

A Methodology for Deblurring and Recovering Conformational States of Biomolecular Complexes from Single Particle Electron Microscopy

Bijan Afsari and Gregory S. Chirikjian

The Johns Hopkins University

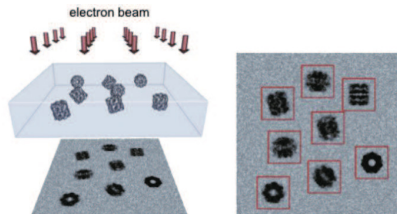
bijan@cis.jhu.edu

October 30, 2015

Introduction and Motivation

- Cryo-Electron Microscopy (EM) is becoming very popular in structural biology as a favorite imaging modality to glean information about large biomolecular complexes in their native states.
- An example of 3-dimensional EM.
- Its resolution is lower than crystallography but alleviates the need for crystallization which is a lengthy and (in some cases) undesirable process.
- Challenges: Extremely high noise, 3D volume reconstruction is very complicated and complexes can be in different states within the sample.
- We introduce a methodology to (possibly) deal with these difficulties in certain applications.
- Remark on the terminology: “single particle” means “not crystallography.” Cryo-EM is a prominent example of single particle EM.

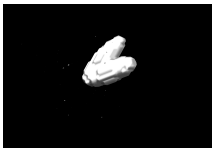
Basics of Cryo-EM



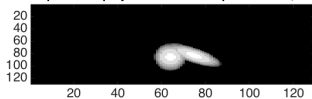
(b) Cryo-EM (from stat.uchicago.edu)

- Copies of a complex lie at **random orientation and positions** in a thin frozen sample.
- Noise is **extremely high** (e.g., $SNR = 1/100$ is common!)
- Thousands of images (projections) are classified (clustered) to similar classes all coming from similar views.
- Images within a class are aligned and averaged to reduce noise.
- As opposed to standard tomography, projection angles are **unknown**.
- Using the “method of common lines” the unknown projection angles can be found and standard “back-projection method” is used to reconstruct the 3D density map.

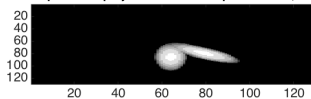
Blurring Effects



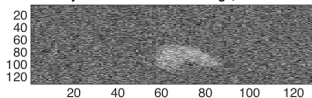
Example of EM projection for two-ellipsoid model, $\theta=30^\circ$



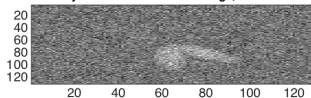
Example of EM projection for two-ellipsoid model, $\theta=45^\circ$



Noisy version of the above image, SNR = 1/10



Noisy version of the above image, SNR = 1/10



- The final 3D structure can be blurred due to high noise and copies of the complex being at different conformational states (conformational blurring).
- Both lead to errors in classification, etc.
- In general, these effects are very complicated to model.

Blurring Effects (cont'ed)

- In general, these effects are very complicated to model.
- We use (**unknown**) $SE(3)$ -kernels to model such blurring effects.
- $SE(3)$: special Linear/Euclidean group, Lie group of rigid body motions in the space.
- 3D density map $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}$ and $SE(3)$ -kernel $f : SE(3) \rightarrow \mathbb{R}$:

$$\tilde{\rho}(\mathbf{r}) = (f \star \rho)(\mathbf{r}) := \int_{SE(3)} f(g) \rho(g^{-1} \cdot \mathbf{r}) dg, \quad (1)$$

where

$$\mathbb{R}^3 \ni \mathbf{r} \mapsto g \cdot \mathbf{r} = R \mathbf{r} + t, \quad (2)$$

is the standard $SE(3)$ action and $(R, t) \in SO(3) \times \mathbb{R}^3$ is the representation of $g \in SE(3)$.

- Group action version of the **convolution** theorem for the prob. density of the **sum** of two random variables.
- As an $SE(3)$ -density: mean and covariance for f can be defined.
- Gaussian $SE(3)$ kernel is only determined by the covariance and mean (to be revisited).

The Case of Multiunit Large Complexes

- In general, deconvolution/deblurring is highly ill-posed.
- We consider a scenario in which the complex has N subunits of known shape.
- Subunits (modeled as rigid bodies) maybe known using crystallography and only their **relative positions** ($\in SE(3)$) maybe unknown and of interest (e.g., relative positions may correlate with the biological function of the complex).

$$\rho(\mathbf{r}) = \sum_{i=1}^N \rho_i(\mathbf{r}), \quad (3)$$

where $\rho_i : \mathbb{R}^3 \rightarrow \mathbb{R}$ is the 3D density of subunit i .

- The blurred 3D reconstructed density:

$$\tilde{\rho}(\mathbf{r}) = \sum_{i=1}^N ((k * f_i) * \rho_i)(\mathbf{r}), \quad (4)$$

where f_i is motional blurring kernel w.r.t. subunit 1 and k is reconstruction blurring kernel.

- Here $*$ is the standard $SE(3)$ convolution

$$(k * f_i)(g) := \int_{SE(3)} k(h) f_i(h^{-1} \circ g) dh$$

- We will treat $k * f_i$ as a single $SE(3)$ kernel.

Blurring a Single Subunit

$$\mathbf{y} = \mathbf{R}\mathbf{r} + \mathbf{t}, \quad \mathbf{r} \in \mathbb{R}^3, \quad \mathbb{E}\{\mathbf{r}\} = \mathbf{0}, \quad \mathbb{E}\{\mathbf{r}\mathbf{r}^\top\} = \mathbf{C}_r, \quad \mathbf{g} = \left[\begin{array}{c|c} \mathbf{R} & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array} \right] \in SE(3), \quad (5)$$

- If \mathbf{r} is Euclidean Gaussian $\sim \mathcal{N}(\mathbf{0}, \mathbf{C}_r)$ and \mathbf{g} is $SE(3)$ Gaussian, what is the density of \mathbf{y} ?
- We use this simple model to model each subunit as an ellipsoid under an $SE(3)$ kernel.
- Rigid body modeling (e.g., with ellipsoids) is common in structural biology.
- The goal is to find the unknown $SE(3)$ kernel from the observed \mathbf{y} and known \mathbf{C}_r .

Parameterization of Gaussian $SE(3)$ Kernels

- Mean: $\mathbf{g} \sim f$ a prob. density on $SE(3)$ its (algebraic or bi-invariant) mean $\mu_{\mathbf{g}}$

$$\mathbb{E}\{\log(\mu_{\mathbf{g}}^{-1}\mathbf{g})\} = \int_{SE(3)} \log(\mu_{\mathbf{g}}^{-1}\mathbf{g})f(\mathbf{g})d\mathbf{g} = 0 \quad (6)$$

- The associated covariance $\Sigma_{\mathbf{g}}$ matrix (6×6)

$$\Sigma_{\mathbf{g}} := \mathbb{E}\{\text{vec}(\Omega_{\mathbf{g}})\text{vec}(\Omega_{\mathbf{g}})^{\top}\} = \int_{SE(3)} \text{vec}(\Omega_{\mathbf{g}})\text{vec}(\Omega_{\mathbf{g}})^{\top} f(\mathbf{g})d\mathbf{g} \quad (7)$$

where $\text{vec} : \mathfrak{se}(3) \rightarrow \mathbb{R}^6$ is an isomorphism between $\mathfrak{se}(3)$ and \mathbb{R}^6 and $\Omega_{\mathbf{g}} = \log(\mathbf{g}\mu_{\mathbf{g}}^{-1})$.

- Total number of parameters: $6 + 21 = 27$.

Matching the Moments

$$\mathbf{y} = \mathbf{R}\mathbf{r} + \mathbf{t}, \quad \mathbf{r} \in \mathbb{R}^3, \quad \mathbb{E}\{\mathbf{r}\} = \mathbf{0}, \quad \mathbb{E}\{\mathbf{r}\mathbf{r}^\top\} = \mathbf{C}_r, \quad \mathbf{g} = \left[\begin{array}{c|c} \mathbf{R} & \mathbf{t} \\ \hline \mathbf{0} & 1 \end{array} \right] \in SE(3), \quad (8)$$

- If \mathbf{g} is with mean $\mu_g = \left[\begin{array}{c|c} \mu_R & \mu_t \\ \hline 0 & 1 \end{array} \right]$ and covariance Σ_g then

$$\mathbb{E}\{\mathbf{y}\} = \mathbb{E}\{\mathbf{t}\} \stackrel{2nd}{=} \left(\mathbf{I} + \frac{1}{2} \mathbb{E}\{\Omega_R^2\} \right) \mu_t \quad (9a)$$

$$\mathbf{C}_y = \mathbb{E}\{\mathbf{R}\mathbf{C}_r\mathbf{R}^\top\} + \mathbf{C}_t \stackrel{2nd}{=} \tilde{\mathbf{C}}_r + \mathbb{E}\{\Omega_R \tilde{\mathbf{C}}_r \Omega_R^\top\} + \frac{1}{2} \mathbb{E}\{\Omega_R^2\} \tilde{\mathbf{C}}_r + \frac{1}{2} \tilde{\mathbf{C}}_r \mathbb{E}\{\Omega_R^2\} + \mathbb{E}\{\Omega_R \mu_t \mu_t^\top \Omega_R^\top\} + \mathbb{E}\{\omega_t \omega_t^\top\} \quad (9b)$$

where $\tilde{\mathbf{C}}_r = \mu_R \mathbf{C}_r \mu_R^\top$ and the expectations of quantities quadratic in Ω_R and ω_t can be expressed in terms of the $SE(3)$ covariance of \mathbf{g} , Σ_g .

- In these equations $\mathbb{E}\{\mathbf{y}\}$ and \mathbf{C}_y can be measured and \mathbf{C}_r is known and μ_g and Σ_g are unknown.
- Still underdetermined!

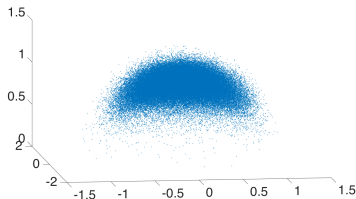
A Simplified Model: Isotropic Blurring Kernel

- At the **Lie algebra** level we assume that all rotations and translations are statistically **independent** (diagonal covariance matrix).
- We assume that all rotational variances are σ_R^2 and all translations variances are σ_t^2 .
- Then the equations simplify as:

$$\mathbb{E}\{\mathbf{y}\} = \mathbb{E}\{\mathbf{t}\} \stackrel{2nd}{=} (1 - \sigma_R^2)\mu_t \quad (10a)$$

$$C_y \stackrel{2nd}{=} \mu_R C_r \mu_R^\top + \sigma_R^2(\text{tr}(C_r)I_3 - 3\mu_R C_r \mu_R^\top) + \sigma_R^2(\|\mu_t\|^2 I_3 - \mu_t \mu_t^\top) + \sigma_t^2 I_3, \quad (10b)$$

- Still nonlinear but not severely!
- Still can be somewhat ill-posed: 9 equations and 8 variables.
- The effect isotropic blurring (banana shape density):



Solving the Inverse Problem and an Example

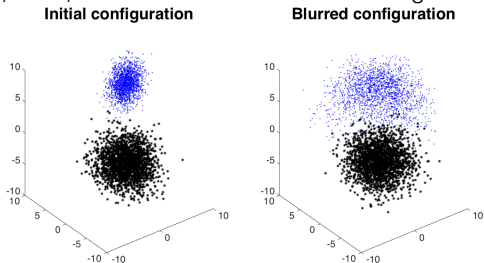
$$F(\mu_R, \sigma_R^2, \sigma_t^2) = \|\mu_R C_r \mu_R^\top + \sigma_R^2 (\text{tr}(C_r) I_3 - 3\mu_R C_r \mu_R^\top) + \sigma_R^2 (\|\mu_t\|^2 I_3 - \mu_t \mu_t^\top) + \sigma_t^2 I_3 - C_y\|_F^2, \quad (11)$$

- We find μ_t from the first eq. assuming σ_R^2 and solve the regularized minimization

$$\min_{\mu_R \in SO(3), \sigma_R^2, \sigma_t^2} F_r(\mu_R, \sigma_R^2, \sigma_t^2; \lambda_R, \lambda_t) \quad (12)$$

where $F_r(\mu_R, \sigma_R^2, \sigma_t^2; \lambda_R, \lambda_t) = F(\mu_R, \sigma_R^2, \sigma_t^2) + \lambda_R (\sigma_R^2)^2 + \lambda_t (\sigma_t^2)^2$ and $\lambda_R, \lambda_t > 0$ are small regularization weights. Repeat till convergence.

- Example: Blurring two bodies with kernels $(\mu_i, \sigma_{R_i}^2, \sigma_{t_i}^2)$.
- Error in means μ_i and t_i are small but for the variances is high!



Conclusions

- In certain applications in structural biology one has prior information about the subunits of a complex under study and the relative motion (position) between subunits may be of interest.
- We gave a plausibility study as to how this information can be used in a inverse problem setting to remove the blurring effects in cryo-EM.
- The blurring can be due to high noise and conformational variations within the copies of the complex in the sample imaged.
- The method was based on parameterization of $SE(3)$ kernels and 2nd order moment matching.
- Higher order moments and information from other imaging modalities such as Small Angle X-ray Scattering (SAXS) can be used to further regularize the inverse problem.

Acknowledgment

National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM113240.

Thank You!