Representing Steerable Bases for cryo-EM in ASPIRE

Christopher Langfield*, **Josh Carmichael***, **Garrett Wright***, **Joakim Andén**†,
Amit Singer*

* Program in Applied and Computational Mathematics, Princeton University
† Department of Mathematics, KTH Royal Institute of Technology

Abstract

An introduction to the mathematical problem of cryo-electron microscopy (cryo-EM) is given, along with an overview of ASPIRE, an open-source Python package for processing cryo-EM image data. ASPIRE uses unique Fourier basis representations for images of cryo-EM particles. The challenge of representing these mathematical structures within the ASPIRE codebase was addressed by building an extensible class hierarchy using mixins.

1. 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. The sample is then placed under a 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) [2].

In this poster we give a brief introduction to ASPIRE [3], a Python package for processing cryo-EM image data, and show an example of how software design principles can be used to efficiently represent mathematical structures used in the algorithms.

2. 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 a molecule given n projection images obtained using the method outlined above. Mathematically, these images are tomographic projections of the target function, and must be combined and oriented to reveal the 3D molecular structure. Several popular programs implement SPR pipelines, using a variety of methods, and active research is still ongoing in the field [4].
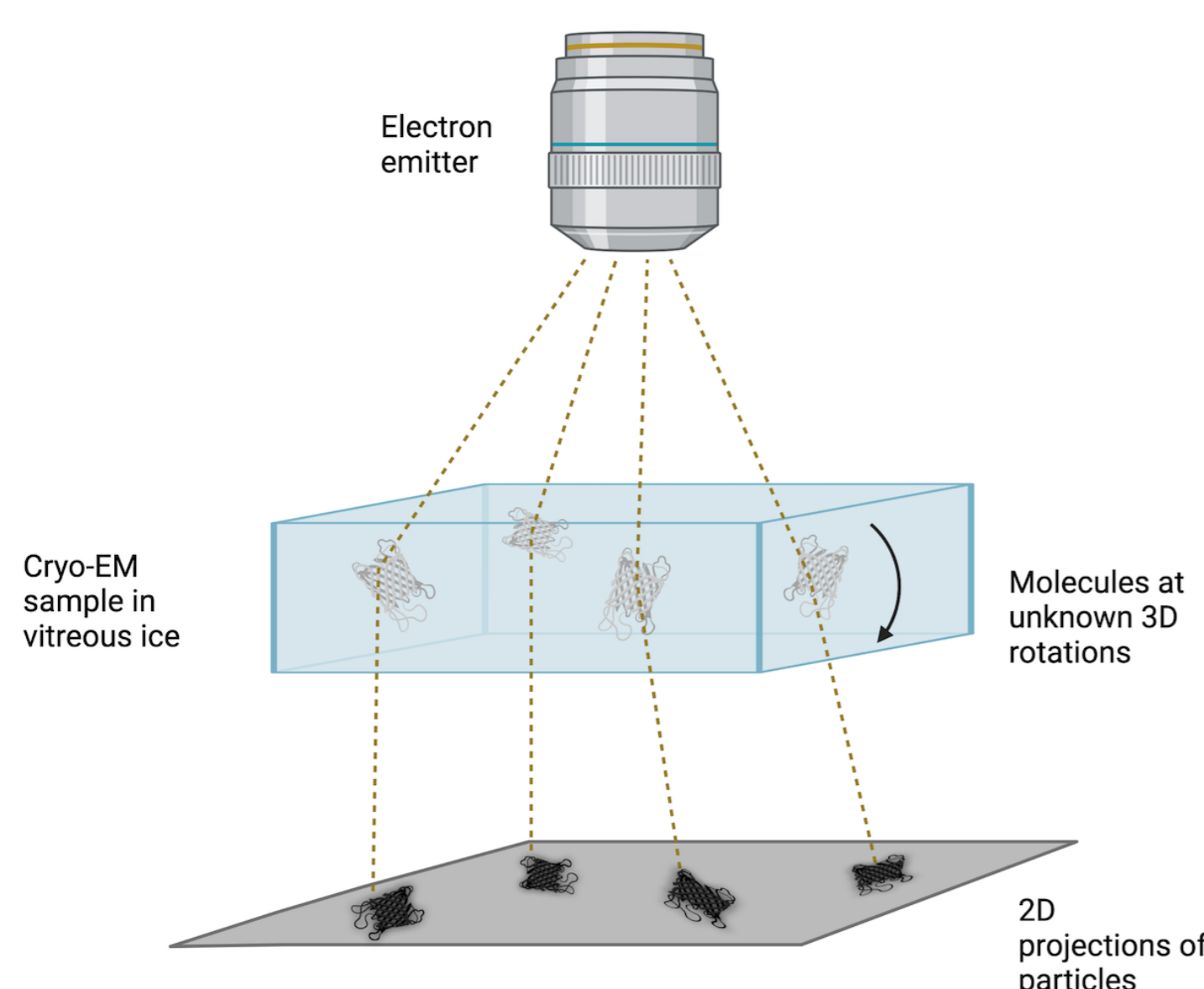

Figure 1: *Image formation model in cryo-EM. Created with BioRender.com.*

3. ASPIRE

ASPIRE (**A**lgorithms for **S**ingle **P**article **Re**construction) [3] is a Python package, installable from the Python Package Index (PyPI), which aims to implement an end-to-end SPR pipeline.[1] It uses novel methods for various steps including particle picking [5], CTF correction [6], particle orientation estimation [7], and image classification [8]. Many common cryo-EM programs, including the widely-used RELION [9], work from an initial guess of the three-dimensional structure of a molecule, and then use maximum likelihood methods to refine this guess into a detailed result [10]. 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.

[1] Source code available at github.com/ComputationalCryoEM/ASPIRE-Python.

4. Efficient Representations of Projection Images in Custom Bases

ASPIRE represents images using "steerable" bases [11] [12]. Images are represented as a list of coefficients associated with a subset of radial and angular functions that span a subspace of bandlimited functions (Fig. 3). 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.
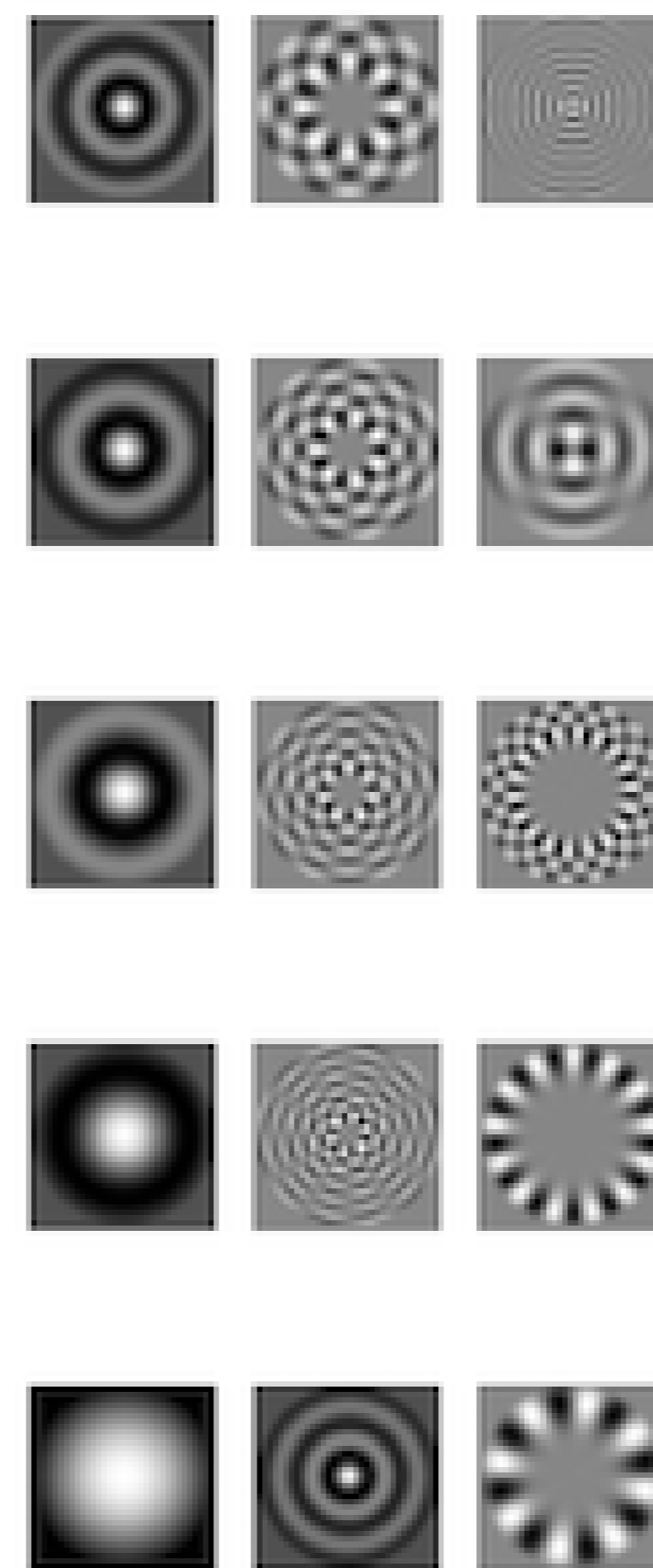

Figure 3: *Some of the 608 Fourier-Bessel functions used to represent a 32x32 image.*

The Fourier-Bessel functions are a set of polar basis functions defined:

$$\psi^{k,q}(\xi,\theta) = \begin{cases} N_{k,q} J_k(R_{k,q}\xi)e^{ik\theta} & \xi \leq 1 \\ 0 & \xi > 1 \end{cases} \quad (1)$$

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. (ξ, θ) is a polar representation of the spatial frequency, with ξ being the radial frequency and θ the angular part. Although this is theoretically an infinite series, an image can be efficiently represented with a truncated Fourier-Bessel sequence using bounds derived in [11] and [12].

The coefficients of the basis representation for a given image $I(x, y)$ are found by numerically integrating

$$a_{k,q} = \int_0^{2\pi} \int_0^1 \hat{I}(\xi,\theta) \overline{\psi^{k,q}(\xi,\theta)} \xi d\xi d\theta \quad (2)$$

The "steerable" property of this basis means that the coefficients $a'_{k,q}$ of the same image rotated by an angle α transform via scalar multiplication (also known as phase modulation):

$$a'_{k,q} = a_{k,q}e^{-ik\alpha} \quad (3)$$

5. Object-oriented Representations of Steerable Bases

Several bases are used at different points in the SPR pipeline: `FBBasis2D`, `FBBasis3D`, and their GPU-accelerated fast analogs `FFBBasis2D`, `FFBBasis3D`. Each basis is represented as an object in an inheritance structure descending from base class `Basis` (Fig. 4).
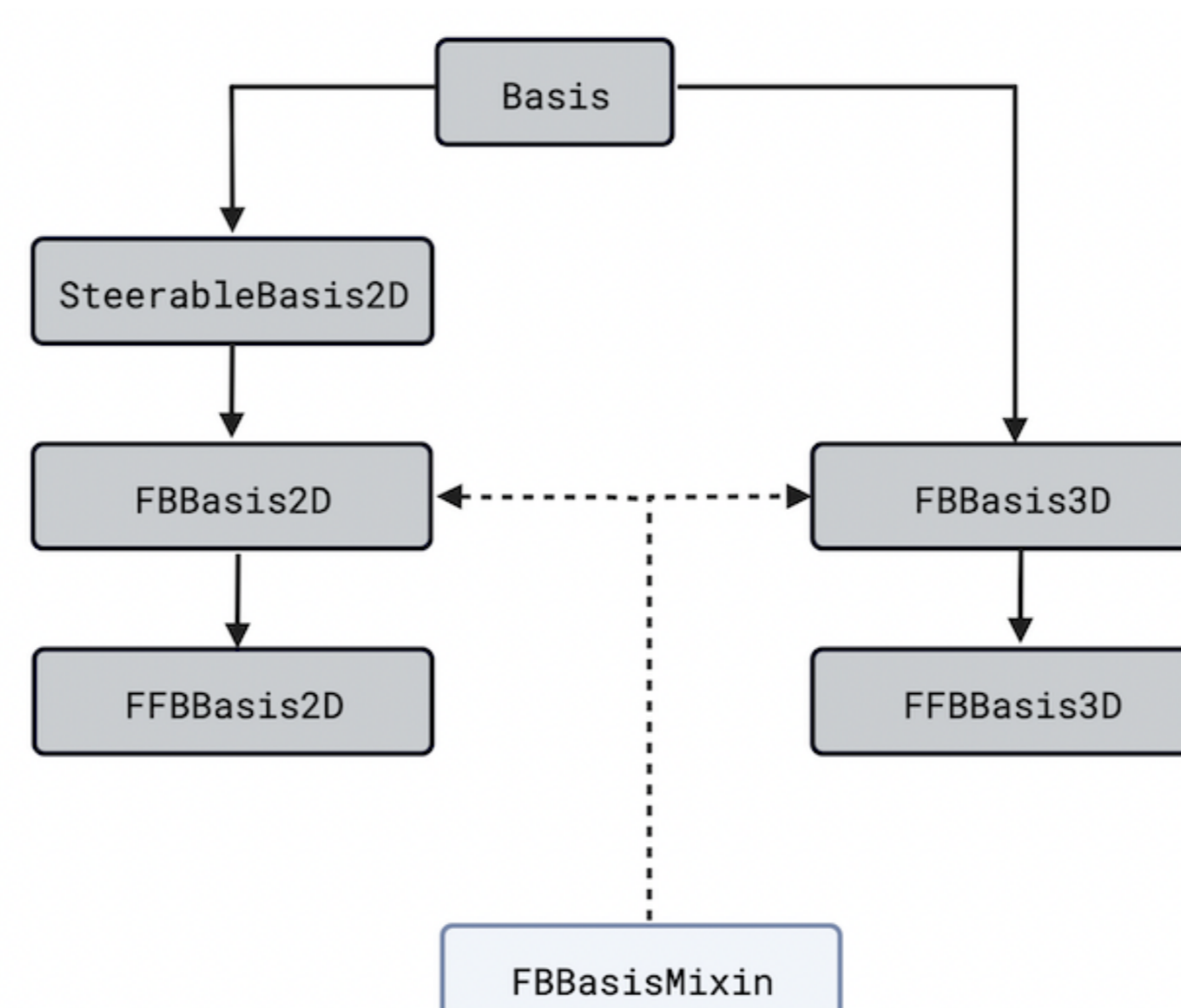

Figure 4: *Object-oriented structure of ASPIRE Fourier-Bessel Bases. Created with BioRender.com.*

Several considerations apply to the design of this inheritance structure. While the two-dimensional Fourier-Bessel and Fast Fourier-Bessel bases implement the "steerable" property mentioned above, their three-dimensional analogs do not. 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.

6. Discussion

Due to working with several similar mathematical structures whose overlapping properties do not fit into a linear structure, ASPIRE's basis libraries originally had a large amount of repeated code and brittle structure. Adding a new type of basis or functionality could be difficult as code would have to be copied or classes rearranged. The typical software design solution is to inherit vertically as much as possible, pulling common methods up into abstract parent classes. However, since the "steerable" and Fourier-Bessel properties do not always overlap, we solved the problem by using a lateral class hierarchy, i.e. a mixin.

References

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

[2] D. Bhella. Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility. *Biophys Rev*, 11:515–519, 2019.

[3] 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. ComputationalCryoEM/ASPIRE-Python v0.9.2. *Zenodo*, 2022. https://doi.org/10.5281/zenodo.6607659.

[4] A. Singer and F. J. Sigworth. Computational methods for single-particle cryo-EM. *Annual Review of Biomedical Data Science*, 3:163–190, 2020.

[5] A. Heimowitz, J. Andén, and A. Singer. APPLE picker: Automatic particle picking, a low-effort cryo-EM framework. *Journal of Structural Biology*, 204(2):215–227, 2018.

[6] A. Heimowitz, J. Andén, and A. Singer. Bias and variance reduction and denoising for CTF estimation. *Ultramicroscopy*, 212, 2020.

[7] Y. Shkolnisky and A. Singer. Viewing direction estimation in cryo-EM using synchronization. *SIAM Journal on Imaging Sciences*, 5(3):1088–1110, 2012.

[8] Z. Zhao and A. Singer. Rotational invariant image representation for viewing direction classification in cryo-EM. *Journal of Structural Biology*, 186(1):153–66, 2014.

[9] S.H. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*, 180(3):519–30, 2012.

[10] F. J. Sigworth, P. C. Doerschuk, J. M. Carazo, and Scheres SH. An introduction to maximum-likelihood methods in cryo-EM. *Methods Enzymol*, 482:263–94, 2010.

[11] Z. Zhao and A. Singer. Fourier-Bessel rotational invariant eigenimages. *The Journal of the Optical Society of America A*, 30(5):871–877, 2013.

[12] Z. Zhao, Y. Shkolnisky, and A. Singer. Fast steerable principal component analysis. *IEEE Transactions on Computational Imaging*, 2(1):1–12, 2016.