

Overview of Crystallographic Structure Refinement

Lynn F. Ten Eyck

San Diego Supercomputer Center
 Department of Pharmacology
 University of California, San Diego

IUCr Computing School, Siena, 2005

Outline

- 1 Introduction
- 2 Problem Statement
- 3 Optimization Methods
- 4 Description and Parameterization of Model
- 5 Implementation Strategy
- 6 Summary

Introduction

- The objective is to obtain the best estimate of what is in the crystal.
 - “Best” means a model that contains all essential elements
 - “Best” also means measures of certainty about what is or is not present
 - “Best” also means quantitative estimates of parameters and uncertainties in parameters
- The method is optimization of the fit of data calculated from the model to the observed data
 - Robust optimization
 - Sensitive optimization
 - Use optimized results to find problems with the model

Problem Statement

We are optimizing a function of our current model and our observations.

$$\Phi(\mathbf{p}) = \sum_{obs} w_i (ky_{i,o} - f(i, \mathbf{p}))^2$$

$$\frac{d}{d\mathbf{p}} \Phi(\mathbf{p}) = -2 \sum_{obs} w_i (ky_{i,o} - f(i, \mathbf{p})) \left(\frac{\partial f(i, \mathbf{p})}{\partial \mathbf{p}} \right)$$

Observations include both diffraction and chemistry, i.e. bond lengths and so forth.

Optimization Methods

All of the methods depend on the Taylor expansion, usually truncated to first order, but sometimes taken to second order

$$\Phi(\mathbf{p}) \approx \Phi_0 + \left\langle \left(\frac{\partial \Phi}{\partial \mathbf{p}} \right) \middle| \Delta \mathbf{p} \right\rangle + \dots$$

- Linearize $\mathbf{F}_c(\mathbf{h})$ in parameters and form normal equations
 - Solve by Conjugate Gradients
 - Solve by Linear Algebra
- Solve directly by non-linear function minimizer

Description of Model

- Description should take advantage of polymeric nature of biological macromolecules
- Description should **NOT** assume proteins are pure polymers!
 - Ligands, covalent and otherwise
 - Cross-links
 - A-B-A interactions
- Constraints are definitions, not refinable parameters
- Restraints are additional observation with weights

Implementation

- Break down into steps
 - Model \Rightarrow Map
 - Map \Rightarrow Structure Factors
 - Structure Factors + Model \Rightarrow Objective Function
 - Objective Function \Rightarrow Gradient wrt SF
 - Gradient wrt SF \Rightarrow Gradient wrt Map
 - Gradient wrt Map \Rightarrow Gradient wrt Model
- Expose each step for possible editing
- Especially note editing of gradients to implement constraints

Summary

- Design for Change
- Keep overall concept SIMPLE
- Efficiency only matters in places
 - In those places, it matters a LOT
- DESIGN FOR VALIDATION