

# Online CryoEM Study Group

Geoffrey Woollard, Vancouver, Canada

February 6, 2022

## Contents

<b>1</b>	<b>Dates and Topics</b>	<b>3</b>
<b>2</b>	<b>General Information</b>	<b>4</b>
2.1	Archived material . . . . .	4
2.2	Audience and Streams . . . . .	4
2.3	Audience . . . . .	4
2.4	Pre-requisites . . . . .	4
2.5	Scope . . . . .	4
2.5.1	Math / Computer science . . . . .	4
2.5.2	Physics . . . . .	5
2.6	Meeting Format . . . . .	5
2.7	Slack . . . . .	5
2.8	Testimonials . . . . .	6
<b>3</b>	<b>Learning Resources</b>	<b>6</b>
<b>4</b>	<b>Upcoming Meetings</b>	<b>7</b>
4.1	11 Feb 2022 - CryoDRGN2 . . . . .	7
4.2	4 Feb 2022 - SO(3) - encoding of S0(3) . . . . .	7
<b>5</b>	<b>Reading list</b>	<b>8</b>
5.1	Advanced treatment of rigid-body rotations (SO(3)) . . . . .	8
5.2	Assorted mathy cryoem papers . . . . .	9
5.3	Advanced microscopy . . . . .	9
<b>6</b>	<b>Past Meetings</b>	<b>11</b>
6.1	21 Jan 2022 - VESPER: global and local cryo-EM map alignment using local density vectors. . . . .	11

6.2	28 Jan 2022 - Guest: Qinwen (Wendy) Huang - Weakly Supervised Learning for Joint Image Denoising and Protein Localization in Cryo-EM . . . . .	12
6.3	4 Feb 2022 - $SO(3)$ - encoding of $S^0(3)$ . . . . .	12

# 1 Dates and Topics

Meetings are generally on Thursdays (morning Pacific time, afternoon Eastern time, evening Europe)

Date	Time	Topic
Fri 21 Jan 2022	9 AM PST	mathy cryoem paper
Fri 28 Jan 2022	9:15 AM PST	Guest: Qinwen (Wendy) Huang - Weakly Supervised Learning for Joint Image Denoising and Protein Localization in Cryo-EM
Fri 4 Feb 2022	9 AM PST	Rotations - encoding of $S0(3)$
Fri 11 Feb 2022	9 AM PST	CryoDRGN2
Fri 18 Feb 2022	9 AM PST	advanced rotations
Fri 25 Feb 2022	9 AM PST	mathy cryoem paper
Fri 4 Mar 2022	9 AM PST	advanced rotations
Fri 11 Mar 2022	9 AM PST	mathy cryoem paper
Fri 18 Mar 2022	9 AM PST	advanced rotations
Fri 25 Mar 2022	9 AM PST	mathy cryoem paper

## 2 General Information

### 2.1 Archived material

Meetings from 2020-2021 are archived here. The audience was a mix of beginners and advanced practitioners, and computational methods developers.

### 2.2 Audience and Streams

Feel free to share this document and direct people to sign up at <https://forms.gle/BUeUW14vV4pyQbDDA> so I have the emails in one place. Online meeting links are emailed to those on this list. **Please join the Slack group and ask questions there, rather than emailing me.**

### 2.3 Audience

In 2022, I am catering to a computational methods development audience. I see this group as a way for computational methods developers to get together in a "pre-competitive" learning environment.

**Practitioners → computational methods developers:** You are a structural biologist, or biochemist, and perhaps an advanced cryo-EM practitioner. You would like to train in computational methods development, either to do very advanced data processing, or develop your own methods.

**Pure computational discipline → cryo-EM computational methods developers:** You have a background in computer science (computer vision, deep learning, statistics, electrical/computer engineering) and would like to develop methods for the "killer application" of cryo-EM.

### 2.4 Pre-requisites

The bar is quite high, and this group is not for all. There are very good resources out there for self-study; see this annotated bibliography. If you have done an undergraduate degree in an advanced computational program (physics, chemistry, computer science, statistics, applied math) or are a PhD student in a computational field, then you are in good company in this group.

### 2.5 Scope

#### 2.5.1 Math / Computer science

1. Amortized inference, model learning
2. Physics aware/inspired/infused deep learning

3. Deep learning of the image formation model (rotation, etc)
4. Computationally modelling uncertainty in the image formation model
5. Geometric deep learning and invariance/equivariance in cryo-EM
6. Computational optimal transport
7. Computational differential geometry
8. Optimization
9. Custom GPU kernels, including gradient for backprop/autodiff

### **2.5.2 Physics**

1. Electron optics
2. Higher order CTF aberrations
3. Multi-slice
4. Sample damage
5. Detector physics
6. Solvation
7. Poisson-Boltzmann equation
8. Modelling choices to encode electrostatic / Coulombic density

## **2.6 Meeting Format**

The meetings are meant to be more informal that is typical in research talks. The point is to learn and discuss with other learners, experienced practitioners, and experts. They are also more comprehensive than typical journal clubs. We may stick with a papers or series of papers for multiple weeks to sufficiently learn the material.

## **2.7 Slack**

We will use the Slack channel 'cryoem\_study\_group' for asynchronous chat. Please join the Slack group and ask questions there, rather than emailing me. You can request a link to join by emailing me.

## 2.8 Testimonials

- *Shayan Shekarforoush, PhD student with Marcus Brubaker and David Fleet, Jan 2021.* I joined this reading group in mid October and I wish I would have done so much earlier. Although I joined when the group was in the middle of reading a fascinating, recently published book in Single-Particle cryoEM, everyone was so welcome that I did not feel I am way behind others. My background is in CS and I do research as a method developer in this field. With that said, I learned a lot from discussions of people with expertise in experimental side of this area. I believe that this group helped me to build a better intuition and now I feel more comfortable with the underlying math and physics of this topic. This group also provided the opportunity to attend talks of prominent researchers in cryoEM where anyone could openly ask their questions and have clear discussions. Looking forward to having more collaborations with the members of this group.

## 3 Learning Resources

1. I have made an annotated bibliography organized thematically here.
2. Coding notebooks to play around with are here. If there is incompatibility between the notebook and the code base in the repo, that is because the code base has been updated. Older version of the code are available via past commits.

## 4 Upcoming Meetings

### 4.1 11 Feb 2022 - CryoDRGN2

–Pre-reading

1. Zhong, E. D., Lerer, A., Davis, J. H., & Berger, B. (2021). CryoDRGN2 : Ab initio neural reconstruction of 3D protein structures from real cryo-EM images. *Iccv*, 4066-4075.

–Questions

1. Write out a schematic (computational graph) of CryoDRGN2. What goes into and out of the encoder and decoder?
2. How could cross validation be incorporated into CryoDRGN? What would count as "out of distribution" for various computational workflows, and how would CryoDRGN perform?

### 4.2 4 Feb 2022 - SO(3) - encoding of S0(3)

–Pre-reading

1. (2013). Hartley et al., Rotation Averaging. <https://link.springer.com/content/pdf/10.1007/s11263-012-0601-0.pdf>
  - Distance between rotations
  - SO(3) encoding

–Questions

1. Table 2 Relationship between the different metrics on SO(3): Code up the various metrics ( $d_{chord}$ ,  $d_{quat}$ ) and numerically verify these relationships on some (arbitrary) rotations

## 5 Reading list

### 5.1 Advanced treatment of rigid-body rotations ( $SO(3)$ )

–Pre-reading

1. (2007). Baldwin, P. R., & Penczek, P. A. The Transform Class in SPARX and EMAN2. *Journal of Structural Biology*, 157(1), 250?261. <http://doi.org/10.1016/j.jsb.2006.06.002>
2. Orthogonal Procrustes problem [https://en.wikipedia.org/wiki/Orthogonal\\_Procrustes\\_problem](https://en.wikipedia.org/wiki/Orthogonal_Procrustes_problem)
3. (2021). Brofos, J. A., Brubaker, M. A., & Lederman, R. R. Manifold Density Estimation via Generalized Dequantization. <https://arxiv.org/pdf/2102.07143.pdf>
4. (2017). State Estimation for Robotics. State Estimation for Robotics. Cambridge: Cambridge University Press. <http://doi.org/10.1017/9781316671528>
5. (2021). Implicit-PDF: Non-Parametric Representation of Probability Distributions on the Rotation Manifold <https://implicit-pdf.github.io/>
6. (2020). A Smooth Representation of Belief over  $SO(3)$  for Deep Rotation Learning with Uncertainty. <https://arxiv.org/pdf/2006.01031.pdf>
7. (2021). Eliminating Topological Errors in Neural Network Rotation Estimation Using Self-selecting Ensembles. <https://dl.acm.org/doi/pdf/10.1145/3450626.3459882>
8. (2021). On the Continuity of Rotation Representations in Neural Networks. <https://arxiv.org/pdf/1812.07035.pdf>
9. (2021). Learning Rotation Invariant Features for Cryogenic Electron Microscopy Image Reconstruction. <https://arxiv.org/pdf/2101.03549.pdf>
10. (2020). SE(3)-Transformers: 3D Roto-Translation Equivariant Attention Networks. <https://arxiv.org/pdf/2006.10503.pdf>
11. (2020). Falorsi, L., de Haan, P., Davidson, T. R., & Forr, P. Reparameterizing distributions on Lie groups. *AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics*, 89. <https://arxiv.org/pdf/1903.02958.pdf>
12. (2020). Dmitry Kostyaev. Hands on Tutorials: Better rotation representations for accurate pose estimation. *Towards Data Science*. <https://towardsdatascience.com/better-rotation-representations-for-accurate-pose-estimation-e890a7e1317f>
  - Optimization
  - $SO(3)$  encoding



## 5.2 Assorted mathy cryoem papers

–Pre-reading

1. Tagare, H. D., Kucukelbir, A., Sigworth, F. J., Wang, H., & Rao, M. (2015). Directly reconstructing principal components of heterogeneous particles from cryo-EM images. *Journal of Structural Biology*, 191(2), 245?262. <http://doi.org/10.1016/j.jsb.2015.05.007>
2. Zivanov, J., Nakane, T., & Scheres, S. H. W. (2019). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ*, 6(1), 5?17. <http://doi.org/10.1107/S205225251801463X>
3. Katsevich, E., Katsevich, A., & Singer, A. (2015). Covariance matrix estimation for the cryo-em heterogeneity problem. *SIAM Journal on Imaging Sciences*, 8(1), 126-185. <http://doi.org/10.1137/130935434>
4. Penczek, P. A. (2010). Resolution Measures in Molecular Electron Microscopy. In *Methods in Enzymology* (1st ed., Vol. 482, pp. 73?100). Elsevier Inc. [http://doi.org/10.1016/S0076-6879\(10\)82003-8](http://doi.org/10.1016/S0076-6879(10)82003-8)
5. Ede, J. M. (2020). Review: Deep learning in electron microscopy. *ArXiv*. <http://doi.org/10.1088/2632-2153/abd614>
6. Zhu, D., Wang, X., Fang, Q., Etten, J. L. Van, Rossmann, M. G., Rao, Z., & Zhang, X. (n.d.). Pushing the resolution limit by correcting the Ewald reconstructions. *Nature Communications*, (2018), 1-7. <http://doi.org/10.1038/s41467-018-04051-9>
7. Maji, S., Liao, H., Dashti, A., Mashayekhi, G., Ourmazd, A., & Frank, J. (2020). Propagation of Conformational Coordinates across Angular Space in Mapping the Continuum of States from Cryo-EM Data by Manifold Embedding. *Journal of Chemical Information and Modeling*, 60(5), 2484?2491. <http://doi.org/10.1021/acs.jcim.9b01115>
8. Zhong, E. D., Lerer, A., Davis, J. H., & Berger, B. (2021). CryoDRGN2 : Ab initio neural reconstruction of 3D protein structures from real cryo-EM images. *Iccv*, 4066-4075.

## 5.3 Advanced microscopy

–Pre-reading

1. Glaeser, R. M., Hagen, W. J. H., Han, B. G., Henderson, R., McMullan, G., & Russo, C. J. (2021). Defocus-dependent Thon-ring fading. *Ultramicroscopy*, 222(October 2020), 113213. <http://doi.org/10.1016/j.ultramic.2021.113213>

2. Russo, C. J., & Egerton, R. F. (2019). Damage in electron cryomicroscopy: Lessons from biology for materials science. *MRS Bulletin*, 44(12), 935?941. <http://doi.org/10.1557/mrs.2019.284>
3. Russo, C. J., & Henderson, R. (2018). Ewald sphere correction using a single side-band image processing algorithm. *Ultramicroscopy*, 187, 26?33. <http://doi.org/10.1016/j.ultramic.2017.11.001>
4. Electron optics textbook chapters (Hawkes and Kasper; Spence; Reimer and Kohl)

## 6 Past Meetings

### 6.1 21 Jan 2022 - VESPER: global and local cryo-EM map alignment using local density vectors.

–Pre-reading

1. Han, X., Terashi, G., Christoffer, C., Chen, S., & Kihara, D. (2021). VESPER: global and local cryo-EM map alignment using local density vectors. *Nature Communications*, 12(1). <http://doi.org/10.1038/s41467-021-22401-y>

–Questions

1. Consider Figure 1 and the Z-score results. When aligning two different maps, and getting DOT scores, how would one re-scale to a Z-score? And to what end?
2. Have you developed any algorithms that start from aligned maps? where did the initial alignment arise from?
3. How do you compare aligning ?by eye? with the algorithms mentioned in the paper, such as the popular fitmap in Chimera. what do you think you are doing in your mind, and how would you break it down into algorithmic steps?
4. Figure 2i (and text on last paragraph of p. 4) shows a case where VESPR did not perform well. two different proteins were predicted to be close in shape ?because they have an overall similar shape and also because these maps are largely hollow inside, and thus inconsistency inside the maps were not much penalized?. Relate this to the DOT score. What other approaches would not suffer from this?
5. In Table 2 there are some cases where the RMSD goes up (worse alignment) with finer grained sampling (rotation angle and translation). Why might this be happening?
6. In the Discussion, the authors explain intuitively how cross correlation (CC) and their DOT score result in different alignments: *In CC, positions with large absolute density values, such as those in a high-density region in a map, influence more to the overall CC value. On the other hand, for the DOT score the contribution of each aligned position pair is essentially the same because the vectors are normalized to the same length. But this also means that the DOT score can be affected by changes in local gradient caused by small structure variations.*

Consider the following cases and think through how the result might be different.

- (a) Continuous conformational heterogeneity: Blurred out flexible region with partial density.
- (b) Discrete conformational heterogeneity: Mixed discrete states in one map vs one of the partners.

- (c) Hinge like displacement (local rigid-body transformation of a domain).
  - (d) Two maps +/- domain.
7. What do you think of the opinion in the Discussion: *Note that, in general, the optimal parameter setting for a method differs for each map and the purpose of the computation. Thus, a perfectly fair comparison is not possible, and the comparison shown in this work is to characterize the performance of VESPER but not to rank the methods.* Hint: see **Responses to Comments by Reviewer #2**, p. 10.
  8. The authors chose a  $7\text{\AA}$  grid spacing. Why? How do you suspect the performance of VESPER to change with other grid spacings? Does the method require the same grid spacing when comparing maps? Hint: read the section **Exploration of parameter combinations**.
  9. Reviewer #1 asked the authors to justify what map-to-map alignment is useful. Why would a researcher want to align maps (vs model-to-model and model-to-map)?

## 6.2 28 Jan 2022 - Guest: Qinwen (Wendy) Huang - Weakly Supervised Learning for Joint Image Denoising and Protein Localization in Cryo-EM

–Pre-reading

1. Qinwen Huang, Ye Zhou, Hsuan-Fu Liu, & Alberto Bartesaghi (2021). Weakly Supervised Learning for Joint Image Denoising and Protein Localization in Cryo-EM [https://www.mlsb.io/papers\\_2021/MLSB2021\\_Weakly\\_Supervised\\_Learning\\_for.pdf](https://www.mlsb.io/papers_2021/MLSB2021_Weakly_Supervised_Learning_for.pdf)

–Questions

1. This paper exploits analytical likelihood-prior-posterior conjugacy between gaussian distributions. How could we extend this to Poisson noise with Poisson-Gamma conjugacy? What would be Poisson, and what would be Gamma?

## 6.3 4 Feb 2022 - SO(3) - encoding of S0(3)

–Pre-reading

1. [https://en.wikipedia.org/wiki/Euler\\_angles#Rotation\\_matrix](https://en.wikipedia.org/wiki/Euler_angles#Rotation_matrix)
2. Gregory G. Slabaugh. Computing Euler angles from a rotation matrix
3. Axis/Angle Representations for Rotations

4. Quine, JR. 4. Orthogonal Transformations and Rotations in Mathematical Techniques in Structural Biology, pp. 25-31
5. Visualizing quaternions (4d numbers) with stereographic projection
6. Quaternions and 3d rotation, explained interactively
7. Visualizing quaternions An explorable video series
8. Wiki: Quaternions and spatial rotation
9. Maxime Tournier's Research Notes. Quaternions

–Questions

1. Problems (e.g. problem 7) from Quine, JR. 4.5 Problems in *Mathematical Techniques in Structural Biology*, pp. 30-31
2. Implement the pseudo-code in Figure 1 of Gregory G. Slabaugh. Computing Euler angles from a rotation matrix, and use it to compute the Euler angles from a rotation matrix.
3. Go to the doublecover interactive, and select the 4D tab on the bottom. Find the quaternion that corresponds to a rotation of 90 deg in the cw direction about the z axis.
4. Quaternions can be used to describe the rotation of the point  $r = \langle r_x, r_y, r_z \rangle \in \mathbb{R}^3$  about angle  $\theta$  in the direction of vector  $\langle q_x, q_y, q_z \rangle$  as follows.  $p_{rotated} = q * p * q^{-1}$ , where  $q = \langle \cos(\theta/2), q_x \sin(\theta/2), q_y \sin(\theta/2), q_z \sin(\theta/2) \rangle$  and  $q^{-1} = \langle \cos(-\theta/2), q_x \sin(-\theta/2), q_y \sin(-\theta/2), q_z \sin(-\theta/2) \rangle$  encode the rotation;  $\vec{r}$  is represented as the quaternion  $p = (0, r_x, r_y, r_z)$ . Question: using pencil and paper, compute the rotation of the point  $r = \langle 1, 0, 0 \rangle$  about the z-axis by  $\theta = 180^\circ$  (ccw or cw, it's the same for this angle). Do you get  $r = \langle -1, 0, 0 \rangle$ ? You will have to use the quaternion multiplication table ( $ij = k$ , etc.), and know how to interconvert quaternions to cartesian 3D vectors.
5. Study the sections Conversion to and from the matrix representation and Performance comparisons on the wiki page for Quaternions and spatial rotation. If we want to rotate many xyz points in space with a 3D rotation, what is the efficient way to do this? Taking into account interconversions between different rotation conventions (Euler angles, rotation matrices, axis-angle, and quaternions) how should we efficiently store rotations versus do rotations?
6. Besides storing and doing rotations, we have to search over rotations. Wikipedia has a section on Differentiation with respect to the rotation quaternion. Unpack the

equation for  $\frac{\partial \mathbf{p}'}{\partial \mathbf{q}}$ . Would this equation ever be useful in cryoEM? When would we want to estimate the rotation by optimization, and how would we do so with this equation?