Three-dimensional structure determination of molecules without crystallization: from electron microscopy to semidefinite programming

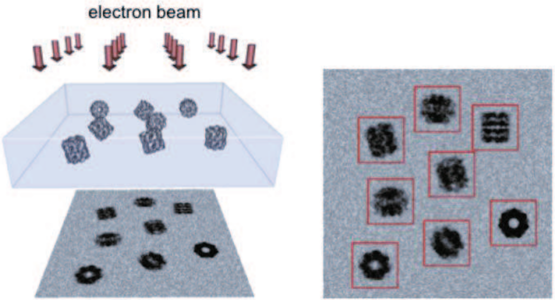
Amit Singer

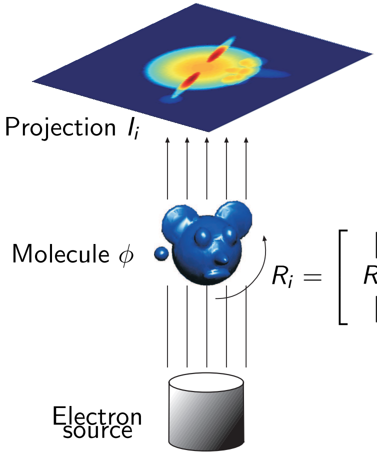
Princeton University, Department of Mathematics and PACM

February 13, 2014

Single Particle Cryo-Electron Microscopy

Drawing of the imaging process:

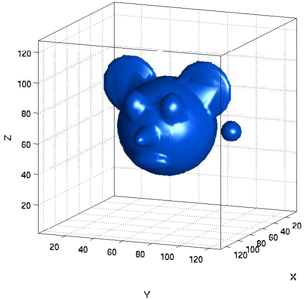


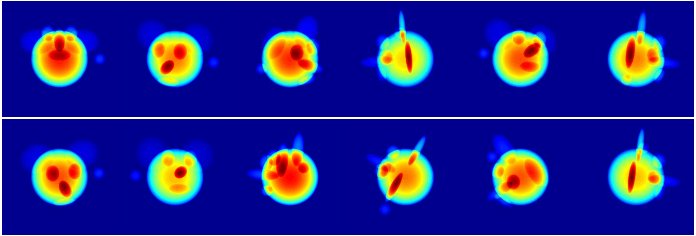


$$R_i = \begin{bmatrix} | & | & | \\ R_i^1 & R_i^2 & R_i^3 \\ | & | & | \end{bmatrix} \in SO(3)$$

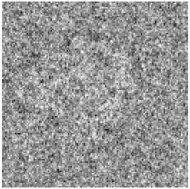
- Projection images $I_i(x,y) = \int_{-\infty}^{\infty} \phi(xR_i^1 + yR_i^2 + zR_i^3)\,dz +$ "noise".
- $\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ is the electric potential of the molecule.
- Cryo-EM problem: Find ϕ and $R_1, \ldots, R_n$ given $I_1, \ldots, I_n$.

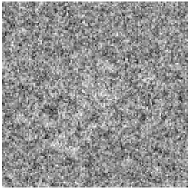
Toy Example

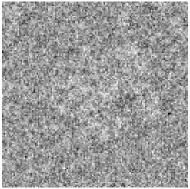


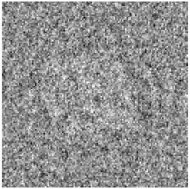


E. coli 50S ribosomal subunit: sample images
Fred Sigworth, Yale Medical School









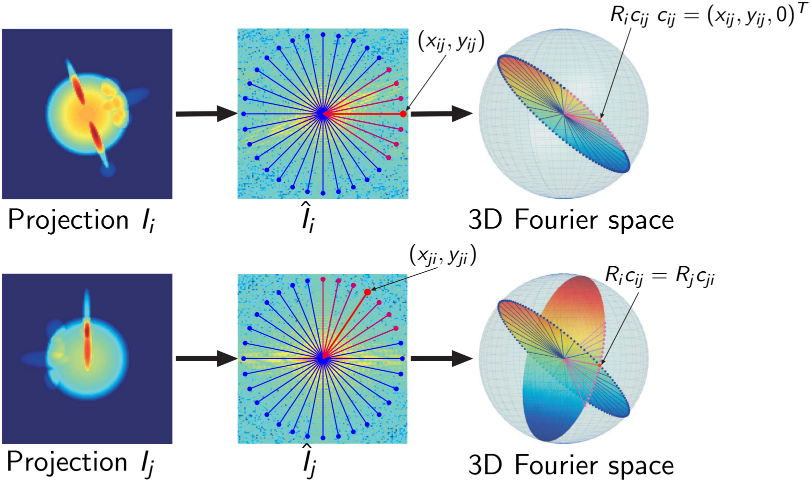
Movie by Lanhui Wang and Zhizhen (Jane) Zhao

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).
- **Orientation Estimation:**
  S, Shkolnisky, SIIMS 2011.
  Bandeira, Charikar, S, Zhu, ITCS 2014.
- **Three-dimensional Reconstruction:**
  a 3D volume is generated by a tomographic inversion algorithm.
- **Iterative Refinement**

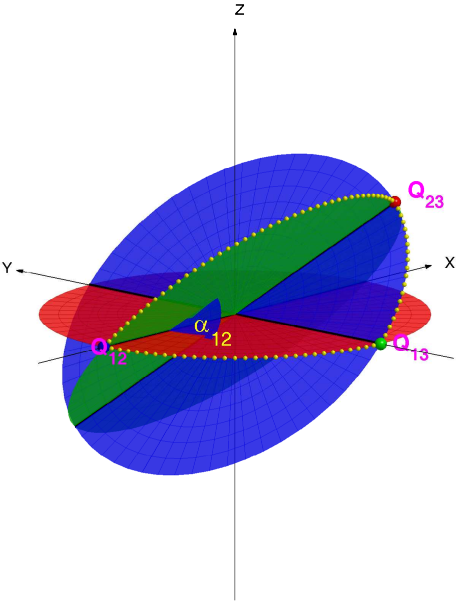
Assumptions for today's talk:

- Trivial point-group symmetry
- Homogeneity: no structural variability



(x_{ij}, y_{ij})
$R_i c_{ij}$ $c_{ij} = (x_{ij}, y_{ij}, 0)^T$
Projection I_i
$\hat{I}_i$
3D Fourier space
(x_{ji}, y_{ji})
$R_i c_{ij} = R_j c_{ji}$
Projection I_j
$\hat{I}_j$
3D Fourier space

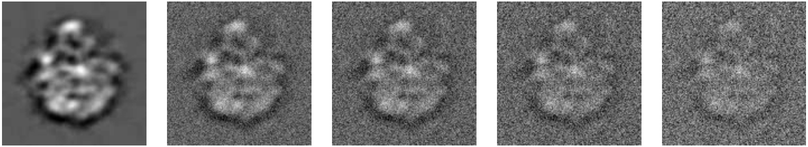
Angular Reconstitution (Van Heel 1987, Vainshtein and Goncharov 1986)



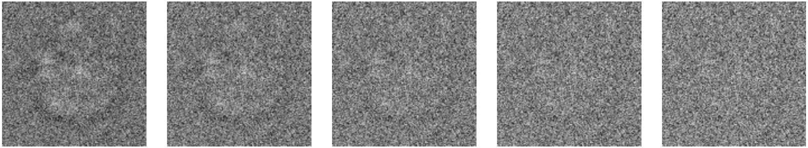
Experiments with simulated noisy projections

- Each projection is 129x129 pixels.

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



(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.

$\log_2(\text{SNR})$	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:

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

- Search space is exponentially large and non-convex.

Quadratic Optimization Under Orthogonality Constraints

- Quadratic cost: $\sum_{i \neq j} \|R_i c_{ij} - R_j c_{ji}\|^2$
- Quadratic constraints: $R_i^T R_i = I_{3\times 3}$
  ($\det(R_i) = +1$ constraint is ignored)
- We approximate the solution using SDP and rounding. Related to:
  - Goemans-Williamson (1995) SDP relaxation for MAX-CUT
  - PhaseLift (Candes et al 2012)
  - Generalized Orthogonal Procrustes Problem (Nemirovski 2007)
  - Non-commutative Grothendick Problem (Naor et al 2013)

"Robust" version – Least Unsquared Deviations (Wang, S, Wen 2013)

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

- Least squares is equivalent to maximizing the sum of inner products:

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

$$\iff \max_{R_1,R_2,\ldots,R_n\in SO(3)} \sum_{i\neq j} \mathrm{Tr}(c_{ji} c_{ij}^T R_i^T R_j) \iff \max_{R_1,R_2,\ldots,R_n\in SO(3)} \mathrm{Tr}(CG)$$

- C is the $2n \times 2n$ matrix ("the common lines matrix") with

$$C_{ij} = \tilde{c}_{ji}\tilde{c}_{ij}^T = \begin{bmatrix} x_{ji} \\ y_{ji} \end{bmatrix} \begin{bmatrix} x_{ij} & y_{ij} \end{bmatrix} = \begin{bmatrix} x_{ji}x_{ij} & x_{ji}y_{ij} \\ y_{ji}x_{ij} & y_{ji}y_{ij} \end{bmatrix}, \quad C_{ii} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

- G is the $2n \times 2n$ Gram matrix $G = \tilde{R}^T\tilde{R}$ with $G_{ij} = \tilde{R}_i^T\tilde{R}_j$:

$$G = \begin{bmatrix} \tilde{R}_1^T \\ \tilde{R}_2^T \\ \vdots \\ \tilde{R}_n^T \end{bmatrix} \begin{bmatrix} \tilde{R}_1 & \tilde{R}_2 & \cdots & \tilde{R}_n \end{bmatrix} = \begin{bmatrix} I_{2\times 2} & \tilde{R}_1^T\tilde{R}_2 & \cdots & \tilde{R}_1^T\tilde{R}_n \\ \tilde{R}_2^T\tilde{R}_1 & I_{2\times 2} & \cdots & \tilde{R}_2^T\tilde{R}_n \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{R}_n^T\tilde{R}_1 & \tilde{R}_n^T\tilde{R}_2 & \cdots & I_{2\times 2} \end{bmatrix}$$

SDP Relaxation and Rounding

$$\max_{R_1,R_2,\ldots,R_n \in SO(3)} \operatorname{Tr}(CG)$$

- **SDP Relaxation**:

$$\max_{G \in \mathbb{R}^{2n\times 2n}} \operatorname{Tr}(CG)$$
$$\text{s.t. } G \succeq 0, \quad G_{ii} = I_{2\times 2}, \ i = 1,2,\ldots,n.$$

- Missing is the non-convex constraint $\operatorname{rank}(G) = 3$.
- **Randomize** a $2n \times 3$ orthogonal matrix Q using (careful) QR factorization of a $2n \times 3$ matrix with i.i.d standard Gaussian entries
- Compute Cholesky decomposition $G = YY^T$
- **Round** using SVD: $(YQ)_i = U_i \Sigma_i V_i^T \implies \tilde{R}_i^T = U_i V_i^T$. Use the cross product to find R_i^T.
- Loss of handedness.

Spectral Relaxation for Uniformly Distributed Rotations

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

- Define 3 vectors of length $2n$

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

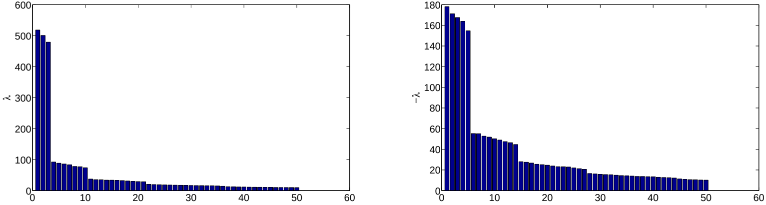
- Rewrite the least squares objective function as

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

- 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: $\lambda_l \approx n \frac{(-1)^{l+1}}{l(l+1)}, \quad l = 1, 2, 3, \ldots$. $\left(\frac{1}{2}, -\frac{1}{6}, \frac{1}{12}, \ldots\right)$
- 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.

Probabilistic Model and Wigner's Semi-Circle Law

- **Simplistic Model:** every common line is detected correctly with probability p, independently of all other common-lines, and with probability $1-p$ the common lines are falsely detected and are uniformly distributed over the unit circle.
- Let C^{clean} be the matrix C when all common-lines are detected correctly ($p=1$).
- The expected value of the noisy matrix C is
$$\mathbb{E}[C] = pC^{\text{clean}},$$
as the contribution of the falsely detected common lines to the expected value **vanishes**.
- Decompose C as
$$C = pC^{\text{clean}} + W,$$
where W is a $2n \times 2n$ zero-mean random matrix.
- The eigenvalues of W are distributed according to Wigner's semi-circle law whose support, up to $O(p)$ and finite sample fluctuations, is $[-\sqrt{2n}, \sqrt{2n}]$.

Threshold probability

- Sufficient condition for top three eigenvalues to be pushed away from the semi-circle and no other eigenvalue crossings:
  (rank-1 and finite rank deformed Wigner matrices,
  Füredi and Komlós 1981, Féral and Péché 2007, ...)

$$p\Delta(C^{\text{clean}}) > \frac{1}{2}\lambda_1(W)$$

- Spectral gap $\Delta(C^{\text{clean}})$ and spectral norm $\lambda_1(W)$ are given by

$$\Delta(C^{\text{clean}}) \approx (\frac{1}{2} - \frac{1}{12})n$$

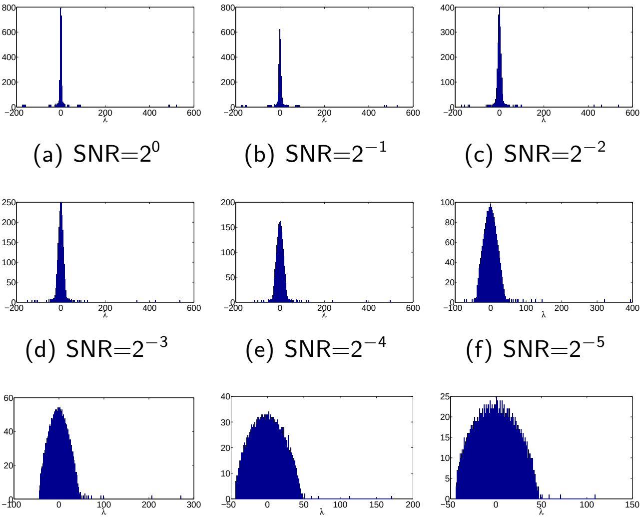
  and

$$\lambda_1(W) \approx \sqrt{2n}.$$

- Threshold probability

$$p_c = \frac{5\sqrt{2}}{6\sqrt{n}}.$$

Numerical Spectra of C, $n = 1000$



(a) SNR=2^0 (b) SNR=2^{-1} (c) SNR=2^{-2}

(d) SNR=2^{-3} (e) SNR=2^{-4} (f) SNR=2^{-5}

(g) SNR=2^{-6} (h) SNR=2^{-7} (i) SNR=2^{-8}

- True rotations: $R_1, \ldots, R_n$.
- Estimated rotations: $\hat{R}_1, \ldots, \hat{R}_n$.
- Registration:

$$\hat{O} = \operatorname*{argmin}_{O \in SO(3)} \sum_{i=1}^{N} \|R_i - O\hat{R}_i\|_F^2$$

- Mean squared error:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} \|R_i - \hat{O}\hat{R}_i\|_F^2$$

MSE for $n = 1000$

SNR	p	MSE
2^{-1}	0.951	0.0182
2^{-2}	0.890	0.0224
2^{-3}	0.761	0.0361
2^{-4}	0.564	0.0737
2^{-5}	0.342	0.2169
2^{-6}	0.168	1.8011
2^{-7}	0.072	2.5244
2^{-8}	0.032	3.5196

- Model fails at low SNR. Why?
- Wigner model is too simplistic – cannot have n^2 independent random variables from just n images.
- $C_{ij} = K(P_i, P_j)$, "random kernel matrix" (Koltchinskii and Giné 2000, El-Karoui 2010).
- Kernel is discontinuous (Cheng, S, 2013)

Maximum Likelihood Solution using SDP

- Main idea: Lift $SO(3)$ to $Sym(\mathbb{S}^2)$
- Suppose $x_1, x_2, \ldots, x_L \in \mathbb{S}^2$ are "evenly" distributed points over the sphere (e.g., a spherical t-design).
- To each $R \in SO(3)$ we can attach a permutation $\pi \in S_L$ via the group action and the assignment/Hungarian algorithm (this does not need to be constructed implicitly).
- Notice: We are discretizing S^2, not $SO(3)$ (substantial gain in computational complexity)
- We will see that the likelihood function is linear in the PSD matrix that encodes the relative permutations, and that $SO(3)$ implies further linear constraints.

Convex Relaxation of Permutations arising from Rotations

- The convex hull of permutation matrices are the doubly stochastic matrices (Birkhoff-von Neumann polytope):

$$\Pi \in \mathbb{R}^{L\times L}, \quad \Pi \geq 0, \quad 1^T\Pi = 1^T, \quad \Pi 1 = 1$$

- Rotation by an element of $SO(3)$ should "map nearby-points to nearby-points". More precisely, $SO(3)$ preserves inner products:

$$X_{ij} = \langle x_i, x_j \rangle, \quad X_{\pi(i),\pi(j)} \stackrel{\epsilon}{=} X_{ij}$$

$$\Pi X \Pi^T \stackrel{\epsilon}{=} X \Longrightarrow \Pi X \stackrel{\epsilon}{=} X \Pi$$

Convex Relaxation of Cycle Consistency: SDP

- Let G be a block-matrix of size $n \times n$ with $G_{ij} \in \mathbb{R}^{L \times L}$.
- We want $G_{ij} = \Pi_i \Pi_j^T$.
- G is PSD, $G_{ii} = I_{L \times L}$, and $\text{rank}(G) = L$ (the rank constraint is dropped).

$$G_{ij} \in \mathbb{R}^{L \times L}, \quad G_{ij} \geq 0, \quad G1 = 1$$

$$G_{ij} X \overset{\epsilon}{=} X G_{ij}$$

Maximum Likelihood: Linear Objective Function

- The common line depends on $R_i R_j^T$.
- Likelihood function is of the form

$$\sum_{i \neq j} f_{ij}(R_i R_j^T)$$

- Nonlinear in $R_i R_j^T$, but linear in G.
- Proof by picture.

Exact Recovery

- Experience with the multireference alignment problem: The solution of the SDP has the desired rank up to a certain level of noise (w.h.p).
- In other words, even though the search-space is exponentially large, we solve ML in polynomial time.
- This is a viable alternative to heuristic methods such as EM and alternating minimization.
- Can be used in a variety of problems where the objective function is a sum of pairwise interactions.
- Need better theoretical understanding for the phase transition behavior and conditions for exact recovery.

Ongoing Research in cryo-EM and Related Applications

- Ab-initio reconstruction without class averaging: scaling the SDP to larger n
- Heterogeneity
- Translations
- Contrast transfer function of the microscope, different defocus groups
- Molecules with symmetries
- Beam induced motion and motion correction
- XFEL (X-ray free electron lasers)
- Structure from motion (computer vision)

References

- A. Singer, Y. Shkolnisky, "Three-Dimensional Structure Determination from Common Lines in Cryo-EM by Eigenvectors and Semidefinite Programming", *SIAM Journal on Imaging Sciences*, **4** (2), pp. 543–572 (2011).
- L. Wang, A. Singer, Z. Wen, "Orientation Determination of Cryo-EM images using Least Unsquared Deviations", *SIAM Journal on Imaging Sciences*, **6**(4), pp. 2450–2483 (2013).
- A. S. Bandeira, M. Charikar, A. Singer, A. Zhu, "Multireference Alignment using Semidefinite Programming", 5th Innovations in Theoretical Computer Science (ITCS 2014).

Thank You!

Funding:

- NIH/NIGMS R01GM090200
- AFOSR FA9550-12-1-0317
- Simons Foundation LTR DTD 06-05-2012