Online CryoEM Study Group

Geoffrey Woollard, Vancouver, Canada

July 14, 2021

Contents

1 Dates and Topics

Given the current distribution of our global audience (90%+ in North America and Europe), we will have the time in the morning, starting around 9 AM Pacific (12 PM Eastern). Meetings are generally on Thursdays.

Date	Time	Stream	Topic
Th 19 Nov 2020	9 AM PST	blur/sharpen	background math & defocus phase contrast
~~Th 26 Nov 2020~~			(cancelled for US Thanksgiving)
Th 3 Dec 2020	9 AM PST	blur	Fourier transform
Th 10 Dec 2020	9 AM PST	sharpen	Fourier transform
Th 17 Dec 2020	9 AM PST	blur	convolution, sampling, Nyquist
Th 7 Jan 2021	9 AM PST	sharpen	convolution, sampling, Nyquist
Th 14 Jan 2020	9 AM PST	blur	phase-contrast in the EM
Th 21 Jan 2021	9 AM PST	sharpen	phase-contrast in the EM
Th 28 Jan 2021	9 AM PST	blur	image formation (forward model)
Th 11 Feb 2021	8 AM PST	sharpen	image formation (forward model), multislice
Th 18 Feb 2021	8 AM PST	sharpen	2D expectation-maximization
Th 25 Feb 2021	8 AM PST	blur	coordinate systems & rotations
Th 4 March 2021	8 AM PST	sharpen	coordinate systems & rotations
Th 11 March 2021	7:30 AM PST	blur	interpolation
Th 18 March 2021	8 AM PST	sharpen	interpolation
Th 25 March 2021	8 AM PST	blur	Guest lecture: D. Dynerman, 3D reconstruction via direct Fourier inversion
Th 1 Apr 2021	8 AM PST	sharpen	3D reconstruction via Fourier inversion
Th 8 Apr 2021	9 AM PST	sharpen	variational autoencoders
Th 15 Apr 2021	8 AM PST	sharpen	cryoDRGN (variational autoencoders) with Ellen Zhong
Th 22 Apr 2021	9 AM PST	blur/sharpen	office hours
Th 29 Apr 2021	9 AM PST	blur/sharpen	Guest Lecture: J. Krieger, normal mode analysis & ProDy
Th 6 May 2021	9 AM PST	sharpen	Guest Lecture: J. Krieger, normal mode analysis & ProDy
Th 13 May 2021	9 AM PST	blur	Guest: Yong Zi Tan. Master class: preferred orientation
Th 20 May 2021	8:30 AM PST	sharpen	Guest: Philip Baldwin. Master class: preferred orientation
Th 27 May 2021	9 AM PST	blur	Fourier transforms and reciprocal space for the beginner (G. Jensen's course)
Th 3 June 2021	9 AM PST	blur/sharpen	office hours
Th 10 June 2021	9 AM PST	blur	Fourier transforms and reciprocal space for the beginner (G. Jensen's course)
Th 17 June 2021	9 AM PST	blur/sharpen	Glaeser, Nogales & Chiu (GNC), 1 Introduction and Overview (1.1, 1.2)
Th 24 June 2021	9 AM PST	blur/sharpen	GNC, 2 Sample preparation
Th 1 July 2021	9 AM PST	blur/sharpen	Office hours
Th 8 July 2021	9 AM PST	sharpen	Guest: Ben Himes. Journal club: multislice
Th 15 July 2021	9 AM PST	blur/sharpen	Journal club: 3DVA
Th 22 July 2021	9 AM PST	blur/sharpen	3DVA practical
Th 29 July 2021	9 AM PST	blur/sharpen	GNC, 3 Data collection (3.2, 3.3)
Th 5 Aug 2021	9 AM PST	blur/sharpen	Office hours
Th 12 Aug 2021	9 AM PST	blur/sharpen	GNC, 3 Data collection (3.4, 3.5)
Th 19 Aug 2021	9 AM PST	blur/sharpen	GNC, 4 Data processing. 4.3 CTF estimation and image correction (restoration)
Th 26 Aug 2021	9 AM PST	blur/sharpen	GNC, 4 Data processing (4.4, 4.5)
Th 2 Sept 2021	9 AM PST	blur/sharpen	Office hours
Th 9 Sept 2021	9 AM PST	blur/sharpen	GNC, 4 Data processing. 4.7 B factors and map sharpening
Th 16 Sept 2021	9 AM PST	blur/sharpen	GNC, 4 Data processing. 4.8 Optical aberrations and Ewald sphere curvature
Th 23 Sept 2021	9 AM PST	blur/sharpen	GNC, 5 Map Validation. 5.2 Measures of resolution: FSC and local resolution
Th 30 Sept 2021	9 AM PST	blur/sharpen	GNC, 5 Map Validation. 5.3 Recognizing the effect of bias and over-fitting
Th 7 Oct 2021	9 AM PST	blur/sharpen	Office hours
Th 14 Oct 2021	9 AM PST	blur/sharpen	GNC, 5 Map Validation. 5.4 Estimates of alignment accuracy
Th 21 Oct 2021	9 AM PST	blur/sharpen	GNC, 6 Model building and validation (6.2, 6.3)
Th 28 Oct 2021	9 AM PST	blur/sharpen	GNC, 6 Model building and validation. 6.4 Quality evaluation of cryo-EM map-derived models
Th 4 Nov 2021	9 AM PST	blur/sharpen	GNC, 6 Model building and validation. 6.5 How algorithms from crystallography are helping electron cryo-microscopy
... 2021 ...	...	...	...

2 General Information

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.**

2.1 Audience and Streams

Over 2/3 of the audience are grad students. There rest of the audience is a mix of long term staff (facility manager, seniour scientist, research associate, unspecified industry position, etc.), principal investigators (industry or academic), postdocs, and undergraduate. Meetings are labelled for the intended audience using the code names **blur** and **sharpen** , referring to map sharpening. Remember that both blurred and sharpened maps are important for being able to see what is going on and build a model in to the map. The goal of this study group is to have both intuition and be able to connect that back to the math, never missing the forest for the trees.

   Blur stream: Beginners and intermediates with may have years of expertise doing sample prep, collecting data on the scope, making maps, and building models. However, many people with such expertise expressed a desire to go deeper into the fundamentals, and understand things better under the hood, and requested help to become more confident in the math. One person put it bluntly, "I am interested in being more than a button pusher."

   Sharpen stream: This stream can be for people with a high degree of expertise that are looking to sharpen their skills, and interact with other experts. For example: methods developers looking to make friends with each other, university instructors, facility managers who have a strong working knowledge but want to go deeper into the foundations and refresh, people with strong math and physics backgrounds that are fairly new but will catch on quickly, experts in crystallography that are switching over to cryoEM but that have a strong theoretical basis in the underlying math already.

2.2 Pre-requisites

This is a sort of lower bound, so that you don't show up and become lost or frustrated, especially in the blur stream. Not all of this background is needed. To learn about trilinear interpolation, you don't need to know anything about electron optics.

2.2.1 Math

You should be able to hit the ground running if you did alright in first year undergraduate calculus or did a second and third year statistics/physics/chemistry course with a quantitative and computational emphasis. If it's been a while you may want to quickly run

through some online videos (khan academy for example) and brush up. You don't have to be strong in pure math (real analysis, proofs)

1. calculus (single and multiple integrals, derivatives)

2. trigonometry (sin, cos, etc.)

3. linear algebra, 2D and 3D cartesian geometry, vectors, matrices

4. statistics and probability (expectation, mean/variance, random variables, common distributions like gaussian/uniform/poisson)

5. summation notation (sigma sums)

Here are some resources from `khanacademy.org` for brushing up on your math

1. Unit: Derivatives: definition and basic rules

2. Unit: Derivatives: chain rule and other advanced topics

3. Unit: Integrals

4. Unit: Series

5. Linear Algebra

6. Unit: Probability

7. Unit: Counting, permutations, and combinations

8. Unit: Random variables

There is also a great series from 3Blue1Brown on The Essence of Linear Algebra, focusing on visual intuition of linear transformations; and the Essence of calculus: "The goal here is to make calculus feel like something that you yourself could have discovered."

2.2.2 Physics

- Some exposure to modern (20th century) physics in a first or second year undergraduate course (relativity, QM). This comes in especially for the forward model.

- Some intuition about what is physically happening in the sample, at room temperature, after vitrification, and during imaging. The 2017 Chemistry Nobel lectures by Henderson, Dubochet and Frank would be more than enough.

2.3   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.

- Lecture with Q&A

- Flipped classroom: i.e. we have a syllabus with pre-reading (textbook chapter, review paper) that we go through before and then discuss online

- Office hours with someone available to answer questions.

2.4   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.

3   Learning Resources

1. A good place to start is Grant Jensen's popular online course 'Getting Started in Cryo-EM' .

2. After getting up to speed on the prerequisites, another good next step is the content developed by Dr Frederick Sigworth and others at `https://cryoemprinciples.yale.edu/video-lectures`. If you are having problems with links, then try viewing his content on YouTube.

3. Also, I highly recommend the interactive learning material developed by Arjen Jakobi, for a course on High-Resolution Imaging at TUDelft: "The practicals are computational assignments in the form of interactive Jupyter notebooks hosted in a virtual learning environment. These notebooks contain code that can be executed to perform certain tasks or visualize results; you do not need any active knowledge of coding to work through the notebook." For the curious, the code that generates the visualizations is available on the repository.

4. I have made an annotated bibliography organized thematically here.

5. 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   15 July 2021 - Sharpen - 3DVA Part I

–Pre-reading

1. Tutorial: 3D Variability Analysis (Part One) `https://guide.cryosparc.com/processing-data/tutorials-and-case-studies/tutorial-3d-variability-analysis-part-one`

2. Sorzano, C. O. S., & Carazo, J. M. (2021). Principal component analysis is limited to low-resolution analysis in cryoEM. Acta Crystallographica Section D Structural Biology, 77(6), 835?839. http://doi.org/10.1107/s2059798321002291

3. Webinar: Resolving flexibility and heterogeneity with 3D Variability Analysis (90 min video).

–Questions

1. What is an eigenvolume?

2. Look at "Eigenimage1" in Fig 1a of Sorzano & Carazo 2021. Is it representative of any "particle" in the dataset?

3. Prior to 3DVA other groups have published similar PCA approaches. Have you come across any? Are they widely known in the cryoEM community?

4. 3DVA webinar (8 min mark): what is a probabilistic graphical model?

5. What are the inputs to 3DVA?

6. 3DVA webinar (20 min mark): Go through every notational aspect of the mathematical equation of the linear subspace model? What does everything refer to (all variables, subscripts, etc). In what sense is the linear subspace model linear?

7. 3DVA webinar (27 min mark): Is there any way to distinguish among possible causes of the heterogeneity with 3DVA: motion, dissociation, ligand binding, floating junk?

8. How could one recognize junk in latent variable space?

9. What can and can't be concluded from the three modes shown for the GPCR in the 3DVA webinar (45 min mark)

4.2   22 July 2021 - Sharpen - 3DVA Part II

–Pre-reading

1. Tutorial: 3D Variability Analysis (Part Two) `https://guide.cryosparc.com/processing-data/tutorials-and-case-studies/tutorial-3d-variability-analysis-part-two`

–Discussion group leaders

• Yuichiro Takagi

5 Past Meetings

5.1 19 Nov 2020 - Background Math & Defocus Phase Contrast

5.1.1 Blur

–Pre-reading

- Complex numbers and the complex exponential (10 min)

- Review of complex numbers (3 pages)

- Defocus phase contrast (35 min)

–Questions

- What is physically happening to the sample when the electron is detected at small diffraction angles vs large diffraction angles? What else can happen to the electron?

5.1.2 Sharpen

–Pre-reading

- Defocus phase contrast (35 min)

- 6.2 The Wave Equation for Fast Electrons in 'Advanced Computing in Electron Microscopy', Kirkland (2020), pp. 156-159.

–Questions

- In Sigworth's derivation of $|\Psi|^2$ in the 'Defocus phase contrast' video, he made various assumptions such as small theta, small epsilon. Under what extreme conditions would they break down. Do these occur in cryoEM? In other experimental regimes besides what is typical in cryoEM?

- Around the 23-25 min mark of the 'Defocus phase contrast' video the CTF seems to oscillate to zeros. Is this observed in practice? Why or why not?

- Work through the derivation with pencil and eraser, justifying each step as best you can. come with your questions to the group study.

- Biological samples are made of atoms that give faint contrast, when compared to samples with higher atomic numbers. Where does the atomic number of the sample come into the equations presented here?

5.2   3 Dec 2020 - Blur - Fourier Transform

–Pre-reading

- The Fourier transform in one dimension (35 min)

- The 1D Fourier Transform (7 pages)

- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks: Introduction, Fourier Series, Frequency Spectrum, 2-D Fourier Analysis. See the links at the bottom. For this week, work through notebooks "ip_basics_part1", "ip_basics_part2", "ip_basics_part3".

–Questions

- Prove the linearity, scale, shift, convolution properties of the FT in 1D.

- Equation 4 in The 1D Fourier Transform expresses how "a narrower function of x transforms into a broader function of u". Can you think of some examples of this in practice? Hint: what happens when you change magnification?

- Assume you have a 128 Åbox size, with pixel size of 1 Å. What is the spacing spacing between bins in Fourier space, if there are 64 bins in the negative direction and 64 bins in the positive direction? What length ranges (in units of Å) do the first few and last few frequency bins correspond to? How many frequency bins are between 10 and 5 Å, versus 5 and 2.5 Å?
  Hint: $\langle ..., [0, 1/128), [1/128, 2/128), ..., [\frac{128/2-2}{128}, \frac{128/2-1}{128}), [\frac{128/2-1}{128}, \frac{128/2}{128}) \rangle$.

5.3   10 Dec 2020 - Sharpen - Fourier Transform

–Pre-reading

- The Fourier transform in two and three dimensions (43 min)

- 2D and 3D Fourier transforms (9 pages)

- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks: Frequency Spectrum, 2-D Fourier Analysis. See the links at the bottom. For this week, play with notebooks "ip_basics_part4", "ip_basics_part5".

- Optional

  – Explanatory Material: "Tutorials and introductions to Fourier transforms and FFTs, in no particular order." `fftw.org`

- Roberta Piroddi and Maria Petrou. Non-Uniform Fourier Transform: A Tutorial
- `https://en.wikipedia.org/wiki/Non-uniform_discrete_Fourier_transform`

–Questions

- Sample a simple function (e.g. exp, sin, gaussian) in 1D or 2D. Then use some library to compute the FFT. Then compute the DFT in your own implementation. How close is the error? Work our the solution analytically for the continuous case. What should the answer be at some discrete points according to the continuous case, and how is it different from what the FFT gave? What is the typical floating point error? What was the speedup of the FFT, and how does this compare to the theoretical limit?

- In practice, how long are FFTs taking (2D and 3D)? How does this compare with other computational bottlenecks like disk I/O and interpolation?

- What are some numerical issues or bottlenecks that can arise when using an implementation of the FFT in practice? What are some ways to overcome them?

5.4   17 Dec 2020 - Blur - Convolution, Sampling, Nyquist

–Pre-reading

- Fourier transform: convolution, sampling and Nyquist (37 min)

- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks: 2-D Fourier Analysis, Convolutions. See the links at the bottom. For this week, work through notebook "ip_basics_part6".

–Questions

- Solve the "homework problem" (without peeking at the answer!) in the first part of Fourier transform: convolution, sampling and Nyquist. If $g(x) = e^{-\pi(x-2)^2} - e^{-\pi(x+2)^2}$, what is $G(u)$?

- By the Convolution theorem $f(x) = g(x) * h(x) \Rightarrow F(u) = G(u)H(u)$. Since $F(u) = G(u)H(u) = H(u)G(u)$ we should have $f(x) = g(x) * h(x) = h(x) * g(x)$. Show explicitly the commutativity of the convolution operator: i.e., $g(x) * h(x) = \int dt \ g(t)h(x-t) = \int dt \ h(t)g(x-t) = h(x) * g(x)$

- Look at the animation in the notebook "ip_basics_part6". Write down the mathematical object that is being visualized, including integration bounds. What does one snapshot of the animation represent?

- When does convolution (physically or computationally) happen in cryoEM happen? Be specific when connecting things back to the math, e.g. what is being convolved, what is it's functional form, and how many dimensions are involved?

- When we convolve with a *broad/wide* gaussian function, what is happening in Fourier space? How does this relate to Fourier filtering? Hint, use the convolution theorem.

5.5   7 Jan 2021 - Sharpen - Convolution, Sampling, Nyquist

–Pre-reading

- Fourier transform: convolution, sampling and Nyquist (37 min)

- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks. For this week, work through notebook "ip_basics_part6".

–Questions

- Around the 16 min mark in Fourier transform: convolution, sampling and Nyquist, we see that the de-convoluted (recovered) signal blows up around the origin, where there is a delta function. From the math, why exactly did this happen? Is there a way to over come this blowing up effect?

- Let $f_a = e^{-ax^2}$; $a > 0$. Show that $f_1 * (f_2 * f_3) = (f_1 * f_2) * f_3$. Note that $\int_{-\infty}^{+\infty} dx\ e^{-ax^2} = \sqrt{\pi/a}$

- The notebook "ip_basics_part6" has an interactive where the size of the kernel (Sobel, Gaussian, etc) can be changed. Convoluting with larger kernels could be a more expensive computation, but we could speed it up by doing this in Fourier space by the convolution theorem. If we take the FT of the image and kernel, there sizes to not match, so how can we multiply them element wise? How would you match the pixel sizes and do things in Fourier space to achieve the same result as the real space convolution?

- In the last part of "ip_basics_part6", the Fourier filtering versus the convolution filtering appear the same for Guassian blur, but different for box blur, and very different for sharpen. What might be causing this? Vary the kernel size and notice the different run times. Is doing things in Fourier space always quicker?

- How can we increase the Nyquist frequency during data collection to get higher resolution information? What is the trade off? What should guide our choice of an optimal Nyquist, given our particular microscope and specimen?

- Code up a simple example illustrating the convolution theorem, where you also actually do the convolution in real space. How close is it to the answer where you did the multiplication in Fourier space? What was the speedup? Now try speeding up the calculation by of the convolution in real space by making the convolution kernel smaller. How good of an approximation is this? Use a 2D projection of a 3D density map and a meaningful kernel (e.g. low pass filter) to build your intuition in a useful way.

5.6   14 Jan 2020 - Blur - Phase-contrast in the EM

–Pre-reading

- Phase-contrast imaging in the EM' (10 pages)

- Interactive coding notebook 'Practical 1 - Fundamentals of Image Processing' in the High-resolution imaging course at TUDelft. Note there are multiple notebooks. Go to the one on the CTF ('ip_basics_part7').

–Questions

1. See panel (b) in the 'signal star' image. What would this image look like if the CTF was applied (i.e. through convolution)?

2. As defined in equation 11 of 'Phase-contrast imaging in the EM', what happens when the B factor is negative versus positive?

3. In practice, on the scope, the CTF is easier to see in an FFT if you are at high mag. Why? Can you relate this back to the math? How is magnification playing into the equation?

4. Can you tell the difference between underfocus and overfocus by eye? Try putting what you see into words? Why do they look different?

5. As defined in equation 11 in 'Phase-contrast imaging in the EM', what physical behaviour is involved in this B factor? What other factors of B are there in cryoEM and what physical behaviour is involved in those B factors? Why are they all called B-factors?

6. With reference to equation 11 in 'Phase-contrast imaging in the EM': "B has units of nm^2 or $\mathring{A}^2$ and is called the 'B-factor' or 'envelope factor'. Good cryo-EM images have B values of $50 - 100 \mathring{A}^2$, but even these values are not so ideal. At $B = 100 \mathring{A}^2$ spatial frequencies of $5 \mathring{A}^2$ are attenuated to 1/e of their original amplitude, and higher spatial frequencies are attenuated even more". Go to 'ip_basics_part7' and play with B. Is there agreement between the B factor explanation and the interactive notebook? Take a closer look at how the interactive notebook codes up the B factor and what pixel size is used/assumed.

7. Play with the amplitude contrast. Do you 'want' more or less amplitude contrast? If the amplitude contrast is off, how will this effect CTF fitting? Can you think of a good reason why many CTF estimators do not fit amplitude contrast? If you are unsure as to the amplitude contrast, what should you do?

8. Play around with 'plot_ctf_img_interactive' in Question 4 of 'ip_basics_part7'. Turn on 'ctf_multiply'. Can you see how the information is delocalized more when the defocus increases? How defocused can the image get? Where does the information delocalize to? When collecting data on a real micrograph (versus computationally applying the CTF to a synthetic image the size of a particle), what issues would you have?

5.7 21 Jan 2021 - Sharpen - Phase-contrast in the EM

–Pre-reading

- Philippsen et al. 2007. "The contrast-imaging function for tilted specimens" `https://www.sciencedirect.com/science/article/pii/S0304399106001525`

- Voortman et al. 2011. "A fast algorithm for computing and correcting the CTF for tilted, thick specimens in TEM" `https://www.sciencedirect.com/science/article/pii/S0304399111000878`

- Voortman et al. 2012. "Fast, spatially varying CTF correction in TEM" `https://pubmed.ncbi.nlm.nih.gov/22728402/`

5.8 28 Jan 2021 - Blur - Image Formation (Forward Model)

–Pre-reading

- NCCAT SPA short course 2020, Lecture 4: Algorithms and foundational math Part I & 2, Fred Sigworth (1:18:10 - 1:28:46, 10 min)

–Questions

1. How is real noise different from Gaussian additive noise? What makes it different?

2. If particles come from different types of grids (thin continuous carbon, single layer graphene) can they be combined into one large dataset to solve a structure?

3. Why don't researchers pool different datasets of the same particle (e.g. from different EMPIAR entries) to solve a high resolution structure with tens of millions of particles?

5.9 11 Feb 2021 - Sharpen - Image Formation (Forward Model), Multislice

–Pre-reading

1. 'Theory' section in Koeck, P. J. B., & Karshikoff, A. (2015). Limitations of the linear and the projection approximations in three-dimensional transmission electron microscopy of fully hydrated proteins. Journal of Microscopy, 259(3), 197-209. `http://doi.org/10.1111/jmi.12253`

2. Section '2. Theory' in Vulović, M., Ravelli, R. B. G., van Vliet, L. J., Koster, A. J., Lazić, I., Löcken, U., Rieger, B. (2013). Image formation modeling in cryo-electron microscopy. Journal of Structural Biology, 183(1), 19?32. `http://doi.org/10.1016/j.jsb.2013.05.008`

   - Supplementary material associated with the article (18 pages)

3. 6.2 The Wave Equation for Fast Electrons in 'Advanced Computing in Electron Microscopy', Kirkland (2020), pp. 156-159.

4. 6.4 The Multislice Solution in 'Advanced Computing in Electron Microscopy', Kirkland (2020), pp. 162-165.

 –Questions

1. Equation 7 in Vulović et al. shows the DQE/NTF being applied. In practice, the DQE/MTF/NTF is often plotted in 1D in papers. What would be the difference between doing the convolution of the DQE/NTF in 2D versus 1D?

2. Equation 5 in Vulović et al. shows the 'complex CTF'. What is the difference between this and the sin CTF that is often used? When is it appropriate to use one versus the other treatment?

3. Simpler models assume additive gaussian noise and pixel independence, for instance see NCCAT SPA short course 2020, Lecture 4: Algorithms and foundational math Part I & 2, Fred Sigworth (1:18:10 - 1:28:46, 10 min). Is this the case in Vulović et al.? If so, what sorts of models would break the pixel independence?

4. For the sake of argument, let's say that our goal is to use more accurate forward models (multislice based, higher order series expansions phase and amplitude object), instead of projection. In practice, how would one go about doing 3D reconstruction with these more accurate forward models? What would be the computational bottlenecks in doing so?

5. What is the difference between the 'weak approximation' and the 'projection assumption'? See the discussion in Koeck & Karshikoff (2015), 'Multisclie and single projection simulations of image formation with amplitude and phase contrast', p 204. Go through the derivation in Koeck & Karshikoff (2015), and point out where precisely the assumptions are made.

5.10 18 Feb 2021 - Sharpen - 2D Expectation-maximization

–Pre-reading

1. Eddy, S. R. (2004). What is Bayesian statistics? Nature Biotechnology, 22(9), 1177?1178. `http://doi.org/10.1038/nbt0904-1177`

2. Nelson, P. C. (2019). Chapter 12 : Single Particle Reconstruction in Cryo-electron Microscopy. In Physical Models of Living Systems (pp. 305?325).

3. Interactive coding notebook 'Practical 2 - 2D/3D reconstruction' in the High-resolution imaging course at TUDelft.

–Questions

1. The primer on Bayesian statistics by S. Eddy concludes: "Using Bayesian methods, we can instead integrate over varying degrees of uncertainty in different aspects of the analysis." What is uncertain in the different aspects of cryoEM analysis? What would be good priors for each of these aspects? In practice, are people taking into account the uncertainty of these parameters, or using point estimates? Why?

2. Some popular cryoEM software does 2D classification in Fourier space. Why do these equations work in Fourier space? What would have to be changed?

3. In Nelson's treatment, there was only one 2D class. To do multiple classes, what would have to be changed? What is happening during a typical 2D classification 'round'? What is constant during a round and what is updated between rounds?

4. How does a prior affect the Log loss function? What functional forms of the priors are convenient for optimization, and what is their interpretation in terms of probability?

5. It can be shown that iid Gaussian additive noise in real space is also Gaussian additive noise in Fourier space. See this dsp.stackexchange.com post for example. Try to generalize this somehow. What happens for iid Gaussian additive noise that is radially dependent in Fourier space, or pixel dependent in real space?

5.11   25 Feb 2021 - Blur - Coordinate Systems & Rotations

–Pre-reading

1. Omega Open Course. 2017. Spherical Coordinate System (With 3D Animation)

2. Andrew Dotson. 2018. Deriving Spherical Coordinates (For Physics Majors)

3. Quine, JR. 4.1.2 Two dimensions, 4.1.3 Three dimensions, 4.2 Complex form of a rotation in *Mathematical Techniques in Structural Biology*, pp. 26-28

4. `https://en.wikipedia.org/wiki/Euler_angles#Rotation_matrix`

5. How is Orientation in Space Represented with Euler Angles?

6. Gregory G. Slabaugh. Computing Euler angles from a rotation matrix

 –Questions

1. Problems 4, 6, 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.

5.12   4 March 2021 - Sharpen - Coordinate Systems & Rotations

–Pre-reading

1. Axis/Angle Representations for Rotations

2. Quine, JR. 4. Orthogonal Transformations and Rotations in Mathematical Techniques in Structural Biology, pp. 25-31

3. Visualizing quaternions (4d numbers) with stereographic projection

4. Quaternions and 3d rotation, explained interactively

5. Visualizing quaternions An explorable video series

6. Wiki: Quaternions and spatial rotation

–Questions

1. all problems from Quine, JR. 4.5 Problems in *Mathematical Techniques in Structural Biology*, pp. 30-31

2. 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.

3. Quaterions can be used to describe the rotation of the point $r = \langle r_x, r_y, r_x \rangle \in \mathbb{R}^3$ about angle θ in the direction of vector $\langle q_x, q_y, q_z \rangle$ as follows. $p_{rotated} = q * p * q^{-1}$, where $q = \langle \cos(\theta/2), q_x \sin(\theta/2), q_y \sin(\theta/2), q_j \sin(\theta/2) \rangle$ 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 $\theta = 180°$ (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.

4. 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?

5. 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?

5.13   11 March 2021 - Blur - Interpolation

–Pre-reading

1. 2016. Resizing Images - Computerphile (bilinear and nearest-neighbour interpolation)

2. 2016. Bicubic Interpolation - Computerphile

3. https://en.wikipedia.org/wiki/Interpolation

 –Questions

1. Go to the Wiki page for interpolation and look at the pictures illustrating different interpolation schemes: nearest-neighbour, linear, polynomial, spline; and 1D vs 2D. How would these schemes differ in terms of "accuracy, cost, number of data points needed, and smoothness of the resulting interpolant function"?

2. Take a look at scipy.interpolate.griddata. When we are interpolating in Fourier space in 3D, what is points, values, and xi? How big are they?

3. Does interpolating with complex numbers cause any problems?

4. Search through the code of some cryoEM software packages and find where they are doing interpolation. You could look for the forward model (simulating data) or for the interpolating of the 3D Fourier map before taking the inverse Fourier transform to get the map in real space. What are the dimensions? Can you tell if it is linear interpolation, or some other method?

5.14   18 March 2021 - Sharpen - Interpolation

–Pre-reading

1. Explanation and Java/C++ implementation of (bi)cubic interpolation

2. `https://en.wikipedia.org/wiki/Bicubic_interpolation`

3. Application of interpolation to elevation samples

 –Questions

1. Go through the source of scipy.interpolate.griddata . What is the bottle neck for interpolation? You can try to read through the code, and also empirically benchmark things (e.g. sample a 2D or 3D Gaussian, which will give you real values, then compute the Gaussian function at these points, and interpolate on a cartesian grid at a specified granularity).

2. Code up your own implementation of some sort of interpolation, that works in a similar way as scipy.interpolate.griddata, ie data point coordinates, data values (at those coordinates), and points at which to interpolate data. What is the bottle neck? How would you speed it up?

3. Investigate how sensitive scipy.interpolate.griddata is to noise, by sampling a function, corrupting it with noise, and seeing how smooth things have to be for the interpolation to work out well. Can you find a situation where there are pathologies?

4. How is scipy.ndimage.map_coordinates different from scipy.interpolate.griddata?

5.15   25 March 2021 - Blur - Guest lecture: Dr. David Dynerman, 3D Reconstruction via Direct Fourier Inversion.

Dr. David Dynerman is currently the Chief Scientific Officer at The Public Health Company. Prior to this he lectured on cryoEM at Berkeley, where he also worked with Eva

Nogales (Molecular & Cell Biology) and Bernd Sturmfels (Mathematics) to develop new algorithms for 2D → 3D reconstruction in cryo-electron microscopy. His lecture will explain direct Fourier inversion (2D → 3D reconstruction).

–Audience
This lecture will be structured so that beginners will (hopefully) learn some useful basics, while folks with more background will gain an orientation of the topic with road signs on where to dive in for more detail.

–Learning objectives

1. understand the basic principles of how direct Fourier inversion produces 3D reconstructions from 2D images

2. understand the basics of two key technical aspects of Fourier inversion: Fourier interpolation and back projection

3. understand what kinds of map artifacts can arise during direct Fourier inversion

–Pre-reading

1. Part 2: 3-D Waves and Transforms - G. Jensen (13 min)

2. Part 4: 3-D Reconstruction - G. Jensen (11 min)

–Questions

1. How could we visualize a single particle slice in a 3D map - either in real space or reciprocal space? How would this type of visualization help you? What other visualizations would be suitable?

2. How does understanding Fourier interpolation and back projection help use data processing software, refine high quality maps, and solve structures?

3. Do you think Fourier inversion will always be used in cryoEM data processing, or will other methods eventually replace it?

4. What other research fields also use similar techniques (medical imaging, materials science, etc), and how much do practitioners in those fields understand about Fourier inversion?

5. How much do you want/need to know about Fourier inversion, and what will help you build intuition?

–Meeting Recording

https://utoronto.zoom.us/rec/share/iUQlpdJvEwJ4Z-GU3bQjJzccyPSJg0iocKhN5RFwyHXMKkuIpR5I3alav3UKfk4T.hA7WqNfNCNPpbqN7

Access Passcode: 041&NOnv+F

5.16   1 Apr 2021 - Sharpen - 3D Reconstruction via Direct Fourier Inversion.

–Learning objectives

1. understand implementation details of 2D $\rightarrow$ 3D Fourier inversion

2. understand what kinds of map artifacts can arise during Fourier inversion

3. play with Fourier inversion of simple shapes, to build intuition

–Pre-reading

1. The Bouman Lectures on Image Processing. A Slecture by Maliha Hossain. Subtopic 6: Fourier Slice Theorem. 2013 (notes and 25 min video)

2. ASTRA Toolbox: Fourier Slice Theorem (3 min)

3. ASTRA Toolbox: Proof of Fourier Slice Theorem (2 min)

4. ASTRA Toolbox: Filtered Backprojection (FBP) (4 min)

–Questions

1. Go to this coding notebook on Fourier inversion. Play around with things to build your intuition. You can look into the code that is doing the Fourier transforms or interpolation to see what is "inside the black box". Are there any artefacts in the reconstruction? What happens where there are less particles, errors in the alignments?

2. Let's say you are trying to do a reconstruction, and you see some artefacts that are difficult to interpret, but that might be caused by angular estimates, preferred orientations, contrast/SNR is not good enough - or a combination of all of those causal factors. How could you play around with a synthetic example to track down what is causing artefacts in your reconstruction?

–Meeting Recording

`https://utoronto.zoom.us/rec/share/I-d6u1KSWACwZrOfwSBjOVGFHTs2DIUeNn2U1VhiNeLd-onFbd7Z3AQHeTq_Imdb.taZbZeiA_6tTUTZa`

Access Passcode: `qdF^I7.^c9`

5.17   8 Apr 2021 - Sharpen - Variational Autoencoders

–Learning objectives

1. understand the theory of variational autoencoders, in order to intelligently use methods such as cryoDRGN

–Pre-reading

1. Variational autoencoder interactive demos with deeplearn.js

    (a) Doersch, C. (2016). Tutorial on Variational Autoencoders, 1-23. `https://arxiv.org/pdf/1606.05908.pdf` (23 pages)

2. Jeremy Jordan - Variational autoencoders

3. Intuitively Understanding Variational Autoencoders

–Questions

1. Describe the "latent space" in your own words. What other ideas does it relate to in structural biology?

2. Regarding the "reparameterization trick": what is being reparameterized and why did we need to do this "trick"?

3. Go through the figures in the blog post "Jeremy Jordan - Variational autoencoders" and explain the take home of each figure.

4. Why do *variational* autoencoders (vs. just plain vanilla autoencoders) allow us to smoothly interpolate between observations?

5. A variational autoencoder reconstructs its own input. I we are interested in reconstructing 3D structures from 2D cryoEM observations, how can we use variational approaches to generate 3D models? Consider how concretely we would compute reconstruction loss

6. Code up a variational autoencoder on some toy data. For example in Keras or PyTorch

5.18 15 Apr 2021 - Sharpen - Guest Lecture with Ellen Zhong (MIT), cryoDRGN & Variational Autoencoders

–Pre-reading

1. cryoDRGN tutorial

2. Ellen Zhong. 2020. CryoDRGN: Deep generative models for reconstructing heterogeneous protein structures from cryo-EM. SBGrid Consortium (40 min).

3. Zhong, E. D., Bepler, T., Berger, B., & Davis, J. H. (2021). CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. Nature Methods, 18(2), 176-185. `http://doi.org/10.1038/s41592-020-01049-4`

(a) Bepler, T., Zhong, E. D., Kelley, K., Brignole, E., & Berger, B. (2019). Explicitly disentangling image content from translation and rotation with spatial-VAE, (NeurIPS 2019).

(b) Zhong, E. D., Bepler, T., Davis, J. H., & Berger, B. (2019). Reconstructing continuous distributions of 3D protein structure from cryo-EM images. ICLR 2020, 1-20. https://arxiv.org/pdf/1909.05215.pdf

–Questions

1. Look through the "Log" from the command

```
cryodrgn train_vae data/128/particles.128.mrcs \
        ——ctf data/ctf.pkl \
        ——poses data/poses.pkl \
        ——zdim 8 \
        −n 50 \
        ——uninvert−data \
        −o tutorial/00_vae128 > tutorial_00.log
```

   Take a deep breath and go through it line by line calmly and patiently. What do you understand? What is confusing? Using the Log output as a guide, can you draw a schematic picture of the architecture?

2. The tutorial suggests 50 epochs for training. Can you learn anything from the other epochs? What would indicate that you should run more/less epochs?

3. Consider the figure with the caption "kmeans20 volumes: vol_000.mrc to vol_019.mrc". What do these results tell you about the particles? What would be some next steps upon seeing these results?

4. Compare the location of the 20 density maps in PCA vs UMAP space. How are they the same and how are they different?

5. The tutorial notes that "Principal component trajectories can also produce nonphysical "motions", e.g. if there is discrete variability in the structure." What evidence form cryoDRGN supports discrete vs continuous variability? What role can domain knowledge play here?

6. The tutorial explains how to filter out junk. What other ways are there to filter out junk and how does cryoDRGN compare with these? When would you use one over another?

7. Look at "rhe resulting 6 clusters" from the GMM. On the right panel, why are there particles scattered/delocalized? Carefully read how the GMM in the tutorial was

done. It might help to look up UMAP (For the curious I would recommend this SciPy talk by one of the original UMAP authors). Note that this is a sort of "trick question".

8. The tutorial reports that "in practice, we often find that junk particles are located as outliers in the distribution of latent encodings." Why do you think this happens? What rules of thumb should you use to decide how many junk particles you should set aside. What problems might this cause, and how could you check if these problems were occurring?

9. It's nice to have fast code, but day and week long experiments put things in perspective. What counts as "long" and "short" run times for data processing. How long are you willing to wait. Be honest, because this can help developers provide methods that people end up using in practice.

10. Compare the density maps with junk at 128 pixels vs high resolution without junk at 256 pixels.

11. Look at the high resolution "PC trajectories" (movies). In your own words, how would you describe this trajectory? What might the biological meaning of it be? Non-biological? What work would we have to do to untangle biological from non-biological conclusions?

12. Look at the "plot of the learning curve (i.e. average loss per epoch)". What can we conclude from this graph? What else might it look like, and what should we do in that case?

13. Let's say you are interested in determining to what extent conformations are discrete or continuous (e.g. an ion channel opening). Describe a workflow, including the generation of an initial reference (volume, pose, ctf), to investigate this. What steps would be involved, what would you look for at each step, and about how long would each step take.

14. Look at how the reconstruction changes with the number of particles. Use the lasso tool to select small and smaller subsets and compare by eye in a 3D viewer.

15. "CryoDRGN's graph traversal algorithm builds a nearest neighbor graph between all the latent embeddings, and then performs dijkstra's algorithm to find the shortest path on the graph between the anchors nodes." What biological meaning might such a path have? What other features of a path would have biological (or should we say physical) meaning? How might we extend Dijkstra's algorithm to find other interesting paths?

16. Let's assume that our sample was experiencing conformational heterogeneity in a defined way: with one degree of freedom, but with two Gaussian distributed clusters along this degree of freedom. For example a protein opening and closing, where there is a Gaussian distribution around the open and closed state, along the coordinate connecting the two. If we used cryoDRGN on such a dataset, what would our results look like? How would things change as the spread of the Gaussians increased so that they bled into each other?

17. A cryoEM dataset is frozen in time, and represents a time average. Sometimes we are interested in inferring the temporal sequence of events, for instance consider the conformational states of Kinesin. But is a temporal sequence always obvious from different states? Imagine different scenarios that might be easier and harder to infer a temporal sequence from the type of data available in cryoEM, and the analysis offered by cryoDRGN.

–Meeting Recording

https://utoronto.zoom.us/rec/share/Zh_SoIGK_JzYBpWjMuOK_oVNIZeBkjBFhjuskbTq-fmxwK3JyIX8LHmG7O2xeo6j.fkEa4BQKud093XyX

Access Passcode: 4nyaZv8AW#

5.19   22 Apr 2021 - Blur/Sharpen - Office Hours

–Pre-reading

1. S2C2 CryoEM Image Processing Workshop: Day 1 (5 hrs). See schedule at the 9:30 min mark. Have a look at:

    (a) 1:01:06 Import movies
    (b) 1:30:55 Motion correction
    (c) 2:00:14 CTF estimation
    (d) 2:17:52 Exposure curation
    (e) 2:39:41 Particle picking
    (f) 3:33:10 2D classification

–Questions

1. Some data processing algorithms have many parameters, and it can be difficult to understand what exactly they are all doing, which ones go together, and when you would use them, how they would change things, and when you'd know that things are working. Find some parameters in a data processing algorithm that you don't understand, and bring it to the office hours to discuss. If you have time, run some jobs by changing a few parameters.

2. Some people would like to know more about how an cryoEM algorithm works, but can't follow the original paper describing. Bring some results you've generated from some cryoEM data processing algorithm to discuss.

3. Some data processing algorithms generate intermediate plots. If you know how to interpret these plots, they can help diagnose if the algorithm it working how you think it should be. Bring some plots that you have difficulty understanding to the office hours to discuss.

4. Some cryoEM discussion forums have "power users" that really understand things thoroughly. If you can follow how they are thinking, you can learn a lot, trouble shoot your data processing problems, and move your projects along. Find some posts and threads with these power users, and bring them to the office hours to discuss.

5.20 29 Apr 2021 - Blur - Guest Lecture with Dr. James Krieger (University of Pittsburgh), normal mode analysis & ProDy

–Pre-reading

1. Bahar, I., Lezon, T. R., Bakan, A., & Shrivastava, I. H. (2010). Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. Chemical Reviews, 110(3), 1463-1497. http://doi.org/10.1021/cr900095e

2. ProDy website

–Questions

1. Have a look at some ProDy tutorials

2. In reality, each single particle image is a slightly different conformation. Let's assume that individual particles are fluctuations around an average reference state, and that the fluctuations are described well by the 'large breathing motions' described by normal mode approaches. How does this type of heterogeneity affect the final reconstruction? If heterogeneous particles go into one structure, what might happen to their alignment estimation?

3. How can physics based models, the type ProDy uses, be incorporated into cryoEM data processing algorithms and downstream analysis?

–Meeting Recording

https://utoronto.zoom.us/rec/share/-nAL6ZajeveRYxzFGjtZZBd5DTlpCe1fDyT1xhtvDSGxxzh36gYMVOPJerrZGO3U.Iyo-PxJegZPkm7qH

Access Passcode: @biLOD3=tw

5.21 6 May 2021 - Sharpen - Guest Lecture with Dr. James Krieger (University of Pittsburgh), normal mode analysis & ProDy

–Pre-reading

1. "2. Theory" section (pp. 1470-1477) in Bahar, I., Lezon, T. R., Bakan, A., & Shrivastava, I. H. (2010). Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins. Chemical Reviews, 110(3), 1463-1497. http://doi.org/10.1021/cr900095e

2. Zhang, Y., Krieger, J., Mikulska-Ruminska, K., Kaynak, B., Sorzano, C. O. S., Carazo, J. M., ... Bahar, I. (2021). State-dependent sequential allostery exhibited by chaperonin TRiC/CCT revealed by network analysis of Cryo-EM maps. Progress in Biophysics and Molecular Biology, 160, 104-120. `http://doi.org/10.1016/j.pbiomolbio.2020.08.006`

3. ProDy source, especially dynamics and the Topology Representing Networks class.

–Questions

1. How can physics based models, the type ProDy uses, be incorporated into cryoEM data processing and analysis?

2. What are the computational bottlenecks (runtime, memory) in the types of approaches in prody.dynamics?

3. Look at the tutorial on Processing Cryo-EM Electron Density Maps. How many pseudo atoms were used to represent the map? Do you think this is reasonable? How many degrees of freedom to M pseudo atoms represent? How many degrees of freedom are needed to encode a map in an NxN real spaced grid? In a low pass Fourier space or Fourier cropped map (e.g. a k_x by k_y cube or sphere of radius $|k|$ pixels)?

5.22 13 May 2021 - Blur/Sharpen - Master class: preferred orientation. Yong Zi Tan (The Hospital for Sick Children, Toronto, Canada)

–Learning Objective:

• Understand, diagnose, and solve the problem of preferred orientation.

–The Problem

• What is preferred orientation? Why do samples suffer from preferred orientation?

1. Blog post by Nanoimaging Services on Preferred Orientation (easy to understand) https://www.nanoimagingservices.com/blogs/post/solving-the-challenges-of-preferred-orientation-in-cryoem-sample-preparation?utm_source=Zoho&utm_medium=social&utm_campaign=May3

2. Noble, A.J., Dandey, V.P., Wei, H., Brasch, J., Chase, J., Acharya, P., Tan, Y.Z., Zhang, Z., Kim, L.Y., Scapin, G. and Rapp, M., 2018. Routine single particle CryoEM sample and grid characterization by tomography. Elife, 7, p.e34257. `https://elifesciences.org/articles/34257`

–Diagnosing the Problem

- How do we know we have a preferred orientation problem? What is the consequence of it? How would we quantify the degree of preferred orientation?

  1. Baldwin, P.R. and Lyumkis, D., 2020. Non-uniformity of projection distributions attenuates resolution in Cryo-EM. Progress in biophysics and molecular biology, 150, pp.160-183. `https://www.sciencedirect.com/science/article/abs/pii/S0079610719301087`
  2. Tan, Y.Z., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carragher, B. and Lyumkis, D., 2017. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. Nature methods, 14(8), pp.793-796. `https://www.nature.com/articles/nmeth.4347`

- What is the 3DFSC? How does it relate to FSC and how does it diagnose preferred orientation issues?

  – **COMPLETE BEFORE MEETING**. Let's analyze some read data with the 3DFSC, using a webserver `https://github.com/geoffwoollard/learn_cryoem_math/blob/master/cryoem-group-study-meeting/20200513_practical_3DFSC.pdf`

–Solving the Problem

- What are the ways to mitigate preferred orientation?

- How does tilting mitigate preferred orientation? Pros and cons?

–Meeting Recording

`https://utoronto.zoom.us/rec/share/ly-8b5zAE_Ms0GDQq6nk0DMqpsVmF0rHH9hpc64oGlue9Cl07Ock5VlzD2YIFaBF.6eDHwBlsaOavK_7c`

Access Passcode: `1h5CGMTg%6`

5.23 20 May 2021 - Sharpen - Master class: preferred orientation. Philip Baldwin (Salk Institute for Biological Studies & Baylor College of Medicine).

–Learning objectives

1. Understand how sampling quantitatively affects resolution

2. Have a visual intuition for what parts of Fourier space are missing in different preferred orientation scenarios.

3. Understand the assumptions underling the sampling compensation factor (SCF), when they hold and when they break down.

–Pre-reading & Questions

- Baldwin, P.R. and Lyumkis, D., 2020. Non-uniformity of projection distributions attenuates resolution in Cryo-EM. Progress in biophysics and molecular biology, 150, pp.160-183. `https://www.biorxiv.org/content/10.1101/635938v2`. *Note that biorxiv preprint and Progress in biophysics and molecular biology article number the equations referred to below identically.*

  1. What is the Spectral signal-to-noise ratio (SSNR) and how is it related to the FSC? For typical FSC curves (sigmoid like), what would the SSNR look like?
  2. Why does the SSNR not increase with more data when there is preferred orientations, and what effect this has on the FSC?
  3. Explain the assumptions going from Eq. 2.10 to 2.11 in your own words.
  4. During iterative refinement, how would you compute the pieces that go into Eq 2.13: $E^2(k), \langle |X|^2 \rangle, \langle |N(\vec{k})|^2 \rangle, \langle \frac{1}{Sp(\vec{k})} \rangle$
  5. What mathematical steps did you not understand, either mathematically, or the physical assumptions justifying certain mathematical leaps?

- 3DFSC webserver `https://github.com/geoffwoollard/learn_cryoem_math/blob/master/cryoem-group-study-meeting/20200513_practical_3DFSC.pdf`

  1. Submit your own job to the 3DFSC webserver (halfmaps, full map, pixel size). Open the .cmd file in Chimera and have it ready for the meeting.

- Baldwin, P. R., & Lyumkis, D. (2021). Tools for visualizing and analyzing Fourier space sampling in Cryo-EM. Progress in Biophysics and Molecular Biology, 160, 53-65. `http://doi.org/10.1016/j.pbiomolbio.2020.06.003`

5.24 27 May 2021 - Blur - Fourier transforms and reciprocal space for the beginner (Grant Jensen's course)

–Pre-reading

1. Part 2: 2-D Waves and Images - G. Jensen (19 min)

2. Part 2: 2-D Transforms and Filters - G. Jensen (33 min)

–Questions

1. In Part 2: 2-D Waves and Images, Grant Jensen shows some visualizations of 2D waves. He enters values for h, k, amplitude, and phase, and a wave appears on the screen. Let's say we wanted to "spin a wave around" the origin. For instance make 360 waves that were each rotated by 1 degree. What values would we enter?

2. If we took the Fourier transform of one of the 2D waves in Part 2: 2-D Waves and Images, what would it look like? What happens as the wave is translated?

3. Sometimes people see "water rings" in the power spectra during CTF estimation. How exactly does it arise, and why is it a ring? What makes it more or less bright? What controls the ring's width/spread?

4. How would you make a checker-board-like pattern (e.g. 8x8) with Jensen's wave program?

5. If we Fourier crop (hard low pass filter) a 3D map (256 pixels, $1\mathring{A}$ / pixel) at $8\mathring{A}$, how many "waves" go into the map?

6. Around 15:20 in Part 2: 2-D Transforms and Filters Jensen remarks that the location of pixels in Fourier space is perpendicular to the location of the crest of the waves in real space. If we are on the scope and take a live FFT of a gridded pattern (pieces of a lattice, rotated in all sorts of directions), we see a ring. What happens to the ring in the FFT when we translate the stage in one direction.

7. Play around a coding notebook on the discrete Fourier transform I made here: `https://github.com/geoffwoollard/learn_cryoem_math/blob/98cce6f1142af0aea8647f7a488d899f7336 nb/dft_sandbox.ipynb`

–Discussion group leaders

- Arunabh Athreya

5.25   3 June May 2021 - Blur/Sharpen - Office Hours

–Pre-reading

1. S2C2 CryoEM Image Processing Workshop: Day 2 Part 1 (3h:20m)

–Questions

1. Some data processing algorithms have many parameters, and it can be difficult to understand what exactly they are all doing, which ones go together, and when you would use them, how they would change things, and when you'd know that things are working. Find some parameters in a data processing algorithm that you don't understand, and bring it to the office hours to discuss. If you have time, run some jobs by changing a few parameters.

2. Some people would like to know more about how an cryoEM algorithm works, but can't follow the original paper describing. Bring some results you've generated from some cryoEM data processing algorithm to discuss.

3. Some data processing algorithms generate intermediate plots. If you know how to interpret these plots, they can help diagnose if the algorithm it working how you think it should be. Bring some plots that you have difficulty understanding to the office hours to discuss.

4. Some cryoEM discussion forums have "power users" that really understand things thoroughly. If you can follow how they are thinking, you can learn a lot, trouble shoot your data processing problems, and move your projects along. Find some posts and threads with these power users, and bring them to the office hours to discuss.

5.26   10 June 2021 - Blur - Fourier transforms and reciprocal space for the beginner (Grant Jensen's course)

–Pre-reading

1. Part 2: 3-D Waves and Transforms - G. Jensen (13 min)

2. Part 2: Convolution and Cross-Correlation - G. Jensen (15 min)

–Questions

1. The Guinier plot fits a slope and reports a B factor. What is on the y-axis, and what is on the x-axis. How do you compute this from a map?

2. Each data point from the Fourier shell correlation curve arises from a comparison of half maps at a given resolution shell. What does a Fourier shell look like in real space? How can making artificially inflate the FSC?

3. Play with a coding notebook I made here: `https://github.com/geoffwoollard/learn_cryoem_math/blob/0d23f7b914ef2bfbcc616e05d2dbc0fa27eaf38d/nb/guinier_fsc_sharpen.ipynb`

5.27   17 June 2021 - blur - Glaeser, Nogales & Chiu (GNC), 1 Introduction and Overview (1.1, 1.2)

–Pre-reading

- Nogales, Eva. (2021). *"Section 1.1 Visualizing biological molecules to understand life's principles* in Glaeser, R. M., Nogales, E., & Chiu, W. (Eds.). (2021). Single-particle Cryo-EM of Biological Macromolecules. IOP Publishing. http://doi.org/10.1088/978-0-7503-3039-8 (abbreviated as GNC)

- Glaeser, Robert. (2021). *"Section 1.2 Recovery of 3D structures from images of weak-phase objects"* in GNC

–Questions

1. Nogales mentions several unique capabilities of cryo-EM

   (a) samples not amenable to other structural biology techniques (polymers and viruses, integral membrane proteins),

   (b) samples where that have properties that makes them amenable to cryoEM (large assemblies)

   (c) samples with a challenge that poses no problem to cryoEM: scarce samples

   (d) areas where data processing can help: compositionally heterogeneous samples, conformationally complex samples

   Discuss some interesting examples that touch on these areas (e.g. work out of your lab, interesting projects you've come across).

2. Structural biology techniques (CryoEM, NMR, SAXS, X-ray crystallography, etc) can be used on the same system to answer different but intertwined questions. Mention some research projects that used multiple structural biology projects, and what biological insight each technique provided.

3. Are there any themes/topics/breakthroughs that you thought should have been included in Nogales overview of cryoEM?

4. An electrons phase changes when it passes through the sample. What is the physical cause of these phase modulations? What would properties of the sample would make the phase modulation be higher or lower?

5. Eq. 1.6 in Glaeser's chapter backs up an approximation of the image wave function that preserved linearity. What exactly is ϵ and why is it small? Look at Eqs. 1.2 and 1.3. What physical assumptions does this underly.

6. For the values in equation 1.3 and 1.4, what can we control experimentally and what is just a physical constant (intrinsic material property of atoms, etc).

7. Look up some reference(s) from the chapter in an area that interests you. What did you learn. If you were writing the chapter, would you have used this reference differently?

8. What sentences in Glaeser's chapter do you not understand? Take note of this and ask for clarification in the meeting or in the Slack group.

5.28 24 June 2021 - blur - Glaeser, Nogales & Chiu (GNC), 2 Sample Preparation Introduction and Overview (2.2 - 2.5)

–Pre-reading

- GNC Chapter 2 Sample Preparation Introduction and Overview (sections 2.2 - 2.5)

–Questions

1. Section 2.2: Do you think negative stain is overused or underused among the people you work with? How do you use it?

2. Section 2.2: Write down the new things you learned about negative stain from this chapter in an email and send it to your cryoEM friends who would appreciate it.

3. Section 2.3: Did you ever use lacey grids, or self-wicking grids? How did they compare with holey grids?

4. Section 2.3: How did you go about optimizing glow discharge? What did you try and what did you look for? How reproducible were the results? Have you done something fancy with treatments of different gasses?

5. Section 2.3: How reproducible is vitrifying for you? Do vitrifying devices behave differently (see Fig 2.6)? Have you noticed any patterns?

6. Section 2.4: At what resolutions does correction for the curvature of the Ewald sphere help? What does this have to do with particle size?

7. Section 2.4: What weird things have you seen in micrographs, and what do you think is causing them? How could you do some measurements on the microscope to track down what is physically happening?

8. Section 2.5: What causes the different behaviour at the support water interface (SWI) versus the air water interface (AWI)? Why do proteins not unfold as much at the SWI versus the AWI?

9. Section 2.5: In Fig 2.10 the edge of the hole is a few lengths the size of the specimen. Is this to scale? How thick is the holey pattern typically?

10. Section 2.5: Since optimization of sample preparation and vitrification conditions is sample dependent, it's hard to develop heuristic "rules-of-thumb". However, perhaps overall strategies can be generalized? What did you try, and in what order? How long did you spend optimizing certain aspects, and what results helped you to move forward?

11. Section 2.5: Do you have any experience with the devices mentioned in "Table 2.2: Cryo-EM sample preparation devices currently under development"?

5.29   3 June May 2021 - Blur/Sharpen - Office Hours

–Pre-reading

1. S2C2 CryoEM Image Processing Workshop: Day 2 Part 2 (3h:40m)

–Questions

1. Some people would like to know more about how an cryoEM algorithm works, but can't follow the original paper describing. Bring some results you've generated from some cryoEM data processing algorithm to discuss.

2. Some cryoEM discussion forums have "power users" that really understand things thoroughly. If you can follow how they are thinking, you can learn a lot, trouble shoot your data processing problems, and move your projects along. Find some posts and threads with these power users, and bring them to the office hours to discuss.

5.30   8 July 2021 - sharpen - Journal club: multisclice. Ben Himes (HHMI Janelia Research Campus)

–Pre-reading

- Himes, B. A., & Grigorieff, N. (2021). Cryo-TEM simulations of amorphous radiation-sensitive samples using multislice wave propagation. BioRxiv.
  `http://doi.org/10.1101/2021.02.19.431636`
  `https://www.biorxiv.org/content/10.1101/2021.02.19.431636v2`

–Questions

1. What are the extensions of this paper compared to past treatments?

2. What is not modelled?

3. What are the computational bottlenecks in simulating images using multislice?

4. Out of all the sources of noise, which ones dominate? How can this be tested/quantified?

5. Why is exact quantitative agreement important (proportionality constant equal to unity)?

6. Who is interested in using multislice simulators and why?

7. Was there any terms you did not understand? (bremsstrahlung, plasmon, etc)

8. Draw a picture illustrating the equations (especially Eqs. 2, 3 (the first one on line 118), 8, 10)

–Meeting Recording

https://utoronto.zoom.us/rec/share/_yXLHqvspeJ5iUGW5FHbTELQs6P9reUzpJiwM-tZK7PB6OAdslPXGac-y4eJhzJh.LvVIo5bYXXGzL40C

Access Passcode: `x5wP8$3$?j`