Online CryoEM Study Group

Geoffrey Woollard, Toronto, Canada

February 25, 2021

1   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.**

1.1   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. After carefully reading the responses, I think it would be best to have two streams each meeting. I have code named them the **blur** and **sharpen** stream, referring to map sharpening. I am counting on your sense of humour and humility for this playful naming schema! Perhaps its a bit more gentle than low and high resolution? 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 with perhaps not that much experience processing datasets from start to finish, 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.

   After reading through the responses a few times, I would estimate the ratio of the streams to be about 3/4 blur to 1/4 sharpen. In the first meeting you can choose your own stream, and go to breakout rooms with this in mind. I will make multiple breakout rooms so that there can be small group discussion. We can use the **white board and share screen features**. If you are an expert feel free to go to the blur stream and help out. If you judge yourself an intermediate-beginner, you can test the waters in the sharpen stream. The two stream format may evolve over time (e.g. three streams, a lecture with everyone together), but let's try out two streams and see how it goes.

1.2   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.

1.2.1   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."

1.2.2 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.

1.3 Meeting Format

We will start with the "flipped classroom" and "breakout room discussions" to get things off the ground with minimal overhead. After some initial meetings we will try out a mix of the following formats listed below, with their surveyed popularity listed next to them. **If you would like to volunteer to give a lecture or be available for office hours then please volunteer yourself through the Slack channel.**

- 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

- Breakout room discussion (small groups organized thematically)

- Office hours with an expert available to answer questions

2 Dates and Topics

Given the current distribution of our global audience (90%+ in North America and Europe), we will have the time at **9 AM PST, which is 12 PM EST, 5 PM GMT, 7 PM GMT+2, 1 AM next day GMT+8, 3 AM next day GMT+10**. I am open to proposals to occasionally have a different time to accommodate a global audience - especially if the presenter is from a less represented time zone.

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	8 AM PST	blur	interpolation
Th 18 March 2021	8 AM PST	sharpen	interpolation
Th 25 March 2021	8 AM PST	blur	Guest lecture: TBC D. Dynerman, 3D reconstruction via direct Fourier inversion
Th 1 Apr 2021	8 AM PST	sharpen	3D reconstruction via direct Fourier inversion
Th 8 Apr 2021	8 AM PST	sharpen	variational autoencoders
Th 15 Apr 2021	8 AM PST	sharpen	cryoDRGN (variational autoencoders) with Ellen Zhong
TBD	8 AM PST	blur/sharpen	Wah Chiu et al, lectures, ...
... 2021 ...	...	...	...

3 Organizational Team

I would appreciate help

- **Giving lectures. You can send a proposal to me including topic, format, time, date.**

- **Being the leader of a breakout room to answer questions. You could list your areas of expertise, so people would know to come to you for those things. A sort of 'office hours'.**

- Developing the syllabus, and suggesting (or making) learning content.

- Formulating challenge problems to test comprehension, build intuition, and conceptual understanding.

- Zoom co-host and admin help.

4   Slack

We will use the Slack channel 'cryoem_study_group' for asynchronous chat. The link to join is `https://join.slack.com/t/cryoemstudygroup/shared_invite/zt-lm8kj6z8-_Xkdk-QowyDD4ewOTrW0hQ`. Please join the Slack group and ask questions there, rather than emailing me, as I may not respond!

5   Syllabus

We will start off drawing heavily from the content developed by Dr Frederick Sigworth, `https://cryoemprinciples.yale.edu/video-lectures`. If you are having problems with links, then try viewing his content on YouTube.

Thanks to 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 visualise 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.

If you haven't already gone through Grant Jensen's popular online course 'Getting Started in Cryo-EM' , now would be a good time to do so. I think enough people have gone through this on their own that we will mainly draw from other material.

6 Upcoming Meetings

6.1   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.

6.2   4 March 2021 - Sharpen - Coordinate Systems & Rotations

–Pre-reading

- Axis/Angle Representations for Rotations

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

- Visualizing quaternions (4d numbers) with stereographic projection

- Quaternions and 3d rotation, explained interactively

- Visualizing quaternions An explorable video series

- Wiki: Quaternions and spatial rotation

–Questions

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

- 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.

- 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.

6.3   11 March 2021 - Blur - Interpolation

–Pre-reading

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

- 2016. Bicubic Interpolation - Computerphile

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

–Questions

- 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?

- Does interpolating with complex numbers cause any problems?

- 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?

6.4   18 March 2021 - Sharpen - Interpolation

–Pre-reading

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

- https://en.wikipedia.org/wiki/Bicubic_interpolation

- Application of interpolation to elevation samples

–Questions

- Go through the https://github.com/scipy/scipy/blob/v1.5.4/scipy/interpolate/interpnd.pyx 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).

- 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?

- 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?

6.5 25 March 2021 - Blur - TBC Guest lecture: Dr. David Dynerman, 3D Reconstruction via Direct Fourier Inversion.

6.6 1 Apr 2021 - Sharpen - 3D Reconstruction via Direct Fourier Inversion.

6.7   8 Apr 2021 - Sharpen - Variational Autoencoders

–Pre-reading

- Variational autoencoder interactive demos with deeplearn.js

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

- 015 Jensen's inequality & Kullback Leibler divergence (10 min)

- Ali Ghodsi, Lec : Deep Learning, Variational Autoencoder, Oct 12 2017 [Lect 6.2] (1 hr)

- Deep Learning Lecture 14: Karol Gregor on Variational Autoencoders and Image Generation (43 min)

- Aaron Courville. 2015. Variational Autoencoder and Extensions. Deep Learning Summer School, Montreal. (1.5 hrs)

–Questions

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

6.8   15 Apr 2021 - Sharpen - Guest Lecture with Ellen Zhong, cryoDRGN & Variational Autoencoders

–Pre-reading

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

- 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).

- 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`

- Nature Methods paper on cryoDRGN (to appear $\sim$ Feb 2021)

–Questions

- Install cryoDRGN and run it on your own data.

6.9   Spring 2021

By this time the anticipated textbook by Wah Chiu, Robert Glaeser, and Eva Nogales will hopefully be out, and we can see how it looks. We could also plan some lectures.

7 Past Meetings

7.1 19 Nov 2020 - Background Math & Defocus Phase Contrast

7.1.1 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?

7.1.2 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?

7.2   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: $\langle ..., [0, 1/128), [1/128, 2/128), ..., [\frac{128/2-2}{128}, \frac{128/2-1}{128}), [\frac{128/2-1}{128}, \frac{128/2}{128}) \rangle$.

7.3   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`

13

   – 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?

7.4   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.

7.5   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.

7.6   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.

16

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?

7.7 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/`

7.8 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?

7.9 11 Feb 2021 - Sharpen - Image Formation (Forward Model), Multi-slice

–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.

7.10 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?