Uniqueness and atomic modelling in cryo-electron microscopy

Joakim Andén, Pär Kurlberg, Ozan Öktem

Cryo-electron microscopy (cryo-EM) is a method for imaging biological macro-molecules (like proteins or viruses) in their natural aqueous environment. A solution containing a large number of isolated particles (copies of the macro-molecule that is to be investigated) is flash-frozen and imaged in a transmission electron microscope (TEM). This yields a 2D image that contains projection images of the particles, each representing the macro-molecule in an unknown 3D orientation. The inverse problem is to computationally recover a 3D model and a motion model of the macro-molecule from these projections.

Many of the advances in cryo-EM that were made during the last three decades were acknowledged with the 2017 Nobel Prize in Chemistry. However, despite many advances, there are still many open questions and this project targets two of these: formal uniqueness of the inverse problem and direct reconstruction of the atomic model.

Research in the above topics touch upon a wide range of mathematical fields, such as integral geometry, mathematical analysis, harmonic analysis, geometry, optimisation, representation theory, and statistics. The project will therefore be jointly led by Joakim Andén, Pär Kurlberg, and Ozan Öktem who all bring complementary expertise to the task. In addition, the project includes collaborations with a doctoral student in Ozan Öktem's group as well as researchers in the groups of Carola-Bibiane Schönlieb (Cambridge), Amit Singer (Princeton), and Roy Lederman (Yale).

Formal uniqueness for single-particle cryo-EM. Formal uniqueness is one of the first questions one seeks to resolve in a mathematical analysis of an inverse problem. Single-particle cryo-EM is an example of a tomographic inverse scattering problem with the (scalar) Schrödinger equation. The standard approach is to consider the eikonal approximation, disregard the TEM optics, and handle phase retrieval by linearising the intensity. This results in a highly simplified model for the TEM image formation where a 2D particle image is modelled as the ray transform (along the electron beam) of the 3D volumetric map. Uniqueness results for single-particle cryo-EM has been proved in this setting [1].

It is however clear that reconstruction methods based on the projection model have reached a "resolution barrier" and going beyond that requires more accurate models. One approach is to model the scattered electron wave by the Born approximation, which amounts to considering the curvature of the "Ewald sphere" [4]. Remarkably, while such a model may be used as a forward operator in reconstruction, there is no rigorous proof of formal uniqueness in this setting. Some results in this direction were recently obtained by Kurlberg and Zickert, but these make several assumptions that can be relaxed. Our goal is to extend these results and establish a mathematical framework for Ewald sphere-corrected single-particle cryo-EM. This would involve extending uniqueness to more accurate models and for non-entire functions as well as in a finite-data setting. These methods will then be implemented algorithmically.

Reconstruction of atomic models. All current approaches for reconstruction in single particle cryo-EM first construct a 3D volumetric model of the molecule. This is followed by a step where one fits an atomic model to this 3D volume, which then allows biological interpretation of the model.

The inverse problem in single particle cryo-EM is severely ill-posed so any reconstruction method for the 3D volume needs to involve regularisation. Typically this is performed using smoothness priors on the 3D volume, i.e., no constraints are imposed on it other than regularity [3]. The images are also very noisy, so this lack of strong priors limits the effectiveness of regularisation. To construct a better prior, we can use the fact that the atomic configurations of biological macro-molecules (such as proteins) are quite restricted. We would then directly reconstruct the atomic models of the molecule from the data. Such models can then be deformed to fit projection images obtained from cryo-EM experiments, which is balanced against minimising the potential energy due to forces induced by the atomic configuration. One can then impose natural constraints on bond lengths and angles, for example, but also on larger-scale structures of amino acids in the case of proteins.

A key component of the project is also to develop domain-adapted deep neural networks that operate on atomic configurations of proteins. Explicit discrete and continuous models based on moduli spaces of flat G-connections on surfaces are suggested in [2]. The final step is to consider uniqueness for the single particle cryo-EM problem in the context of recovering the atomic model of the molecule.

[1] L. Lamberg. Unique recovery of unknown projection orientations in three-dimensional tomography. *Inverse Probl. Imaging*, 2(4): 47–575 (2008).
[2] R. C. Penner. Moduli spaces and macromolecules. *Bull. Am. Math. Soc.*, 53: 217–268 (2016).
[3] S. H. W. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.*, 180(3): 519–530 (2012).
[4] O. Öktem. Mathematics of electron tomography. O. Scherzer (ed.), *Handbook of Mathematical Methods in Imaging*: 937–1031 (2015).