

Image formation model

- Images in Fourie space:

$$\mathcal{X} = \{\mathbb{X}_i\}_1^n = \mathbb{X}_1, \dots, \mathbb{X}_i, \dots, \mathbb{X}_n \in \mathbb{C}^{l^2} \text{ where } l = \text{boxsize}$$

- Volume in Fourie space:

$$\mathbb{V} \in \mathbb{C}^{l^3}$$

- Projection operator:

$$\mathbb{H}_{q_i} \in \mathbb{C}^{l^2 \times l^3}$$

where $q \in \mathbb{Q} := \text{SE}(3)$ describe Euler angles and translation.

$$\mathbb{X}_i = \mathbb{H}_{q_i} \mathbb{V}$$

If we assume CTF effects are linear, it could be describe in \mathbb{H} .

Probability model

- Some important formulae

$$a \in A, b \in B$$

$$P(A) = \sum_{a \in A} P(a)$$

$$P(\{a_i\}_1^n) = P(\mathcal{A}) = \prod_i^n P(a_i)$$

$$\begin{aligned} P(a) &= \sum_{b_i \in B} P(a, b) \\ &= \int_B P(a, b) \, db \end{aligned}$$

$$P(a|b) = \frac{P(a, b)}{P(b)} = \frac{P(a, b)}{\sum_A P(a, b)}$$

We then seek the maximum a posteriori (MAP) estimate of \mathbb{V} by maximizing the following

$$P(\mathbb{V}|\mathcal{X}) = \frac{P(\mathbb{V}, \mathcal{X})}{P(\mathcal{X})} = \frac{P(\mathcal{X}|\mathbb{V})P(\mathbb{V})}{P(\mathcal{X})}$$

and regularized log-likelihood form is following

$$\begin{aligned} \ell(\mathbb{V}|\mathcal{X}) &= \log P(\mathbb{V}|\mathcal{X}) = \log \left(P(\mathcal{X}|\mathbb{V}) \frac{P(\mathbb{V})}{P(\mathcal{X})} \right) \\ &= \log P(\mathcal{X}|\mathbb{V}) + \log \frac{P(\mathbb{V})}{P(\mathcal{X})} \end{aligned}$$

As this is not a computable form about $P(\mathcal{X}|\mathbb{V})$, we use EM-algorithm to explicitly include the projection model.

$$\begin{aligned}
\log P(\mathcal{X}|\mathbb{V}) &= \log \sum_{q \in Q} P(\mathcal{X}, q|\mathbb{V}) \\
&= \log \sum_{q \in Q} p(q) \frac{P(\mathcal{X}, q|\mathbb{V})}{p(q)} \\
(\text{Jensen's inequality}) &\geq \sum_{q \in Q} p(q) \log \frac{P(\mathcal{X}, q|\mathbb{V})}{p(q)} \\
\mathcal{L}(\mathcal{X}|\mathbb{V}) &:= \sum_{q \in Q} p(q) \log \frac{P(\mathcal{X}, q|\mathbb{V})}{p(q)} \\
\log P(\mathcal{X}|\mathbb{V}) - \mathcal{L}(\mathcal{X}|\mathbb{V}) &= \sum_{q \in Q} p(q) \left\{ \log P(\mathcal{X}|\mathbb{V}) - \log \frac{P(\mathcal{X}, q|\mathbb{V})}{p(q)} \right\} \\
&= \sum_{q \in Q} p(q) \left\{ \log p(q) \frac{P(\mathcal{X}|\mathbb{V})}{P(\mathcal{X}, q|\mathbb{V})} \right\} \\
&= \sum_{q \in Q} p(q) \left\{ \log \frac{p(q)}{P(q|\mathcal{X}, \mathbb{V})} \right\} \geq 0 \\
\text{KL}(p|P) &:= \sum_{q \in Q} p(q) \left\{ \log \frac{p(q)}{P(q|\mathcal{X}, \mathbb{V})} \right\} \geq 0
\end{aligned}$$

Then in E-step, we calculate following

$$\hat{P}_j = \underset{p(q)}{\operatorname{argmax}} P(\mathcal{X}|\mathbb{V}_j)$$

If $\hat{P}_j = \Gamma(q) = P(q|\mathcal{X}, \mathbb{V}_j)$, $\text{KL}(p|P) = 0 \Rightarrow \log P(\mathcal{X}|\mathbb{V}_j) = \mathcal{L}(\mathcal{X}|\mathbb{V}_j)$, and $\log P(\mathcal{X}|\mathbb{V}_j)$ don't depends on $p(q)$, its value is constant.

$$\begin{aligned}
\Gamma(q) &= \frac{P(\mathcal{X}, q|\mathbb{V})}{P(\mathcal{X}|\mathbb{V})} = \frac{P(\mathcal{X}, q|\mathbb{V})}{\int_Q P(\mathcal{X}, q|\mathbb{V})} \\
&= \frac{P(\mathcal{X}|q, \mathbb{V})P(q|\mathbb{V})}{\int_Q P(\mathcal{X}|q, \mathbb{V})P(q|\mathbb{V})}
\end{aligned}$$

We can model this form by following

$$\begin{aligned}
P(\mathbb{X}_i|q, \mathbb{V}) &= \frac{C}{\sqrt{\det(\sigma^2)}} \exp\left(-\frac{1}{2}\|\sigma^{-1}(\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V})\|^2\right) \\
\text{where } \sigma &:= \operatorname{diag}\left(\{\sigma_i\}_1^{l^2}\right) \in \mathbb{R}_+^{l^2 \times l^2} \\
P(\mathbb{V}) &= \frac{C}{\sqrt{\det(\tau^2)}} \exp\left(-\frac{1}{2}\|\tau^{-1}\mathbb{V}\|^2\right) \\
\text{where } \tau &:= \operatorname{diag}\left(\{\tau_i\}_1^{l^3}\right) \in \mathbb{R}_+^{l^3 \times l^3}
\end{aligned}$$

The noise parameter σ are often modeled as the resolution-dependent variance.

As $P(q|\mathbb{V}) \approx P(q)$ is a probability for angle and movement, it is intuitively a uniform distribution, but some assumptions may be made to speed up the calculation (i.e. local search, branch and bound).

The prior $P(\mathbb{V})$ expresses how likely that model is given the prior information. Since the protein is made up of atoms, the amplitude in reciprocal space is the sum of these numbers. The phase represents the position of each atom and, assuming that the positions of the atoms in the protein are random, the phase of the whole protein can be regarded as equivalent to a random walk in a two-dimensional plane and can therefore be modelled by a Gaussian distribution with zero mean and resolution-dependent variance τ .

So we summarize E-step.

$$\begin{aligned}\hat{\ell}(\mathbb{V}|\mathcal{X}) &= \sum_{q \in Q} \Gamma(q)_j \log \frac{P(\mathcal{X}, q|\mathbb{V})}{\Gamma(q)_j} + \log \frac{P(\mathbb{V})}{P(\mathcal{X})} \\ &= \sum_{q \in Q} \Gamma(q)_j \log \frac{P(\mathcal{X}|q, \mathbb{V})P(q)}{\Gamma(q)_j} + \log \frac{P(\mathbb{V})}{P(\mathcal{X})}\end{aligned}$$

Next M-step is following

$$\hat{\mathbb{V}}_{j+1} = \operatorname{argmax}_{\mathbb{V}} \hat{\ell}(\mathbb{V}|\mathcal{X})$$

and solve $\nabla_{\mathbb{V}} \hat{\ell}(\mathbb{V}|\mathcal{X}) = 0$ to find \mathbb{V} .

$$\begin{aligned}\nabla_{\mathbb{V}} \hat{\ell}(\mathbb{V}|\mathcal{X}) &= \frac{\partial}{\partial \mathbb{V}} \hat{\ell}(\mathbb{V}|\mathcal{X}) \\ &= \frac{\partial}{\partial \mathbb{V}} \left\{ \sum_{q \in Q} \Gamma(q)_j \log P(\mathcal{X}|q, \mathbb{V}) + \log P(\mathbb{V}) \right\} \\ &= \frac{\partial}{\partial \mathbb{V}} \left\{ - \sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \left\| \sigma^{-1}(\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V}) \right\|^2 - \left\| \tau^{-1} \mathbb{V} \right\|^2 \right\} \\ &= \sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \mathbb{H}_q^{*,i} \sigma^{-2} (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V}) - \tau^{-2} \mathbb{V}\end{aligned}$$

Then, in the Maximization step we solve for the parameters of a \mathbb{V} , which yields the closed-form solution:

$$\mathbb{V}_{j+1} = \left(\sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \mathbb{H}_q^{*,i} \sigma^{-2} \mathbb{H}_q^i + \tau^{-2} \right)^{-1} \left(\sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \mathbb{H}_q^{*,i} \sigma^{-2} \mathbb{X}_i \right) := \mathbb{K}^{-1} \mathbb{B}$$

The $\mathbb{H}^* \sigma^{-2} \mathbb{H}$ in \mathbb{K} means backprojected sigma. Note atleast we want keep 3 volume in VRAM. (reference, backprojection volume, backprojection weight = \mathbb{K})

With same manner, we solve σ_{j+1}, τ_{j+1} .

$$\begin{aligned}
\frac{\partial}{\partial \sigma^2} \hat{\ell}(\mathbb{V}|\mathcal{X}) &= \frac{\partial}{\partial \sigma^2} \sum_{q \in Q} \Gamma(q)_j \log P(\mathcal{X}|q, \mathbb{V}) \\
&= \frac{\partial}{\partial \sigma^2} \left\{ - \sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \left(\frac{1}{2} \|\sigma^{-1}(\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V})\|^2 + \log \det(\sigma^2) \right) \right\} \\
&= \sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \left(\frac{1}{2} \sigma^{-2} (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V}) (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V})^* \sigma^{-2} - \sigma^{-2} \right) \\
&= \sigma^{-2} \sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \left(\frac{1}{2} \sigma^{-2} (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V}) (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V})^* - \mathbb{I} \right) \\
\\
\frac{\partial}{\partial \tau^2} \hat{\ell}(\mathbb{V}|\mathcal{X}) &= \frac{\partial}{\partial \tau^2} \log P(\mathbb{V}) \\
&= \frac{\partial}{\partial \tau^2} \left\{ - \left(\frac{1}{2} \|\tau^{-1} \mathbb{V}\|^2 + \log \det(\tau^2) \right) \right\} \\
&= \frac{1}{2} \tau^{-2} \mathbb{V} \mathbb{V}^* \tau^{-2} - \tau^{-2} \\
\\
\sigma_{j+1}^2 &= \frac{1}{2n} \left(\sum_i^n \sum_{q \in Q} \Gamma_i(q)_j (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V}) (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V})^* \right) \\
\tau_{j+1}^2 &= \frac{1}{2} \mathbb{V} \mathbb{V}^*
\end{aligned}$$

Summary

$$\begin{aligned}
P(\mathbb{X}_i|q, \mathbb{V}) &= \frac{C}{\sqrt{\det(\sigma^2)}} \exp \left(-\frac{1}{2} \|\sigma^{-1}(\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V})\|^2 \right) \\
\Gamma_i(q) &= \frac{P(\mathbb{X}_i|q, \mathbb{V}) P(q|\mathbb{V})}{\int_Q P(\mathbb{X}_i|q, \mathbb{V}) P(q|\mathbb{V})} \\
\\
\mathbb{V}_{j+1} &= \left(\sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \mathbb{H}_q^{*,i} \sigma^{-2} \mathbb{H}_q^i + \tau^{-2} \right)^{-1} \left(\sum_i^n \sum_{q \in Q} \Gamma_i(q)_j \mathbb{H}_q^{*,i} \sigma^{-2} \mathbb{X}_i \right) \\
\sigma_{j+1}^2 &= \frac{1}{2n} \left(\sum_i^n \sum_{q \in Q} \Gamma_i(q)_j (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V}) (\mathbb{X}_i - \mathbb{H}_q^i \mathbb{V})^* \right) \\
\tau_{j+1}^2 &= \frac{1}{2} \mathbb{V} \mathbb{V}^*
\end{aligned}$$