Noise-resistant tomographic reconstruction under unknown angles

EPFL, under the guidance of Professor Victor M. Panaretos

December 10, 2018

1 Introduction

Reconstructing the structure of an object from its tomographic projections is a fundamental research problem that arises in diverse fields, such as medical imaging and reconstruction in cryo-electron-microscopy (referred to hereafter as 'Cryo-EM'). If the viewing orientations are known *a priori*, standard algorithms such as the Filtered Backprojection Algorithm or variants thereof for different acquisition geometries [1], can be used to reconstruct the image. However, there are many scenarios where this may not be the case. One such example is Cryo-EM where the objective is to determine the structure of a molecule from its projections which essentially appear in various unknown orientations [2, 3]. Other examples include insect tomography [4], or tomography of objects performing unknown rigid motion [5], which is equivalent to performing a tomographic reconstruction on a fixed object, with the viewing directions being unknown. Uncertainty in viewing angles may also occur due to patient motion in medical imaging, even though to a lower degree. To tackle these problems we explore two algorithms. The first of these algorithms is RELION, which stands for REgularized LIkelihood OptimizatioN [6].

2 RELION Algorithm

2.1 Description of the Bayesian Optimization

RELION interprets the problem as a Bayesian interpretation problem, where smoothness in the reconstructed density is imposed through a Gaussian prior in the Fourier domain. The reconstruction problem is formulated as finding the model with parameter set Θ that has the highest probability of being the correct one in the light of both the observed data X and the prior information Y. According to Bayes' law, this so-called posterior distribution factorizes into two components:

$$P(\Theta|X,Y) \propto P(X|\Theta,Y)P(\Theta|Y) \tag{1}$$

where the *likelihood* $P(X|\Theta,Y)$ quantifies the probability of observing the data given the model, and the *prior* $P(\Theta|Y)$ expresses how likely that model is given the prior information.

For this problem, we use a linear image formation model in the Fourier space:

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

where,

- X_{ij} is the j^{th} component, with $j = 1, 2, .., J$ of the one-dimensional (1D) Fourier transform of the i^{th} experimental projection, with $i = 1, 2.., N$.
- CTF_{ij} is the j^{th} component of the contrast transfer function for the i^{th} image.
- V_l is the l^{th} component, with $l = 1, 2, .., L$ of the 2D Fourier transform V of the underlying structure in the data set.
- $\mathbf{P}^{\phi}$ is a $J \times L$ matrix of elements P_{jl}^{ϕ}. The operation $\sum_{l=1}^{L} P_{jl}^{\phi} V_l$ for all j extracts a slice out of the 3D Fourier transform of the underlying structure. Similarly, the operation $\sum_{j=1}^{J} P^{\phi_{lj}^T} X_{ij}$ for all l places the 1D Fourier transform of an experimental projection back into the 2D transform.

Or we can equivalently write,

$$V_l = \sum_{j=1}^{J} \frac{P^{\phi_{lj}^T}}{CTF_{ij}}(X_{ij} - N_{ij}) \tag{3}$$

which can be thought of as the back-projection operation.

We assume that all the noise components N_{ij} are independent and Gaussian distributed. The variance σ_{ij}^2 of these noise components is unknown and must be estimated from the data. The assumption of independence in the noise allows the probability of observing an image given its orientation and the model to be calculated as a multiplication of Gaussians over all its Fourier components:

$$P(X_i|\phi, \Theta, Y) = \prod_{j=1}^{J} \frac{1}{2\pi\sigma_{ij}^2} exp\left(\frac{|X_{ij} - CTF_{ij}\sum_{l=1}^{L} P_{jl}^{\phi} V_l|^2}{-2\sigma_{ij}^2}\right) \tag{4}$$

The corresponding marginal likelihood of observing the entire data set X is then given by:

$$P(X|\Theta, Y) = \prod_{i=1}^{N} \int_{\phi} P(X_i|\phi, \Theta, Y) P(\phi|\Theta, Y) d\phi \tag{5}$$

where $P(\phi|\Theta, Y)$ expresses prior information about the distribution of orientations. Calculation of the prior relies on the assumption of smoothness in the reconstruction. Smoothness is encoded in the assumption that all Fourier components V_l are independent and Gaussian distribution with zero mean and unknown variance τ_l^2, so that:

$$P(\Theta|Y) = \prod_{l=1}^{L} \frac{1}{2\pi\tau_l^2} exp\left(\frac{|V_l^2|}{-2\tau_l^2}\right) \tag{6}$$

We use expectation maximization to find the optimal values of V_l, τ_l, and σ_{ij}. For the expectation maximization algorithm we treat the orientation (Φ) of all the projections as hidden variables. In the *E-step* of the algorithm we basically have to compute the conditional expectation $E_{\mathbf{\Phi}|\mathbf{X},\Theta_n} lnP(\mathbf{X}, \phi|\Theta)$. Next, in the *M-step*, we maximize this expression along with the prior with respect to Θ. Using this, we get the optimal values of V_l, τ_l, and σ_{ij} as follows:

$$\begin{aligned}
V_{l^*} &= \frac{\sum_{i=1}^{N} \int_{\phi} K_{\phi}^{n} \sum_{j=1}^{J} P^{\phi_{jl^*}^T} \frac{CTF_{ij} X_{ij}}{\sigma_{ij}^2} d\phi}{\sum_{i=1}^{N} \int_{\phi} K_{\phi}^{n} \sum_{j=1}^{J} P^{\phi_{jl^*}^T} \frac{CTF_{ij}^2}{\sigma_{ij}^2} d\phi + \frac{1}{\tau_{l^*}^2}} \\
\tau_{l^*}^2 &= \frac{1}{2}|V_{l^*}|^2 \\
\sigma_{ij}^* &= \frac{1}{2}\int_{\phi} K_{\phi}^{n} |X_{ij} - CTF_{ij}\sum_{l=1}^{L} P_{jl}^{\phi} V_l|^2 d\phi
\end{aligned} \tag{7}$$

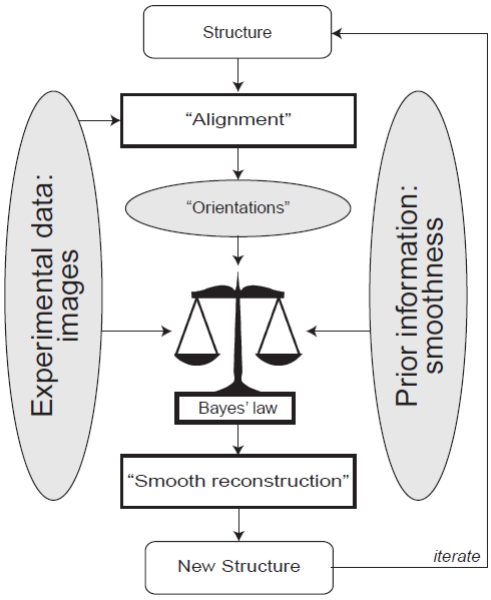
where

$$K_{\phi}^{n} = \prod_{i=1}^{N} \Gamma_{i\phi}^{n} \tag{8}$$

and

$$\Gamma_{i\phi}^{n} = \frac{P(X_i|\phi, \Theta_n) P(\phi|\Theta_n)}{P(X_i|\Theta_n)} \tag{9}$$

For the derivation of these formulas refer to Supplementary Article 1 (7).
We implemented this algorithm in an iterative manner, where we first calculate the probability of each orientation for each projection - $\Gamma_{i\phi}^{n}$ after which we calculate the new estimates for V_l, τ_l, and σ_{ij}. The overview of the algorithm can be seen below



It is important to note that the algorithm outlined above is a local optimizer. Thereby, the outcome of the refinement depends on the suitability of the starting model, and grossly incorrect starting models may lead to suboptimal results. Typically, to reduce bias to a possibly incorrect starting model, one applies a strong low-pass filter to the starting model.

2.2 Implementation

An integral component of the reconstruction algorithm is the back-projection and the projection operation. The success of the algorithm relies on how accurate the backprojection and projection operation are and whether these operations are reversible or not. These operations need to be entirely implemented in the Fourier domain to preserve the scale of the values and prevent the assumption of approximations that are often used in the Fourier domain.

2.2.1 Backprojection operation

For this operation, a fine radial grid is created and the projections are places along the appropriate orientations. If two or more projections have the same orientations, they are first ordered with respect to the other projections cyclically and then equally distributed in the gap in between. Interpolation plays an important role in the back projection operation. After the projections are placed in the 2D transform of the object the remaining missing elements are filled by solving a direct linear system of equations.

2.2.2 Projection operation

This operation is simple as we just have to extract the projection along the given direction. Having a universal grid makes this operation very simple and systematic.

There are also a number of modifications which need to be made to Equations 7, before they can be implemented in practice. A straightforward implementation of Equations 7 would lead to unsatisfactory results.Therefore we make two important modifications.

- The derivation of the formulas depends on the assumption of independence between Fourier components of the signal. This assumption is known to be a poor one because the image has a limited support in real space. Therefore, in the calculations presented, all estimates for τ_l^2 were multiplied by a constant, T = 20, in an attempt to account for the correlations between Fourier components in the signal.
- The summation over all back-projected images in the numerator of Eq. (7) results in a severely non-uniformly sampled 2D transform. This transform must be properly weighted before the actual reconstruction is obtained by an inverse Fourier transform operation, since straightforward division by the weights in the

denominator of Eq. (7) would lead to unsatisfactory results. Therefore we iteratively estimate the weights using the methodology described in Pipe and Menon, 1999 [7]

Some other minor implementations, but necessary for the success of the RELION algorithm was to decrease the maximum error in the orientations as the iterations progressed. This signifies our confidence in the refinement and reduces the effect of a drifting projection - A projection whose angle estimate drifts in the wrong direction.

2.3 Experiment

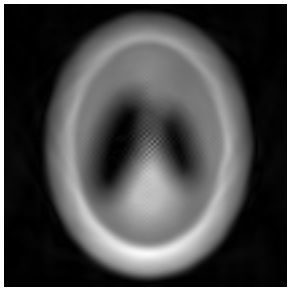
In all the experiments, a 100×100 image was used. 180 projections were taken in various orientations the accuracy of which we only knew up to $\pm 5^{\circ}$. Each of those projections could even have a random unknown shift up to ± 2 pixels. On top of this all projections have 5% Gaussian noise in them. The RELION algorithm is designed to incorporate all of these inconsistencies and correct them as the iterations progress.

The initial model is constructed using the back-projection operation on shifted projections using the incorrect angle estimates. To reduce model bias a low-pass filter is applied to create a low resolution image. This serves as the initial model of our algorithm.

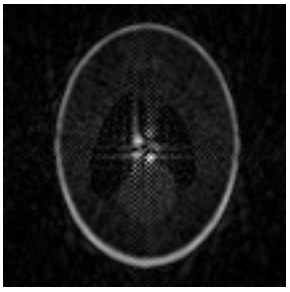
This is the original image, the structure of which we will attempt to reconstruct.

Figure 1: Original image

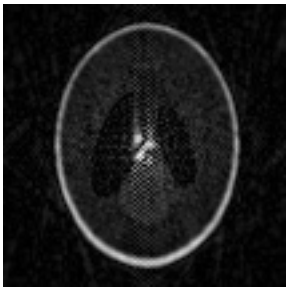
This is how the successive reconstructions of the algorithms look like. Every alternate iteration is shown to reduce the redundancy of information and focus on the improvements.



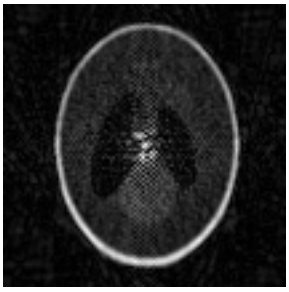
(a) Initial model



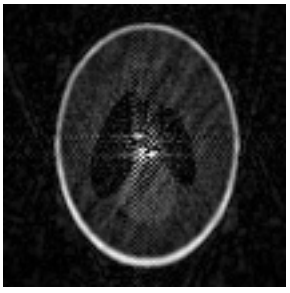
(b) Iteration 2



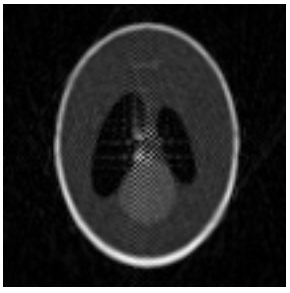
(c) Iteration 4



(d) Iteration 6



(e) Iteration 8



(f) Iteration 10

Figure 2: The progress of the algorithm

At the end of the 10^{th} iteration, we have correctly converged on to the original structure we wanted to estimate. This is what the error, compared to the original image, looks like as the iterations progress. The error metric used it the Mean Squared Error between the reconstructed image and the test image.

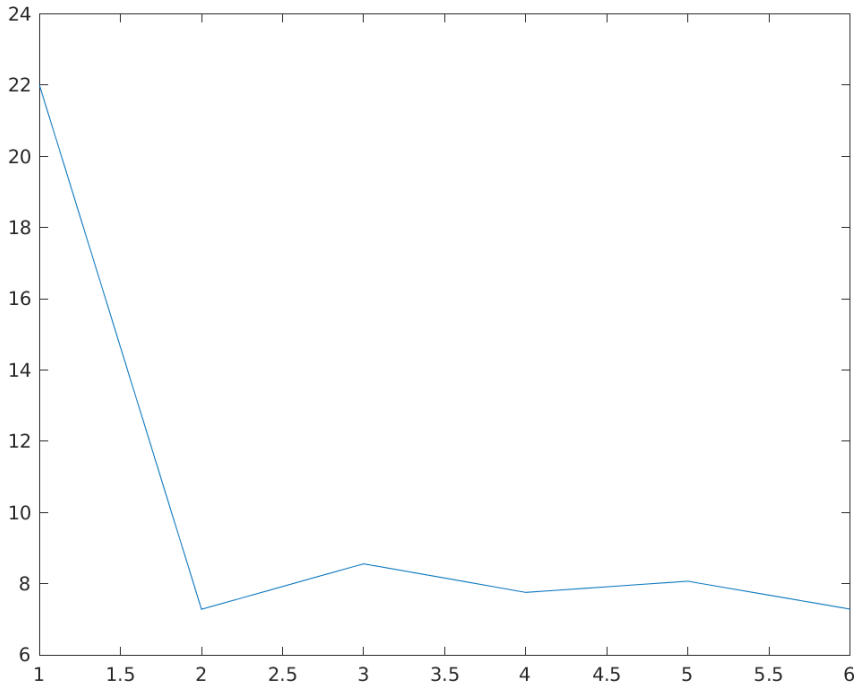


Figure 3: Error convergence

Thus we can conclude, using a Bayesian optimization framework and an Expected-Maximization algorithm we can converge onto an accurate estimate of the test structure provided we have a reasonably accurate initial estimate of the image and the orientations. The next logical step of our work is to see, if we can arrive at an good estimate of the structure without any prior information about the structure or the orientation. This is exactly what the next algorithm - *CryoSPARC* is going to tackle.

3 CryoSPARC

CryoSPARC [8] is an algorithm for rapid unsupervised cryo-EM structure determination. It makes possible to perform unsupervised *ab initio* structure determination without any prior structural knowledge.

Like the **RELION** algorithm, even cryoSPARC frames the optimization problem in the Bayesian likelihood framework [9, 10].

$$\arg\max_{V_{1,..,K}} = logp(V_{1,..,K}|X_{1,..,N}) = \arg\max_{V_{1,..,K}} \sum_{i=1}^{N} log \sum_{j=1}^{K} \frac{1}{K} \int p(X_i, \phi_i|V_j) d\phi_i + logp(V_{1,..,K}) \tag{10}$$

The aim of this optimization is to find the structures ($V_{1,..,K}$) that best explain the observed projections. In cryoSPARC an **SGD - Stochastic Gradient descent** [11] optimization scheme is used to quickly identify low-resolution structures that are consistent with the set of observed images. An SGD optimization scheme is proposed because the optimization problem above is a non-convex algorithm and starting with any random model may quickly converge to a local optima. Sensitivity to local optima is seen in most optimization algorithms. To circumvent these problems SGD is proposed as a key tool for the optimization of nonconvex functions.

3.1 Stochastic Gradient Descent

SGD iteratively optimizes an objective function by computing approximate gradients and taking steps in the parameter space according to those gradients.

Objective Function - The optimization function is the log posterior probability distribution over K 3-D densities, given N particle projections, i.e,

$$\begin{aligned}
&= \arg \max_{V_1,V_2,..,V_K} logp(V_1, V_2, .., V_K|X_1, X_2, .., X_N) \\
&= \arg \max_{V_1,V_2,..,V_K} logp(X_1, X_2, .., X_N|V_1, V_2, .., V_K) + logp(V_1, V_2, .., V_K) \\
&= \arg \max_{V_1,V_2,..,V_K} \sum_{i=1}^{N} logp(X_i|\mathbf{V}) + \sum_{j=1}^{K} logp(V_j) \\
&= \arg \max_{\mathbf{V}} f(\mathbf{V})
\end{aligned} \tag{11}$$

In Equation 2, the second term is a joint prior over the structures. This prior can be set, for example, to restrict density to be strictly positive or to penalize high-frequency noise in structures. In cryoSPARC, the prior is assumed to be independent over the structures, meaning that it can factor over the structures as we did in Equation 2.

We know that,

$$\begin{aligned}
p(X_i|\mathbf{V}) &= \sum_{j=1}^{K} \pi_j p(X_i|V_j) \equiv U_i \\
p(X_i|V_j) &= \int p(X_i|\phi, V_j)p(\phi)d\phi
\end{aligned} \tag{12}$$

Therefore,

$$f(\mathbf{V}) = \sum_{i=1}^{N} log\left(\sum_{j=1}^{K} \pi_j \int p(X_i|\phi, V_j)p(\phi)d\phi \right) + \sum_{j=1}^{K} logp(V_j) \tag{13}$$

The mixing probabilities between the different structures are given by π_j and in cryoSPARC, the mixing probabilities are assumed to be uniform over all classes. i.e., $\pi_j = K^{-1}$. Also a prior over poses p_ϕ can be specified, and in this work an uniform distribution is again used.

Gradient - SGD optimizes the objective function in Equation 4 by iteratively updating the parameters $V_1, .., V_K$. The gradient of Equation 4 with respect to each structure is computed in order to take steps. The gradient is

$$\begin{aligned}
\frac{\partial f}{\partial V_k} &= \sum_{i=1}^{N} \frac{1}{U_i} \frac{\partial U_i}{\partial V_k} + \frac{\partial}{\partial V_k} logp(V_k) \\
&= \sum_{i=1}^{N} \frac{1}{U_i} \pi_k \int \frac{\partial}{\partial V_k} p(X_i|\phi, V_k)p(\phi)d\phi + \frac{\partial}{\partial V_k} logp(V_k)
\end{aligned} \tag{14}$$

Approximate gradient - The sum giving the gradient in Equation 5 is over all projections in the dataset. In SGD, the sum is approximated using subsampling. At each iteration, SGD selects a subset of the projections and uses only those projections to approximate Equation 5. The size of each batch $\mathbf{M}$ can vary over iterations and in cryoSPARC, it is set automatically based on the current resolution and the number of classes K. The approximate gradients are given by,

$$\frac{\partial f}{\partial V_k} \approx G_k \equiv \frac{N}{M} \sum_{i \in M} \frac{1}{U_i} \pi_k \int \frac{\partial}{\partial V_k} p(X_i|\phi, V_k)p(\phi)d\phi + \frac{\partial}{\partial V_k} logp(V_k) \tag{15}$$

SGD update rule with momentum - The approximate gradient in Equation 6 points in a direction within the space of structure that will, in expectation over random selections of minibatches, improve the objective functions in Equation 4. It is well known that in the general case, optimization of non convex functions like Equation 4

is difficult and SGD only provides guarantees of local convergence. Nevertheless, in practice, SGD performs well and finds the correct 3D structures. SGD computes the update at the current iteration, $dV_k^{(n-1)}$, by scaling the current gradient $G_k^{(n)}$ with a step size η_k and combining this linearly with the previous update in a ration given by μ. This linear averaging is known as momentum [12] and serves to smooth the noisy approximate gradient directions in SGD.

$$\begin{aligned} dV_k^{(n)} &= (\mu) dV_k^{(n-1)} + (1-\mu)(\eta_k) G_k^{(n)} \\ V_k^{(n+1)} &= V_k^{(n)} + dV_k^{(n)} \end{aligned} \tag{16}$$

3.2 Implementation

In this work, CryoSPARC is implemented in a 2-D setting, where the structure is a 2-D image and the projections are in 1-D. The back-projection and the projection operation are implemented in the same way as in RELION described in Section 2.2.1 and 2.2.2.

With some algebraic manipulations, the stochastic gradient calculated in Equation 15 can be expressed as combination of the back-projection and projection operation. It was observed that the convergence of the SGD descent is pretty robust to the mini-batch size and taking 30 to 60 projections out of 180 have an accurate reconstruction.

Step size - In most gradient-descent algorithms, setting the step-size can often require tuning. Step-sizes that are too small yield slow convergence, while step-sizes that are too large can cause divergence of the algorithm. SGD generally has a similar property, but analysis of a particular optimization problem can yield methods for automatically setting the step-sizes to appropriate values. In this work the step-sizes are set using an approximation of the second-order curvature of the objective function $f(\mathbf{V})$. In this work the step-sizes are set using an approximation of the second-order curvature of the objective function $f(\mathbf{V})$. The step-size is calculated by calculating an approximate of the Hessian matrix. Approximate Hessian for $f(\mathbf{V})$ gives diagonal Fourier space curvature information about the objective function. Concretely, the inverse step-size for each structure is given by

$$\frac{1}{\eta_k} = \left\| \sum_{i \in M} \pi_k \int p(\phi | X_i, V_k) P_\phi^T \frac{C_i^2}{\sigma^2} d\phi \right\|_\infty \tag{17}$$

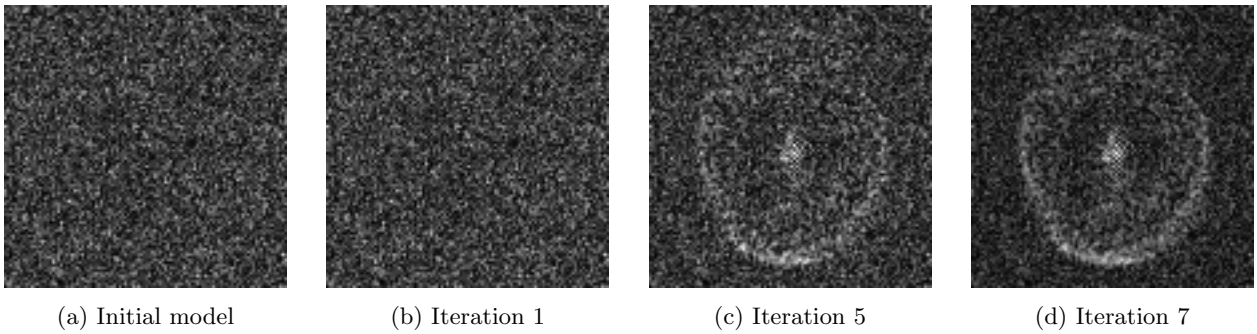
3.3 Experiment

In the experiment, we have no prior structural information or any information about the orientations of the projections. We only have 180 projections taken in various unknown directions. We used an image initialized randomly (noise) with the only caveat that the scale of the values is same as that of the projections which makes sense because these projections are a slice of the 2-D Fourier transform of the image. All the projections have 5% Gaussian noise (Gaussian noise with variance 5% of the average value of the projections).

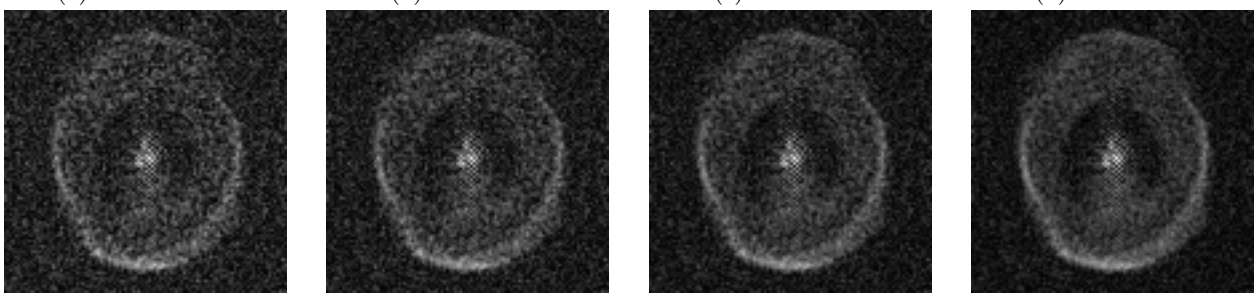
This is the original image, the structure of which we will attempt to reconstruct.

Figure 4: Original image

Next we will show the reconstructed image as the SGD algorithm progresses. Some iterations have been skipped to focus on the improvement of the algorithm.



(a) Initial model (b) Iteration 1 (c) Iteration 5 (d) Iteration 7



(e) Iteration 9 (f) Iteration 11 (g) Iteration 14 (h) Iteration 20

Figure 5: The progress of the algorithm

We can see how the SGD algorithm gradually converges to the correct structure of the original object. The plot below shows the error with the original image as the iterations progress. The error metric used is the mean squared error between the reconstructed image and the test image.

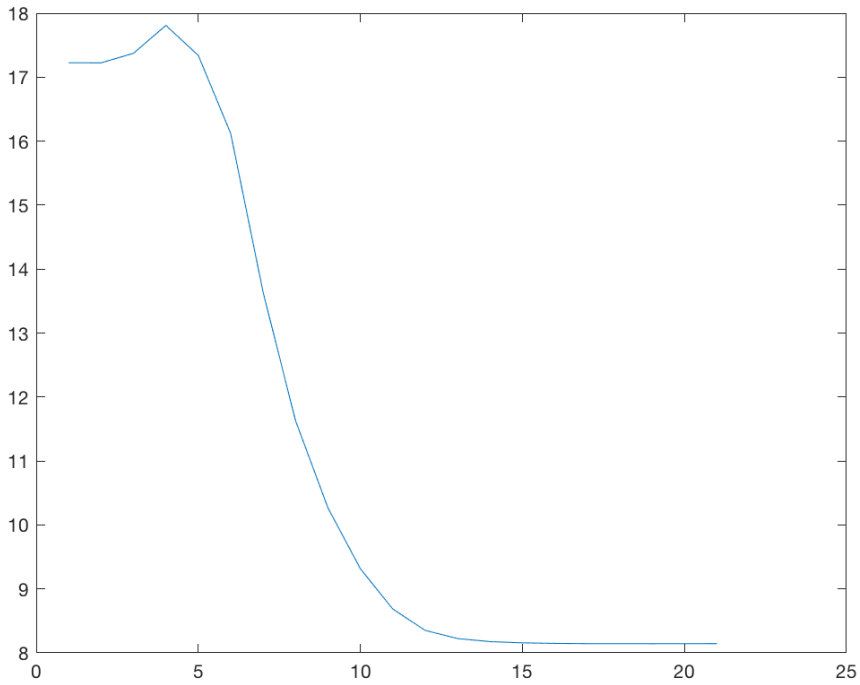


Figure 6: Error convergence

We can see that initially this error increases as SGD is exploring the space and once it hits the semblance of the correct structure is quickly converges down to the right structure.

4 Future work and conclusion

In this work, two algorithms have been implemented, one which can be used for the refinement of a good initial model and the other which can used for generating the initial model. It will be interesting to see if RELION can be used to refine the images produced by cryoSPARC. RELION with a fine resolution and a wider search domain for each orientation can be used to generate better models and this work can hence become a complete pipeline to generate an accurate model.

RELION assumes that all the Fourier components are mean 0 and just iteratively improves the variance at each iteration. It would be interesting to see if stronger priors can be applied in subsequent iterations to generate better models from lesser accurate initial models.

Last but not the least, this work needs to be applied on actual cryo-EM datasets and we need to validate whether all our aforementioned hypothesis actually work in the real world or not.

5 Github Repository

All the work presented above can be reproduced using the scripts present in EPFL-CryoEM (7). For a complete overview of the file structure please refer README.md (7).
All results are generated using MATLAB^{TM} 2018a on 2.3GHz quad-core desktop.

6 Acknowledgements

This work would not have been possible without the guidance and motivation of professor Victor Panaretos (7) and the resources provided to me by EPFL (7). I am grateful to all the authors of the RELION and the CryoSPARC algorithm for being the founding stone on which this work is based.

7 Links

In case, the inbuilt links are not active, you can find all the mentioned links over here.

- Supplementary Article 1 - https://www.overleaf.com/read/bmrmktdyyppw
- EPFL-CryoEM Github repository - https://github.com/Arunabh98/EPFL-CryoEM
- README.md - https://github.com/Arunabh98/EPFL-CryoEM/blob/master/README.md
- Victor Panaretos - http://smat.epfl.ch/victor/
- EPFL - https://www.epfl.ch/index.en.html

References

[1] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612, 6 1984.

[2] Joachim Frank. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Oxford University Press, 3 2006.

[3] Fred J. Sigworth. Principles of cryo-EM single-particle image processing. *Microscopy*, 65(1):57–67, 2 2016.

[4] Jorge F. Genise and Gerardo Cladera. Application of computerized tomography to study insect traces. *Ichnos*, 4(1):77–81, 4 1995.

[5] Hengyong Yu and Ge Wang. Data Consistency Based Rigid Motion Artifact Reduction in Fan-Beam CT. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 26(2), 2007.

[6] Sjors H.W. Scheres. A Bayesian View on Cryo-EM Structure Determination. *Journal of Molecular Biology*, 415(2):406–418, 1 2012.

[7] J G Pipe and P Menon. Sampling density compensation in MRI: rationale and an iterative numerical solution. *Magnetic resonance in medicine*, 41(1):179–86, 1 1999.

[8] Ali Punjani, John L Rubinstein, David J Fleet, and Marcus A Brubaker. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3):290–296, 3 2017.

[9] Sjors H.W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3):519–530, 12 2012.

[10] F.J. Sigworth. A Maximum-Likelihood Approach to Single-Particle Image Refinement. *Journal of Structural Biology*, 122(3):328–339, 1998.

[11] Léon Bottou. Large-Scale Machine Learning with Stochastic Gradient Descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Physica-Verlag HD, Heidelberg, 2010.

[12] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning, 2013.