PIMS workshop: Mathematical and Computational Challenges in CryoEM

May 5 (UBC Earth Sciences Building 4133)

9:00-9:05: Introductory remarks
9:05-9:35: Roy Lederman (Yale): Continuous heterogeneity in cryo-EM
9:35-9:55: Shayan Shekarforoush (University of Toronto): Continuous Conformational 3D Reconstruction with Invertible ResNet
9:55-10:15: Bogdan Toader (Yale): Ab initio reconstruction of CryoEM volumes using Markov Chain Monte-Carlo sampling

10:15-10:30: Coffee break

10:30-11:00: Joe Kileel (UT Austin): Analyzing molecular conformation spaces by manifold learning with respect to non-Euclidean norms
11:00-11:20: Aryan T. Riahi (UBC): Optimal transport based algorithm for fast 3D alignment of EM maps
11:20-11:40: Axel Lévy (Stanford): CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images
11:40-12:00: Calvin Yip (UBC): Challenges in structural investigations of flexible proteins and assemblies by single-particle EM

12:00-1:45 Lunch break

1:45:2:30: Sriram Subramaniam (UBC): Cryo-EM of dynamic molecular assemblies
2:30-2:50: Geoffrey Woollard (UBC): Cryo-EM Simulation
2:50-3:20: Sonya Hanson (Flatiron Institute): Exploring the quaternion adjugate approach to the pose estimation problem

3:20-3:30: Coffee Break

3:30-4:30: Panel (Sriram Srubamaniam, Frédéric Poitevin (SLAC National Lab), Marcus Brubaker (University of Toronto), Sonya Hanson)
4:30-5:30: Open discussion in PIMS lounge

May 6 (UBC LSK 121)

9:00-9:30: Frédéric Poitevin (SLAC National Lab): Introduction to compSPI: building a community for reliable and efficient method development in cryoEM

9:30-10:00: Nina Miolane (UC Santa Barbara): Github and coding best practices

10:00-4:00: Coding

Remark: Coffee and snacks will be available during breaks and 15 min before the morning session

Abstracts

Roy Lederman (Yale): Continuous heterogeneity in cryo-EM

One of the great promises of cryo-EM is to explore multiple conformations of biomolecules: while other methods for structure determination, such as x-ray crystallography and NMR, measure ensembles of molecules, cryo-electron microscopes produce images of individual particles; therefore, cryo-EM could potentially be used to study mixtures of conformations of molecules. We will discuss a range of recent methods for analyzing the geometry of molecular conformations using cryo-EM data, and point to some of the mathematical and practical challenges ahead.

Shayan Shekarforoush (University of Toronto): Continuous Conformational 3D Reconstruction with Invertible ResNet

In cryo-EM reconstruction problem, we are given a large set of extremely noisy images each with each image corresponding to the projection of a randomly oriented structure of a macromolecular complex such as proteins. In general, proteins have intrinsic flexibilities, allowing them to perform structural deformations during sample preparation. Traditional homogeneous reconstruction methods does not take into account such possible conformational variations. We approach this by factorizing reconstruction into learning a canonical structure, shared during optimization, and a deformation vector field modeling local displacements of density. Motivated by prior knowledge about feasible deformations, an Invertible Residual Network, conditioned on image latent variables, generates new locations of density after deformation. With Intrinsic invertibility of this architecture, we avoid learning physically unrealistic deformations of structure, including collapsing of densities. Moreover, we show, both theoretically and empirically, that the learned vector field is locally smooth. Finally, we demonstrate the performance of the models on synthetic datasets.

Bogdan Toader (Yale): Ab initio reconstruction of CryoEM volumes using Markov Chain Monte-Carlo sampling

Most refinement algorithms for single particle reconstruction in Cryo-EM require a low resolution initial model and an approximation of the poses in each particle image as a good initialization. Therefore, having a high quality ab initio model is of crucial importance. Often, this is computed using class averages, obtained by splitting the particle images into different classes corresponding to different orientations and averaging each cluster, a process which leads to denoised particle images, but also to loss of information. In this proof-of-concept work, we show how Markov Chain Monte-Carlo (MCMC) sampling methods can be used to obtain an initial model directly from the noisy particle images, bypassing the need for class averaging. In this talk, we describe useful sampling

strategies of the volume and the poses, highlighting the benefits of this method and show results on both simulated and experimental data.

Joe Kileel (UT Austin): Analyzing molecular conformation spaces by manifold learning with respect to non-Euclidean norms

I will discuss a mathematical framework that has arisen from the need to analyze molecular conformation spaces in cryo-EM. Specifically, I'll talk about manifold learning when affinities are measured by non-Euclidean norms. I'll determine the limiting differential operator, and contrast it with the usual Laplace-Beltrami operator. Synthetic experiments on the ATP synthase conformation space show that it can be advantageous to use transport-based distances instead of Euclidean ones. Joint with Amit Moscovich, Nathan Zelesko and Amit Singer.

Aryan T. Riahi (UBC): AlignOT: Optimal transport based algorithm for Cryo-EM density map alignment

The alignment of voxelized 3D electron density maps, obtained by Cryo-EM, is a core algorithmic subroutine in studying multiple conformations of a biomolecule. This step remains costly and challenging, with standard alignment tools getting potentially stuck in local minima. We propose a new procedure, called AlignOT, which uses computational optimal transport (OT) to align Cryo-EM maps in 3D space. In this talk, we describe a method that embeds a fast estimation of OT plan matrices within a stochastic gradient descent algorithm, and searches for a rigid body alignment of the maps that minimizes their Wasserstein distance. We quantify the precision and accuracy of the alignment and show that AlignOT can outperform the standard local alignment methods with an increased range of rotation angles leading to proper alignment.

Axel Lévy (Stanford): CryoAI: Amortized Inference of Poses for Ab Initio Reconstruction of 3D Molecular Volumes from Real Cryo-EM Images

Cryo-electron microscopy (cryo-EM) has become a tool of fundamental importance in structural biology, helping us understand the basic building blocks of life. The algorithmic challenge of cryo-EM is to jointly estimate the unknown 3D poses and the 3D electron scattering potential of a biomolecule from millions of extremely noisy 2D images. Existing reconstruction algorithms, however, cannot easily keep pace with the rapidly growing size of cryo-EM datasets due to their high computational and memory cost. We introduce cryoAI, an ab initio reconstruction algorithm for homogeneous conformations that uses direct gradient-based optimization of particle poses and the electron scattering potential from single-particle cryo-EM data. CryoAI combines a learned encoder

that predicts the poses of each particle image with a physics-based decoder to aggregate each particle image into an implicit representation of the scattering potential volume. This volume is stored in the Fourier domain for computational efficiency and leverages a modern coordinate network architecture for memory efficiency. Combined with a symmetrized loss function, this framework achieves results of a quality on par with state-of-the-art cryo-EM solvers for both simulated and experimental data, one order of magnitude faster for large datasets and with significantly lower memory requirements than existing methods.

Calvin Yip (UBC): Challenges in structural investigations of flexible proteins and assemblies by single-particle EM

Although advances in microscopy hardware and image processing algorithms have enabled cryo-EM to become a mainstream structural determination method, many technical challenges still exist in studying conformationally flexible and less "globular" proteins and protein assemblies. In this presentation, I will illustrate, from a biochemist perspective, impediments from specimen preparation, data acquisition, particle picking, to 3D reconstruction, using examples from our laboratory.

Sriram Subramaniam (UBC): Cryo-EM of dynamic molecular assemblies

Recent breakthroughs in the field of cryo-electron microcopy (cryo-EM) provide new prospects for determination of the structures of a variety of medically important macromolecular assemblies. The prospect that the determination of protein structures at atomic resolution will no longer be limited by size, or by the need for crystallization represents a significant and exciting horizon in structural biology. In my talk, I will discuss selected applications of cryo-EM methods for studying a range of dynamic protein assemblies including metabolic enzymes, nucleic acid-protein complexes and the SARS-COV-2 spike protein, with emphasis on the potential use of cryo-EM for the development of improved therapeutic agents.

Geoffrey Woollard (UBC): Cryo-EM simulators

The forward model of image formation in single particle cryo-EM is unique among imaging applications in two senses: the physics of the specimen, and the physics of the imaging process. While voxelized representations of both the specimen and the imaging process have received much attention from a digital signal processing perspective, atom-based coordinate representations are more familiar to (bio)chemists, and are a data format closer to how structural biologists communicate and share information among different imaging modalities. Here I explore an image formation model in real space that perturbs, rotates,

and projects atom centers. The latent space includes conformational heterogeneity, 3D pose, and CTF defocus with no linear interpolation in voxelized Fourier space, or Fourier slicing. I frame the stochastic measurement process as a probabilistic program, where each latent is sampled from a prior distribution. Such structured probabilistic approaches hold promise to be generalized to incorporate prior knowledge of microscope and detector physics of advanced models of image formation, and be suitable for inference. I will also highlight the other simulators in compSPI/simSPI repository, and advocate for design features such as reproducibility, modularity, and interpretability.

Sonya Hanson (Flatiron Institute): Exploring the quaternion adjugate approach to the pose estimation problem

Quaternions are important for a wide variety of rotation-related problems in scientific imaging. We study the nontrivial geometry of the relationship between quaternions and rotation matrices by exploiting a powerful construct of linear algebra optimization algorithms known as the adjugate matrix. We argue that quaternions parameterized by their corresponding rotation matrices cannot be expressed, for example, in machine learning tasks, as single-valued functions: the quaternion solution must instead be treated as a manifold, with different algebraic solutions for each of several single-valued sectors represented by the adjugate matrix. We present novel constructions exploiting the quaternion adjugate variables to reformulate classic pose estimation applications: point-cloud matching and point-cloud-to-projection matching, and also extend the usefulness of this formalism to machine learning approaches to these same problems.