CryoEM Report 08/08 Derivation of EM algorithm

August 8, 2018

Definition of projection matrix P^ϕ

We assume the 3D structure $\boldsymbol{f}$ in Fourier space is on a Cartesian grid based space:

$$\Omega = \{\boldsymbol{n} \cdot w_0 \ : \ \boldsymbol{n} = (n_1, n_2, n_3) \in Z^3, -\frac{N}{2} \leq n_i < \frac{N}{2}, \ i = 1, 2, 3, \ w_0 > 0\}$$

where w_0 is the grid space. So $\Omega = \{\boldsymbol{n} \cdot w_0\}$ is large enough to cover the support of $\boldsymbol{f}$. In other words, we pre-assume a discrete 3D density map with $N \times N \times N$ arrays. The density value of each voxel is unknown. This 3D grid size N is consistent with 2D image size N. The 3D grid space w_0 is also consistent with 2D grid space Δ. Note grid size and grid space have different definition.

Extend the coordinates $\boldsymbol{n} \cdot w_0$ of this 3D density map to an ordered series $\{(x_i, y_i, z_i) : for \ i = 1, 2, \cdots, \ L, \ for \ L = N \times N \times N\}$. Then continuous 3D structure $\boldsymbol{f}$ can be approximated by a linear combination of a finited set of known and fixed basis function $b(\cdot)$,

$$\vec{f}(x, y, z) = \sum_{l=1}^{L} c_l b(x - x_l, y - y_l, z - z_l) \tag{1}$$

where $x, y, z \in R$, $-\frac{N}{2} w_0 \leq x, y, z \leq \frac{N}{2} w_0$. $\{c_l\}$ are unknown coefficients we want to estimate. The conventional choice of basis function $b(.)$ would be voxel basis function $b(x, y, z)$, defined as

$$b(x, y, z) = \begin{cases} 1 & for \ |x|, |y|, |z| \leq \frac{w_0}{2} \\ 0 & otherwise \end{cases}$$

Compare the equation(1) with the model in RELION paper,

$$X_{ij} = CTF_{ij} \sum_{l=1}^{L} P_j \sum_{l=1}^{L} P_{jl}^{\phi} V_l + N_{ij} \tag{2}$$

Consider the coordinate of X_{ij} is known as $(\omega_{ij,0}, \omega_{ij,1})$. It is extended to 3D vector $\boldsymbol{\omega_{ij}} = (\omega_{ij,0}, \omega_{ij,1}, 0)$ in 3D space. The coordinate of corresponding points in 3D Fourier space would be $\boldsymbol{R_i^{-1}}\boldsymbol{\omega_{ij}}$, where $\boldsymbol{R_i^{-1}}$is the rotation matrix if given sampling direction $\boldsymbol{\phi_i}$. Then we would rewrite model (2) as

$$X_{ij}(\boldsymbol{\omega}) = CTF_{ij}(\boldsymbol{\omega}) \sum_{l=1}^{L} c_l b(\boldsymbol{R_i^{-1}}\boldsymbol{\omega_{ij}} - (x_l, y_l, z_l)) + N_{ij} \tag{3}$$

Specifically, we have the definition of projection matrix $P^{\boldsymbol{\phi}}$ as

$$P_{jl}^{\boldsymbol{\phi}} = b(\boldsymbol{R_i^{-1}}\boldsymbol{\omega_{ij}} - (x_l, y_l, z_l))$$

The difference between $P_{jl}^{\boldsymbol{\phi}}$ and $P_{jk}^{\boldsymbol{\phi}}$ would be (x_l, y_l, z_l) and (x_k, y_k, z_k) in the arguments. By the definition of $b(.)$, grid space and $|x_l - x_k| + |y_l - y_k| + |z_l - z_k| \geq 1$, for any $l \neq k$, we have

$$P_{jl}^{\boldsymbol{\phi}} P_{jk}^{\boldsymbol{\phi}} = \begin{cases} 0 & if \ l \neq k \\ P_{jl}^{\boldsymbol{\phi}} & if \ l = k \end{cases}$$

That means the value of each element of projection matrix $P^{\boldsymbol{\phi}}$, depends on sampling direction $\boldsymbol{\phi_i}$, fixed grid points $\{(x_l, y_l, z_l) : l = 1, 2, \cdots, L\}$and coordinate $\boldsymbol{\omega_{ij}}$.

Assumptions of distribution

We have following assumptions.

$$X_{ij} \sim N\left(CTF_{ij} \sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l, \ \sigma_{ij}^2\right)$$

$$\Rightarrow \qquad P(\boldsymbol{X_i} | \ \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{CTF}) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left(\frac{(X_{ij} - \mathrm{CTF}_{ij} \sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l)^2}{-2\sigma_{ij}^2}\right)$$

where $\boldsymbol{\theta} = (\boldsymbol{V}, \boldsymbol{\tau}, \boldsymbol{\Sigma})$ is the set of unkown model parameters. $\boldsymbol{\Sigma}$ is the matrix generated by σ_{ij}.
Marginal likelihood:

$$P(\boldsymbol{X} \mid \boldsymbol{\theta}, \ \mathbf{CTF}) = \prod_{i=1}^{N} \int_{\boldsymbol{\phi}} P(\boldsymbol{X_i} \mid \boldsymbol{\phi}, \boldsymbol{\theta}, \mathbf{CTF}) P(\boldsymbol{\phi} \mid \boldsymbol{\theta}, \mathbf{CTF}) d\boldsymbol{\phi}$$

Above will integrate oritation $\boldsymbol{\phi}$ out. And we will utilize it in the E-step.
Prior of $\boldsymbol{\theta}$:

$$P(\boldsymbol{\theta}) = \prod_{l=1}^{L} \frac{1}{\sqrt{2\pi\tau_l^2}} \exp\left(\frac{V_l^2}{-2\tau_l^2}\right)$$

Prior of $\boldsymbol{\phi}$:
We denote it be $f_\Phi(\boldsymbol{\phi})$. This distribution of the orietations include three Euler angle parameters and two x-y offest parameters. It follows the HEALPix framework, which is an approximation of uniform sampling of the sphere.

EM algorithm:

We start from denoting the complete data set to be

$$(\boldsymbol{X_i},\ \Phi_i),\ \ for\ i = 1, 2, \cdots, N$$

Φ_i is the hidden variable, which represents the orietation of image $\boldsymbol{X_i}$. Our goal is to find $\boldsymbol{\theta}$ to maximize posterior $P(\boldsymbol{\theta} \mid \boldsymbol{X}, \mathbf{CTF})$ In the following contex we will write $P(\boldsymbol{\theta} \mid \boldsymbol{X})$. The dependence on CTF will be suppressed since all distributions depend on this observed covariate. Also the suppression will be applied to all conditional distributions.

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta}\in\Omega} logP(\boldsymbol{\theta} \mid \boldsymbol{X}) = \arg\max_{\boldsymbol{\theta}\in\Omega} logP(\boldsymbol{X} \mid \boldsymbol{\theta}) + logP(\boldsymbol{\theta}) \tag{1}$$

For convenience, denote

$$\begin{aligned} f(\boldsymbol{X_i} \mid \boldsymbol{\phi}, \boldsymbol{\theta}) &= P(\boldsymbol{X_i} \mid \boldsymbol{\theta}, \boldsymbol{\phi}) \\ f_\Phi(\boldsymbol{\phi}) &= P(\boldsymbol{\phi} \mid \boldsymbol{\theta}) = P(\boldsymbol{\phi}) \end{aligned} \tag{2}$$

Eq.(2) holds beacause the orietation $\boldsymbol{\phi}$ of each 3D structure is not affected by $\boldsymbol{\theta}$ and constant **CTF**.

We write ritht side of Eq.(1) as function of 3 unknown parameters are $\boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\tau}$.

$$\begin{aligned} L(\boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\tau}) &= logP(\boldsymbol{X} \mid \boldsymbol{\theta}) + logP(\boldsymbol{\theta}) \\ &= \sum_{i=1}^{N} P(\boldsymbol{X_i} \mid \boldsymbol{\theta}) + logP(\boldsymbol{\theta}) \\ &= \sum_{i=1}^{N} log \int_{\boldsymbol{\phi}} P(\boldsymbol{X_i} \mid \boldsymbol{\phi}, \boldsymbol{\theta}) P(\boldsymbol{\phi} \mid \boldsymbol{\theta}) d\boldsymbol{\phi} + logP(\boldsymbol{\theta}) \\ &= \sum_{i=1}^{N} log \int_{\boldsymbol{\phi}} f_{\boldsymbol{X} \mid \Phi, \ominus, \mathbf{CTF}}(\boldsymbol{X_i} \mid \boldsymbol{\phi}, \boldsymbol{\theta}) f_\Phi(\boldsymbol{\phi} \mid \boldsymbol{\theta}) d\boldsymbol{\phi} + logP(\boldsymbol{\theta}) \end{aligned}$$

Thus, we turn Eq.(1) into following problem

$$\begin{aligned} &maximize\ L(\boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\tau}) \\ &\quad subject\ to\ \boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\tau} \end{aligned} \tag{3}$$

Before we start EM algorithm, we have some denotions first.

$$f(\boldsymbol{\phi} \mid \boldsymbol{X}_i, \boldsymbol{\theta}) = \frac{f(\boldsymbol{X}_i \mid \boldsymbol{\phi}, \boldsymbol{\theta}) f_{\boldsymbol{\phi}}(\boldsymbol{\phi})}{\int f(\boldsymbol{X}_i \mid \boldsymbol{\theta}, \boldsymbol{\phi}) f_{\boldsymbol{\phi}}(\boldsymbol{\phi}) d\boldsymbol{\phi}} \tag{4}$$

We will replace unknown parameters $\boldsymbol{\theta}$, f_Φ in Eq.(4) by current guess. Denote

$$\theta^{old} = (\boldsymbol{V}^{old}, \boldsymbol{\tau}^{old}, \boldsymbol{\Sigma}^{\boldsymbol{old}})$$

At the same time, denote the posterior distribution of $\boldsymbol{\phi}$ given $\boldsymbol{X}$, $\boldsymbol{\theta}$ as $g(\boldsymbol{\phi},\ \boldsymbol{X_i})$.

$$g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i}) = f_\Phi^{old}(\boldsymbol{\phi} \mid \boldsymbol{X_i}, \boldsymbol{\theta}) \tag{5}$$

Next we use EM algorithm to solve the optimization of (3).

E-step:

Derive the Q-function

$$Q(\boldsymbol{\theta}, f_\Phi \mid \theta^{old}, f_\Phi^{old}) = \sum_{i=1}^{N} \int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i}) \log(f_{\boldsymbol{X} \ \mid \Phi, \ominus, \mathbf{CTF}}(\boldsymbol{X_i} \mid \boldsymbol{\phi}, \boldsymbol{\theta}) f_\Phi(\boldsymbol{\phi})) d\boldsymbol{\phi}$$

M-step:

We can do M-step separately to obtain estimate of $(\boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\tau})$

(1)

Maximization over $\boldsymbol{\Sigma}$

That is to maximize

$$\begin{aligned} L_3(\boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\tau}) &= \sum_{i=1}^{N} \int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i}) \log(f_{\boldsymbol{X} \ \mid \Phi, \ominus, \mathbf{CTF}}(\boldsymbol{X_i} \mid \boldsymbol{\phi}, \boldsymbol{\theta}) d\boldsymbol{\phi} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{J} \int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i}) \{-\frac{(X_{ij} - \mathrm{CTF}_{ij} \sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l)^2}{2\sigma_{ij}^2} - \log\sqrt{2\pi}\sigma_{ij}\} d\boldsymbol{\phi} \end{aligned}$$

Take derivative of L_3 by σ_{ij}

$$\frac{dL_3}{d\sigma_{ij}} = 0$$

$$\Rightarrow \int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\{-\frac{(X_{ij} - \mathrm{CTF}_{ij}\sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l)^2}{\sigma_{ij}^3} - \frac{1}{\sigma_{ij}}\}d\boldsymbol{\phi} = 0$$

$$\sigma_{ij}^2 = \int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X}_i)(X_{ij} - \mathrm{CTF}_{ij}\sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l)^2 d\boldsymbol{\phi} \tag{9}$$

Here we deviate from E-M algorithm. We take current guess V_l^{old} in Eq.(9).

$$(\sigma_{ij}^2)^{new} = \int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})(X_{ij} - \mathrm{CTF}_{ij}\sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l^{old})^2 d\boldsymbol{\phi} \tag{10}$$

(2)

Maximization over $\boldsymbol{V}$

That is to maximize

$$\begin{aligned} L_4(\boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\tau}) &= L_3(\boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\tau}) + \log P(\boldsymbol{\theta}) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\{-\frac{(X_{ij} - \mathrm{CTF}_{ij}\sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l)^2}{2\sigma_{ij}^2} - \log\sqrt{2\pi}\sigma_{ij}\}d\boldsymbol{\phi} \\ &+ \sum_{l=1}^{L}(-\frac{V_l^2}{2\tau_l^2}) - \sum_{l=1}^{L}\log\sqrt{2\pi}\tau_l^2 \end{aligned}$$

Tke derivative of $L_4$by V_l

$$\frac{dL_4}{dV_l} = 0$$

$$\Rightarrow \sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_I})\{-\frac{(X_{ij} - \mathrm{CTF}_{ij}\sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l)}{\sigma_{ij}^2}\times(-\mathrm{CTF}_{ij}P_{jl}^{\boldsymbol{\phi}})\}d\boldsymbol{\phi} - \frac{V_l}{\tau_l^2} = 0$$

$$\sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\{P_{jl}^{\boldsymbol{\phi}}\frac{\mathrm{CTF}_{ij}X_{ij}}{\sigma_{ij}^2}\}d\boldsymbol{\phi} = \frac{V_l}{\tau_l^2} + \sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\frac{\mathrm{CTF}_{ij}^2}{\sigma_{ij}^2}\times P_{jl}^{\boldsymbol{\phi}}(\sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l)d\boldsymbol{\phi}$$

Accordding to the defition of $\boldsymbol{P_j^{\phi}}$

$$P_{jl}^{\boldsymbol{\phi}} P_{jk}^{\boldsymbol{\phi}} = \begin{cases} 0 & if\ l \neq k \\ P_{jl}^{\boldsymbol{\phi}} & if\ l = k \end{cases}$$

$$\Rightarrow \sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\{P_{jl}^{\boldsymbol{\phi}} \frac{CTF_{ij}X_{ij}}{\sigma_{ij}^2}\}d\boldsymbol{\phi} = V_l\{\frac{1}{\tau_l^2} + \sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\frac{\mathrm{CTF}_{ij}^2}{\sigma_{ij}^2} \times P_{jl}^{\boldsymbol{\phi}} d\boldsymbol{\phi}\} \tag{11}$$

Just like in Eq.(10) we replace σ_{ij}^{old} and τ_l^{old} to replace σ_{ij} nd τ_l in Eq.(11), we will have

$$V_l^{new} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\{P_{jl}^{\boldsymbol{\phi}} \frac{\mathrm{CTF}_{ij}X_{ij}}{(\sigma_{ij}^2)^{old}}\}d\boldsymbol{\phi}}{\frac{1}{(\tau_l^2)^{old}} + \sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\frac{\mathrm{CTF}_{ij}^2}{(\sigma_{ij}^2)^{old}} \times P_{jl}^{\boldsymbol{\phi}} d\boldsymbol{\phi}} \tag{12}$$

(3)

Maxizization over $\boldsymbol{\tau}$

We only need to maximize $\log P(\boldsymbol{\theta})$ over $\boldsymbol{\tau}$

$$lopP(\boldsymbol{\theta}) = \sum_{l=1}^{L} -\frac{V_l^2}{2\tau_l^2} - \sum_{l=1}^{L}\log\sqrt{2\pi}\tau_l$$

$$\frac{d\log P(\boldsymbol{\theta})}{d\tau_l} = 0$$

$$\Rightarrow \ -\frac{V_l^2}{\tau_l^3} - \frac{1}{\tau_l} = 0$$

$$(\tau_l^2)^{new} = (V_l^{new})^2 \tag{13}$$

for $l = 1, \cdots, L$

Result

After combining Eq.(5), (8), (10), (12), (13), we have following updating algorithm,

$$V_l^{new} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})\{P_{jl}^{\boldsymbol{\phi}} \frac{\mathrm{CTF}_{ij}X_{ij}}{(\sigma_{ij}^2)^{old}}\}d\boldsymbol{\phi}}{\frac{1}{(\tau_l^2)^{old}} + \sum_{i=1}^{N}\sum_{j=1}^{J}\int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X}_i)\frac{\mathrm{CTF}_{ij}^2}{(\sigma_{ij}^2)^{old}} \times P_{jl}^{\boldsymbol{\phi}} d\boldsymbol{\phi}} \tag{14}$$

$$(\sigma_{ij}^2)^{new} = \int_{\boldsymbol{\phi}} g^{old}(\boldsymbol{\phi}, \boldsymbol{X_i})(X_{ij} - \mathrm{CTF}_{ij}\sum_{l=1}^{L} P_{jl}^{\boldsymbol{\phi}} V_l^{old})^2 d\boldsymbol{\phi} \tag{15}$$

$$(\tau_l^2)^{new} = (V_l^{new})^2 \tag{16}$$

where the update of g^{new}would be

$$\begin{aligned} g^{new}(\boldsymbol{\phi},\ \boldsymbol{X_i}) &= f^{new}(\boldsymbol{\phi}\ |\boldsymbol{X_i}, \boldsymbol{\theta}) \\ &= \frac{f(\boldsymbol{X_i} \mid \boldsymbol{\phi}, \boldsymbol{\theta^{new}}) f_\Phi(\boldsymbol{\phi})}{\int_{\boldsymbol{\phi}} f(\boldsymbol{X_i} \mid \boldsymbol{\phi}, \boldsymbol{\theta}^{new}) f_\Phi(\boldsymbol{\phi}) d\boldsymbol{\phi}} \end{aligned} \quad 17$$

Above all is the derivation of EM algorithm to estimate the 3D map of macro-molecular. Note, (1) it is not the exact EM algorithm, since in Eq.(10) we take curre guess $\boldsymbol{V}^{old}$ rather than $\boldsymbol{V}^{new}$ to update σ_{ij}. And it does not guarantee that $L(\boldsymbol{\Sigma}^{old}, \boldsymbol{V}^{old}) \leq L(\boldsymbol{\Sigma}^{new}, \boldsymbol{V}^{new})$. However, the experiment from (Scheres, 2007) shows it nearly holds in practice. (2) The EM algorithm sometimes will get stuck in the local maximum sometime. That whether result is good depends on the intial value of $\boldsymbol{\theta}$. New approach cryosparc (Ali, John, et al, 2017) used stochastic gradient descent (SGD) to solve this issue. (3) This algorithm needs large computation capability. Updating formula Eq.(17) will take most of the computation burden, because it needs to integral all the orietations. Numerically, we will do summation of all sampling orietations to approximate the integration.