurate}.

\begin{itemize}
  \item sawtooth
  \item harmonic
  \item crenel
\end{itemize}

\begin{itemize}
  \item \(\bigcirc\) \(\bullet\) \(\bigotimes\) = atoms, groups, or molecules in a regular 3-dimensional structure
\end{itemize}
Modulated Structures

The average structure can usually be determined from the principal reflections.

The direction and periodicity of the modulation can be determined from the diffraction images.

The task for the crystallographer is to propose a parameterised model for the modulated perturbation of the average structure. This can be in terms of atomic positions, element types and/or atomic displacement factors.
Data Modelling

Fig. 409.
Modelling in Crystallography

• What is X-ray crystal structure analysis?

• It is an attempt to find a model which predicts Structure Amplitudes which are in “good agreement” with the observed values.

David Watkin
Chemical Crystallography Laboratory, Oxford
Modelling

Why do we need modelling?

• To enable us to create a realistic representation of some natural object or effect. A description is a model in words.
• To provide a parameterised representation of some natural object or effect.

Train: Courtesy of Howard Flack
Realistic Representation

• Measurements made by hand or instrument can be converted into a representation of the original object.
• The conversion can be quite primitive with values processed point by point.
• More sophisticated processing takes groups of points and processes them together *with the assumption that something is known about their relationship.*
Realistic Representation

• Aerial Surveying

• A plane flies over the land, recording optical or radar images.
Realistic Representation

• Simple processing. The elevations at each point on a grid are simply joined by straight lines.
Neutron Powder Profiles

The points on the profile for each temperature are joined by a smooth line, but each temperature is kept separate.
More Realistic Representation

• The elevations are processed in groups using splines to create a ‘realistic’ image. The values of the parameters in the spline are of no interest – only the final image is wanted.
Neutron Powder Profiles

All the points are in both theta and temperature are smoothed together into a continuous surface.
Crystallographic Application

- To compute or view this Fourier map we only need \( F_0 \) and \( \alpha_c \). We don’t need the atomic coordinates.
Crystallographic Application

• However, we may need the coordinates in order to be able to compute $\alpha_c$. 
Simple Example

Determination of the volume of irregular objects of a given material

• We know that big objects are heavier than small ones.
• Determine the weight of some regular objects of known volume.
• Plot a graph of volume against weight.
Analogue Representation

This **calibration curve** enables the weight and volume of objects to be related.
Just weigh the irregular object, and determine its volume from the graph.
Analogue Representation

• Few phenomena show sharp discontinuities – the points might be better joined with a curve.
Some Antique French Curves

• Smooth curves can be drawn by hand using templates (French Curves), or mathematically using *splines*. A large number of parameters are used in computing the curve, but they are never looked at.
Analogue Representations

- Later, a new observation is added to the curve – clearly something is wrong.
• A large number of parameters are needed to fit a smooth curve (spline) through these points – is there any point in trying?

• We need to see the error bars on the observations
Analogue Representations

- When we add error bars, it becomes evident that the fluctuations may be artefacts of the measurements. We have over interpreted the data.
Analogue Representations

- In fact, a straight trend line passes quite close to all the observations. A good digital representation of the data is

\[ y = 3.2x - 3.0 \]

The 9 pairs of observations can be represented by just 2 parameters,
Cautions

- Extrapolation to small volume samples shows that they have negative weights! Is there a zero-point error, a non-linearity with the instruments, serious under-estimates of the errors or is the relationship really not linear?
Cautions

- Extrapolation to small volume samples shows that they have negative weights! Is there a zero-point error, a non-linearity with the instruments, serious under-estimates of the errors or is the relationship really not linear?
Data Richness

Very many research projects require a hypothesis to be tested against physical observations.

The greater the number of observations, the higher the probability of unsuitable hypotheses being rejected.
Anscombe's Quartet

These are four sets of eleven pairs of numbers.

They were constructed in 1973 by the statistician F.J. Anscombe to demonstrate both the importance of graphing data before analyzing it and the effect of outliers on statistical properties.

## Anscombe’s Quartet

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>x</th>
<th>y</th>
<th>x</th>
<th>y</th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4.3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5.4</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>5.7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5.7</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>7.2</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6.1</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>4.8</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>6.4</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>6.8</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>8.8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7.1</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>7.5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>8.3</td>
<td>11</td>
<td>9</td>
<td>11</td>
<td>7.8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>12</td>
<td>9</td>
<td>12</td>
<td>8.2</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>7.6</td>
<td>13</td>
<td>9</td>
<td>13</td>
<td>13</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>14</td>
<td>8</td>
<td>14</td>
<td>8.8</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>
The smooth curve is computed by Excel using splines.

What will the Excel trend-line be?
Anscombe's Quartet

\[ y = 0.5001x + 3.0001 \]

\[ R^2 = 0.6665 \]
Modelling

Anscombe’s Quartet

What will the Excel trend-line be?
Anscombe’s Quartet

\[ y = 0.4997x + 3.0025 \]

\[ R^2 = 0.6663 \]
Anscombe's Quartet

\[ y = 0.4997x + 3.0025 \]

\[ R^2 = 0.6663 \]

Is the peak just an outlier, or a sharp feature on a smooth background?

More observations might resolve the problem.
What will the Excel trend-line be?
Anscombe's Quartet

\[ y = 0.5x + 3.0009 \]

\[ R^2 = 0.6662 \]
Anscombe's Quartet

0 2 4 6 8 10 12 14 16 18 20

What will the Excel trend-line be?
Anscombe's Quartet

$y = 0.4999x + 3.0017$

$R^2 = 0.6667$

Make some more observations.
Anscombe's Quartet

\[y = 0.4999x + 3.0017\]

\[R^2 = 0.6667\]
Anscombe’s Quartet

Summary Statistics.

Three of the four sets of data points clearly represent quite different phenomena, yet they can all be represented equally well by the same “model”:

\[ y = mx + c \]

Random: \[ y = 0.5001x + 3.0001 \quad \text{R2} = 0.6665 \]

Trend: \[ y = 0.5000x + 3.0009 \quad \text{R2} = 0.6662 \]

Outlier: \[ y = 0.4997x + 3.0025 \quad \text{R2} = 0.6663 \]

Bunched: \[ y = 0.4999x + 3.0017 \quad \text{R2} = 0.6667 \]
Anscombe’s Quartet

Anscombe’s Quartet demonstrates the possibility of fitting a false model to a limited set of observed data points.

A well-designed experiment should aim to reduce the possibility of accepting a false solution.

Graphical representations of the data (or some function of the data) can provide insights into the suitability of the model.
How much Data do we Need?

It depends upon what else you know.

1. If you know what the model should look like, you need enough data to determine the parameters in the model.

   If you also know that the data has large random errors, you need more data so that they ‘cancel out’.

2. If you don’t know what the model should look like, you need enough data to resolve one model from another.

   More data will not remove systematic errors unless you can do something special with the extra data.
Lots of Observations

- In the experiment we looked at earlier, we saw that the observations could be represented by a straight line.
- A straight line can be defined by two error-free points.
- If there are errors, more observations may help to reduce their influence, or enable us to estimate their effect upon the equation (redundancy).
- More points near the ends of the line may help define it more precisely (leverage).
Crystallographic Implication

• In crystallography we get a model which represents the observed structure factors.

• There is no guarantee that if we could measure some more structure factors, the model would also represent them.
Completeness

- To reduce the risk of using an incorrect model, we try to collect a full asymmetric unit of the reciprocal lattice.

- However, if we are confident that the model is parametrically correct, we can use very incomplete data sets.
Data Extent

Big spheres are heavier than small ones.

Try plotting mass against radius.

\[ m = k \cdot R \]

This looks OK \textbf{but} ….

The observational space should be well covered to reduce the risk of using an invalid model.

Measure some more spheres.
Modelling

Data Extent

Plot of mass vs radius

High and low order data help confirm/refute the basic model.

The straight line was a poor model. It was under parameterised.

\[ m = kR^3 \]

\[ m = k \cdot R^3 \]

is better.
• Demo – size of a piece of paper
Applications in Crystallography

- Uses of calibration curves in crystallography:
  - Dead time corrections for quantum counters
  - Spatial correction files for ccds with tapers
  - Dark current correction
  - Temperature indications from thermocouples
  - 9 parameter representation of scattering factors
Strength in Numbers

• You can put any kind of curve through one point.
• A straight line will pass through 2 points, but so also will a sine wave or circle.
• The more observations one has, the less the chance of fitting an inappropriate function.
• Don’t forget that the points have errors.
• *The better ones knowledge of the model, the less observations needed to parameterise it.*
Some Numbers
(Why crystallography is so reliable)

- Imagine a triclinic cell 10x10x10 Angstrom
  - Its volume is 1,000 cubic Angstrom
- The volume of an atom is 20 cubic Angstrom
  - Thus the cell contains 50 atoms
- Atoms can be represented by 9 parameters each.
  - Thus there are ~500 parameters to be optimised.

- An electron density map of this cell computed at 0.3Å would require 30,000 density points.
Fourier Map

This Fourier section, contoured at $\frac{1}{4}$ A°, requires 1,400 data points to be plotted.

The anisotropic model requires 72 variables (parameters)
More Numbers

- For a 10 x 10 x 10 Å cell, the observable
- h, k & l max are all greater than 10
- and h, k & l min are all less than –10
  - Thus total number of reflections is 20 x 20 x 10
- Number of observations/number of parameters
  \[ = \frac{4000}{500} = 8 \]
- This is equivalent to putting a straight line through 16 observations.
What have we learned so far?

• A parameterised model may be a very efficient way of summarising data.
• Parameterised models may be concealed in other representations of data.
• Over-parameterisation may simply model errors in the data.
• Under-parameterisation may fail to represent the data.
• Neglect of data from some part of the observational space may lead to problems being overlooked.
Refinement

- Why do we need refinement?
- To enable us to use a model to smooth out errors in the data.
- To enable us to estimate values for data we cannot measure.
- To provide a parameterised model for an event or object.
Crystallographic Refinement

• Today, refinement in crystallography usually implies least-squares or a related technique.

• However, for difficult problems it is useful to remember modelling and Fourier refinement.

• These have different sensitivities to errors on the data and errors in the model.
Model Building

• In a difficult structure there is no point doing endless cycles of Fourier and Least Squares cycling if simple manual model building will solve the problem
Model Building

This structure could not be refined by the original researcher.

With model building, a good result was achieved.
Modelling a triphenylphosphine
• Demo  Nat
Conclusions

• In a difficult case, just doing more cycles of refinement will probably not solve the problem.
• An un-reasonable model is probably wrong.
• A mathematically reasonable model may also be wrong.
• Use non-crystallographic methods to validate the model.
• Use graphical methods to investigate the data.
At 3 points per Angstrom,
Diffraction

Single Crystal diffraction experiments are *data rich*.

For “ordinary chemicals”, single crystal X-ray diffraction yields about 90 independent observations *for each non-hydrogen atom*.
Diffraction

Thus the “O” atom in a hydroxyl group is represented by the information contained in the equivalent of 90 observations.

Unfortunately, the information is not concentrated into 90 identifiable reflections, but is distributed in varying proportions throughout all the reflections.

*It is not possible to say that such-and-such a reflection is due to any particular atom.*
The diffraction image represents a small rotation of the crystal. The black “windscreen wiper” scans round the whole image sweeping the individual Bragg reflections into a continuous powder profile.
High Resolution Powder Profile

In powder diffraction, the 3D data is compressed into a 1D profile. Inevitably, there is overlap of individual Bragg reflections, and hence a degradation of the information content.

The data quality is often sample (not instrument) limited
High Resolution Powder Profile

The overlap of reflections makes indexing difficult for low symmetry cells.

The overlap of reflections makes it difficult to determine systematic absences (and hence space group).

The overlap of reflections reduces the number of independent observations.

Powder diffraction is important in industry for quality control, and for screening for new polymorphs.

The peak shape (and hence overlap) is often limited by the quality of the microcrystalline sample.
Powders

The white powder is freeze dried and non-crystalline.

The blue powder is microcrystalline

NEVER vac your sample to dryness on a rotary evaporator
## Single Crystal and Powder X-ray Structure Analysis

<table>
<thead>
<tr>
<th>Powder</th>
<th>Single Crystal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requires microcrystalline sample</td>
<td>Requires a good quality single crystal</td>
</tr>
<tr>
<td>May be difficult to determine the unit cell</td>
<td>Unit cell determination is often simple</td>
</tr>
<tr>
<td>May be difficult to determine space group</td>
<td>Space group determination is sometimes simple</td>
</tr>
<tr>
<td>Structure Solution usually more difficult</td>
<td></td>
</tr>
<tr>
<td>Refinement usually more difficult</td>
<td></td>
</tr>
<tr>
<td>Data is from a bulk sample</td>
<td>Data is from a single example</td>
</tr>
<tr>
<td>Twinning does not introduce additional problems</td>
<td>Twinning introduces additional experimental and computational problems</td>
</tr>
</tbody>
</table>
## X-ray Structure Analysis

### Strengths and Weaknesses

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reveals the full connectivity</td>
<td>Requires a good quality single crystal or micro-crystalline powder</td>
</tr>
<tr>
<td>Enables atom types to be determined (more or less)</td>
<td>Structure is determined in the solid state</td>
</tr>
<tr>
<td>Enables molecular parameters (distances etc) to be determined</td>
<td></td>
</tr>
<tr>
<td></td>
<td>May fail totally</td>
</tr>
</tbody>
</table>

*nmr is often able to confirm that the structure in solution is the same as that in the crystal.*
Refinement
Refinement

Methods for solving the phase problem led to atomic parameters which are precise to within $5/100$ths of an Angstrom.

The values of the adps are usually just estimated from the temperature of the experiment, typically $0.03 – 0.05\,\text{Å}^2$.

Refinement techniques seek to obtain parameter values which are both *precise* (are a good representation of the data) and *accurate* (are a good representation of “reality”).
Quality Indicators

Measures of “goodness” of the model

• For a molecular material, the connectivity should make chemical sense

• For a molecular material, the bond lengths and angles should make sense.

• There should be no large differences between the observed and the calculated structure factors.

• There should be no systematic differences between the observed and the calculated structure factors.

• The Fourier transform of the residuals should be featureless
Quality Indicators

The estimation of reflection phases by Direct Methods is a statistical process.

Usually, it either fails catastrophically, or yields connectivity which can be interpreted by an experienced chemist.

*The program has assigned bond types only on the basis of bond lengths and atom assignments.*
R factors

The crystallographic R factors are one-parameter estimates of reliability.

\[
R = \frac{\sum_{\text{data}} (|F_o| - |F_c|)}{\sum_{\text{data}} |F_o|}
\]

\[
wR_2 = \sqrt{\frac{\sum_{\text{data}} w(F_o^2 - F_c^2)^2}{\sum_{\text{data}} w(F_o^2)^2}}
\]

In day-to-day conversation, the conventional R factor is the usual measure of the goodness of a structure, often with a value between 5 and 10%.

Because the modulus operation is non-differentiable, statistical analysis is usually carried out with \(wR_2\). \(wR_2\) is always higher than R

\(w\) is the weight (measure of confidence) assigned to a reflection.
Weights

The observations do not all have the same accuracy or precision.
The weights may be related to the esds (sus) of the reflections, or simply be estimated mathematically.
To a first approximation:

\[ w = 1/\sigma^2(I) \]

We can decrease the significance of reflections we are unsure about by giving them a low weight in the refinement.
Single Value Quality Indicators

Other Measures of Agreement between observed and calculated structure factors.

The summations are over all data.

The weighted residual or Minimisation Function,

\[ M = \sum w(F_o^2 - F_c^2)^2 \]

or \[ M' = \sum w'(F_o - F_c)^2 \]  

The Goodness of Fit,

\[ S = \sqrt{\frac{\sum w(F_o - F_c)^2}{n - m}} \approx 1. \]

Where \( n \) is the number of observed reflections, and \( m \) is the number of parameters in the model \( (n-m \text{ is the number of degrees of freedom}) \).

*If an analysis has been properly carried out and proper weights used, \( S \) should be about unity.*
Quality Indicators

One-value indicators are insensitive to details.

Descriptive Statistics (often plots) sometimes reveal hidden problems.

Plot of Fo vs Fc for ordinary quality data.

Plot of Fo vs Fc for data from a poor crystal. There is a mass of weak intensity data, and some outliers.

Plot of Fo vs Fc for data from a twinned crystal, showing a characteristic distribution.
Binned Residuals

Red bars are unweighted mean residuals. 
Blue bars are weighted mean residuals. 

\[
\langle (F_o - F_c)^2 \rangle \quad \langle w(F_o - F_c)^2 \rangle
\]
Quality Indicators - the Difference Synthesis

\[ \rho_{xyz} = \frac{1}{V} \sum \sum \sum |F|_{hkl} e^{-2\pi i(hx+ky+lz-\alpha_{hkl})} \]

We saw earlier that given the magnitude of the structure factors, \( |F| \), and the phases, \( \alpha \), we can reconstruct the electron density.

Replacing the argument by (Fo-Fc) gives a map which shows features in the data which are not reproduced in the current model.

When all atoms have been placed and refined, a final difference map should be featureless.
Quality Indicators - the Difference Synthesis

The difference map enables us to see defects in the current model, such as missing atoms, inadequate adp modelling and the effect of some systematic errors such as uncorrected absorption.

*When all atoms have been placed and refined, a final difference map should be featureless.*
When is a Structure Wrong?

Ged Parkin has studied this series of compounds formed with different halides.
Wrong Structure

Look at these structures. Comments?

Not-chloride

Chloride
Wrong Structure

Both solutions
Wrong Structure

- Although the R factors for both solutions are similar, there is no path for least squares between them.
Crystallographic Refinement

Refinement path of a single atom in the x-y hyper-surface, all other parameters remaining fixed. The vertical axis is the minimisation function.

If refinement is started from the red arrow, convergence is to a local minimum.
Simulated Annealing

1) Generate a model
2) Set counter, 'n' to zero
3) Compute 'Cost Function', $E_1$
LOOP:

4) Generate quasi-random perturbations in model
5) Compute new cost function, $E_2$
6) Compute acceptability

$$B = \exp \left[ - \frac{(E_2 - E_1)}{T} \right]$$

c.f. 

$$\text{Prob}(E) = \exp \left[ - \frac{E}{kT} \right]$$

7) increment n
8) If $n > n_{\text{max}}$ then decrease $T$ and reset $n = 0$
9) If $E_2 < E_1$ or

$$E_2 > E_1 \text{ and } B > \text{RAN (0:1)}$$

then Accept new model and set $E_2 = E_1$
10) Goto LOOP
Simulated Annealing

Shift $x_1$ always acceptable
Shift $x_2$ conditionally acceptable
Simulated Annealing

Problems

Designing Cost function
Generating perturbations
Relating cooling regime to perturbation scheme

Cost Functions – anything which gets lower as the model improves:

- Packing Energy.
- Internal Energy.
- Agreement with observations \((I_0-I_c)^2\).
- Pair Distribution Function
Simulated Annealing
Generating Perturbations
The perturbations generally reflect our knowledge of the system.

Trial model - two molecules to be positioned in cell
Simulated Annealing
Generating Perturbations

For protein work, perturbations are generated by simplified molecular mechanics algorithms. The atoms are joined by “springs” so that a perturbation in one atom affects its neighbours.

In powder work, “genetic algorithms” swap features of one model (genes) with features in another model. The annealing leads to survival of the fittest.
Simulated Annealing
Generating Perturbations

Random Atomic Perturbations.
The risk is that the model may become nonsensical.

Random Molecular Perturbations (orientation and translation).
If the model has internal flexibility it may be necessary to perturb the torsion angles.
Simulated Annealing

The trial model is perturbed and evaluated against a cost function which tries to maintain connectivity while maximising 2D spatial separation.
Simulated Annealing

Demo
(Graph)
Refinement

Optimising the agreement between Fo and Fc

\[
\sum \text{Occ}_j \cdot f_j \cdot e^{(-8\pi^2 s^2 U_{ij})} \cdot e^{2\pi i (hx_j + ky_j + lz_j)} = Fc \approx Fo
\]

The parameters \(x, y, z, U[ij]\) (and sometimes \(\text{Occ}\)) are adjusted to get a best agreement between Fo and Fc.

The adjustment is made by the method of Least Squares.
Least Squares in Crystallography

• Least squares is a mathematical technique which tries to modify the parameters in a model so that functions computed from these parameters best fit the observations.

Refinement

- Structure development and refinement are often the most time consuming stage of difficult analyses.

- Refinement is often taken to mean “doing some cycles (of least squares).” To get the best results, you need to have a feeling for how LSQ works, and have a feeling for what problem you are trying to solve. If refinement were easy, it would have been automated by now.

- Difficulties with a refinement are sometimes pointers to mistakes made earlier in the process.
## Overview

### Typical Time scales

<table>
<thead>
<tr>
<th>Task</th>
<th>Clock-on-wall time</th>
<th>Your time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal Growing</td>
<td>1 day – 1 year</td>
<td>2 hours-10 hours</td>
</tr>
<tr>
<td>Crystal selection</td>
<td>15 min</td>
<td>15 min</td>
</tr>
<tr>
<td>Data Collection</td>
<td>40 min – 3 days</td>
<td>15 min – 60 min</td>
</tr>
<tr>
<td>Data Processing</td>
<td>15 min – 60 min</td>
<td>15 min – 60 min</td>
</tr>
<tr>
<td>Structure Solution</td>
<td>2 min – 2 days</td>
<td>2 min – 1 day</td>
</tr>
<tr>
<td><strong>Basic Refinement</strong></td>
<td><strong>15 min – 2 hours</strong></td>
<td><strong>15 min – 2 hours</strong></td>
</tr>
<tr>
<td><strong>Fiddling About</strong></td>
<td><strong>0 min – 1 week</strong></td>
<td><strong>0 min – 1 week</strong></td>
</tr>
<tr>
<td>Tables, cif</td>
<td>5 min</td>
<td>5 min</td>
</tr>
<tr>
<td>Pictures</td>
<td>5 min – 1 hour</td>
<td>5 min – 1 hour</td>
</tr>
</tbody>
</table>
Introduction to Least Squares refinement

Based on notes by Peter Main
Numbers

Single crystal crystallography is a “number rich” science unlike, for example, nmr or powder diffraction.

It is important to have a “feeling” for how to use the numbers, even if you let a computer process them for you.
Getting the “best” values
Mean and standard uncertainty
(Average and estimated standard deviation)

For a set of measurements $x_i$, $i = 1..n$, the mean is defined as:

$$ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i $$

The variance is a measure of the spread of values and is given by:

$$ \sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 $$

The square root of the variance is the standard deviation, $\sigma$. 
Weighted mean

If some measurements are considered to be more reliable than others, they may be weighted according to their reliability:

weighted mean: \[ \bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \]

variance: \[ \sigma^2 = \frac{n \sum_{i=1}^{n} w_i (x_i - \bar{x})^2}{n - 1 \sum_{i=1}^{n} w_i} \]

To minimise the variance: \[ w_i \propto 1/(\text{expected error in } x_i)^2 \]
Normal (Gaussian) probability distribution

\[ P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left( -\frac{(x-\bar{x})^2}{2\sigma^2} \right) \]

where \( \bar{x} = \text{mean}; \quad \sigma^2 = \text{variance} \)

This distribution is important because it is produced by a combination of a sufficiently large number of other distributions (Central Limit Theorem).

- The significance of \( \sigma \) is that 68% of observations are expected to be within one standard deviation of the mean.
- 99.7% of observations are expected to be within 3\( \sigma \) of the mean.
- It is common practice to write a mean of 1.56 and a standard deviation of 0.03 as 1.56(3).
Mean (Average) value

When computing a mean value, it is **vital** to ensure that the quantities being added together are all measures of the same thing.

What is the “average” fruit?

Fig. 9.5.2.3. Resolution of the bimodal distribution of C—N bond lengths in C$_{ar}$—N(C$_{sp}^3$)$_2$, fragments: (a) complete distribution, (b) distribution for planar N, mean valence angle at N > 117.5°, (c) distribution for pyramidal N, mean valence angle at N in the range 108–114°.
A scientific experiment

Note: Most texts on the solution of linear systems of equations solve for x in the matrix equation $Ax=b$. Hence “$b=x_1a+x_2$” above for consistency.
A mean calculation
A 1-dimensional model

7 measurements of the length of a football pitch:

85.9, 86.0, 86.1, 86.2, 86.4, 86.5, 87.0 m

the mean is:

\[ x = \frac{1}{7} \sum_{i=1}^{7} x_i = \frac{604.1}{7} = 86.3 \text{ m} \]

length = 86.3(4)m

variance:

\[ \sigma^2 = \frac{7}{6} \sum_{i=1}^{7} (x_i - 86.3)^2 = \frac{0.84}{6} = 0.14 \text{ m}^2 \]
Linear Least Squares: Residuals

Equation of the line (model):

\[ b = x_1a + x_2 \]

We measure \( b_i \) at known values of \( a_i \) to give the set of *observational equations*:

\[ b_i = x_1a_i + x_2 \]

for \( i \) observations.

The *residual* of each equation is:

\[ \varepsilon_i = b(\text{obs})_i - b(\text{calc})_i = b_i - (x_1 a_i + x_2) \]

which is the *vertical distance* of each point from the line.
Linear Least Squares: Observations

Observational equations:

\[ b_i = x_1 a_i + x_2 \]

We measure \( b_i \) at known values of \( a_i \)

Set of observational equations:

\[
\begin{pmatrix}
 a_1 & 1 \\
 a_2 & 1 \\
 \vdots & \vdots \\
 a_n & 1 \\
\end{pmatrix}
\begin{pmatrix}
 x_1 \\
 x_2 \\
 \vdots \\
 x_{n-1} \\
 x_n \\
\end{pmatrix}
= 
\begin{pmatrix}
 b_1 \\
 b_2 \\
 \vdots \\
 b_{n-1} \\
 b_n \\
\end{pmatrix}
\]

Just to check - expand back out to give:

\[
 a_1 x_1 + x_2 = b_1 \\
 a_2 x_1 + x_2 = b_2 \\
 \vdots \\
 a_n x_1 + x_2 = b_n \\
\]

Or, as a matrix equation, \( Ax = b \)
Linear Least Squares: Observations

We want to minimise some function of the residuals $\varepsilon_i$. $\varepsilon_i$ can be $^+$ve and $^-$ve.

$\Sigma \varepsilon_i$ has a minimum of $-\infty$  \(\text{X no good}\)

$\Sigma |\varepsilon_i|$, the least modulus, tends towards zero.

$\Sigma \varepsilon_i^2$, the least squares, tends towards zero

The "best" values of $x_1$ and $x_2$ are those which minimize $\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (b_i - (x_1 a_i - x_2))^2$

Write the observational equations in vector notation:

\[
\begin{pmatrix}
  a_1 & 1 \\
  a_2 & 1 \\
  \vdots & \vdots \\
  a_n & 1
\end{pmatrix}
\begin{pmatrix}
  x_1 \\
  x_2
\end{pmatrix}
=
\begin{pmatrix}
  b_1 \\
  b_2 \\
  \vdots \\
  b_n
\end{pmatrix}
\Rightarrow A x = b
\]

A is known as the \textit{design matrix}. We can choose the values in A by careful design of the experiment. This equation must be solved for the unknowns, $x$:

$x = A^{-1} b$

But … the inverse of $A$, ($A^{-1}$), can only be formed if $A$ is a square matrix.
Linear Least Squares: Normal equations

- There are generally more observations than unknowns, so $A$ is not square.

- The mathematical trick of multiplying both sides of the equation by $A$ transpose leads to a square matrix, which can be inverted.

$$Ax = b$$

\[ A^TAx = A^Tb \]

\[ x = (A^TA)^{-1}A^Tb \]

These are the normal equations of least squares: $x = (A^TA)^{-1}Ab$

\[ x = [A^TA]^{-1}A^Tb = VA^Tb \Rightarrow \]

\[ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \Rightarrow \]

\[ x_1 = v_{11}(a_1b_1 + a_2b_2 + \cdots + a_nb_n) + v_{12}(b_1 + b_2 + \cdots + b_n) \]

\[ x_2 = v_{21}(a_1b_1 + a_2b_2 + \cdots + a_nb_n) + v_{22}(b_1 + b_2 + \cdots + b_n) \]

which give the values of the “best” (i.e. least squares) fit model: $x_1$ and $x_2$. 

29-Nov-12
Least Squares Example

Given

\[
\begin{align*}
2m + c &= 7 \\
7m + c &= 17 \\
5m + c &= 13
\end{align*}
\]

Find \( m \) and \( c \)

\[
\begin{pmatrix}
x_1 & 1 \\
x_2 & 1 \\
\vdots & \vdots \\
x_n & 1
\end{pmatrix}
\begin{pmatrix}
m \\
c
\end{pmatrix} =
\begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix}
\]

\[
A^T A x = A^T b
\]

\[
\begin{pmatrix}
x_1 & x_2 & \cdots & x_n \\
1 & 1 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
x_1 & 1 \\
x_2 & 1 \\
\vdots & \vdots \\
x_n & 1
\end{pmatrix}
\begin{pmatrix}
m \\
c
\end{pmatrix} =
\begin{pmatrix}
x_1 & x_2 & \cdots & x_n \\
1 & 1 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix}
\]
Least Squares

\[
\begin{bmatrix}
2 & 7 & 5 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
m \\
c
\end{bmatrix}
=
\begin{bmatrix}
2 & 7 & 5 \\
1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
7 \\
13
\end{bmatrix}
\]

\[
\begin{bmatrix}
2x2 + 7x7 + 5x5 \\
2x1 + 7x1 + 5x1 \\
2x1 + 7x1 + 5x1 \\
2x1 + 7x1 + 5x1 \\
1x1 + 1x1 + 1x1
\end{bmatrix}
\begin{bmatrix}
m \\
c
\end{bmatrix}
=
\begin{bmatrix}
2x7 + 7x17 + 5X13 \\
1x7 + 1x17 + 1x13
\end{bmatrix}
\]

\[
\begin{bmatrix}
78 & 14 \\
14 & 3
\end{bmatrix}
\begin{bmatrix}
m \\
c
\end{bmatrix}
=
\begin{bmatrix}
198 \\
37
\end{bmatrix}
\equiv
\begin{bmatrix}
234 & 42 \\
196 & 42
\end{bmatrix}
\begin{bmatrix}
m \\
c
\end{bmatrix}
=
\begin{bmatrix}
594 \\
518
\end{bmatrix}
\]

\[
m = ?
\]

\[
c = ?
\]
Least Squares

\[\begin{bmatrix} 2 & 7 & 5 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 2 & 7 & 5 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 7 \\ 17 \end{bmatrix} \]

\[\begin{bmatrix} 2x2 + 7x7 + 5x5 \\ 2x1 + 7x1 + 5x1 \\ 2x1 + 7x1 + 5x1 \\ 1x1 + 1x1 + 1x1 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 2x7 + 7x17 + 5X13 \\ 1x7 + 1x17 + 1x13 \end{bmatrix} \]

\[\begin{bmatrix} 78 & 14 \\ 14 & 3 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 198 \\ 37 \end{bmatrix} = \begin{bmatrix} 234 & 42 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 594 \\ 518 \end{bmatrix} \]

\[38m = 76\]
\[m = 2\]
\[c = 3\]
Use of weights

The observational equations may be given weights $w_i$ according to their reliability, where $w_i \propto 1/(\text{expected value of } \varepsilon_i)^2$

The least squares process will then minimise $\sum_{i=1}^{n} w_i \varepsilon_i^2$

This is done by multiplying the observational equations by a weight matrix, which is normally a diagonal matrix with the weights on the diagonal: $WAx = Wb$

The normal equations of least squares are: $A^TWAx = A^TWh$

Expressing the least squares solution of the equations in matrix notation allows the use of as many equations and unknowns as desired.
The variance-covariance matrix

observational equations: \( Ax = b \)

normal equations: \( A^T W A x = A^T W b \)

parameters \( x \) will minimise \( \sum_{i=1}^{n} w_i \varepsilon_i^2 \)

the variance-covariance matrix is: \( M = \frac{n}{n - p} \frac{\sum w_i \varepsilon_i^2}{\sum w_i} (A^T W A)^{-1} \)

\( n = \) number of observational equations
\( p = \) number of parameters
\( n - p = \) number of degrees of freedom

\( M \) contains variances on the diagonal, covariances off-diagonal

e.g. \( M = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \mu_{12} \\ \sigma_1 \sigma_2 \mu_{12} & \sigma_2^2 \end{pmatrix} \)
# Variance and covariance

If a quantity $x$ is derived from two measured parameters $a$ and $b$ as:

$$x = \alpha a + \beta b$$

the variance of $x$ is:

$$\sigma_x^2 = \alpha^2 \sigma_a^2 + \beta^2 \sigma_b^2 + 2 \alpha \beta \sigma_a \sigma_b \mu_{ab}$$

(1)

where $\sigma_a \sigma_b \mu_{ab} = \text{covariance of } a \text{ and } b$

$$-1 \leq \mu_{ab} \leq +1$$

$\mu_{ab} = \text{correlation coefficient}$

- measures how strongly errors in $a$ and $b$ are related
**Variance and covariance**

If we do not have the covariance, equation (1) cannot be properly calculated. This is a problem for CIFcheck and PLATON

\[
\sigma_x^2 = \alpha^2 \sigma_a^2 + \beta^2 \sigma_b^2 + 2\alpha \beta \sigma_a \sigma_b \mu_{ab} \quad (1)
\]

\[
\sigma_x^2 = \alpha^2 \sigma_a^2 + \beta^2 \sigma_b^2
\]

\[-1 \leq \mu_{ab} \leq +1\]

**Simplifying, we can see that the value of \( \mu \) can either double the variance of \( x \), or reduce it to zero**

\[
\sigma_x^2 = 2\alpha^2 \sigma_a^2 \cdot (1 + \mu_{ab})
\]
The Surveyors Tale

(How restraints and constraints can help a poor experiment)

Mediaeval Fields
The Surveyors Tale
A student surveyor measures the angles of a triangular field. When he gets home, he finds that the sum of the angles is $174\degree$. What can he do to fix the error?
Restraints

We know that the sum of the interior angles must be 180°.
We can apply this rule approximately.

Observational equations:
\[
\begin{align*}
\alpha &= 73° \\
\beta &= 46° \\
\gamma &= 55°
\end{align*}
\]

Restraint equation:
\[
\alpha + \beta + \gamma = 180°
\]

They are added to the observational equations:
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\gamma
\end{pmatrix}
= 
\begin{pmatrix}
73 \\
46 \\
55 \\
180
\end{pmatrix}
= 
\begin{pmatrix}
1\alpha + 0 + 0 \\
0 + 1\beta + 0 \\
0 + 0 + 1\gamma \\
1\alpha + 1\beta + 1\gamma
\end{pmatrix}
= 
\begin{pmatrix}
73 \\
46 \\
55 \\
180
\end{pmatrix}
\]

i.e.
The normal equations are formed by:

\[
\begin{pmatrix}
1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 \\
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\gamma \\
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & 1 & 73 \\
0 & 1 & 0 & 1 & 46 \\
0 & 0 & 1 & 1 & 55 \\
1 & 1 & 1 & 1 & 180 \\
\end{pmatrix}
\]

which gives:

\[
\begin{pmatrix}
2 & 1 & 1 & 253 \\
1 & 2 & 1 & 226 \\
1 & 1 & 2 & 235 \\
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\gamma \\
\end{pmatrix}
= 
\begin{pmatrix}
253 \\
226 \\
235 \\
\end{pmatrix}
\]

and the solution \( \alpha = 74.5^\circ \quad \beta = 47.5^\circ \quad \gamma = 56.5^\circ \) is:

The previous sum of angles \( 174^\circ \) was

The new sum of angles \( 178.5^\circ \)
Use of weights

Assume the expected error in $\alpha$ is $\frac{1}{2}$ the expected errors in $\beta$ and $\gamma$.

$$w_i \propto \frac{1}{(\text{expected error})^2}$$

In addition, we will apply the restraint more strongly.

The weighted observational and restraint equations will therefore be:

$$w=1/(\frac{1}{2})^2 = 4$$

\[
\begin{pmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 4
\end{pmatrix}
\begin{pmatrix}
4 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 4
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\gamma
\end{pmatrix}
= 
\begin{pmatrix}
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 4
\end{pmatrix}
\begin{pmatrix}
4 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 4
\end{pmatrix}
\begin{pmatrix}
73 \\
46 \\
55 \\
180
\end{pmatrix}
\]
Multiplying the matrices gives the normal equations of least squares:

\[
\begin{pmatrix}
8 & 4 & 4 \\
4 & 5 & 4 \\
4 & 4 & 5
\end{pmatrix}
\begin{pmatrix}
\alpha \\
\beta \\
\gamma
\end{pmatrix}
=
\begin{pmatrix}
1012 \\
766 \\
775
\end{pmatrix}
\]

and their solution is \( \alpha = 73.6^\circ, \beta = 48.4^\circ, \gamma = 57.4^\circ \)

(cf 73, 46, 55)

- Notice that \( \alpha \) has moved away from its measured value by a smaller amount than either \( \beta \) or \( \gamma \) because of its assumed smaller error.

- Notice also that the sum of angles is now greater than its previous value of \( 179.4^\circ \) because of the stronger restraint.
Further observations

It appears that any desired result can be obtained by adjusting the weights to some appropriate value.

If you are to trust the numbers you get from the calculation, you must put genuine information into it.

**The weights must therefore reflect the actual reliability of the observations.**

If you put rubbish into your calculation, you will get rubbish out.
Constraints

The sum of the angles of a triangle is exactly 180 degrees so it is sensible to impose this restriction rigorously.

This is done using constraints which are satisfied exactly, whereas restraints are satisfied only approximately.

LaGrange Multipliers \((\text{rarely used in crystallography})\)

To solve the equations \(Ax = b\)

subject to the constraints \(Gx = f\)

introduce a new set of variables \(\lambda\) such that

\[ Ax + G^T \lambda = b \]
\[ Gx \quad = f \]

These equations are solved for \(x\) and \(\lambda\).

The constraints will be satisfied exactly and the observational equations will be satisfied approximately.
Reduce the number of parameters

Crystallographic least squares programs will typically use the constraints to eliminate some of the parameters to be determined.

The constraint \( \alpha + \beta + \gamma = 180 \)

can be rearranged to give \( \gamma = 180 - \alpha - \beta \)

Now use a reduced set of unknowns \( u, v \) such that

\[
\begin{pmatrix}
\alpha \\
\beta \\
\gamma
\end{pmatrix} = \begin{pmatrix}
1 & 0 \\
0 & 1 \\
-1 & -1
\end{pmatrix} \begin{pmatrix}
u \\
v
\end{pmatrix} + \begin{pmatrix} 0 \\
0 \\
180
\end{pmatrix}
\]

Physical Parameters
Least Squares Parameters
The observational equations will then become:

\[
\begin{pmatrix}
\alpha \\
\beta \\
\gamma
\end{pmatrix}
= \begin{pmatrix}
1 & 0 \\
0 & 1 \\
-1 & -1
\end{pmatrix}
\begin{pmatrix}
u \\
v
\end{pmatrix}
+ \begin{pmatrix}
0 \\
0 \\
180
\end{pmatrix}
= \begin{pmatrix}
73 \\
46 \\
55
\end{pmatrix}
\]

so that

\[
\begin{pmatrix}
1 & 0 \\
0 & 1 \\
-1 & -1
\end{pmatrix}
\begin{pmatrix}
u \\
v
\end{pmatrix}
= \begin{pmatrix}
73 \\
46 \\
55
\end{pmatrix}
- \begin{pmatrix}
0 \\
0 \\
180
\end{pmatrix}
= \begin{pmatrix}
73 \\
46 \\
-125
\end{pmatrix}
\]

Premultiply by the lhs matrix to give the normal equations:

\[
\begin{pmatrix}
2 & 1 \\
1 & 2
\end{pmatrix}
\begin{pmatrix}
u \\
v
\end{pmatrix}
= \begin{pmatrix}
198 \\
171
\end{pmatrix}
\]

The solution is \( u = 75, \ v = 48 \)

so that \( \alpha = 75^\circ, \ \beta = 48^\circ, \ \gamma = 180 - (\alpha + \beta) = 57^\circ \)
Non-linear least squares

If the sides of the triangle are known, they can be added to the observational equations:  \( a = 21 \text{ m}, \quad b = 16 \text{ m}, \quad c = 19 \text{ m}. \)

The sides are related to the angles through the sine and cosine rules,

\[
\frac{a}{\sin(\alpha)} = \frac{b}{\sin(\beta)} = \frac{c}{\sin(\gamma)}
\]

e.g.

which can be used as restraints or constraints.
Non-linear least squares contd.

Observational equations:
\[
\begin{align*}
\alpha &= 73 \\
\beta &= 46 \\
\gamma &= 55 \\
a &= 21 \\
b &= 16 \\
c &= 19
\end{align*}
\]

Restraint equations:
\[
\begin{align*}
a \sin(\beta) - b \sin(\alpha) &= 0 \\
b \sin(\gamma) - c \sin(\beta) &= 0 \\
\alpha + \beta + \gamma &= 180
\end{align*}
\]

The equations are now non-linear and no general method of solving non-linear simultaneous equations exists.

It is necessary to start with approximate values for the parameters and calculate shifts that will reduce the value of \[\sum_{i} w_i \varepsilon_i^2\]

**Care must be taken with the weights as the angles and lengths are completely different quantities.**
Calculation of parameter shifts

Restraint equation: \[ a \sin(\beta) - b \sin(\alpha) = 0 \]

With approximate parameter values, the equation leaves a residual \( \varepsilon \):
\[ f(\alpha, \beta, a, b) = a \sin(\beta) - b \sin(\alpha) = 0 + \varepsilon \]

Adjust the values of \( \alpha, \beta, a, b \) to reduce \( \varepsilon \) to zero. This means the shifts to the parameters should satisfy:
\[
\frac{\partial f}{\partial a} \Delta a + \frac{\partial f}{\partial \beta} \Delta \beta + \frac{\partial f}{\partial b} \Delta b + \frac{\partial f}{\partial \alpha} \Delta \alpha = -\varepsilon
\]

i.e. \[ \sin(\beta) \Delta a + a \cos(\beta) \Delta \beta - \sin(\alpha) \Delta b - b \cos(\alpha) \Delta \alpha = -\varepsilon \]

which is a linear equation in the shifts \( \Delta \alpha, \Delta \beta, \Delta a \) and \( \Delta b \)
Parameter shifts contd.

Repeating this for all observational and restraint equations gives a system of linear simultaneous equations:

\[ A \Delta x = -\epsilon \]

\[ \text{matrix of} \quad \text{vector of} \quad \text{vector} \]
\[ \text{of} \quad \text{derivatives} \quad \text{shifts} \quad \text{residuals} \]

A weighted least squares solution of these equations is obtained from

\[ A^T W A \Delta x = -A^T W \epsilon \]

Then the parameters are updated: \( x_{new} = x_{old} + \Delta x \)

and the new values are used to repeat the calculation to obtain further shifts until the shifts become negligibly small.
Ill-conditioning
Solution of Simultaneous Equations

\[
\begin{bmatrix}
3 & 6 \\
4 & 2 \\
\end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 21 \\ 22 \end{bmatrix}
\]

(ask for volunteer to solve these equations)
Try solving 2 more simultaneous equations

\[
\begin{bmatrix}
3 & 6 \\
2 & 4
\end{bmatrix}
\begin{bmatrix}
x \\
y
\end{bmatrix}
=
\begin{bmatrix}
21 \\
14
\end{bmatrix}
\]
Ill - conditioning

Consider the equations:

\[ 23.3 \, x + 37.7 \, y = 14.4 \]
\[ 8.9 \, x + 14.4 \, y = 5.5 \]

The exact solution is:

\[ x = -1, \quad y = +1 \]

Now change the rhs of the first equation to 14.39 and solve again:

\[ 23.3 \, x + 37.7 \, y = 14.39 \]
\[ 8.9 \, x + 14.4 \, y = 5.5 \]

This gives:

\[ x = 13.4, \quad y = -7.9 \]

The tiny change in one of the equations has made the solution unrecognisable.
The reason for the ill-conditioning is that the second equation is an almost exact copy of the first, multiplied by a scale factor.

Putting them on the same scale gives:

\[ 23.300 x + 37.700 y = 14.400 \]
\[ 23.300 x + 37.699 y = 14.399 \]

The second equation adds almost no new information to the first.
Symptoms of ill-conditioning

1. A small change to any of the equations results in a large change in the solution.

2. The determinant of the lhs matrix has a small value:

\[
\begin{vmatrix}
23.3 & 37.7 \\
8.9 & 14.4 \\
\end{vmatrix} = -0.01
\]

3. The elements of the inverse matrix are large:

\[
\left(\begin{array}{cc}
23.3 & 37.7 \\
8.9 & 14.4 \\
\end{array}\right)^{-1} = \left(\begin{array}{cc}
-1440 & 3770 \\
890 & -2330 \\
\end{array}\right)
\]

4. The estimated standard deviations of the parameters obtained from a least squares solution are large.

\[
M = \frac{n}{n-p} \frac{\sum w_i \varepsilon_i^2}{\sum w_i} \left(A^TWA\right)^{-1}
\]
Generalised Linear Least Squares

\[ y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \]

- The observable quantity \( y \) is linearly dependent upon the unknown quantities \( x \).
- We make the observations \( y \) under the conditions \( a \).
- \textbf{We can generally choose the values of} \( a \).
Application to the Crystallographic Model

In P-1, the structure factor expression is

\[ F_{calc} = \sum f_j e^{-8\pi^2 s^2 U_j} \cdot \cos(hx) \]

This is similar in form to

\[ y_i = \sum_j a_{ji} x_j \]

However, the cosine and exponentiation functions are \textit{non-linear}, so we cannot directly use linear least squares.
Taylor expansion: find a linear function, which locally represents your non-linear function

We can work out a linear approximation to \( \cos(x) \) if we know an approximate value of \( x \).
Similarly for exponential and other functions.

To carry out least squares, we must linearise the function. To do this we need a starting guess for each \( x \).
Given this guess, \( a \), we can approximate the function with a linear one by using only linear terms of a Taylor series expansion:

\[
 f(x) \approx f(a) + f'(a)(x - a)
\]

\( e.g. \) \( \cos(x) \) at \( x=3.77 \)

\[
 f(x) \approx \cos(3.77) - \sin(3.77)(x - 3.77)
 = -0.81 - (-0.59(x - 3.77))
 = 0.59x - 3.03
\]

The graph (left) shows linear approximations of \( \cos(x) \) at \( x=1.63 \) and \( x=3.77 \).
The problem is now that the linear approximation depends on the current estimate of \( x \). We must iterate towards a stable solution.
Non-linear least squares

The Taylor series expansion of $F_{\text{calc}}$ w.r.t. each of the parameters:

\[ f(x) \approx f(a) + f'(a)(x - a) \]

\[ F_{\text{calc}} \approx F_{\text{calc}}^k + \sum_j \frac{dF}{dx_j} (x_j - x_j^k) = F_{\text{calc}}^k + \sum_j \frac{dF}{dx_j} \Delta x_j \]

The residual:

\[ \varepsilon_i = F_{\text{obs}} - \left( F_{\text{calc}}^k + \sum_j \frac{dF}{dx_j} \Delta x_j \right) \]

\[ \sum_j \frac{dF}{dx_j} \Delta x_j = F_{\text{obs}} - F_{\text{calc}}^k \]

The observational equations. This time we solve for $\Delta x$. Our ‘observations’ are $\Delta F$ and the design matrix is all the first derivatives of the structure factor expression (linearised).

\[ x_j \approx x_j^k + \Delta x_j \]

$\Delta x_j$ are the increments to $x_j$, which make successive improvements.

$k$ is the iteration number.
Problems with Crystallographic LSQ

- The right-hand side is not $F_o$, it is $F_o - F_c$, with $F_c$ computed from the current model.

- The left-hand side is a derivative, also computed from a model. *We cannot choose it.*

\[
\sum \frac{dF_{calc}}{dx} \Delta x = F_{obs} - F_{calc}
\]
Getting a Good Model

One definition of a good model is one that minimises the difference between the observations and values computed from the model.

Small variations in the details of the calculation enable one to minimise different Minimisation Functions, $M$.

\[
M = \sum |Fo - Fc|
\]

\[
M = \sum (Fo - Fc)^2
\]

\[
M = \sum (Fo^2 - Fc^2)^2
\]
Problems with Crystallographic LSQ

Consequences of the Taylor Expansion:

- **We need a starting model** to be able to compute $F_c$ and the derivatives.
- The starting model must be pretty good - no positional errors of more than ½ Angstrom.
- The calculated shifts, $\Delta x$, are only approximate (because the derivatives are approximate), so the refinement is iterative.
- **Refinement converges to a local minimum** – there may be a more profound global minimum somewhere else.
Crystallographic Refinement: Local minima

Refinement optimises hundreds of parameters at the same time. Let’s consider just two – the x and y coordinate of a chlorine atom.

How does the minimisation function change as x and y are altered?

\[ M = \sum (F_o - F_c)^2 \]
Crystallographic Refinement: Local minima

It is not necessarily downhill all the way to a global minimum.

Refinement path of a single atom in the x-y hyper-surface, all other parameters remaining fixed. *The vertical axis is the minimisation function.*

If refinement is started from the red arrow, convergence is to a local minimum.
Refinement Problems

Sometimes the X-ray data just isn’t good enough to permit the full refinement of a structure. This can happen for several reasons:

• The crystals are of poor quality
• The crystal only diffracts to low resolution
• There is a lot of poorly ordered solvent
• There are large adps in the main molecule (motion or disorder)
Refinement Problems

Sometimes the model is so poor that refinement yields an unreasonable structure:

• Very large isotropic or anisotropic adp may indicate inclusion of a spurious atom
• Very extended anisotropic adp may indicate disorder or even omission of a neighbouring atom
• A difference map has large residual peaks – missing atoms, solvent or disorder.
Refinement with Supplementary Conditions

If the X-ray data themselves are inadequate to fully resolve a problem, additional non-diffraction information can be introduced into the calculation.

This can include both geometrical and vibrational information.

- Distances can be set equal to established values, eg C-H to 0.95Å, C-C in a phenyl group to 1.395Å.
- If the molecule is expected to exhibit symmetry, related bonds or angles can be set equal to each other.
- The adps of adjacent atoms can be set to be similar to each other.
Refinement with Supplementary Conditions

The supplementary information can be introduced in such a way that it is *totally enforced*. This is called *constrained* refinement. Examples include forcing groups to be planar, forcing hydrogen atoms to obey idealised geometries, forcing the occupancies in disordered fragments to sum to unity.

Alternatively (and more commonly) the information can be *tentatively introduced*, as advice for the calculation. This is called *restrained* refinement. Examples include setting groups of atoms to be approximately planar, or bonds to be approximately a given length.
Restraints

Example: Bond length restraint

Compare

\[
F_h = \sum_{1}^{n} f_n \cos 2\pi (hx_n + ky_n + lz_n)
\]

with

\[
d^2 = (x_a - x_b)^2 + (y_a - y_b)^2 + (z_a - z_b)^2
\]

Value = function of structural parameters.

*This is an equation of restraint*

In practice it is a little more difficult because ‘d’ must be computed from orthogonal parameters.
Equation of Restraint

- The equations of restraint (e.g. distances) are just set up in the same way as the structure factor equations and added in to the minimisation function with suitable weights.

\[
T_{obs} - T_{calc} = \sum \frac{\partial T_{calc}}{\partial \mathbf{x}} \cdot \Delta \mathbf{x}
\]

\[
M = \sum \text{All observations} \left( F_{obs}^2 - F_{calc}^2 \right)
\]

\[
+ \sum \text{All restraints} \left( T_{obs} - T_{calc} \right)^2
\]

\[
\begin{bmatrix}
w_{11} & w_{12} & \ldots & w_{1n} \\
w_{21} & w_{22} & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots \\
w_{n1} & \ldots & \ldots & w_{nn}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\frac{\partial F_{calc}}{\partial x_1} & \frac{\partial F_{calc}}{\partial x_m} \\
\frac{\partial F_{calc}}{\partial x_1} & \frac{\partial F_{calc}}{\partial x_m} \\
\frac{\partial F_{calc}}{\partial x_1} & \frac{\partial F_{calc}}{\partial x_m} \\
\frac{\partial F_{calc}}{\partial x_1} & \frac{\partial F_{calc}}{\partial x_m}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\delta x_1' \\
\delta x_2' \\
\vdots \\
\delta x_q'
\end{bmatrix}
\]

\[
\begin{bmatrix}
F_{obs} - F_{calc} \\
F_{obs} - F_{calc} \\
\vdots \\
F_{obs} - F_{calc}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\frac{\partial F_{calc}}{\partial x_1} & \frac{\partial F_{calc}}{\partial x_m} \\
\frac{\partial F_{calc}}{\partial x_1} & \frac{\partial F_{calc}}{\partial x_m} \\
\frac{\partial F_{calc}}{\partial x_1} & \frac{\partial F_{calc}}{\partial x_m} \\
\frac{\partial F_{calc}}{\partial x_1} & \frac{\partial F_{calc}}{\partial x_m}
\end{bmatrix}
\]

\[
\begin{bmatrix}
\delta x_1' \\
\delta x_2' \\
\vdots \\
\delta x_q'
\end{bmatrix}
\]

\[
\begin{bmatrix}
T_{obs} - T_{calc} \\
T_{obs} - T_{calc} \\
\vdots \\
T_{obs} - T_{calc}
\end{bmatrix}
\]
Equations of Restraint

The most commonly used restraints relate the atomic coordinates to some geometric parameter (distances, planarity), or satisfy some physical law (Hirshfeld Condition on adps)
Distance Restraint

BOLD = X-ray minima
Faint = restraint minima
Thermal Restraints
Hirshfeld Condition

The components of the adp in selected directions can be set to be similar.
Restrainted Refinement

We saw this structure earlier.

In the 5-membered ring C114, C115, C118, C120, O119, the oxygen could be on either side of the plane through the carbon atoms.

This uncertainty has been modelled by a long, elliptical adp.

The consequence of this disorder is more evident in the adjoining azide group, which has been modelled with pairs of alternative atom sites.

Here, it is likely that the crystallographer has constrained the elements of the adps of equivalent atoms to be exactly the same.

It is also likely that the bond lengths and angles in each moiety have been restrained to be very similar.
Are Restraints Cheating?

Queen Victoria’s Penny
Restraints

Restraints run the risk of introducing false information into a refinement if the assumptions are wrong.

However, there is a considerable advantage over constraints because existing data can still resist their influence – investigate carefully when this happens.
Refinement

To date, no one has been able to automate the refinement of anything but trivial problems.

As a chemist, you should be aware that there is no definitive refinement procedure.

If fine details of a structure are important to you, you should look carefully at the quality of the data and the refinement protocol used.
Least squares: the complete scheme

Tableau for damped, weighted, constrained, restrained Least Squares
Crystallographic Refinement

In crystallography the design matrix $A$ has one column for each variable (9 per anisotropic atom) and one row for each reflection.

Thus, for a 30 atom structure there are 270 variables, and there will generally be about 10 times this number of reflections, 2700.

The matrix $A$ contains 700,000 elements.

The matrix $A^tA$ only contains 70,000 elements, but each element is the sum of 2700 products.

200 million operations plus all the Sines, Cosines and exponentials is a big calculation.
Key Points

• A valid structure should fit accepted chemical and physical concepts.
• A valid structure should give calculated structure factors which match the observations.
• A very poor structure cannot be refined by least squares.
• Least squares cannot indicate missing parameters.
• Least squares may reveal surplus parameters.
• The final difference map should be featureless.
• Least squares is robust against random errors, but is vulnerable to systematic errors.
• Least squares is computationally intensive.