Solving the 3D Puzzle of Rotation Assignment in Single Particle Cryo-Electron Microscopy

Amit Singer

Princeton University

Department of Mathematics and Program in Applied and Computational Mathematics

May 15, 2015
Optimization problem:

$$\min_{g_1, g_2, \ldots, g_n \in G} \sum_{i,j=1}^{n} f_{ij}(g_i g_j^{-1})$$

$G = SO(3)$ is the group of rotations in space.

Parameter space $G \times G \times \cdots \times G$ is exponentially large.
High SNR: Matched filtering, phase-correlation

Low SNR: Failure of pairwise comparisons.

How to use information in all signals?

Invariants: Power spectrum, higher order spectra (bispectrum).

Is maximum likelihood estimation (MLE) computationally tractable?
Let $G$ be a group of transformations acting on a vector space $X$, e.g., $G = SO(2)$, $X \subset L^2(S^1)$, $(g(\alpha) \cdot x)(\theta) = x(\theta - \alpha)$.

Estimate $x \in X$ and $g_1, \ldots, g_n \in G$ from $n$ measurements of the form

$$y_i = P g_i \cdot x + \epsilon_i, \quad i = 1, \ldots, n$$

where $\epsilon_i$ are independent noise terms.

$P : X \rightarrow Y$ is a sampling operator, $Y$ is the “measurement” space and $X$ is the “object” space.

Quasi MLE

$$\min_{g_1, \ldots, g_n \in G} \sum_{i, j=1}^{n} \|g_i^{-1} \circ y_i - g_j^{-1} \circ y_j\|^2 \iff \min_{g_1, \ldots, g_n \in G} \sum_{i, j=1}^{n} f_{ij}(g_i g_j^{-1})$$
Non-Unique Games over Compact Groups

\[
\min_{g_1, g_2, \ldots, g_n \in G} \sum_{i,j=1}^{n} f_{ij}(g_i g_j^{-1})
\]

- For \( G = \mathbb{Z}_2 \) this encodes Max-Cut, little Grothendieck, and Stochastic Block Model Clustering.
- Max-2-Lin(\( \mathbb{Z}_L \) ) formulation of Unique Games (Khot et al 2005): find \( x_1, \ldots, x_n \in \mathbb{Z}_L \) that satisfy as many difference eqs as possible

\[
x_i - x_j = b_{ij} \mod L, \quad (i,j) \in E
\]

- This corresponds to \( G = \mathbb{Z}_L \) and

\[
f_{ij}(x_i - x_j) = -\delta(x_i - x_j - b_{ij})
\]

- Our games are non-unique in general, and the group is not necessarily finite.
Single Particle Reconstruction using cryo-EM

Schematic drawing of the imaging process:

The cryo-EM problem:
The Resolution Revolution

Werner Kühlbrandt

Precise knowledge of the structure of macromolecules in the cell is essential for understanding how they function. Structures of large macromolecules can now be obtained at near-atomic resolution by averaging thousands of electron microscope images recorded before radiation damage accumulates. This is what Amunts et al. have done in their research article on page 1485 of this issue (1), reporting the structure of the large subunit of the mitochondrial ribosome at 3.2 Å resolution by electron cryo-microscopy (cryo-EM). Together with other recent high-resolution cryo-EM structures (2–4) (see the figure), this achievement heralds the beginning of a new era in molecular biology, where structures at near-atomic resolution are no longer the prerogative of x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy.

Ribosomes are ancient, massive protein-RNA complexes that translate the linear genetic code into three-dimensional proteins.
Projection images $I_i(x, y) = \int_{-\infty}^{\infty} \phi(xR_i^1 + yR_i^2 + zR_i^3) \, dz + \text{"noise"}$.

- $\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ is the electric potential of the molecule.
- Cryo-EM problem: Find $\phi$ and $R_1, \ldots, R_n$ given $I_1, \ldots, I_n$. 

$R_i = \begin{bmatrix} - & R_i^1 & - \\ - & R_i^2 & - \\ - & R_i^3 & - \end{bmatrix} \in SO(3)$
Toy Example
E. coli 50S ribosomal subunit

27,000 particle images provided by Dr. Fred Sigworth, Yale Medical School

3D reconstruction by S, Lanhui Wang, and Jane Zhao
Open source toolbox, publicly available:
http://spr.math.princeton.edu/
Main Algorithmic Challenges

1. **Orientation assignment**

2. **Particle picking**

3. **Heterogeneity (resolving structural variability)**

4. **2D Class averaging (de-noising)**

5. **Symmetry detection**

6. **Motion correction**
Standard procedure is iterative refinement.

Alternating minimization or expectation-maximization, starting from an initial guess $\phi_0$ for the 3-D structure

$$l_i = P(R_i \cdot \phi) + \epsilon_i, \quad i = 1, \ldots, n.$$  

- $R_i \cdot \phi(r) = \phi(R_i^{-1}r)$ is the left group action
- $P$ is integration in the $z$-direction and grid sampling.

Converges to a local optimum, not necessarily the global one.

Model bias is a well-known pitfall

Is “reference free” orientation assignment and reconstruction possible?
Assumptions for today’s talk

- Trivial point-group symmetry
- Homogeneity
Orientation Estimation: Fourier projection-slice theorem

Projection $I_i$

Projection $I_j$

$R_i c_{ij} = (x_{ij}, y_{ij}, 0)^T$

$R_i c_{ij} = R_j c_{ji}$

$\hat{I}_i$

$\hat{I}_j$

3D Fourier space

3D Fourier space

Amit Singer (Princeton University)
Angular Reconstitution (Vainshtein and Goncharov 1986, Van Heel 1987)
Experiments with simulated noisy projections

- Each projection is 129x129 pixels.

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

(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
Define common line as being correctly identified if both radial lines deviate by no more than $10^\circ$ from true directions.

<table>
<thead>
<tr>
<th>$\log_2(\text{SNR})$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.997</td>
</tr>
<tr>
<td>0</td>
<td>0.980</td>
</tr>
<tr>
<td>-1</td>
<td>0.956</td>
</tr>
<tr>
<td>-2</td>
<td>0.890</td>
</tr>
<tr>
<td>-3</td>
<td>0.764</td>
</tr>
<tr>
<td>-4</td>
<td>0.575</td>
</tr>
<tr>
<td>-5</td>
<td>0.345</td>
</tr>
<tr>
<td>-6</td>
<td>0.157</td>
</tr>
<tr>
<td>-7</td>
<td>0.064</td>
</tr>
<tr>
<td>-8</td>
<td>0.028</td>
</tr>
<tr>
<td>-9</td>
<td>0.019</td>
</tr>
</tbody>
</table>
Least Squares Approach

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

- Search space is exponentially large and non-convex.
- Spectral and semidefinite programming relaxations (non-commutative Grothendieck, Max-Cut)
- S, Shkolnisky (*SIAM Journal on Imaging Sciences, 2011*)
The images contain more information than that expressed by optimal pairwise matching of common lines.

Algorithms based on pairwise matching can succeed only at “high” SNR.

(Quasi) Maximum Likelihood: We would like to try all possible rotations $R_1, \ldots, R_n$ and choose the combination for which the agreement on the common lines (implied by the rotations) as observed in the images is maximal.

Computationally intractable: exponentially large search space, complicated cost function.
Quasi MLE

Common line equation: \( R_i c_{ij} = R_j c_{ji} = \frac{R_i e_3 \times R_j e_3}{\| R_i e_3 \times R_j e_3 \|} \) with \( e_3 = (0, 0, 1)^T \).

\[
\begin{align*}
    c_{ij} &= R_i^{-1} \frac{R_i e_3 \times R_j e_3}{\| R_i e_3 \times R_j e_3 \|} = \frac{e_3 \times R_i^{-1} R_j e_3}{\| e_3 \times R_i^{-1} R_j e_3 \|} \\
    c_{ji} &= R_j^{-1} \frac{R_i e_3 \times R_j e_3}{\| R_i e_3 \times R_j e_3 \|} = \frac{R_j^{-1} R_i e_3 \times e_3}{\| R_i^{-1} R_i e_3 \times e_3 \|}
\end{align*}
\]

Quasi MLE

\[
\min_{R_1, \ldots, R_n \in SO(3)} \sum_{i,j=1}^n \| \hat{I}_i(\cdot, c_{ij}) - \hat{I}_j(\cdot, c_{ji}) \|^2 \iff \min_{g_1, \ldots, g_n \in G} \sum_{i,j=1}^n f_{ij}(g_i g_j^{-1})
\]
Fourier transform over $G$

- Recall for $G = SO(2)$

$$f(\alpha) = \sum_{k=-\infty}^{\infty} \hat{f}(k)e^{ik\alpha}$$

$$\hat{f}(k) = \frac{1}{2\pi} \int_{0}^{2\pi} f(\alpha)e^{-ik\alpha} \, d\alpha$$

- In general, for a compact group $G$

$$f(g) = \sum_{k=0}^{\infty} d_k \text{Tr} \left[ \hat{f}(k)\rho_k(g) \right]$$

$$\hat{f}(k) = \int_{G} f(g)\rho_k(g)^* \, dg$$

- Here
  - $\rho_k$ are the unitary irreducible representations of $G$
  - $d_k$ is the dimension of the representation $\rho_k$
    (e.g., $d_k = 1$ for $SO(2)$, $d_k = 2k + 1$ for $SO(3)$)
  - $dg$ is the Haar measure on $G$
Linearization of the cost function

- Introduce matrix variables

\[ X_{ij}^{(k)} = \rho_k (g_i g_j^{-1}) \]

- Fourier expansion of \( f_{ij} \)

\[ f_{ij}(g) = \sum_{k=0}^{\infty} d_k \text{Tr} \left[ \hat{f}_{ij}(k) \rho_k(g) \right] \]

- Linear cost function

\[ f(g_1, \ldots, g_n) = \sum_{i,j=1}^{n} f_{ij}(g_i g_j^{-1}) = \sum_{i,j=1}^{n} \sum_{k=0}^{\infty} d_k \text{Tr} \left[ \hat{f}_{ij}(k) X_{ij}^{(k)} \right] \]
Constraints on the variables $X_{ij}^{(k)} = \rho_k(g_i g_j^{-1})$

1. $X^{(k)} \succeq 0$
2. $X_{ii}^{(k)} = I_{d_k}$, for $i = 1, \ldots, n$
3. $\text{rank}(X^{(k)}) = d_k$

$$X_{ij}^{(k)} = \rho_k(g_i g_j^{-1}) = \rho_k(g_i) \rho_k(g_j^{-1}) = \rho_k(g_i) \rho_k(g_j)^*$$

$$X^{(k)} = \begin{bmatrix}
\rho_k(g_1) \\
\rho_k(g_2) \\
\vdots \\
\rho_k(g_n)
\end{bmatrix}
\begin{bmatrix}
\rho_k(g_1)^* & \rho_k(g_2)^* & \cdots & \rho_k(g_n)^*
\end{bmatrix}$$

- We drop the non-convex rank constraint.
- The relaxation is too loose, as we can have $X_{ij}^{(k)} = 0$ (for $i \neq j$).
- Even with the rank constraint, nothing ensures that $X_{ij}^{(k)}$ and $X_{ij}^{(k')} = X_{ij}^{(k')} \rho_k(g_i g_j^{-1})$ correspond to the same group element $g_i g_j^{-1}$.  

Amit Singer (Princeton University)
May 2015 24 / 30
Additional constraints on $X_{ij}^{(k)} = \rho_k(g_ig_j^{-1})$

- The delta function for $G = SO(2)$
  $$
  \delta(\alpha) = \sum_{k=-\infty}^{\infty} e^{ik\alpha}
  $$

- Shifting the delta function to $\alpha_i - \alpha_j$
  $$
  \delta(\alpha - (\alpha_i - \alpha_j)) = \sum_{k=-\infty}^{\infty} e^{ik\alpha} e^{-ik(\alpha_i - \alpha_j)} = \sum_{k=-\infty}^{\infty} e^{ik\alpha} X_{ij}^{(k)*}
  $$

- The delta function is non-negative and integrates to 1:
  $$
  \sum_{k=-\infty}^{\infty} e^{ik\alpha} X_{ij}^{(k)*} \geq 0, \quad \forall \alpha \in [0, 2\pi)
  $$
  $$
  \frac{1}{2\pi} \int_{0}^{2\pi} \sum_{k=-\infty}^{\infty} e^{ik\alpha} X_{ij}^{(k)*} d\alpha = X_{ij}^{(0)*} = 1
  $$
In practice, we cannot use infinite number of representations to compose the delta function.

Simple truncation leads to the Dirichlet kernel which changes sign

\[ D_m(\alpha) = \sum_{k=-m}^{m} e^{ik\alpha} \]

This is also the source for the Gibbs phenomenon and the non-uniform convergence of the Fourier series.

The Fejér kernel is non-negative and leads to uniform convergence

\[ F_m(\alpha) = \frac{1}{m} \sum_{k=0}^{m-1} D_k(\alpha) = \sum_{k=-m}^{m} \left( 1 - \frac{|k|}{m} \right) e^{ik\alpha} \]

The Fejér kernel is the first order Cesàro mean of the Dirichlet kernel.
Finite truncation via Fejér-Riesz factorization

- Non-negativity constraints over $SO(2)$

$$
\sum_{k=-m}^{m} \left(1 - \frac{|k|}{m}\right) e^{i k \alpha} X_{ij}^{(k)*} \geq 0, \quad \forall \alpha \in [0, 2\pi)
$$

- Fejér-Riesz: $P$ is a non-negative trigonometric polynomial over the circle, i.e. $P(e^{i \alpha}) \geq 0 \ \forall \alpha \in [0, 2\pi)$ iff $P(e^{i \alpha}) = |Q(e^{i \alpha})|^2$ for some polynomial $Q$.

- Leads to semidefinite constraints on $\{X_{ij}^{(k)}\}_k$ for each $i,j$.

- Similar non-negativity constraints hold for general $G$ using the delta function over $G$

$$
\delta(g) = \sum_{k=0}^{\infty} d_k \text{Tr} [\rho_k(g)]
$$

- For example, Fejér proved that for $SO(3)$ the second order Cesàro mean of the Dirichlet kernel is non-negative.
Tightness of the semidefinite program

- We solve an SDP for the matrices $X^{(0)}, \ldots , X^{(m)}$.
- Numerically, the solution of the SDP has the desired ranks up to a certain level of noise (w.h.p).
- In other words, even though the search-space is exponentially large, we typically find the MLE in polynomial time.
- This is a viable alternative to heuristic methods such as EM and alternating minimization.
- The SDP gives a certificate whenever it finds the MLE.
Final Remarks

- Loss of handedness ambiguity in cryo-EM: If $g_1, \ldots, g_n \in SO(3)$ is the solution, then so is $Jg_1 J^{-1}, \ldots, Jg_n J^{-1}$ for $J = \text{diag}(-1, -1, 1)$.

- Define $X_{ij}^{(k)} = \frac{1}{2} \left[ \rho_k(g_i g_j^{-1}) + \rho_k(Jg_i g_j^{-1} J^{-1}) \right]$

- Splits the representation: $2k + 1 = d_k = k + (k + 1)$, reduced computation

- Point group symmetry (cyclic, dihedral, etc.): reduces the dimension of the representation (invariant polynomials)

- Translations and rotations simultaneously: $SE(3)$ is a non-compact group, but we can map it to $SO(4)$. 
