On Minimization Targets and Algorithms

Dale E. Tronrud
Howard Hughes Medical Institute
Institute of Molecular Biology
University of Oregon, USA
http://www.uoxray.uoregon.edu/dale/

Refinement and Rebuilding

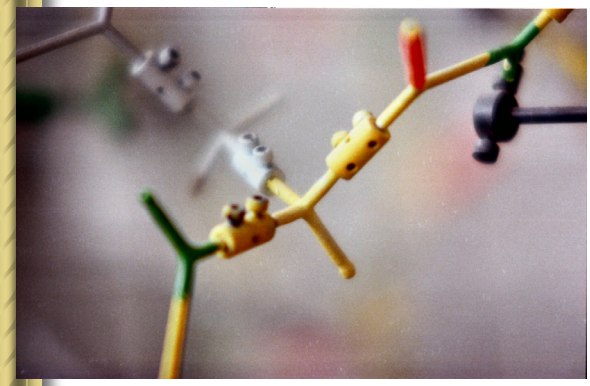
- Refinement and model rebuilding cannot be considered separately. Each is useless without the other.
- "Refinement" is a process where continuous changes are made to the model with consideration given to the observations.
- "Rebuilding" requires consideration of the context of each portion of the model with an eye towards making large changes, e.g. rotomer changes, peptide flips, as well as adding and deleting atoms.
- Rebuilding is usually manually controlled, although some protocols, such as arp/warp, are automated.

What is Refinement?

- Refinement is the optimization of a function by changing the parameters of a model.
- What are the parameters?
  - Usually everyone agrees that the parameters are the position, B factor, and occupancy of each atom.
  - In most cases other parameters are added.
  - In some special cases different parameters are chosen.
- What is the function?
  - Maximum likelihood in REFMAC, CNS, BUSTER
  - Least squares in SHELX and TNT
  - Empirical energy in X-PLOR, CNS

- What minimization method?
  - Choice depends upon the function and the model.
  - Usually a variety of methods will be used on the same problem.
  - A variety of methods are used; several being used in the same project.
    - Simulated Annealing
    - Conjugate Gradient
    - Preconditioned Conjugate Gradient
    - Sparse Matrix
    - Full Matrix
    - . . .

What Parameters to Refine?



Elaborate Parameterization

- If one holds to the traditional parameterization of position and isotropic B's
  - High resolution and more precise lower resolution data will be wasted.
- The cost of elaboration is an increase in the total number of parameters.
  - Too many parameters leads to "overfitting" and poor refinement results.
  - Finding the right balance is difficult and will vary from case to case.
- The recent success stories are:
  - Torsion angle simulated annealing at low resolution
  - TLS anisotropic B's at midrange resolution

Who's Got What?

- All packages can refine positions and isotropic B's and atomic occupancies.
- All packages can perform rigid body refinement.
- SHELLX, REFMAC, and Restrain have individual anisotropic B's.
- REFMAC has TLS anisotropic B's built as a "tree".
- CNS has a "torsion-angle" parameterization which greatly improves simulated annealing refinement.
- Each program has many options to vary parameterization. Check the documentation for the options available to you.

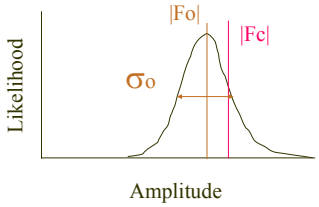
What Function to Optimize?

Maximum Likelihood

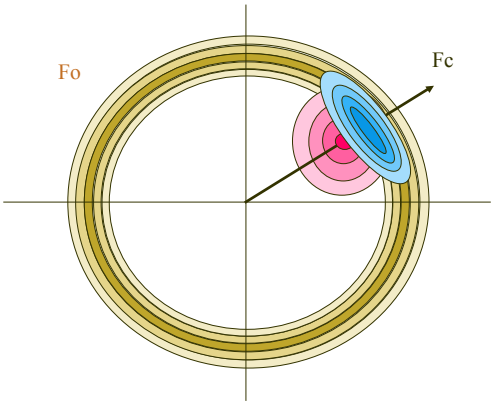
- Is a generalized statistical framework for optimizing models
  - Which means it is somewhat vague.
- The best set of parameters is the one for which the probability is the greatest that the experiment performed would result in the measured values.
- The character of the errors in your model are specifically built into the target function.
- If the errors in the experiment and the model obey certain assumptions Least Squares is the proper Maximum Likelihood method for the problem.
- Generally, the larger the errors the less applicable Least Squares becomes.

Maximum Likelihood (cont.)

- Least Squares assumes that $\frac{|F_o| - |F_c|}{\sigma_o}$ for all reflections obeys a Normal distribution with a mean of zero and a standard deviation of one.
- Least Squares view of the world:



Maximum Likelihood's View



Difficulties in Maximum Likelihood

- What is the character of the uncorrected errors in the model?
  - Existing programs assume the errors behave like randomly, and Normally distributed displacements of atomic positions and B factors.
  - Buster offers the option of non-uniform distribution of errors
    - It has a two state error model, where one part is treated in the usual way, but another part is identified only by a region of space and an elemental composition.
- How does one estimate the quantity of error in the model?
  - All ML programs use the agreement of the model to the test set to calibrate the error level.

The Least-squares Function

- $Q_o(i)$ Observed quantity i
- $\sigma_o^2(i)$ Observed variance of quantity i
- $\mathbf{p}$ Parameters of a model
- $Q_c(i,\mathbf{p})$ Corresponding quantity inferred from the current model

- The best $\mathbf{p}$ is that which minimizes

$$f = \sum_i \frac{(Q_o(i) - Q_c(i,\mathbf{p}))^2}{\sigma_i^2(i)}$$

Different Classes of Observations

- This equation can easily handle observations of many different classes.

$$f = \sum_i \frac{1}{\sigma_i^2(i)} (Q_o(i) - Q_c(i,\mathbf{p}))^2 + \sum_i \frac{1}{\sigma_i^2(i)} (Q_o(i) - Q_c(i,\mathbf{p}))^2 + \cdots$$

- Examples of classes are:
  - Structure factor amplitudes
  - Bond length and angles
  - Torsion angles and planarity
  - Non-bonded contacts
  - Stereochemical B factor correlation
  - MIR Phases
  - Noncrystallographic symmetry

Major Limitation of this Equation

- The equation assumes that the observations are statistically independent. Often this is not the case.
  - Some programs use non-independent stereochemical restraint categories.
  - Many particular stereochemical targets are correlated.
  - The presence of noncrystallographic symmetry creates dependencies between some (many) reflections.
- This limitation also exists in all Maximum Likelihood implementations to date.

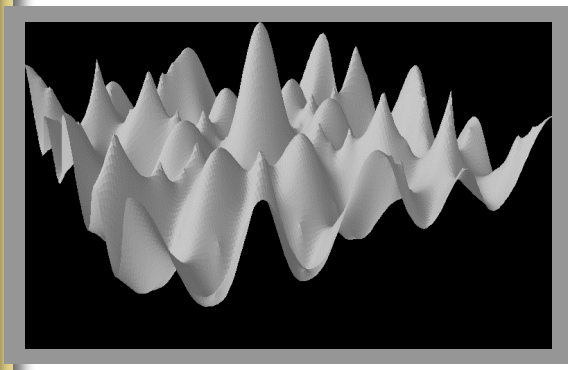
Energy Minimization

- The best model with the one with the lowest energy.
- How does one calculate the “energy” of a model?
- How does the diffraction data become “energy”?
- How does one reconcile the instantaneous nature of energy with the time averaged nature of the diffraction data?
- Why bother when a statistically based method has answers to all these questions?

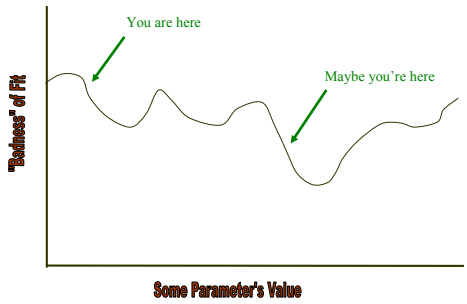
How to Optimize the Function?

Methods of Minimization

- Methods using no function derivatives
  - Simulated Annealing, Monte Carlo, Simplex, Metropolis
- Methods using first derivatives
  - Steepest Descent, Conjugate Gradient
- Methods using first and second derivatives
  - Full matrix, Block diagonal, Diagonal, Preconditioned Conjugate Gradient (, and Conjugate Gradient II)



Simulated Annealing



Full Matrix Minimization

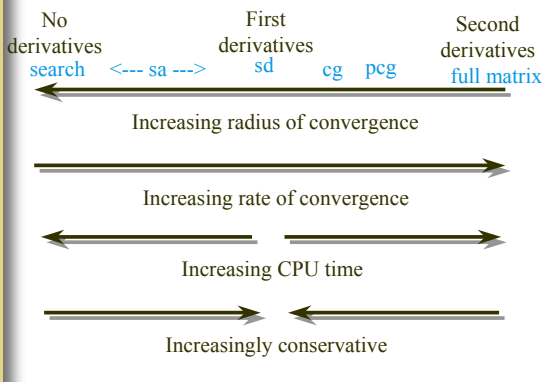
$$-\left|\frac{d^2 f(\mathbf{p})}{d\mathbf{p}d\mathbf{p}}\right|_{\mathbf{p}=\mathbf{p}_0}^{-1}\left|\frac{df(\mathbf{p})}{d\mathbf{p}}\right|_{\mathbf{p}=\mathbf{p}_o} = \mathbf{s}$$

- If the function is not quadratic
  - more than one cycle is required to reach the minimum.
  - an initial guess for the parameters is required.
- The second derivative matrix is huge
  - very time consuming to calculate and invert.
- The power of convergence is great.
- The radius of convergence is very poor.
- It absolutely requires an overdetermined problem.

Approximations to Full Matrix

- Sparse Matrix
  - Only large matrix elements are used
- Block Diagonal
  - Assumes the parameters can be categorized
- Preconditioned Conjugate Gradient
  - Assumes all off diagonal elements are zero, but learns the truth from experience
- Gradient / Curvature
  - Assumes all off diagonal elements are zero, and is pig-headed about it.
- Conjugate Gradient
  - Assumes all diagonal elements are equal, but learns from experience
- Steepest Descent
  - Assumes all diagonal elements are equal

The Minimization Continuum



Who's Got What?

- Full matrix refinement is available
  - SHELLX, REFMAC, and RESTRAIN.
  - Full matrix programs usually allow for the exclusion of subsets of off-diagonal elements.
  - They will also offer a selection of means to approximate the inverse of the Normal matrix.
- Ignoring all off-diagonal elements leads to the diagonal approximation and preconditioned conjugate gradient.
  - TNT and BUSTER
- Ignoring the diagonal elements leads to steepest descent and conjugate gradient.
  - X-PLOR and CNS

In Conclusion...

- This talk should help you to ask the right questions to find out how one refinement package differs from another.
- One must know the differences between the packages to make a rational decision about which to use in your project.
- One must also know the limitations of their package and what problems to watch for. No package is perfect.