

ASPIRE: Cryo-Electron Microscopy Image Processing in Python

Christopher Langfield

Program in Applied and Computational Mathematics
Princeton University
Princeton, NJ
langfield@princeton.edu

Joshua Carmichael

Program in Applied and Computational Mathematics
Princeton University
Princeton, NJ
carmichael@princeton.edu

Garrett Wright

Program in Applied and Computational Mathematics
Princeton University
Princeton, NJ
gbwright@princeton.edu

Joakim Andén

Department of Mathematics
KTH Royal Institute of Technology
Stockholm, Sweden
janden@kth.se

Amit Singer

Program in Applied and Computational Mathematics
Princeton University
Princeton, NJ
amits@math.princeton.edu

Abstract—An introduction to the mathematical problem of cryo-electron microscopy (cryo-EM) is given, along with an overview of ASPIRE, a software package for processing cryo-EM image data. ASPIRE uses unique Fourier basis representations for images of cryo-EM particles. The background and the implementation of these bases in the code is discussed.

I. INTRODUCTION

Cryo-electron microscopy (cryo-EM) is a molecular imaging technique used to estimate the three dimensional structure of a molecule or protein. [1] In cryo-EM, a sample is purified in an aqueous solution and then rapidly frozen into a thin layer. This process, which results in "vitreous" ice, ensures both that the sample is trapped in close to its physiological conformation and that no ice crystals form, which is crucial to the imaging process. [2]

The sample is then placed under an beam of electrons, whose trajectories are deflected and scattered by the electrostatic fields of the individual molecules, creating an image on the detector beneath. (Fig. 1)

In this poster we give a brief introduction to ASPIRE [8], a Python package for processing cryo-EM image data, and show an example of how software design principles are used to represent mathematical structures used in the algorithms.

II. SINGLE-PARTICLE RECONSTRUCTION IN CRYO-EM

The goal of single-particle reconstruction (SPR) is to estimate the electrostatic potential $\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ of the molecule given n projection images obtained using the method outlined above. The image formation model for the projection of some

particle I_i is assumed to be in the form of (1), where h is the point-spread function of the electron microscope and $R_i \in \text{SO}(3)$ is the unknown rotation matrix representing the given particle's orientation in the frozen sample. The line integral along one Cartesian direction is also known as "tomographic projection". The point-spread function h is more commonly referred to by its Fourier transform, called the Contrast Transfer Function (CTF). One of the principal challenges in cryo-EM reconstruction is the very signal-to-noise ratio of raw data. [3]

$$I_i(x, y) = h \star \int_{-\infty}^{\infty} \phi(R_i(x, y, z)^T) dz + \text{"noise"} \quad (1)$$

Several programs implement SPR pipelines, including particle-picking, denoising, CTF correction, and volume refinement steps, [4] [5] [6] and active research is still ongoing in the field. [7]

III. ASPIRE

A. Overview

ASPIRE (Algorithms for Single Particle Reconstruction) [8] is a Python package, installable from the Python Package Index (PIP), which aims to implement an end-to-end SPR pipeline. It uses novel methods for various steps including particle picking [9], CTF correction [10], particle orientation estimation [11], and image classification. [12] [13]

Many common cryo-EM programs, including the widely-used RELION [4], work from an initial guess at the three-

dimensional structure of a molecule, and then use maximum likelihood methods to refine this guess into a detailed result. [14] In addition to introducing new algorithms and research avenues, ASPIRE aims to contribute to the cryo-EM software ecosystem by providing a robust initial volume estimate directly from experimental data, a so-called *ab-initio* reconstruction which can be further refined with other tools.

IV. EFFICIENT REPRESENTATIONS OF PROJECTION IMAGES IN CUSTOM BASES

A crucial step in the single-particle reconstruction pipeline is to perform principal component analysis on the set of particle projection images. Since projections from similar viewing directions can appear with arbitrary in-plane rotations relative to each other, it is important to be able to efficiently rotate projection images to determine whether they belong in the same class. To accomplish this, ASPIRE represents images using "steerable" bases. [12] [13] Images are represented as a list of coefficients associated with a subset of radial and angular functions that span the unit disk in Fourier space. The properties of these bases allow the coefficients for a rotation of the original image to be quickly computed compared to rotating an image represented in Cartesian coordinates.

The Fourier-Bessel functions are a set of polar basis functions spanning the disk of radius c in Fourier space. (2)

$$\psi_c^{k,q}(\xi, \theta) = \begin{cases} N_{k,q} J_k(R_{k,q}\xi) e^{ik\theta} & \xi \leq c \\ 0 & \xi > c \end{cases} \quad (2)$$

Here J_k is the k 'th Bessel function of the first kind, $R_{k,q}$ is the q 'th zero of J_k , and $N_{k,q}$ is a normalization constant. Although this is theoretically an infinite series, an image can be efficiently represented with a truncated Fourier-Bessel sequence using bounds derived in [12] and [13]. Fig. 2 shows the Fourier-Bessel basis functions.

V. OBJECT-ORIENTED REPRESENTATIONS OF STEERABLE BASES

Several bases are used at different points in the SPR pipeline. Each basis is represented as an object in an inheritance structure descending from base class `Basis`. (Fig. 3)

`Basis` objects convert between Cartesian images and their basis coefficient representations via `evaluate` and `evaluate_t` methods, which are analogous to the linear transformations converting between the two coordinate systems. (Fig. 4)

Several considerations apply to the design of this inheritance structure. Firstly, while the two-dimensional Fourier-Bessel and Fast Fourier-Bessel bases have the "steerable" property mentioned above, their three-dimensional analogs do not. Specifically, this indicates the ability to rotate the image via complex exponential multiplication. However, all four Fourier-Bessel bases share certain methods related to computing Bessel function zeros and coefficients. Because of this, they should all share certain class characteristics.

The solution to this is to use a "mixin" class, `FBBasisMixin`, which can sit outside of the linear class

hierarchy and lend functionality to classes which may not be amenable to having the same place in that hierarchy. This creates a class structure which is extensible and adds flexibility for situations like that of the Fourier-Bessel bases.

ACKNOWLEDGMENT

The authors would like to thank all of the researchers, software engineers, and students who have contributed to ASPIRE. We would also like to thank the Research Software Engineering group at Princeton and Ian Cosden.

REFERENCES

- [1] F. J. Sigworth, "Principles of cryo-EM single-particle image processing," *Microscopy*, vol. 65, no. 1, pp. 57–67, 2016.
- [2] Bhella, D. Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility. *Biophys Rev* 11, 515–519, 2019.
- [3] Bendory T, Bartesaghi A, Singer A. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Process Mag.* 2020 Mar;37(2):58-76. doi: 10.1109/msp.2019.2957822. Epub 2020 Feb 27. PMID: 32395065; PMCID: PMC7213211.
- [4] Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol.* 2012 Dec;180(3):519-30. doi: 10.1016/j.jsb.2012.09.006. Epub 2012 Sep 19. PMID: 23000701; PMCID: PMC3690530.
- [5] Zhong, E.D., Lerer, A., Davis, J.H. and Berger, B., 2021. CryoDRGN2: Ab initio neural reconstruction of 3D protein structures from real cryo-EM images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4066-4075).
- [6] G. Tang, L. Peng, P.R. Baldwin, D.S. Mann, W. Jiang, I. Rees and S.J. Ludtke. (2007) EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol.* 157, 38-46. PMID: 16859925
- [7] A. Singer and F.J. Sigworth, *Computational Methods for Single-Particle Cryo-EM*, Annual Review of Biomedical Data Science, 3, pp. 163-190 (2020).
- [8] Garrett Wright, Joakim Andén, Vineet Bansal, Junchao Xia, Chris Langfield, Josh Carmichael, Robbie Brook, Yunpeng Shi, Ayelet Heimowitz, Gabi Pragier, Itay Sason, Amit Moscovich, Yoel Shkolnisky, and Amit Singer. (2022). *ComputationalCryoEM/ASPIRE-Python: v0.9.2 (v0.9.2)*. Zenodo. <https://doi.org/10.5281/zenodo.6607659>
- [9] A. Heimowitz, J. Andén, A. Singer, APPLE Picker: Automatic Particle Picking, a Low-Effort Cryo-EM Framework. *Journal of Structural Biology*, 204(2), pp. 215-227, 2018. PMID: 30134153 [PubMed] PMCID: PMC6183064.
- [10] A. Heimowitz, J. Andén, A. Singer, Bias and variance reduction and denoising for CTF Estimation, *Ultramicroscopy*, 212, 112950, 2020.
- [11] Y. Shkolnisky, A. Singer, Viewing Direction Estimation in Cryo-EM Using Synchronization, *SIAM Journal on Imaging Sciences*, 5 (3), pp. 1088-1110 (2012). PMID: 24363820 [PubMed] PMCID: PMC3868463
- [12] Z. Zhao, A. Singer, Fourier-Bessel Rotational Invariant Eigenimages, *The Journal of the Optical Society of America A*, 30 (5), pp. 871-877 (2013). PMID: 23695317 [PubMed] PMCID: PMC3711886.
- [13] Z. Zhao, Y. Shkolnisky, A. Singer, Fast Steerable Principal Component Analysis, *IEEE Transactions on Computational Imaging*, 2 (1), pp. 1-12 (2016). PMID: 27570801 [PubMed] PMCID: PMC4996495.
- [14] Sigworth FJ, Doerschuk PC, Carazo JM, Scheres SH. An introduction to maximum-likelihood methods in cryo-EM. *Methods Enzymol.* 2010;482:263-94. doi: 10.1016/S0076-6879(10)82011-7. PMID: 20888965; PMCID: PMC3051286.

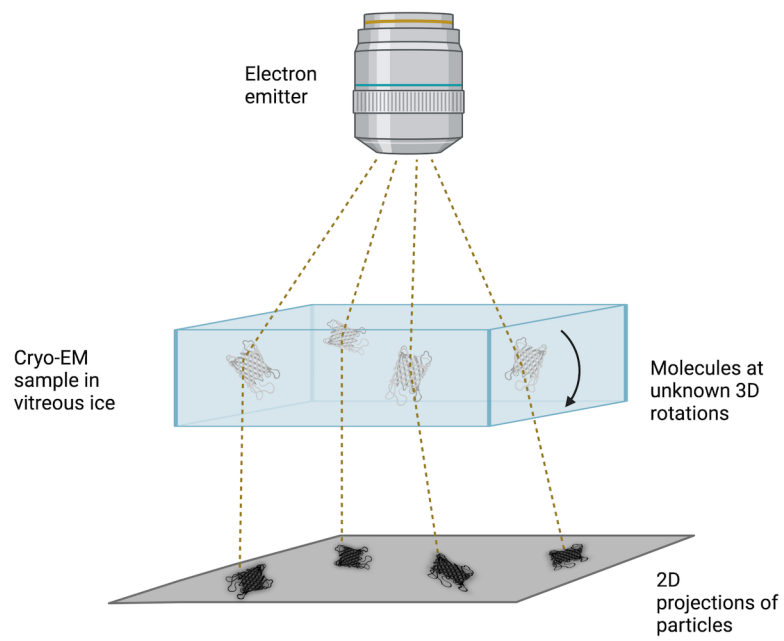


Fig. 1. Schematic of cryo-EM imaging process. Created with BioRender.com.

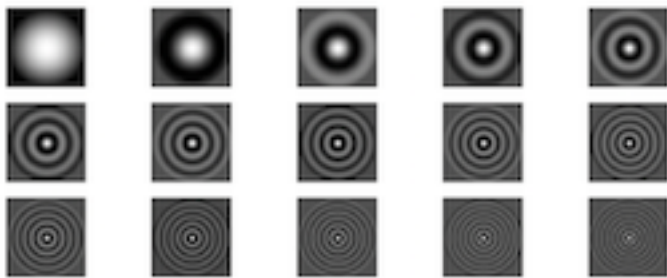


Fig. 2. The first 15 of the 608 Fourier-Bessel functions used to represent a 32x32 image.

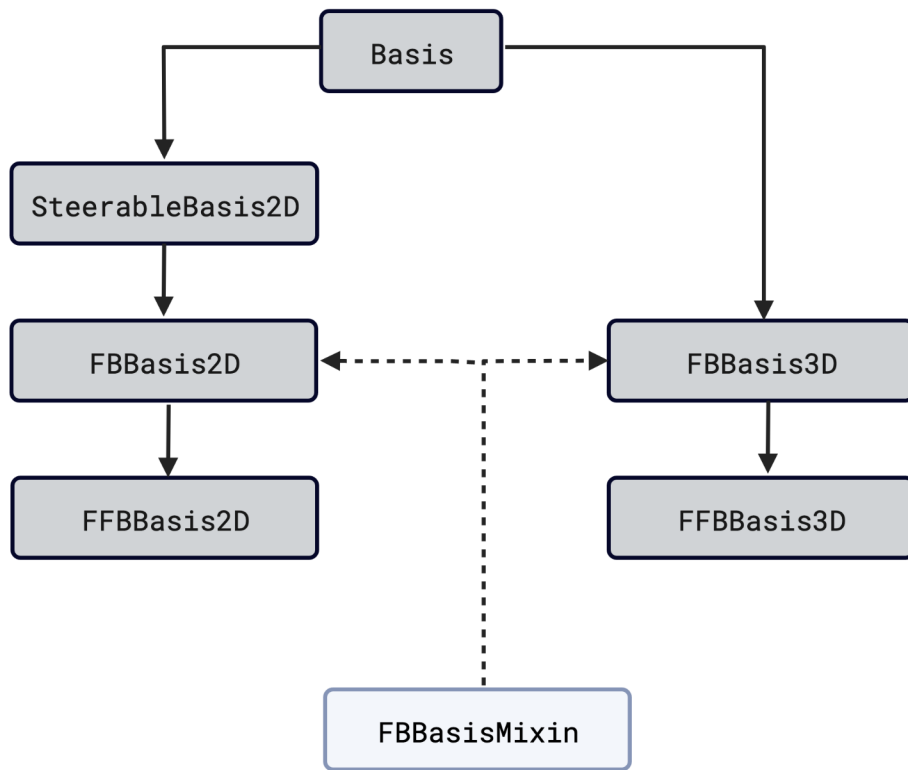


Fig. 3. Object-oriented structure of ASPIRE Fourier-Bessel Bases. Created with BioRender.com.

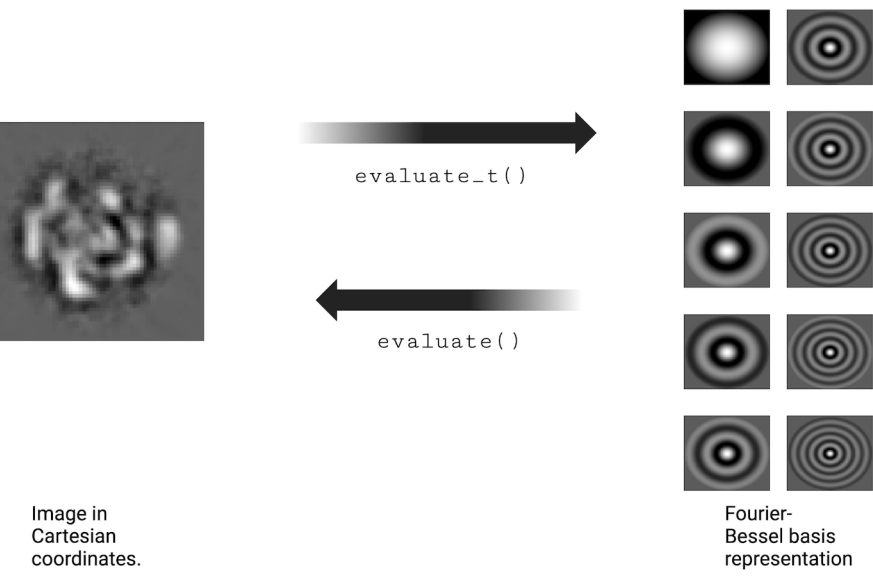


Fig. 4. Schematic of ASPIRE's basis methods. Created with BioRender.com.