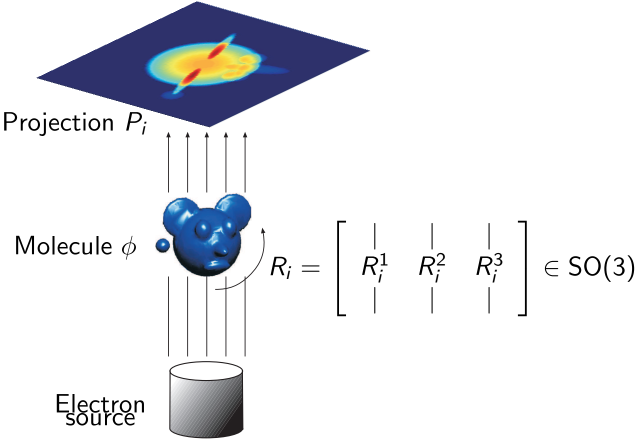
The Inverse Problem of Cryo-Electron Microsopy

Amit Singer

Princeton University, Department of Mathematics and PACM

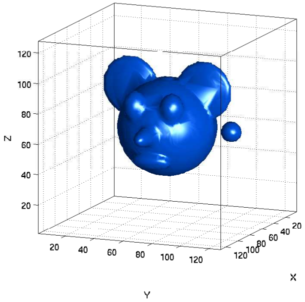
Bridging the Gap Between the Geosciences and Mathematics,
Statistics, and Computer Science
October 2, 2012

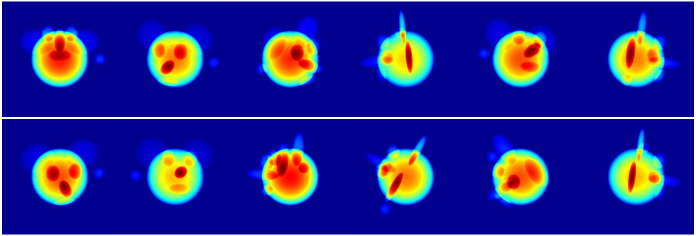
Single Particle Cryo-Electron Microscopy: Model



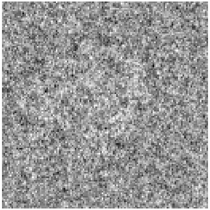
- Projection images $P_i(x,y) = \int_{-\infty}^{\infty} \phi(xR_i^1 + yR_i^2 + zR_i^3)\, dz +$ "noise".
- $\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ is the electric potential of the molecule.
- Cryo-EM problem: Find ϕ and $R_1, \ldots, R_n$ given $P_1, \ldots, P_n$.

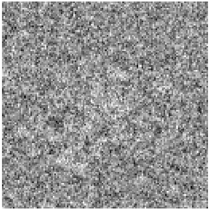
Toy Example

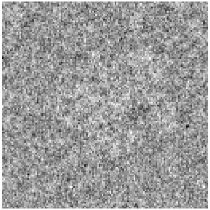


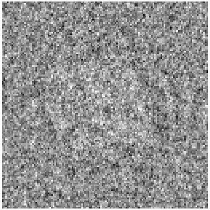


E. coli 50S ribosomal subunit: sample images
Fred Sigworth, Yale Medical School

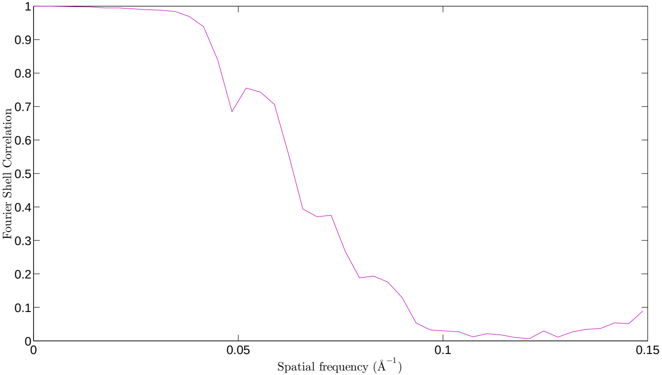








Movie by Lanhui Wang and Zhizhen (Jane) Zhao



Algorithmic Pipeline

- **Particle Picking:** manual, automatic or experimental image segmentation.
- **Class Averaging:** classify images with similar viewing directions, register and average to improve their signal-to-noise ratio (SNR).
  S, Zhao, Shkolnisky, Hadani, SIIMS, 2011.
- **Orientation Estimation:**
  S, Shkolnisky, SIIMS, 2011.
- **Three-dimensional Reconstruction:** a 3D volume is generated by a tomographic inversion algorithm.
- **Iterative Refinement**

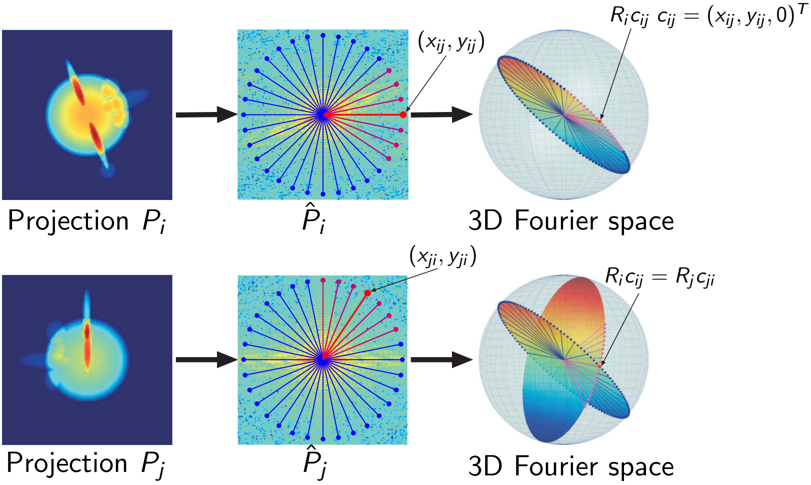
Assumptions for Today's talk:

- Trivial point-group symmetry
- Homogeneity

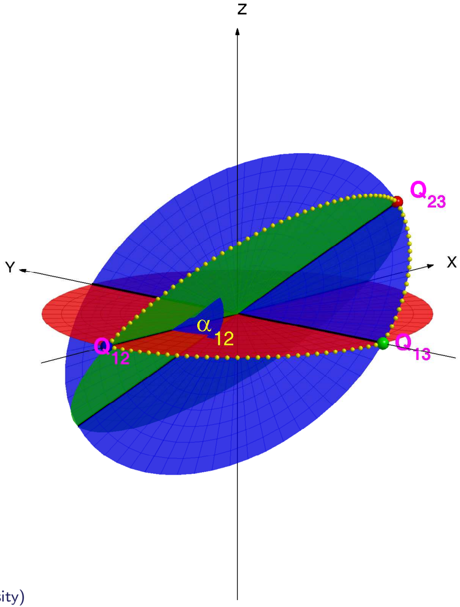
What mathematics do we use to solve the inverse problem?

- Tomography
- Convex optimization and semidefinite programming
- Random matrix theory (in several places)
- Representation theory of SO(3) (spherical harmonics)
- Spectral graph theory, (vector) diffusion maps
- Fast randomized algorithms
- ...

Orientation Estimation: Fourier projection-slice theorem



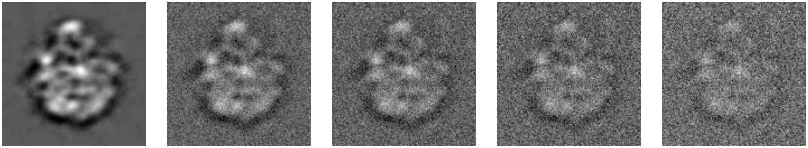
Angular Reconstitution (Van Heel 1987, Vainshtein and Goncharov 1986)



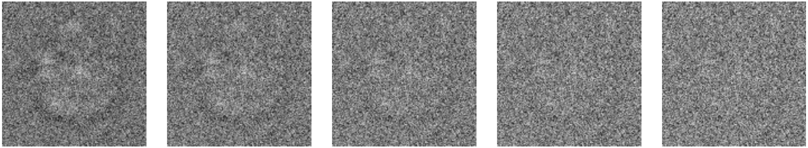
Experiments with simulated noisy projections

- Each projection is 129x129 pixels.

$$\text{SNR} = \frac{\text{Var}(Signal)}{\text{Var}(Noise)},$$



(a) Clean (b) SNR=2^{0} (c) SNR=2^{-1} (d) SNR=2^{-2} (e) SNR=2^{-3}



(f) SNR=2^{-4} (g) SNR=2^{-5} (h) SNR=2^{-6} (i) SNR=2^{-7} (j) SNR=2^{-8}

Fraction of correctly identified common lines and the SNR

- Define common line as being correctly identified if both radial lines deviate by no more than 10° from true directions.
- Fraction p of correctly identified common lines increases by PCA

$\log_2(\text{SNR})$	p
20	0.997
0	0.980
-1	0.956
-2	0.890
-3	0.764
-4	0.575
-5	0.345
-6	0.157
-7	0.064
-8	0.028
-9	0.019

Least Squares Approach

- Consider the unit directional vectors as three-dimensional vectors:

$$\begin{aligned} c_{ij} &= (x_{ij}, y_{ij}, 0)^T, \\ c_{ji} &= (x_{ji}, y_{ji}, 0)^T. \end{aligned}$$

- Being the common-line of intersection, the mapping of c_{ij} by R_i must coincide with the mapping of c_{ji} by R_j: $(R_i, R_j \in SO(3))$

$$R_i c_{ij} = R_j c_{ji}, \text{ for } 1 \leq i < j \leq n.$$

- Least squares or Energy minimization:

$$\min_{R_1, R_2, \ldots, R_n \in SO(3)} \sum_{i \neq j} \| R_i c_{ij} - R_j c_{ji} \|^2$$

- Non-convex... Exponentially large search space...

Spectral Relaxation for Uniformly Distributed Rotations

$$\tilde{R}_i = \begin{bmatrix} | & | \\ R_i^1 & R_i^2 \\ | & | \end{bmatrix} = \begin{bmatrix} x_i^1 & x_i^2 \\ y_i^1 & y_i^2 \\ z_i^1 & z_i^2 \end{bmatrix}, \quad i = 1, \ldots, n.$$

- Define 3 vectors of length $2n$

$$\begin{aligned} x &= \begin{bmatrix} x_1^1 & x_1^2 & x_2^1 & x_2^2 & \cdots & x_n^1 & x_n^2 \end{bmatrix}^T \\ y &= \begin{bmatrix} y_1^1 & y_1^2 & y_2^1 & y_2^2 & \cdots & y_n^1 & y_n^2 \end{bmatrix}^T \\ z &= \begin{bmatrix} z_1^1 & z_1^2 & z_2^1 & z_2^2 & \cdots & z_n^1 & z_n^2 \end{bmatrix}^T \end{aligned}$$

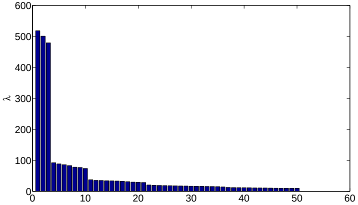
- Rewrite the objective function as

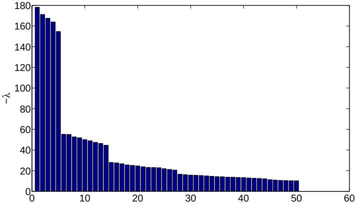
$$\max_{R_1,\ldots,R_n \in SO(3)} \sum_{i \neq j} \langle R_i c_{ij}, R_j c_{ji} \rangle = \max_{R_1,\ldots,R_n \in SO(3)} x^T C x + y^T C y + z^T C z$$

- By **symmetry**, if rotations are uniformly distributed over $SO(3)$, then the top eigenvalue of C has multiplicity 3 and corresponding eigenvectors are x, y, z from which we recover $R_1, R_2, \ldots, R_n$!

Spectrum of C

- Numerical simulation with $n = 1000$ rotations sampled from the Haar measure; no noise.
- Bar plot of positive (left) and negative (right) eigenvalues of C:





- Eigenvalues: $\lambda_l \approx n\frac{(-1)^{l+1}}{l(l+1)}, \quad l = 1, 2, 3, \ldots \ (\frac{1}{2}, -\frac{1}{6}, \frac{1}{12}, \ldots)$
- Multiplicities: $2l + 1$.
- Two basic questions:
  1. Why this spectrum? Answer: Representation Theory of SO(3) (Hadani, S, 2011)
  2. Is it stable to noise? Answer: Yes, due to random matrix theory.