EPFL

Thèse n° 7540

Reconstruction Methods for Cryo-Electron Microscopy: From Model-based to Data-driven

Présentée le 18 septembre 2020

à la Faculté des sciences et techniques de l'ingénieur
Laboratoire d'imagerie biomédicale
Programme doctoral en génie électrique

pour l'obtention du grade de Docteur ès Sciences

par

Laurène DONATI

Acceptée sur proposition du jury

Prof. J.-Ph. Thiran, président du jury
Prof. M. Unser, Dr D. Sage, directeurs de thèse
Prof. L. Blanc-Féraud, rapporteuse
Dr C. O. Sorzano, rapporteur
Prof. D. Van de Ville, rapporteur

École polytechnique fédérale de Lausanne

2020

A mes parents, Graziano et Isabelle,
pour leur amour inconditionnel.

Abstract

The topic of this thesis is the development of new reconstruction methods for cryo-electron microscopy (cryo-EM). Cryo-EM has revolutionized the field of structural biology over the last decade and now permits the regular discovery of biostructures. Yet, the technical challenges associated to cryo-EM are still numerous, and the measurements remain notoriously difficult to process. This calls for fast and robust algorithms that can reliably handle the challenging reconstruction task at hand.

In this thesis, we investigated two reconstruction paradigms: model-based and data-driven. *Model-based* methods formulate the reconstruction task as an inverse problem and rely on a faithful model of the acquisition physics. By contrast, the central philosophy of *data-driven* approaches is to let the reconstruction algorithm be guided by the measured data through some learning procedure. Both paradigms share a tight link in all our works: their reliance on a rigorous mathematical formulation of the cryo-EM imaging model.

The first cryo-EM method we considered is scanning transmission electron tomography (STET), a modality whose primary concern is to reduce the electron dosage required for accurate imaging. To handle this, we developed a tailored acquisition-reconstruction STET framework that relies on the principles of compressed sensing. This scheme permits high-quality reconstruction from a reduced number of measurements, hence greatly preserving the sample.

We then designed several reconstruction algorithms for single-particle analysis (SPA), a popular cryo-EM method that enables the determination of structures at near-atomic resolution. A key challenge for the deployment of robust, iterative reconstruction methods in SPA is that they usually come with a prohibitive computational cost if not carefully engineered. To circumvent this problem, we developed a regularized reconstruction scheme whose cost-dominant operation is recast as a discrete convolution, which makes the use of our robust scheme feasible in SPA. Building on this development, we devised a joint optimization framework that efficiently alternates between the reconstruction and the estimation of the unknown orientations.

We then explored a learning-based method to estimate the unknown orientations in SPA directly from the acquired dataset of projections. Capitalizing on our ability to model the cryo-EM procedure, we generated large synthetic SPA datasets to train a function—parametrized as a neural network—to predict the relative orientation between two projections based on their similarity. The framework relies on the postulate that it is possible to recover, from these estimated orientation distances, the orientations themselves through an appropriate

minimization scheme, as supported by preliminary tests.

Finally, we developed a completely new paradigm for SPA reconstruction that leverages the remarkable capability of deep neural networks to capture data distribution. The proposed algorithm uses a generative adversarial network to learn the 3D structure that has simulated projections that most closely match the real data in a distributional sense. By doing so, it can resolve a 3D structure in a single algorithmic run using only the dataset of projections and CTF estimations as inputs. Hence, it bypasses many processing steps that are necessary in the usual cryo-EM reconstruction pipeline, which opens new perspectives for reconstruction in SPA.

Keywords: cryo-electron microscopy, tomographic reconstruction, single-particle analysis, scanning transmission electron tomography, model-based, regularized inverse problems, compressed sensing, data-driven, neural networks.

Résumé

Cette thèse est consacrée au développement de nouvelles méthodes de reconstruction pour la cryo-microscopie électronique (cryo-EM en anglais). Cette technique d'imagerie a révolutionné la biologie structurelle ces dix dernières années et permet désormais la découverte régulière de nouvelles structures biologiques. Les difficultés liées à cette méthode restent toutefois nombreuses ; en particulier, le traitement algorithmique des images acquises représente toujours une tâche extrêmement complexe. Le développement d'algorithmes robustes et rapides permettant une reconstruction fiable des structures tridimensionnelles est donc d'une importance capitale.

Dans cette thèse, nous avons exploré deux paradigmes de reconstruction pour la cryo-EM : les méthodes dites *model-based* (basées sur un modèle) et les méthodes dites *data-driven* (guidées par les données). Les approches *model-based* formulent la reconstruction comme un problème inverse, s'appuyant pour ce faire sur une modélisation précise du processus d'acquisition. Par contraste, la philosophie des méthodes *data-driven* est de laisser les algorithmes de reconstruction exploiter les données acquises à travers un processus d'apprentissage. Dans nos travaux, les deux paradigmes sont étroitement liés en ce qu'ils s'appuient tout deux sur une formalisation mathématique rigoureuse du modèle d'acquisition d'image propre à la cryo-EM.

La première modalité cryo-EM sur laquelle nous avons travaillé est la microscopie électronique en transmission à balayage (STET en anglais). Le défi principal en STET est de minimiser la dose de radiation nécessaire à une acquisition de qualité suffisante. Pour résoudre ce problème, nous avons développé un protocole d'acquisition-reconstruction pour la STET qui exploite les principes de la théorie de l'« échantillonnage compressé ». Cette approche permet d'obtenir des reconstructions de haute qualité à partir d'un nombre restreint de mesures, ce qui garantit une meilleure préservation des échantillons biologiques.

Nous avons également développé une série d'algorithmes de reconstruction pour la modalité dite d'« analyse des particules isolées » (SPA en anglais). La SPA est une technique d'imagerie cryo-EM très prisée qui permet la caractérisation tridimensionnelle des structures biologiques à une résolution quasi atomique. En SPA, un frein important au déploiement de méthodes de reconstruction itératives — pourtant d'une très grande robustesse — est qu'elles s'accompagnent généralement d'un coût computationnel particulièrement élevé. Pour contourner ce problème, nous avons développé un algorithme de reconstruction régularisée dont l'opération la plus coûteuse est reformulée comme une convolution rapide, ce qui rend l'utilisation d'une telle méthode concevable en SPA. En s'appuyant sur cet algorithme,

nous avons ensuite conçu une stratégie d'optimisation jointe qui alterne de manière efficace entre la reconstruction de la structure et l'estimation des orientations des particules, qui sont initialement inconnues.

Nous avons ensuite exploré une méthode *data-driven* dont le but est d'estimer les orientations directement à partir des mesures acquises par le microscope. Capitalisant sur notre capacité à modéliser précisément le processus d'acquisition d'image en SPA, nous avons généré de nombreuses données synthétiques pour apprendre une fonction — paramétrée par un réseau de neurones — qui prédise l'orientation relative entre deux images de mesure en se basant uniquement sur leur similarité. Ce protocole s'appuie sur le postulat qu'il est possible de retrouver, à partir des orientations relatives, l'ensemble des orientations absolues en utilisant un algorithme d'optimisation approprié, comme le démontrent une série de tests préliminaires.

Pour finir, nous avons proposé un paradigme de reconstruction complètement novateur pour la SPA qui repose sur la capacité remarquable des réseaux de neurones à « capturer » les distributions sous-jacentes aux données. Notre approche utilise un réseau antagoniste génératif pour apprendre progressivement la structure tridimensionnelle dont la distribution des projections simulées s'approche le plus de celle des vraies données. Cela permet la détermination de structures en une seule exécution algorithmique, et cela en se basant uniquement sur les mesures acquises et sur une estimation des paramètres optiques du microscope. Notre approche permet d'éviter nombre d'étapes coûteuses inhérentes aux algorithmes standards, ce qui ouvre de nouvelles perspectives dans le domaine de la reconstruction en SPA.

Mots clefs : cryo-microscopie électronique, reconstruction tomographique, analyse de particules isolées, microscopie électronique en transmission à balayage, problèmes inverses régularisés, échantillonnage compressé, réseaux de neurones.

Contents

Notations

Sets

$\mathbb{Z}$	Integers	
$\mathbb{R}$	Real numbers	
$\mathbb{C}$	Complex numbers	
$\mathbb{H}$	Quaternions	
$\mathbb{U}$	Unit quaternions	$\mathbb{U} = \{q \in \mathbb{H} \mid \lvert q\rvert = 1\}$
$\mathbb{Z}^d$	Multi-integers of dimension d	
$\mathbb{R}^d$	Vectors of real numbers of dimension d	
$\mathbb{S}^d$	Unit sphere in $\mathbb{R}^d$	$\mathbb{S}^d = \{\mathbf{x} \in \mathbb{R}^d \mid \lVert\mathbf{x}\rVert_2 = 1\}$

Sequence and Function Spaces

$\ell_1(\mathbb{Z}^d)$	Absolutely-summable sequences
$\ell_2(\mathbb{Z}^d)$	Finite-energy sequences
$L_1(\mathbb{R}^d)$	Absolutely-integrable measurable functions
$L_2(\mathbb{R}^d)$	Finite-energy functions

Cryo-EM Forward Model

f, $f(\boldsymbol{x})$	3D object	$\mathbb{R}^3 \to \mathbb{R}$
c, $c[\boldsymbol{k}]$	Coefficient sequence	$\mathbb{Z}^3 \to \mathbb{R}$
φ	Basis function	$\mathbb{R}^3 \to \mathbb{R}$
$V(\varphi)$	Reconstruction space generated by φ	
s	Scaling factor	$\mathbb{R}_+$
$\mathbf{SO}(3)$	Rotation group in $\mathbb{R}^3$	
$\boldsymbol{\theta}$	Euler angles with $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$	$\Omega_{\boldsymbol{\theta}} = [0; 2\pi) \times [0; \pi] \times [0; 2\pi)$
$\mathscr{P}_{\boldsymbol{\theta}}$	X-ray transform for a 3D geometry given by $\boldsymbol{\theta}$	$L_2(\mathbb{R}^3) \to L_2(\mathbb{R}^2)$
$\vartheta_{\boldsymbol{\theta}}$	Projection direction	$\mathbb{S}^2$
$\mathbf{M}_{\boldsymbol{\theta}^\perp}$	Hyperplane projection matrix	$\mathbb{R}^{2\times 3}$
h	Point-spread function (PSF)	$\mathbb{R}^2 \to \mathbb{R}$
N	Number of voxels of the 3D object	$\mathbb{N}^*$
M	Number of pixels of a 2D projection	$\mathbb{N}^*$
P	Number of acquired 2D projections	$\mathbb{N}^*$
$\mathbf{c}$	Discrete object coefficients	$\mathbb{R}^N$

$\mathbf{b}^{\boldsymbol{\theta}_p}, \mathbf{b}^p$	2D measurement vector for an orientation $\boldsymbol{\theta}_p$	$\mathbb{R}^M$
$\mathbf{b}$	Measurement dataset such that $\mathbf{b} = \left\{\mathbf{b}^p\right\}_{p=1}^P$	$\left(\mathbb{R}^M\right)^P = \mathbb{R}^{MP}$
$\mathbf{H}$	System matrix	$\mathbb{R}^{MP \times N}$
$\mathbf{H}^T$	Transpose of $\mathbf{H}$	$\mathbb{R}^{N \times MP}$
$\mathbf{n}$	Additive noise vector on $\mathbf{b}$	$\mathbb{R}^{MP}$
$\Omega_{\boldsymbol{\theta}}$	Domain of the Euler angles $\boldsymbol{\theta}$	$[0;2\pi) \times [0;\pi] \times [0;2\pi)$
$\Omega_{2\mathrm{D}}$	Support of the 2D vector $\mathbf{b}^p$	$\Omega_{2\mathrm{D}} \subset \mathbb{Z}^2$ and $\sharp\Omega_{2\mathrm{D}} = M$
$\Omega_{3\mathrm{D}}$	Support of the 3D vector $\mathbf{c}$	$\Omega_{3\mathrm{D}} \subset \mathbb{Z}^3$ and $\sharp\Omega_{3\mathrm{D}} = N$
$\boldsymbol{j}$	Indices of 3D vectors	$\Omega_{2\mathrm{D}}$
$\boldsymbol{k}$	Indices of 2D vectors	$\Omega_{3\mathrm{D}}$

Model-Based Algorithms

$\tilde{f}$	Approximation of f in $V(\varphi)$	$\mathbb{R}^3 \to \mathbb{R}$
$\widehat{\mathbf{c}}$	Estimate of $\mathbf{c}$	$\mathbb{R}^N$
$\|\cdot\|_p$	p-norm of a vector with $p = 1, 2$	$\mathbb{R}^d \to \mathbb{R}$
λ	Regularization parameter	$\mathbb{R}_+$
$\mathscr{R}$	Convex regularization functional	$\mathbb{R}^Q \to \mathbb{R}$
$\mathbf{L}$	Regularization operator	$\mathbb{R}^{Q \times N}$
∇	Discrete gradient operator	$\mathbb{R}^{3N \times N}$
$i_{\geq 0}$	Nonnegativity constraint	$\mathbb{R}^N \to \mathbb{R}$
$\mathbf{u}, \mathbf{v}$	Auxiliary variables	$\mathbb{R}^Q$
$\mathscr{L}$	Augmented Lagrangian functional	$\mathbb{R}^N \times \mathbb{R}^Q \times \mathbb{R}^Q \to \mathbb{R}$
$\boldsymbol{\alpha}$	Lagrangian multipliers	$\mathbb{R}^Q$
ρ	Penalty parameter	$\mathbb{R}_+$
$\mathrm{prox}_{\mathscr{R}}$	Proximal operator of $\mathscr{R}$	$\mathbb{R}^Q \to \mathbb{R}^Q$

Data-Driven Algorithms

q	Unit quaternion	$\mathbb{U}$
d_b	Distance between two projections	$\mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}_+$
d_q	Geodesic distance between unit quaternions	$\mathbb{U} \times \mathbb{U} \to [0, \pi]$
$\mathscr{G}_\Phi$	Generator network with weights Φ	
$\mathscr{D}_\Phi$	Discriminator network with weights Φ	

Miscellaneous

δ	Dirac delta function	$S'(\mathbb{R}^d)$
$\mathscr{F}$	Fourier transform operator	$L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$
$\widehat{f}$	Fourier transform of f	$\widehat{f} = \mathscr{F}\{f\}$
$f * g$	Continuous convolution	$(f * g)(\boldsymbol{x}) = \int_{\mathbb{R}^d} f(\boldsymbol{\tau}) g(\boldsymbol{x} - \boldsymbol{\tau}) \mathrm{d}\boldsymbol{\tau}$
$c \star d$	Discrete convolution	$(c \star d)[\boldsymbol{k}] = \sum_{\boldsymbol{l} \in \mathbb{Z}^d} c[\boldsymbol{l}] d[\boldsymbol{k} - \boldsymbol{l}]$
$f^\vee$	Reflection of a function	$f^\vee(\boldsymbol{x}) = f(-\boldsymbol{x})$
$(\cdot)$	Spaceholder to define mappings	*e.g.*, $f(\cdot/s) := \left(\boldsymbol{x} \mapsto f(\boldsymbol{x}/s)\right)$

Abbreviations

ADMM	alternating-direction method of multipliers
CG	conjugate gradient
CS	compressed sensing
Cryo-EM	cryogenic electron microscopy
CT	computed tomography
CTF	contrast transfer function
CNN	convolutional neural network
DFR	direct Fourier reconstruction
DNN	deep neural networks
EM	electron microscopy
ET	electron tomography
FBP	filtered back projection
FFT	fast Fourier transform
FSC	Fourier shell correlation
GAN	generative adversarial network
KBWF	Kaiser-Bessel window function
MAP	maximum a posteriori
PDF	probability density function
PSF	point spread function
RB-STET	random-beam scanning transmission electron tomography
TEM	transmission electron microscope
TET	transmission electron tomography
TV	total variation
SGD	stochastic gradient descent
SNR	signal-to-noise ratio
STET	scanning transmission electron tomography
SPA	single-particle analysis (also stands for single-particle cryo-EM)

Introduction

The topic of this thesis is the development of new reconstruction methods for cryo-electron microscopy (cryo-EM). We investigate two reconstruction paradigms: model-based and data-driven. *Model-based* methods formulate the reconstruction task as an inverse problem and rely on a faithful model of the acquisition physics. By contrast, the central philosophy of *data-driven* approaches is to let the reconstruction algorithm be guided by the measured data through some learning procedure. In this introduction, we give an overview of the scientific background in which our work takes place. We then summarize its main contributions in that context. Relevant chapters are mentioned throughout this introduction. A roadmap of the thesis is given in Figure 1.

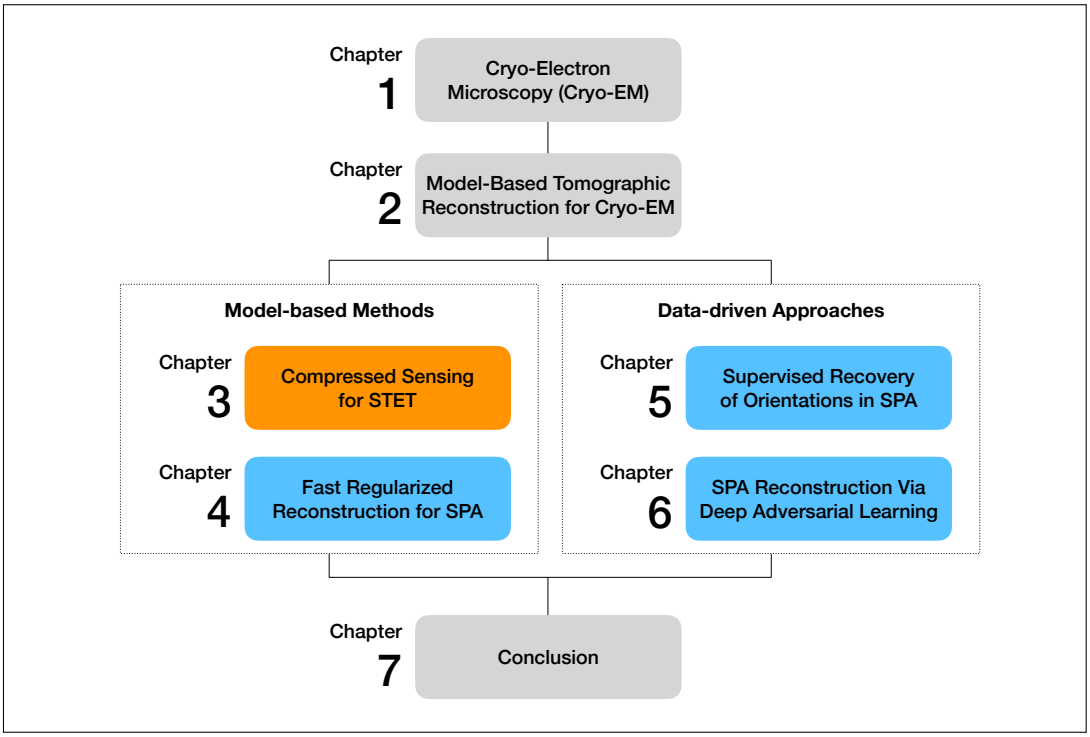


Figure 1 – Roadmap of the thesis. The work dedicated to electron tomography is highlighted in orange, and those dedicated to single-particle analysis in blue.

From Jacques Dubochet to Neural Networks

Two scientific milestones coincided with this PhD thesis and left their mark on it: 1) Jacques Dubochet received the 2017 Nobel Prize in Chemistry for his work on cryo-EM [1], putting Lausanne in the imaging spotlight, and 2) deep neural networks took over the world of signal processing, for the better [2] and the worse [3]. Hence, by both circumstances and conception, the subject of the thesis is very much rooted in the framework of *computational imaging*—the central philosophy being to exploit the tight link between physics and algorithms to produce imaging pipelines with enhanced capabilities.

It is hard to overstate how much the imaging method in question, cryo-EM, has revolutionized the field of structural biology [4]. The use of electron beams to image ice-embedded samples has permitted the recovery of 3D bio-structures at unprecedented resolution (Chapter 1). This revolution was made progressively possible by key scientific advances in sample preparation, detector hardware, and reconstruction algorithms [5]. The advent of atomic-resolution cryo-EM has had a tremendous impact in biomedical research, providing invaluable insights into the biological processes that underlie many current diseases. Cryo-EM has for example become vital in the quest toward prevention and treatment of age-related disorders such as Alzheimer's or Parkinson's diseases [6].

Yet, despite its countless successes, cryo-EM imaging remains notoriously riddled with technical challenges, for which scientists have spent the better part of the last 30 years looking for ingenuous solutions. For one, most samples of interest are highly radiation-sensitive. This results in extremely noisy measurements that are difficult to process. In addition, information on key imaging parameters is commonly lacking. The situation is even more complex for single-particle analysis, an EM modality that requires one to handle copious amounts of data, the unknown orientation of samples, and their possible conformational heterogeneity. These difficult imaging conditions impose strong constraints on the quality and robustness of the deployed image-processing methods.

Most cryo-EM variants for bioimaging are tomographic setups, which implies that the measurements are the 2D projections of a 3D object. Hence, computational frameworks are necessary to retrieve the desired structure from the acquired data (Chapter 2). The task of solving such inverse problems is not straightforward [7]. Those problems are generally *ill-posed*, meaning that one cannot rely on the measurements alone to robustly recover the 3D object. Indeed, even small perturbations on the measurements can massively impact the recovered solution—a major difficulty in cryo-EM.

In practice, many cryo-EM software packages still rely on direct inversion algorithms [8, 9, 10]. Those methods have the crucial advantage of being fast, so that the processing of large datasets can be done in a reasonable amount of time. Unfortunately, direct methods tend to be sensitive to heavy noise and/or to limited measurements, which can reduce their performance in challenging imaging situations.

A more refined—and now classical—approach to overcome ill-posedness in imaging consists in imposing suitable constraints on the solution [11]. One thereby formulates an energy functional that measures the fidelity of the solution to the measured data, while regularizing the solution to have some desired properties—typically, sparsity [12]. The sought solution minimizes this functional (Chapter 2.3). Such variational methods rely on a faithful modelling of the imaging system, and tend to be more robust than direct algorithms. However, model-based methods usually require iterative schemes that come with a prohibitive computational cost if not carefully engineered, which has limited their wider use in cryo-EM.

Finally, the past years have seen an explosion of data-driven techniques in imaging applications. In particular, the deployment of deep-learning models has had a profound influence in reason of their remarkable ability to autonomously capture data representations [13]. Cryo-EM has been positively impacted too, although learning-based methods developed for the field mostly address the preprocessing steps (*e.g.,* measurement denoising [14]) rather than the 3D reconstruction itself. This is likely to see future developments in view of the good match of cryo-EM for data-driven applications and its very large datasets, although concerns about their robustness and reproducibility could linger given the high stakes.

Considered Cryo-EM Methods

Several cryo-EM imaging variants permit the imaging of samples from the micro to the nano-scale. In this thesis, the focus is on two high-resolution tomographic modalities: scanning transmission electron tomography (STET) and single-particle analysis (SPA). For the sake of clarity, we provide here a concise introduction to both. These methods are presented in greater details in Chapter 1.

STET is dedicated to the imaging of thin 3D slices of biological samples, typically cells or tissues [15]. A narrow beam of electrons is used to raster-scan the specimen tilted at different orientations. The collected 2D measurements are then used to reconstruct the desired volume. The primary concern in STET is to reduce the electron radiation dosage—which is highly damaging to biosamples—while maintaining an acceptable image quality. Robust algorithms are also necessary to handle the heavily degraded measurements and the so-called "missing wedge of information" in the Fourier domain.

SPA aims at characterizing the atomic structure of individual proteins themselves [16]. The procedure starts with the imaging of numerous 3D clones of the object of interest positioned at random unknown orientations. A high-resolution 3D reconstruction is then obtained by letting intricate algorithmic schemes process the 2D projection measurements. The key challenges in SPA reconstruction include heavy noise corruption, the estimation of the unknown orientation of each projection, the handling of massive datasets, and the possible heterogeneity of samples.

Contributions

An overarching theme of this thesis is the aim to integrate available information on the cryo-EM imaging procedure into the reconstruction algorithms, for the purpose of increasing their performance. The most powerful ways for doing so are to rely on a realistic model of the acquisition physics, or to let the algorithms exploit the large datasets through some learning procedure. In this context, our contributions are two-fold.

1. The development of variational *model-based* methods for cryo-EM that can be easily deployed in practice, with a particular focus on the increase in speed and robustness to ill-posed imaging conditions. This is done for both STET (Chapter 3) and SPA (Chapter 4).
2. The proposal of novel *data-driven* methods based on deep neural networks that can be efficiently and robustly used in SPA. Both supervised (Chapter 5) and unsupervised (Chapter 6) learning procedures are explored.

The link between these two paradigms is strong. The proposed learning-based methods directly benefit from our ability to model the cryo-EM procedure, either for the generation of realistic training datasets, or by using it as a module inside the networks architectures themselves. The objective is to leverage the best of both worlds: the guaranteed robustness of model-based approaches, and the remarkable capabilities of learning-based methods at exploiting data statistics.

We hereafter provide a short summary for each of our main research works. All rely on a sound mathematical framework that formalizes the practical objectives at hand, which is also a contribution in itself. It goes without saying that these works are the result of fruitful collaborative efforts. Throughout the manuscript, we indicate in footnotes the publication(s) related to each work, and detail the contribution of each collaborator whenever the author of this thesis is not first author.

Compressed Sensing for STET (Chapter 3)

We present a tailored acquisition-reconstruction STET framework [17] that exploits the principles of compressed sensing and permits to minimize the electron dosage required for high-quality STET imaging. More precisely, we propose to scan only a small random subset of pixels at every sample orientation, and present a regularized reconstruction scheme to recover the 3D sample from these strongly undersampled data. We demonstrate on simulated and real data that reconstruction can then be achieved with very few measurements, hence strongly reducing the effect of sample degradation and preserving image quality.

Related Publication:

L. Donati, M. Nilchian, S. Trépout, C. Messaoudi, S. Marco, and M. Unser, “Compressed Sensing for

STEM Tomography," *Ultramicroscopy*, vol. 179, pp. 47–56, 2017.

Fast Regularized Reconstruction Scheme for SPA (Chapter 4)

We present a fast regularized reconstruction framework for SPA that relies on a rigorous mathematical modeling of the cryo-EM physics [18]. To make the use of such iterative method feasible in SPA, we formulate its cost-dominant step as a discrete convolution whose computational cost does not depend on the number of projections. In addition, the representation of 3D objects with scaled basis functions enables the reconstruction of volumes at any desired scale in the real space. This brings increased robustness and permits further gains in computational speed.

We then use this fast algorithm as a building block of a new refinement framework for SPA that alternates between the reconstruction and the estimation of the unknown orientations [19]. This joint optimization scheme benefits from the robustness brought by our regularized reconstruction scheme, which positively impacts its convergence. The orientations are updated in the continuum through a semi-coordinate-wise gradient descent, which removes the need for the computationally expensive step of classical refinement procedures. We experimentally demonstrate that this joint optimization framework can efficiently refine high-resolution 3D structures from projections with initially inaccurate orientations.

Finally, we further diminish the cost per iteration of our reconstruction algorithm by ingeniously splitting its objective function [20].

Related Publications:

L. Donati, M. Nilchian, C. O. S. Sorzano, and M. Unser, "Fast Multiscale Reconstruction for Cryo-EM," *Journal of Structural Biology*, vol. 204, no. 3, pp. 543–554, 2018.

M. Zehni, L. Donati, E. Soubies, Z. J. Zhao, and M. Unser, "Joint Angular Refinement and Reconstruction for Single-Particle Cryo-EM," *IEEE Transactions on Image Processing*, 2020.

L. Donati, E. Soubies, and M. Unser, "Inner-Loop Free ADMM for Cryo-EM," IEEE International Symposium on Biomedical Imaging (ISBI), 2019.

Supervised Recovery of Orientations SPA: Learning from Projections (Chapter 5)

We present the outline of an ongoing research project for SPA that capitalizes on the powerful learning capabilities of neural networks, yet still fundamentally relies on our ability to faithfully model the cryo-EM imaging process for the generation of the training dataset. Its target is the design of a learning-based method that estimates the unknown orientations in SPA directly from the acquired dataset of projections, *i.e.*, without relying on any intermediate reconstruction procedure.

Our approach relies on the well-known observation that the more similar two projections, the more likely they originated from two 3D particles that adopted close orientations in the ice

layer prior to imaging. Taking this line of thought further, we train a function—parametrized as a neural network—to predict the relative orientation between two projections based on their similarity. Using this trained network, we can then estimate the relative orientations between pairs of projections in any SPA dataset. Our postulate is that we can finally recover, from these estimated relative distances, the orientations themselves through an appropriate minimization scheme.

A new Paradigm for SPA Reconstruction via Deep Adversarial Learning (Chapter 6)

We present CryoGAN, a completely new paradigm for SPA reconstruction that leverages the remarkable capability of deep neural networks to capture data distribution. CryoGAN uses a generative adversarial network (GAN) to learn the 3D structure that has simulated projections that most closely match the real data *in a distributional sense*.

The architecture of CryoGAN resembles that of standard GAN, with the twist that the generator network is replaced by a cryo-EM physics simulator. Thanks to this adversarial scheme, CryoGAN can resolve a 3D structure in a single algorithmic run using only the dataset of projections and CTF estimations as inputs. The algorithm is completely unsupervised, does not rely on an initial volume estimate, and requires minimal user interaction. It is also backed up by a comprehensive mathematical framework that provides guarantees on the recovery of the volume under a given set of assumptions.

Experiments on synthetic and real datasets demonstrate the ability of CryoGAN to capture and exploit real-data statistics in challenging imaging conditions. While the spatial resolution of the CryoGAN reconstructions from real data is not yet competitive with the state-of-the-art, we expect these results to improve in the near future, along with the ongoing progresses in deep learning architecture. In the meantime, the preliminary results obtained with CryoGAN are encouraging and demonstrate the potential of adversarial-learning schemes in image reconstruction.

Related Preprint:

H. Gupta, M. T. McCann, L. Donati, M. Unser. “CryoGAN: A New Reconstruction Paradigm for Single-particle Cryo-EM Via Deep Adversarial Learning”.
https://www.biorxiv.org/content/10.1101/2020.03.20.001016v1

1 Cryo-Electron Microscopy (cryo-EM)

[1]Cryo-electron microscopy (cryo-EM) encompasses a broad range of imaging methods that exploit the wave-like behavior of electrons in vacuum to produce a high-resolution visualization of biological structures. At the heart of all cryo-EM disciplines lies the use of a transmission electron microscope (TEM) to image radiation-sensitive samples under cryogenic conditions. Cryo-EM has been intensively used over the past decades to study cellular and molecular biology at the nanometer scale [21, 22]. It reveals the architecture of cells and proteins, which then brings key insights about their functions. Hence, cryo-EM is a fundamental tool in the search of diagnostic, preventive and curative solutions for a large spectrum of medical diseases and disorders.

In this chapter, we present the operating principles underlying cryo-EM and its variants of interests (STET, SPA), as well their specificity and the key associated challenges. We do not intend here to provide a comprehensive review of those topics. Rather, we aim to equip non-expert readers with the basic knowledge of cryo-EM necessary to an appropriate understanding of our work. For more complete reviews on the subject, we refer to [23, 24, 25].

1.1 Cryo-EM in a Nutshell

1.1.1 The Transmission Electron Microscope (TEM)

Common to all cryo-EM modalities is the use of electron beams for imaging, which permits a significant gain in resolution compared to other microscopy techniques thanks to the extremely short wavelength of electrons (Figure 1.1). Such imaging is performed using a TEM (Figure 1.2a), a device that sends parallel electrons beams through a 3D sample to form a 2D *projection* image[2]. In short, electrons are generated and accelerated by an electron gun, focused in beams by magnetic lenses, and sent toward the specimen inside the TEM column. The electrons then interact with the sample as they propagate through it, which provides

[1]This chapter uses content from our works [17, 18].

[2]The image-formation model behind cryo-EM is detailed in Section 1.2.

indirect information about the 3D structure of interest once magnified by lenses and captured by direct electron detectors (DED). The exact design of a TEM is extremely complex and varies depending on the application, the model, and the manufacturer.

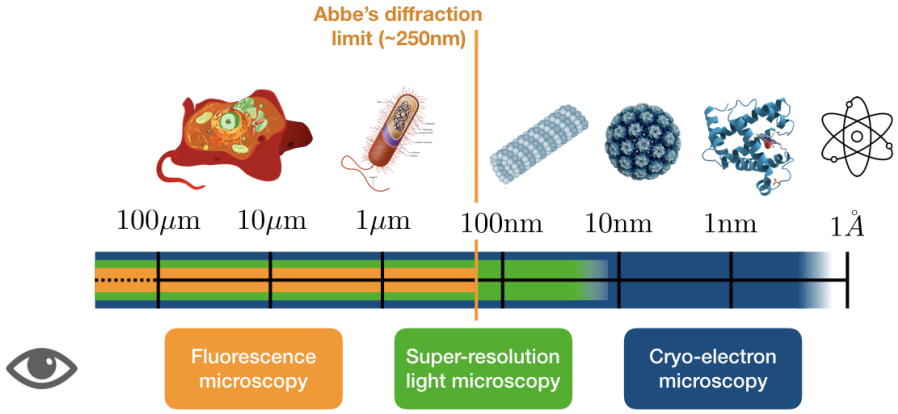
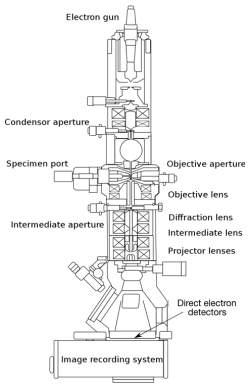


Figure 1.1 – The microscopy scale. Cryo-EM (blue) permits imaging up to the nano-scale, whereas fluorescence microscopy (orange) and super-resolution light microscopy (green) are limited to lower resolutions. However, light-based methods enable the imaging of living samples, which is impossible in cryo-EM.

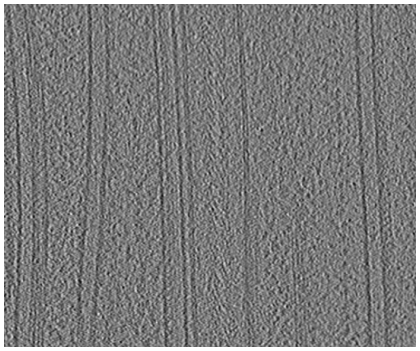
1.1.2 Radiation Sensitivity and Cryogenization of Bio-Samples

The high resolving power of TEM comes at the cost of extreme constraints being imposed on the imaged specimen. For one, high-vacuum conditions are required inside the microscope for the electron beams to travel in straight lines and interact uniquely with the sample. This necessitates to fix the specimen while trying to maximally preserve its native structure—a non-trivial feat [1]. In addition, the inelastic scattering of electrons with the specimen causes extensive radiation damage. This effect is particularly damageable to biological samples. This is problematic because high-resolution imaging requires dense sampling with high electron dosage.

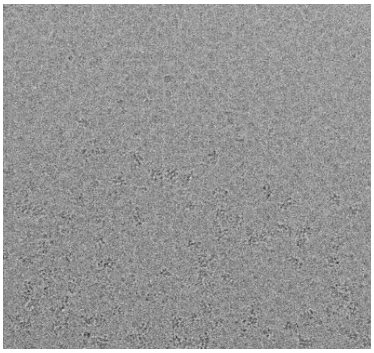
A major technical advance with respect to those two problems has been the development of the plunge-freezing method in the 1980s by Jacques Dubochet and his colleagues [1]. Their *cryogenization* technique permits the imaging of fully hydrated specimen with preserved structure. This is achieved by embedding the sample in a thin layer of water and rapidly freezing it at cryogenic temperature (about -180°C) so that no ice crystal is formed. An additional advantage of the technique is that the imaging of samples at very low temperatures helps mitigate the extent of radiation damage. The effect is rather limited though, and electron-induced radiation damages remains a serious problem in cryo-EM. Hence, high electron voltages have to be sparingly used to image bio-samples. This results in challenging measurements with extremely low signal-to-noise ratio (SNR), as illustrated in Figure 1.2b-c.



(a) Layout of a TEM structure. Image from *wikipedia.org/TEM*.



(b) A cryo-electron tomography measurement [14].



(c) A single-particle cryo-EM micrograph [28].

Figure 1.2 – The **(a)** architecture of a TEM permits high-resolution imaging, but the **(b)-(c)** measurements typically show high noise degradation, which makes their processing difficult.

Remark 1 (Jacques Dubochet)**.** *The Swiss biophysicist and EPFL-graduate Jacques Dubochet is a recurring character in this thesis. In 2017, the Vaudois received, along with Joachim Frank and Richard Henderson, the Nobel Prize in Chemistry for his work on the cryogenization of samples in electron microscopy. This rewarded years of developments in cryo-EM that launched a new era in structural biology, characterized by a boom of interdisciplinary collaborations [26, 27].*

1.2 Image-Formation Model

The perfect modeling of the cryo-EM imaging system is an impossible task. Nevertheless, it is of great practical importance to approximate, as faithfully as can be, the physics behind the acquisition process. The quest for computationally tractable entities is also central for the design of applicable algorithms. Hence, a compromise between accuracy and computability has to be found when modeling physical processes. In this section, we derive the cryo-EM image-formation model that serves as the building block for all our works (Chapters 3- 6).

1.2.1 Geometry of the Imaging Procedure

We start by defining the necessary mathematical objects and the 3D geometry associated to our imaging model. We formulate the complete problem in the continuum and later discuss discretization aspects in Chapter 2.

We model the 3D object being imaged—*i.e.*, a physical entity that varies over space—as a function $f : \mathbb{R}^3 \to \mathbb{R}$ that maps a point $\boldsymbol{x} \in \mathbb{R}^3$ to a point $f(\boldsymbol{x}) \in \mathbb{R}$. In cryo-EM, this function

corresponds to the electric field of the bio-sample of interest, called its Coulomb potential. The function f is often assumed to be compactly supported as the biological objects are spatially localized.

Inside the microscope, the object is positioned at a given 3D orientation with respect to the detector plane $\boldsymbol{y} = (y_1, y_2) \in \mathbb{R}^2$, as illustrated in Figure 1.3 (left). Hence, the geometry of the imaging procedure is described by the geometrical transformation that maps the object coordinate system (x_1, x_2, x_3) to the measurement coordinate system (y_1, y_2). In order to represent all the possible orientations of the 3D object in space, we consider that this mapping corresponds to a rotation in $\mathbf{SO}(3)$, the group of all 3D rotations about the origin of $\mathbb{R}^3$.

In $\mathbf{SO}(3)$, every rotation can be described by a 3×3 orthogonal matrix with determinant 1. As is standard in cryo-EM, we use Euler angles[3] to parametrize the rotation matrix that relates our two coordinate systems [29]. The Euler angles, which we denote as $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, are a set of three angles that describes a sequence of three rotations about three fixed axes (Figure 1.3 (right)). The angle $\theta_1 \in [0; 2\pi)$, called the rotational angle, describes the first rotation around the $\boldsymbol{x}_3$ axis. The angle $\theta_2 \in [0; \pi]$, called the azimutal angle, represents the second rotation around the $\boldsymbol{x}_2$ axis. Finally, the third angle $\theta_3 \in [0; 2\pi)$ corresponds to the in-plane rotation in the measurement plane. We then denote the domain of the Euler angles as $\Omega_{\boldsymbol{\theta}} = [0; 2\pi) \times [0; \pi] \times [0; 2\pi)$. For the sake of conciseness, we hereafter use the term "with/at orientation $\boldsymbol{\theta}$" when referring to an entity (*e.g.*, 3D object, 2D measurement) that is considered in an imaging geometry parametrized by $\boldsymbol{\theta}$.

1.2.2 Object Projection

Due to their extremely short wavelength ($\sim$ 2pm at 300kV), electrons approximately travel in straight lines in vacuum. The initial intensity I_0 of parallel electron beams decays as they propagate through the 3D sample. The transmitted intensity I ends up being recorded as a 2D measurement $b^{\boldsymbol{\theta}}$ following the Beer-Lambert's law [30],

$$b^{\boldsymbol{\theta}}(\boldsymbol{y}) = -\log\left(\frac{I(\boldsymbol{y}, \boldsymbol{\theta})}{I_0(\boldsymbol{y})}\right). \tag{1.1}$$

Under the weak-phase object approximation, the relationship between the 3D object f and its 2D measurement $b^{\boldsymbol{\theta}}$ in (1.1) is mathematically best described through the X-ray transform $\mathscr{P}_{\boldsymbol{\theta}} : L_2(\mathbb{R}^3) \to L_2(\mathbb{R}^2)$ [31]. The X-ray transform is a linear operator that maps a 3D function into its 2D line-integral image, called the *projection* image, along the projection direction $\vartheta_{\boldsymbol{\theta}} = (\cos\theta_1 \sin\theta_2, \sin\theta_1 \sin\theta_2, \cos\theta_2) \in \mathbb{S}^2$, with $\mathbb{S}^2 = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\|_2 = 1\}$ the unit sphere in $\mathbb{R}^3$. Hence, we model an idealized cryo-EM measurement as

$$b^{\boldsymbol{\theta}}(\boldsymbol{y}) = \mathscr{P}_{\boldsymbol{\theta}}\{f(\boldsymbol{x})\}(\boldsymbol{y}) = \int_{\mathbb{R}} f(t\vartheta_{\boldsymbol{\theta}} + \mathbf{M}_{\boldsymbol{\theta}^\perp}^T \boldsymbol{y})\mathrm{d}t, \tag{1.2}$$

[3] Although commonly used in 3D cryo-EM, the Euler angles convention comes with a few technical challenges [29], as we shall later discuss in Chapter 5.

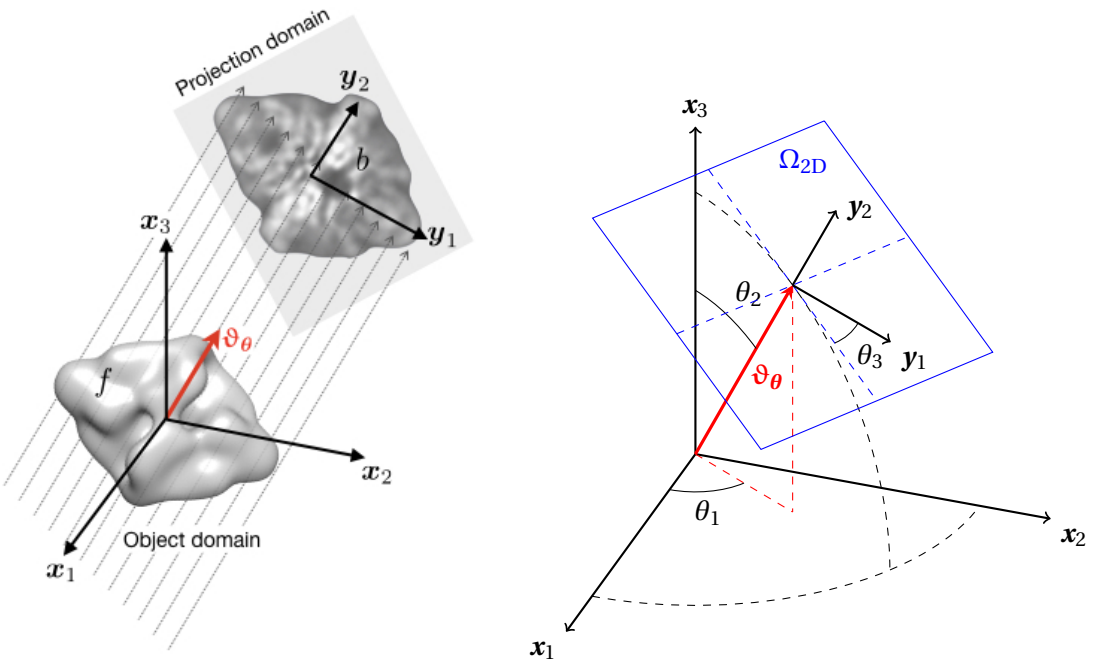


Figure 1.3 – Geometry of the 3D imaging model. The 3D object f in the coordinate system (x_1, x_2, x_3) is imaged along the projection direction $\vartheta_{\boldsymbol{\theta}}$ to produce the 2D projection b in the coordinate system (y_1, y_2). The Euler angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \Omega_{\boldsymbol{\theta}}$ compactly represent the 3D rotation that maps the object coordinate system to the projection coordinate system. The angles θ_1, θ_2, and θ_3, respectively correspond to the rotation, the tilt, and the in-plane rotation in the projection plane. The set Ω_{2D} denotes the support of the projection.

where $\mathbf{M}_{\boldsymbol{\theta}^\perp}^T$ is the adjoint of the hyperplane projection matrix $\mathbf{M}_{\boldsymbol{\theta}^\perp} \in \mathbb{R}^{2\times 3}$ given by

$$\mathbf{M}_{\boldsymbol{\theta}^\perp} = \begin{pmatrix} \cos\theta_1\cos\theta_2\cos\theta_3 - \sin\theta_1\sin\theta_3 & \cos\theta_1\cos\theta_2\sin\theta_3 + \sin\theta_1\cos\theta_3 & -\cos\theta_1\sin\theta_2 \\ -\sin\theta_1\cos\theta_2\cos\theta_3 - \cos\theta_1\sin\theta_3 & -\sin\theta_1\cos\theta_2\sin\theta_3\cos\theta_1\cos\theta_3 & \sin\theta_1\sin\theta_2 \end{pmatrix}. \quad (1.3)$$

The matrix $\mathbf{M}_{\boldsymbol{\theta}^\perp}$ has rows that specify the normal basis of the hyperplane perpendicular to the projection direction $\vartheta_{\boldsymbol{\theta}}$. Intuitively, this matrix expresses how a point in the object domain gets geometrically positioned in the projection domain, while its adjoint $\mathbf{M}_{\boldsymbol{\theta}^\perp}^T$ maps the reciprocal relationship.

1.2.3 Optical Effects and Detection

Equation (1.2) is a good starting point in the modelling of the cryo-EM imaging procedure, but it only tells a limited part of the story. In practice, the beams exiting the specimen undergo complex optical effects in the TEM microscope before they reach the electron detectors. This results in an alteration of the projection frequency content that is accounted for through a convolution with the shift-invariant 2D point-spread function (PSF) $h : \mathbb{R}^2 \to \mathbb{R}$ of the TEM

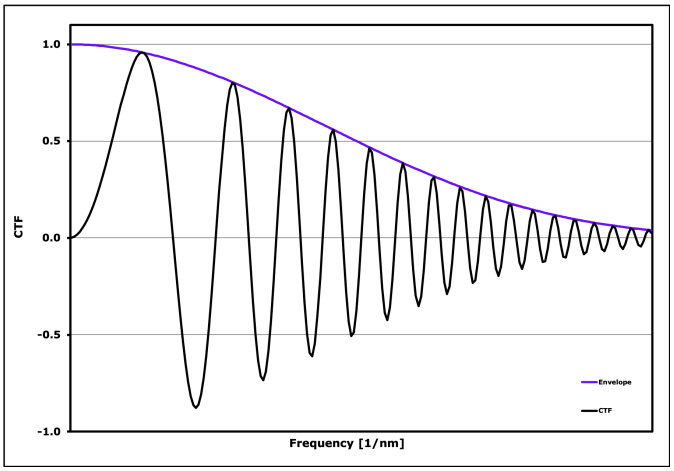


Figure 1.4 – Example of a CTF in TEM (radial plot). The function (in Fourier space) typically displays rapid oscillations, multiple zero-crossings, and is dampened by a decreasing envelope function. Image simulated with *www.c-cina.org/tools/soft/ctf-simulation/*.

(*i.e.*, its impulse response), such that

$$b^{\boldsymbol{\theta}}(\boldsymbol{y}) = \big(\mathscr{P}_{\boldsymbol{\theta}}\{f\} * h\big)(\boldsymbol{y}). \tag{1.4}$$

In cryo-EM, these optical effects are mostly expressed in the Fourier domain as a multiplication with the Fourier transform $\widehat{h}$ of the PSF, called the contrast transfer function (CTF). The CTF is a rapidly oscillating and decreasing function with multiple zero-crossings (Figure 1.4). Its theoretical model derives from the phase-object approximation [32] and is most often given for $\boldsymbol{\omega} \in \mathbb{R}^2$ by

$$\widehat{h}(\boldsymbol{\omega}) = E(\boldsymbol{\omega})\, A(\boldsymbol{\omega}) C(\boldsymbol{\omega}), \tag{1.5}$$

where E is called the envelope function, A corresponds the objective aperture function, and C represents the phase contrast transfer function. The first two functions are mostly responsible for the limitation in resolution, while the third one explains the oscillatory behavior of the CTF. More details on the modeling of the CTF are given in Appendix A.2.

These CTF features—dampening, oscillations and zero-crossings—typically distort the projection amplitude and affect the resolution if left uncorrected. In practice, CTF-correction is often performed on every 2D projection prior to 3D reconstruction[4]. This is most often done by estimating the parameters of the theoretical CTF model (1.5) from the acquired measurements (*e.g.*, by analyzing their power spectra). This is a challenging procedure by itself given the high level of noise that degrades the cryo-EM data (see Section 1.2.4 below).

The imaging model in (1.4) describes the measurements in their continuous form. In

[4] Some cryo-EM variants consider the possibility for each projection to be affected by a different (shift-invariant) CTF. We shall discuss this point in due time.

practice, this information is captured at a large number of equally spaced discrete points by electron detectors. Hence, we assume that the measurements $b^{\boldsymbol{\theta}}(\boldsymbol{y})$ are acquired at the sampled points $\boldsymbol{y}_{\boldsymbol{j}} = \boldsymbol{j}\Delta$ for $\boldsymbol{j} \in \Omega_{2\mathrm{D}}$, where the set $\Omega_{2\mathrm{D}} \subset \mathbb{Z}^2$ denotes the support of the projection and is constituted of $M = \sharp\Omega_{2\mathrm{D}}$ elements. For the sake of clarity, we consider $\Delta = 1$ (without loss of generality) and we denote as $\widetilde{\mathbf{b}}^{\boldsymbol{\theta}} \in \mathbb{R}^M$ the discrete noiseless measurement vector for an object with orientation $\boldsymbol{\theta}$.

1.2.4 Noise Degradation

The final consideration is to account for random perturbations that degrade the captured information. In cryo-EM, different types of noise can undesirably affect the measurements [24], but the two major noise sources originate from the interaction of electrons with ice and from the electron counting by the TEM detectors. This counting process, which is by nature discrete, leads to a noise that is usually dominated by Poisson statistics. However, when large number of detection events arise, the electrons are no longer individually observed and the Poisson distribution approaches a Gaussian one. We thus assume a cryo-EM projection $\widetilde{\mathbf{b}}^{\boldsymbol{\theta}}$ to be corrupted by a substantial additive Gaussian noise $\mathbf{n}^{\boldsymbol{\theta}} \sim \mathcal{N}(0, \sigma^2 \cdot Id)$ as suggested in [33, 34], with $\mathbf{n}^{\boldsymbol{\theta}} \in \mathbb{R}^M$, such that

$$\mathbf{b}^{\boldsymbol{\theta}}_{\boldsymbol{j}} = \widetilde{\mathbf{b}}^{\boldsymbol{\theta}} + \mathbf{n}^{\boldsymbol{\theta}}_{\boldsymbol{j}} = \big(\mathscr{P}_{\boldsymbol{\theta}}\{f\} * w\big)(\boldsymbol{j}) + \mathbf{n}^{\boldsymbol{\theta}}_{\boldsymbol{j}}. \tag{1.6}$$

More precise noise models than in (1.6) exist (*e.g.*, which include noise sources affected by the CTF), but those also tend to significantly complexify the subsequent processing steps. Finally, it is worth insisting that, whatever the model considered, noise degradation in cryo-EM remains a massive processing challenge: the measurements typically have one of the lowest SNR of any imaging modality [35].

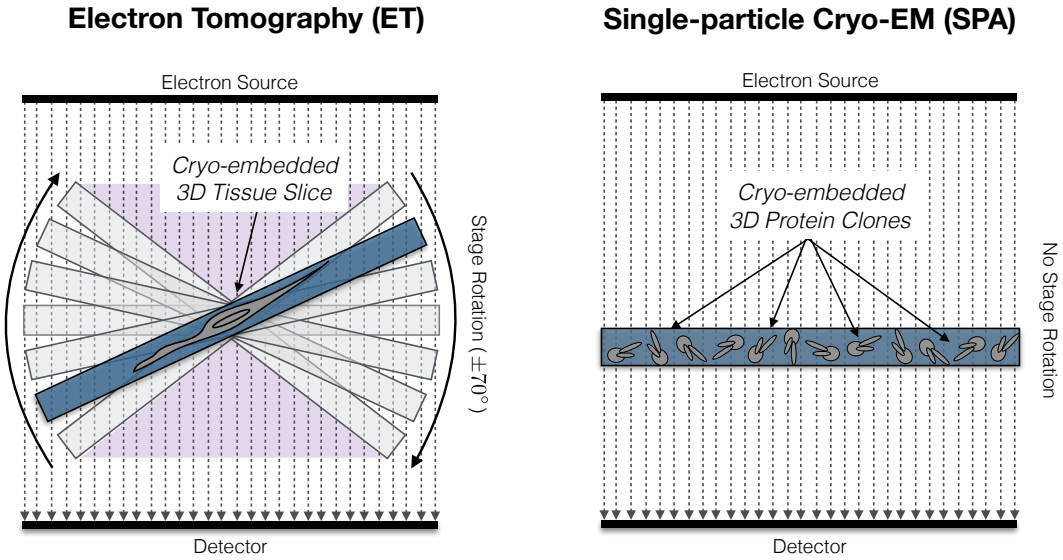
Equipped with the imaging model (1.6) in which the X-ray transform plays a central role, the idea is then to image the object under different orientations, and use *tomographic* reconstruction algorithms to obtain a 3D visualization of its structure (Chapter 2). In practice, this reconstruction task is always digitally handled. Hence, careful discretization of the object f is necessary for practical applications; this is the topic of Section 2.2.

1.2.5 Tomographic Variants in Cryo-EM

Several cryo-EM operating modes exist for bio-imaging, and most tomographic setups fall in one of two categories (Figure 1.5):

1. **Electron tomography (ET)**, in which a 3D tissue slice is imaged by sequentially tilting the sample-stage around a single-axis of rotation (Section 1.3). Hence, each 2D measurement corresponds to the projection of the 3D sample at a given stage orientation, altered by some optical effects and noise.

2. **Single-particle analysis (SPA)**, in which hundreds of thousands of protein clones with random orientations are rapidly freezed in a thin layer of ice and subsequently imaged (Section 1.4). Each acquired 2D image thus corresponds to the projection of a 3D clone under an unknown orientation, altered by some optical effects and noise.



(a) Schematic ET setup. The purple area indicates the angular region left uncovered due to technical constraints, which results in a missing cone of information in the Fourier domain.

(b) Schematic SPA setup. Numerous clones of a molecule, each with its unknown orientation, are freezed in ice and subsequently imaged.

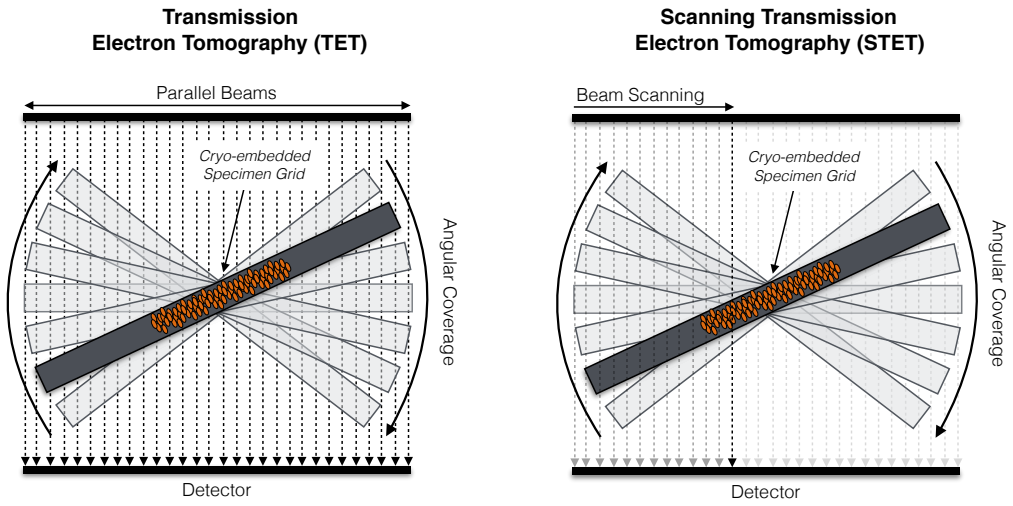
Figure 1.5 – Illustrations of the cryo-EM variants for tomographic imaging: (a) ET and (b) SPA.

ET and SPA share a number of similarities, most notably the cryogenization of samples, the use of a TEM for imaging, heavy noise-degradation of the measurements, and the reliance on tomographic reconstruction algorithms. Hence, the image-formation model is the same for both variants and is given by (1.6).

That being said, ET and SPA do differ in some important aspects, in particular regarding the sample preparation, the tilting strategy, and the achieved resolution. Those differences (Table 1.1) lead to reconstruction challenges that are specific to each imaging method.

	Sample	**Resolution**	**Stage Rotation**	**Missing Information**
ET	Tissue/Cell Slice	Nano (1-10nm)	Single-axis (±70°)	Unimaged Tilt Wedge
SPA	Protein Clones	Near-atomic (1-10Å)	None	Proteins Orientations

Table 1.1 – Main experimental differences between ET and SPA.



(a) Cryo-TET set-up. Parallel beams of electrons are simultaneously sent through the tilted sample to form a projection image.

(b) Cryo-STET set-up. Lenses focus the electrons in a single beam and the tilted sample is rastered to form a projection image.

Figure 1.6 – Illustrations of two cryo-ET modalities: (a) TET and (b) STET.

1.3 Electron Tomography (ET)

The most generic form of ET is transmission electron tomography (TET), which we briefly present in Section 1.3.1. In this thesis, the ET method of interest is scanning transmission electron tomography (STET[5]), a variant of TET that we describe in Section 1.3.2.

1.3.1 Transmission Electron Tomography (TET)

The core principle in TET is to incrementally tilt, around a single axis of rotation and over a large range of tilt angles, a thin 3D sample through which parallel electron beams are sent at each tilt angle (Figure 1.6a). The recorded 2D projections, called the tilt-views, compose the tilt-series, which is then processed to obtain a 3D reconstruction.

In conventional TET systems, the thickness of the sample and the geometry of the specimen holder preclude imaging through a full 180° angular range (see purple area in Figure 1.5a) [36]. Artifacts consequent to this missing wedge of information in the Fourier domain may appear on the reconstructed image if the lack of angular coverage is excessive and/or if the reconstruction algorithm is not sufficiently robust. TET is currently restricted to the imaging of thin bio-samples (<350nm), which ensures correct propagation of the electrons even at high tilt ($\pm 70°$).

[5]The modality is also known as STEM in the community.

1.3.2 Scanning Transmission Electron Tomography (STET)

The main difference between TET and STET is that in STET, the specimen is *scan-rastered* by a focused electron beam, and the transmitted radiation is gradually detected (Figure 1.6b). Thanks to this scanning approach, STET offers several advantages over TET for the imaging of bio-samples, including an increased resolving power, an improved contrast, and better SNR [37, 38, 39]. It also broadens the scope of acceptable biological specimens [15]. Yet, STET still suffers from the problems of missing information and radiation-sensitivity. Hence, a trade-off between the acquisition quality (*i.e.*, wide and numerous high-SNR acquisitions) and the sample integrity (*i.e.*, low electron dosage) must be considered during imaging.

In that respect, the flexibility in the choice of the scanning approach offers some interesting opportunities. The design of an acquisition-reconstruction STET framework that combines a tailored scanning pattern and a reconstruction algorithm with increased robustness [17] is the subject of Chapter 3. We briefly review below the approaches related to this work.

Literature Review

Several works have proposed reconstruction methods for STET that address the missing-wedge problem and permit lower-dose acquisition without introducing significant degradation in the final volume. These methods are categorized in two main classes depending on whether the dose reduction is achieved by tilt/angular downsampling or by image/spatial downsampling [40].

Tilt-downsampling (T-DS) techniques refer to algorithms that perform reconstruction from a reduced number of tilt images. T-DS is generally performed by acquiring fewer projections over the widest possible angular range. To compensate for the induced lack of information, a standard approach is to incorporate prior knowledge into the reconstruction process (see Section 2.3). Such advanced T-DS reconstruction algorithms primarily include discrete algebraic reconstruction techniques (DART) [41, 42] and compressing sensing (CS) based methods [43, 44, 45, 46].

Image-downsampling (I-DS) techniques take a different route and reduce the electron beam coverage of individual tilt images. This can be achieved by decreasing the frame size, the per-pixel dwell time, or the beam current density [47]. Alternatively, one can scan only a fraction of the image pixels following a certain downsampling pattern. The recovery of *random* I-DS projections has been the topic of several publications. Most notably, Stevens *et al.* [48] applied a Bayesian dictionary-learning technique to restore such measurements, while Saghi *et al.* [40] used total variation (TV) to inpaint the I-DS measurements prior to feeding them to a CS-based reconstruction algorithm. To the best of our knowledge, full 3D tomographic reconstruction directly from the unrestored random I-DS STET data had not yet been demonstrated prior to our work.

1.4 Single-Particle Analysis (SPA)

Single-particle analysis (SPA[6]) differs from ET in that its final goal is to characterize the *atomic model* of proteins, *i.e.*, the spatial organization of each atom in their chains of amino acids (Figure 1.7c). The motivation is that proteins carry out crucial functions in cells, and alterations in their structure usually affect their ability to perform these tasks. Hence, SPA plays a major role in structural biology and pharmaceutical research.

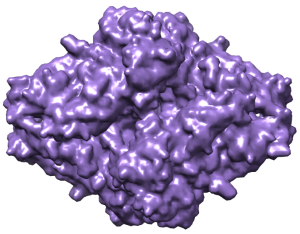
A typical end-to-end SPA pipeline consists of three main phases [23].

1. The preparation of the sample and its imaging in cryogenic conditions with a TEM (Figures 1.8a).

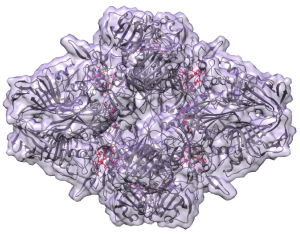
2. The reconstruction of the high-resolution 3D density map[7] (Figure 1.7a) from the 2D projection measurements using digital image-processing algorithms (Figures 1.8b-c).

3. The determination of the atomic model (Figure 1.7c) based on the reconstructed density map and prior knowledge on the physics of protein folding.

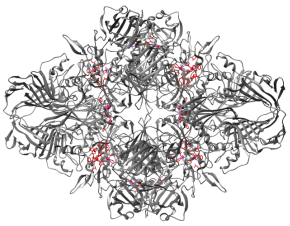
In this thesis, our focus is on the design of algorithms that reconstruct the density map (Step 2). As the challenges faced by reconstruction algorithms in SPA derive from the specificity of the imaging procedure (Step 1), we start by briefly presenting its key aspects. The procedure of atomic modelling (Step 3) is beyond the scope of our work; for further details, we refer to [49].



(a) Density map of β-gal at a resolution of 7Å.



(b) Superimposition of the atomic model and the density map of β-gal for visualisation purpose.



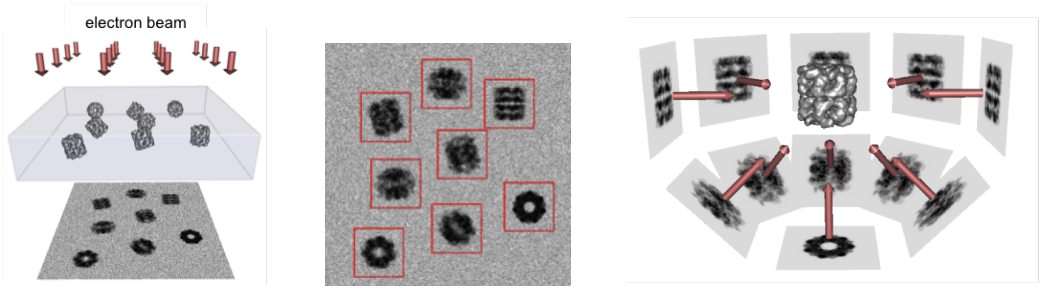
(c) Atomic model of β-gal.

Figure 1.7 – Visualization of the density map and the atomic model of the β-galactosidase enzyme (PDB-5a1a) [50]. Images produced with Chimera [51]

[6] The modality is also known in the field as single-particle cryo-EM. To ensure coherence with our published works, we shall use the term SPA in this thesis.

[7] In practice, the beams of electrons interact with the electron cloud of the protein. Hence, the imaging procedure actually provides a direct representation of its 3D electron density, called the *density map*. In the next chapters, we shall use the more generic term of *3D structure* for the sake of clarity.

Remark 2 (The Boom of SPA)**.** *SPA is not a young imaging method: its development began in the 1970s. However, its popularity has rocketed in recent years due to technical advances in detector technology and software algorithms. This launched a "resolution revolution" [5] that saw the determination of structures progress up to near-atomic resolution. This momentum culminated in 2017 with the Nobel Prizes of its pioneers Dubochet, Frank and Henderson.*



(a) Imaging Procedure.

(b) Pre-processing of the micrographs.

(c) Reconstruction of the density map from the projections.

Figure 1.8 – Illustrations of three key steps in SPA for the reconstruction of the density map. Note that the contrast and SNR of the measurements in the micrograph have been exaggerated for visualization purposes. In practice, the particles are barely visible. Images from *people.csail.mit.edu/gdp/cryoem*.

1.4.1 Imaging Procedure

The key to appreciate the specificity of SPA is to first understand that the conventional tomographic approach—*i.e.*, to rotate a single object and take measurements at every stage rotation—is not a viable option for the imaging of single proteins. The reason is that their radiation sensitivity is so high that they cannot withstand the repetitive exposures to electron beams necessary in the standard setup.

The twist to get around this is to work with numerous 3D copies of the same protein, called "particles", which are supposed to be structurally identical. The idea is to let these particles adopt random positions and orientations in a thin layer of water, freeze them at cryogenic temperatures, and image all particles simultaneously with parallel electron beams in a minimal number of exposures (Figure 1.8a). The collected measurements are called micrographs (Figure 1.8b), and contain thousands of projections of the protein under different orientations, which can then be used for tomographic reconstruction (Figure 1.8c).

The imaging procedure in SPA is incredibly ingenious, but its implementation comes with a series of serious challenges. The most obvious one is that the orientations taken by the particles in the ice layer are random and thus unknown; yet, they are essential for reconstruction. In addition, interactions with the ice surface can sometimes drive particles

to favor certain orientations, which leads to a nonuniform angular coverage of the structure. Moreover, the acquired projections are always extremely numerous, heavily degraded by noise, and affected by complex optical effects, which makes their processing difficult.

Another major challenge with more recent (and thus trickier) samples is that many proteins are actually dynamic entities: they perform functions in cells by changing their configurations upon receiving some stimuli. The range of configurations adopted by the protein, called its conformational landscape, means that the micrographs may actually contain a mix of projections from distinct 3D structures [52]. If left untreated, this can lead to a blurred incorrect reconstruction.

To summarize, the main difficulties that have to be considered when designing reconstruction algorithms for SPA are:

1. The unknown orientation of each projection.
2. The strong optical effects that modulate the projections.
3. The extremely low contrast and SNR of the projections.
4. The massive amount of measurements to process.
5. (Possible) The nonuniform distribution of the particles' orientations.
6. (Possible) The presence of degraded, unusable data in the measurements.
7. (Possible) The structural heterogeneity of the sample.

This collection of problems makes the reconstruction task in SPA an enduring technical challenge: scientists have spent the better part of the last 30 years designing a solid pipeline that can reliably deliver 3D structures with atomic resolution. The result is an intricate multi-steps procedure that permits the regular discovery of new structures, but that can still be prone to overfitting and irreproducibility.

1.4.2 Reconstruction of the Density Map

Prior to 3D reconstruction, a series of essential preprocessing steps are first needed to localize, extract and normalize the individual 2D projections from the micrographs [53, 54, 55, 56, 57]. Only then can one use intricate algorithmic schemes to combine the large set of measurements, with the ultimate goal of producing a high-resolution 3D reconstruction (*i.e.*, smaller than 3-4Å). To handle the strongly incomplete data (in particular, the unknown orientations), most software packages [58, 8, 9, 10, 59, 60, 61, 62] implement a so-called *iterative-refinement* pipeline during which information is gradually added to a rough initial volume[8].

[8] The search for an appropriate initialization is known in SPA as *ab initio* modelling [63]. It is a very active field of research in itself [64, 65, 66, 62, 67], which is not surprising given the stakes.

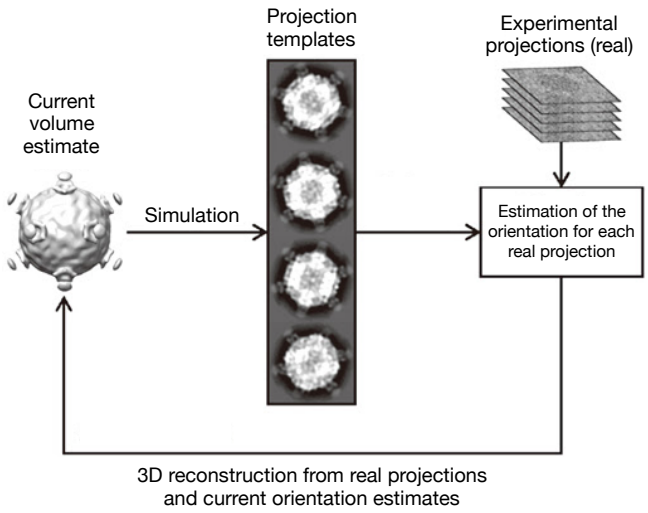


Figure 1.9 – Schematic of the projection-matching approach. Image adapted from *cpb.iphy.ac.cn*.

Remark 3 (Terminology)**.** *The iterative-refinement pipeline in SPA sometimes makes the use of the term "reconstruction" confusing. Depending on the context, it can either refer to the reconstruction of the final density map as a global process, to the action of the reconstruction algorithm as a local procedure, or to any 3D structure obtained throughout the pipeline. We shall try to make its meaning clear when the context does not.*

Iterative-Refinement Pipelines

Projection-matching approaches (Figure 1.9) refine the initial volume by alternating between estimation of the orientations and 3D reconstruction from the projections given the current (however inaccurate) parameter estimates [68, 69]. The first rough 3D structure is often computed from high-SNR class averages—a complicated task in itself given the challenging imaging conditions. From this first volume, one produces a finite number of synthetic projections with uniformly distributed orientations. Those projection templates are used to predict the relative orientation of every 2D projection in the experimental dataset through some appropriate angular-assignment method [70]. This process is then repeated with an increasing number of distinct synthetic projection templates until the optimization fulfills some convergence criterion.

Bayesian approaches [71, 72, 73] differ in the way they handle the unknown orientations during this alternating refinement pipeline: they do not constrain the 2D projections to be assigned a unique orientation estimate. In practice, they rely on the formulation of a maximum marginalized *a posterior* (MAP) estimation that is solved by expectation maximization.

The process starts with the calculation of the posterior probabilities for all possible orientations based on statistical noise models. These probabilities are then used to weight the contribution of each 2D projection to every orientation class (as defined by the projection templates). This treatment of the unknown orientations, called marginalization, is less sensitive to the initial model and brings increased robustness in high-noise regimes compared to projection-matching procedures. However, the marginalization process is often computationally expensive. A recent work notably improved the computational efficiency of the MAP scheme by solving it through stochastic average gradient descent (SGD) and using importance sampling to further reduce the cost of computing the marginalized likelihood [74].

For both approaches, the global refinement process equates to a high-dimensional non-convex optimization problem with numerous local minima, whose outcome is still predicated on the quality of the initial reconstruction [75, 76]. Moreover, both are computationally challenging pipelines that can demand large resources or rely on approximations (*e.g.*, the reliance on sums instead of integrals during the marginalization procedure) [77].

Tomographic Reconstruction Algorithms

All iterative-refinement pipelines rely on a reconstruction procedure at every step of the pipeline. This reconstruction is carried out independently of the orientation estimation. In most instances, software packages use direct algorithms based on the central-slice theorem (Section 2.1) to perform this reconstruction [78, 79, 70]. Direct methods work adequately when the projections are of good quality and sufficiently numerous, and their speed is a key advantage. Unfortunately, their performance can be limited in adverse imaging situations (Section 2.3.1).

A more sophisticated and robuster reconstruction approach is to formulate it as a regularized inverse problem (see Section 2.3) that is solved iteratively [80, 81, 82, 83, 84]. Some approaches also take into account the blurring of each projection by the CTF of the microscope [85, 32]. Those iterative methods permit high-quality reconstruction but often require very large computational resources if not carefully engineered. The design of a fast, regularized reconstruction framework for SPA that is robust to the challenging imaging conditions is the subject of Chapter 4.

We conclude this introduction to SPA by mentioning that, although deep-learning models have already had a profound impact in a wide range of image reconstruction applications [86, 87, 88], their current utilization in SPA is mostly restricted to preprocessing steps, such as micrograph denoising [89] or particle picking [90, 91, 92, 93, 94]. The design of a method that learns to estimate the unknown orientation associated to each projection in a SPA dataset is the topic of Chapter 5. In Chapter 6, we present a completely new paradigm for SPA reconstruction based on deep adversarial learning, which we believe to be the first demonstration of a deep-learning architecture able to perform the full SPA reconstruction procedure without any prior training.

2 Model-based Tomographic Reconstruction for Cryo-EM

2.1 Context

[1]The technique of mathematically combining 2D projections acquired from various directions into a 3D volume is known as *tomography*. Tomography is used in a variety of imaging methods beyond cryo-EM (*e.g.*, in X-ray computed tomography (CT) [95]) to gather information about the internal structure of objects through indirect measurements obtained with penetrating waves.

Remark 4 (EM and tomography)**.** *The story between tomography and EM goes back a long time. Interestingly, it even precedes the development of X-ray CT by Hounsfield and Cormack [95]. In their landmark 1968 paper [96], Aaron Klug and David DeRosier elucidated the structure of the T4-bacteriophage by combining a small set of 2D EM images—the dawn of tomographic EM! A former student of the brilliant crystallography pioneer Rosalind Franklin, Klug developed reconstruction methods based on the central-slice theorem (or, as Klug put it, "a theorem familiar to crystallographers") that are still of relevance today (see Section 2.1.1). Sir Aaron Klug later received the Nobel Prize in Chemistry for his outstanding contributions, which set the stage for a long series of Nobel Prizes dedicated to this revolutionary imaging method.*

2.1.1 Direct Inversion Methods

Tomographic reconstruction is most often performed using direct inversion methods. A prominent approach is the filtered back projection (FBP) algorithm [97], which implements a discrete version of the analytically-derived X-ray inversion formula [31]. In FBP, the projections are filtered with an appropriate convolution kernel before being back projected in the object domain. As the X-ray inversion formula is also an integral operator, FBP relies on the approximation of integrals by sums and the use of interpolation to evaluate quantities

[1]This chapter uses content from our works [17, 18, 20].

between samples whenever needed.

In cryo-EM, many software packages rather rely on direct Fourier-reconstruction (DFR) methods [78, 79] that are based on the *central slice theorem*. This theorem relates the Fourier transform of a 2D projection to an hyperplane (orthogonal to the projection direction) in the Fourier transform of the 3D object. Formally, this writes as

$$\mathscr{F}_{3\mathrm{D}}\{f\}\big(\mathbf{M}_{\boldsymbol{\theta}^{\perp}}^{T}\boldsymbol{\omega}\big) = \mathscr{F}_{2\mathrm{D}}\{\mathscr{P}_{\boldsymbol{\theta}}\{f\}\}(\boldsymbol{\omega}), \tag{2.1}$$

where $\mathscr{F}_{2\mathrm{D}} : L_2\left(\mathbb{R}^2\right) \to L_2\left(\mathbb{R}^2\right)$ and $\mathscr{F}_{3\mathrm{D}} : L_2\left(\mathbb{R}^3\right) \to L_2\left(\mathbb{R}^3\right)$ denote the 2D and 3D Fourier transform operators respectively, and $\boldsymbol{\omega} \in \mathbb{R}^2$. We recall that $\mathbf{M}_{\boldsymbol{\theta}^{\perp}} \in \mathbb{R}^{2\times 3}$ is the hyperplane projection matrix whose expression is given in (1.3). A popular DFR approach uses interpolation kernels in the Fourier domain to bring irregularly-distributed samples onto a regular grid for each 2D projection, before applying an inverse 3D fast Fourier transform (FFT) to the set of regridded projections [70].

The success of direct inversion methods owes primarily to their speed and their satisfactory performance when handling numerous, high-quality measurements. However, the X-ray inversion formula is based on assumptions that can get violated when the measurements are lacking (*e.g.*, missing wedge of information in STET) or heavily degraded by noise. Hence, the efficiency of direct methods is usually limited in adverse imaging situations, which can be especially problematic in cryo-EM. An alternative is to rely on more robust iterative optimization approaches. These methods necessitate the modeling of the imaging process with a discrete *system matrix* $\mathbf{H}$ that is physically accurate and can be handled computationally. This entity is at the core of our works and permits the formulation of the discrete *forward model* which relates the acquired 2D measurements to the unknown 3D structure of interest.

2.2 Discrete Forward Model in Cryo-EM

We first recall that the cryo-EM image-formation model, which we fully derived in the continuum in Section 1.2, is given by

$$\mathbf{b}_{\boldsymbol{j}}^{\boldsymbol{\theta}} = \big(\mathscr{P}_{\boldsymbol{\theta}}\{f\} * h\big)(\boldsymbol{j}) + \mathbf{n}_{\boldsymbol{j}}^{\boldsymbol{\theta}}, \tag{2.2}$$

where

- $f : \mathbb{R}^3 \to \mathbb{R}$ is the 3D object being imaged;
- $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \Omega_{\boldsymbol{\theta}}$ parametrizes the problem geometry in $\mathbf{SO}(3)$;
- $\mathscr{P}_{\boldsymbol{\theta}} : L_2\left(\mathbb{R}^3\right) \to L_2\left(\mathbb{R}^2\right)$ is the X-ray transform operator for an orientation $\boldsymbol{\theta}$;
- $h : \mathbb{R}^2 \to \mathbb{R}$ is the PSF of the TEM microscope;

- $\mathbf{b}^{\boldsymbol{\theta}} \in \mathbb{R}^M$ is the 2D measurement vector (*i.e.*, the projection[2]) for an orientation $\boldsymbol{\theta}$;
- $\mathbf{n}^{\boldsymbol{\theta}} \in \mathbb{R}^M$ is the additive Gaussian noise vector associated to $\mathbf{b}^{\boldsymbol{\theta}}$;
- $\boldsymbol{j} \in \Omega_{\text{2D}} \subset \mathbb{Z}^2$ indicate the indices of the M-sized vectors, with $M = \sharp\Omega_{\text{2D}}$.

For practical reasons, we make the choice to look for an approximation $\tilde{f} : \mathbb{R}^3 \to \mathbb{R}$ of f that accurately represents the 3D object in a computationally-convenient, shift-invariant basis.

2.2.1 Discretization Scheme

Following the generalized sampling scheme [98], we parametrize $\tilde{f}$ in the reconstruction space

$$V(\varphi) = \left\{ \tilde{f} = \sum_{\boldsymbol{k} \in \mathbb{Z}^3} c[\boldsymbol{k}] \varphi(\cdot - \boldsymbol{k}) : c \in \ell_2\left(\mathbb{Z}^3\right) \right\}, \tag{2.3}$$

where $\varphi \in L_2(\mathbb{R}^3)$ is a suitable basis function. In practice, the coefficient sequence c is restricted to a finite number of coefficients as the object $\tilde{f}$ and the basis function φ we consider (see below) are compactly supported. We write this vector of coefficients as $\mathbf{c} = \left(c[\boldsymbol{k}]\right)_{\boldsymbol{k} \in \Omega_{\text{3D}}}$. Here, the set $\Omega_{\text{3D}} \subset \mathbb{Z}^3$ corresponds to the support of the coefficients required to represent the object $\tilde{f}$, and the number of elements in Ω_{3D} is $N = \sharp\Omega_{\text{3D}}$.

In practice, the coefficients $\mathbf{c} \in \mathbb{R}^N$ are optimized during the reconstruction procedure (Section 2.3). Once the optimization is done, the obtained coefficients can be re-expanded in the space $V(\varphi)$ through (2.3) to obtain the continuous representation of the reconstructed volume.

A suitable choice for the basis function φ is the optimized Kaiser-Bessel window function (KBWF) [99, 100] defined as

$$\varphi(\boldsymbol{x}) = \begin{cases} \dfrac{\left(1 - \left(\frac{\|\boldsymbol{x}\|}{a}\right)^2\right)^{\frac{m}{2}} I_m\left(\alpha\sqrt{1 - \left(\frac{\|\boldsymbol{x}\|}{a}\right)^2}\right)}{I_m(\alpha)} & \text{if } 0 \le \|\boldsymbol{x}\| \le a, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The KBWF depends on three parameters: 1) the order m of the modified Bessel function I_m, 2) the window taper α, and 3) the support radius a. KBWFs satisfy several key desirable properties of basis functions: they form a Riesz basis [101], they are compactly supported, and they have minimal deviation from the partition-of-unity condition[3] [98] for specific sets of parameter values (*e.g.*, $m = 2$, $\alpha = 10.83$, $a = 2$) [100]. Moreover, their isotropic property allows for a significant reduction in computational costs, as we shall later illustrate in Chapter 4.

[2] Unless otherwise indicated, we shall use the term "projection" throughout this thesis to concisely refer to a noisy, PSF-convolved 2D cryo-EM projection measurement.

[3] It was shown in [100] that the properties of "compact support" and "partition-of-unity" were mutually exclusive for isotropic basis functions.

2.2.2 System Matrix and Forward Model

We now make the assumption that the measurement $\mathbf{b}^{\boldsymbol{\theta}}$ in (2.2) are taken from the approximation $\tilde{f}$ of f given by (2.3), and we aim at recovering its coefficients $\mathbf{c}$. Using the linearity and shift-invariance properties of the X-ray transform (see Appendix A.1), we rewrite the cryo-EM image-formation model as

$$\mathbf{b}_{\boldsymbol{j}}^{\boldsymbol{\theta}} = \big(\mathscr{P}_{\boldsymbol{\theta}}\{\tilde{f}\} * h\big)(\boldsymbol{j}) + \mathbf{n}_{\boldsymbol{j}}^{\boldsymbol{\theta}}, \tag{2.5}$$

$$= \sum_{\boldsymbol{k}\in\Omega_{3\mathrm{D}}} c[\boldsymbol{k}]\big(\mathscr{P}_{\boldsymbol{\theta}}\{\varphi\} * h\big)(\boldsymbol{j} - \mathbf{M}_{\boldsymbol{\theta}^{\perp}}\boldsymbol{k}) + \mathbf{n}_{\boldsymbol{j}}^{\boldsymbol{\theta}}. \tag{2.6}$$

This gives the entries of the system matrix $\mathbf{H}^{\boldsymbol{\theta}}$ for a single orientation $\boldsymbol{\theta}$ as

$$[\mathbf{H}^{\boldsymbol{\theta}}]_{\boldsymbol{j},\boldsymbol{k}} = \big(\mathscr{P}_{\boldsymbol{\theta}}\{\varphi\} * h\big)(\boldsymbol{j} - \mathbf{M}_{\boldsymbol{\theta}^{\perp}}\boldsymbol{k}). \tag{2.7}$$

In practice, the tomographic setup produces a series of P projections, each with orientation $\boldsymbol{\theta}_p$, where $p = 1,\dots,P$. For the sake of conciseness, we hereafter use the notations $\mathbf{b}^{\boldsymbol{\theta}_p} = \mathbf{b}^p$, $\mathbf{n}^{\boldsymbol{\theta}_p} = \mathbf{n}^p$, and $\mathbf{H}^{\boldsymbol{\theta}_p} = \mathbf{H}^p$. By concatenating these entities for P measurements, we obtain the discrete formulation of the complete cryo-EM forward model as

$$\mathbf{b} = \mathbf{H}\mathbf{c} + \mathbf{n}, \tag{2.8}$$

where

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}^1 \\ \mathbf{b}^2 \\ \vdots \\ \mathbf{b}^P \end{bmatrix} \in \big(\mathbb{R}^M\big)^P = \mathbb{R}^{MP}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{H}^1 \\ \mathbf{H}^2 \\ \vdots \\ \mathbf{H}^P \end{bmatrix} \in \mathbb{R}^{MP\times N}, \quad \text{and } \mathbf{n} = \begin{bmatrix} \mathbf{n}^1 \\ \mathbf{n}^2 \\ \vdots \\ \mathbf{n}^P \end{bmatrix} \in \mathbb{R}^{MP}. \tag{2.9}$$

Finally, we remark that the direct storage of the matrix $\mathbf{H}$ is not feasible due to its size: its number of entries is typically on the order of billions for cryo-EM. To circumvent this problem, we follow the approach presented in [102]. Taking advantage of the isotropy of KBWFs, we store the oversampled values of $(\mathscr{P}_{\boldsymbol{\theta}}\{\varphi\} * h)$ for a single orientation $\boldsymbol{\theta} = \boldsymbol{\theta}_p$ in a look-up table, from which we then derive all the entries of $\mathbf{H}$ whenever needed.

2.3 Model-based Reconstruction with Sparsity Constraint

2.3.1 Inverse Problems and Ill-Posedness

Equipped with the discrete linear cryo-EM forward model, the reconstruction task now consists in recovering $\mathbf{c}$ from $\mathbf{b}$ in (2.8). In other words, the goal is to "reverse" the acquisition process to reconstruct the 3D structure of interest, a procedure that falls in the scope of *inverse-problem* theory. This process can be more or less arduous depending on the imaging

conditions. Formally, an inverse problem is said to be well-posed if a solution exists, is unique, and continuously depends on the measurements [103, 104]. By contrast, *ill-posed* inverse problems refer to problems that do not satisfy one (or more) of these conditions.

Ill-posed problems often arise in practice. For one, the system matrix $\mathbf{H}$ is often non-invertible as the object size tends to exceed the measurements size. Even when $\mathbf{H}$ is invertible, it is usually ill-conditioned, *i.e.*, the ratio of its maximal and minimal singular values is large. Hence, even small perturbations on the measurements $\mathbf{b}$ massively impact the recovered solution, which makes direct inversion futile in presence of noise. Most inverse problems in biological imaging—and certainly cryo-EM—are ill-posed to some degree, meaning that one cannot rely on the measurements alone to uniquely and/or stably recover the desired object.

2.3.2 Sparsity-Based Variational Formulation

A standard solution to handle ill-posed imaging conditions is to rely on *regularized variational* methods that inject prior information into the reconstruction procedure. When the noise distribution is assumed to be Gaussian, the data-fidelity term that best measures the fidelity of the reconstruction to the acquired data is quadratic [105]. Then, variational methods take the form

$$\widehat{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \lambda \mathscr{R}(\mathbf{L}\mathbf{c}), \tag{2.10}$$

where $\widehat{\mathbf{c}} \in \mathbb{R}^N$ is the recovered volume estimate, $\mathscr{R} : \mathbb{R}^Q \to \mathbb{R}$ is a regularization functional, $\mathbf{L} \in \mathbb{R}^{Q\times N}$ is a regularization operator, and $\lambda > 0$ is the regularization parameter that sets the strength of the regularization.

Different constraints based on prior knowledge can be applied to the reconstructed 3D object through the term $\mathscr{R}(\mathbf{L}\mathbf{c})$. A popular approach is to exploit the sparsity of natural objects, *i.e.*, the fact that they can be represented with a small number of nonzero (or near-zero) coefficients in a suitable basis specified by the operator $\mathbf{L}$ [106]. The inclusion of a sparsity-promoting constraint in (2.10) is classically achieved by using the convex ℓ_1 norm as a regularization functional, *i.e.*, $\mathscr{R} = \|\cdot\|_1$. In this case, the cost function in (2.10) is non-differentiable, which notably precludes the use of gradient-based minimization methods [107].

In this thesis, we often use the popular edge-preserving total-variation (TV) regularization as our sparsity-based prior [12]. TV regularization is obtained by combining the discrete gradient operator $\boldsymbol{\nabla} \in \mathbb{R}^{3N\times N}$ with the mixed (ℓ_1/ℓ_2)-norm defined, for all $\mathbf{u} \in \mathbb{R}^{3N}$, as

$$\|\mathbf{u}\|_{2,1} = \sum_{n=1}^{N} \|u_{n,.}\|_2 \tag{2.11}$$

with $u_{n,.} \in \mathbb{R}^3$. Hence, we set $\mathscr{R}(\mathbf{L}\mathbf{c}) = \|\mathbf{c}\|_{\mathrm{TV}} = \|\boldsymbol{\nabla}\mathbf{c}\|_{2,1}$ in (2.10).

2.3.3 ADMM-based Minimization Algorithm

Many efficient iterative algorithms have emerged over the years to solve non-differentiable, convex optimization problems [108, 109, 110, 111, 112]. A powerful splitting-based algorithm with prime convergence speed that solves (2.10) is the alternating direction method of multipliers (ADMM) [113, 114]. The philosophy behind ADMM is to decompose a relatively intricate optimization problem into a set of simpler sub-problems that are solved recursively until convergence.

We start by introducing an auxiliary variable $\mathbf{u} = \mathbf{L}\mathbf{c}$, $\mathbf{u} \in \mathbb{R}^Q$, to rewrite (2.10) as a constrained optimization problem

$$\widehat{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \lambda\mathscr{R}(\mathbf{u}) \quad s.t. \quad \mathbf{u} = \mathbf{L}\mathbf{c}. \tag{2.12}$$

Its augmented Lagrangian function is given by

$$\mathscr{L}(\mathbf{c},\mathbf{u},\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{b}-\mathbf{H}\mathbf{c}\|_2^2 + \lambda\mathscr{R}(\mathbf{u}) + \boldsymbol{\alpha}^T(\mathbf{L}\mathbf{c}-\mathbf{u}) + \frac{\rho}{2}\|\mathbf{L}\mathbf{c}-\mathbf{u}\|_2^2, \tag{2.13}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^Q$ is the vector of Lagrange multipliers, and $\rho > 0$ is a penalty parameter that influences the convergence speed.

ADMM theory [113] then asserts that (2.12) can be decomposed into simpler problems by alternatively minimizing (2.13) with respect to $\mathbf{c}$ and $\mathbf{u}$, and updating $\boldsymbol{\alpha}$ with a gradient ascent on (2.13), such that

$$\begin{cases} \mathbf{c}^{(k+1)} & = \underset{\mathbf{c}}{\arg\min}\ \mathscr{L}\big(\mathbf{c},\mathbf{u}^{(k)},\boldsymbol{\alpha}^{(k)}\big) & (2.14a)\\ \mathbf{u}^{(k+1)} & = \underset{\mathbf{u}}{\arg\min}\ \mathscr{L}\big(\mathbf{c}^{(k+1)},\mathbf{u},\boldsymbol{\alpha}^{(k)}\big) & (2.14b)\\ \boldsymbol{\alpha}^{(k+1)} & = \boldsymbol{\alpha}^{(k)} + \rho(\mathbf{L}\mathbf{c}^{(k+1)} - \mathbf{u}^{(k+1)}). & (2.14c) \end{cases}$$

Hence, the ADMM sub-solvers of (2.12) are given by

$$\begin{cases} \mathbf{c}^{(k+1)} & = \underset{\mathbf{c}}{\arg\min}\ \frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \frac{\rho}{2}\|\mathbf{L}\mathbf{c}-\mathbf{u}^{(k)} + \frac{\boldsymbol{\alpha}^{(k)}}{\rho}\|_2^2 & (2.15a)\\ \mathbf{u}^{(k+1)} & = \underset{\mathbf{u}}{\arg\min}\ \lambda\mathscr{R}(\mathbf{u}) + \frac{\rho}{2}\|\mathbf{L}\mathbf{c}^{(k+1)}-\mathbf{u} + \frac{\boldsymbol{\alpha}^{(k)}}{\rho}\|_2^2 & (2.15b)\\ \boldsymbol{\alpha}^{(k+1)} & = \boldsymbol{\alpha}^{(k)} + \rho(\mathbf{L}\mathbf{c}^{(k+1)} - \mathbf{u}^{(k+1)}). & (2.15c) \end{cases}$$

The pseudo code corresponding to this cyclic three-steps minimization scheme is given in Algorithm 1.

The first ADMM step (2.15a) is a quadratic minimization with respect to $\mathbf{c}$ that requires one to solve

$$\boldsymbol{\nabla}_{\mathbf{c}}\mathscr{L}\big(\mathbf{c}^{(k+1)},\mathbf{u}^{(k)},\boldsymbol{\alpha}^{(k)}\big) = \mathbf{0}, \tag{2.16}$$

which is equivalent to solving the linear system

$$\left(\rho\mathbf{L}^T\mathbf{L}+\mathbf{H}^T\mathbf{H}\right)\mathbf{c}^{(k+1)} = \mathbf{H}^T\mathbf{b}+\rho\mathbf{L}^T\left(\mathbf{u}^{(k)}+\frac{\boldsymbol{\alpha}^{(k)}}{\rho}\right) \tag{2.17}$$

in terms of $\mathbf{c}^{(k+1)}$. The difficulty in our case is that we cannot explicitly build the inverse of the matrix $\left(\rho\mathbf{L}^T\mathbf{L}+\mathbf{H}^T\mathbf{H}\right)$ due to the size of $\mathbf{H}^T\mathbf{H}$. We thus rely on an iterative algorithm—classically, the conjugate gradient (CG) method [107]—to solve (2.17). This inner CG algorithm includes iterates that necessitate the application of the normal matrix $\mathbf{H}^T\mathbf{H}$ to the current volume estimate, which can be very costly if this normal matrix is not carefully engineered and/or if multiple CG loops are needed. The matrix $\mathbf{L}^T\mathbf{L}$ often takes the form of a (diagonalizable) convolution operator, assuming that proper boundary conditions are used. Even when this is not the case, its application often has negligible computational cost compared to that of $\mathbf{H}^T\mathbf{H}$.

We recognize in the second ADMM step (2.15b) the proximal operator corresponding to the regularization functional $\mathscr{R}$. We recall that the proximal operator, $\text{prox}_\Phi : \mathbb{R}^Q \to \mathbb{R}^Q$, of a function, $\Phi : \mathbb{R}^Q \to \mathbb{R}$, is defined by

$$\text{prox}_\Phi(\mathbf{u};\mu) = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{z}-\mathbf{u}\|_2^2 + \mu\Phi(\mathbf{u}) \tag{2.18}$$

for $\mu > 0$. In many instances, the proximal operator admits a closed-form expression that can be efficiently computed [115]. In the case of TV regularization, where $Q = 3N$, the proximal operator associated to the $\|\cdot\|_{2,1}$ norm corresponds to a rapid pointwise thresholding operation [116]

$$\left[\text{prox}_{\text{TV}}(\mathbf{u};\mu)\right]_{n,\cdot} = \left(1-\frac{\mu}{\|u_{n,\cdot}\|_2}\right)_+ u_{n,\cdot} \tag{2.19}$$

with $\mathbf{u} \in \mathbb{R}^{3N}$, $u_{n,\cdot} \in \mathbb{R}^3$, and $(\cdot)_+ = \max(\cdot, 0)$. The step (2.15b) is often said to be a generic (*i.e.*, application-independent) statistical step, as it affects the reconstruction quality via the regularization constraints imposed on the solution. Finally, the third ADMM step (2.15c) is a simple update of the Lagrange multipliers that corresponds to a gradient ascent on (2.13).

A definite advantage of ADMM is that it reaches a solution with acceptable precision in relatively few iterations [117]. Different stopping criteria can be chosen for the ADMM and CG algorithms. In our works, we either fix the maximal number of outer ADMM loops and inner CG loops, or we stop the run when the difference between successive iterations drops below a certain threshold. The complete algorithmic scheme derived in this section is implemented in the GlobalBioIm library, an open source MATLAB library for solving inverse problems (see Remark 5). The library being particularly flexible, one can easily introduce additional constraints (*e.g.*, positivity constraint), implement different splitting schemes, or use alternative minimization algorithms.

Algorithm 1 ADMM with inner CG loops

Inputs: $\mathbf{c}^0 \in \mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^{MP}$, $\lambda > 0$, $\rho > 0$

$\mathbf{u}^0 = \mathbf{L}\mathbf{c}^0$, $\boldsymbol{\alpha}^0 = \mathbf{u}^0$
$k = 0$
while (ADMM stopping criteria not met) **do**
    while (CG stopping criteria not met) **do**
        $\mathbf{c}^{(k+1)} \leftarrow$ solve (2.17) using CG
    end while
    $\mathbf{u}^{(k+1)} = \text{prox}_{\mathscr{R}}\left(\mathbf{L}\mathbf{c}^{(k+1)} - \frac{\boldsymbol{\alpha}^{(k)}}{\rho}; \frac{\lambda}{\rho}\right)$
    $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + \rho\left(\mathbf{L}\mathbf{c}^{(k+1)} - \mathbf{u}^{(k+1)}\right)$
    $k = k + 1$
end while
return $\mathbf{c}^{(k+1)}$

Remark 5 (GlobalBioIm)**.** *The optimization methods that decouple the physical aspects of the problem from the imposition of prior constraints on the signal are of special interest in inverse problems. They permit a particularly modular implementation that unifies modality-specific reconstruction algorithms under a common reconstruction framework [118]. The user-friendly and open-source GlobalBioIm library [119, 120] capitalizes on this unified multiplicity of modalities to offer users a standardized resolution of a wide range of imaging problems.*

Although they are more robust than direct methods, most iterative reconstruction schemes come with a prohibitive computational cost if not carefully engineered. This is an important challenge in cryo-EM given the amount of data that needs to be processed. The dominant computational cost in our reconstruction algorithm is the linear step (2.17), which requires the repeated application of the normal matrix $\mathbf{H}^T\mathbf{H}$. Note that, although the discrete product $\mathbf{H}^T\mathbf{b}$ in (2.17) needs only be computed once during the whole optimization procedure, this can also be costly in its own right. The quest for computationally tractable formulations for $\mathbf{H}^T\mathbf{H}$ and $\mathbf{H}^T\mathbf{b}$ is discussed in Chapter 4.

3 Compressed Sensing (CS) for STET

3.1 Introduction

[1]This work is dedicated to STET imaging (Figure 3.1, left), which we introduced in Section 1.3.2. As explained then, the primary concern in STET is to reduce the electron dosage—which is damaging to the sample (Figure 3.1, right)—while maintaining an acceptable reconstruction quality. Our contributions to this conundrum is the design of a robust acquisition-reconstruction scheme for STET that permits high-quality reconstruction from limited STET measurements.

To set the context, we recall that popular downsampling strategies for low-dose STET imaging include the two following approaches.

- *Uniform tilt-downsampling (T-DS)*, where one acquires fewer equally-spaced, fully-scanned 2D projections over the widest possible angular range. This is equivalent to increasing the angular increment between two successive tilts.
- *Random image-downsampling (I-DS)*, where one scans only a fraction of the 3D sample at each orientation following a random downsampling pattern, resulting in incomplete 2D projections.

In [40], Saghi *et al.* combined the two downsampling techniques and demonstrated that further reduction of radiation damage could be achieved by doing so. Interestingly, they choose to perform the reconstruction in two distinct steps. They first filled the missing pixels in each undersampled 2D projection through TV-inpainting [121]. They then recovered the 3D volume from these restored projections with a TV-regularized tomographic reconstruction scheme.

The reconstruction task in such low-dose STET imaging conditions is highly ill-posed due

[1]The content of this chapter is based on [17]: **Laurène Donati**, Masih Nilchian, Sylvain Trépout, Cédric Messaoudi, Sergio Marco, Michael Unser. *Compressed Sensing for STEM Tomography.* Ultramicroscopy, 2017.

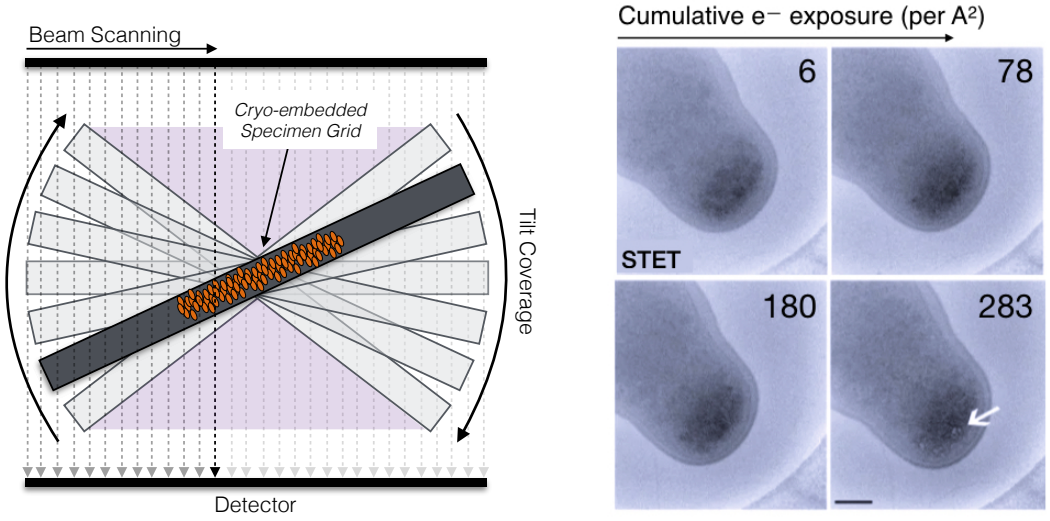


Figure 3.1 – **(Left)** In a typical STET setup, the lenses focus the electrons in a single beam and the tilted sample is scan-rastered to form a projection image. The purple area indicates the angular region left uncovered due to technical constraints, which results in a missing cone of information in the Fourier domain. **(Right)** A primary concern in STET is to reduce the electron damage that results from an increased electron exposure (image adapted from [122]).

to the multiple sources of missing information. Those include a limited number of projections, unscanned pixels in the projections, a missing wedge of information in the Fourier domain, and strong noise degradation of the projections. This adverse situation advocates for the development of more robust reconstruction methods than the one proposed in [40].

A particularly powerful paradigm to robustly handle reconstructions from very few measurements is compressed sensing (CS) [123]. CS is extremely relevant to biomedical imaging and has been applied with great success to multiple imaging modalities [124, 125, 126, 127]. In short, the CS theory assesses that sparse signals can be acquired with a reduced number of measurements provided that proper acquisition and recovery methods are used.

In this work, we present a tailored acquisition-reconstruction STET framework that relies on the principles of CS and permits to minimize the electron dosage required for high-quality STET imaging. More precisely, we propose to scan only a small random subset of pixels at every sample orientation, and present a regularized reconstruction scheme to recover the 3D sample from these strongly undersampled data. We demonstrate on simulated and real data that reconstruction can then be achieved with very few measurements, hence strongly reducing the effect of sample degradation and preserving image quality.

The chapter is organized as follows. In Section 3.2, we recall the principles of CS theory and three of its key ingredients: data sparsity, incoherent sensing, and ℓ_1-regularized signal recovery. We then demonstrate the applicability of each principle to STET in the subsequent

sections. In Section 3.3, we discuss the sparsity of biological STET samples. In Section 3.4, we demonstrate the incoherence property of the random I-DS scheme associated with the wavelet representation basis. In Section 3.5, we describe the nonlinear reconstruction scheme that permits the robust recovery of volumes from measurements acquired in low-dose imaging conditions. We present our experimental results in Section 3.6 and our conclusions in Section 3.7.

3.2 CS Theory

As a preliminary, we briefly describe here the three theoretical components of CS theory that are of central importance to our work.

3.2.1 Data Sparsity

The theory of CS relies on the notion of sparsity [123]. A signal is said to be *sparse* if it has a concise representation in some basis. The mathematical formulation is as follows. Let the expansion of a signal $\mathbf{c} \in \mathbb{R}^N$ in an orthonormal representation basis $\mathbf{\Psi} = [\boldsymbol{\psi}_1 \boldsymbol{\psi}_2 \cdots \boldsymbol{\psi}_N] \in \mathbb{R}^{N \times N}$ be given by

$$\mathbf{c} = \sum_{n=1}^{N} a_n \boldsymbol{\psi}_n, \tag{3.1}$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_N) \in \mathbb{R}^N$ is the sequence of expansion coefficients of $\mathbf{c}$, with $a_n = \langle \mathbf{c}, \boldsymbol{\psi}_n \rangle$. The implication of sparsity is that, in a suitable sparsifying basis $\mathbf{\Psi}$, the signal $\mathbf{c}$ is represented by very few nonzero coefficients a_n (in comparison to N). Those nonzero coefficients thus concentrate most of the signal information. In practice, most objects of interest are not exactly sparse, but rather approximately sparse (*i.e.*, compressible). For these signals, most of the coefficients in the sparsifying basis $\mathbf{\Psi}$ have near-zero values instead of strictly null ones. We say that the K-sparse approximation of the signal is obtained when one keeps only the $K \ll N$ larger coefficients and discards the rest.

3.2.2 Incoherent Sensing

It is well known in signal processing that sampling a signal below the Nyqvist frequency introduces aliasing artifacts that manifest themselves as periodizations in the Fourier domain [98]. A key idea in CS is to rely on a sensing matrix $\mathbf{A}$ (*i.e.*, an acquisition process[2]) that introduces *incoherent* artifacts in the sparsifying representation basis $\mathbf{\Psi}$. These artifacts are said to be incoherent because they spread uniformly throughout the basis $\mathbf{\Psi}$ in a noise-like manner. Hence, the few non-zero entries of $\mathbf{a}$ stand out from the introduced incoherent artifacts, and can then be recovered through sparsity-promoting reconstruction [123].

[2] We use here the notation $\mathbf{A}$ to describe a more generic acquisition process than the cryo-EM one introduced in Chapter 2 and modeled by $\mathbf{H}$.

The smallest sampling frequency that can accurately capture the signal information is directly determined by the incoherence of the sensing matrix $\mathbf{A}$ with respect to the sparsifying representation basis $\mathbf{\Psi}$ [128]. The coherence of a matrix relates to the amount of cross-correlation between its column vectors [129]. In CS theory, the incoherence condition is fulfilled when $\mathbf{A}$ respects the restricted isometry property [130]. Simply said, the larger the incoherence of the sensing-representation bases pair $\mathbf{A}/\mathbf{\Psi}$, the fewer the number of samples needed for proper signal reconstruction.

Herein lies the key practical challenge in CS: to design a physical acquisition procedure that sufficiently fulfills the incoherence condition when paired with an appropriate representation basis. In that respect, an important consideration for the present work is that the column vectors of random matrices are largely incoherent with any fixed basis. Hence, random sensing matrices can very efficiently capture information on sparse signals with a minimal number of measurements, which we shall exploit for the design of an incoherent STET acquisition scheme (Section 3.4).

3.2.3 ℓ_1-Regularized Signal Recovery

Let now consider the task of reconstructing the signal $\mathbf{c} \in \mathbb{R}^N$ from its measurements $\mathbf{b} \in \mathbb{R}^{MP}$ collected through the sensing matrix $\mathbf{A} \in \mathbb{R}^{MP \times N}$, *i.e.* $\mathbf{Ac} = \mathbf{b}$. Assuming that the conditions of sparsity and incoherence are met, CS theory assesses that it is possible to correctly recover the original signal from its undersampled measurements. More precisely, the transform coefficients $\mathbf{a}$ of $\mathbf{c}$ in (3.1) are recovered by solving the optimization problem [131]

$$\widehat{\mathbf{a}} = \underset{\mathbf{a} \in \mathbb{R}^L}{\arg\min} \|\mathbf{A\Psi a} - \mathbf{b}\|_2^2 \quad s.t. \quad \|\mathbf{a}\|_0 < K, \tag{3.2}$$

with $\widehat{\mathbf{a}} \in \mathbb{R}^N$ and $K > 0$. The ℓ_0 seminorm $\|\mathbf{a}\|_0$ counts the number of nonzero elements of $\mathbf{a}$, while the quadratic term enforces data consistency. Hence, the minimization (3.2) yields the K-sparse solution that is the most compatible with the measured data [123, 132]. In addition, the incoherence condition ensures that (3.2) can be relaxed into the more computationally-tractable problem [131]

$$\widehat{\mathbf{a}} = \underset{\mathbf{a} \in \mathbb{R}^N}{\arg\min} \|\mathbf{A\Psi a} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{a}\|_1, \tag{3.3}$$

where the ℓ_1-norm promotes the sparsity of $\mathbf{c}$ in $\mathbf{\Psi}$ with a strength set by $\lambda > 0$. The estimated signal $\widehat{\mathbf{c}}$ can then be re-expanded through $\widehat{\mathbf{c}} = \sum_{n=1}^{N} \widehat{a}_n \boldsymbol{\psi}_n$.

3.3 Sparsity of STET Samples

We start our application of the CS principles to STET by exposing the sparsity of the considered bio-samples in the discrete Haar wavelet transform (Haar DWT) domain. In preamble, we recall that it has long been established that most natural images or volumes are (approxi-

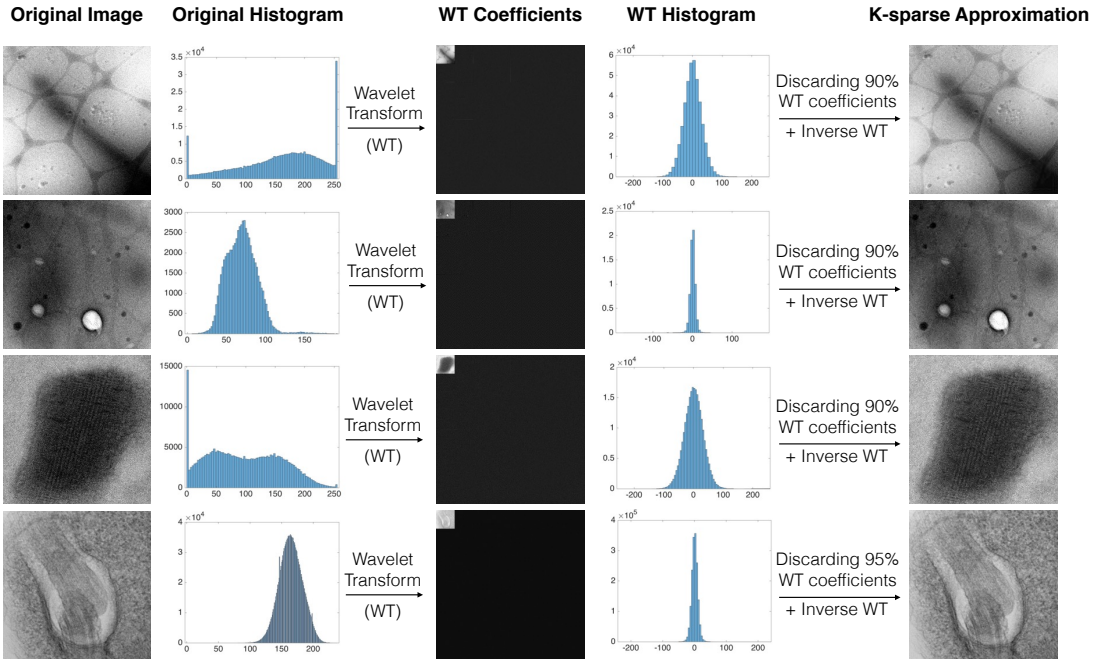


Figure 3.2 – Illustration of the sparsity of biological 2D STET images in the Haar DWT. We consider typical STET samples (here in 2D) and compute their Haar DWT coefficients using the ImageJ software [141]. We then discard most of these coefficients (90%-95%) and compute the inverse transform to get the K-sparse approximation of the original images. In the image domain, a large range of nonzero coefficients is observed for all samples (left histograms). In contrast, most coefficients in the sparsifying domain have near-zero values (right histograms). As a result, discarding of 90%-95% of the transform coefficients does not lead to significant perceptual losses in the inverse-transformed images (far right).

mately) sparse in appropriate domains, such as the discrete cosine transform (DCT) and the DWT [133]. Similarly, the sparsity of various signals of relevance in biomedical imaging has been demonstrated by multiple works [124, 134, 135, 125]. In [136], Anderson *et al.* assessed the sparsity of various EM samples in the block-DCT domain. For cryo-ET, Song *et al.* showed that the projections exhibit sparsity in the DCT domain [137]. The suitability of the DWT for sparsely representing ET data has also been demonstrated in a number of empirical studies [138, 139, 140]. Even though the aforementioned works certainly provide compelling arguments for the sparsity of STET samples, we confirm this property in Figure 3.2 for the sake of completeness.

3.4 Incoherence Analysis for Random-Beam STET (RB-STET)

We now restrict our interest to the low-dose STET regime achieved through random I-DS, motivated by the observation that random matrices are largely incoherent with any fixed basis. For the sake of conciseness, we hereafter refer to this acquisition setup as random-beam

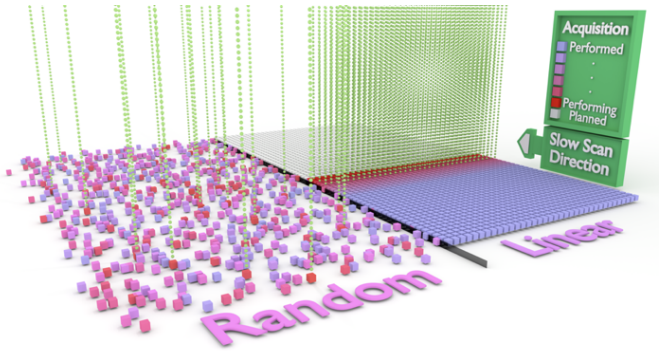


Figure 3.3 – **(Left)** The RB-STET acquisition scheme. **(Right)** A fully-sampled linear-scanning scheme. Image from Sylvain Trépout.

STET (RB-STET), for which an illustration is provided in Figure 3.3 (left). In this section, we first derive the discrete forward model for RB-STET. We then experimentally demonstrate the incoherence of RB-STET sensing associated to the DWT representation basis $\mathbf{\Psi}$. As discussed, this amounts to verifying that the artifacts introduced by the RB-STET acquisition scheme behave in a "noise-like" manner in $\mathbf{\Psi}$.

3.4.1 Discrete Forward Model of RB-STET

We now derive the discrete forward model that relates the 3D object of interest $\mathbf{c}$ to its RB-STET measurements $\mathbf{b}$. This acquisition process corresponds to the straight-line transmission of the electron beam through the sample at positions dictated by the random I-DS scanning pattern. Hence, we can largely rely on the building blocks for cryo-EM modelling introduced in Section 2.2. In particular, we use the system matrix $\mathbf{H}^p$ defined in (2.7) to model the transmission process for a sample at orientation $\boldsymbol{\theta}_p \in \Omega_{\boldsymbol{\theta}}$. The random scanning for this projection $\mathbf{b}^p$ is then handled by introducing the diagonal matrix $\mathbf{S}^p \in \mathbb{R}^{M\times M}$, whose M-sized diagonal has entries equal to ones at the locations on the projection plane scanned by the electron beam, and zeros elsewhere. We assume here that this random-scanning regime follows a uniform distribution.

By concatenating all the relevant entities and assuming an additive Gaussian noise model, we obtain the discrete formulation of the RB-STET forward model for P projections as

$$\mathbf{b} = \mathbf{A}\mathbf{c} + \mathbf{n} = \mathbf{S}\mathbf{H}\mathbf{c} + \mathbf{n}, \tag{3.4}$$

with $\mathbf{c} \in \mathbb{R}^N$,

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}^1 \\ \mathbf{b}^2 \\ \vdots \\ \mathbf{b}^P \end{bmatrix} \in \mathbb{R}^{MP},\ \mathbf{S} = \begin{bmatrix} \mathbf{S}^1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{S}^P \end{bmatrix} \in \mathbb{R}^{MP\times MP},\ \mathbf{H} = \begin{bmatrix} \mathbf{H}^1 \\ \mathbf{H}^2 \\ \vdots \\ \mathbf{H}^P \end{bmatrix} \in \mathbb{R}^{MP\times N},\ \text{and}\ \mathbf{n} = \begin{bmatrix} \mathbf{n}^1 \\ \mathbf{n}^2 \\ \vdots \\ \mathbf{n}^P \end{bmatrix} \in \mathbb{R}^{MP}. \tag{3.5}$$

The treatment of optical effects deserves some discussion here. An inherent problematic of optical imaging systems in the context of CS is that their PSF can negatively affect the incoherence as such sensing systems due to the correlative nature of the convolution. In our case, we make the assumption that this deviation from the ideal CS world is manageable and does not preclude the use of a CS-scheme based in STET. Hence, we do not include the effect of PSF in our RB-STET forward model and set $h = \delta$ in (2.2), with δ the Dirac delta distribution. We investigate the soundness of this assumption in Section 3.6.2 by running experiments with real STET projections.

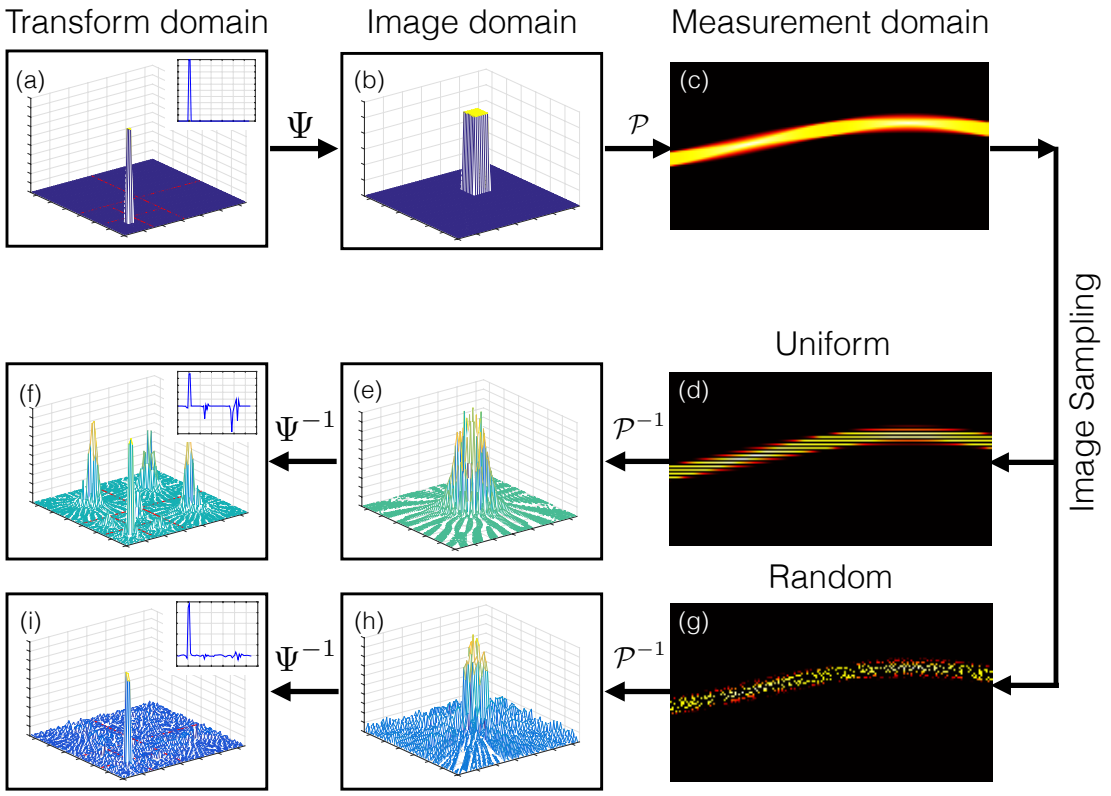


Figure 3.4 – Comparison of the reconstruction of a sparse 2D signal (top row) from two different downsampling approaches: uniform sinogram sampling (middle row) and random sinogram sampling, which mimics RB-STET (bottom row). The projections (*c*) of the spatial signal (*b*) harbor an "impulse-like" sparse expansion (*a*) in the wavelet domain $\mathbf{\Psi}$. Uniform undersampling of the projections (*d*) is associated with a coherently aliased TPSF (*f*). In contrast, random undersampling of the projections (*g*) results in an incoherently aliased TPSF (*i*). As a corollary, FBP reconstruction (indicated by $\mathscr{P}^{-1}$) of the randomly sampled 1D projections (*h*) is more accurate (*b*) than the signal reconstructed from uniform sampling (*e*). The small upper-right boxes in (*a*), (*f*), and (*i*) display profile views of the respective sparse representations.

3.4.2 Experimental Incoherence-Analysis

We now verify the incoherence of the RB-STET sensing matrix $\mathbf{A} = \mathbf{SH} \in \mathbb{R}^{MP \times N}$ in (3.5) with respect to the Haar DWT representation basis $\mathbf{\Psi}$ by using the transform point-spread function (TPSF) analysis proposed by Lustig *et al.* [124]. The TPSF generalizes the notion of PSF in that it assesses "how a single transform coefficient of the underlying object ends up influencing other transformed coefficients of the measured undersampled object".

The results of our incoherence analysis (performed in 2D) are displayed in Figure 3.4. The evaluation is performed on a (64×64) synthetic image representing a single square surrounded by zero values (Figure 3.4b). For the sake of simplicity, we use the FBP[3] algorithm to reconstruct the signal from 1800 noise-free projections equally-spaced in $[0; 2\pi)$ (Figure 3.4c). A uniform I-DS scheme (Figure 3.4d-f) is also considered to permit comparison with the RB-STET regime (Figure 3.4g-i).

The results empirically confirm the incoherence between the RB-STET sensing and the Haar DWT. Indeed, the introduced artifacts have a strongly incoherent behaviour in $\mathbf{\Psi}$ (Figure 3.4i). In contrast, a uniform I-DS scheme leads to a much less suitable outcome (Figure 3.4f), with the appearance of severe aliasing artifacts. These results suggest that the signal of interest can indeed be recovered from undersampled RB-STET measurements assuming that a proper non-linear recovery scheme is used.

3.5 Reconstruction Scheme for RB-STET

Our focus so far has been on the Haar DWT as the sparsifying basis for our RB-STET measurements. In this framework, the recovery scheme (3.3) actually corresponds to a wavelet-shrinkage problem [142], which can be solved through the cycle-spinning technique [143, 144, 145]. An alternative is to link the wavelet-shrinkage algorithm to the closely-related TV-regularised scheme [146, 147], which is the approach we take in this work. Hence, our RB-STET reconstruction scheme takes the variational form

$$\widehat{\mathbf{c}} = \underset{\mathbf{c} \in \mathbb{R}^N}{\arg\min} \, \|\mathbf{SHc} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{\nabla c}\|_1, \tag{3.6}$$

where $\mathbf{\nabla} : \mathbb{R}^{3N \times N}$ is the discretized gradient operator. We solve (3.6) using the ADMM-based minimization algorithm presented in Section 2.3.2. The regularization functional $\|\cdot\|_1 : \mathbb{R}^{3N} \to \mathbb{R}$ corresponds here to the ℓ_1 norm, which also admits a fast pointwise proximal operator [148].

[3] As mentioned in Section 2.1, FPB demonstrates satisfactory performance when the measurements are sufficiently numerous, noise-free, and uniformly covering the Fourier domain.

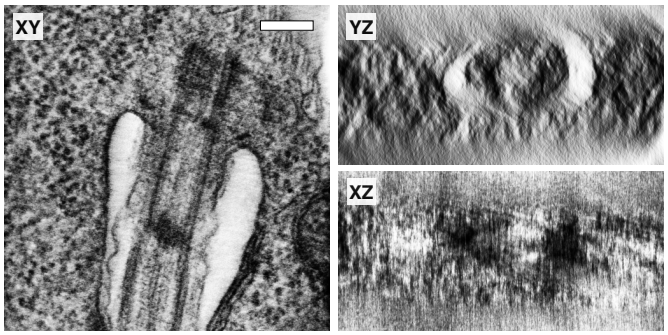


Figure 3.5 – Orthoslices of our (512 × 512 × 256) ground-truth depicting the flagellar pocket of a trypanosome [149]. The scale bar indicates 250nm.

3.6 Experiments

We assess the performance of our RB-STET acquisition-reconstruction scheme with experiments on synthetic and real STET data. In the synthetic case, we use as ground-truth a 3D reconstruction obtained from a set of real fully-scanned STET projections. We then simulate RB-STET projections from this ground-truth using our forward model (3.4). In the real-case, we directly work with the aforementioned real STET projections and randomly select a subset of pixels to mimic real RB-STET projections. This allows us to assess the efficacy of our framework in more realistic, challenging conditions that notably include optical effects.

3.6.1 Synthetic-Data Experiment

Simulation Conditions

Our ground-truth volume is a 3D visualization of size (512 × 512 × 256) of the flagellar pocket of a trypanosome (Figure 3.5) [149]. We simulate the RB-STET acquisition process through (3.4) and (2.7), using a KBWF with parameters $a = 2$, $\alpha = 10.8$, $m = 2$ [100]. To mimic the missing wedge of information, we restrict ourselves to an angular coverage of $(-70°; +70°)$, with a 1° increment (*i.e.*, 140 projections). For comparison purposes, two downsampling strategies are simulated in the experiments: uniform T-DS and random I-DS. Uniform T-DS condition is achieved by uniformly increasing the angular distance between two acquired projections. Random I-DS is achieved by applying a uniformly randomized subsampled binary mask over the 140 initial projections.

Unless otherwise indicated, the reconstruction is performed by solving (3.6) using the ADMM-based algorithm presented in Section 2.3.2. For the reconstruction, we use a different KBWF ($a = 4$, $\alpha = 19$, $m = 2$) as basis function to reduce the risk of inverse crime[4]. Finally, the

[4] An inverse crime [150, 151] occurs when one uses the same theoretical ingredients to simulate the measurements and reconstruct the sample. In this thesis, we reduce this risk by always using KBWFs with different sets of parameters for generating the 2D projections and for reconstructing the 3D object. The sets of KBWF parameters we consider both ensure minimal deviation of KBWFs from the partition of unity condition [100].

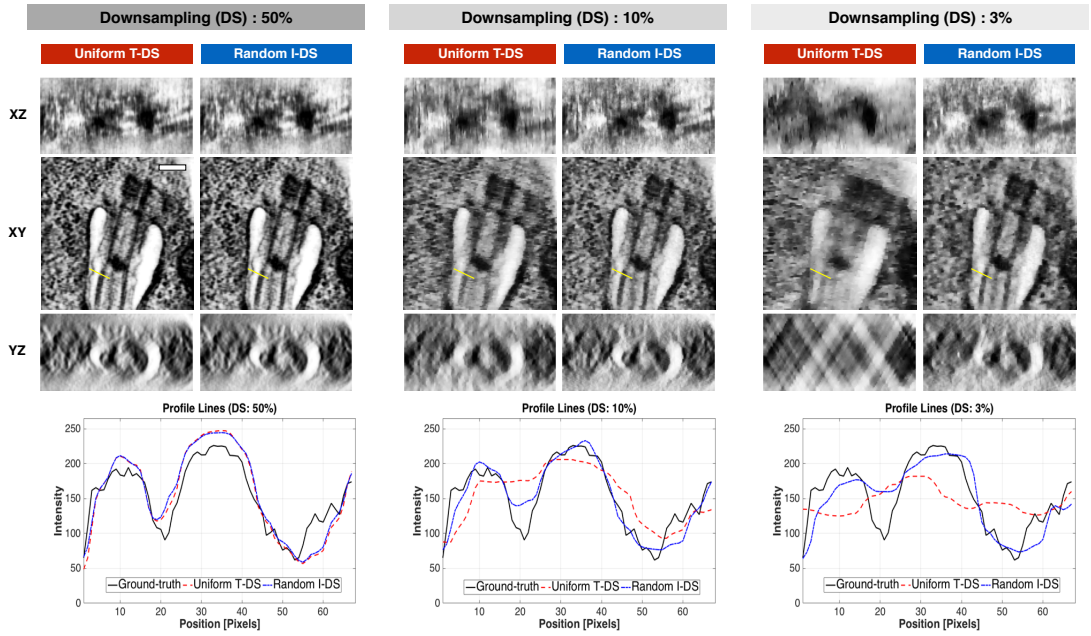


Figure 3.6 – (Top) Cross-sectional slices through the flagellar pocket reconstructed from uniformly T-DS synthetic projections and from random I-DS synthetic projections at various downsampling levels (50%, 10%, 3%). (Bottom) Profile lines taken on the XY-orthoslices of the reconstructed volumes. The position of the profile line is indicated in yellow on the orthoslices. The scale bar indicates 250nm.

optimization of the regularization parameter λ in (3.6) is done by visual assessment.

Uniform T-DS vs Random I-DS

We first demonstrate the superiority of random I-DS strategies over uniform T-DS techniques at various levels of electron-dosage reduction. We simulate both downsampling conditions and compare the reconstructions from the two synthetic datasets.

Figure 3.6 shows the orthoslices of the reconstructions when only 50%, 10%, and 3% of the initial electron dosage is used. When the dose is reduced by half (left column), high-quality reconstructions can still be obtained with both frameworks, as confirmed by the clear overlapping of their corresponding profile lines. However, when the ratio of scanned areas falls below 10% (middle column), the reconstruction from the uniform T-DS measurements starts showing strong degradation whereas the reconstruction from random I-DS acquisitions remains of satisfactory quality. The profile lines on the 3% reconstructions (righ column) further illustrate the superior robustness of the random I-DS strategy in low-dose imaging conditions: Whereas the T-DS approach fails to retrieve any important change in intensity, random I-DS still permits the visual delimitation of the main trypanosome structures.

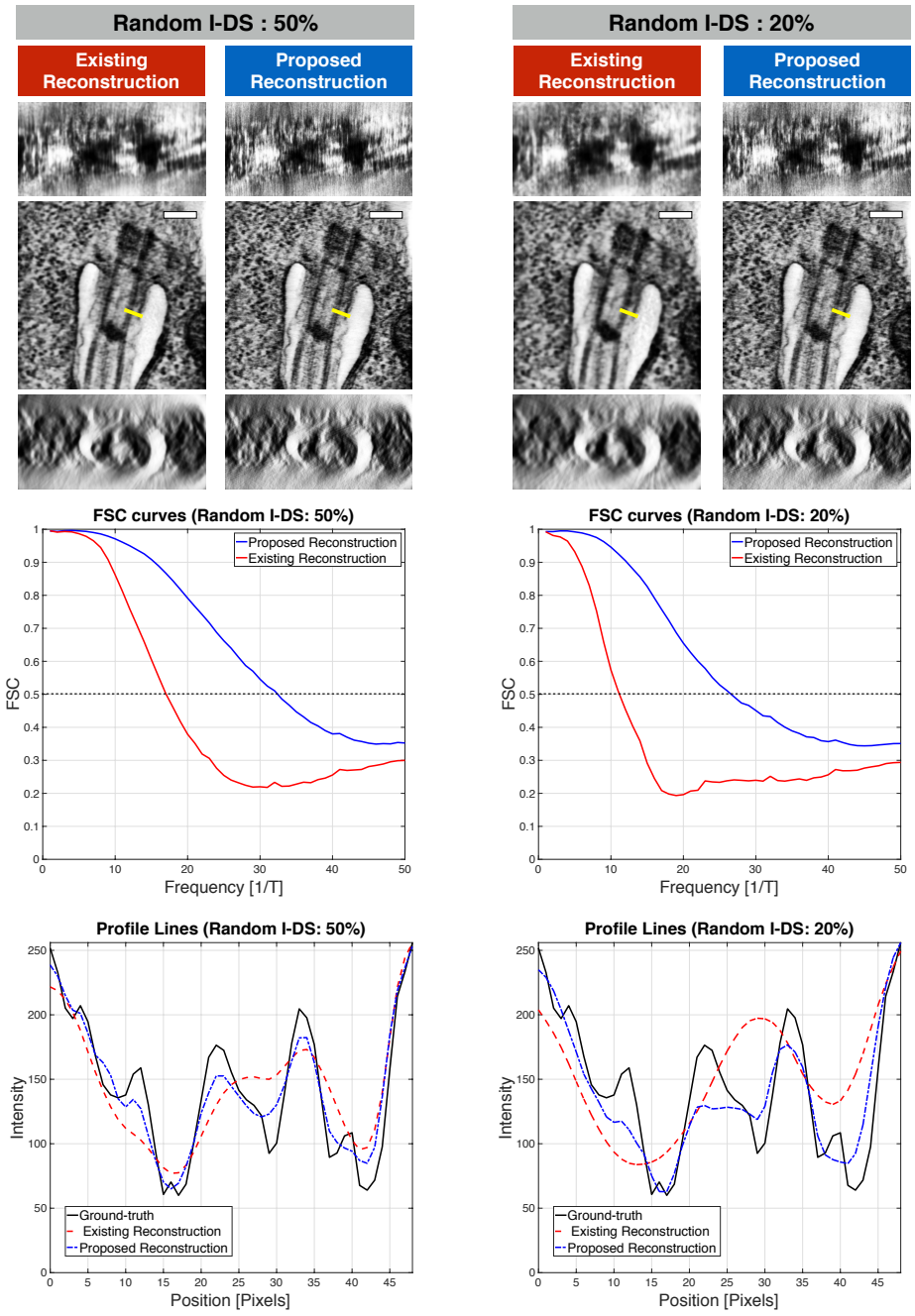


Figure 3.7 – Comparison of the proposed RB-STET algorithm with the existing one [40] for the reconstruction of synthetic random I-DS projections of *T. brucei* [149]. (Top) Cross-sectional slices through the reconstructed flagellar pocket for both algorithms at 50% and 20% downsampling. (Middle) FSC curves between the ground-truth and the reconstruction for both algorithms at 50% and 20% downsampling. (Bottom) Profile lines taken on the XY-orthoslices of the reconstructions. The position of the profile line is indicated in yellow on the orthoslices. The scale bar indicates 250nm.

Reconstruction Algorithms for Random I-DS Data: Existing [40] vs Proposed (RB-STET)

We now compare our full sensing-reconstruction RB-STET framework to the pioneering reconstruction approach[5] proposed by Saghi *et al.* [40] for the reconstruction of random I-DS measurements. As previously mentioned, the authors performed the reconstruction in two distinct steps: They filled in the missing unscanned pixels in the projections through TV-inpainting, and then applied an iterative tomographic algorithm with TV-regularization on the restored projections. We have reimplemented their algorithm as described in [40]. For both reconstruction schemes, the optimization of the hyper-parameters is performed by visual assessment.

Figure 3.7 presents orthoslices of the reconstructions of random I-DS data achieved by both frameworks at 50% and 20% downsampling ratios (top). The corresponding Fourier shell correlation (FSC[6]) are also displayed (middle), as well as the profile lines taken on the XY-orthoslices of the reconstructed volumes (bottom). Visual and quantitative analysis of these results indicate that, at equivalent dose reduction, the proposed RB-STET reconstruction algorithm outperforms the algorithm from [40]. In particular, our algorithm can retrieve finer details (*e.g.*, filament-like structures) and achieve higher resolution at both downsampling levels, as indicated by the FSC curves and the profile lines.

Several reasons can be put forward to explain these improvements. First, the proposed RB-STET algorithm performs the tomographic reconstruction in a single global fashion, as prescribed by the theory of CS. Significant advantages follow: The combination of more data gives additional information about the object of interest, and the influence of sparsity increases with the dimensionality of the reconstruction procedure. Furthermore, as explained by the authors in their discussion [40], their reliance on a delicate intermediate TV-inpainting step limits their capacity to reconstruct fines structures when only few pixels are scanned, as TV-inpainting tends to introduce strong staircase artifacts on the heavily-downsampled projections. Finally, our approach simplifies the optimization procedure in that it only requires the tuning of a unique hyper-parameter (the regularization parameter λ).

3.6.2 Real-Data Experiment

To get some insight on the performance of the RB-STET scheme in real conditions, we generate a RB-STET dataset from real STET projection measurements. To do this, we apply a uniformly-randomised subsampling binary mask to the real projection dataset. The difference with the previous synthetic experiment is important as the use of real STET projections 1) involves optical effects that may jeopardize the incoherence of our sensing-representation basis, and 2) increases the ill-posedness of the reconstruction problem.

[5] This was the state-of-the-art approach at the time of publication.

[6] The FSC is a popular metric in cryo-EM [152]. It provides a measure of resolution by comparing the Fourier transforms of two volumes (*e.g.*, the ground-truth and the reconstruction) at different frequencies. The spatial frequency at which the FSC curve falls below a certain criterion—commonly fixed at FSC=0.5 or FSC=0.143 in the cryo-EM community—indicates the achieved resolution. For more details, see Appendix A.3.

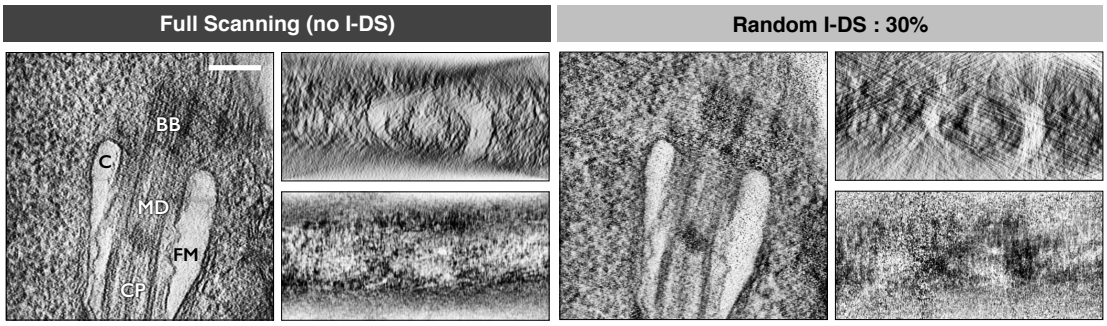


Figure 3.8 – Cross-sectional slices through the reconstructed flagellar pocket of *T. brucei* from (top) real fully-scanned STET projections and from (bottom) real STET projections with 30% random I-DS. The indicated biological structures are the basal body (BB), the collarette (C), the central pair (CP), the flagellar membrane (FM) and the microtubules doublets (MD). The scale bar indicates 250nm.

Sample Preparation and Data Acquisition

T. brucei cells are cultured in SDM79 medium supplemented with haemin and 10% foetal calf serum [153], and fixed directly in the culture flask with 2.5% glutaraldehyde and 4% PFA for 30 min at room temperature. The cells are then rinsed three times in PBS and subsequently post-fixed in 1% OsO_4 for 30 min at room temperature. They are then dehydrated in baths of increasing ethanol concentrations at 4° C and included in Epon resin. Sections of 500 nm are prepared using an ultramicrotome and are mounted on electron microscopy copper grids for observation.

The tomographic tilt-series are acquired using the bright-field STET detector (camera length: 60 cm; magnification: 150,000×; probe size: 1.5 nm; convergence semi-angle of the beam: 25 mrad; collection semi-angle of the detector: 6.667 mrad) on a JEOL 2200FS field emission gun 200 kV electron microscope (JEOL© Ltd.). A total of 100 projections are acquired following a Saxton scheme [154] from -70° up to +70° with tilt increments varying between 1° (at the highest tilt angles) and 2° (at the lowest tilt angles around 0°). Images are recorded using the Recorder software (JEOL© Ltd.). A total of five images with different foci are collected for each tilt angle and merged as described in [149], enabling the recovery of information at focus through the whole sample depth.

Reconstruction of Randomly-Downsampled Real STET Projections

The reconstruction is performed by solving (3.6) using the ADMM-based algorithm presented in Section 2.3.2. Figure 3.8 compares the $(512 \times 512 \times 256)$ reconstruction from the unsampled real STET dataset to the $(512 \times 512 \times 256)$ reconstruction obtained when a 30% random I-DS scheme is applied to these projections.

The results show that the RB-STET reconstruction scheme can recover the main Try-

panosome structures even when only a third of the initial information is available in the measurements. In particular, the structures annotated in Figure 3.8 can still be located and distinguished from one another in the reconstruction (Figure 3.8-bottom). The contours of the collarette and the flagellar membrane are also correctly recovered. The analysis of the central structures of the flagellar pocket (*i.e.*, the basal body, the microtubules doublets and the central pair) underlines the difficulty of recovering the higher-frequency details. Nevertheless, important visual information on the shape and texture of those structures can be retrieved from the reconstructed volume. Finally, these positive experimental results tend to confirm that the presence of optical effect does not severely impact the incoherence of the RB-STET acquisition scheme.

3.7 Discussion

We presented a regularized tomographic reconstruction framework to recover high-quality volumes from randomly-downsampled STET projections (*i.e.*, RB-STET data). This acquisition-reconstruction framework was built upon the demonstration that, in contrast to uniform downsampling methods, RB-STET fulfills the incoherence condition required by the CS theory when associated to the Haar DWT representation basis. Its predicted superiority over T-DS approaches was then demonstrated through simulations on synthetic RB-STET data. We also showed that the proposed algorithm outperforms the pioneering approach by [40] for the reconstruction of randomly downsampled STET measurements. The experiments with real projections confirmed the robustness of the RB-STET in more realistic (and thus more challenging) imaging conditions.

Overall, this work establishes the potential of RB-STET to produce quality reconstructions of detailed biological objects in low-dose imaging conditions. The development of RB-STET could enable the study of highly electron-sensitive biological samples through less electron-intensive methods. The major development ahead for RB-STET is the development of a physical setup that implements the random I-DS acquisition scheme. The project, led by our collaborator Sylvain Trépout at the Institut Curie, is an intricate one. The main technical challenges include the need for fast but accurate beam-positioning, and the necessity of reducing sample displacement during the random-scanning procedure to prevent reconstruction artifacts. By occurring at a timepoint where cryo-STET studies are emerging [15], this new development could contribute to the feasibility and popularisation of RB-STET in biological sciences.

4 Fast Regularized Reconstruction Framework for SPA

In this chapter, we build upon the comprehensive mathematical framework introduced in Chapter 2 and present a fast, regularized reconstruction framework for SPA[1]. More specifically, our contributions to SPA are three-fold. The first and central part is the design of a fast multicale reconstruction scheme (Section 4.1) that makes the use of our regularized iterative method feasible in SPA. We then use this fast algorithm as a building block of a new joint optimization framework (Section 4.2) that alternates between the reconstruction and the estimation of the unknown orientations. Finally, we propose an alternate splitting scheme for our ADMM-based algorithm (Section 4.3) that removes the need for the nested CG loops.

4.1 Fast Multiscale Reconstruction Scheme

4.1.1 Overview

[2]The determination of a high-resolution 3D structure in SPA is a highly challenging procedure. This is due to a multiplicity of complicating factors, which include the unknown projection orientations, an extreme noise level on the projections, their low contrast, the conformational heterogeneity of molecules, an incomplete knowledge of the imaging physics, and the large number of measurements to process.

To handle this, most SPA packages implement a so-called 3D *iterative-refinement* procedure during which information is gradually added to a rough initial volume estimate. In most instances, the reconstruction at every iteration of this refinement pipeline is carried out independently of the *angular estimation*, which aims at estimating the unknown orientation $\boldsymbol{\theta}_p$ associated to each projection $\mathbf{b}^p$. The reconstruction is usually performed with direct inversion methods, that are fast but lack robustness in adverse imaging conditions.

[1] SPA is introduced in Section 1.4.2.

[2] The content of this section is based on [18]: **Laurène Donati**, Masih Nilchian, Carlos Oscar Sorzano, Michael Unser. *Fast Multiscale Reconstruction for Cryo-EM.* Journal of Structural Biology, 2018. This article was part of the 2018 Special Issue on *Cryo-EM Structure Map and Model Challenges.*

In this work, we propose a multiscale iterative-reconstruction framework for SPA based on the mathematical framework we derived in Section 2.2. By permitting full 3D regularization, our algorithm is a robust alternative to direct methods for performing reconstruction in adverse imaging conditions. We handle the reconstruction of volumes at any desired spatial scale by representing 3D objects with scaled basis functions. The controlled dilation of these basis functions then gives us the possibility to adjust the scale of the representation to the quality of the measurements.

The reconstruction task itself is formulated as a regularized optimization problem that is solved iteratively, as described in Section 2.3.2. To make the use of such an iterative method feasible in SPA, we introduce a fast formulation for the cost-dominant step of the reconstruction. The cost of this new operation does not depend on the number of projection orientations, which results in a substantial speedup. Moreover, the PSF is included inside the reconstruction scheme at no extra computational cost.

This multiscale reconstruction tool generates interesting opportunities for the stabilization of the 3D iterative refinement procedure in SPA, whose convergence still heavily depends on the quality of the initial reconstruction [75, 76]. In particular, we show that reconstructing at a coarse scale increases the robustness to angular misestimations and leads to gains in computational speed. We present reconstructions obtained at different scales from a real dataset of the 2015/2016 EMDataBank Map Challenge[3].

This chapter is structured as follows: the principles behind our multiscale framework and its relevance in the context of SPA are explained in Section 4.1.2. We detail the iterative multiscale reconstruction scheme in Section 4.1.3. In Sections 4.1.4 and 4.1.5, we derive fast formulations for the cost-dominant operations of our iterative scheme. The resulting reduction in computational cost is discussed in Section 4.1.6. Finally, our results are presented in Section 4.1.7 and discussed in Section 4.4.

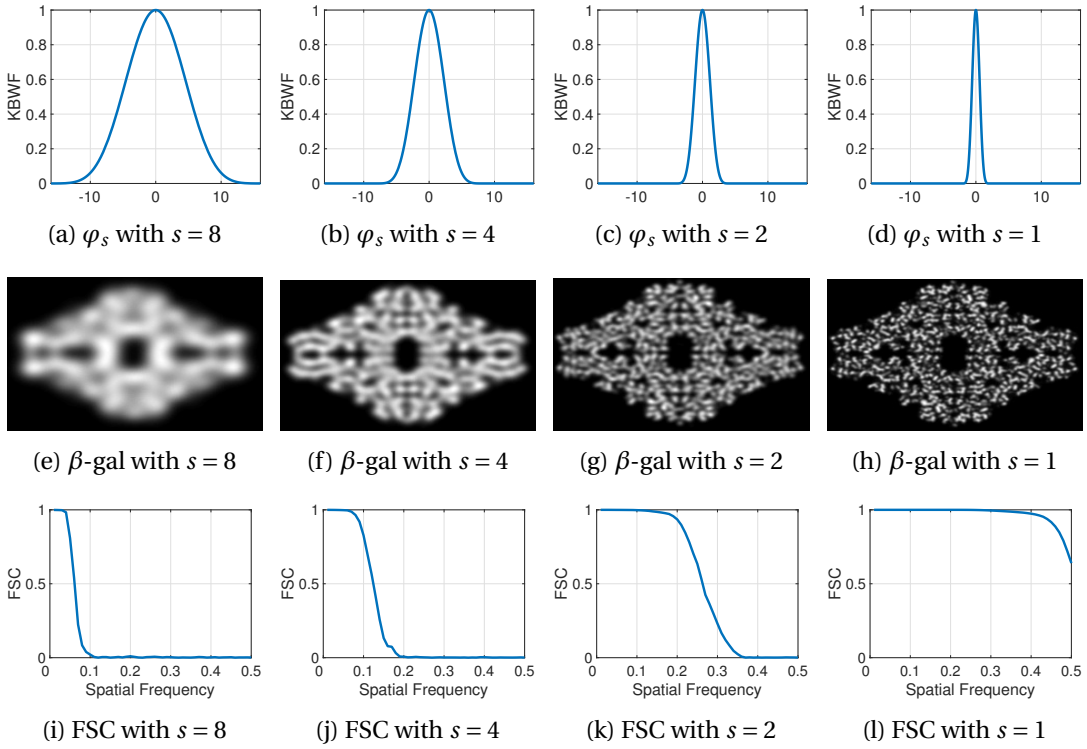
4.1.2 Multiscale Representation

Multiscale for Solving Ill-Posed Inverse Problems

The idea behind multiscale processing is to process signals over a certain range of scales when executing multistep procedures. An advantage is that operations performed at coarse scale are usually more robust and permit gains in computational speed [155]. They come useful when 1) only incomplete and degraded information is available as input, and 2) a low-resolution output is acceptable for further processing. A benefit is that this robustness of the initial coarse estimate can positively impact the convergence of all subsequent steps in the procedure.

This observation is the motivation behind the so-called pyramid approaches [156, 157] that solve complex imaging problems iteratively using a multiscale representation of volumes.

[3] `https://challenges.emdataresource.org/2015_map_challenge`



(a) φ_s with $s = 8$ (b) φ_s with $s = 4$ (c) φ_s with $s = 2$ (d) φ_s with $s = 1$

(e) β-gal with $s = 8$ (f) β-gal with $s = 4$ (g) β-gal with $s = 2$ (h) β-gal with $s = 1$

(i) FSC with $s = 8$ (j) FSC with $s = 4$ (k) FSC with $s = 2$ (l) FSC with $s = 1$

Figure 4.1 – Impact of the controlled dilatation of the basis function φ_s by the scaling parameter $s > 0$ on the reconstruction. This simple dataset consists of 2,000 noiseless projections a $(256 \times 256 \times 256)$ beta-galactosidase volume with equidistributed orientations. The scale is decreased from left ($s = 8$) to right ($s = 1$), with intermediate values $s = 4$ and $s = 2$. **(a-d)** Optimized isotropic KBWF φ_s (with $a = 2$, $\alpha = 10.8$, $m = 2$) are dilated by increasing s. **(e-h)** Central orthogonal slices of the volume reconstructed with the proposed approach at different scales. Volumes are reexpanded on a fine grid through (4.1). **(i-l)** Corresponding FSC curves, which indicate at which frequencies the reconstructions correlates with the ground truth (see Appendix A.3).

Several works have successfully used pyramid processing in biomedical imaging for handling strongly ill-posed optimization problems with abundant local minima [158, 159]. This approach has also been favorably used in alternate minimization frameworks, for example in blind deconvolution works [160, 161].

Multiscale-based approaches have been used in SPA to improve the angular-estimation procedure [162, 163]. In [163], Sorzano *et al.* use a coarse-to-fine discrete wavelet transform to compute the correlation between the measurements and the reference projections. A coarse scale indeed suffices to discriminate projections that come from different orientations. However, if the projections match at a coarse scale, then the analysis is pursued at a finer scale. This multiscale wavelet-space matching algorithm provide gains both in speed and in robustness.

Several conditions specific to SPA advocate for the use of pyramid-like approaches inside

the global refinement procedure itself. For example, it has been shown that the determination of the unknown orientations essentially depends on low-resolution information [164, 165]. Thus, a coarse representation of volumes is more desirable at the early stages of the refinement process. Indeed, its resolution proves sufficient for further processing while it remains robust to the incomplete information.

Multiscale Representation for SPA

We thus look for the s-scaled function $\tilde{f}_s : \mathbb{R}^3 \to \mathbb{R}$ that accurately approximates the 3D object f in a shift-invariant basis. The important aspect here is to consider the *scaled* reconstruction space

$$V_s(\varphi) = \left\{ \tilde{f}_s = \sum_{\boldsymbol{k} \in \mathbb{Z}^3} c_s[\boldsymbol{k}] \varphi_s(\cdot - s\boldsymbol{k}) : c_s \in \ell_2\left(\mathbb{Z}^3\right) \right\} \tag{4.1}$$

specified by the scaled basis function

$$\varphi_s = \varphi(\cdot / s) \in L_2(\mathbb{R}^3), \tag{4.2}$$

where $s > 0$ denotes the scaling parameter. Similarly to Section 2.2.1, we use the KBWF as our basis function φ. This choice is greatly motivated here by the isotropy of KBWFs, which allows for a significant reduction in computational costs as we shall later illustrate in Sections 4.1.4 and 4.1.5.

The coefficient sequence c_s in (4.1) corresponds to the discrete s-scaled representation of the reconstruction $\tilde{f}_s$ in the space $V_s(\varphi)$. This sequence is restricted to a finite number of coefficients as the object f (hence $\tilde{f}_s$) and the considered basis function φ are compactly supported. We write this finite sequence as the vector of coefficients $\mathbf{c}_s = \left(c_s[\boldsymbol{k}]\right)_{\boldsymbol{k} \in \Omega^s_{3\mathrm{D}}}$. Here, the set $\Omega^s_{3\mathrm{D}} \subset \mathbb{Z}^3$ corresponds to the support of the coefficients required to represent the scaled reconstruction $\tilde{f}_s$. After the optimization procedure, the obtained coefficients are reexpanded in the space $V_s(\varphi)$ through (4.1) to obtain the scaled reconstruction $\tilde{f}_s$.

The scaling parameter s in (4.2) is the central element of our multiscale representation. It dilates the basis function φ_s and thus controls the coarseness of the representation $\tilde{f}_s$. A large value of s enforces a coarse ("low-resolution") volume, while a small value of s imposes a fine ("high-resolution") representation. This is illustrated in Figure 4.1, where the coarseness of the reconstructions increases with the scaling parameter.

Finally, the scale s also influences the size of the set of coefficients $\mathbf{c}_s$. More precisely, the number of coefficients decreases as the scale increases, with a size of $\mathbf{c}_s$ given by $\sharp\Omega^s_{3\mathrm{D}} = N/s^3$. The scaling thus strongly impacts the speed of the reconstruction algorithm, as the procedure is then performed on significantly smaller objects. For the sake of clarity, we hereafter use the notation $N_s = N/s^3$, such that $\mathbf{c}_s \in \mathbb{R}^{N_s}$.

4.1.3 Iterative Multiscale Reconstruction

Imaging Model with Multiscale Representation

We recall that our cryo-EM image-formation model is given by

$$\mathbf{b}_{\boldsymbol{j}}^{p} = \big(\mathscr{P}_{\boldsymbol{\theta}_p}\{f\} * h^{p}\big)(\boldsymbol{j}) + \mathbf{n}_{\boldsymbol{j}}^{p}, \tag{4.3}$$

and is detailed in (2.2). We recall that $\mathbf{b}^p$ is the 2D measurement of the 3D object f imaged at orientation $\boldsymbol{\theta}_p$ through a process that involves straight-line electron transmission (via $\mathscr{P}_{\boldsymbol{\theta}_p}$), optical effects (via h^p), and additive Gaussian noise degradation (via $\mathbf{n}^p$). In SPA, the object f actually corresponds to P (supposedly identical) copies of the same structure of interest, with P in the orders of 10^5-10^6. Each of these 3D copies (called "particles") gets imaged once, which yields P projections $\mathbf{b}^p$.

We now assume that the s-scaled approximation $\tilde{f}_s$ of f in (4.1) yielded the measurements given by (4.3). Using again the properties of the X-ray transform, we rewrite (4.3) as

$$\begin{aligned}\mathbf{b}_{\boldsymbol{j}}^{p} &= \big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\tilde{f}_s\} * h^{p}\big)(\boldsymbol{j}) + \mathbf{n}_{\boldsymbol{j}}^{p} && (4.4)\\ &= \sum_{\boldsymbol{k}\in\Omega_{3\mathrm{D}}^{s}} c_s[\boldsymbol{k}]\big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^{p}\big)(\boldsymbol{j} - s\mathbf{M}_{\boldsymbol{\theta}_p{}^{\perp}}\boldsymbol{k}) + \mathbf{n}_{\boldsymbol{j}}^{p}. && (4.5)\end{aligned}$$

This gives us the entries of the imaging matrix $\mathbf{H}_s^p$ as

$$[\mathbf{H}_s^p]_{\boldsymbol{j},\boldsymbol{k}} = \big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^{p}\big)(\boldsymbol{j} - s\mathbf{M}_{\boldsymbol{\theta}_p^{\perp}}\boldsymbol{k}), \tag{4.6}$$

with $\boldsymbol{j} \in \Omega_{2\mathrm{D}}$ and $\boldsymbol{k} \in \Omega_{3\mathrm{D}}^{s}$.

Hence, the discrete forward model for P projections in our multiscale framework is given by

$$\mathbf{b} = \mathbf{H}_s\mathbf{c}_s + \mathbf{n}, \tag{4.7}$$

with $\mathbf{c}_s \in \mathbb{R}^{N_s}$,

$$\mathbf{b} = \begin{bmatrix}\mathbf{b}^1\\ \mathbf{b}^2\\ \vdots\\ \mathbf{b}^P\end{bmatrix} \in \mathbb{R}^{MP}, \quad \mathbf{H}_s = \begin{bmatrix}\mathbf{H}_s^1\\ \mathbf{H}_s^2\\ \vdots\\ \mathbf{H}_s^P\end{bmatrix} \in \mathbb{R}^{MP\times N_s}, \quad \text{and} \quad \mathbf{n} = \begin{bmatrix}\mathbf{n}^1\\ \mathbf{n}^2\\ \vdots\\ \mathbf{n}^P\end{bmatrix} \in \mathbb{R}^{MP}. \tag{4.8}$$

Multiscale Reconstruction Framework

The task at hand is then to recover the set of coefficients $\mathbf{c}_s$ that best explains the complete vector $\mathbf{b}$ of 2D measurements obtained through (4.7). We assume here that we have existing estimates (however inaccurate they may be) of the orientation $\boldsymbol{\theta}_p$ and the PSF h^p for $p =$

$1,\dots,P$. Following our sparsity-based formulation (Section 2.3), we perform the reconstruction by solving the regularized variational problem

$$\widehat{\mathbf{c}}_s = \underset{\mathbf{c}_s\in\mathbb{R}^{N_s}}{\arg\min}\frac{1}{2}\|\mathbf{H}_s\mathbf{c}_s-\mathbf{b}\|_2^2+\lambda\|\mathbf{L}_s\mathbf{c}_s\|_1, \tag{4.9}$$

where $\mathbf{L}_s\in\mathbb{R}^{Q\times N_s}$ is the regularization operator, and $\lambda>0$ is the regularization parameter.

The ADMM subsolvers of (4.9) are given in Algorithm 1 in Section 2.3. As we discuss there, the computational bottleneck of our ADMM-based minimization algorithm is the quadratic step (2.15a) that requires one to solve $n_{\text{ADMM}}>0$ times the linear system

$$\underbrace{\left(\rho\mathbf{L}_s^T\mathbf{L}_s+\mathbf{H}_s^T\mathbf{H}_s\right)}_{\mathbf{A}}\mathbf{c}_s^{k+1}=\mathbf{H}_s^T\mathbf{b}+\rho\mathbf{L}_s^T\left(\mathbf{u}^k+\frac{\boldsymbol{\alpha}^k}{\rho}\right), \tag{4.10}$$

in terms of $\mathbf{c}_s^{k+1}$. Since the matrix $\mathbf{A}=\left(\rho\mathbf{L}_s^T\mathbf{L}_s+\mathbf{H}_s^T\mathbf{H}_s\right)$ cannot be explicitly inverted due to the size of $\mathbf{H}_s^T\mathbf{H}_s$, we solve (4.10) using a conjugate-gradient (CG) scheme at every k-th ADMM iteration. Hence, our ADMM minimization scheme requires the repeated matrix multiplication of $\mathbf{H}_s^T\mathbf{H}_s$ with the updated vector $\mathbf{c}_s$. If not carefully engineered, this operation comes at a heavy computational cost ($\mathcal{O}(N^2)$), which then makes the use of such an iterative reconstruction algorithm out of practical reach for SPA.

4.1.4 Fast Implementation of $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s$

In this section, we provide an alternate mathematical formulation of $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s$ that significantly reduces the computational cost of this operation. As a preliminary, let us define a condition on basis functions (with unit step size) that is required for further calculations.

Definition 1 (Radial Nyquist criterion)**.** *The function $\varphi\in L_2(\mathbb{R}^3)$ satisfies the radial Nyquist criterion (RNC) with respect to the grid $\mathbb{Z}^3$ if $\widehat{\varphi}(\boldsymbol{\omega})=0$ for all $\|\boldsymbol{\omega}\|\geq\pi$.*

A function thus satisfies the RNC if its Fourier transform is zero outside of a ball of radius π. Note that if $\varphi_{s_0}=\varphi(\cdot/s_0)$ satisfies the RNC, then so does $\varphi_s=\varphi(\cdot/s)$ for all $s\geq s_0$, due to the scaling property of the Fourier transform.

In addition, for any pair of functions (f,g) that satisfy the RNC, it holds through Shannon's theorem and the orthonormality of sinc functions [98] that

$$\sum_{\mathbf{n}\in\mathbb{Z}^2}f(\boldsymbol{n})g(\boldsymbol{n})=\int_{\mathbb{R}^2}f(\boldsymbol{x})g(\boldsymbol{x})\mathrm{d}\boldsymbol{x}. \tag{4.11}$$

Finally, we recall that $\mathbf{c}_s=\left(c_s[\boldsymbol{k}]\right)_{\boldsymbol{k}\in\Omega_{3\text{D}}^s}$ is the finite vector of coefficients of the sequence c_s.

We now provide in Theorem 1 our result on the fast computation of $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s$ for 3D objects.

Theorem 1. *Assume that $\varphi_s(\boldsymbol{x}) = \varphi(\boldsymbol{x}/s)$, with $\boldsymbol{x} \in \mathbb{R}^3$ and $s > 0$, satisfies the RNC for all $s \geq s_0$. Moreover, let the imaging matrix $\mathbf{H}_s$ be as defined in* (4.8)*, and $P \in \mathbb{N}^*$ denote the number of projections. Then, for all $s \geq s_0$, the discrete product $\mathbf{H}_s^T \mathbf{H}_s \mathbf{c}_s$ can be computed for $\boldsymbol{k} \in \Omega_{3\mathrm{D}}^s$ as the discrete convolution*

$$[\mathbf{H}_s^T \mathbf{H}_s \mathbf{c}_s]_{\boldsymbol{k}} = (c_s \star r_s)[\boldsymbol{k}], \tag{4.12}$$

with the kernel sequence

$$r_s[\boldsymbol{k}] = |s|^6 \sum_{p=1}^{P} \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\} * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi^\vee\} * q_{1/s}^p\right)(\mathbf{M}_{\boldsymbol{\theta}_p{}^\perp} \boldsymbol{k}). \tag{4.13}$$

Here, the function $q_{1/s}^p(\boldsymbol{y}) = (h^p * (h^p)^\vee)(s\boldsymbol{y})$, with $\boldsymbol{y} \in \mathbb{R}^2$, corresponds to the scaled autocorrelation function of $h^p(\boldsymbol{y})$.

Proof. Through a series of steps (detailed further down the proof), we rewrite the entries of the discrete product $\mathbf{H}_s^T \mathbf{H}_s \mathbf{c}_s$ for $\boldsymbol{k} \in \Omega_{3\mathrm{D}}^s$ as

$$\begin{aligned}
\left[\mathbf{H}_s^T \mathbf{H}_s \mathbf{c}_s\right]_{\boldsymbol{k}} &= \Big[\sum_{p=1}^{P} (\mathbf{H}_s^p)^T \mathbf{H}_s^p \mathbf{c}_s\Big]_{\boldsymbol{k}} \\
&\overset{(i)}{=} \sum_{p=1}^{P} \sum_{\boldsymbol{l} \in \mathbb{Z}^3} c_s[\boldsymbol{l}] \sum_{\boldsymbol{j} \in \mathbb{Z}^2} \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p\right)(\boldsymbol{j} - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp} \boldsymbol{l}) \cdot \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p\right)(\boldsymbol{j} - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp} \boldsymbol{k}) \\
&\overset{(ii)}{=} \sum_{p=1}^{P} \sum_{\boldsymbol{l} \in \mathbb{Z}^3} c_s[\boldsymbol{l}] \int_{\mathbb{R}^2} \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p\right)(\boldsymbol{y} - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp} \boldsymbol{k}) \cdot \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p\right)(\boldsymbol{y} - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp} \boldsymbol{l}) \mathrm{d}\boldsymbol{y} \\
&\overset{(iii)}{=} \sum_{p=1}^{P} \sum_{\boldsymbol{l} \in \mathbb{Z}^3} c_s[\boldsymbol{l}] \int_{\mathbb{R}^2} \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p\right)(\widetilde{\boldsymbol{y}}) \cdot \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p\right)(\widetilde{\boldsymbol{y}} - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}(\boldsymbol{l} - \boldsymbol{k})) \mathrm{d}\widetilde{\boldsymbol{y}} \\
&\overset{(iv)}{=} \sum_{p=1}^{P} \sum_{\boldsymbol{l} \in \mathbb{Z}^3} c_s[\boldsymbol{l}] \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p * \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p\right)^\vee\right)(s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}(\boldsymbol{l} - \boldsymbol{k})) \\
&\overset{(v)}{=} \sum_{\boldsymbol{l} \in \mathbb{Z}^3} c_s[\boldsymbol{l}] \sum_{p=1}^{P} \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s^\vee\} * h^p * (h^p)^\vee\right)(s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}(\boldsymbol{l} - \boldsymbol{k})).
\end{aligned} \tag{4.14}$$

We thus have that the product $\mathbf{H}_s^T \mathbf{H}_s \mathbf{c}_s$ can be computed as the discrete convolution

$$\left[\mathbf{H}_s^T \mathbf{H}_s \mathbf{c}_s\right]_{\boldsymbol{k}} = (c_s \star r_s)[\boldsymbol{k}], \tag{4.15}$$

with the kernel sequence $r_s \in \ell_2(\mathbb{Z}^3)$ given by

$$r_s[\boldsymbol{k}] = \sum_{p=1}^{P} \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s^\vee\} * q^p\right)(s\mathbf{M}_{\boldsymbol{\theta}_p^\perp} \boldsymbol{k}), \tag{4.16}$$

where the function $q^p(\boldsymbol{y}) = (h^p * (h^p)^\vee)(\boldsymbol{y})$, with $(h^p)^\vee$ the reflection of the function h^p and $\boldsymbol{y} \in \mathbb{R}^2$, corresponds to the autocorrelation function of the PSF $h^p(\boldsymbol{y})$.

Equality (i) derives from the definition of $\mathbf{H}_s^p$ given by (4.6) and from the compact support of c_s and $\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\}$. Equality (ii) results from (4.11), which can be invoked here as the function $(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p)(\boldsymbol{n})$ with $\boldsymbol{n} \in \mathbb{Z}^2$ satisfies the RNC. Indeed, as the basis function φ_s satisfies the RNC by hypothesis, so does $\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\}$ through the central-slice theorem, and then so does $\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * h^p$ through the convolution theorem. Equality (iii) is obtained through a simple change of variables, while Equalities (iv) and (v) use properties of the continuous convolution.

We then rewrite the kernel $r_s[\boldsymbol{k}]$ given by (4.16) as:

$$
\begin{aligned}
r_s[\boldsymbol{k}] &= \sum_{p=1}^{P} \big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\} * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s^\vee\} * q^p\big)(s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}) \\
&\overset{(vi)}{=} |s|^2 \sum_{p=1}^{P} \big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\}(\cdot/s) * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi^\vee\}(\cdot/s) * q^p(\cdot)\big)(s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}) \\
&\overset{(vii)}{=} |s|^4 \sum_{p=1}^{P} \big((\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\} * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi^\vee\})(\cdot/s) * q^p_{1/s}(\cdot/s)\big)(s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}) \\
&\overset{(viii)}{=} |s|^6 \sum_{p=1}^{P} \big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\} * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi^\vee\} * q^p_{1/s}\big)(\mathbf{M}_{\boldsymbol{\theta}_p{}^\perp}\boldsymbol{k}),
\end{aligned}
\tag{4.17}
$$

where $q^p_{1/s}(\boldsymbol{y}) = q(s\boldsymbol{y})$. Equality (vi) is obtained after applying twice the scale-invariance property of the X-ray transform

$$
\mathscr{P}_{\boldsymbol{\theta}}\{f(\cdot/s)\}(\boldsymbol{y}) = |s| \cdot \mathscr{P}_{\boldsymbol{\theta}}\{f\}(\boldsymbol{y}/s), \tag{4.18}
$$

where $f \in L_2(\mathbb{R}^3)$ and $s \neq 0$. Equalities (vii) and (viii) are both derived by using the fact that the convolution of two functions $\mathscr{P}_{\boldsymbol{\theta}}\{f\}, \mathscr{P}_{\boldsymbol{\theta}}\{g\} \in L_2(\mathbb{R}^2)$ scaled by a factor $s \neq 0$ satisfies

$$
\big(\mathscr{P}_{\boldsymbol{\theta}}\{f\}(\cdot/s) * \mathscr{P}_{\boldsymbol{\theta}}\{g\}(\cdot/s)\big)(\boldsymbol{y}) = |s|\big(\mathscr{P}_{\boldsymbol{\theta}}\{f\} * \mathscr{P}_{\boldsymbol{\theta}}\{g\}\big)(\boldsymbol{y}/s). \tag{4.19}
$$

□

The benefit of Theorem 1 is that the costly step $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s$ can now be quickly computed as a pointwise multiplication in the Fourier domain. In practice, the discrete convolution $(c_s \star r_s)$ in (4.12) only needs to be computed for $\boldsymbol{k} \in \Omega^s_{\text{3D}}$. We efficiently do this by convolving a padded $\mathbf{c}_s$ with a kernel of finite support $\mathbf{r}_s = \big(r_s[\boldsymbol{k}]\big)_{\boldsymbol{k}\in\Omega^s_{\text{3D}}}$ using FFTs and periodic boundary conditions. This kernel is obtained by first convolving the autocorrelation function of $\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\}$ with the scaled autocorrelation function of the PSF h^p, and then interpolating its value at the sampling points $\mathbf{M}_{\boldsymbol{\theta}_p{}^\perp}\boldsymbol{k}$. Another benefit of Theorem 1 is that the cost of (4.12) does not depend on the number of projections. These two aspects lead to a massive gain in speed for the computation of $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s$, as illustrated in Table 4.1 and quantified in Section 4.1.6.

The RNC criterion given in Definition 1 is a necessary condition for the theoretical result obtained in Theorem 1. We ran practical tests to ensure that the use of the KBWF—a basis

Computation	$\mathbf{H}_s\mathbf{c}$	$\mathbf{H}_s^T\mathbf{b}$	$\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}$	fast $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}$ (4.12)
Runtime [s]	127.70	132.08	257.41	0.81

Table 4.1 – Comparaison of the runtimes for applying the system matrix $\mathbf{H}$ and its variants $\mathbf{H}_s^T$, $\mathbf{H}_s^T\mathbf{H}$, and our fast $\mathbf{H}_s^T\mathbf{H}$ (4.12) for a given experimental setup with $s = 1$. The system matrix (2.7) computes 500 projections of a $(128 \times 128 \times 128)$ volume. The computation of the kernel (4.13) for the fast $\mathbf{H}_s^T\mathbf{H}_s$ takes $118.07s$, and is only needed once for the whole procedure.

function that does not verify the RNC but that is "effectively localized in Fourier" [99]—had negligible impact on the quality of the computation of $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s$ through the discrete convolution (4.12). The choice of KBWF as basis functions is motivated by their isotropy, which further reduces the cost of (4.12) as the autocorrelation of $\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\}$ only needs to be computed once for a randomly selected $\boldsymbol{\theta}_p$.

Note that when no PSF effect is considered, *i.e.* $h^p = \delta$ for $p = \{1, \dots, P\}$, the kernel (4.13) reduces to

$$r_s[\boldsymbol{k}] = |s|^4 \sum_{p=1}^{P} \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\} * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi^\vee\}\right)(\mathbf{M}_{\boldsymbol{\theta}_p{}^\perp}\boldsymbol{k}). \tag{4.20}$$

The special case when $s = 1$ in (4.20) was introduced in X-ray tomography (with a fixed rotation axis) by McCann *et al.* [166], and other works have proposed alternative speedups in parallel-beam tomography [167, 168, 169].

Finally, we remark that, because the fast $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s$ computation (4.15) requires the zero-padding of $\mathbf{c}_s$ prior to applying the FFT, the linear step (4.10) cannot be directly inverted in the Fourier domain through the circular convolution theorem. Hence, the iterative ADMM scheme still has to rely on an inner CG scheme to solve (4.10).

4.1.5 Fast Implementation of $\mathbf{H}_s^T\mathbf{b}$

The linear step (4.10) also requires the computation of the discrete product $\mathbf{H}_s^T\mathbf{b}$. Even though this only needs to be done once for the whole reconstruction procedure, it can also be costly in its own rights ($\mathscr{O}(M^2P^2)$). We present in Theorem 2 a fast way of computing this vector.

Theorem 2. *Assume that $\varphi_s(\boldsymbol{x}) = \varphi(\boldsymbol{x}/s)$, with $\boldsymbol{x} \in \mathbb{R}^3$ and $s > 0$, satisfies the RNC for all $s \geq s_0$. Moreover, let the imaging matrix $\mathbf{H}_s^p$ be as defined in* (4.6)*, the measurements $\mathbf{b}$ be as defined in* (4.3)*, and $P \in \mathbb{N}^*$ denote the number of projections. Then, for all $s \geq s_0$, the matrix-vector product $\mathbf{H}_s^T\mathbf{b}$ can be computed for $\boldsymbol{k} \in \Omega_{3\mathrm{D}}^s$ as*

$$[\mathbf{H}_s^T\mathbf{b}]_{\boldsymbol{k}} = \sum_{p=1}^{P} \left(b^p * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s^\vee\} * (h^p)^\vee\right)(s\mathbf{M}_{\boldsymbol{\theta}_p{}^\perp}\boldsymbol{k}). \tag{4.21}$$

Proof. Through a series of steps (detailed further down the proof), we rewrite the entries of the product $\mathbf{H}_s^T\mathbf{b}$ for $\boldsymbol{k} \in \Omega_{3\mathrm{D}}^s$ as

$$
\begin{aligned}
\left[\mathbf{H}_s^T\mathbf{b}\right]_{\boldsymbol{k}} &= \Big[\sum_{p=1}^{P}(\mathbf{H}_s^p)^T\mathbf{b}^p\Big]_{\boldsymbol{k}} \\
&\overset{(i)}{=} \sum_{p=1}^{P}\sum_{\boldsymbol{j}\in\Omega_{2\mathrm{D}}}\mathbf{b}^p[\boldsymbol{j}]\Big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\}(\cdot - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}) * h^p\Big)(\boldsymbol{j}) \\
&\overset{(ii)}{=} \sum_{p=1}^{P}\int_{\mathbb{R}^2} b^p(\boldsymbol{y})\Big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\}(\cdot - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}) * h^p\Big)(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\
&\overset{(iii)}{=} \sum_{p=1}^{P}\int_{\mathbb{R}^2} b^p(\boldsymbol{y})\int_{\mathbb{R}^2}\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\}(\widetilde{\mathbf{y}} - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k})h^p(\boldsymbol{y}-\widetilde{\mathbf{y}})\mathrm{d}\widetilde{\mathbf{y}}\mathrm{d}\boldsymbol{y} \\
&\overset{(iv)}{=} \sum_{p=1}^{P}\int_{\mathbb{R}^2}\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\}(\widetilde{\mathbf{y}} - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k})\int_{\mathbb{R}^2} b^p(\boldsymbol{y})h^p(\boldsymbol{y}-\widetilde{\mathbf{y}})\mathrm{d}\boldsymbol{y}\mathrm{d}\widetilde{\mathbf{y}} \\
&\overset{(v)}{=} \sum_{p=1}^{P}\int_{\mathbb{R}^2}\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s\}(\widetilde{\mathbf{y}} - s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k})\big(b^p * (h^p)^\vee\big)(\widetilde{\mathbf{y}})\mathrm{d}\widetilde{\mathbf{y}} \\
&\overset{(vi)}{=} \sum_{p=1}^{P}\Big(b^p * \mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi_s^\vee\} * (h^p)^\vee\Big)(s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}).
\end{aligned}
\tag{4.22}
$$

Equality (i) is obtained by applying the definition of $\mathbf{H}_s^p$ given by (4.6). Equality (ii) results from (4.11). Equalities (iii), (v), and (vi) are all obtained by using the definition of the continuous convolution. Equality (iv) is a simple rearrangement of the two integrals. □

The product $\mathbf{H}_s^T\mathbf{b}$ can thus be obtained by computing the convolutions in (4.21) on a fine grid via 2D FFTs, storing the result in a lookup table, and evaluating it at the points $s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}$ through interpolation. This significantly reduces the cost of the operation, as we shall explain in Section 4.1.6. Another benefit is that the computation of $\mathbf{H}_s^T\mathbf{b}$ can be easily parallelized over the set of P projections.

4.1.6 Computational Cost and Implementation

We assume that the task is to reconstruct a volume of size $N = n^3$ from a set of P measurements, each of size $M = m^2$. We also consider the support of our unscaled basis function φ to be W^3, and the scale of the basis function φ_s along each dimension to be s. For simplicity, we do not consider the cost of the PSF inclusion here.

The computation of the $\mathbf{H}_s^T\mathbf{H}_s$ kernel in Theorem 1 requires the autocorrelation of the projection of the unscaled basis function, which is precomputed and stored in a lookup table. Since we are using isotropic basis functions, this lookup table is relatively small. The evaluation of the kernel at the points $\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}$ then requires a cost of $\mathcal{O}(PW^2n)$. Finally, the convolution of the kernel with the current coefficient sequence $\hat{\mathbf{c}}_s$ only requires two FFTs. This comes at a cost of $\mathcal{O}(N\log(n)/s^3)$ at every ADMM iteration.

Algorithm 2 Fast ADMM iterative reconstruction at scale s

Inputs: $\mathbf{c}^0 \in \mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^{MP}$, $s > 0$, $\lambda > 0$, $n_{\text{ADMM}} > 0$, $n_{\text{CG}} > 0$, $\rho > 0$, $\mathbf{h} \in \mathbb{R}^{MP}$ (if $s = 1$)

$\varphi_s \leftarrow$ Initialize based on (4.2)
$\mathbf{r}_s \leftarrow$ Compute based on (4.13)
$\mathbf{H}_s^T\mathbf{b} \leftarrow$ Compute based on (4.21)
$\mathbf{u}^0 = \mathbf{L}_s\mathbf{c}^0$, $\boldsymbol{\alpha}^0 = \mathbf{u}^0$
$k = 0$
while $(k < n_{\text{ADMM}})$ **do**
    $\mathbf{c}_s^{(k+1)} \leftarrow$ Solve (4.10) with CG for n_{CG} iterations using Theorem 1, $\mathbf{r}_s$ and $\mathbf{H}_s^T\mathbf{b}$
    $\mathbf{u}^{(k+1)} = \text{prox}_{\mathscr{R}}\left(\mathbf{L}_s\mathbf{c}_s^{(k+1)} - \frac{\boldsymbol{\alpha}^{(k)}}{\rho}; \frac{\lambda}{\rho}\right)$
    $\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{\alpha}^{(k)} + \rho\left(\mathbf{L}_s\mathbf{c}_s^{(k+1)} - \mathbf{u}^{(k+1)}\right)$
    $k = k + 1$
end while
return $\mathbf{c}_s^{(k+1)}$

The computation of $\mathbf{H}_s^T\mathbf{b}$ through Theorem 2 requires to perform discrete convolutions via 2D FFTs, at a cost in the order of $\mathscr{O}(PM\log(m))$. We store the result in a lookup table. Then, its evaluation at the points $s\mathbf{M}_{\boldsymbol{\theta}_p^\perp}\boldsymbol{k}$ comes at a cost of $\mathscr{O}(PN/s^3)$.

Whether the cost of our approach is comparable to that of direct methods based on Fourier regridding[4] (Section 2.1.1) depends on the specific experimental conditions, especially on the scale desired for the reconstruction. In particular, the coarser the representation of the image, the quicker our reconstruction procedure. For example, when the reconstruction is performed at the scale $s = 4$, the cost of computing $\mathbf{H}_s^T\mathbf{H}_s\mathbf{c}_s$ is reduced by a factor $4^3 = 64$ compared to a fine-scale reconstruction with $s = 1$, which is a massive computational gain.

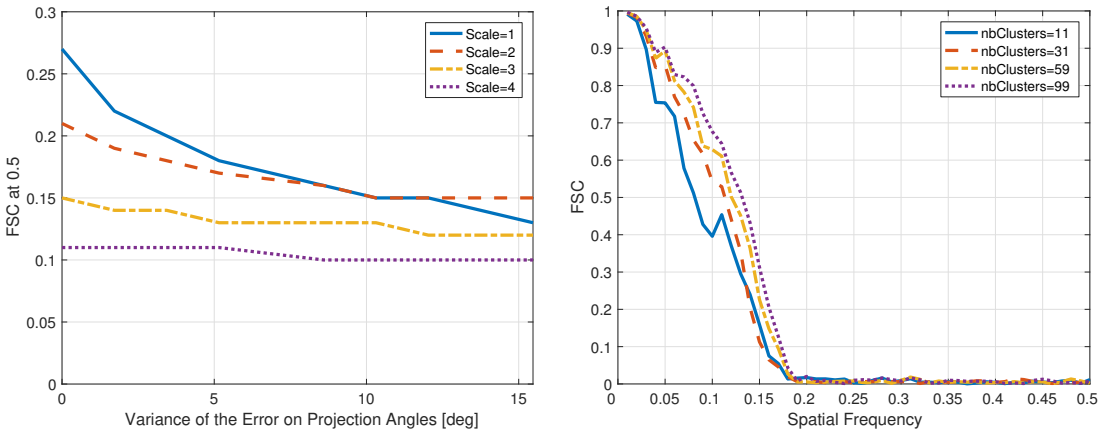
A pseudo-code of the proposed reconstruction algorithm is provided in Algorithm 2. As inputs, the algorithm requires the projections and the desired scale for the reconstruction. Optionally, a stack that contains the PSFs associated to every projection can be given as input. This is only relevant when reconstructing at the finest scale ($s = 1$), as the correction of high-frequency information is pointless at coarse scales ($s > 1$) (see Figure 4.1, bottom row). Another more practical reason for the optionality of this inclusion is that many SPA packages actually deconvolve the projections (*i.e.*, correct the effect of the PSF) prior to reconstruction.

4.1.7 Experiments

We evaluate the performance of our fast multiscale algorithm on simulated and real datasets. In particular, we explore how the scaling impacts the robustness of the reconstruction to angular misestimation, *i.e.*, the incorrect estimation of the orientations $\{\boldsymbol{\theta}_p\}_{p=1}^P$. We describe

[4] By comparison, standard DFR methods require that one 2D FFT be computed per projection ($\mathscr{O}(PM\log(m))$), followed by some interpolation procedure. Assuming that the support of the interpolating function is similar to that of our basis function, this yields a cost of $\mathscr{O}(PW^3M)$. One then needs to apply one 3D inverse Fourier transform to get back to the spatial domain, which comes at a cost of $\mathscr{O}(N\log(n))$.

the experiments performed with simulated data in the next two subsections, and present the results obtained with real data from the 2015/16 EMDataBank Map Challenge in the final subsection.



(a) (b)

Figure 4.2 – **(a)** Effect of the scale on the robustness to angular misestimation for β-galactosidase reconstruction. The reported resolution is the FSC resolution estimate at the threshold value of 0.5. **(b)** FSC curves of β-galactosidase reconstructed at scale $s = 3$ from 20,000 projections clustered in a varying number of classes (n=11,31,59,99) before averaging. The variance of the error on the orientations is 15.47 deg^2.

Simulation Conditions and Implementation Details

We use as ground truth a $(128 \times 128 \times 128)$ β-galactosidase volume with 2.2Å resolution [170]. We compute 20,000 randomly equidistributed projections of the ground truth using the imaging matrix (4.6) with a KBWF with parameters $a = 2$, $\alpha = 10.8$, $m = 2$. We then generate multiple datasets by adding (1) Gaussian noise on the projections, such that their SNR is equal to 1, and (2) a uniformly distributed zero-mean random variable to the ground truth orientations.

For every dataset, we cluster the projections in n distinct classes, as is commonly done in projection matching during the iterative-refinement procedure (see Section 1.4.2). To do so, we generate n projections of the ground truth with uniformly distributed orientations, which serve as "class templates". Each projection in the dataset is then assigned to the class whose template has the closest orientation. Then, for every class, the projections are aligned to the template by rigid transformation (using the `imregister` function in MATLAB®), and averaged. From these class averages, we reconstruct the 3D structure at different scales using our algorithm. We consider small class sizes (from 10 to 100) to mimic the processing conditions during the early stages of the iterative refinement procedure.

For the reconstruction, we solve (4.9) with Algorithm 2. We use TV regularization, *i.e.*, $\mathscr{R}(\mathbf{L}\mathbf{c}_s) = \|\mathbf{c}_s\|_{\mathrm{TV}} = \|\boldsymbol{\nabla}\mathbf{c}_s\|_{2,1}$, with the gradient operator $\boldsymbol{\nabla}$ and the mixed (ℓ_1/ℓ_2)-norm $\|\cdot\|_{2,1}$

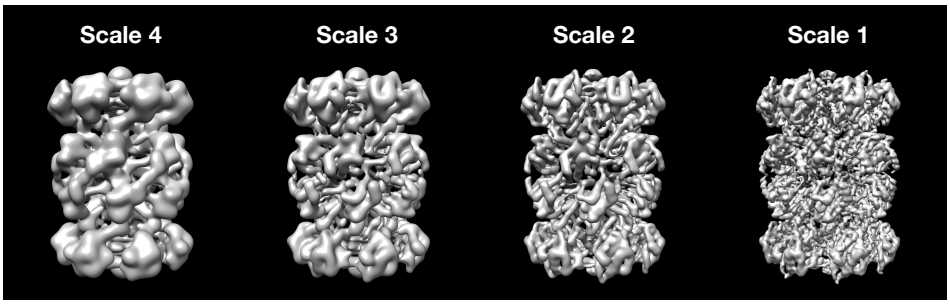


Figure 4.3 – Reconstructions at different scales ($s = 1,2,3,4$) of the T20S Proteasome from the 2015/2016 EMDatabank Challenge.

defined in Section 2.3. We set φ as a KBWF with parameters $a = 4$, $\alpha = 19$, $m = 2$ that are different from those used to generate the measurements in order to reduce the risk of inverse crime. We do not consider the effect of the PSF in these simulations[5].

The number of (outer) ADMM iterations is set to $n_{\text{ADMM}} = 30$, and the number of (inner) CG iterations to $n_{\text{CG}} = 7$. The regularization parameter is chosen by spanning a range of powers of ten (*i.e.*, five values from $\lambda = 10^{-2}$ to $\lambda = 10^{2}$) and selecting the best output in terms of FSC. Since this does not require the recomputation of the kernel $\mathbf{r}_s$ nor of the product $\mathbf{H}_s^T\mathbf{b}$ to do so, the cost of this search is acceptable. The penalty parameter ρ is set proportional to λ. The algorithm is implemented using the GlobalBioIm library [119]. The accuracy of the reconstructions is evaluated by computing the FSC with respect to the ground truth, and considering a threshold value of FSC = 0.5 when necessary.

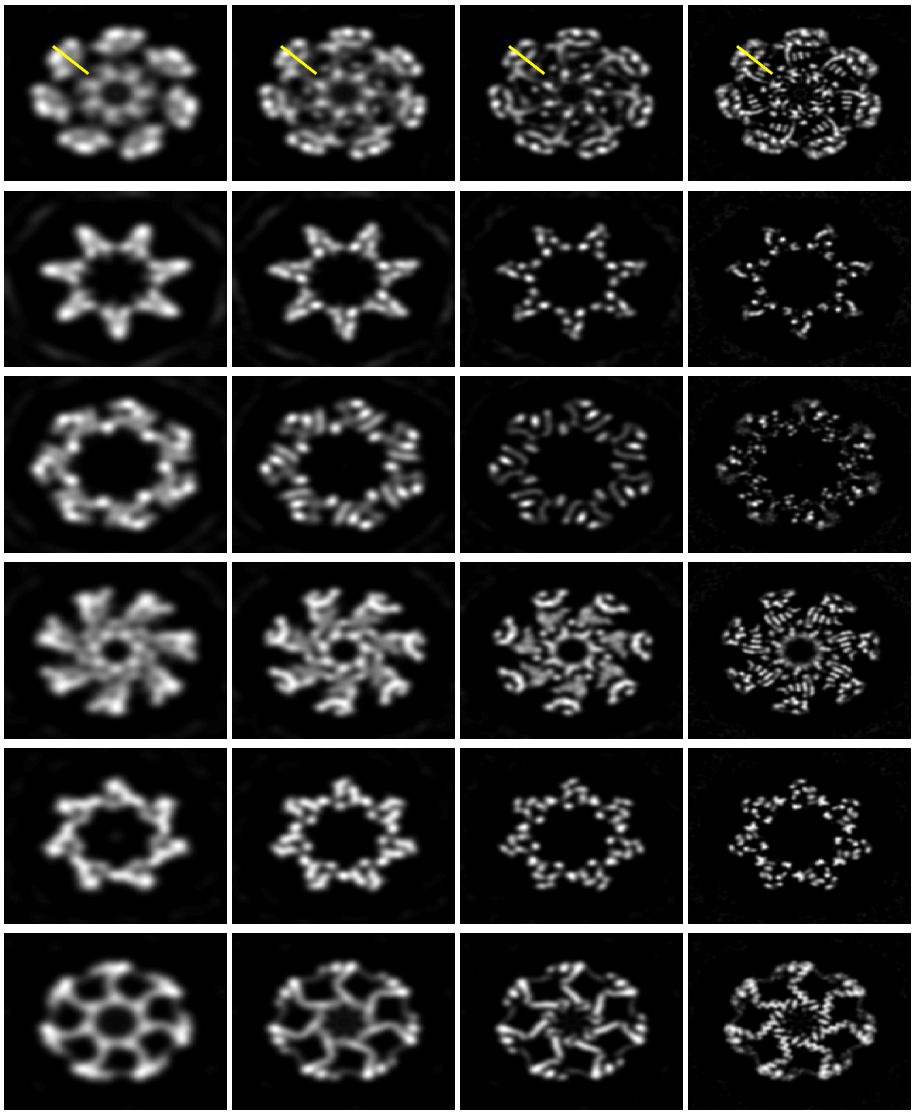
Robustness to Angular Misestimations

We analyze here how angular misestimations influence the quality of the reconstruction at different scales ($s = 1,2,3,4$). To do so, we add an increasing amount of error to the orientations prior to clustering the projections in 80 equidistributed classes and averaging them. We then reconstruct the volumes at scales $s = 1,2,3,4$. For each one, we return the resolution reached at the threshold value FSC = 0.5.

The results presented in Figure 4.2a indicate that, as expected, robustness to angular misestimations increases as the reconstruction is coarsened. At fine scale ($s = 1$), the algorithm performs appropriately when the error level is low, but its performance quickly degrades when the error on the orientations increases. This behavior is much less pronounced at coarser scales ($s = 2,3,4$), where much better robustness against angular misestimations is observed.

We then explore how the scaling influences the quality of the reconstruction in the extreme case where only a very small number of projections classes (*i.e.*, 11) is available and contains angular misestimations (with var=15.47 deg^2). This situation mimics the early stages of the

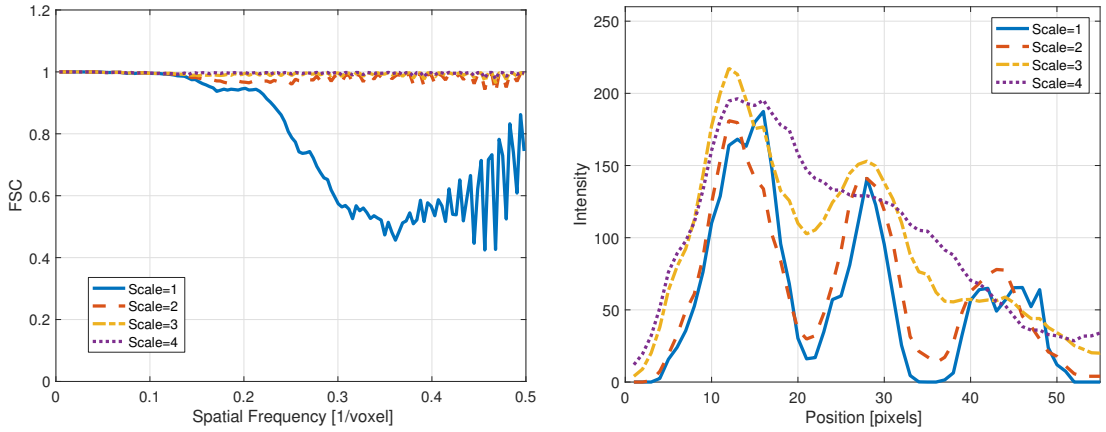
[5]The beneficial impact of including the PSF for producing high-resolution reconstructions has been established in multiple works [85, 32].



(a) Scale 4 (b) Scale 3 (c) Scale 2 (d) Scale 1

Figure 4.4 – Cross sections of T20S Proteasome reconstructions at different scales ($s = 1,2,3,4$). Slices number from top row to bottom row: 75, 100, 125, 150, 175, 200. The yellow lines (top row) indicate the location of the intensity profiles displayed in Figure 4.5b are measured.

iterative-refinement procedure. The results are presented in Figure 4.2b and show that, when performing reconstruction with such a limited amount of data, all output volumes ($s = 1,2,3,4$) have a roughly similar information content. This is not surprising: at all scales, there is only a limited resolution that can be achieved with such blurred and incomplete data, which advocates for multiscale approaches during these early stages of the iterative-refinement procedure. Indeed, the limited available frequency content is preserved at coarse scales, and the associated gains in computational speed can be substantial (see Section 4.1.6).



(a) (b)

Figure 4.5 – **(a)** FSC curves for the T20S Proteasome reconstructions at different scales ($s = 1,2,3,4$). For each scale, the curve was obtained by comparing reconstructions from two distinct halves of set of projections. **(b)** Profile lines taken on a cross-section of the T20S Proteasome reconstructions at different scales ($s = 1,2,3,4$). The position of the measured pixels is indicated by the yellow lines in Figure 4.4

Real Data from the EMDataBank Map Challenge

We now deploy our reconstruction algorithm on a real dataset from the EMDataBank Map Challenge: the T20S Proteasome [171]. The dataset consists in 22,884 projections extracted from 196 micrographs. We classify those into 32 classes using the `CL2D` protocol [172] in the Scipion software package [61]. From those classes, we construct an initial volume with the algorithm in [173] assuming D7 symmetry. We then apply five iterations of the `reconstruct_highres` protocol in Scipion, using the standard direct reconstruction method. The parameters for running those five iterations are identical[6] to those used in [174].

We then perform the final iteration of `reconstruct_highres` using our multiscale reconstruction method, for the scales $s = 1,2,3,4$. For all scales, `reconstruct_highres` separates the dataset in two halves, and performs on each half one iteration of angular estimation and one of reconstruction. We run the `reconstruct_highres` protocol with its default parameters and all the `Post-Processing` options disabled. For all scales, we use the same set of parameters for our algorithm: $\lambda = 100$, $n_{\text{ADMM}} = 30$, $n_{\text{CG}} = 7$. The results are displayed in Figures 4.3, 4.4 and 4.5.

Figure 4.3 presents the reconstructions obtained at the four scales, while Figure 4.4 displays several of their cross-sections. Line profiles taken from a cross-section of the reconstructions are given in Figure 4.5b. These figures illustrate how the scaling influences the coarseness of the reconstructions while preserving their key structural features. The FSC curves of the reconstructed volumes are presented in Figure 4.5a. For each scale, the curve is obtained by

[6] The parameters are described in a paper from the same special issue.

comparing the reconstructions from the two halves of dataset. These results testify to the increase in robustness achieved by coarse reconstructions, which is, once again, consistent with observations made by other multiscale approaches in various biomedical applications [155]. The possible extensions of this framework are discussed in Section 4.4

4.2 Joint Angular Estimation and Reconstruction Scheme

4.2.1 Overview

[7]In this section, we build on the fast reconstruction scheme introduced in Section 4.1 and present a novel optimization framework for SPA that jointly recovers the 3D structure and the projection orientations.

As discussed, most software packages in SPA recover a high-resolution structure by iteratively refining a rough initial volume. Currently, many state-of-the-art iterative-refinement techniques [175, 176, 177, 178] alternate between

1. The reconstruction of the 3D structure for a given set of (however inaccurate) projection orientations using direct methods (see Section 2.1.1);
2. The estimation of the orientations (i.e., *angular estimation*) for all the 2D projections based on the previously reconstructed 3D volume.

For angular estimation, the most commonly-used method is projection matching [68, 69]. This approach compares every projection against a set of "clean" templates generated from the current 3D reconstruction and estimates their orientations based on their closest match. Projection matching thus performs the angular estimation on a *discretized* orientation space. As a consequence, the quality of the angular estimation depends on the fineness of this discretization (and on the quality of the current volume estimate). The main problem is that a fine discretization of $\mathbf{SO}(3)$ implies the generation of a large set of templates. Hence, projection matching is a computationally-expensive procedure. Moreover, iterative-refinement methods based on projection matching tend to perform poorly in low SNR regimes [179]. Alternatively, one can follow a Bayesian approach to formulate the refinement problem as a marginalized MAP estimation [175] (Section 1.4), but this still involves some form of discretization of $\mathbf{SO}(3)$.

In this work, we present a novel iterative-refinement approach for SPA that circumvents the aforementioned problems. More precisely, we formulate the refinement procedure as a

[7]The content of this section is based on [19]: Mona Zehni, **Laurène Donati**, Emmanuel Soubies, Zhizhen J. Zhao, Michael Unser. *Joint Angular Refinement and Reconstruction for Single-Particle Cryo-EM*. IEEE Transaction in Image Processing, 2020. This publication follows from a four-months internship Mona Zehni (University of Illinois at Urbana-Champaign) carried out in the Biomedical Imaging Group, EPFL, under the guidance of Laurène Donati and Emmanuel Soubies. The three were involved in the conception of the project, the development of the theory, the planning of the experiments, and the redaction of the manuscript. MZ carried out most of the implementation and performed the numerical experiments.

joint optimization scheme that alternates between

1. The reconstruction of the 3D structure (for a given set of projection orientations) using the fast regularized scheme introduced in Section 4.1;

2. The estimation of the projection orientations through a semi-coordinate-wise gradient descent, based on the reconstructed volume. More specifically, we use an explicit derivation of the gradient of our objective function to optimize the orientations in the continuum.

The gain with our variational approach is two-fold. First, as discussed in Section 4.1, the use of our ADMM-based reconstruction algorithm brings increased robustness in highly ill-posed conditions, such as those faced at the early stages of refinement for instance. This can then positively impact the convergence of the subsequent steps in the refinement procedure. Second, our angular-estimation method avoids the need for a fine discretization of the orientation space $\mathbf{SO}(3)$. Hence, it does away with the computationally expensive procedure of projection matching. Numerical results demonstrate that our method can efficiently refine a high-resolution 3D structure from projections with initially inaccurate orientations.

This section is organized as follows. In Section 4.2.2, we describe our novel joint optimization framework that is based on fast and robust algorithms. We then compare the performance of our approach to several baselines in Section 4.2.3. We present the obtained results in Section 4.2.4.

4.2.2 Joint Optimization Framework

Discrete Forward Model with In-Plane Translations

For this work, we slightly extend the image-formation model given in (1.6). In particular, we consider the possibility for a p-th projection $\mathbf{b}^p$ to be shifted[8] from its image center by an in-plane translation $\mathbf{t}_p = (t_{p,1}, t_{p,2}) \in \mathbb{R}^2$. We also include sampling steps Δ_1 and Δ_2 of the projection domain. Hence, our image-formation model formulates as

$$\mathbf{b}_{\boldsymbol{j}}^p = \left(\mathscr{P}_{\boldsymbol{\theta}_p}\{f\} * h^p\right)(\boldsymbol{\Lambda}\boldsymbol{j} - \mathbf{t}_p) + \mathbf{n}_{\boldsymbol{j}}^p, \tag{4.23}$$

where $\boldsymbol{\Lambda} = \mathbf{diag}(\Delta_1, \Delta_2) \in \mathbb{R}^{2\times 2}$ is a diagonal matrix formed out of Δ_1 and Δ_2. We then look for an (unscaled) approximation $\tilde{f} : \mathbb{R}^3 \to \mathbb{R}$ of f, and we assume that it yielded the measurements (4.23). Using the generalized sampling scheme (2.3), we then parametrize $\tilde{f}$ to obtain

[8] In SPA, shifts can arise when extracting the projections from their micrographs (see Section 1.4). Shifts can be pre-corrected prior to 3D reconstruction, but this requires methods capable of directly estimating them from the noisy dataset.

the discrete forward model for an orientation $\boldsymbol{\theta}_p$

$$\mathbf{b}_{\boldsymbol{j}}^{p} = \sum_{\boldsymbol{k}\in\Omega_{3\mathrm{D}}} c[\boldsymbol{k}]\big(\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\} * h^{p}\big)(\boldsymbol{\Lambda}\boldsymbol{j} - \mathbf{M}_{\boldsymbol{\theta}^{\perp}}\boldsymbol{k} - \mathbf{t}_p) + \mathbf{n}_{\boldsymbol{j}}^{p}. \tag{4.24}$$

We concisely write (4.24) as

$$\mathbf{b}^{p} = \mathbf{H}(\boldsymbol{\theta}_p, \mathbf{t}_p)\mathbf{c} + \mathbf{n}^{p}, \tag{4.25}$$

where $\mathbf{H}(\boldsymbol{\theta}_p, \mathbf{t}_p) \in \mathbb{R}^{M\times N}$ is the discrete system matrix for an orientation $\boldsymbol{\theta}_p \in \Omega_{\boldsymbol{\theta}}$ and an in-plane translation $\mathbf{t}_p$. The use of the new notation $\mathbf{H}(\boldsymbol{\theta}_p, \mathbf{t}_p)$ for the system matrix is driven by the desire to explicit the *latent variables* of our joint optimization scheme in the next subsection.

We then write the full dataset of P projections as $\mathbf{b} = \{\mathbf{b}^{p}\}_{p=1}^{P} \in \mathbb{R}^{MP}$. Similarly, the set of all orientations is defined as $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_p\}_{p=1}^{P} \in (\Omega_{\boldsymbol{\theta}})^{P}$, and the set of all in-plane translations as $\boldsymbol{\Gamma} = \{\mathbf{t}_p\}_{p=1}^{P} \in (\mathbb{R}^2)^{P}$. Our global forward model is then given by

$$\mathbf{b} = \mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma})\mathbf{c} + \mathbf{n}, \tag{4.26}$$

with $\mathbf{c} \in \mathbb{R}^{N}$,

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}^{1} \\ \vdots \\ \mathbf{b}^{P} \end{bmatrix} \in \mathbb{R}^{MP}, \quad \mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma}) = \begin{bmatrix} \mathbf{H}(\boldsymbol{\theta}_1, \mathbf{t}_1) \\ \vdots \\ \mathbf{H}(\boldsymbol{\theta}_P, \mathbf{t}_P) \end{bmatrix} \in \mathbb{R}^{MP\times N}, \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}^{1} \\ \vdots \\ \mathbf{n}^{P} \end{bmatrix} \in \mathbb{R}^{MP}. \tag{4.27}$$

For the sake of clarity, we hereafter use the notations $\mathbf{H}^{T}\mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma}) = (\mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma}))^{T}\mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma})$ and $\mathbf{H}^{T}(\boldsymbol{\Theta}, \boldsymbol{\Gamma}) = (\mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma}))^{T}$.

Joint Optimization Scheme

Our goal is then to jointly estimate the unknown variables in (4.26), namely the coefficients of the 3D structure $\mathbf{c}$, the projection orientations $\boldsymbol{\Theta}$, and the in-plane translations $\boldsymbol{\Gamma}$. To do so, we express this refinement procedure for $\big(\mathbf{c}, \boldsymbol{\Theta}, \boldsymbol{\Gamma}\big) \in \mathbb{R}^{N} \times (\Omega_{\boldsymbol{\theta}})^{P} \times (\mathbb{R}^2)^{P}$ as the regularized minimization

$$\big(\widehat{\mathbf{c}}, \widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\Gamma}}\big) \in \left\{ \arg\min_{\mathbf{c}, \boldsymbol{\Theta}, \boldsymbol{\Gamma}} \mathscr{J}\big(\mathbf{c}, \boldsymbol{\Theta}, \boldsymbol{\Gamma}\big) \right\}, \tag{4.28}$$

with

$$\mathscr{J}\big(\mathbf{c}, \boldsymbol{\Theta}, \boldsymbol{\Gamma}\big) = \frac{1}{2}\|\mathbf{b} - \mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma})\,\mathbf{c}\|_2^2 + \lambda\mathscr{R}(\mathbf{L}\mathbf{c}), \tag{4.29}$$

where the regularization term $\mathscr{R}(\mathbf{L}\mathbf{c})$ consists of the functional $\mathscr{R} : \mathbb{R}^{Q} \to \mathbb{R}$, the operator $\mathbf{L} \in \mathbb{R}^{Q\times N}$, and the parameter $\lambda > 0$, as described in Section 2.3.2.

Algorithm 3 Joint-Optimization Framework (Global Scheme)

Inputs: $\mathbf{c}^0, \boldsymbol{\Theta}^0, \boldsymbol{\Gamma}^0$

1: $k = 0$
2: **while** not converged **do**
    ▷ Update the 3D structure:
3: $\mathbf{c}^{k+1} = \text{ADMM}\left(\mathscr{J}(\cdot, \boldsymbol{\Theta}^k, \boldsymbol{\Gamma}^k), \mathbf{c}^k\right)$
    ▷ Update the latent variables:
4: $(\boldsymbol{\Theta}^{k+1}, \boldsymbol{\Gamma}^{k+1}) = \text{GD}\left(\mathscr{J}(\mathbf{c}^{k+1}, \cdot, \cdot), \boldsymbol{\Theta}^k, \Gamma^k\right)$
5: $k \leftarrow k+1$
6: **end while**

Return: $\mathbf{c}^k, \boldsymbol{\Theta}^k, \Gamma^k$

Algorithm 4 ADMM (Fast Update of the 3D Structure)

Inputs: $\mathbf{b}$, $\boldsymbol{\Theta}$, $\boldsymbol{\Gamma}$, $\mathbf{c}^0$, $\lambda > 0$, $\rho > 0$, $K_{\text{ADMM}} > 0$

1: $\mathbf{u}^0 = \mathbf{L}\mathbf{c}^0$, $\boldsymbol{\alpha}^0 = \mathbf{u}^0$
2: $k = 0$
3: **while** $k < K_{\text{ADMM}}$ **do**
4: $\mathbf{u}^{k+1} = \text{prox}_{\mathscr{R}}\left(\mathbf{L}\mathbf{c}^k - \boldsymbol{\alpha}^k/\rho; \lambda/\rho\right)$
5: $\mathbf{g} = (\mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma}))^T \mathbf{b} + \rho \mathbf{L}^T (\mathbf{u}^{k+1} - \boldsymbol{\alpha}^k/\rho)$
6: $\mathbf{c}^{k+1} = \left(\mathbf{H}^T \mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma}) + \rho \mathbf{L}^T \mathbf{L}\right)^{-1} \mathbf{g}$
7: $\boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k + \rho\left(\mathbf{u}^{k+1} - \mathbf{L}\mathbf{c}^{k+1}\right)$
8: $k \leftarrow k+1$
9: **end while**

Return: $\mathbf{c}^{K_{\text{ADMM}}}$

To solve (4.28), we alternate between the minimization over the coefficients $\mathbf{c}$ and the minimization over the latent variables $\boldsymbol{\Theta}$ and $\boldsymbol{\Gamma}$. Although the objective function in (4.28) is convex with respect to $\mathbf{c}$, it is not convex with respect to $\boldsymbol{\Theta}$ and $\boldsymbol{\Gamma}$. However, it can be shown[9] that the objective function is differentiable with respect to $\boldsymbol{\Theta}$ and $\boldsymbol{\Gamma}$, but that is usually nondifferentiable (due to $\mathscr{R}$) with respect to $\mathbf{c}$.

These observations dictate the choice of two distinct minimization procedures within our global alternating scheme. For the minimization of $\mathscr{J}$ with respect to $\mathbf{c}$, we use the fast and robust ADMM scheme presented in Section 4.1.6, whose pseudocode is given in Algorithm 4. For the minimization of $\mathscr{J}$ with respect to $\boldsymbol{\Theta}$ and $\boldsymbol{\Gamma}$, we use a gradient descent with line search, as detailed in the next subsection.

The outline of this joint optimization procedure is given in Algorithm 3 and is implemented within the GlobalBioIm library [180]. Note that, at Line 3 of Algorithm 3, we use the notation $\text{ADMM}\left(\mathscr{J}(\cdot, \boldsymbol{\Theta}^k, \boldsymbol{\Gamma}^k), \mathbf{c}^k\right)$ to refer to the minimization of $\mathscr{J}(\cdot, \boldsymbol{\Theta}^k, \boldsymbol{\Gamma}^k)$ using ADMM initialized with $\mathbf{c}^k$. We do the same for the gradient-descent algorithm (see Line 4).

[9] While not completely trivial due to the nonlinearity of $\mathbf{H}(\boldsymbol{\Theta}, \boldsymbol{\Gamma})$ with respect to $\boldsymbol{\Theta}$, this can be handled using the continuity of $\mathscr{P}_{\boldsymbol{\theta}}$ and the chain rule.

Update of the Latent Variables Θ and Γ

We first remark that the least-squares term in (4.29) can be written as

$$\frac{1}{2}\|\mathbf{b}-\mathbf{H}(\boldsymbol{\Theta},\boldsymbol{\Gamma})\,\mathbf{c}\|_2^2=\frac{1}{2}\sum_{p=1}^{P}\|\mathbf{b}^p-\mathbf{H}(\boldsymbol{\theta}_p,\mathbf{t}_p)\,\mathbf{c}\|_2^2. \tag{4.30}$$

Hence, when $\mathbf{c}$ is fixed, the minimization of $\mathscr{J}(\mathbf{c},\cdot,\cdot)$ amounts to solving

$$(\widehat{\boldsymbol{\theta}}_p,\widehat{\mathbf{t}}_p)\in\left\{\arg\min_{\boldsymbol{\theta},\mathbf{t}}\ \mathscr{J}_p(\boldsymbol{\theta},\mathbf{t})\right\} \tag{4.31}$$

for all $p\in\{1,\dots,P\}$, where $\mathscr{J}_p:(\boldsymbol{\theta},\mathbf{t})\mapsto\mathbb{R}$ is defined as

$$\mathscr{J}_p(\boldsymbol{\theta},\mathbf{t})=\frac{1}{2}\|\mathbf{b}^p-\mathbf{H}(\boldsymbol{\theta},\mathbf{t})\,\mathbf{c}\|_2^2. \tag{4.32}$$

As the objective function $\mathscr{J}_p$ is differentiable, the minimization (4.31) can be achieved using gradient-descent steps (assuming proper initializations). Hence, we first need to compute the gradients

$$\nabla_{\boldsymbol{\theta}}\mathscr{J}_p(\boldsymbol{\theta},\mathbf{t})=\left(\frac{\partial\mathscr{J}_p}{\partial\theta_1}(\boldsymbol{\theta},\mathbf{t}),\frac{\partial\mathscr{J}_p}{\partial\theta_2}(\boldsymbol{\theta},\mathbf{t}),\frac{\partial\mathscr{J}_p}{\partial\theta_3}(\boldsymbol{\theta},\mathbf{t})\right) \tag{4.33}$$

$$\nabla_{\mathbf{t}}\mathscr{J}_p(\boldsymbol{\theta},\mathbf{t})=\left(\frac{\partial\mathscr{J}_p}{\partial t_1}(\boldsymbol{\theta},\mathbf{t}),\frac{\partial\mathscr{J}_p}{\partial t_2}(\boldsymbol{\theta},\mathbf{t})\right). \tag{4.34}$$

The explicit expressions of these quantities are provided in Theorem 3 for the case when φ is an isotropic basis function. As a preliminary, we remark that due to this isotropy we can write $\mathscr{P}_{\boldsymbol{\theta}_p}\{\varphi\}=\mathscr{P}\{\varphi\}$ for all $1<p<P$. We also recall that $\boldsymbol{\Lambda}$ is a diagonal matrix whose entries are the sampling steps Δ_1 and Δ_2, and that b^p denotes the continuous counterpart of $\mathbf{b}^p$ through (1.4). Finally, we shall denote with $\star$ the discrete convolution between two periodized sequences.

Theorem 3. *Let φ be an isotropic basis function, and $\mathbf{H}(\boldsymbol{\theta},\mathbf{t})\in\mathbb{R}^{M\times N}$ be defined through* (4.24). *Then, for $v\in\{\theta_1,\theta_2,\theta_3,t_1,t_2\}$, there exist $\mathbf{r}_v\in\mathbb{R}^N$ and $\mathbf{q}_v\in\mathbb{R}^N$ such that*

$$\frac{\partial\mathscr{J}_p}{\partial v}(\boldsymbol{\theta},\mathbf{t})=\frac{1}{2}\mathbf{c}^T\left(\mathbf{r}_v\star\mathbf{c}-2\mathbf{q}_v\right), \tag{4.35}$$

where, with $\psi:\mathbf{y}=(y_1,y_2)\mapsto\left(h^p*\mathscr{P}\{\varphi\}\right)(\mathbf{y})$, for all $\boldsymbol{k}\in\Omega_{3\mathrm{D}}$, we have that

- *if $v=\theta_i$, for $i\in\{1,2,3\}$, then*

$$r_v[\boldsymbol{k}]=\frac{1}{\det(\boldsymbol{\Lambda})}\left(\frac{\partial\mathbf{M}_{\boldsymbol{\theta}^\perp}}{\partial\theta_i}\boldsymbol{k}\right)^T\nabla\left(\psi*\psi^\vee\right)(\mathbf{M}_{\boldsymbol{\theta}^\perp}\boldsymbol{k}), \tag{4.36}$$

$$q_v[\boldsymbol{k}]=\frac{1}{\det(\boldsymbol{\Lambda})}\left(\frac{\partial\mathbf{M}_{\boldsymbol{\theta}^\perp}}{\partial\theta_i}\boldsymbol{k}\right)^T\nabla(b^p*\psi^\vee)(\mathbf{M}_{\boldsymbol{\theta}^\perp}\boldsymbol{k}+\mathbf{t}), \tag{4.37}$$

- *if* $v = t_j$, *for* $j \in \{1,2\}$, *then*

$$r_v[\boldsymbol{k}] = 0, \tag{4.38}$$

$$q_v[\boldsymbol{k}] = \frac{1}{\det(\boldsymbol{\Lambda})} \frac{\partial (b^p * \psi^\vee)}{\partial y_j} (\mathbf{M}_{\boldsymbol{\theta}^\perp} \boldsymbol{k} + \mathbf{t}), \tag{4.39}$$

with $\psi^\vee(\mathbf{y}) = \psi(-\mathbf{y})$, *and* $\frac{\partial \mathbf{M}_{\boldsymbol{\theta}^\perp}}{\partial \theta_i} \in \mathbb{R}^{2\times 3}$ *which contains the entrywise derivatives of* $\mathbf{M}_{\boldsymbol{\theta}^\perp}$.

The proof of Theorem 3 is given in Appendix A.4.1. Among others, it includes the derivation of the gradients of $\left(\psi * \psi^\vee\right)$ and $\left(b_p * \psi^\vee\right)$. In particular, we show that they depend on $\mathscr{P}\{\varphi\}$ and $\frac{\partial \mathscr{P}\{\varphi\}}{\partial y_j}$, whose expressions are provided in Proposition 1 for the specific case when φ is a KBWF.

Proposition 1. *For the case when* φ *is the KBWF defined in* (2.4), *we have that*

$$\mathscr{P}\{\varphi\}(\mathbf{y}) = aA\beta_a(\|\mathbf{y}\|)^{m+\frac{1}{2}} I_{m+\frac{1}{2}}\left(\alpha\beta_a(\|\mathbf{y}\|)\right), \tag{4.40}$$

$$\frac{\partial \mathscr{P}\{\varphi\}}{\partial y_j}(\mathbf{y}) = -\frac{\alpha y_j A}{a}\beta_a(\|\mathbf{y}\|)^{m-\frac{1}{2}} I_{m-\frac{1}{2}}(\alpha\beta_a(\|\mathbf{y}\|)), \tag{4.41}$$

where α, a, *and* m *are the KBWF parameters,* $\beta_a(r) = \sqrt{1-(r/a)^2}$, $A = \frac{\sqrt{2\pi/\alpha}}{I_m(\alpha)}$, *and* I_m *is the modified Bessel function of order* m.

The proof of Proposition 1 is given in Appendix A.4.2.

Equipped with those gradient expressions, we then deploy a semi-coordinate-wise gradient-descent to solve (4.31), as summarized in Algorithm 5. At each iteration, the parameters $\boldsymbol{\theta}_p$ and $\mathbf{t}_p$ are updated sequentially. This permits the use of different stepsizes $\gamma_{\boldsymbol{\theta}}$ and $\gamma_{\mathbf{t}}$ for the orientations and the in-plane translations, respectively, which is crucial to account for the different dynamics between the two variables. Moreover, we use adaptive steps selected according to a backtracking line-search method [181, 182]. Given an initial value, the step is decreased by a factor of the parameter $\eta \in (0,1)$ until the cost corresponding to the updated variable is smaller than its current value (*i.e.*, conditions checked in Steps 6 and 12 in Algorithm 5).

Finally, we divide the projection dataset $\mathbf{b}$ into mini-batches and process them in parallel to further accelerate the update of the latent variables $\boldsymbol{\Theta}$ and $\boldsymbol{\Gamma}$. This parallelization is possible due to the separability of the objective function (4.31) with respect to the (independent) projections.

Comparison of Computational Complexity

We now compare the computational complexity of the projection-matching approach to that of the proposed latent-variable update. Let $N = n^3$ and $M = m^2$, so that $\mathbf{c} \in \mathbb{R}^{n\times n\times n}$ and

$\mathbf{b}^p \in \mathbb{R}^{m\times m}$, with m usually close to n in practice.

1. Projection Matching

Each iteration consists of two steps.

- *Generation of Clean Templates.* Given the current structure $\mathbf{c}$, evaluate $\mathbf{H}(\boldsymbol{\theta},\mathbf{0})\,\mathbf{c}$ for $N_{\theta_1}N_{\theta_2}$ different values of $\boldsymbol{\theta} = (\theta_1,\theta_2,0)$ obtained by sampling $[0,2\pi)$ with N_{θ_1} points and $[0,\pi]$ with N_{θ_2} points. Let $C_H(n,m)$ be the cost of one such evaluation that we shall shortly quantify. Then, the computational complexity of this first step is $\mathcal{O}(N_{\theta_1}N_{\theta_2}C_H(n,m))$.
- *Matching the Projections.* For all $p \in \{1,\dots,P\}$, compare the projection $\mathbf{b}^p$ against the $N_{\theta_1}N_{\theta_2}$ clean templates. This requires rotation and in-plane-translation alignment, whose complexity is $\mathcal{O}(m^2\log(m))$ if done efficiently using polar Fourier transforms [183], spherical harmonics [184], or steerable basis functions [185]. Doing it for every projection and every template thus results in a complexity of $\mathcal{O}(N_{\theta_1}N_{\theta_2}Pm^2\log(m))$.

The complexity of projection matching is thus given by $\mathcal{O}(N_{\theta_1}N_{\theta_2}(Pm^2\log(m)+C_H(n,m)))$, where the cost $C_H(n,m)$ depends on the implementation of the forward model. For instance, some fast implementations rely on the central-slice theorem and nonuniform FFTs. This

Algorithm 5 GD (Update of the Latent Variables)

Inputs: $\gamma_{\boldsymbol{\theta}}^0 > 0$, $\gamma_{\mathbf{t}}^0 > 0$, $K_{\mathrm{GD}} > 0$, $\eta \in (0,1)$, $\boldsymbol{\Theta}^0$, $\boldsymbol{\Gamma}^0$, $\mathbf{c}$

for $p = 1,\dots,P$
  $k = 0$
  while $k < K_{\mathrm{GD}}$
    ▷ Update $\boldsymbol{\theta}_p$
    $\gamma_{\boldsymbol{\theta}} \leftarrow \gamma_{\boldsymbol{\theta}}^0$
    $\boldsymbol{\theta}_p^{k+1} = \boldsymbol{\theta}_p^k - \gamma_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}\mathscr{J}_p(\boldsymbol{\theta}_p^k,\mathbf{t}_p^k)$
    while $\mathscr{J}_p(\boldsymbol{\theta}_p^{k+1},\mathbf{t}_p^k) > \mathscr{J}_p(\boldsymbol{\theta}_p^k,\mathbf{t}_p^k)$
      $\gamma_{\boldsymbol{\theta}} \leftarrow \eta\gamma_{\boldsymbol{\theta}}$
      $\boldsymbol{\theta}_p^{k+1} = \boldsymbol{\theta}_p^k - \gamma_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}\mathscr{J}_p(\boldsymbol{\theta}_p^k,\mathbf{t}_p^k)$
    end while
    ▷ Update $\mathbf{t}_p$
    $\gamma_{\mathbf{t}} \leftarrow \gamma_{\mathbf{t}}^0$
    $\mathbf{t}_p^{k+1} = \mathbf{t}_p^k - \gamma_{\mathbf{t}}\nabla_{\mathbf{t}}\mathscr{J}_p(\boldsymbol{\theta}_p^{k+1},\mathbf{t}_p^k)$
    while $\mathscr{J}_p(\boldsymbol{\theta}_p^{k+1},\mathbf{t}_p^{k+1}) > \mathscr{J}_p(\boldsymbol{\theta}_p^{k+1},\mathbf{t}_p^k)$
      $\gamma_{\mathbf{t}} \leftarrow \eta\gamma_{\mathbf{t}}$
      $\mathbf{t}_p^{k+1} = \mathbf{t}_p^k - \gamma_{\mathbf{t}}\nabla_{\mathbf{t}}\mathscr{J}_p(\boldsymbol{\theta}_p^{k+1},\mathbf{t}_p^k)$
    end while
    $k \leftarrow k+1$
  end while
end for

Return: $\boldsymbol{\Theta}^{K_{\mathrm{GD}}}, \boldsymbol{\Gamma}^{K_{\mathrm{GD}}}$

strategy roughly requires one 3D-FFT of $\mathbf{c}$, one interpolation step to extract the central slice perpendicular to the projection direction, and one inverse 2D-FFT of this slice. This then gives $C_H(n,m) = n^3\log(n^3) + m^2 + m^2\log(m^2)$, and an overall complexity of $\mathcal{O}(N_{\theta_1}N_{\theta_2}(Pm^2\log(m) + n^3\log(n)))$.

2. Proposed Update Scheme

As stated by (4.35) in Theorem 3, the evaluation of the partial derivative $\partial\mathscr{J}_p/\partial\upsilon$ can be done at the cost of one 3D convolution (only required when $\upsilon = \theta_i$), one componentwise subtraction, and one scalar product. This gives a complexity of $\mathcal{O}(n^3\log(n))$, *i.e.*, the cost of computing the 3D convolution in the Fourier domain. We then have to add the cost of computing $\mathbf{r}_\upsilon$ and $\mathbf{q}_\upsilon$ in Theorem 3. First, let us remark that $\frac{\partial\mathbf{M}_{\boldsymbol{\theta}^\perp}}{\partial\theta_i}$ is known in closed form from (1.3). Hence, the complexity of computing $\frac{\partial\mathbf{M}_{\boldsymbol{\theta}^\perp}}{\partial\theta_i}\boldsymbol{k}$ for all $\boldsymbol{k}\in\Omega_{3\mathrm{D}}$ is $\mathcal{O}(n^3)$. We then distinguish two situations:

- *The explicit expressions of* $\nabla\left(\psi * \psi^\vee\right)$ *and* $\nabla(b_p * \psi^\vee)$ *are known.* Given $\boldsymbol{\theta}$ and $\mathbf{t}$, the computation of $\mathbf{r}_\upsilon$ and $\mathbf{q}_\upsilon$ amounts to evaluating these gradients at the points $\mathbf{M}_{\boldsymbol{\theta}^\perp}\boldsymbol{k}$ (or $\mathbf{M}_{\boldsymbol{\theta}^\perp}\boldsymbol{k}+\mathbf{t}$), for $\boldsymbol{k}\in\Omega_{3\mathrm{D}}$, followed by an inner product with $\frac{\partial\mathbf{M}_{\boldsymbol{\theta}^\perp}}{\partial\theta_i}\boldsymbol{k}$, which results in an overall complexity of $\mathcal{O}(n^3)$.
- *The explicit expressions of* $\nabla\left(\psi * \psi^\vee\right)$ *and* $\nabla(b_p * \psi^\vee)$ *are unknown.* Due to their independence with respect to $\boldsymbol{\theta}$ and $\mathbf{t}$, these quantities need only be evaluated once (optionally, with some upsampling) on the grid $\Omega_{2\mathrm{D}}$ using (A.18) with Proposition 1 and discrete convolutions (complexity of $\mathcal{O}(m^2\log(m))$). With this precomputed quantity saved as a lookup table, the evaluation of $\nabla\left(\psi * \psi^\vee\right)$ and $\nabla(b_p * \psi^\vee)$ at the points $\mathbf{M}_{\boldsymbol{\theta}^\perp}\boldsymbol{k}$ (or $\mathbf{M}_{\boldsymbol{\theta}^\perp}\boldsymbol{k}+\mathbf{t}$) is obtained through interpolation. Hence, the computational complexity is $\mathcal{O}(n^3)$ here as well.

Considering that we have P projections and that K_{GD} iterations of gradient descent are performed at each latent-variable update (see Algorithm 3), we obtain an overall complexity of $\mathcal{O}(PK_{\mathrm{GD}}n^3\log(n))$ for our proposed approach.

In practice, K_{GD} is usually kept small, *e.g.*, $K_{\mathrm{GD}} = 3$ in our experiments. As for projection matching, it is recommended [179] to set $N_{\theta_1}N_{\theta_2}$ in the order of n^2 to maintain a precise estimation of the orientations. Hence, the proposed method offers an interesting improvement in runtime over projection matching.

4.2.3 Experiments

Datasets

We assess the performance of our algorithm on two synthetic datasets. The first dataset corresponds to the Holliday junction complex (HJC) structure, while the second corresponds to the Human patched 1 (PTCH1) protein. For each dataset, we generate the synthetic ground-truth from their atomic model [186, 187] in the Protein Data Bank using Chimera [51]. The

sizes of the HJC and PTCH1 volumes used in our simulations are $(90\times90\times90)$ and $(84\times84\times84)$, with voxel sizes of 2.867Å and 1.8Å, respectively. We also synthesize a higher-resolution version of HJC with size $(124\times124\times124)$ and voxel size 2Å. The first two volumes are used in our proof of concept simulations; the last volume is used in the experiment that mimics more realistic cryo-EM conditions.

From the ground truths, we generate P projections using (4.23). We sample the orientation space using P points in an equi-distributed fashion over $\{(\theta_{1,p},\theta_{2,p})\}_{p=1}^{P}$. The in-plane rotations $\{\theta_{3,p}\}_{p=1}^{P}$ are generated by uniformly sampling P points on a $[0,2\pi)$ interval. We apply those in-plane translations by moving the center of the projections randomly by at most $m_{\mathbf{t}}$ pixels in either the horizontal or vertical directions. In our experiments, we use at most 20,000 projections to demonstrate the feasibility of our method. Finally, the projections are corrupted by additive Gaussian noise with zero mean and variance σ^2. The average SNR across all projections is then given by $\mathrm{SNR}_{\mathrm{data}} = 10\log\big(\frac{1}{P}\sum_{p=1}^{P}\frac{\|\widetilde{\mathbf{b}}^p\|_2^2}{\sigma^2}\big)$, where $\widetilde{\mathbf{b}}^p$ corresponds to the noiseless measurement vector (see Section 1.2).

Initial Structure, Orientations, and In-Plane Translations

As an initial volume for the refinement procedures, we use the volume generated by the 3D *ab-initio* model in Relion [188]. For the orientations, we consider two possible initializations.

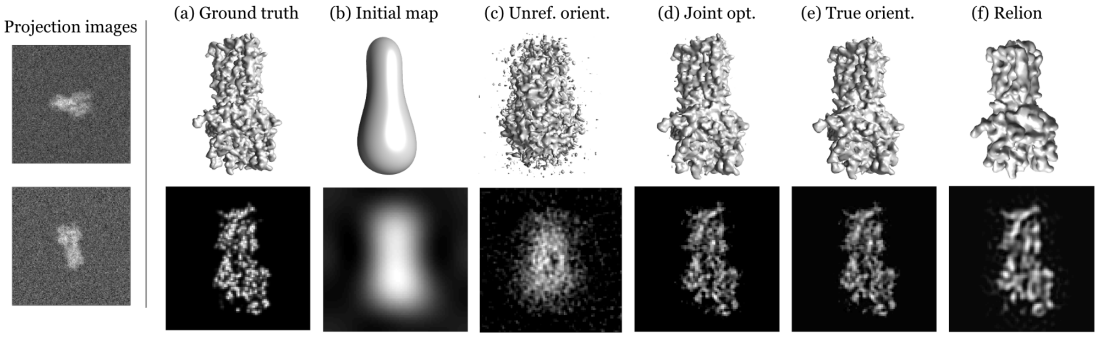
- Model *Init-1*: We add a zero-mean random variable uniformly distributed in $[-e_\theta, e_\theta]$ to the ground-truth orientations, *i.e.*, $\boldsymbol{\theta}_p^{\mathrm{init}} = \boldsymbol{\theta}_p^{\mathrm{true}} + \boldsymbol{\varepsilon}_{\theta,p}$ where $\varepsilon_{\theta,p,j} \sim \mathrm{Unif}(-e_\theta, e_\theta)$, for $p \in \{1,\dots,P\}$ and $j \in \{1,2,3\}$.

- Model *Init-2*: We use the orientations estimated by the 3D *ab-initio* model in Relion as initial orientations $\boldsymbol{\theta}_p^{\mathrm{init}}$, $p \in \{1,\dots,P\}$.

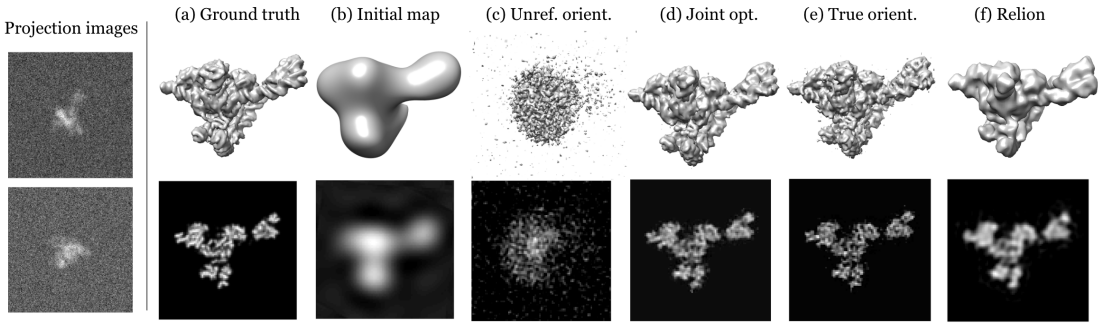
The in-plane translations $\{\mathbf{t}_p\}_{p=1}^{P}$ are all initialized at zero. Our different initializations are thus $\boldsymbol{\Theta}^{\mathrm{init}} = \{\boldsymbol{\theta}_p^{\mathrm{init}}\}_{p=1}^{P}$, $\boldsymbol{\Theta}^{\mathrm{true}} = \{\boldsymbol{\theta}_p^{\mathrm{true}}\}_{p=1}^{P}$, $\boldsymbol{\Gamma}^{\mathrm{init}} = \{\mathbf{0}\}_{p=1}^{P}$, and $\boldsymbol{\Gamma}^{\mathrm{true}} = \{\mathbf{t}_p^{\mathrm{true}}\}_{p=1}^{P}$.

Tuning of the Algorithm Parameters

The following parameters need to be set in our algorithm: λ, ρ, and K_{ADMM}, used in the update of the structure (Algorithm 4), and $\gamma_{\boldsymbol{\Theta}}$, $\gamma_{\mathbf{t}}$, and K_{GD}, used in the update of the latent variables (Algorithm 5). In our experiments, we use $K_{\mathrm{ADMM}} = 2$ or $K_{\mathrm{ADMM}} = 5$, along with $K_{\mathrm{GD}} = 3$, $\gamma_{\boldsymbol{\theta}} = 10^{-7}$, $\gamma_{\mathbf{t}} = 10^{-5}$, and $\eta = 0.25$. The parameters λ and ρ grow like the noise level σ. We use the same set of parameters for the two molecules. We set the parameters of the KBWF to $a = 4$, $\alpha = 19$, and $m = 2$ [100].



(a) PTCH1, $P = 500$, $f_c = 0.03$, $\mathrm{SNR}_{\mathrm{data}} = 3.5781\,(\mathrm{dB})$, $m_{\mathbf{t}} = 0$.



(b) HJC, $P = 500$, $f_c = 0.02$, $\mathrm{SNR}_{\mathrm{data}} = -0.5733\,(\mathrm{dB})$, $m_{\mathbf{t}} = 3$.

Figure 4.6 – Reconstructions of PTCH1 and HJC. Left: Samples of the noisy projections. Top row: 3D structures. Bottom row: Intensity maps of the central slice of the structures. The presented volumes are (a) the ground truth, (b) the initial volume, (c) the reconstruction with unrefined projection orientations, (d) the output of our joint refinement approach, (e) the reconstruction with the true projection orientations, (f) the output of Relion after postprocessing. For both experiments, the latent variables are initialized following the *Init-1* model with $e_\theta = 0.7$ [rad].

Metrics

We assess the quality of the reconstructions with the FSC metric. For the orientations, we visualize the deviations of the estimated angles from their ground-truth values. More precisely, we examine the histogram of the differences $\{\theta^{\mathrm{true}}_{i,p} - \theta^{\mathrm{rec}}_{i,p}\}_{p=1}^{P}$ for $i \in \{1,2,3\}$, and compare it to that of $\{\theta^{\mathrm{true}}_{i,p} - \theta^{\mathrm{init}}_{i,p}\}_{p=1}^{P}$. When the difference between the orientations is small (up to some global rotations), the histogram of the differences is more concentrated around zero. On the contrary, the histogram is more spread for estimations that are further away from their ground-truth values.

Compared Methods

We compare our joint-optimization method to the following approaches:

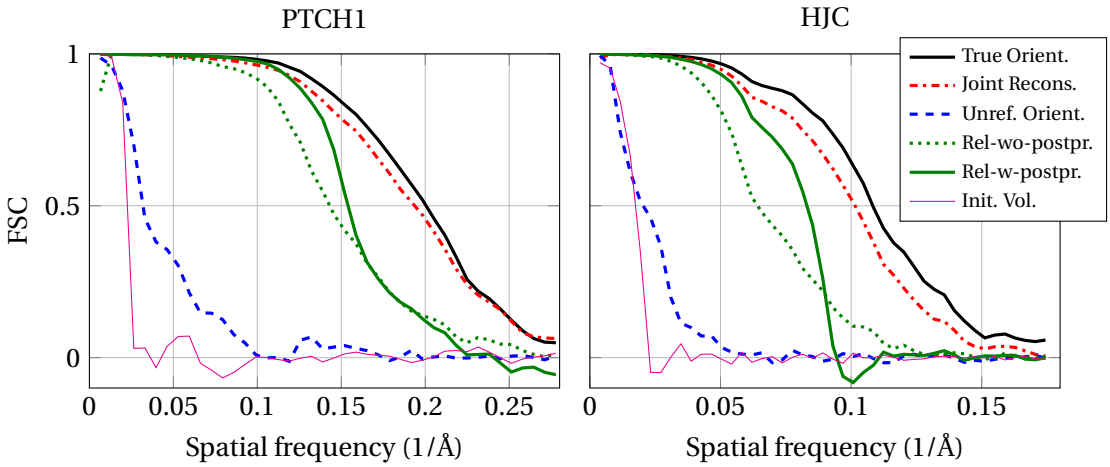


Figure 4.7 – Comparison of the FSC between the 3D structure from various baselines and the ground-truth 3D structure. *Rel-w-postpr* (solid green curve) and *Rel-wo-postpr* (dashed green curve) refer to the Relion results with and without postprocessing, respectively. Note that *Rel-wo-postpr* is obtained after averaging the two half maps. The experimental setups are identical to the ones used in Figure 4.6.

1. *Reconstruction with Unrefined Orientations.* We do not refine the initial orientations and directly reconstruct the 3D structure. This gives an indication of the quality of reconstruction prior to any refinement procedure.

2. *Reconstruction with True Orientations.* We reconstruct the structure with the ground-truth orientations and in-plane translations. This serves as an oracle benchmark that allows us to quantify the improvement brought by the refinement procedures.

3. *Reconstruction with the Relion Package* [175]. We run the *3D auto-refine* function in Relion (version 2.1.0) using its default parameters. For the experiments with the initial in-plane translations set to zero, the *Initial offset range* and *Initial offset step* parameters are set to their minimum values: 0 and 0.1; otherwise, they are set to 4 and 0.5. All the reconstructions obtained with Relion are postprocessed. To reduce the impact of noise, we mask the projections with a soft circular mask with a diameter proportional to the support of the structure. We then apply a low-pass to filter the volumes with a cutoff frequency that corresponds to the final FSC between the two half maps using the postprocessing function in Relion.

4.2.4 Results

Visual Comparison

In Figure 4.6, we compare the 3D structures obtained using our join-optimization scheme (Figure 4.6-(d)) and the other methods (Figure 4.6-(c,e,f)). In this experiment, the latent variables are initialized following the *Init-1* model. As expected, the reconstruction fails when

the unrefined orientations are used (Figure 4.6-(c)), which confirms that angular estimation is required to achieve a successful reconstruction. By contrast, a perfect knowledge of the latent variables leads to a successful reconstruction (Figure 4.6-(e)). The reconstruction obtained with our method (Figure 4.6-(d)) closely resembles that obtained with the perfectly known latent variables, which demonstrates its ability to correctly refine both the 3D structure and the latent variables. Figure 4.6-(f) displays the 3D structure refined by Relion, which is visually less similar to the ground-truth.

FSC Curves

The FSC curves of the reconstructed volumes are sketched in Figure 4.7. These curves confirm that our joint-optimization approach correctly refines the initial low-resolution structure. Indeed, its resolution closely approaches that of the reconstruction obtained with a perfect knowledge of the latent variables. Moreover, our framework outperforms Relion with and without postprocessing.

Quality of Angular Refinement

We display in Figure 4.8 the histogram of the differences between 1) the true and the initial orientations, 2) the true orientations and the ones refined by our method, and 3) the true orientations and the ones refined by Relion. The optimal histogram is obtained when all the differences are zero, up to a global rotation. The corresponding curve then resembles a delta function. We thus observe that our proposed method performs well in refining the orientations, and outperforms Relion in doing so.

Figure 4.9 compares the histogram of the differences between 1) the true and the initial in-plane translations, 2) the true in-plane translations and the ones refined by our method, and 3) the true in-plane translations and the ones refined by Relion. Here as well, the figure demonstrates the ability of our method to refine in-plane translations, and its superior performance compared to Relion.

Convergence Results

The evolution of the volume during our joint-optimization procedure is presented in Figure 4.10. The evolution of the resolution of the reconstruction is shown in Figure 4.11 for our framework and two other baselines. There, r_c marks the radial frequency at which the FSC between the true and the reconstructed structures equals 0.5.

When the orientations are perfectly known, the reconstruction process reaches a high resolution in twenty iterations. Importantly, our framework converges to an almost equally high resolution structure starting from poor angular estimations. Once more, we observe that refining the latent variables is vital to achieve a high-quality reconstruction of the structure.

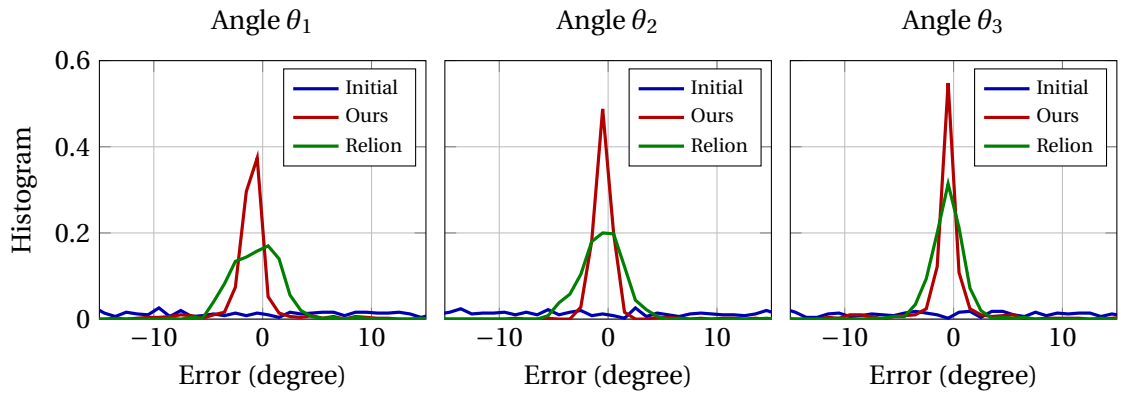


Figure 4.8 – Histogram of the differences between the true and estimated orientations for our method (red curves) and for Relion (green curves). For comparison purposes, the histogram of the differences between the true and the initial unrefined orientations is provided (blue curves). The experimental setup is identical to Figure 4.6 (HJC). The x-axis is truncated between -15 and 15 degrees for visualization purposes.

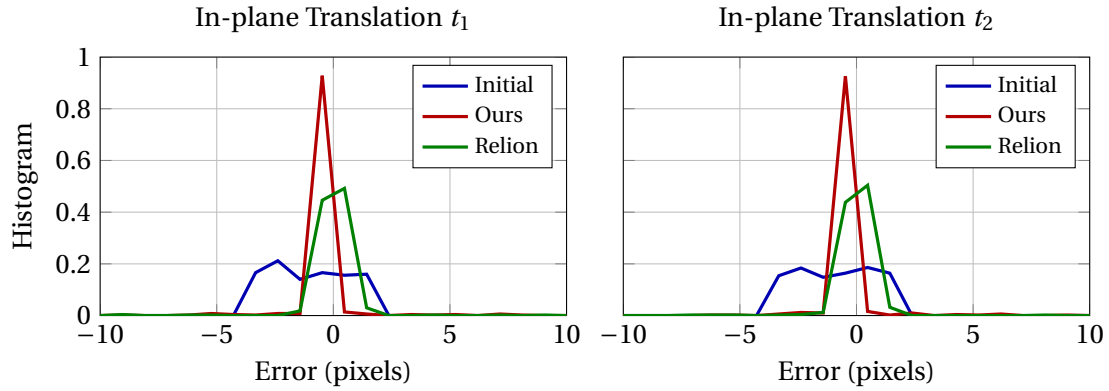


Figure 4.9 – Histogram of the differences between the true and estimated translations for our method (red curves) and for Relion (green curves). For comparison, the histogram of the differences between the true and the initial unrefined translations is provided (blue curves). The experimental setup is identical to Figure 4.6(HJC).

Indeed, reconstruction fails when the set of unrefined orientations is used.

Simulation of a Real Scenario

We mimic here a real scenario in which the output of the 3D *ab initio* method provided by Relion is used to initialize both the structure and the orientations. The in-plane translations are initialized with zeros. We use the HJC structure to synthesize a volume with size $(124\times124\times124)$ and with a voxel size of 2Å. The number of projections is $20,000$ and the average SNR of the projections is -14.2dB. Samples of the projections are provided in Fig. 4.12.

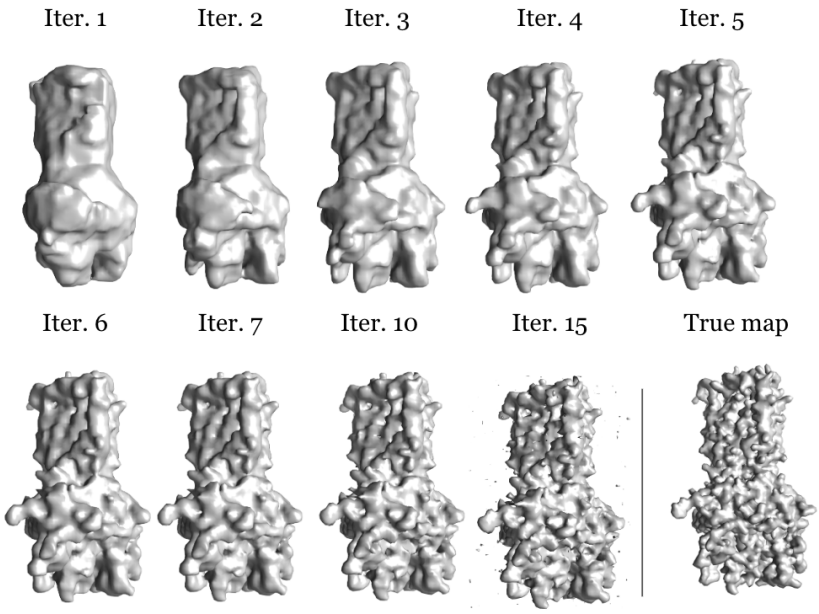


Figure 4.10 – Evolution through iterations for the refinement of PTCH1. The experimental setup is the same as in Figure 4.6.

We split the projection dataset in two halves and refine each half separately, starting from the same initial volume. Independent refinement of the two halves—the outputs of which are called "half maps"—is a common practice in SPA and has two main goals. First, a convergence criterion is obtained by comparing the two refined half maps against each other. More precisely, we stop the refinement when the FSC between the half maps fails to improve from one iteration to the next. Second, it reduces overfitting, especially in high-noise regimes.

A visual comparison of the structures refined by our method and by Relion is presented in Fig. 4.12 (c)-(e). Both results are postprocessed by combining the half maps and filtering out frequencies beyond the final FSC by applying a soft tight mask. A quantitative comparison between our method and Relion is provided in Fig. 4.13 and shows that our method outperforms Relion both with and without postprocessing.

We compare the histogram of the differences between the refined and the ground-truth orientations for our method and Relion in Figures 4.14. We do the same for the in-plane translations in Figure 4.15. In both cases, we observe that our method enjoys comparable performance with Relion for the refinement of the orientations and of the in-plane translations. That being said, while our joint-optimization framework offers a substantial gain over Relion in the proof-of-concept experiments, the difference here is less significant for a larger volume and a noisier regime. Nevertheless, several improvements could further improve our method; we will discuss those in Section 4.4.

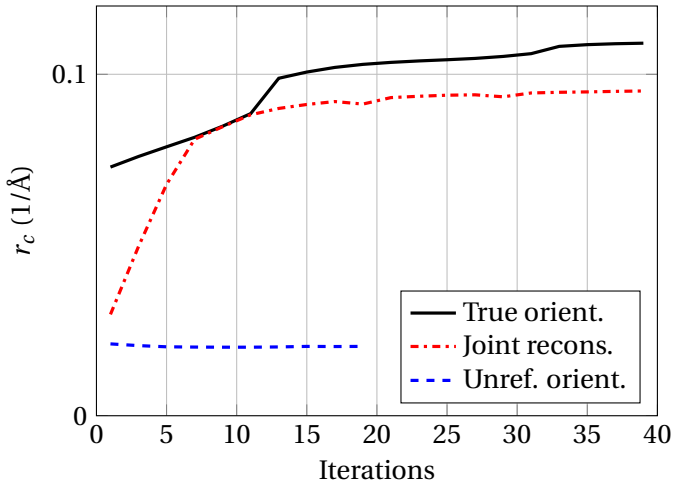


Figure 4.11 – Evolution of r_c, the radial frequency at which the FSC equals 0.5. Comparison between our joint-reconstruction framework (dash-dotted), reconstruction with true orientation projections (solid), and reconstruction with unrefined orientation projections (dashed). The experimental setup is identical to Figure 4.6 (HJC).

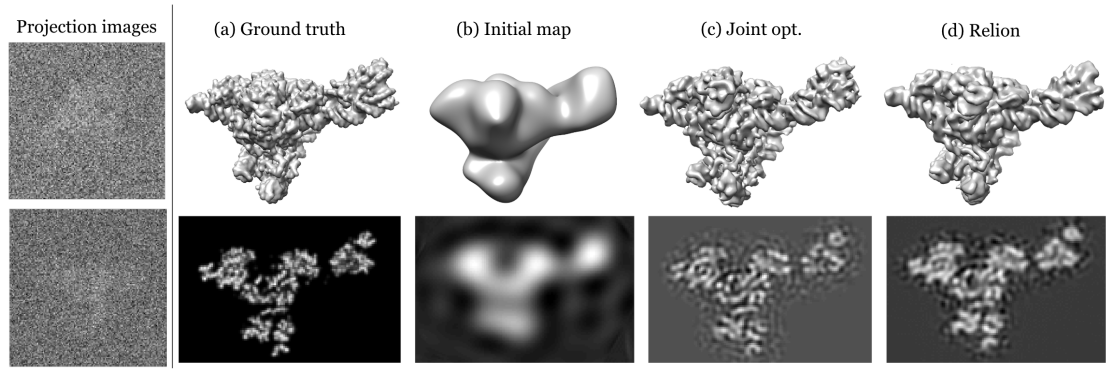


Figure 4.12 – Reconstructions of HJC. Left: Samples of the noisy projections. Top row: 3D structures. Bottom row: Intensity maps of the central slice of the structures. The presented volumes are (a) the ground truth, (b) the initial volume, (c) the output of our joint refinement approach after postprocessing, (d) the output of Relion after postprocessing. For this experiments, the latent variables are initialized following the *Init-2* model using the results from Relion 3D ab-initio modeling. The parameters of this experiment are: $P = 20000$, $\mathrm{SNR}_{\mathrm{data}} = -14.2\,\mathrm{dB}$, $m_{\mathbf{t}} = 3$.

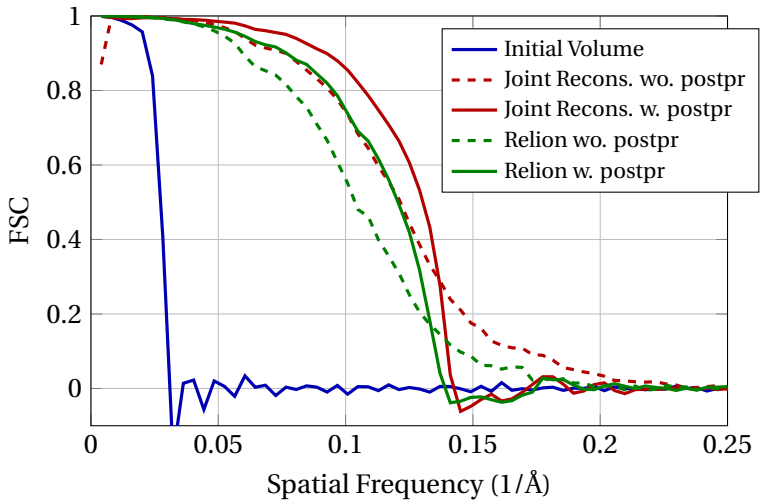


Figure 4.13 – Comparison between the FSC of the structures obtained from several baselines and the ground-truth structure. *Relion-w-postpr* (solid green curve) and *Relion-wo-postpr* (dashed green curve) refer to the Relion results with and without postprocessing, respectively. Note that, *Relion-wo-postpr* is obtained after averaging the two half maps. The experimental setups are identical to the ones used in Figure 4.12.

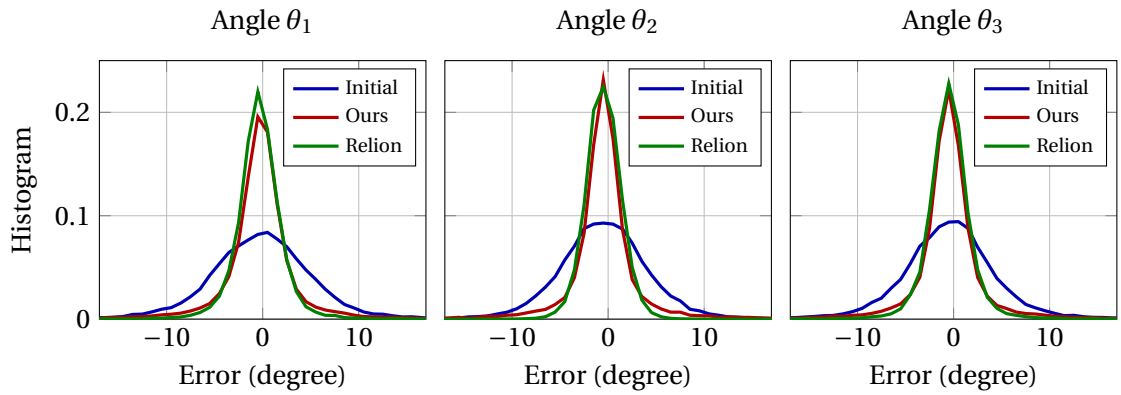


Figure 4.14 – Histogram of the differences between the true and estimated orientations for our method (red curves) and for Relion (green curves). For comparison purposes, the histogram of the differences between the true and the initial unrefined orientations is provided (blue curves). The experimental setup is identical to Figure 4.12. The x-axis is truncated between -17 and 17 degrees for visualization purposes.

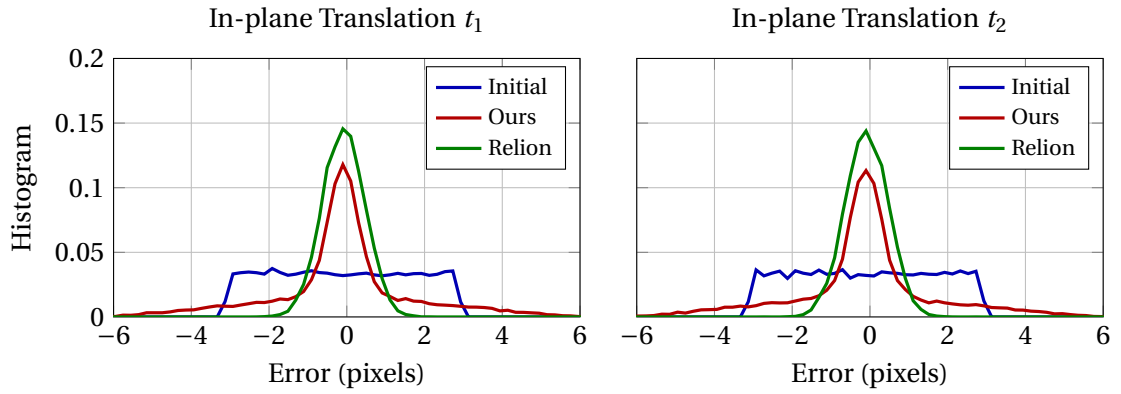


Figure 4.15 – Histogram of the differences between the true and estimated translations for our method (red curves) and for Relion (green curves). For comparison, the histogram of the differences between the true and the initial unrefined translations is provided (blue curves). The experimental setup is identical to Figure 4.12.

4.3 Inner-loop-free ADMM

4.3.1 Overview

[10]In Section 4.1, we derived a fast formulation for the cost-dominant term of our ADMM-based iterative scheme. As discussed then, this development is essential to the deployment of our iterative reconstruction scheme in SPA, and to the design of the joint-optimization framework presented in Section 4.2. Yet, our "standard" ADMM loop still relies on an inner CG scheme to solve the linear step in its minimization procedure, which leads to an algorithmic cost that grows prohibitively with the number of inner CG iterations required for the reconstruction.

In this work, we present an inner-loop-free ADMM algorithm for cryo-EM inspired by [189]. By using an appropriate splitting scheme, we eliminate the need for a nested CG to solve the linear step. Experiments show that this improvement leads to a noticeable increase in algorithmic speed.

4.3.2 ADMM Without Inner CG Loops

We recall that our SPA reconstruction framework formulates as

$$\widehat{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \lambda\mathscr{R}(\mathbf{L}\mathbf{c}) + i_{\geq 0}(\mathbf{c}), \tag{4.42}$$

where $\mathscr{R}(\mathbf{L}\mathbf{c})$ injects prior knowledge into the reconstruction process with a strength set by $\lambda > 0$, with $\mathscr{R} : \mathbb{R}^Q \to \mathbb{R}$ and $\mathbf{L} \in \mathbb{R}^{Q\times N}$. We also consider here the additional constraint $i_{\geq 0} : \mathbb{R}^N \to \mathbb{R}$ that imposes the nonnegativity of the reconstruction. In our standard ADMM formulation (Section 2.3), the splitting scheme writes as

$$\widehat{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \lambda\mathscr{R}(\mathbf{u}) + i_{\geq 0}(\mathbf{c}) \quad s.t. \quad \mathbf{u} = \mathbf{L}\mathbf{c}, \quad \mathbf{v} = \mathbf{c}, \tag{4.43}$$

with $\mathbf{u} \in \mathbb{R}^Q$, $\mathbf{v} \in \mathbb{R}^N$ the two auxiliary variables. As previously discussed, the cost-dominant operation of the ADMM scheme that derives from (4.43) is the linear step

$$\Big(\rho_{\mathrm{u}}\mathbf{L}^T\mathbf{L} + \rho_{\mathrm{v}}\mathbf{I} + \mathbf{H}^T\mathbf{H}\Big)\mathbf{c}^{k+1} = \mathbf{H}^T\mathbf{b} + \rho_{\mathrm{u}}\mathbf{L}^T\Big(\mathbf{u}^{k+1} + \frac{\boldsymbol{\alpha}_{\mathrm{u}}^k}{\rho_{\mathrm{u}}}\Big) + \rho_{\mathrm{v}}\Big(\mathbf{v}^{k+1} + \frac{\boldsymbol{\alpha}_{\mathrm{v}}^k}{\rho_{\mathrm{v}}}\Big), \tag{4.44}$$

that needs to be solved in terms of $\mathbf{c}^{k+1}$. Here, $\mathbf{I} \in \mathbb{R}^{N\times N}$ is the identity matrix, $\boldsymbol{\alpha}_{\mathrm{u}} \in \mathbb{R}^Q$, $\boldsymbol{\alpha}_{\mathrm{v}} \in \mathbb{R}^N$, are the Lagrangian multipliers, and $\rho_{\mathrm{u}}, \rho_{\mathrm{v}} > 0$ are their associated penalty parameters. In all that precedes, we relied on an iterative CG to solve (4.44) since we cannot explicitly build the inverse of the matrix $\big(\rho_{\mathrm{u}}\mathbf{L}^T\mathbf{L} + \rho_{\mathrm{v}}\mathbf{I} + \mathbf{H}^T\mathbf{H}\big)$ nor can we perform the inversion in the Fourier domain due to the padding of $\mathbf{c}^{k+1}$ prior to the convolution.

[10]The content of this section is based on [20]: **Laurène Donati**, Emmanuel Soubies, Michael Unser. *Inner-Loop-Free ADMM for Cryo-EM*. IEEE International Symposium on Biomedical Imaging (ISBI), 2019.

To avoid the use of CG, we propose a novel splitting of (4.42) inspired by [189]. The idea is to add a tailored auxiliary variable $\mathbf{w} \in \mathbb{R}^N$ in the splitting scheme to simplify the standard ADMM algorithm and makes its linear step directly solvable. We formulate this new splitting scheme as

$$\widehat{\mathbf{c}} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \lambda\mathscr{R}(\mathbf{u}) + i_{\geq 0}(\mathbf{v}) \quad s.t. \quad \mathbf{u}=\mathbf{L}\mathbf{c}, \quad \mathbf{v}=\mathbf{c}, \quad \mathbf{w}=\mathbf{T}\mathbf{c}, \tag{4.45}$$

where $\mathbf{T} = (\mathbf{A}-\mathbf{H}^T\mathbf{H})^{1/2} \in \mathbb{R}^{N\times N}$ with $\mathbf{A}$ a symmetric positive semidefinite matrix such that $\mathbf{A} \succeq \mathbf{H}^T\mathbf{H}$. In other words, we impose that the matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ dominate the term $\mathbf{H}^T\mathbf{H}$. In this work, we set $\mathbf{A} = \|\mathbf{H}\|^2\mathbf{I}$.

The augmented Lagrangian for (4.45) is then given by

$$\begin{aligned}\mathscr{L}\big(\mathbf{c},\mathbf{u},\boldsymbol{\alpha}_{\mathrm{u}},\mathbf{v},\boldsymbol{\alpha}_{\mathrm{v}},\mathbf{w},\boldsymbol{\alpha}_{\mathrm{w}}\big) = &\frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \lambda\mathscr{R}(\mathbf{u}) + \langle\boldsymbol{\alpha}_{\mathrm{u}},\mathbf{u}-\mathbf{L}\mathbf{c}\rangle + \frac{\rho_{\mathrm{u}}}{2}\left\|\mathbf{u}-\mathbf{L}\mathbf{c}\right\|_2^2 \\ &+ i_{\geq 0}(\mathbf{v}) + \langle\boldsymbol{\alpha}_{\mathrm{v}},\mathbf{v}-\mathbf{c}\rangle + \frac{\rho_{\mathrm{v}}}{2}\|\mathbf{v}-\mathbf{c}\|_2^2 + \langle\boldsymbol{\alpha}_{\mathrm{w}},\mathbf{w}-\mathbf{T}\mathbf{c}\rangle + \frac{\rho_{\mathrm{w}}}{2}\left\|\mathbf{w}-\mathbf{T}\mathbf{c}\right\|_2^2,\end{aligned} \tag{4.46}$$

where $\boldsymbol{\alpha}_{\mathrm{w}} \in \mathbb{R}^N$ is the additional vector of Lagrangian multipliers and $\rho_{\mathrm{w}} > 0$ its associated penalty parameter. The ADMM alternating scheme is thus (temporarily) composed of 7 steps.

- Minimization *w.r.t.* the auxiliary variables $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$:

$$\mathbf{u}^{k+1} = \mathrm{prox}_{\mathscr{R}}\Big(\mathbf{L}\mathbf{c}^k - \frac{\boldsymbol{\alpha}_{\mathrm{u}}^k}{\rho_{\mathrm{u}}}; \frac{\lambda}{\rho_{\mathrm{u}}}\Big) \tag{4.47}$$

$$\mathbf{v}^{k+1} = \mathrm{prox}_{i_{\geq 0}}\Big(\mathbf{c}^k - \frac{\boldsymbol{\alpha}_{\mathrm{v}}^k}{\rho_{\mathrm{v}}}; \frac{1}{\rho_{\mathrm{v}}}\Big) \tag{4.48}$$

$$\mathbf{w}^{k+1} = \underset{\mathbf{w}\in\mathbb{R}^N}{\arg\min} \frac{\rho_{\mathrm{w}}}{2}\left\|\mathbf{w}-\mathbf{T}\mathbf{c}^k + \frac{\boldsymbol{\alpha}_{\mathrm{w}}^k}{\rho_{\mathrm{w}}}\right\|_2^2 = \mathbf{T}\mathbf{c}^k - \frac{\boldsymbol{\alpha}_{\mathrm{w}}^k}{\rho_{\mathrm{w}}}. \tag{4.49}$$

- Minimization *w.r.t.* to the object $\mathbf{c}$:

$$\begin{aligned}\mathbf{c}^{k+1} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} &\frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \frac{\rho_{\mathrm{v}}}{2}\left\|\mathbf{v}^{k+1}-\mathbf{c}+\frac{\boldsymbol{\alpha}_{\mathrm{v}}^k}{\rho_{\mathrm{v}}}\right\|_2^2 \\ &+ \frac{\rho_{\mathrm{u}}}{2}\left\|\mathbf{u}^{k+1}-\mathbf{L}\mathbf{c}+\frac{\boldsymbol{\alpha}_{\mathrm{u}}^{k+1}}{\rho_{\mathrm{u}}}\right\|_2^2 + \frac{\rho_{\mathrm{w}}}{2}\left\|\mathbf{w}^{k+1}-\mathbf{T}\mathbf{c}+\frac{\boldsymbol{\alpha}_{\mathrm{w}}^k}{\rho_{\mathrm{w}}}\right\|_2^2.\end{aligned} \tag{4.50}$$

- Update of the dual variables $\boldsymbol{\alpha}_{\mathrm{u}}$, $\boldsymbol{\alpha}_{\mathrm{v}}$ and $\boldsymbol{\alpha}_{\mathrm{w}}$:

$$\boldsymbol{\alpha}_{\mathrm{u}}^{k+1} = \boldsymbol{\alpha}_{\mathrm{u}}^k + \rho_{\mathrm{u}}\big(\mathbf{u}^{k+1}-\mathbf{L}\mathbf{c}^{k+1}\big) \tag{4.51}$$

$$\boldsymbol{\alpha}_{\mathrm{v}}^{k+1} = \boldsymbol{\alpha}_{\mathrm{v}}^k + \rho_{\mathrm{v}}\big(\mathbf{v}^{k+1}-\mathbf{c}^{k+1}\big) \tag{4.52}$$

$$\boldsymbol{\alpha}_{\mathrm{w}}^{k+1} = \boldsymbol{\alpha}_{\mathrm{w}}^k + \rho_{\mathrm{w}}\big(\mathbf{w}^{k+1}-\mathbf{T}\mathbf{c}^{k+1}\big). \tag{4.53}$$

The operations (4.47), (4.48), (4.51) and (4.52) are identical to the ones found in the standard

ADMM algorithm. As stated earlier, the proposed algorithm thus necessitates two steps more than the standard ADMM. The trick to reduce the number of steps is that we can plug (4.49) into (4.50), which then simplifies to:

$$\mathbf{c}^{k+1} = \underset{\mathbf{c}\in\mathbb{R}^N}{\arg\min} \frac{1}{2}\|\mathbf{H}\mathbf{c}-\mathbf{b}\|_2^2 + \frac{\rho_\mathrm{u}}{2}\left\|\mathbf{u}^{k+1} - \mathbf{L}\mathbf{c} + \frac{\boldsymbol{\alpha}_\mathrm{u}^{k+1}}{\rho_\mathrm{u}}\right\|_2^2 + \frac{\rho_\mathrm{v}}{2}\left\|\mathbf{v}^{k+1} - \mathbf{c} + \frac{\boldsymbol{\alpha}_\mathrm{v}^{k}}{\rho_\mathrm{v}}\right\|_2^2 + \frac{\rho_\mathrm{w}}{2}\left\|\mathbf{T}(\mathbf{c}^k - \mathbf{c})\right\|_2^2. \tag{4.54}$$

Operations (4.49) and (4.53) can then be set aside as they are not used in any of the other alternating steps. Moreover, if one sets $\rho_\mathrm{w} = 1$, the minimizer of (4.54) satisfies

$$\left(\rho_\mathrm{u}\mathbf{L}^T\mathbf{L} + \rho_\mathrm{v}\mathbf{I} + \mathbf{A}\right)\mathbf{c}^{k+1} = \mathbf{H}^T\mathbf{b} + \rho_\mathrm{u}\mathbf{L}^T(\mathbf{u}^{k+1} + \boldsymbol{\alpha}_\mathrm{u}^k/\rho_\mathrm{u}) + \rho_\mathrm{v}(\mathbf{v}^{k+1} + \boldsymbol{\alpha}_\mathrm{v}^k/\rho_\mathrm{v}) + (\mathbf{A} - \mathbf{H}^T\mathbf{H})\mathbf{c}^k. \tag{4.55}$$

The elegance of this formulation is that (4.55) can be solved directly without relying on a nested CG. Indeed, the matrix $\left(\rho_\mathrm{u}\mathbf{L}^T\mathbf{L} + \rho_\mathrm{v}\mathbf{I} + \mathbf{A}\right)$ is easily invertible in the Fourier domain.

For a volume of size N, this new scheme leads to a complexity of $\mathcal{O}(2N\log(N))$ per ADMM iteration. By comparison, the standard ADMM with K inner CG loops has a complexity of $\mathcal{O}(2(K+1)N\log(N))$ per iteration. A pseudocode of our inner-loop-free ADMM is presented in Algorithm 6.

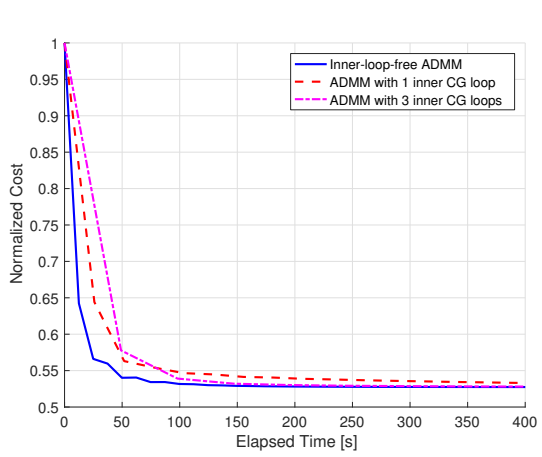
4.3.3 Experiments

We compare the convergence speed of the standard ADMM scheme and that of the proposed inner-loop-free algorithm by reconstructing a synthetic dataset. We use as our ground truth a $(256 \times 256 \times 256)$ β-galactosidase volume [170]. We put ourselves in challenging imaging conditions, such as those faced at the beginning of the refinement procedure (*i.e.*, few data,

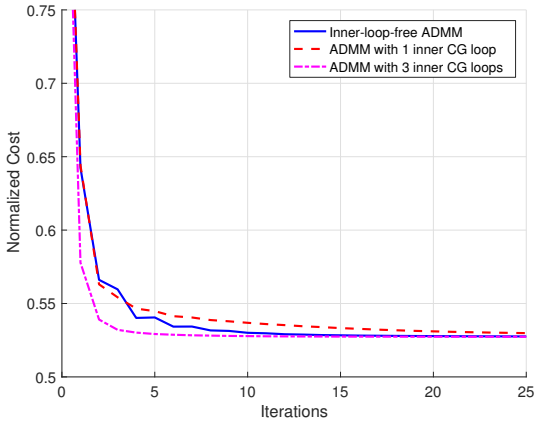
Algorithm 6 Inner-loop-free ADMM (proposed)

Require: $\mathbf{b} \in \mathbb{R}^{MP}$, $\mathbf{c}^0 \in \mathbb{R}^N$, $\mathbf{A} = \|\mathbf{H}\|^2\mathbf{I} \in \mathbb{R}^{N\times N}$, $\lambda > 0$, $\rho > 0$

$\rho_\mathrm{u} = \rho_\mathrm{v} = \rho$
$\mathbf{u}^0 = \mathbf{L}\mathbf{c}^0$, $\boldsymbol{\alpha}_u^0 = \mathbf{u}^0$
$\mathbf{v}^0 = \mathbf{c}^0$, $\boldsymbol{\alpha}_v^0 = \mathbf{v}^0$
$k = 1$
while (not converged) **do**
    $\mathbf{u}^{k+1} = \mathrm{prox}_{\mathscr{R}}\left(\mathbf{L}\mathbf{c}^k - \frac{\boldsymbol{\alpha}_\mathrm{u}^k}{\rho_\mathrm{u}}; \frac{\lambda}{\rho_\mathrm{u}}\right)$
    $\mathbf{v}^{k+1} = \mathrm{prox}_{i_{\geq 0}}\left(\mathbf{c}^k - \frac{\boldsymbol{\alpha}_\mathrm{v}^k}{\rho_\mathrm{v}}; \frac{1}{\rho_\mathrm{v}}\right)$
    $\mathbf{c}^{k+1} \leftarrow$ solve (4.55) directly using fast $\mathbf{H}^T\mathbf{H}$
    $\boldsymbol{\alpha}_\mathrm{u}^{k+1} = \boldsymbol{\alpha}_\mathrm{u}^k + \rho_\mathrm{u}\left(\mathbf{u}^{k+1} - \mathbf{L}\mathbf{c}^{k+1}\right)$
    $\boldsymbol{\alpha}_\mathrm{v}^{k+1} = \boldsymbol{\alpha}_\mathrm{v}^k + \rho_\mathrm{v}\left(\mathbf{v}^{k+1} - \mathbf{c}^{k+1}\right)$
    $k \leftarrow k+1$
end while
return $\mathbf{c}^{k+1}$



(a) Convergence with respect to time.



(b) Convergence with respect to number of iterations.

Figure 4.16 – Comparisons of the empirical convergence of the objective function (4.42) for three ADMM schemes: the standard ADMM with inner CG loops (resp. 1 and 3) and the proposed inner-loop-free ADMM. The proposed algorithm provides a noticeable gain in algorithmic speed for cryo-EM reconstruction.

noisy projections, multiple angular misestimations). From the ground truth, we compute 30 equidistributed projections using the imaging model in (2.7) with a KBWF as basis function. We then degrade these projections with Gaussian noise, setting their SNR to 1, and we add some error on the projection orientations prior to reconstruction. For the sake of simplicity, we do not consider the effect of the PSF. The reconstruction algorithms are implemented using the GlobalBioIm library [118].

We compare three ADMM configurations: two standard ADMM setups with 1 and 3 inner CG loops, respectively, and the inner-loop-free ADMM. The algorithm parameters (*i.e.*, the regularization parameter λ and the penalty parameter ρ with $\rho_{\mathrm{u}} = \rho_{\mathrm{v}} = \rho$) are tuned empirically to get the best reconstruction with the fastest convergence. We run each algorithm for 200 iterations and monitor the evolution of the cost with respect to the elapsed time (Figure 4.16.a) and the number of iterations (Figure 4.16.b).

The results show that the proposed inner-loop-free ADMM provides a noticeable gain in algorithmic speed compared to the standard approach. As expected, the convergence of the standard ADMM accelerates with the number of CG iterations, but this also severely increases the execution time of the algorithm. The cost per iteration of the proposed ADMM algorithm is lower and permits faster computations.

4.4 Discussion

It is worth recalling that the use of our iterative reconstruction framework in SPA is made entirely possible by the fast formulations of its cost-dominant operations (Section 4.1). Without

these mathematical contributions, the applicability of our framework to this data-intensive modality would be rather limited. Another key feature of the proposed approach is its ability to inject prior information into the reconstruction procedure. Through its regularization term, our scheme is a robust alternative for handling reconstructions for which direct methods fail to yield satisfactory results.

Beyond these contributions, our multiscale scheme can be especially beneficial at the early stages of the iterative-refinement procedure. During these first moments, the optimization problem is highly ill-posed due the almost complete lack of information on the orientations and the underlying 3D structure. Reconstruction algorithms are thus likely to get trapped into abundant local minima if they lack robustness, which then undermines the rest of the iterative-refinement procedure. Indeed, several works have demonstrated that the accuracy of the initial volume estimates can considerably affect the final high-resolution reconstruction [76, 75]. In that sense, our multiscale reconstruction scheme could provide novel ways of stabilizing the early ill-posed stages. A judicious approach could be to start the process with coarse reconstructions that are more robust to errors on the orientations, but yet contain all the necessary information for their further refinement [164, 165]. One could then repeat the process by slowly increasing the scale as the estimation of the orientations is refined.

In the same line of thought, the multiscale scheme could be combined with the joint-optimization framework we introduced in Section 4.2 to improve the iterative-refinement procedure. The idea would be similar here, *i.e.*, to perform first the orientation estimation at coarser scales in order to benefit from the increased robustness to noise and to unknowns. This extension would not only improve the performance of the joint-optimization method, it would also increase its computational attractiveness by reducing its numerical complexity. Another development for our joint-orientation framework could be to add regularization for the latent-variable estimation themselves.

Finally, the aforementioned contributions open the door for other types of developments in SPA, such as the inclusion of novel constraints, a different handling of specimens with symmetries, or the use of promising learning-based approaches [190, 191].

5 Supervised Recovery of Orientations in SPA: Learning from Projections

5.1 Context

[1]In the previous chapters, we focused on model-based methods for cryo-EM that formulate the reconstruction task as an inverse problem and rely on a faithful model of the acquisition physics. Yet, the past years have also seen an explosion of *data-driven* approaches, whose central philosophy is to let the reconstruction algorithms be guided by the measured data through some learning procedure. In this chapter, we present the outline and the preliminary results of an ongoing research project for SPA that capitalizes on the powerful learning capabilities of neural networks, yet still fundamentally relies on our ability to faithfully model the cryo-EM imaging process (for the generation of the training dataset). As we shall shortly detail, its target is the design of a method that learns to estimate the unknown orientation associated to each projection in a SPA dataset. The method is still at its proof-of-concept stage, and several interesting developmental steps lie ahead.

We already discussed that, in SPA, the 3D particles adopt a random orientation in the ice layer before being imaged with parallel electron beams. Hence, the projection geometry associated to each 2D projection is unknown. Yet, this knowledge is essential for tomographic reconstruction. To handle this, a popular approach used by most SPA software packages is to alternatively refine the 3D structure and the orientation estimation. This is also the philosophy behind our novel joint-optimization scheme presented in Section 4.2. Iterative refinement procedures are extremely powerful and have permitted the determination of numerous biological structures up to near-atomic resolution [5]. Unfortunately, the outcome of these methods is predicated on the quality of the initial reconstruction, or, equivalently, on the initial estimation of the orientations [75, 76]. Several methods have been designed to produce a first rough structure [64, 65, 66, 62, 67], but this remains a notoriously arduous challenge in SPA.

[1]This work is linked to a student project that is currently undertaken by Jelena Bancac (MSc Student, EPFL) under the joint supervision of Laurène Donati and Michaël Defferrard (LTS2, EPFL). LD and MD conceived the original idea, devised the project, and planned the experiments. JB carried out most of the implementation and performed the numerical experiments.

5.2 Outline of the Proposed Method

In this work, we present a learning-based method to estimate the unknown orientations in SPA directly from the acquired dataset of projections, *i.e.*, without relying on any intermediate reconstruction procedure.

Our approach relies on the well-known observation that the greater the similarity between two projections, the more likely they originated from two 3D particles that adopted close orientations in the ice layer prior to imaging[2]. This principle guides a number of applications in SPA, including that of projection matching [68].

Taking this line of thought further, we train a function—parametrized as a neural network—to predict the relative orientation between two projections based on their similarity. To make such training possible, we capitalize on our ability to model the cryo-EM imaging procedure to generate a large, representative synthetic dataset using publicly available 3D atomic models.

Using this trained distance function, we can estimate the relative orientations between pairs of projections in any real dataset. Our postulate is that we can then recover, from these estimated relative orientations, the orientations themselves through an appropriate minimization scheme. This two-steps pipeline is illustrated in Figure 5.1.

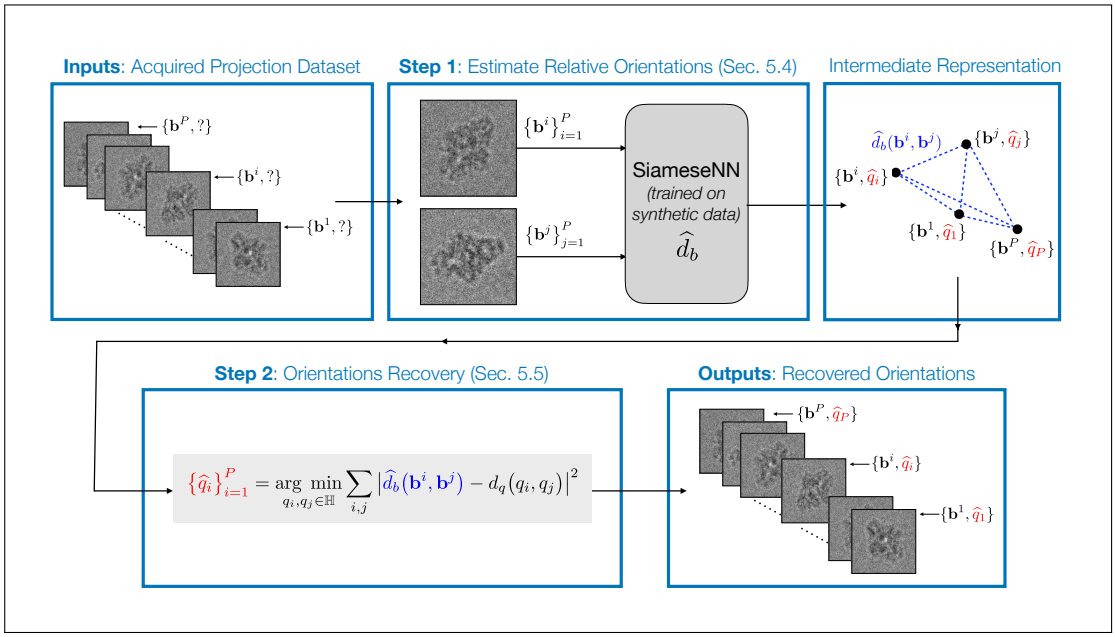


Figure 5.1 – Overview of the proposed two-steps method: 1) estimate the relative orientations between projection pairs through a learned distance $\widehat{d}_b$, and 2) recover the orientations from the estimated relative orientations. As done previously, we denote a pth projection by $\mathbf{b}^p$. In this work, we use the notation q_p to symbolize the orientation of $\mathbf{b}^p$, for reasons that we shall further detail in Section 5.3. The geodesic distance between two orientations is denoted by d_q.

[2] Up to some possible intrinsic symmetries of the objects, which we discuss later.

The task of recovering points based on their relative distances has been extensively studied in the literature, mostly within the framework of dimensionality reduction and primarily for the case of *Euclidean* embedding spaces[3] [192, 193, 194, 195, 196] . In that respect, the short example given by Dokmanic *et al.* in [196] efficiently illustrates the philosophy behind such methods (see the "An Analogy" box).

An Analogy: Mapping the Position of Swiss Cities with a Train Timetable [196]

In this toy problem, the authors aim at estimating the position of five cities on the Swiss map based not on the spatial distances between them, but on the time it takes to travel by train between them. Those time data (in minutes) are collected in the following timetable:

	L	G	Z	N	B
Lausanne	0	33	128	40	66
Geneva	33	0	158	64	101
Zürich	128	158	0	88	56
Neuchâtel	40	64	88	0	34
Bern	66	101	56	34	0

Remarkably, even though these time data only roughly correlate with the physical distances between the cities, one can still obtain a remarkably good estimate of their positions on the Swiss map (up to some symmetries of the embedding space) using a multidimensional scaling algorithm.

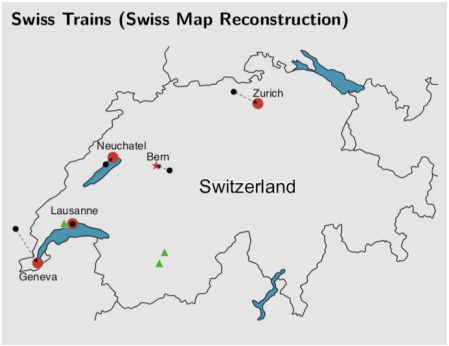


Figure 5.2 – Image adapted from [196]. The red signs indicate the correct city locations. The black dots denote the recovered city locations. The green triangles indicate the places where the present thesis has been written.

This example, if rather simple, nevertheless underlines well the key ingredients of methods that aim at retrieving points from distances that may not be directly measurable:

1. *An appropriate proxy for the "real" distance.* In the above example, the proxy for the spatial distance between two cities is the time taken to travel by train between them. In our case, we shall consider the similarity between two projections to be a good proxy for their relative orientation.

2. *A sufficiently rich collection of proxy distance data.* In this example, these data are provided by the (complete) train timetable. In our approach, we shall estimate the relative orientations between numerous pairs of projections based on the aforementioned proxy distance.

3. *An efficient recovery scheme.* In [196], the embedding space being Euclidean, the theo-

[3] An "embedding space" corresponds to the lower-dimensional space in which data is embedded, *i.e.*, mapped to in such a way that the relative distances between its points are preserved as much as possible.

retical framework of the Euclidean distance matrices (EDMs) guarantees that one can retrieve the desired points from the collected distances. In our case, as we shall shortly explain, we aim to embed the estimated relative orientations on $\mathbf{SO}(3)$, the space of 3D rotations. Unfortunately, the extension of the EDM theory to such manifold is all but straightforward.

There is no straightforward way to "handcraft" a proxy distance that would robustly predict the similarity between two projections. Hence, we resort to *learning* this distance function by parametrizing it as a neural network and capitalizing on 1) the public availability of large datasets of 3D atomic models[4], and 2) our ability to model the cryo-EM imaging process (Chapter 2). This is the topic of Section 5.4.

Equipped with this learned distance, the idea is then to apply the aforementioned two-steps method (see Figure 5.1) for any projection dataset. This is detailed in Section 5.5. As we just mentioned, we cannot rely on the theoretical framework of EDMs since our embedding space is non-Euclidean. Despite this lack of theoretical guarantees, we are able to appropriately minimize our objective function using a gradient-based algorithm, as we experimentally demonstrate in Section 5.4.

As a preamble, we discuss the need in this work for an alternative representation of orientations in $\mathbf{SO}(3)$ that relies on unit quaternions.

5.3 Unit Quaternions and the Geodesic Distance

As mentioned, our objective is to recover unknown 3D orientations by embedding their estimated relative distances on the $\mathbf{SO}(3)$ space. As we shall explain in the next sections, this embedding requires the efficient computation of the relative distance between two rotations $\mathbf{R}_1, \mathbf{R}_2 \in \mathbf{SO}(3)$, which corresponds to the rotation $\mathbf{R}_* \in \mathbf{SO}(3)$ such that $\mathbf{R}_1 = \mathbf{R}_* \mathbf{R}_2$.

We have been working so far with Euler angles to describe the orientation of a 3D object in the electron microscope, as is standard in SPA. In other words, we relied on the Euler angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \Omega_{\boldsymbol{\theta}}$ to encode the 3D rotation that relates the object coordinate system to the projection coordinate system (see Chapter 1). Unfortunately, the relative distance between two rotations $\mathbf{R}(\boldsymbol{\theta}_1)$, $\mathbf{R}(\boldsymbol{\theta}_2)$, parametrized by Euler angles cannot be directly computed from $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, but requires the computation of the rotation matrices, which is computationally inefficient[5]. Hence, we resort to a more convenient representation of 3D rotations that relies on unit quaternions.

The algebra of quaternions was introduced in the mid-nineteenth century by Hamil-

[4] `https://www.ebi.ac.uk/pdbe/emdb`

[5] Another technical challenge with Euler angles is that they suffer from the so-called gimbal lock problem, which arises when $\theta_2 = 0$ and restricts the number of rotational degrees of freedom to one even though θ_1 and θ_3 have not yet been fixed [197].

ton [198]. A quaternion $q \in \mathbb{H}$ takes the form

$$q = a\mathbf{1} + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}, \tag{5.1}$$

where $(a,b,c,d) \in \mathbb{R}^4$, and $\mathbf{1}$, $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ are the fundamental quaternion units

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{i} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \boldsymbol{j} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \boldsymbol{k} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \tag{5.2}$$

with i the imaginary unit. Any quaternion q can thus be represented by its set of coefficients $(a,b,c,d) \in \mathbb{R}^4$. The algebra $\mathbb{H}$ is similar to the algebra of complex numbers $\mathbb{C}$, with the exception of the multiplication operation being noncommutative.

In this work, we restrict our interest to unit quaternions $q \in \mathbb{U}$, with $\mathbb{U} = \{q \in \mathbb{H} \mid |q| = 1\}$, which identify the $\mathbb{S}^3$ hypersphere in $\mathbb{R}^4$. Unit quaternions concisely and elegantly represent the elements of the $\mathbf{SO}(3)$ group. More precisely, a unit quaternion $q \in \mathbb{U}$ parametrizes a rotation $\mathbf{R} \in \mathbf{SO}(3)$ through

$$\mathbf{R}(q) = \begin{pmatrix} a^2+b^2-c^2-d^2 & 2bc-ad & 2bd+2ac \\ 2bc+2ad & a^2-b^2+c^2d^2 & 2cd-2ab \\ 2bd-2ac & 2cd+2ab & a^2-b^2-c^2+d^2 \end{pmatrix}. \tag{5.3}$$

The geodesic distance $d_q : \mathbb{U} \times \mathbb{U} \rightarrow [0,\pi]$ between two unit quaternions $q_i, q_j \in \mathbb{H}$ is then defined as

$$d_q(q_i,q_j) = 2\arccos\left(|\langle q_i, q_j \rangle|\right), \tag{5.4}$$

with the inner product between quaternions given by

$$\langle q_i, q_j \rangle = a_i a_j + b_i b_j + c_i c_j + d_i d_j. \tag{5.5}$$

The distance (5.4) is the shortest distance between q_i and q_j on the surface of $\mathbb{S}^3$. As $\mathbb{S}^3$ is isomorphic to the universal cover of $\mathbf{SO}(3)$, this distance corresponds to the magnitude of the relative orientation $\mathbf{R}_*$ between $\mathbf{R}(q_i)$ and $\mathbf{R}(q_j)$ in $\mathbf{SO}(3)$ [199]. In other words, the relative distance between two rotations encoded by unit quaternions can be efficiently computed from the unit quaternions themselves through (5.4), which is of key practical importance for this work.

For the sake of conciseness, we shall use the term "with orientation q" to refer to 2D/3D objects considered in an imaging geometry parametrized by q, similar to what we previously did with Euler angles.

5.4 Estimating Relative Orientations from Projections

Equipped with the geodesic distance d_q, our goal is now to find a "projection distance" d_b that is a good predictor of d_q. Before discussing the different options, we briefly describe the synthetic datasets used in this work.

5.4.1 Experimental Dataset

We consider two proteins as ground truths: the β-galactosidase, a protein with a dihedral (D2) symmetry, and the lambda excision HJ intermediate (HJI), an asymmetric protein. Their deposited PDB atomic models are 5a5a [50] and 5j0n [186], respectively. For each atomic model, we generate the ground truth by fitting a 5Å density map in Chimera [200]. We thus obtain a volume of size $(117 \times 117 \times 117)$ for the β-galactosidase, and a volume of size $(275 \times 275 \times 275)$ for the HJI.

From these ground truths, we generate $5,000$ synthetic projections of size (117×117) and (275×275), respectively, using the ASTRA projector[6] [201]. The projection orientations are sampled from a uniform distribution over half the $\mathbf{SO}(3)$ space, which suffices to generate all the possible projections of a volume. For the sake of simplicity, the projections are currently kept unblurred and noiseless. Whenever training neural networks, we split the datasets into a distinct training set (50%), validation set (22%), and testing set (33%), to ensure that the results can generalize to unseen projections. The complete pipeline is implemented in Tensorflow [202].

5.4.2 Baseline Test with the Euclidean Distance

As a baseline, we first evaluate the suitability of the Euclidean distance as a projection distance d_b to predict d_q. For the two aforementioned datasets, we randomly select $1,000$ pairs of projections. For each pair, we compute the Euclidean distance between the projections $d_b(\mathbf{b}^i, \mathbf{b}^j) = \|\mathbf{b}^i - \mathbf{b}^j\|_2$ and their relative orientation $d_p(q_i, q_j)$ through (5.4). We then report the (d_q, d_b) relationship for all pairs in Figure 5.3, for both the asymmetric protein (left) and the symmetric one (right).

Two principal observations can be made from this experiment. First, as suspected, the Euclidean distance between projections fails to be a consistent predictor of their relative orientation distance, even in the simple imaging conditions considered here (no noise and no effect of the PSF). In particular, the larger the relative distance d_q, the poorer the predictive ability of the Euclidean distance as d_b. The other interesting observation is that the intrinsic symmetry of the β-galactosidase protein (5a1a) appears in its (d_q, d_b) plot.

[6] The open-source ASTRA toolbox implements an accelerated projector for 3D tomography. The code is provided in Python, which is convenient at the current stage of development.

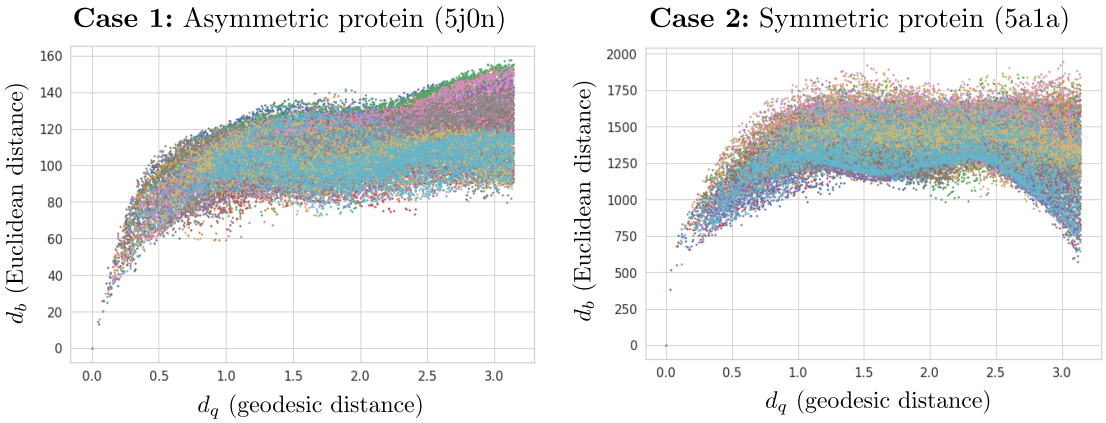


Figure 5.3 – Plotting the Euclidean distance between two projections versus their actual relative orientation (measured by the geodesic distance) for **(left)** the asymmetric protein (5j0n) dataset, and **(right)** the symmetric protein (5a1a) dataset.

5.4.3 Learning $\widehat{d}_b$ with a Siamese Neural Network

As previously discussed, we make the choice to *learn* a good approximation $\widehat{d}$ on a synthetic training dataset $\{\mathbf{b}^{*p}, q_p^*\}_{p=1}^{N_t}$ through

$$\widehat{d}_b = \underset{d_b}{\arg\min} \sum_{i,j} \left| d_b(\mathbf{b}^{*i}, \mathbf{b}^{*j}) - d_q(q_i^*, q_j^*) \right|^2, \tag{5.6}$$

with d_q defined in (5.4), and where N_t indicates the number of projection-orientation pairs in the training dataset. More precisely, we parametrize the distance function d_b in (5.6) as a Siamese neural network (SiameseNN) [203], and resort to learning its weights w, as illustrated in Figure 5.4.

SiameseNNs, also termed "twin networks", are commonly used in the field of deep metric learning to learn similarity functions [204]. They are usually constituted of two sister neural networks that work in tandem and share the exact same architecture and weights. Their role (once trained) is to extract the projection features that are the most relevant to predict the relative orientation between two projections. The weights w of the two sister networks are progressively learned by 1) comparing the difference of their projection feature vectors to the magnitude of the corresponding relative orientations, and 2) back-propagating this error (via the derivative chain rule) to the weights.

Generating a Proper Training Dataset for the SiameseNN

The success of the SiameseNN as a faithful predictor of relative orientations eventually relies on our capacity to generate a synthetic training dataset that is both large and representative of SPA measurements. In other words, we need to create a training set whose data distribution is diverse enough to cover that of unseen projection datasets. The objective is for the SiameseNN

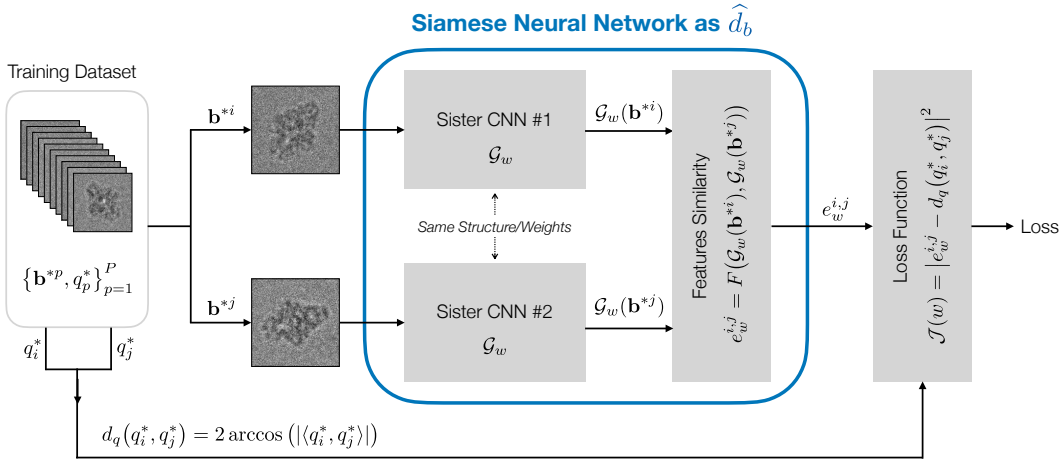


Figure 5.4 – Training a Siamese neural network (SiameseNN) to become a faithful predictor of the relative orientation between two input projections. In other words, we train the SiameseNN to serve as a projection distance $\widehat{d}_b$ that correctly approximates the orientation distance d_q. The training is performed with a synthetic dataset that contains thousands of projections with their associated orientation.

to be able to handle projections acquired in all sorts of imaging conditions and originating from 3D volumes it has never been trained on.

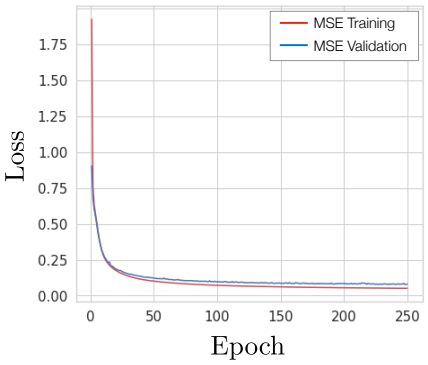
We shall create such comprehensive training dataset by capitalizing on two favorable conditions. First, there exists a large publicly-available database of deposited atomic models of proteins, which gives us access to thousands of different 3D ground truths. Then, we shall take advantage of our ability to model the cryo-EM imaging procedure (Chapter 2) to generate, from these ground truths, a synthetic dataset that contains a massive amount of realistic projections whose orientations are, by definition, all known.

Note that an interesting aspect of SiameseNNs for the present application is that they intrinsically predict the *relationship* between objects. Hence, a well-trained SiameseNN could be relatively robust to the change of volumes. In the same line of thought, our SiameseNN will likely benefit from the profound structural similarity shared by proteins—after all, they all derived from just the same 21 amino acids.

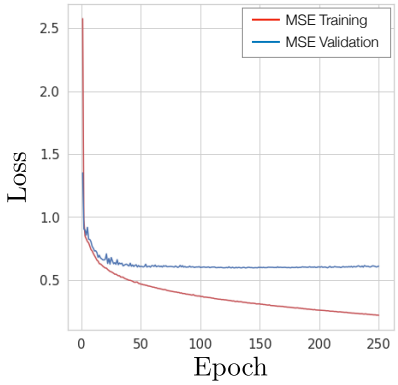
Preliminary Training Results

We present here a preliminary evaluation of the ability of SiameseNNs to learn a projection distance $\widehat{d}_b$ that correctly approximates the orientation distance d_q.

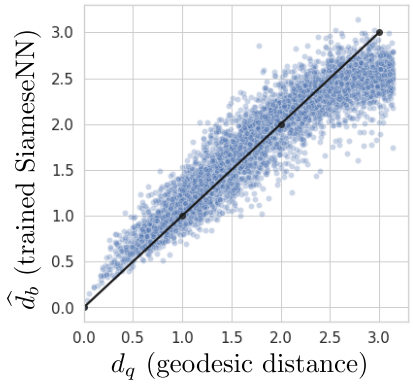
SiameseNNs come with a variety of more or less powerful architectures. At the current stage of development, we work with a simple one. Our SiameseNN is composed of two convolutional neural networks (CNNs) with shared weights. Their output features vectors



(a) Training losses of the SiameseNN on the asymmetric protein (5j0n) training and validation datasets.



(b) Training losses of the SiameseNN on the symmetric protein (5a1a) training and validation datasets.



(c) Relative orientations predicted by the trained SiameseNN from projections in the asymmetric protein (5j0n) testing dataset.

Figure 5.5 – Training results for the SiameseNN.

are compared through an Eulidean distance, *i.e.*, $F(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2$ in Figure 5.4. The detailed architecture of this SiameseNN is given in Figure A.1 in Appendix A.5.

For each protein, we train the SiameseNN on its training dataset for 250 epochs (~10 hours) using an Adam optimizer [205], a learning rate of 10^{-3}, and a batch size of 256 projections. The evolution of the training and validation losses are presented in Figure 5.5a for the asymmetric protein (5j0n), and in Figure 5.5b for the symmetric one (5a1a). The results demonstrate that the SiameseNN succeeds at learning a proxy distance for the asymmetric protein dataset, as convergence is reached in about 50 epochs (~ 2 hours).

However, the current SiameseNN architecture fails at learning the distance for the dataset 5a1a, which is very likely due to the symmetry of the β-galactosidase protein. Indeed, its

synthetic dataset contains pairs of projections that share the same d_b, yet differ in their d_q. This simply advocates for the restriction to non-overlapping areas on $\mathbf{SO}(3)$ when sampling the orientations used to generate the SiameseNN training dataset. The latter would then only contain projection pairs with a linear (d_q, d_b) relationship, which should ensure a successful training of the network. For the rest of the experiments, we use the asymmetric protein (5j0n) dataset.

We then feed to the trained SiameseNN $1,000$ pairs of projections randomly selected from the 5j0n testing dataset, and report the $(d_q, \widehat{d}_b)$ relationship of each pair in Figure 5.5c. These results confirm that, for this protein at least, the SiameseNN is able to predict the orientation distance d_q using only the projections as inputs. Moreover, it clearly outperforms the Euclidean distance at doing so. These preliminary results are encouraging, as much has yet to be gained from improving upon the rather primitive SiameseNN architecture we currently use.

5.5 Orientation Recovery

Equipped with an appropriately learned $\widehat{d}_b$, the objective is then to recover the unknown unit quaternions $\{q_p\}_{p=1}^P$ associated to the projections $\{\mathbf{b}^p\}_{p=1}^P$ in any given dataset.

5.5.1 Minimization Scheme

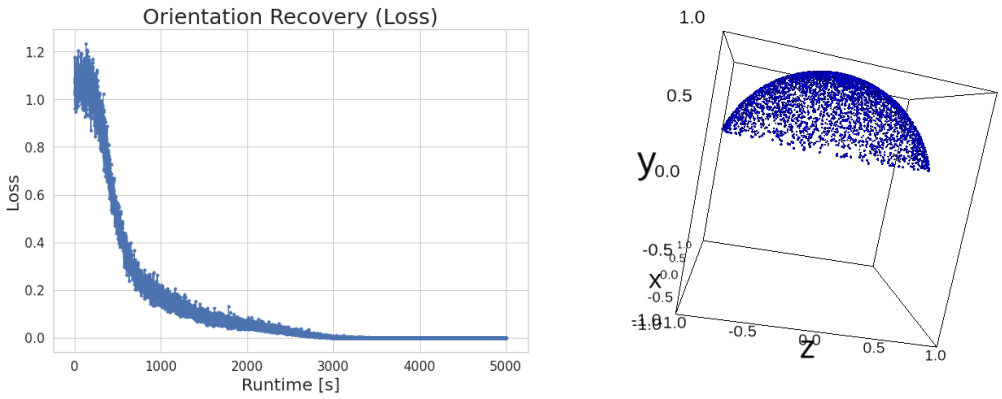
We propose to start this process by computing of a great number of pairwise projection distances $\{\widehat{d}_b(\mathbf{b}^i, \mathbf{b}^j)\}_{i,j=1}^P$ through $\widehat{d}_b$. Then, our postulate is that we can recover the orientations from theses distances by solving

$$\{\widehat{q}_p\}_{p=1}^P = \underset{q_i \in \mathbb{U}}{\arg\min} \sum_{i,j} \left| \widehat{d}_b(\mathbf{b}^i, \mathbf{b}^j) - d_q(q_i, q_j) \right|^2, \tag{5.7}$$

as is illustrated in Figure 5.1.

In practice, one cannot possibly evaluate (5.7) for every pair of orientations $\{q_i, q_j\}_{i,j=1}^P$ given the extremely large size of SPA datasets, with P typically in the order of dozens of thousands. Hence, we need to partially sample the projection dataset. We experimentally demonstrate in Section 5.5.2 that this does not affect the performance of our recovery scheme.

As previously discussed, we are not yet aware of any guarantee of convergence for (5.7). Similarly, we do not know of any theoretical characterization of the behaviour of (5.7) in ill-posed conditions, such as when pairwise distances are misestimated, for instance. Hence, we rely for now on experimental demonstrations to 1) ensure feasibility, and 2) indicate where efforts need to be invested.



(a) (b)

Figure 5.6 – Results of the orientation recovery scheme when using the perfect orientation distances for the asymmetric protein (5j0n). **(a)** Evolution of the loss of (5.7) during minimization. **(b)** Coverage of $\mathbf{SO}(3)$ after the orientation recovery from the perfect relative distances.

5.5.2 Feasibility Check: Recovery from the Exact Relative Distances

Our first investigation is to verify that it is at all possible to recover the orientations through (5.7) from their true relative distances (*i.e.*, using d_q and not a proxy d_b).

We use the 5,000 projections from the asymmetric protein (5j0n) dataset. Out of the possible 25 mio possible pairs, we randomly select only 5,000 of them and compute their geodesic distance through (5.4). We then minimize (5.7) using the SGD Adam optimizer [205] for $30K$ steps (~1 hour) with a learning rate of 0.1.

The results are given in Figure 5.6. They confirm that it is possible to recover the orientations from their true relative distances, even though the embedding space is non-Euclidean. As previously discussed, this is not a straightforward result. The results also demonstrate that a large subsampling of the projection pairs does not affect the convergence of (5.7), which is in straight line with the observations made by numerous Euclidean-based dimensionality reduction works [192, 193, 194, 195].

5.5.3 Robustness of Recovery to (Additive) Errors on the Relative Distances

We now go one step further and evaluate the behaviour of (5.7) when the true relative distances are corrupted by additive Gaussian noise.

The experimental conditions are the same than in the previous section, except that we add an error with increasing variance on the relative distances prior to the minimization. The results are presented in Figure 5.7 (red curve).

Before discussing the results, we remark that one cannot really quantify the performance

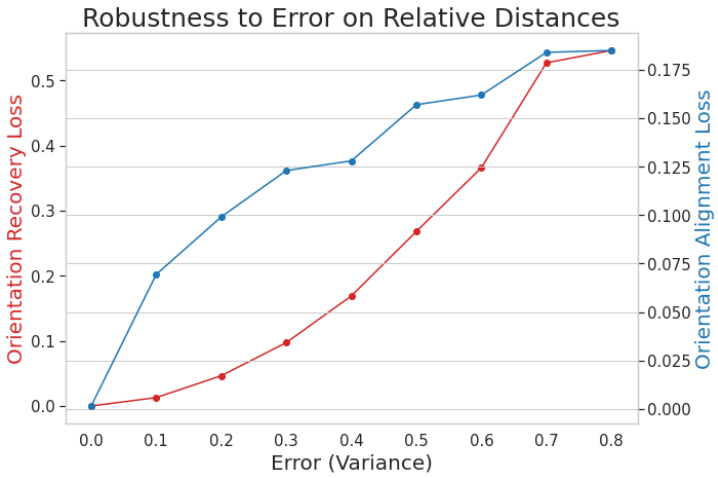


Figure 5.7 – Results of the recovery scheme (red curve) and the alignment procedure (blue curve) when an increasing amount of errors is added to the true relative distances.

of (5.7) through its loss. Unfortunately, it is also not appropriate to directly compute the error between the recovered orientations $\{\widehat{q}_p\}_{p=1}^P$ and the true ones $\{q_p\}_{p=1}^P$. The reason is that the recovery of orientation through (5.7) is up to a global rotation, *i.e.*, any global rotation of the set of recovered orientations is as valid as any other. This is not a problem for the ultimate application of our scheme, but it complicates the quantitative evaluation of its performance in synthetic experiments. We circumvent this problem by 1) aligning the true and recovered orientation sets, and 2) computing their distance after alignment. The alignment is performed by searching for the orthogonal matrix (with determinant ± 1) $\mathbf{T} \in \mathbb{R}^4$ that minimizes

$$\widehat{\mathbf{T}} = \underset{\mathbf{T}\in\mathbb{R}^4}{\arg\min} \sum_{i,j} \left| d_q(q_i, q_j) - d_q(\mathbf{T}\widehat{q}_i, \mathbf{T}\widehat{q}_j) \right|^2. \tag{5.8}$$

For all variances, the distance after alignment is reported in Figure 5.7 (blue curve).

These results demonstrate that the performance of the minimization scheme (5.7) linearly depends on the quality of the relative distances, which advocates for a proper and extensive training of the SiameseNN in the next stages of development. Another interesting output of Figure 5.7 is that it indicates that the error of the orientation recovery behaves as a monotonic function of its loss. Hence, it suggests that the loss can be used as a good indicator of its performance, which has obvious practical implications for our future works on real data.

5.6 Discussion

The results obtained so far provide some key insights on the viability of the proposed method. They guarantee that each component of this new paradigm can handle the task it is designed for, and they provide some early indications on the robustness of the overall scheme. Unfortunately, we have not yet had the time to run an end-to-end pipeline and evaluate the

performance of (5.7) when the trained SiameseNN is used as learned $\widehat{d}_b$; this is very clearly the next line of research.

The obtained results underline the importance of learning an accurate proxy distance $\widehat{d}_b$. In this regard, we could improve the performance of the SiameseNN in several ways, for instance by further tuning the architecture of its twin CNNs. We could also parametrize the function F, which compares the similarity of the features outputs (see Figure 5.4), as a feed-forward neural network instead of the current Euclidean distance, and learn its weights as well.

Once these improvements are made, we shall enhance the training dataset for the SiameseNN and test its predictive ability in more challenging situations. To achieve this, we shall rely on our powerful, expressive forward model of the cryo-EM procedure to generate realistic projections from thousands of atomic models in the PDB database. This will also allow us to test different options for the challenging handling of proteins with multiple conformational states.

6 A New Reconstruction Paradigm for SPA via Deep Adversarial Learning

6.1 Context

[1]We extensively discussed in all that precedes that a major challenge in SPA is that the imaged 3D particles have unknown orientations. Current reconstruction techniques either estimate the orientations through an iterative-refinement procedure [68] or marginalize them in likelihood-based optimization procedures [10]. Both are computationally challenging approaches that can demand very large resources or rely on approximations that may limit the attainable resolution (see Section 1.4).

In Chapter 4, we introduced a refinement-based approach that brings increased robustness and does away with the computationally expensive procedure of projection matching. Yet, like all iterative-refinement approaches, its outcome depends on the quality of the initial reconstruction. In this regard, the learning-based method presented in Chapter 5—whose target is to estimate the orientations directly from the acquired dataset of projections—could prove highly useful. However, the method is still at its proof-of-concept stage and several key developmental steps lie ahead.

In this chapter, we introduce CryoGAN, a completely new paradigm for SPA reconstruction based on deep adversarial learning. The CryoGAN algorithm can resolve a 3D structure in a single algorithmic run using only the dataset of projections and CTF estimations. Moreover, it does not rely on a first low-resolution volume estimate and can be initialized with a zero-valued volume. While the spatial resolution of the CryoGAN reconstructions from real data is not yet competitive with the state-of-the-art, we expect these results to improve in the near future. In particular, we believe that the ongoing progresses in deep-learning architectures will help enrich the CryoGAN algorithm and improve the resolution of its reconstructions.

[1]This chapter uses content from a recent preprint: Harshit Gupta, Michael T. McCann, **Laurène Donati**, Michael Unser. *CryoGAN: A New Reconstruction Paradigm for Single-particle Cryo-EM Via Deep Adversarial Learning*. `https://www.biorxiv.org/content/10.1101/2020.03.20.001016v1`. HG and MM conceived the original idea and implemented the pipeline. HG carried out most of the experiments. LD provided expertise on cryo-EM related aspects. All authors were involved in the planning of the experiments and discussed the results. LD took the lead in writing, with inputs from all authors. MU supervised the project.

In the meantime, the preliminary results obtained with CryoGAN are encouraging and demonstrate the potential of adversarial-learning schemes in image reconstruction. As far as we know, CryoGAN is the first demonstration of a deep-learning architecture able to perform the full SPA reconstruction procedure without any prior training. The framework is a prime illustration of the immense gain that can derive from combining the robustness of model-based approaches with the powerful abilities of learning-based schemes.

6.2 Overview: The CryoGAN Paradigm

CryoGAN is an unsupervised reconstruction algorithm for SPA that exploits the remarkable ability of generative adversarial networks (GANs) to capture data distributions [206]. Similar to GANs, CryoGAN is driven by the competitive training of two entities: one that captures the distribution of real data, and another that discriminates between generated samples and samples from the real dataset (Figure 6.1a). In a classical GAN, the two entities are each a convolutional neural network (CNN). They are known as the generator and the discriminator and are trained simultaneously using backpropagation.

The important twist with CryoGAN is that we replace the generator network by a cryo-EM physics simulator (Figure 6.1b). This simulator implements a mathematical model of the imaging procedure to produce a synthetic measurement based on 1) the current volume estimate and 2) a random projection orientation. As we have done so far, this image-formation model considers that the cryo-EM 2D measurement is the projection of the volume at that orientation, modulated by microscopy-related effects and corrupted by substantial additive noise.

The cryo-EM physics simulator is then paired with a discriminator network whose architecture is similar to that of standard GANs. The role of the discriminator in CryoGAN is to encourage the simulator to learn the volume whose simulated dataset distribution matches that of the real dataset, while it simultaneously gets better at evaluating the simulated projections for authenticity. This novel adversarial-learning scheme allows CryoGAN to output the 3D structure whose simulated projections are the most consistent, in a distributional sense, with the real projections in a 2D measurement set.

The CryoGAN architecture represents a complete change of paradigm for SPA reconstruction. No estimation of the orientations is attempted during the learning procedure; rather, the reconstruction is obtained through a distributional matching performed in a likelihood-free manner. Due to this innovative setting, CryoGAN sidesteps many cumbersome processing stages such as 2D alignment and 2D/3D classification. It also avoids many of the computational drawbacks associated with likelihood-based methods, in particular the need to marginalize over all orientations via numerical integration.

In practice, CryoGAN requires no prior knowledge of the 3D structure; its learning process is purely unsupervised and data-driven. The user needs only to feed the projections and

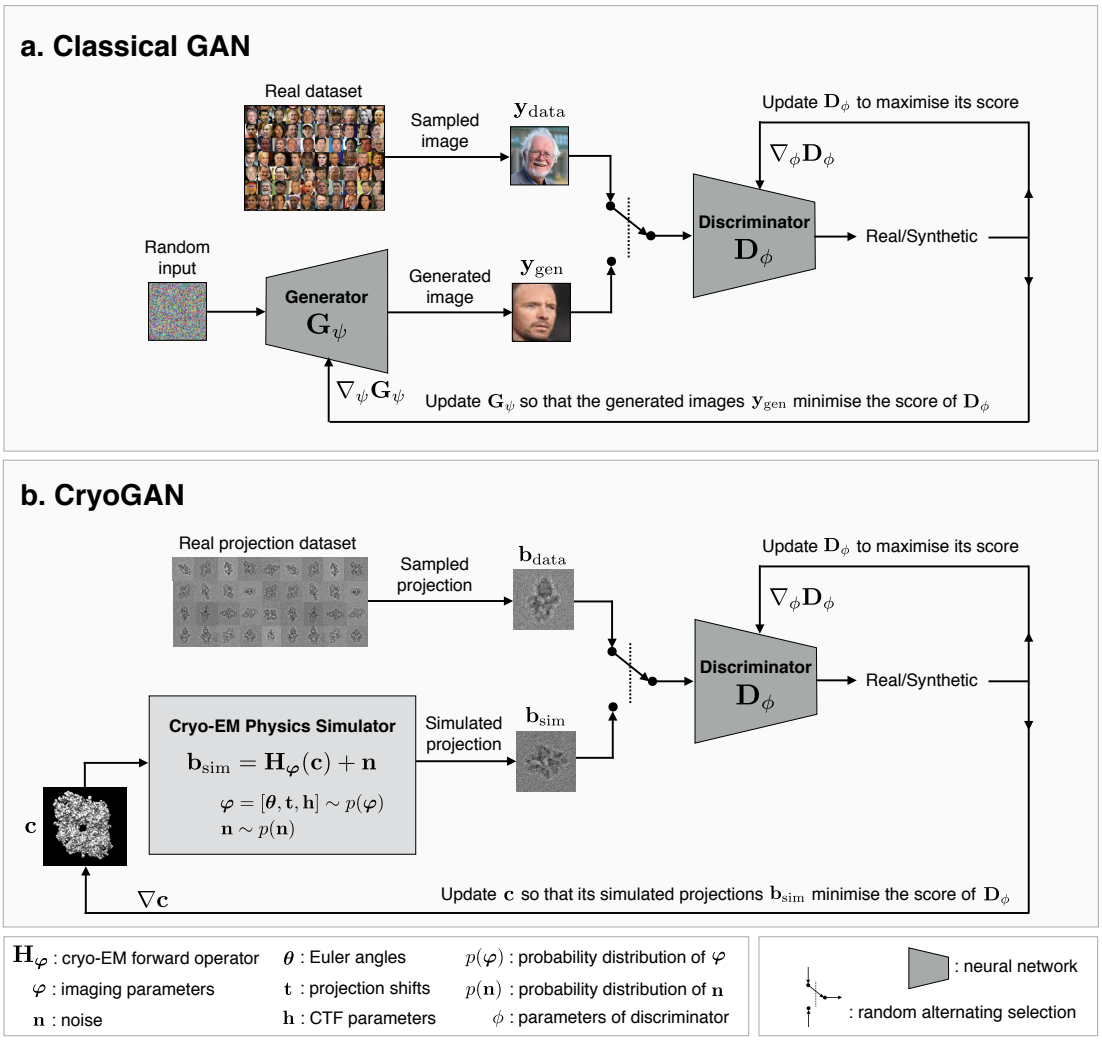


Figure 6.1 – A schematic comparison between **(a)** a classical GAN architecture and **(b)** the CryoGAN architecture. Both frameworks rely on a deep adversarial learning scheme to capture the distribution of real data. CryoGAN exploits this ability to look for the volume whose simulated measurements have a distribution that matches this real data distribution. This is achieved by adding a "cryo-EM physics simulator" that produces synthetic measurements following a mathematical model of the cryo-EM imaging procedure. Importantly, CryoGAN does not rely on a first low-resolution volume estimate, but is initialized with a zero-valued volume. Note that, for both architectures, the updates involve backpropagating through the neural networks; those actions are not indicated here for the sake of clarity.

estimates of the contrast transfer function (CTF). No initial estimate of the volume is needed: the algorithm starts with a volume initialized with zeros.

As we shall now detail, the CryoGAN algorithm is based on a sound mathematical framework that provides guarantees on the recovery of the volume under a given set of assumptions that are often met in practice, at least to some degree of approximation. Its adversarial learning scheme falls under the framework of Wasserstein GANs (WGANs), with some key architectural differences. The CryoGAN algorithm is implemented in PyTorch [207], and mostly relies on stochastic gradient descents (SGD) to alternate between the updates of the discriminator and of the volume.

6.3 Mathematical Framework of CryoGAN

Similar to what we have done so far[2], we model the imaging procedure in SPA through the linear relationship

$$\mathbf{b} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{c} + \mathbf{n}, \tag{6.1}$$

where $\mathbf{b} \in \mathbb{R}^M$ is a 2D projection, $\mathbf{c} \in \mathbb{R}^N$ corresponds the 3D volume of interest, $\mathbf{H}_{\boldsymbol{\varphi}} \in \mathbb{R}^{M\times N}$ is the forward operator with imaging parameters $\boldsymbol{\varphi}$, and $\mathbf{n} \in \mathbb{R}^M$ represents an additive noise that follows a distribution $p_{\mathbf{n}}$.

The set of imaging parameters is denoted as $\boldsymbol{\varphi} = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast}})$. It includes the Euler angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \Omega_{\boldsymbol{\theta}}$, the projection shifts $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$, and the CTF parameters $\mathbf{h} = (d_1, d_2, \alpha_{\text{ast}}) \in \mathbb{R}^3$, where d_1 is the defocus-major, d_2 is the defocus-minor, and α_{ast} is the angle of astigmatism. We then model the forward operator $\mathbf{H}_{\boldsymbol{\varphi}}$ as

$$\mathbf{H}_{\boldsymbol{\varphi}} = \mathbf{C}_{\mathbf{h}}\mathbf{S}_{\mathbf{t}}\mathbf{P}_{\theta}, \tag{6.2}$$

where $\mathbf{P}_{\boldsymbol{\theta}} : \mathbb{R}^N \to \mathbb{R}^M$ is the projection operator, $\mathbf{S}_{\mathbf{t}} : \mathbb{R}^M \to \mathbb{R}^M$ is the shift operator, and $\mathbf{C}_{\mathbf{h}} : \mathbb{R}^M \to \mathbb{R}^M$ corresponds to the convolution operator with the PSF, which, we recall, is the inverse Fourier transform of the CTF with parameters $\mathbf{h}$.

6.3.1 The Quest for Distributional Matching

The ultimate goal of SPA reconstruction is to estimate a 3D structure $\mathbf{c}_{\text{rec}}$ whose projections are consistent with the observed projections (*i.e.*, the acquired data) of the true structure $\mathbf{c}_{\text{true}}$.

We can write the conditional probability density function of a measurement $\mathbf{b}$ given a

[2] The only noteworthy change of notation in this chapter is the use of $\mathbf{b}$ to denote a single projection rather than a stack of projections, which helps alleviate the subsequent notations.

volume $\mathbf{c}$ by marginalizing over the imaging parameters

$$p(\mathbf{b}|\mathbf{c}) = \int_{\boldsymbol{\varphi}} p_{\mathbf{n}}(\mathbf{b} - \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{c})\, p_{\boldsymbol{\varphi}}(\boldsymbol{\varphi}) \mathrm{d}\boldsymbol{\varphi}, \tag{6.3}$$

where $p_{\boldsymbol{\varphi}}$ is the distribution of the imaging parameters $\boldsymbol{\varphi}$. We denote $\mathbf{b}_{\text{noiseless}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{c}$.

In our formulations, the projections in the real dataset are samples of a distribution p_{data}. We then make the assumption that the distribution $p(\mathbf{b}|\mathbf{c}_{\text{true}})$ corresponds to the distribution $p(\mathbf{b}|\mathbf{c}_{\text{true}}) = p_{\text{data}}(\mathbf{b})$ of the real dataset, which is reasonable if the image-formation model faithfully mimics the cryo-EM physics.

We demonstrate in Theorem 4 in Appendix A.6.2 that two 3D volumes $\mathbf{c}_1$ and $\mathbf{c}_2$ have identical conditional distributions $p(\mathbf{b}|\mathbf{c}_1) = p(\mathbf{b}|\mathbf{c}_2)$ if and only if $\mathbf{c}_1$ is equal to $\mathbf{c}_2$, up to rotation and reflection. Hence, Theorem 4 implies that, for the reconstruction $\mathbf{c}_{\text{rec}}$ to satisfy $\mathbf{c}_{\text{rec}} = \mathbf{c}_{\text{true}}$, it must also satisfy $p(\mathbf{b}|\mathbf{c}_{\text{rec}}) = p(\mathbf{b}|\mathbf{c}_{\text{true}})$.

This is a mathematical result of importance as it means we can formulate the reconstruction task as the minimization problem

$$\begin{aligned} \mathbf{c}_{\text{rec}} &= \arg\min_{\mathbf{c}} D\big(p(\mathbf{b}|\mathbf{c}), p(\mathbf{b}|\mathbf{c}_{\text{true}})\big) \\ &= \arg\min_{\mathbf{c}} D\big(p_{\mathbf{c}}(\mathbf{b}), p_{\text{data}}(\mathbf{b})\big), \end{aligned} \tag{6.4}$$

where D is some distance between two distributions. In essence, (6.4) states that the appropriate reconstruction is the 3D structure whose theoretical projection set is the most similar to the real data set *in a distributional sense*. For the sake of conciseness, we shall henceforth use the notation $p(\mathbf{b}|\mathbf{c}) = p_{\mathbf{c}}(\mathbf{b})$.

We use for (6.4) the Wasserstein distance defined as

$$D(p_1, p_2) = \inf_{\gamma \in \Pi(p_1, p_2)} \mathbb{E}_{(\mathbf{b}_1, \mathbf{b}_2)\sim\gamma}[\|\mathbf{b}_1 - \mathbf{b}_2\|], \tag{6.5}$$

where $\Pi(p_1, p_2)$ is the set of all the joint distributions $\gamma(\mathbf{b}_1, \mathbf{b}_2)$ whose marginals are p_1 and p_2, respectively. Our choice is driven by works demonstrating that the Wasserstein distance is more stable than other popular distances (*e.g.*, total-variation or Kullback-Leibler) for this kind of applications [208].

Using (6.5), the minimization problem (6.4) expands as

$$\mathbf{c}_{\text{rec}} = \arg\min_{\mathbf{c}} \inf_{\gamma \in \Pi(p_{\mathbf{c}}, p_{\text{data}})} \mathbb{E}_{(\mathbf{b}_1, \mathbf{b}_2)\sim\gamma}[\|\mathbf{b}_1 - \mathbf{b}_2\|]. \tag{6.6}$$

By using the formalism of [208, 209, 210], this minimization problem can also be stated in its

dual form

$$\mathbf{c}_{\text{rec}} = \underset{\mathbf{c}}{\operatorname{argmin}} \max_{f:\|f\|_L \leq 1} \Big(\mathbb{E}_{\mathbf{b}\sim p_{\text{data}}}[f(\mathbf{b})] - \mathbb{E}_{\mathbf{b}\sim p_{\mathbf{c}}}[f(\mathbf{b})] \Big), \tag{6.7}$$

where the function f belongs to the set of functions with Lipschitz value $\|f\|_L$ less than 1.

6.3.2 CryoGAN and the Connection with WGANs

A key observation is that (6.7) falls under the framework of the generative adversarial networks (GANs) [206] called WGANs (for Wasserstein-GANs) [208].

In the classical WGAN representation, the function f is parameterized by a neural network $\mathbf{D}_\phi$ with parameters ϕ that is called the discriminator. The task of the discriminator is to learn to differentiate between samples that originate from real data and samples that originate from synthetic data. These synthetic data are produced by another neural network called the generator, which aims at producing data realistic enough to delude the discriminator. This adversarial-learning scheme drives the WGAN to capture the distribution of the real data.

In CryoGAN, we also learn the volume $\mathbf{c}$ whose simulated projections follow the real-data distribution, as captured through adversarial learning. The role of the cryo-EM physics simulator is to produce synthetic projections of a volume estimate $\mathbf{c}$ using (6.1). These simulated projections then follow a distribution $\mathbf{b} \sim p_{\mathbf{c}}$. Hence, (8) translates into

$$\mathbf{c}_{\text{rec}} = \underset{\mathbf{c}}{\operatorname{argmin}} \max_{\mathbf{D}_\phi:\|\mathbf{D}_\phi\|_L \leq 1} \Big(\mathbb{E}_{\mathbf{b}\sim p_{\text{data}}}[\mathbf{D}_\phi(\mathbf{b})] - \mathbb{E}_{\mathbf{b}\sim p_{\mathbf{c}}}[\mathbf{D}_\phi(\mathbf{b})] \Big). \tag{6.8}$$

As proposed in [211], the Lipschitz constraint $\|\mathbf{D}_\phi\|_L \leq 1$ is best enforced by penalizing the norm of the gradient of $\mathbf{D}_\phi$ with respect to its input. This gives the final formulation of our reconstruction problem as

$$\mathbf{c}_{\text{rec}} = \underset{\mathbf{c}}{\operatorname{argmin}} \max_{\mathbf{D}_\phi} \Big(\mathbb{E}_{\mathbf{b}\sim p_{\text{data}}}[\mathbf{D}_\phi(\mathbf{b})] - \mathbb{E}_{\mathbf{b}\sim p_{\mathbf{c}}}[\mathbf{D}_\phi(\mathbf{b})] - \lambda \cdot \mathbb{E}_{\mathbf{b}\sim p_{\text{int}}}[(\|\nabla_{\mathbf{b}}\mathbf{D}_\phi(\mathbf{b})\| - 1)^2] \Big). \tag{6.9}$$

Here, p_{int} describes the uniform distribution along the straight line between points sampled from p_{data} and $p_{\mathbf{c}}$, while $\lambda \in \mathbb{R}_+$ is an appropriate penalty coefficient (see [211], Section 4).

6.4 The CryoGAN Algorithm

6.4.1 The CryoGAN Adversarial Learning Scheme

Equation (6.9) is a min-max optimization problem. By replacing the expected values with their empirical counterparts (*i.e.*, sums) [211], we reformulate it as the minimization of

$$L_S(\mathbf{c}, \mathbf{D}_\phi) = \sum_{n\in S} \mathbf{D}_\phi(\mathbf{b}^n_{\text{data}}) - \sum_{n\in S} \mathbf{D}_\phi(\mathbf{b}^n_{\text{sim}}) - \lambda \sum_{n\in S} (\|\nabla_{\mathbf{b}}\mathbf{D}_\phi(\mathbf{b}^n_{\text{int}})\| - 1)^2), \tag{6.10}$$

where

- S consists of either the full dataset $S_{\text{full}} = \{1, \dots, N_{\text{tot}}\}$ or a batch $B \subseteq S_{\text{full}}$;
- $\mathbf{b}^n_{\text{data}}$ is a real projection sampled from the acquired dataset;
- $\mathbf{b}^n_{\text{sim}} \sim p_{\mathbf{c}}$ is a synthetic projection of the current estimate $\mathbf{c}$ generated by the cryo-EM physics simulator;
- $\mathbf{b}^n_{\text{int}} = \alpha_n \cdot \mathbf{b}^n_{\text{data}} + (1-\alpha_n) \cdot \mathbf{b}_{\text{sim}}$, where α_n is sampled from a uniform distribution between 0 and 1.

In practice, we minimize (6.10) through SGD using batches. We alternatively update the discriminator $\mathbf{D}_\phi$ (in n_{discr} iterations) using an Adam optimizer [205] and the volume $\mathbf{c}$ (in 1 iteration) using the appropriate gradients of $L_S(\mathbf{c}, \mathbf{D}_\phi)$. The pseudo-code and the schematic view of the CryoGAN algorithm are given in Algorithm 7 and Figure 6.1b, respectively. The architecture of the CryoGAN discriminator is presented below (see Section 6.4.3).

6.4.2 The Cryo-EM Physics Simulator

The goal of the physics simulator is to sample $\mathbf{b}_{\text{sim}} \sim p_{\mathbf{c}}(\mathbf{b})$. We do this in three steps.

1. Sample the imaging parameters $\boldsymbol{\varphi}$ from the distribution $p_{\boldsymbol{\varphi}}$: $\boldsymbol{\varphi} \sim p_{\boldsymbol{\varphi}}$.
2. Generate noiseless CTF-modulated and shifted projections from the current volume estimate with $\mathbf{H}_{\boldsymbol{\varphi}}(\mathbf{c})$.
3. Sample the noise model to simulate noisy projections $\mathbf{b} = \mathbf{H}_{\boldsymbol{\varphi}}(\mathbf{c}) + \mathbf{n}$, where $\mathbf{n} \sim p_{\mathbf{n}}$.

The pseudo-code of the cryo-EM physics simulator is given in Algorithm 8.

Step 1: Sampling of the Imaging Parameters

We recall that the set of imaging parameters is given by $\boldsymbol{\varphi} = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast}})$.

We first sample the Euler angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ from a distribution $p_{\boldsymbol{\theta}}$ decided *a priori* based on the acquired dataset. Similarly, the projection shifts $\mathbf{t} = (t_1, t_2)$ are also sampled from the prior distribution $p_{\mathbf{t}}$.

The CTF parameters $\mathbf{h} = (d_1, d_2, \alpha_{\text{ast}})$ are sampled from the prior distribution $p_{\mathbf{h}}$. For example, a uniform distribution over the defocus range can be assumed. In practice, we exploit the fact that the CTF parameters can often be efficiently estimated for the micrographs. We then uniformly sample from the whole set of extracted CTF parameters.

Algorithm 7 CryoGAN

Parameters: number n_{train} of training iterations; number n_{discr} of iterations of the discriminator per training iteration; size N of the batches used for SGD; penalty parameter λ.

for n_{train}**do**
for n_{discr} **do**

- Sample real projections: $\{\mathbf{b}^1_{\text{batch}},\dots,\mathbf{b}^N_{\text{batch}}\} = \{\mathbf{b}^n_{\text{data}}\}_{n\in B}$.
- Sample projections simulated from current $\mathbf{c}$: $\{\mathbf{b}^1_{\text{sim}},\dots,\mathbf{b}^N_{\text{sim}}\} \sim p_{\mathbf{c}}$ (see Algorithm 2).
- Sample $\{\alpha_1,\dots,\alpha_n\} \sim U[0,1]$.
- For all $n \in \{1,\dots,N\}$, compute $\mathbf{b}^n_{\text{int}} = \alpha_n \cdot \mathbf{b}^n_{\text{batch}} + (1-\alpha_n)\cdot \mathbf{b}^n_{\text{sim}}$.
- Update the parameters ϕ of the discriminator $\mathbf{D}_\phi$ by ascending its stochastic gradient

$$\nabla_\phi L_B(\mathbf{c},\mathbf{D}_\phi) = \nabla_\phi \left(\sum_{n=1}^N \mathbf{D}_\phi(\mathbf{b}^n_{\text{batch}}) - \sum_{n=1}^N \mathbf{D}_\phi(\mathbf{b}^n_{\text{sim}}) - \lambda \sum_{n=1}^N (\|\nabla_{\mathbf{b}} \mathbf{D}_\phi(\mathbf{b}^n_{\text{int}})\| - 1)^2 \right). \tag{6.11}$$

end for

- Sample $\{\mathbf{b}^1_{\text{sim}},\dots,\mathbf{b}_{\text{sim}}{}^N\} \sim p_{\mathbf{c}}$;
- Update of the volume $\mathbf{c}$ by descending its stochastic gradient

$$\nabla_{\mathbf{c}} L_B(\mathbf{c},\mathbf{D}_\phi) = \nabla_{\mathbf{c}} \left(-\sum_{n=1}^N \mathbf{D}_\phi(\mathbf{b}^n_{\text{sim}}) \right). \tag{6.12}$$

end for

Step 2: Synthesis of Noiseless Cryo-EM Projections

We generate noiseless projections $\mathbf{b}_{\text{noiseless}}$ by applying $\mathbf{H}_{\boldsymbol{\varphi}}$ to the current volume estimate $\mathbf{c}$. The projection operator $\mathbf{P}_{\boldsymbol{\theta}}$ in (6.2) is implemented using the ASTRA toolbox [201].

Step 3: Sampling of the Noise Model

The precise modeling of the noise is a particularly challenging feat in cryo-EM. To produce noise realizations that are as realistic as possible, we extract random background patches directly from the micrographs themselves, at locations where 3D particles do not appear. For consistency, the noise patch added to a given noiseless projection is taken from the same micrograph that was used in Step 1 to estimate the CTF parameters previously applied to that specific projection. Additional details for this implementation are given in Appendix A.6.6.

Algorithm 8 Cryo-EM Physics Simulator

Inputs: current volume estimate $\mathbf{c}$

1. Sample the imaging parameters $\boldsymbol{\varphi} = [\boldsymbol{\theta}, \mathbf{t}, \mathbf{h}]$.
   - Sample the Euler angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \sim p_{\boldsymbol{\theta}}$.
   - Sample the 2D shifts $\mathbf{t} = (t_1, t_2) \sim p_{\mathbf{t}}$.
   - Sample the CTF parameters $\mathbf{h} = (d_1, d_2, \alpha_{\text{ast}}) \sim p_{\mathbf{h}}$.
2. Generate a synthetic noiseless projection based on (6.2), with $\mathbf{b}_{\text{noiseless}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{c}$.
3. Sample the noise $\mathbf{n} \sim p_{\mathbf{n}}$. Add to the projection as $\mathbf{b}_{\text{sim}} = \mathbf{b}_{\text{noiseless}} + \mathbf{n}$.

6.4.3 The CryoGAN Discriminator Network

The role of the discriminator is to differentiate between projections from the real dataset and projections synthesized by the cryo-EM physics simulator. The gradients from the discriminator (see (6.12) in Algorithm 7) carry information on the difference between real and synthesized images at a given run-time. Those gradients are used by the simulator to update itself, thus improving on the realism of the projections.

The discriminator network takes an image as input and outputs a scalar value. Its architecture is illustrated in Figure 6.2. It is composed of 8 layers: 6 convolutional blocks followed by 2 fully connected (FC) layers. Each convolutional block is made up of a convolutional layer followed by a max-pooling and a leaky ReLU (with negative slope of 0.1). The number of channels in each convolutional layer is 96, 192, 384, 768, 1536, and 3072, respectively. The filters in these layers are of size 3, and the padding size is 1. The max-pooling layer uses a kernel of size 2 with a stride of 2. This leads to a downsampling by a factor of 2. The output of the final convolutional block is then reshaped, fed into the FC layer with 10 neurons, and finally processed by a leaky ReLU. The resulting activations are fed to the last FC layer to output a scalar.

6.5 Results

6.5.1 Performance on a Synthetic Dataset

We first assessed the viability and performance of CryoGAN on a synthetic dataset that consists of 41,000 β-galactosidase projections. To generate this dataset of "real projections", we fitted the PDB entry (*5a1a*) of the protein with a 5Å-resolution structure and applied the forward model (6.2) to obtain thousands of projections modulated by CTF effects and corrupted by noise. We then randomly divided this dataset in two and applied the CryoGAN algorithm separately on both halves to generate half maps. The details of the experimental setup are

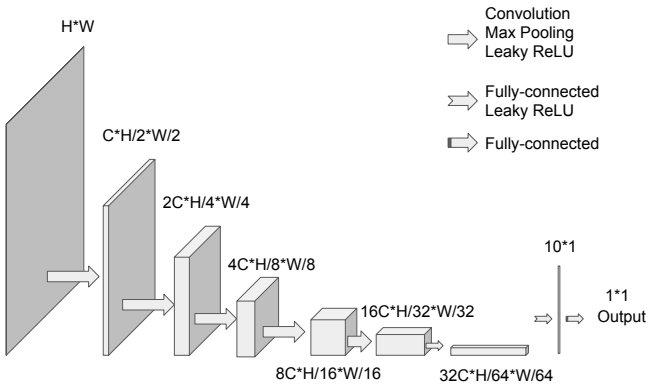


Figure 6.2 – Architecture of the discriminator. The parameter for the channel size is $C = 96$ in every experiment. The input image with size $H \times W$ is successively processed and downsampled to output a scalar.

given in Appendix A.6.5.

We ran the CryoGAN algorithm for 160 minutes on an NVIDIA P100 GPU and obtained a 7.58Åresolution reconstruction (Figure 6.3.a). Starting from a zero-valued volume, CryoGAN progressively updates the 3D structure so that its simulated projections, generated by the cryo-EM physics simulator and displayed in Figure 6.3.b, reach a distribution that matches that of the particles dataset. These gradual updates are at the core of the deep adversarial learning scheme of CryoGAN. At each iteration of the algorithm, the gradients from the discriminator (Figure 6.1.b) carry information about the current difference between the picked particles ("real data") and the generated projections ("fake data"). These gradients are used by the cryo-EM physics simulator to update its parameters and learn a volume that improves the fidelity of the projections. Hence, at the end of its run, the volume learned by CryoGAN has projections (Figure 6.3.c, Rows 1-3) that are similar to the picked particles (Figure 6.3.c, Row 4) in a *distributional* sense. The evolution of the FSC between the reconstructed half maps (Figure 6.3.d) testifies to the progressive increase in resolution that derives from this adversarial learning scheme.

6.5.2 Results on Real Data

We then deployed CryoGAN on 41,123 β-galactosidase projections (EMPIAR-10061) to assess its capacity to reconstruct real, experimental data. This dataset is much more challenging and puts CryoGAN to the test. Here as well, we randomly divided the dataset in two and applied CryoGAN separately on both halves. The details of this experimental setup are given in Appendix A.6.6.

We ran CryoGAN for 160 minutes to obtain a 8Åresolution reconstruction using an NVIDIA

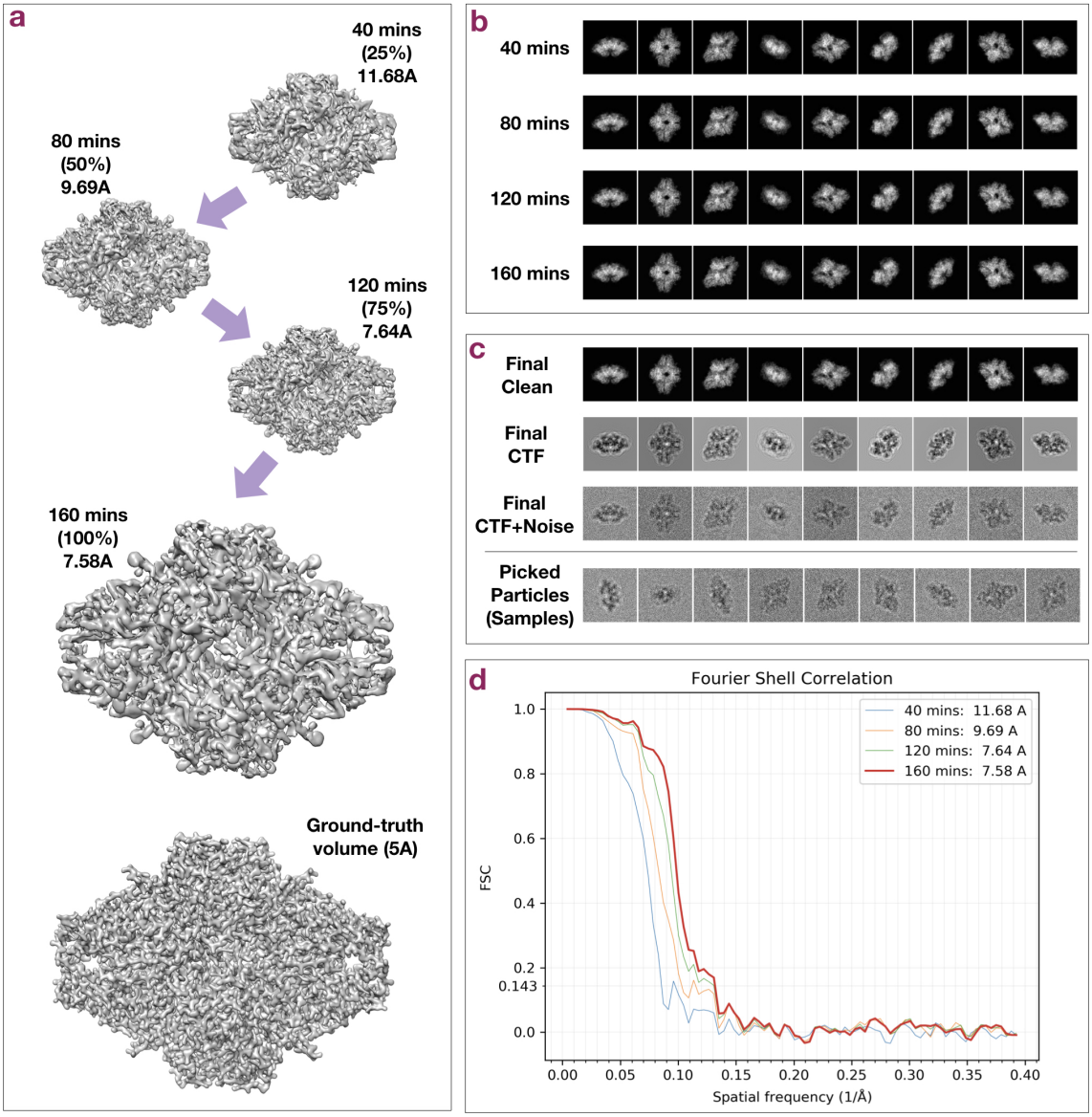


Figure 6.3 – CryoGAN is applied on a synthetic dataset (dubbed "picked particles") generated from a 5Å β-galactosidase volume. **(a)** Starting with zero-values, the volume is progressively updated to produce projections whose distribution matches that of the picked particles. **(b)** Evolution during training of some "clean" projections (*i.e.*, before CTF and noise) generated by the cryo-EM physics simulator. **(c)** *Row 1*: Clean projections (before CTF and noise) generated at the final stage of training. *Row 2*: CTF-modulated projections (before noise) generated at the final stage of training. *Row 3*: Realistic projections (with CTF and noise) generated at the final stage of training. *Row 4*: Samples of picked particles, for comparison. **(d)** FSC curves between the two reconstructed half maps at different training times.

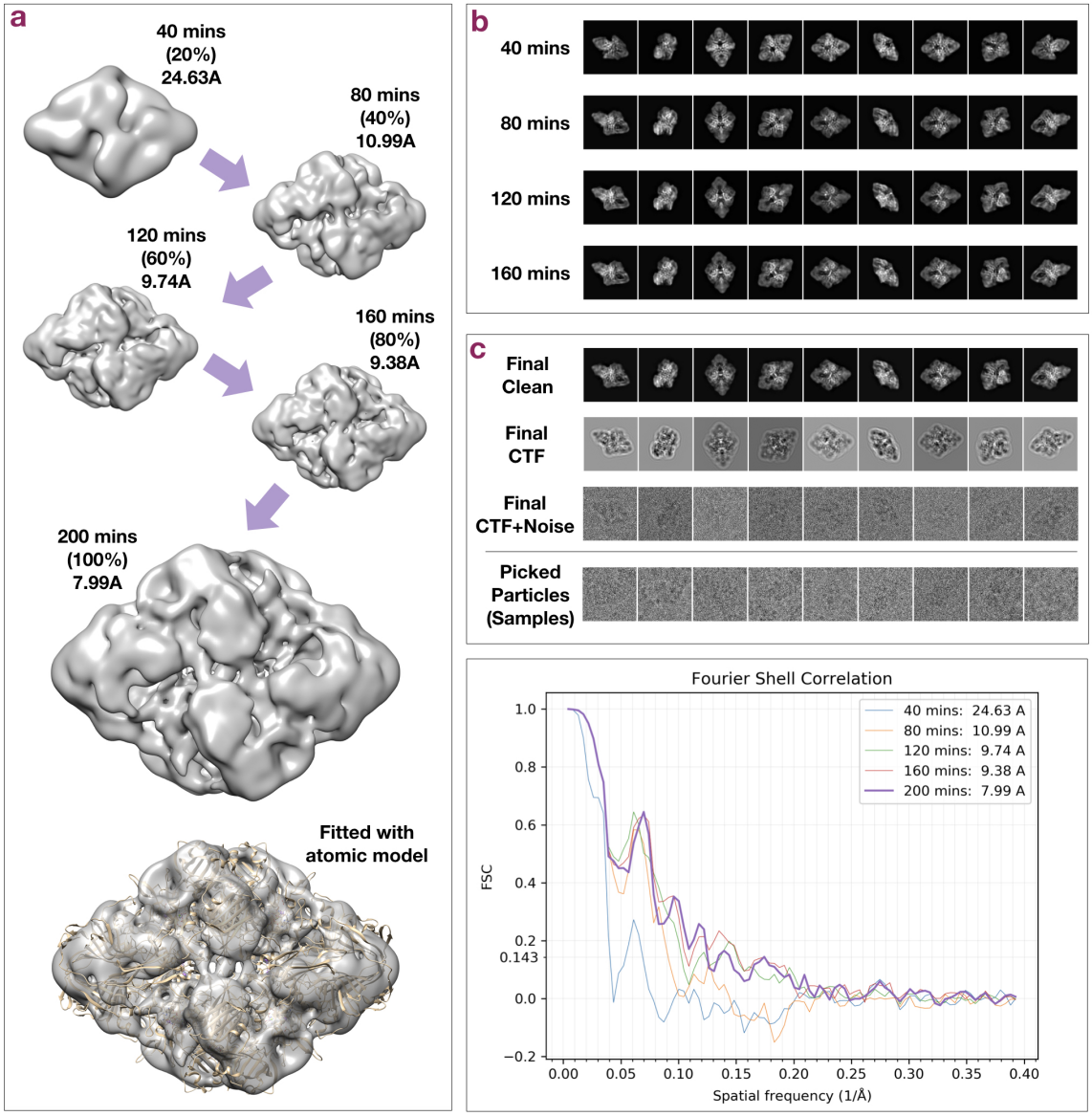


Figure 6.4 – Evolution of CryoGAN while reconstructing the real β-galactosidase dataset from [50]. **(a)** Starting with zero-values, the volume is progressively updated to produce projections whose distribution matches that of the real dataset. **(b)** Evolution during the training of clean projections (*i.e.*, before CTF and noise) generated by the cryo-EM physics simulator. **(c)** *Row 1*: Clean projections generated at the final stage of training. *Row 2*: CTF-modulated projections (before noise) generated at the final stage of training. *Row 3*: Realistic projections (with CTF and noise) generated at the final stage of training. *Row 4*: Samples of the real dataset. **(d)** FSC between the two reconstructed half maps at different training times.

P100 GPU. The results are displayed in Figure 6.4. The flexible architecture of CryoGAN permits the straightforward injection of prior knowledge on this specific imaging procedure into the reconstruction pipeline (*e.g.*, the assumption of uniform-pose distribution). Using this prior knowledge and its adversarial learning scheme, CryoGAN converges toward the reconstruction that best explains the statistics of the dataset (Figure 6.4.a). As with the synthetic experiment, this is achieved by exploiting the gradients of the discriminator to update the simulator and the current volume estimate, so that the projections generated at later iterations (Figure 6.4.b) follow a distribution that better approaches that of the real dataset. Higher-resolution details are thus progressively introduced in the estimated volume throughout the run, as illustrated by the FSC curves between successive reconstructed half maps (Figure 6.4.d). For this particular run, this resulted in a 7.99Å β-galactosidase structure whose synthetic projections closely resemble the real projections, both visually (Figure 6.4.c) and—more importantly—statistically.

6.6 Discussion

We demonstrated the ability of CryoGAN to autonomously reconstruct 3D structures through a purely data-driven adversarial learning scheme, which represents a complete change of paradigm for SPA reconstruction. Capitalizing on the ability of deep-learning models to capture data distributions, the CryoGAN algorithm looks for the reconstruction that is most consistent with the measurements in a *distributional* sense. Hence, it is able to avoid the angular-assignment procedure altogether by directly exploiting the statistics of the provided dataset. CryoGAN is a truly unsupervised algorithm that requires minimal prior information and user input. It is backed up by a sound mathematical framework that gives guarantees on the recovery, provided the image-formation model is valid. When the assumptions are met, our main theorem (see Appendix A.6.2) asserts that CryoGAN samples the proper probability distribution and recovers the correct 3D volume.

An important point is that CryoGAN bypasses angular assignments in a likelihood-free manner, which is in contrast with the likelihood-based approaches used, for instance, in the CryoSPARC software [62]. This permits CryoGAN to avoid marginalization over the angles, a complex task that is inherent in likelihood-based approaches but that is undesirable because it requires the approximation of integrals by sums. CryoGan also sidesteps many cumbersome processing steps such as 2D alignment or 2D/3D classification, which further reduces the need for user-dependent inputs.

Our synthetic experiments demonstrate the ability of CryoGAN to resolve a structure so that its simulated projection distribution matches that of the experimental projections. These results validate the CryoGAN paradigm and the viability of its current implementation. Indeed, without any prior training and starting from a zero-valued volume, the algorithm is able to autonomously capture the relevant statistical information from the dataset of noise-corrupted, CTF-modulated projections, and to learn the volume that best explains these statistics. The results on the real β-galactosidase dataset further demonstrate the capacity of CryoGAN to

perform reconstruction in challenging real imaging conditions.

Roadmap for Future Work

The implementation of the CryoGAN algorithm is bound to further improve. We expect that several interesting developmental steps lie ahead. In particular, we are convinced that the ongoing progresses in deep-learning architectures will help enrich the fast-evolving CryoGAN algorithm, providing additional gains in resolution in the near future.

The current implementation of CryoGAN is at the proof-of-concept stage. Among several algorithmic refinements, we expect the speed could be increased by using a fast Fourier-transform-based projector. Some fine tuning of the global CryoGAN architecture (*e.g.*, the number of convolutional layers in the discriminator) could further improve its performance, leading again to faster convergence. This would consequently improve the resolution that can be obtained in a given amount of time. The use of a larger discriminator and/or size of the dataset size, GPU permitting, would likely have the same effect.

Like all reconstruction algorithms, CryoGAN can fail if the dataset contains too many corrupted projections, possibly with broken structures or strong optical aberrations. Several solutions could be deployed to handle excessive outliers in the data distribution. One approach would be to include a step that automatically spots and discards corrupted data so that the discriminator never sees them. Another interesting option is to directly simulate the patches of nonaligned micrographs/frames (rather than picked projections) and match their distribution to that of the raw dataset. Doing so would also allow cryoGAN to bypass additional preprocessing tasks, in particular, particle picking. Finally, recent deep-learning-based approaches able to track outliers in data could prove useful.

Similar to the likelihood-based methods, the CryoGAN algorithm requires the specification of the distribution of orientations. One could also parameterize it and learn its parameters during the reconstruction procedure [212]. A similar approach could be used to calibrate the distribution of the translations of the projections.

On the theoretical side, we currently have mathematical guarantees on the recovery of volumes for which the assumed distribution of orientations (be it uniform or not) matches the distribution of the real data. Moreover, we have prior mathematical indications that this can also be achieved when there is a certain mismatch between the assumed distribution of orientations and the actual one, given that an appropriate GAN loss is used.

The performance of the cryo-EM physics simulator should improve hand-in-hand with our ability to precisely model the physics behind SPA with computationally tractable entities. At the moment, CryoGAN assumes that the noise is additive in its image-formation model. One could alternatively consider a Poisson-noise-based forward model [213, 214]. This would, however, require backpropagating through a Poisson distribution, a nontrivial operation at this stage of development.

Another promising direction of research is the use of a coarse-to-fine strategy to reconstruct the volume progressively as the resolution improves. The motivation is that increased robustness during the low-resolution regime tends to have a positive impact on the convergence of the higher-resolution steps. Several GAN architectures rely on such approaches, such as the progressive GANs [215] and the styleGANs [216]. The benefits of multi-scale refinement could be considerable for CryoGAN given the extremely challenging imaging conditions that prevail in SPA and make the convergence of optimization algorithms sometimes fail. The core idea here would be to have the discriminator learn to differentiate between real and synthetic distributions first at a low resolution, and then at successively higher ones. The impact on CryoGAN could be as important as the one it had on GANs, which progressed in just a few years from generating blurry facial images [206] to synthetic images indistinguishable from real facial images [215, 216]. More generally, the upcoming tools and extensions in GAN architectures could bring significant gain in resolution to the CryoGAN implementation.

While the spatial resolution of the CryoGAN reconstructions from real data is not yet competitive with the state-of-the-art, the algorithm is already able to steadily perform the hardest part of the job, which is to obtain a reasonable structure by using nothing save the projection dataset and CTF estimations. We believe that the aforementioned developments will help to bring the CryoGAN algorithm to the stage where it becomes a relevant contributor for high-resolution reconstruction in SPA. Moreover, we have laid out a roadmap of future improvements that should get us to this stage. Our hope is that this new take on GANs will foster developments beyond the present application in cryo-EM.

7 Conclusion: On Getting the Best of Both Worlds

Our mathematical formulation of the cryo-EM imaging model (Chapter 2) is the foundational layer of all the works presented in this thesis. In particular, it is at the core of the variational methods we developed for STET (Chapter 3) and SPA (Chapter 4). By incorporating prior knowledge into the reconstruction process, model-based methods demonstrate superior performance in ill-posed conditions. Hence, they permit high-quality reconstruction even in the challenging, low-dose imaging conditions faced in cryo-EM. In addition, model-based methods are backed up by solid theoretical foundations that guarantee their robustness.

By contrast, the power of data-driven approaches lies in their (not yet fully understood) ability to exploit information directly from the datasets, which would be hardly accessible otherwise. In this regard, the two data-driven approaches we designed for SPA also completely rely on our ability to model the cryo-EM procedure. Our supervised method for orientation recovery (Chapter 5) depends on this model for the creation of a realistic training dataset, while in CryoGAN (Chapter 6) it plays a key role as a module inside the adversarial scheme itself. To the best of our knowledge, CryoGAN is the first demonstration of a deep-learning architecture able to singlehandedly perform the full SPA reconstruction procedure without prior training, which testifies to the potential of such hybrid approaches.

SPA has revolutionized structural biology over the last decade and will remain an active topic of research for years to come. Beyond the technical challenges addressed in this thesis, the most notable remaining difficulty with the current pipeline is the non-straightforward handling of structural heterogeneity. It that respect, the flexible architecture of CryoGAN could facilitate its extension to the treatment of structures with a continuum of conformational states, a feat often perceived as the greatest challenge ahead for SPA [52].

In any case, the most impactful forthcoming developments in cryo-EM will likely rely on a combination of model-based and learning-based elements, and leverage the best of both worlds. Ultimately, the better the model, the stronger the learning, the higher the performance of the algorithm. This foresees an intense, powerful interplay between the two paradigms in the near future—from model-based to data-driven, and back.

A Appendices

A.1 Properties of the X-ray Transform

The X-ray transform fulfills several important properties. We state here the main ones in 3D. For the proofs, we refer to [31].

Linearity. For any $\alpha, \beta \in \mathbb{R}$ and $f, g \in L_2(\mathbb{R}^3)$, we have that

$$\mathscr{P}_{\boldsymbol{\theta}}\{\alpha f + \beta g\}(\mathbf{y}) = \alpha \cdot \mathscr{P}_{\boldsymbol{\theta}}\{f\}(\mathbf{y}) + \beta \cdot \mathscr{P}_{\boldsymbol{\theta}}\{g\}(\mathbf{y}). \tag{A.1}$$

This follows directly from the linearity of integral operators.

Pseudo-shift-invariance. For any $\mathbf{x}_0 \in \mathbb{R}^3$, we have that

$$\mathscr{P}_{\boldsymbol{\theta}}\{f(\cdot - \mathbf{x}_0)\}(\mathbf{y}) = \mathscr{P}_{\boldsymbol{\theta}}\{f\}(\mathbf{y} - \mathbf{M}_{\boldsymbol{\theta}^\perp}\mathbf{x}_0), \tag{A.2}$$

where $\mathbf{M}_{\boldsymbol{\theta}^\perp}$ is given in (1.3).

Scale-Invariance. For any $\alpha > 0$, we have that

$$\mathscr{P}_{\boldsymbol{\theta}}\{f(\alpha\cdot)\}(\mathbf{y}) = \alpha^{-1}\mathscr{P}_{\boldsymbol{\theta}}\{f\}(\alpha\mathbf{y}). \tag{A.3}$$

Pseudo-Distribution w.r.t. Convolution. For any $f, g \in L_2(\mathbb{R}^3)$, we have that

$$\mathscr{P}_{\boldsymbol{\theta}}\{f * g\}(\mathbf{y}) = \big(\mathscr{P}_{\boldsymbol{\theta}}\{f\}(\cdot) * \mathscr{P}_{\boldsymbol{\theta}}\{g\}(\cdot)\big)(\mathbf{y}). \tag{A.4}$$

A.2 CTF Model in Cryo-EM

We model the CTF in cryo-EM as

$$\widehat{w}(\boldsymbol{\omega}) = A(\boldsymbol{\omega})E(\boldsymbol{\omega})C(\boldsymbol{\omega}), \tag{A.5}$$

where A is the objective aperture function, E describes the spatial and chromatic envelop function, and C is the phase contrast transfer function [23]. More precisely, we have that

- $A : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\hat{A}(\boldsymbol{\omega}) = \begin{cases} 1, & \|\boldsymbol{\omega}\| \leq \omega_{\text{cutoff}} \\ 0, & \|\boldsymbol{\omega}\| > \omega_{\text{cutoff}}, \end{cases} \tag{A.6}$$

  where $\omega_{\text{cutoff}} = \frac{2\pi d_{\text{ap}}}{f_l \lambda}$ is the cutoff frequency, f_l is the focal length of the objective lens, and d_{ap} corresponds to the diameter of the aperture;

- $E : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$\hat{E}(\boldsymbol{\omega}) = \exp\left(-B(\|\boldsymbol{\omega}\|^2)\right), \tag{A.7}$$

  where $B(\|\boldsymbol{\omega}\|^2)$ is a function influenced by chromatic aberration and spatial incoherence;

- $C : \mathbb{R}^2 \to \mathbb{R}$takes the form

$$C(\boldsymbol{\omega}) = -\sqrt{1 - A^2}\sin(\gamma(\boldsymbol{\omega})) - A^2\cos(\gamma(\boldsymbol{\omega})), \tag{A.8}$$

  with

$$\gamma(\boldsymbol{\omega}) = \pi\lambda\left(z(\alpha)\|\boldsymbol{\omega}\|^2 - \frac{1}{4}\lambda^3 c_s \|\boldsymbol{\omega}\|^4\right), \tag{A.9}$$

  where λ is the electron wavelength, c_s is third-order spherical aberration constant, α is the phase of the vector $\boldsymbol{\omega}$, and $z(\alpha)$ is the defocus arising at the phase α. This defocus is given as

$$z(\alpha) = z_u \cos^2(\alpha - \alpha_0) + z_v \sin^2(\alpha - \alpha_0), \tag{A.10}$$

  where z_u and z_v are the horizontal and vertical defocus and α_0 is the reference angle that defines the azimuthal direction of axial astigmatism. All these parameters are part of the experimental setup.

A.3 Fourier Shell Correlation (FSC) Metric

To assess the quality of a reconstruction, we use the Fourier shell correlation (FSC) between the reconstructed volume V^{rec} and the ground-truth V^{gt}, as defined by

$$\text{FSC}(r) = \frac{\sum\limits_{r_i \in \mathbf{r}} \hat{V}^{\text{rec}}(r_i)\hat{V}^{\text{gt}}(r_i)^*}{\sqrt{\sum\limits_{r_i \in \mathbf{r}} |\hat{V}^{\text{rec}}(r_i)|^2 \sum\limits_{r_i \in \mathbf{r}} |\hat{V}^{\text{gt}}(r_i)|^2}}. \tag{A.11}$$

where $\mathbf{r} = \{(x_i, y_i, z_i) : |\sqrt{x_i^2 + y_i^2 + z_i^2} - r| \leq \varepsilon_r\}$, for $\varepsilon_r > 0$, denotes the set of all points in the discrete Fourier domain that lie in a spherical shell with inner radius $r - \varepsilon_r$ and outer radius $r + \varepsilon_r$, centered at origin. The FSC thus computes the correlation between two corresponding spherical shells of the density maps in the Fourier domain.

A.4 Proofs for Section 4.2

A.4.1 Proof of Theorem 3

Let us expand $\mathscr{J}_p(\boldsymbol{\theta},\mathbf{t})$ in (4.32) as

$$\begin{aligned}\mathscr{J}_p(\boldsymbol{\theta},\mathbf{t}) &= \frac{1}{2}\mathbf{c}^T\mathbf{H}^T\mathbf{H}(\boldsymbol{\theta},\mathbf{t})\mathbf{c} - \mathbf{c}^T\mathbf{H}^T(\boldsymbol{\theta},\mathbf{t})\mathbf{b}^p + \frac{1}{2}\|\mathbf{b}^p\|^2 \\ &= \frac{1}{2}\mathbf{c}^T(\mathbf{w}(\boldsymbol{\theta}) * \mathbf{c}) - \mathbf{c}^T\mathbf{H}^T(\boldsymbol{\theta},\mathbf{t})\mathbf{b}^p + \frac{1}{2}\|\mathbf{b}^p\|^2, \end{aligned} \tag{A.12}$$

where $\mathbf{w}(\boldsymbol{\theta}$ corresponds to one term in the kernel $\mathbf{w}(\boldsymbol{\Theta}) \in \mathbb{R}^N$ is given by, $\forall \mathbf{k} \in \Omega_{3\mathrm{D}}$,

$$[\mathbf{w}(\boldsymbol{\Theta})]_{\mathbf{k}} = \frac{1}{\det(\boldsymbol{\Lambda})}\sum_{p=1}^{P}\left(\psi_{\Theta_p} * \psi_{\Theta_p}^{\vee}\right)(\mathbf{M}_{\Theta_p^{\perp}}\mathbf{k}), \tag{A.13}$$

with $\psi_{\Theta_p} = h * \mathscr{P}_{\Theta_p}(\varphi)$. Note that the kernel $\mathbf{w}(\boldsymbol{\Theta})$ does not depend on the in-plane translations $\boldsymbol{\Gamma}$.

Moreover, because φ is isotropic, we have that $\mathscr{P}_{\boldsymbol{\theta}}(\varphi) = \mathscr{P}(\varphi)$, a quantity that does not depend on $\boldsymbol{\theta}$. Hence,

$$[\mathbf{w}(\boldsymbol{\theta})]_{\boldsymbol{k}} = \frac{1}{\det(\boldsymbol{\Lambda})}\left(\psi * \psi^{\vee}\right)(\mathbf{M}_{\boldsymbol{\theta}^{\perp}}\boldsymbol{k}), \tag{A.14}$$

with $\psi = h * \mathscr{P}(\varphi)$.

Then, from (A.12), one easily sees that, for all $v \in \{\theta_1, \theta_2, \theta_3, t_1, t_2\}$,

$$\frac{\partial \mathscr{J}_p}{\partial v}(\boldsymbol{\theta},\mathbf{t}) = \frac{1}{2}\mathbf{c}^T\left(\mathbf{r}_v * \mathbf{c} - 2\mathbf{q}_v\right), \tag{A.15}$$

where

$$\mathbf{r}_v = \frac{\partial \mathbf{w}(\boldsymbol{\theta})}{\partial v} \text{ and } \mathbf{q}_v = \frac{\partial \mathbf{H}^T(\boldsymbol{\theta},\mathbf{t})\mathbf{b}^p}{\partial v}. \tag{A.16}$$

We now distinguish two cases.

Case $v = \theta_i$ for $i \in \{1,2,3\}$

From (A.14) and the chain rule, we get that

$$r_v[\boldsymbol{k}] = \frac{1}{\det(\boldsymbol{\Lambda})}\left(\frac{\partial \mathbf{M}_{\boldsymbol{\theta}^{\perp}}}{\partial \theta_i}\boldsymbol{k}\right)^T \nabla\left(\psi * \psi^{\vee}\right)(\mathbf{M}_{\boldsymbol{\theta}^{\perp}}\boldsymbol{k}), \tag{A.17}$$

where $\frac{\partial \mathbf{M}_{\boldsymbol{\theta}^{\perp}}}{\partial \theta_i} \in \mathbb{R}^{2\times 3}$ contains the entry-wise derivatives with respect to θ_i of the matrix $\mathbf{M}_{\theta^{\perp}}$ given in (1.3). Moreover, from the definition of $\psi : \mathbf{y} \mapsto (h * \mathscr{P}(\varphi))(\mathbf{y})$, with $\mathbf{y} = (y_1, y_2) \in \mathbb{R}^2$, and

from the derivation property of the convolution, we have that

$$\nabla\left(\psi * \psi^{\vee}\right) = \begin{pmatrix} \dfrac{\partial h}{\partial y_1} * \mathscr{P}(\varphi) * \psi^{\vee} \\ \dfrac{\partial h}{\partial y_2} * \mathscr{P}(\varphi) * \psi^{\vee} \end{pmatrix} = \begin{pmatrix} h * \dfrac{\partial \mathscr{P}(\varphi)}{\partial y_1} * \psi^{\vee} \\ h * \dfrac{\partial \mathscr{P}(\varphi)}{\partial y_2} * \psi^{\vee} \end{pmatrix}. \tag{A.18}$$

Note that we could have also differentiated $\psi^{\vee}$ (instead of h or $\mathscr{P}(\phi)$).

From [18], we have that

$$\left[\mathbf{H}^T(\boldsymbol{\Theta},\boldsymbol{\Gamma})\mathbf{g}\right]_{\mathbf{k}} = \frac{1}{\det(\boldsymbol{\Lambda})} \sum_{p=1}^{P} \left(b_p * \psi^{\vee}_{\theta_p}\right)(\mathbf{M}_{\theta_p^{\perp}}\mathbf{k} + \mathbf{t}_p). \tag{A.19}$$

For $\mathbf{q}_v$, we get from (A.19) that

$$\begin{aligned} q_v[\boldsymbol{k}] &= \frac{1}{\det(\boldsymbol{\Lambda})} \frac{\partial (b_p * \psi^{\vee})(\mathbf{M}_{\boldsymbol{\theta}^{\perp}}\boldsymbol{k} + \mathbf{t})}{\partial \theta_i} \\ &= \frac{1}{\det(\boldsymbol{\Lambda})} \left(\frac{\partial \mathbf{M}_{\boldsymbol{\theta}^{\perp}}}{\partial \theta_i}\boldsymbol{k}\right)^T \nabla (b_p * \psi^{\vee})(\mathbf{M}_{\boldsymbol{\theta}^{\perp}}\boldsymbol{k} + \mathbf{t}), \end{aligned} \tag{A.20}$$

where $\nabla\left(b_p * \psi^{\vee}\right)$ is obtained in the same way as (A.18), with differentiation on $\psi^{\vee}$ instead of b_p.

Case $v = t_j$ for $j \in \{1,2\}$

As $\mathbf{w}(\boldsymbol{\theta})$ does not depend on the in-plane translation $\mathbf{t}$, we have that $\mathbf{r}_v = \mathbf{0}_{\mathbb{R}^N}$. For $\mathbf{q}_v$, as in (A.20), we get that

$$q_v[\boldsymbol{k}] = \frac{1}{\det(\boldsymbol{\Lambda})} \frac{\partial (b_p * \psi^{\vee})}{\partial y_j}(\mathbf{M}_{\boldsymbol{\theta}^{\perp}}\boldsymbol{k} + \mathbf{t}). \tag{A.21}$$

A.4.2 Proof of Proposition 1

The closed-form expression of the x-ray transform of the KBWF φ is provided in [217] as

$$\mathscr{P}\{\varphi\}(\mathbf{y}) = \frac{a\sqrt{2\pi/\alpha}}{I_m(\alpha)} \beta_a(\|\mathbf{y}\|)^{m+\frac{1}{2}} I_{m+\frac{1}{2}}\big(\alpha\beta_a(\|\mathbf{y}\|)\big), \tag{A.22}$$

where $\beta_a(r) = \sqrt{1-(r/a)^2}$ and I_m is the modified Bessel function of order m. Now, let us introduce the function $f(u) = (\alpha u)^{m+\frac{1}{2}} I_{m+\frac{1}{2}}(\alpha u)$ whose derivative is $f'(u) = \alpha(\alpha u)^{m+\frac{1}{2}} I_{m-\frac{1}{2}}(\alpha u)$. Then, we can write (A.22) as

$$\mathscr{P}\{\varphi\}(\mathbf{y}) = \frac{a\sqrt{2\pi/\alpha}}{I_m(\alpha)} \frac{1}{\alpha^{m+\frac{1}{2}}} f(\beta_a(\|\mathbf{y}\|)) \tag{A.23}$$

and, for all $v \in \{1,2\}$, obtain that

$$\frac{\partial \mathscr{P}\{\varphi\}}{\partial y_v}(\mathbf{y}) = \frac{a\sqrt{2\pi/\alpha}}{I_m(\alpha)\alpha^{m+\frac{1}{2}}} \frac{y_v}{\|\mathbf{y}\|} \beta'_a(\|\mathbf{y}\|) f'(\beta_a(\|\mathbf{y}\|)). \tag{A.24}$$

Finally, the injection of f' and $\beta'_a(r) = \left(-\frac{r}{a^2}\left(1-(r/a)^2\right)^{-\frac{1}{2}}\right) = \left(-\frac{r}{a^2\beta_a(r)}\right)$ into (A.24) leads to

$$\begin{aligned}
\frac{\partial \mathscr{P}(\varphi_{\alpha,a})}{\partial y_v}(\mathbf{y}) &= -\frac{a\sqrt{2\pi/\alpha}}{I_m(\alpha)\alpha^{m+\frac{1}{2}}} \frac{y_v}{\|\mathbf{y}\|} \frac{\|\mathbf{y}\|\alpha(\alpha\beta_a(\|\mathbf{y}\|))^{m+\frac{1}{2}}}{a^2\beta_a(\|\mathbf{y}\|)} \\
&\quad \times I_{m-\frac{1}{2}}(\alpha\beta_a(\|\mathbf{y}\|)), \\
&= -\frac{\alpha y_v\sqrt{2\pi/\alpha}}{aI_m(\alpha)} \beta_a(\|\mathbf{y}\|)^{m-\frac{1}{2}} \\
&\quad \times I_{m-\frac{1}{2}}(\alpha\beta_a(\|\mathbf{y}\|)),
\end{aligned} \tag{A.25}$$

which completes the proof.

A.5 Architecture of the SiameseNN for Distance Learning

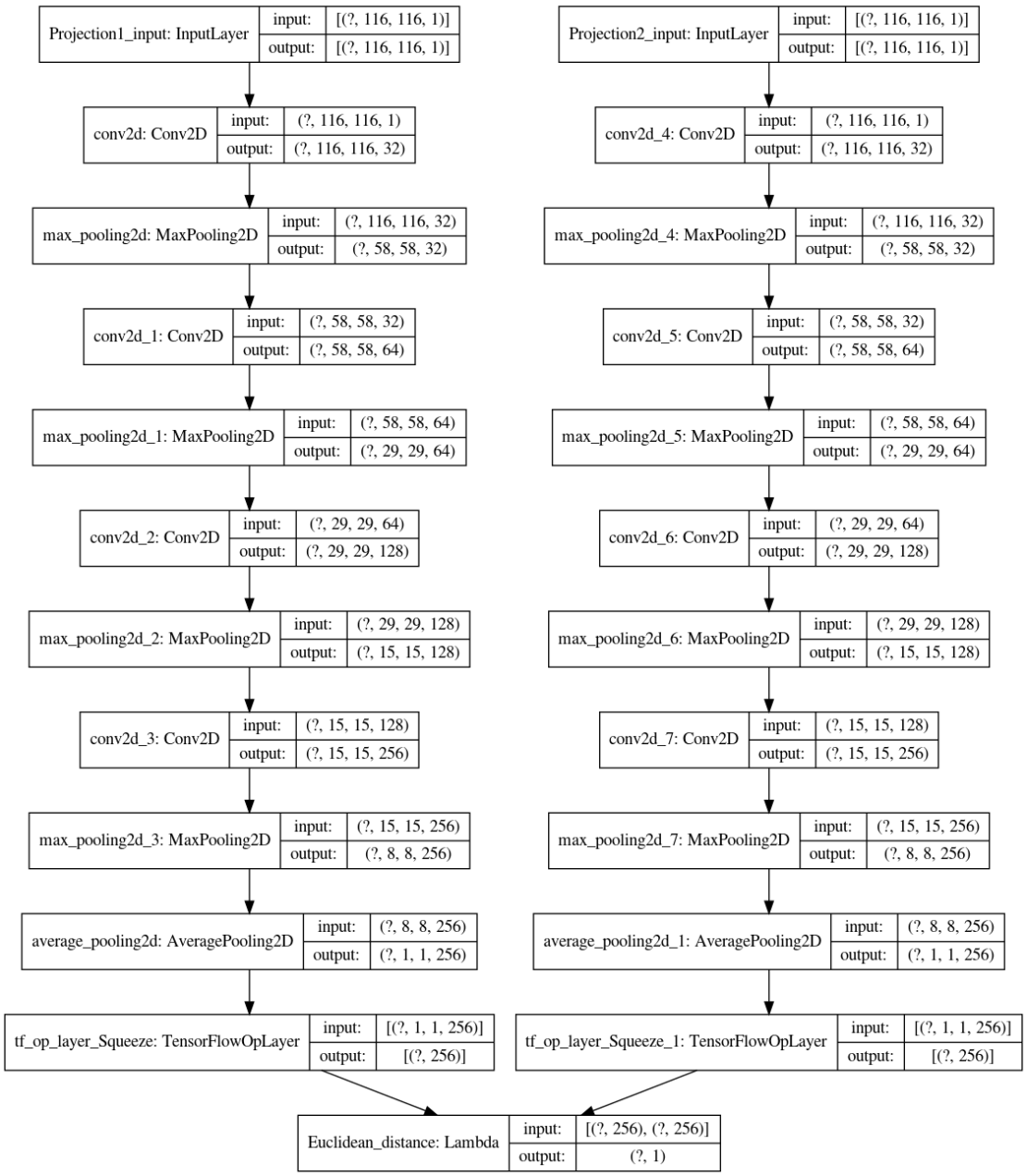


Figure A.1 – Architecture of our SiameseNN. The question mark indicates the number of images per batch, as parallel computations are used to speed up the training.

A.6 Supplementary Materials for Chapter 6

In this section, we present the theoretical results associated to the CryoGAN paradigm, as well as the experimental conditions of the experiments presented in Chapter 6. The notations used in the mathematical proofs correspond to those used in the submitted preprint manuscript[1].

A.6.1 Forward Model in CryoGAN

We indicate here the notations used to model the SPA imaging procedure in CryoGAN. This is done through the linear relationship

$$\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x} + \mathbf{n}, \tag{A.26}$$

where

- $\mathbf{y} \in \mathbb{R}^M$ is a 2D projection of size $M = M_1 \times M_2$;
- $\mathbf{x} \in \mathbb{R}^V$ is the 3D density map of size $V = V_1 \times V_2 \times V_3$;
- $\mathbf{H}_{\boldsymbol{\varphi}} \in \mathbb{R}^{M \times V}$ denotes the forward operator (see (A.27)) with parameters $\boldsymbol{\varphi}$;
- $\boldsymbol{\varphi} = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast}})$ is the set of imaging parameters. It includes the projection (Euler) angles $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, the projection shifts $\mathbf{t} = (t_1, t_2)$, and the CTF parameters $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}})$, where d_1 is the defocus-major, d_2 is the defocus-minor, and α_{ast} is the angle of astigmatism;
- $\mathbf{n} \in \mathbb{R}^M$ represents an additive noise following a distribution $p_{\mathbf{n}}$.

The forward operator $\mathbf{H}_{\boldsymbol{\varphi}}$ is given by

$$\mathbf{H}_{\boldsymbol{\varphi}} = \mathbf{C}_{\mathbf{c}}\mathbf{S}_{\mathbf{t}}\mathbf{P}_{\theta}. \tag{A.27}$$

It is composed of the projection operator $\mathbf{P}_{\boldsymbol{\theta}} : \mathbb{R}^V \to \mathbb{R}^M$, the shift operator $\mathbf{S}_{\mathbf{t}} : \mathbb{R}^M \to \mathbb{R}^M$, and the convolution operator $\mathbf{C}_{\mathbf{c}} : \mathbb{R}^M \to \mathbb{R}^M$.

A.6.2 Theoretical Guarantee of Recovery

The proposed paradigm is supported by Theorem 4, a major contribution of this work. Recall from (A.26) and (A.27) that $\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x} + \mathbf{n}$ is the 2D measurement obtained from the 3D volume $\mathbf{x}$. The operator is $\mathbf{H}_{\boldsymbol{\varphi}} = \mathbf{C}_{\mathbf{c}}\mathbf{S}_{\mathbf{t}}\mathbf{P}_{\boldsymbol{\theta}}$, where $\mathbf{P}_{\boldsymbol{\theta}}$ is the projection operator, $\mathbf{S}_{\mathbf{t}}$ is the shift operator, and $\mathbf{C}_{\mathbf{c}}$ is the convolution operator.

Let $f : \mathbb{R}^d \to \mathbb{R}$. Then its support is $\text{Support}\{f\} = \{\mathbf{x} \in \mathbb{R}^d : f(\mathbf{x}) \neq 0\}$. If $\text{Support}\{f\} = \mathbb{R}^d$, then f is said to have a full support.

[1] https://www.biorxiv.org/content/10.1101/2020.03.20.001016v1

Theorem 4. *Let* $\mathbf{y} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x} + \mathbf{n}$ *as given in* (A.27) *with* $\boldsymbol{\varphi} = (\theta_1, \theta_2, \theta_3, t_1, t_2, d_1, d_2, \alpha_{\text{ast}})$*, where* $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ *are the projection angles,* $\mathbf{t} = (t_1, t_2)$ *are the shifts, and* $\mathbf{c} = (d_1, d_2, \alpha_{\text{ast}})$ *are the CTF parameters (defocus-major, defocus-minor, and angle of astigmatism, respectively),* $\mathbf{x} \in \mathbb{R}^{\mathbf{V}}$ *is the vectorized 3D volume, and* $\mathbf{y}, \mathbf{n} \in \mathbb{R}^M$ *are vectorized 2D images. Let* $\boldsymbol{\theta} \sim p_{\boldsymbol{\theta}}$, $\mathbf{c} \sim p_{\mathbf{c}}$, $\mathbf{t} \sim p_{\mathbf{t}}$, *and* $\mathbf{n} \sim p_{\mathbf{n}}$*. Moreover, assume that*

1. *the Fourier transform* $\hat{p}_{\mathbf{n}}$ *of the noise distribution* $p_{\mathbf{n}}$ *has a full support;*
2. *the support of* $p_{\mathbf{c}}$ *is such that, for any* $\mathbf{c}_1, \mathbf{c}_2 \in \text{Support}\{p_{\mathbf{c}}\}$ *and* $\mathbf{c}_1 \neq \mathbf{c}_2$*, the support* $\mathscr{F}\{\mathbf{C}_{\mathbf{c}_1} + \mathbf{C}_{\mathbf{c}_2}\}$ *is full;*
3. *the volume* $\mathbf{x}$ *is nonnegative everywhere and has a bounded support;*
4. *the probability distributions* $p_{\boldsymbol{\theta}}$*,* $p_{\mathbf{c}}$*, and* $p_{\mathbf{t}}$ *are bounded.*

Then, it holds that

$$p(\mathbf{y}|\mathbf{x}_1) = p(\mathbf{y}|\mathbf{x}_2) \Leftrightarrow \mathbf{x}_1 = \mathbf{G}(\mathbf{x}_2), \tag{A.28}$$

where $\mathbf{G}$ *is some member of the set of rotation-reflection operations.*

Proof. We first comment on the assumptions. Assumption 1) is true for many noise distributions, including the Gaussian distribution. Assumption 2) is generally true as well. In fact, it is used to justify the application of Wiener filter to the clustered projections in classical cryo-EM reconstruction pipelines. Assumption 3) is true since the volume represents the Coulomb potential, which is nonnegative. Also, the biological structures considered in cryo-EM have finite sizes.

We denote $\mathbf{y}_{\text{noiseless}} = \mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$ with distribution $p_{\text{noiseless}}(\cdot|\mathbf{x})$. We shall prove the following in sequence:

1. $p(\cdot|\mathbf{x}_1) = p(\cdot|\mathbf{x}_2) \Leftrightarrow p_{\text{noiseless}}(\cdot|\mathbf{x}_1) = p_{\text{noiseless}}(\cdot|\mathbf{x}_2)$,
2. $p_{\text{noiseless}}(\cdot|\mathbf{x}_1) = p_{\text{noiseless}}(\cdot|\mathbf{x}_2) \Leftrightarrow \mathbf{x}_2 = \mathbf{G}(\mathbf{x}_1)$.

For the first part we progress by noting that $\mathbf{y} = \mathbf{y}_{\text{noiseless}} + \mathbf{n}$. Since the distribution of the addition of two random variables is equal to the convolution of the distributions of the two random variables, we have that

$$p(\mathbf{y}|\mathbf{x}) = (p_{\text{noiseless}}(\cdot|\mathbf{x}) * p_{\mathbf{n}})(\mathbf{y}), \tag{A.29}$$

$$\mathscr{F}\{p(\cdot|\mathbf{x})\} = \mathscr{F}\{p_{\text{noiseless}}(\cdot|\mathbf{x})\}\mathscr{F}\{p_{\mathbf{n}}\}. \tag{A.30}$$

By Assumption (1, Theorem 4), we can now write that

$$p_{\text{noiseless}}(\cdot|\mathbf{x}) = \mathscr{F}^{-1}\left\{\frac{\mathscr{F}\{p(\cdot|\mathbf{x})\}}{\mathscr{F}\{p_{\mathbf{n}}\}}\right\}. \tag{A.31}$$

From (A.31), it is easy to see that $p(\cdot|\mathbf{x}_1) = p(\cdot|\mathbf{x}_2) \Leftrightarrow p_{\text{noiseless}}(\cdot|\mathbf{x}_1) = p_{\text{noiseless}}(\cdot|\mathbf{x}_2)$. This concludes our first part.

For the second part, we use the result from Theorem 6, which is based on a continuous-domain volume. Note that $\mathbf{x}$ actually represent a continuous domain volume. Given Assumption (4, Theorem 4), the continuous-domain representation of $\mathbf{x}$ is

$$f_{\mathbf{x}}(\cdot) = \sum_{k_1=1}^{V_1} \sum_{k_2=1}^{V_2} \sum_{k_3=1}^{V_3} c_{k_1,k_2,k_3}\, \beta(\cdot - s\mathbf{k}), \tag{A.32}$$

where β is a compactly supported basis function, s is the size of a voxel in unit length, and c is such that $f_{\mathbf{x}}(s\mathbf{k}) = \mathbf{x}[\mathbf{k}]$ for $\mathbf{k} \in \mathbb{Z}^3$. We assume a β such that $f_{\mathbf{x}}$ is nonnegative. The operator $\mathbf{H}_{\boldsymbol{\varphi}}\mathbf{x}$ has the same effect as the application of the continuous-domain forward operator, given in (**??**), to $f_{\mathbf{x}}$ and then discretization and vectorization. We can now invoke Theorem 6 which claims that, given $f_{\mathbf{x}_1}$ and $f_{\mathbf{x}_2}$, the support of their corresponding $p_{\text{noiseless}}$ is identical if and only if $f_{\mathbf{x}_1}$ and $f_{\mathbf{x}_2}$ are identical up to a rotation-reflection operation. The latter is equivalent to $\mathbf{x}_1 = \mathbf{G}(\mathbf{x}_2)$, where $\mathbf{G}$ is some rotation-reflection operation. This concludes that $p_{\text{noiseless}}(\cdot|\mathbf{x}_1) = p_{\text{noiseless}}(\cdot|\mathbf{x}_2) \Leftrightarrow \mathbf{x}_1 = \mathbf{G}(\mathbf{x}_2)$.

A.6.3 Theoretical Guarantee of Recovery in the Continuous Domain

In the absence of CTF and shifts the recoverability of $f : \mathbb{R}^3 \to \mathbb{R}$ from its 2D projections obtained at unknown random poses is guaranteed by [218, Theorem 3.1]. We first go through the notations described in [218] before we state the required foundational result. We then extend [218, Theorem 3.1] to the case when the CTF and shifts are present.

Notations and Preliminaries

Let $SO(3)$ be the space of the special orthonormal matrices and $\mathscr{D}$ be the Borel $\sigma-$algebra induced using the standard Riemannian metric on $SO(3)$. Then, $(SO(3), \mathscr{D})$ describes the measurable space of orthonormal matrices. Let $\Delta_N^W = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\|_2 \leq W\}$ for some $W \in \mathbb{R}^+$. By $(\mathscr{L}_2, \mathscr{B})$, we denote the measurable space of all the square-integrable functions supported in Δ_2^W with Borel $\sigma-$algebra $\mathscr{B}$ induced by the L_2-norm. We denote by $\mathbb{F}$ the set of all the functions supported in Δ_3^W, which are nonnegative and essentially bounded.

For any $f \in \mathbb{F}$ and $\mathbf{A} \in SO(3)$, we denote $\mathbf{y} = \mathscr{P}_{\mathbf{A}}\{f\} = \int_{-\infty}^{\infty} \mathbf{A}f(x_1, x_2, x_3)\, x_3$ where $\mathbf{A}f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{x})$. Let $p_{\mathbf{A}}$ be a probability density on the space $(SO(3), \mathscr{D})$. Note that there is a bijective mapping from $\boldsymbol{\theta}$ in Theorem 1 and $\mathbf{A}$. In fact, $\mathbf{A}$ represents the rotation matrix associated with the projection angle $\boldsymbol{\theta}$.

We denote by Ψ the normalized Haar measure on $(SO(3), \mathscr{D})$ and by $\Psi_{\mathbf{A}}$ the measure associated with $p_{\mathbf{A}}$ such that $\Psi_{\mathbf{A}}[\cdot] = \int_{(\mathbf{a} \in \cdot)} p_{\mathbf{A}}(\mathbf{a}) \Psi[\mathbf{a}]$.

For a given $f \in \mathbb{F}$, the density $p_{\mathbf{A}}$ induces a probability measure $\mathbb{P}_{\text{proj}}(\cdot|f)$ on the space

$(\mathscr{L}_2, \mathscr{B})$ through the mapping $\mathscr{P}_{\mathbf{A}}\{f\}$ such that

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \Psi_{\mathbf{A}}[\{\mathbf{A} \in SO(3) : \mathscr{P}_{\mathbf{A}}\{f\} \in \cdot\}]. \tag{A.33}$$

When $p_{\mathbf{A}}$ is uniform on $SO(3)$, one has that

$$\mathbb{P}_{\text{proj}}(\cdot|f) = \mathbb{P}_{\text{proj}}(\cdot|\mathbf{R}f),\ \forall f \in \mathbb{F} \text{ and } \mathbf{R} \in O(3), \tag{A.34}$$

where O_3 is the space of all orthogonal matrices such that $\det\mathbf{A} \in \{-1, 1\}$. The invariance in (A.34) is true since

$$\begin{aligned}
\mathbb{P}_{\text{proj}}(\cdot|f) &= \Psi[\{\mathbf{A} \in SO(3) : \mathscr{P}_{\mathbf{A}}\{f\} = \cdot\}] \\
&= \Psi[\{\mathbf{A} \in SO(3) : \mathscr{P}_{\mathbf{R}^{-1}\mathbf{A}}\{\mathbf{R}f\} = \cdot\}] \\
&= \Psi[\{\mathbf{R}\mathbf{A}' \in SO(3) : \mathscr{P}_{\mathbf{A}'}\{\mathbf{R}f\} = \cdot\}] \\
&= \Psi[\{\mathbf{A}' \in SO(3) : \mathscr{P}_{\mathbf{A}'}\{\mathbf{R}f\} = \cdot\}]
\end{aligned} \tag{A.35}$$

where $\mathbf{A}' = \mathbf{R}^{-1}\mathbf{A}$ and the last equality follows from the right invariance of Haar measure. We define $G\{\mathbb{F}\} = \{\gamma_A : A \in O_3\}$ such that

$$(\gamma_A f)(\cdot) = f(\mathbf{A}^{-1}\cdot), \forall \mathbf{A} \in O(3), f \in \mathbb{F}. \tag{A.36}$$

We define the shape $[f]$ as an orbit of f under the influence of G such that $[f] = \{\gamma_{\mathbf{A}} f : \gamma_A \in G\}$. When $p_{\mathbf{A}}$ is uniform, the shape $[f]$ is composed of all the rotations and reflections of f.

We can now restate [218, Theorem 3.1]. We discuss here the sketch of the proof given in [218].

Theorem 5 ([218, Theorem 3.1])**.** *Let $p_{\mathbf{A}}$ be any bounded distribution on $SO(3)$ and let the assumptions of Theorem 4 be true; then, $\forall f, g \in \mathbb{F}$,*

$$[f] \neq [g] \Rightarrow \mathbb{P}_{\text{proj}}(\cdot|f) \perp \mathbb{P}_{\text{proj}}(\cdot|g). \tag{A.37}$$

Sketch of the Proof. Without loss of generality, we provide the sketch of the proof for the case when $p_{\mathbf{A}}$ is uniform. For the case when $p_{\mathbf{A}}$ is nonuniform the argument remains the same provided that $\Psi_{\mathbf{A}}$ associated with the non-uniform distribution $p_{\mathbf{A}}$ is absolutely continuous with respect to Ψ ($\Psi_{\mathbf{A}} \ll \Psi$). This has been stated in [218]. Since we assume $p_{\mathbf{A}}$ to be bounded, this condition is satisfied. The only difference here with respect to the uniform distribution is that the orbit of f and g are more restricted than O(3).

The proof first uses in [219, Proposition 7.8] which we restate here as Proposition 2.

Proposition 2 ([219, Proposition 7.8])**.** *Let $f \in \mathbb{F}$ and let $S_{\mathbf{A}}$ be an uncountably infinite subset of $SO(3)$, then f is determined by the collection $\{\mathscr{P}_A\{f\}\}_{\mathbf{A} \in S_{\mathbf{A}}}$ ordered with respect to $\mathbf{A} \in S_{\mathbf{A}}$.*

Note that this proposition assumes that the angle of the projections are known. Although in our case the angles are unknown, we shall see that this proposition will be useful.

We now want to determine how different $\mathbb{P}_{\text{proj}}(.|f)$ and $\mathbb{P}_{\text{proj}}(.|g)$ are for any given f and g. For this, we use the equality

$$\text{TV}(\mathbb{P}_1, \mathbb{P}_2) = 2 \inf_{\gamma \in \Pi(\mathbb{P}_1, \mathbb{P}_2)} \mathbb{E}_{(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma}[1_{\mathbf{y}_1 \neq \mathbf{y}_2}], \tag{A.38}$$

where TV is the total variation distance and $\Pi(\mathbb{P}_1, \mathbb{P}_2)$ is the set of all the joint distributions $\gamma(\mathbf{y}_1, \mathbf{y}_2)$ whose marginals are $\mathbb{P}_1$ and $\mathbb{P}_2$ [209]. In fact, $\mathbb{E}[1_{\mathbf{y}_1 \neq \mathbf{y}_2}]$ is equal to the probability of the event $\mathbf{y}_1 \neq \mathbf{y}_2$. In our context, this translates into

$$\text{TV}(\mathbb{P}_{\text{proj}}(.|f), \mathbb{P}_{\text{proj}}(.|g)) = 2 \inf_{\gamma \in \Pi(\mathbb{P}_{\text{proj}}(.|f), \mathbb{P}_{\text{proj}}(.|g))} \text{Prob}(\mathbf{y}_1 \neq \mathbf{y}_2), \text{ where } (\mathbf{y}_1, \mathbf{y}_2) \sim \gamma. \tag{A.39}$$

The optimum is achieved at the extremas which are sparse joint distributions and are such that the variable $\mathbf{y}_2$ is a function of $\mathbf{y}_1$. For any arbitrary joint distribution (or coupling) of this form, the proof then assigns a measurable function $h : SO(3) \to SO(3)$ such that $(\mathbf{y}_1, \mathbf{y}_2) = (\mathscr{P}_{\mathbf{A}}\{f\}, \mathscr{P}_{h(\mathbf{A})}\{g\})$ for $\mathbf{A} \sim p_{\mathbf{A}}$.

We can then write that

$$\Psi[\{\mathbf{A} \in SO(3) : \mathscr{P}_{h(\mathbf{A})}\{g\} \in \cdot\}] = \mathbb{P}_{\text{proj}}(\cdot|g). \tag{A.40}$$

The task now is to estimate $\text{Prob}(\mathbf{y}_1 \neq \mathbf{y}_2)$, where $(\mathbf{y}_1, \mathbf{y}_2) = (\mathscr{P}_{\mathbf{A}}\{f\}, \mathscr{P}_{h(\mathbf{A})}\{g\})$ for $\mathbf{A} \sim p_{\mathbf{A}}$.

(Continuous h). When h is continuous, Proposition 2 implies that, if $[f] \neq [g]$, then

$$\Psi[\{\mathbf{A} \in SO(3) : \|\mathscr{P}_{\mathbf{A}}\{f\} - \mathscr{P}_{h(\mathbf{A})}\{g\}\|_2 > 0\}] = 1. \tag{A.41}$$

(General h). When the function h is discontinuous, the proof uses Lusin's theorem to approximate h by a continuous function. Lusin's theorem states that, for any $\delta > 0$, there exists an h_δ such that $h(\mathbf{A}) = h_\delta(\mathbf{A}), \forall \mathbf{A} \in \mathscr{H}_\delta$ and $\Psi[SO(3)|\mathscr{H}_\delta] < \delta$. This then leads to

$$\begin{aligned} \Psi[\{\mathbf{A} \in SO(3) : \|\mathscr{P}_{\mathbf{A}}\{f\} - \mathscr{P}_{h(\mathbf{A})}\{g\}\|_2 > 0\}] &\geq \Psi(H_\delta) \\ &\geq 1 - \delta. \end{aligned} \tag{A.42}$$

Since δ is arbitrarily small, the event $\{\mathscr{P}_{\mathbf{A}}\{f\} \neq \mathscr{P}_{h(\mathbf{A})}\{g\}\}$ has probability 1.

In conclusion, for any arbitrary coupling, the event $\{\mathscr{P}_{\mathbf{A}}\{f\} \neq \mathscr{P}_{h(\mathbf{A})}\{g\}\}$ has probability 1 if $[f] \neq [g]$. This implies that, when $[f]$ and $[g]$ are not the same, the total-variation distance between $\mathbb{P}_{\text{proj}}(\cdot|f)$ and $\mathbb{P}_{\text{proj}}(\cdot|g)$ is 2. This ensures that the two probability measures are mutually singular meaning that the intersection of their support has zero measure. This concludes the proof.

A.6.4 Noiseless CTF-Modulated Projections

We now extend the previous result to the case when the CTF is present. We assume that $\mathbf{c} \sim p_{\mathbf{c}}$ such that the support of $p_{\mathbf{c}}$ is in some bounded region $\mathscr{C} \subset \mathbb{R}^3$. We denote $\Psi_{\mathbf{c}}[\cdot]$ as the measure associated with $p_{\mathbf{c}}$ on the space $\mathscr{C}$.

We denote by $(SO(3) \times \mathscr{C})$ the product space of $SO(3)$ and $\mathscr{C}$, while we denote by $\Psi_{\mathbf{A},\mathbf{c}}$ the measure on this product space. We then define

$$\mathbb{P}_{\text{proj,CTF}}(\cdot|f) = \Psi_{\mathbf{A},\mathbf{c}}[\{(\mathbf{A},\mathbf{c}) \in (SO(3) \times \mathscr{C}) : \mathbf{C}_{\mathbf{c}} * \mathscr{P}_{\mathbf{A}}\{f\} \in \cdot\}], \tag{A.43}$$

where $\mathbf{C}_{\mathbf{c}}$ is the space-domain CTF.

Theorem 6. *Let $p_{\mathbf{A}}$ be a bounded probability distribution on $SO(3)$, $p_{\mathbf{c}}$ be a distribution of the CTF with parameters $\mathbf{c} \in \mathscr{C}$, and let the assumptions of Theorem 4 be true; then, $\forall\, f, g \in \mathbb{F}$,*

$$[f] \neq [g] \Rightarrow \mathbb{P}_{\text{proj,CTF}}(\cdot|f) \perp \mathbb{P}_{\text{proj,CTF}}(\cdot|g). \tag{A.44}$$

Proof. Similarly to the previous proof, we show that the TV distance between $\mathbb{P}_{\text{proj,CTF}}(\cdot|f)$ and $\mathbb{P}_{\text{proj,CTF}}(\cdot|g)$ is 2 when $[f]$ and $[g]$ are distinct. For simplification, we assume that $p_{\mathbf{A}}$ is uniform. (When this is not the case the proof essentially remains the same.) We need to show that $\text{Prob}(\mathbf{y}_1 \neq \mathbf{y}_2) = 1$, where $(\mathbf{y}_1, \mathbf{y}_2) \sim \gamma$ for any arbitrary coupling γ of $\mathbb{P}_{\text{proj,CTF}}(\cdot|f)$ and $\mathbb{P}_{\text{proj,CTF}}(\cdot|g)$. For an arbitrary coupling γ such that $\text{Prob}(\mathbf{y}_1 \neq \mathbf{y}_2)$ is minimum, we again assign $h : (SO(3) \times \mathscr{C}) \to (SO(3) \times \mathscr{C})$ such that

$$(\mathbf{y}_1, \mathbf{y}_2) = (\mathbf{C}_{\mathbf{c}} * \mathscr{P}_{\mathbf{A}}\{f\}, \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathscr{P}_{h_0(\mathbf{A},\mathbf{c})}\{g\}), \tag{A.45}$$

where $\mathbf{A} \sim p_{\mathbf{A}}, \mathbf{c} \sim p_{\mathbf{c}}$ and where $h_0 : (SO(3) \times \mathscr{C}) \to SO(3)$ and $h_1 : (SO(3) \times \mathscr{C}) \to \mathscr{C}$ are such that $h(\mathbf{A},\mathbf{c}) = (h_0(\mathbf{A},\mathbf{c}), h_1(\mathbf{A},\mathbf{c}))$. This implies that

$$\mathbb{P}_{\text{proj,CTF}}(\cdot|g) = \Psi_{\mathbf{A},\mathbf{c}}[\{(\mathbf{A},\mathbf{c}) \in (SO(3) \times \mathscr{C}) : \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathscr{P}_{h_0(\mathbf{A},\mathbf{c})}\{g\} \in \cdot\}]. \tag{A.46}$$

We now show that, for any h, the event $\{\mathbf{y}_1 \neq \mathbf{y}_2\}$ has probability 1.

(Continuous h). We first assume that h is continuous and use the same kind of technique as in the proof of [218, Theorem 3.1].

Since $SO(3)$ is transitive, we can write that

$$h(\mathbf{A},\mathbf{c}) = (\mathbf{A}\Gamma_{\mathbf{A},\mathbf{c}}, h_1(\mathbf{A},\mathbf{c})). \tag{A.47}$$

As h is continuous, so is $\Gamma_{\mathbf{A},\mathbf{c}}$. Let $\{\mathscr{A}_n^m \times \mathscr{C}_n^m\}_{m=1}^n$ be a collection of n disjoint sets which creates the partition of $(SO(3) \times \mathscr{C})$. These partitions are such that for any m, there exists a k_m such that $\{\mathscr{A}_{n+1}^m \times \mathscr{C}_{n+1}^m\} \subset \{\mathscr{A}_n^{k_m} \times \mathscr{C}_n^{k_m}\}$. This means that, as n increases, the partitions become

finer. We now define

$$h_n(\mathbf{A},\mathbf{c}) = (\mathbf{A}\Gamma_n^m, h_{n,1}^m(\mathbf{A},\mathbf{c}))\ \forall\ (\mathbf{A},\mathbf{c}) \in \{\mathscr{A}_n^m \times \mathscr{C}_n^m\}, \tag{A.48}$$

such that

$$\Gamma_n^m = \arg\min \Gamma \in \{\Gamma_{\mathbf{A},\mathbf{c}} : (\mathbf{A},\mathbf{c}) \in \{\bar{\mathscr{A}}_n^m \times \bar{\mathscr{C}}_n^m\}\} \min_{(\mathbf{A},\mathbf{c})\in\{\bar{\mathscr{A}}_n^m \times \bar{\mathscr{C}}_n^m\}} \|\mathscr{P}_{\mathbf{A}}\{f\} - \mathscr{P}_{\mathbf{A}\Gamma}\{g\}\|, \tag{A.49}$$

where $\bar{\mathscr{A}}_n^m$ and $\bar{\mathscr{C}}_n^m$ are the closures of $\mathscr{A}_n^m$ and $\mathscr{C}_n^m$, respectively. The sequence h_n converge to h as $n \to \infty$. We denote

$$\mathscr{K} = \{(\mathbf{A},\mathbf{c}) \in (SO(3)\times\mathscr{C}) : \|\mathbf{C}_{\mathbf{c}} * \mathscr{P}_{\mathbf{A}}\{f\} - \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathscr{P}_{\mathbf{A}\Gamma_{\mathbf{A}}}\{g\}\| > 0\}, \tag{A.50}$$

$$\mathscr{K}_n = \{(\mathbf{A},\mathbf{c}) \in (\mathscr{A}_n^m \times \mathscr{C}_n^m) : \|\mathbf{C}_{\mathbf{c}} * \mathscr{P}_{\mathbf{A}}\{f\} - \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathscr{P}_{\mathbf{A}}\{\Gamma_n^m g\}\| > 0\}. \tag{A.51}$$

Similarly to [218, Theorem 3.1], we can then show that

$$\Psi_{\mathbf{A},\mathbf{d}}[\mathscr{K}] = \lim_{n\to\infty} \sum_{m=1}^{m=n} \Psi_{\mathbf{A},\mathbf{d}}[\mathscr{K}_m]. \tag{A.52}$$

We invoke Proposition 3, which gives that $\Psi_{\mathbf{A},\mathbf{c}}[\mathscr{K}_n] = \Psi_{\mathbf{A},\mathbf{c}}[(\mathscr{A}_n^m \times \mathscr{C}_n^m)]$. Therefore, $\Psi_{\mathbf{A},\mathbf{d}}[\mathscr{K}] = \Psi_{\mathbf{A},\mathbf{c}}[(SO(3)\times\mathscr{C})] = 1$. This means that, when h is continuous, the event $\{\mathbf{y}_1 \neq \mathbf{y}_2\}$ has probability 1 if $[f] \neq [g]$.

(General h). When h is discontinuous, we can invoke Lusin's theorem to claim the same, similarly to Theorem 5. This means that, for any h, if $[f] \neq [g]$, then the probability of the event $\{\mathbf{y}_1 \neq \mathbf{y}_2\}$ is 1. Therefore, the TV distance between $\mathbb{P}_{\text{proj,CTF}}(\cdot|f)$ and $\mathbb{P}_{\text{proj,CTF}}(\cdot|g)$ is 2, yielding that $\mathbb{P}_{\text{proj,CTF}}(\cdot|f) \perp \mathbb{P}_{\text{proj,CTF}}(\cdot|g)$. This concludes the proof.

Proposition 3. *Let $f, g \in \mathbb{F}$, $\mathscr{A}' \subseteq SO(3)$, $\mathscr{C}' \subseteq \mathscr{C}$, $\Gamma \in SO(3)$, and*

$$\mathscr{K}' = \{(\mathbf{A},\mathbf{c}) \in (\mathscr{A}' \times \mathscr{C}') : \|\mathbf{C}_{\mathbf{c}} * \mathscr{P}_{\mathbf{A}}\{f\} - \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathscr{P}_{\mathbf{A}}\{\Gamma g\}\| > 0\}. \tag{A.53}$$

Let the assumptions from Theorem 4 be true. Then, if $[f] \neq [g]$, it holds that

$$\Psi_{\mathbf{A},\mathbf{c}}[\mathscr{K}'] = \Psi_{\mathbf{A},\mathbf{c}}[(\mathscr{A}' \times \mathscr{C}')]. \tag{A.54}$$

Proof. We show that $\Psi_{\mathbf{A},\mathbf{c}}[\mathscr{K}'^c] = 0$, where $(\mathscr{K}'^c \cup \mathscr{K}') = (\mathscr{A}' \times \mathscr{C}')$. We define the set $S_{\mathbf{A}} = \{\mathbf{c} \in \mathscr{C}' : \|\mathbf{C}_{\mathbf{c}} * \mathscr{P}_{\mathbf{A}}\{f\} - \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * \mathscr{P}_{\mathbf{A}}\{\Gamma g\}\| = 0\}$. We define $S_{\mathscr{A}''} = \cup_{\mathbf{A}\in\mathscr{A}''} S_{\mathbf{A}}$ for any $\mathscr{A}'' \subseteq \mathscr{A}'$. We define

$$\mathscr{A}_1' = \{\mathbf{A} \in \mathscr{A}' : S_{\mathbf{A}} \text{ is an uncountable set}\}, \tag{A.55}$$

$$\mathscr{A}_2' = \{\mathbf{A} \in \mathscr{A}' : S_{\mathbf{A}} \text{ is a countable non-empty set}\}. \tag{A.56}$$

Note that $\mathscr{K}'^{c} = \cup_{k=1}^{2} \cup_{\mathbf{A}\in\mathscr{A}'_k} (\mathbf{A} \times S_{\mathbf{A}})$. Then,

$$\Psi_{\mathbf{A},\mathbf{c}}[\{\mathscr{K}'^{c}] = \sum_{k=1}^{2} \Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathscr{A}'_k}(\mathbf{A} \times S_{\mathbf{A}})\}] \tag{A.57}$$

We now look at the two cases.

- *(When $S_{\mathbf{A}}$ is uncountable).* For this case, we show that $\Psi[\mathscr{A}'_1] = 0$. The main argument is that if this is not true, then it contradicts $[f] \neq [g]$.

  For the sake of conciseness, we denote $\mathscr{P}_{\mathbf{A}}\{f\}$ by I_f and $\mathscr{P}_{\mathbf{A}}\{\Gamma g\}$ by I_g. For any $\mathbf{A} \in \mathscr{A}'_1$, it holds that

  $$\mathbf{C}_{\mathbf{c}} * I_f = \mathbf{C}_{h_1(\mathbf{A},\mathbf{c})} * I_g, \ \forall \mathbf{c} \in S_{\mathbf{A}}, \tag{A.58}$$
  $$\widehat{\mathbf{C}}_{\mathbf{c}} \cdot \hat{I}_f = \widehat{\mathbf{C}}_{h_1(\mathbf{A},\mathbf{c})} \cdot \hat{I}_g, \ \forall \mathbf{c} \in S_{\mathbf{A}}, \tag{A.59}$$

  where $\widehat{\mathbf{C}}$, $\hat{I}_f$, and $\hat{I}_g$ are the Fourier transforms of $\mathbf{C}$, I_f, and I_g, respectively.

  We define $ze(\hat{I}) = \{\omega \in \mathbb{R}^2 : \hat{I}(\omega) = 0\}$, $\omega_\alpha = \{[(r\cos\alpha, r\sin\alpha)] : r > 0\}$, and $ze_\alpha(\hat{I}) = ze(\hat{I}) \cap \omega_\alpha$. From (A.59), we can write that

  $$ze(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup ze(\hat{I}_f) = ze(\widehat{\mathbf{C}}_{h_1(\mathbf{A},\mathbf{c})}) \cup ze(\hat{I}_g), \ \forall \mathbf{c} \in S_{\mathbf{A}}. \tag{A.60}$$

  Two remarks are in order. Firstly, by assumption 2 of Theorem 4, $ze(\widehat{\mathbf{C}}_{\mathbf{c}_1}) \cap ze(\widehat{\mathbf{C}}_{\mathbf{c}_2}) = \emptyset$ for $\mathbf{c}_1 \neq \mathbf{c}_2$. (Remember that $ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}})$ for any $\alpha \in [0,\pi]$ is nonempty (see "Image Formation Theory).) Secondly, by assumption 3 of Theorem 4, the supports of f and g are compact and nontrivial, so are the supports of I_f and I_g. This means that their Fourier transforms $\hat{I}_f$ and $\hat{I}_g$ are analytic functions, which implies that there are infinitely many α such that the cardinality of the sets $ze_\alpha(\hat{I}_f)$ and $ze_\alpha(\hat{I}_f)$ is countable. We call the set of such α as $\mathscr{S}_\alpha$. Now, we have that

  $$\begin{aligned}
  &ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup ze_\alpha(\hat{I}_f)) = ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap (ze_\alpha(\widehat{\mathbf{C}}_{h_1(\mathbf{A},\mathbf{c})}) \cup ze_\alpha(\hat{I}_g)),\\
  &ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\widehat{\mathbf{C}}_{h_1(\mathbf{A},\mathbf{c})})) \cup (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g)),\\
  &ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g)
  \end{aligned} \tag{A.61}$$

  for all $\mathbf{c} \in S_{\mathbf{A}}$ and $\alpha \in [0,\pi]$.

  We can now write that

  $$\cup_{\mathbf{c}\in S_{\mathbf{A}}} ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cup (ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_f)) = \cup_{\mathbf{c}\in S_{\mathbf{A}}} ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}}) \cap ze_\alpha(\hat{I}_g). \tag{A.62}$$

  for any $\alpha \in S_\alpha$. The set on the left hand side of (A.62) has an uncountably infinite cardinality since there are uncountably many $\mathbf{c} \in \mathscr{S}_{\mathbf{A}}$ and for each $\mathbf{c}$ there are distinct $ze_\alpha(\widehat{\mathbf{C}}_{\mathbf{c}})$. In return, the set in the right hand side of (A.62) is countable for a given $\alpha \in S_\alpha$. Therefore, for any $\alpha \in S_\alpha$, the two sets have different cardinality, which raises a contradiction. The only possible scenario in which (A.60) is true is when $h_1(\mathbf{A},\mathbf{c}) = \mathbf{c}$.

Using (A.59), we infer that $\mathscr{P}_{\mathbf{A}}\{f\} = \mathscr{P}_{\mathbf{A}}\{\Gamma g\}$. Therefore, for any $\mathbf{A} \in \mathscr{A}_1'$, $\mathscr{P}_{\mathbf{A}}\{f\} = \mathscr{P}_{\mathbf{A}}\{\Gamma g\}$. However, $\Psi[\mathscr{A}_1'] = 0$ since, if this is not true, then $[f] = [g]$ by Proposition 2.

Now note that

$$\Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathscr{A}_1'}(\mathbf{A}\times S_{\mathbf{A}})] \leq \underbrace{\Psi[\mathscr{A}_1']}_{0}\underbrace{\Psi_{\mathbf{c}}[\cup_{\mathbf{A}\in\mathscr{A}_1'}S_{\mathbf{A}}]}_{\text{finite}} = 0. \tag{A.63}$$

- *(When $S_{\mathbf{A}}$ is countable and nonempty).* Since $S_{\mathbf{A}}$ is a countable set in this case, its elements have a bijection with natural numbers. We denote this bijection by $b : \mathbb{Z}\times\mathscr{A}_2' \to S_{\mathbf{A}}$. We denote by $q(z) = \cup_{\mathbf{A}\in\mathscr{A}_2'}(\mathbf{A}, b_{\mathbf{A}}(z))$, $\forall z \in \mathbb{Z}$. Note that $q(z)$ is a graph of the function $b(z,\cdot)$. Since it is a graph, $\Psi_{\mathbf{A},\mathbf{c}}[q(z)] = 0$.

  We also have that $\Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathscr{A}_2'}(\mathbf{A}\times S_{\mathbf{A}})] = \Psi_{\mathbf{A},\mathbf{c}}[\sum_{z\in\mathbb{Z}} q(z)]$. The latter vanishes since it is the measure of a countable addition of sets of measure zero. Hence, $\Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathscr{A}_2'}(\mathbf{A}\times S_{\mathbf{A}})] = 0$.

This gives that $\Psi_{\mathbf{A},\mathbf{c}}[\mathscr{K}'^{c}] = \sum_{k=1}^{2}\Psi_{\mathbf{A},\mathbf{c}}[\cup_{\mathbf{A}\in\mathscr{A}_k'}(\mathbf{A}\times S_{\mathbf{A}})] = 0$, which concludes the proof.

A.6.5 Synthetic-Data Experiment

Experimental Dataset. We construct a synthetic cryo-EM dataset that mimics the real β-galactosidase dataset (EMPIAR-10061) from [50]. We generate 41,000 synthetic β-galactosidase particles using our cryo-EM image-formation model (see *Online Methods*). The ground-truth volume is generated by fitting a 5Å density map on the PDB-*5a1a* atomic model in Chimera [200]. This gives a volume of size $(302\times233\times163)$ with pixel size 0.637Å. The volume is then padded, averaged, and downsampled to size $(180\times180\times180)$ with pixel size 1.274Å. This corresponds to a Nyquist resolution of 2.548 Å for the reconstructed volume.

The projection orientations are sampled from a uniform distribution over $SO(3)$, where $SO(3)$ is the group of 3D rotations around the origin of $\mathbb{R}^3$. For the CTF, a micrograph from the EMPIAR-10061 dataset is randomly selected and its CTF parameters are extracted using Relion [10]. We then apply the CTF with these parameters to the clean projections. The parameter B of the envelope function of the CTF (see (A.7)) is chosen such that it decays to 0.2 at the Nyquist frequency.

Noisy projections are obtained by adding a randomly selected background patch from the same micrograph to each noiseless projection. The noise patch is first normalized to zero-mean and scaled. The scaling is such that the ratio of the energy of the signal to the energy of the noise (SNR) is kept at 0.55, which is equivalent to -5.2 dB.

The dataset is randomly divided into two halves. The algorithm is applied separately on both halves to generate the half-maps.

Generator Settings. We reconstruct a volume of size $(180\times180\times180)$ voxels for each half

dataset. The pixel size is 1.274 Å. The volumes are initialized with zeros. The D2 symmetry of β-galactosidase is enforced during reconstruction.

We use our image-formation model to generate realistic projections from the current volume estimate at every CryoGAN iteration. The distribution of the imaging parameters is identical to the one used to generate the dataset. To add the noise on the CTF-modulated projections, we keep the same approach as the one used to generate the dataset. However, we assume that the final SNR of each projection is unknown, leading us to learn the scaling parameter that controls the ratio between the projections and the noise patches.

We apply a binary spherical mask of size $(171 \times 171 \times 171)$ on the learned volume. To handle the sharp transitions at the mask borders, we enforce some clipping constraints on the masked volume. The clipping value increases linearly with the distance from the center of the projection to the border of the mask, while its minimum value at the center increases linearly from 0 to 10% of the maximum protein value with the number of epochs (*i.e.*, a full pass through each half-dataset). This promotes nonnegativity during the initial phases of reconstruction and increases the stability of the algorithm.

Discriminator Architecture. The architecture of the discriminator network is detailed in *Online Methods*. The discriminator is initialized identically for the two half datasets. All projections (*i.e.*, the picked particles and the ones generated by the simulator) are normalized to zero-mean and unit standard-deviation before being fed to the discriminator.

General Settings. The adversarial learning scheme is implemented in Pytorch [207]. For the optimization, we use [205] ($\beta_1 = 0.5$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$) with a learning rate of 10^{-3} and a batch size of 8. The learning rate decreases by 8% at every epoch. The parameter for the gradient-penalty term is kept to $\lambda = 0.001$. The discriminator is trained 4 times for every training of the generator ($n_{\text{discr}} = 4$ in Algorithm 7).

For the back-propagations, the norm of the gradients for the discriminator are clipped to a maximal value of 10^6. For the generator, the gradients for each pixel are clipped to a maximal value of 10^3. The clipping values increase linearly from zero to those maxima in the first two epochs. Doing so improves on the stability of the adversarial learning scheme in the start, in particular, on that of the discriminator. All parameters are tuned for a fixed value range that follows from the normalization of all projections.

Computational Resources. The reconstruction is run on a Nvidia P100 GPU with 18GB memory. Each epoch lasts 10 minutes. The algorithm is run for 16 epochs which, in the current implementation, takes 160 minutes.

A.6.6 Real-Data Experiment

Experimental dataset. The dataset consists of 41,123 β-galactosidase (EMPIAR-10061) particle images extracted from 1,539 micrographs [50]. Particle images of size (384×384) are

downsampled to (192×192), with pixel size of 1.274 Å. This corresponds to a Nyquist resolution of 2.548 Å for a reconstructed volume of size $(180 \times 180 \times 180)$.

The dataset is randomly divided into two halves. The algorithm is applied separately on both halves to generate the half-maps. The defocus and astigmatism parameters of the CTF are estimated from each micrograph using Relion.

Generator Settings. We reconstruct a volume of size $(180 \times 180 \times 180)$ voxels for each half dataset. The pixel size is 1.274 Å. The volumes are initialized with zeros. The D2 symmetry of β-galactosidase is enforced during reconstruction. A uniform distribution is assumed for the orientations. The CTF parameters estimated in Relion are used in the forward model of the CryoEM physics simulator. The parameter B of the envelope function of the CTF (see (A.7)) decays to 0.4 at the Nyquist frequency. The translations (vertical and horizontal) are sampled independently from triangle-shaped distributions.

To handle the noise, we randomly extract (prior to the learning procedure) 41,123 patches of size (384×384) from the background of the micrographs at locations where particles do not appear; this is done by identifying patches with the lowest variance. We extract as many noise patches per micrograph as we have particle images. Each noise patch is then downsampled to size (192×192) and normalized. Then, during run-time, the noise patches are sampled from this collection, scaled, and added to the simulated projections. For consistency, the noise patch added to a given simulated projection is taken from the same micrograph as the one that was used to estimate the CTF parameters previously applied to that specific projection. The scaling operation weighs the contribution of the noise with respect to the projection signal. This is handled by multiplying the pixel values of the noise images and the projection images by two scalars that are learnt throughout the procedure. These two scalar values are the same for every pair of noise/projection images, so that the same amount of extracted noise is added to every simulated projection.

We apply a binary spherical mask of size $(171 \times 171 \times 171)$ on the learned volume. To handle the sharp transitions at the mask borders, we enforce the same clipping constraints on the masked volume as in the synthetic experiment.

Discriminator Architecture. The architecture of the discriminator network is detailed in *Online Methods*. The discriminator is initialized identically for the two half datasets. The projection images (real and synthesized) are smoothed with a Gaussian kernel before being fed to the discriminator. The width of the kernel is initially set at 2 and decreases by 2% at every epoch.

General Settings. The adversarial learning scheme is implemented in Pytorch [207]. For the optimization, we use [205] ($\beta_1 = 0.5$, $\beta_2 = 0.9$, $\epsilon = 10^{-8}$) with a learning rate of 10^{-3} and a batch size of 8. The learning rate decreases by 8% at every epoch. The parameter for the gradient-penalty term is kept to $\lambda = 1$. The discriminator is trained 4 times for every training of the generator ($n_{\text{discr}} = 4$ in Algorithm 7).

For this dataset, the algorithm is first run for 8 epochs (with translation search switched off) to produce a stable low-resolution reconstruction (15Å). The process is then restarted using this volume and run for 12 epochs to obtain a high-resolution volume. In this second stage, we limit the refinement to the high-frequency components above 15Å.

For the back-propagations, the norm of the gradients for the discriminator are clipped to a maximal value of 10^6. For the generator, the gradients for each pixel are clipped to a maximal value of 10^3. The clipping values increase linearly from zero to those maxima in the first two epochs. Doing so improves on the stability of the adversarial learning scheme in the start, in particular, on that of the discriminator. The gradients that correspond to the learning of the scaling ratios between the noise and projection images are clipped to a value of 10.

Computational Resources. The reconstruction is run on a Nvidia P100 GPU with 18GB memory. Each epoch lasts 10 minutes. The algorithm is run for 200 minutes.

Bibliography

[1] J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, and P. Schultz, "Cryo-electron microscopy of vitrified specimens," *Quarterly reviews of biophysics*, vol. 21, no. 2, pp. 129–228, 1988.

[2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[3] V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen, "On instabilities of deep learning in image reconstruction-Does ai come at a cost?," *arXiv preprint arXiv:1902.05300*, 2019.

[4] "Method of the year 2015," *Nature Methods*, vol. 13, no. 1, 2016.

[5] W. Kühlbrandt, "The resolution revolution," *Science*, vol. 343, no. 6178, pp. 1443–1444, 2014.

[6] R. Guerrero-Ferreira, N. M. Taylor, D. Mona, P. Ringler, M. E. Lauer, R. Riek, M. Britschgi, and H. Stahlberg, "Cryo-EM structure of alpha-synuclein fibrils," *Elife*, vol. 7, p. e36402, 2018.

[7] J. Hadamard, *Le probleme de Cauchy et les équations aux dérivées partielles linéaires hyperboliques*, vol. 220. Paris Russian translation), 1932.

[8] C. Sorzano, R. Marabini, J. Velázquez-Muriel, J. R. Bilbao-Castro, S. H. Scheres, J. M. Carazo, and A. Pascual-Montano, "XMIPP: A new generation of an open-source image processing package for electron microscopy," *Journal of Structural Biology*, vol. 148, no. 2, pp. 194–204, 2004.

[9] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, "EMAN2: An extensible image processing suite for electron microscopy," *Journal of Structural Biology*, vol. 157, no. 1, pp. 38–46, 2007.

[10] S. H. Scheres, "RELION: Implementation of a bayesian approach to cryo-EM structure determination," *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012.

[11] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, vol. 375. Springer Science & Business Media, 1996.

[12] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[14] T.-O. Buchholz, M. Jordan, G. Pigino, and F. Jug, "Cryo-CARE: Content-aware image restoration for cryo-transmission electron microscopy data," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 502–506, IEEE, 2019.

[15] S. G. Wolf, L. Houben, and M. Elbaum, "Cryo-scanning transmission electron tomography of vitrified cells," *Nat. Methods*, vol. 11, pp. 423–428, feb 2014.

[16] J. Frank, *Electron tomography: methods for three-dimensional visualization of structures in the cell*. Springer Science & Business Media, 2008.

[17] L. Donati, M. Nilchian, S. Trépout, C. Messaoudi, S. Marco, and M. Unser, "Compressed Sensing for STEM Tomography," *Ultramicroscopy*, vol. 179, pp. 47–56, 2017.

[18] L. Donati, M. Nilchian, C. O. S. Sorzano, and M. Unser, "Fast Multiscale Reconstruction for Cryo-EM," *Journal of Structural Biology*, vol. 204, no. 3, pp. 543–554, 2018.

[19] M. Zehni, L. Donati, E. Soubies, Z. J. Zhao, and M. Unser, "Joint Angular Refinement and Reconstruction for Single-Particle Cryo-EM," *IEEE Transactions on Image Processing*, 2020.

[20] L. Donati, E. Soubies, and M. Unser, "Inner-Loop Free ADMM for Cryo-EM," IEEE International Symposium on Biomedical Imaging (ISBI), 2019.

[21] C. V. Robinson, A. Sali, and W. Baumeister, "The molecular sociology of the cell," *Nature*, vol. 450, no. 7172, pp. 973–982, 2007.

[22] J. Mahamid, S. Pfeffer, M. Schaffer, E. Villa, R. Danev, L. K. Cuellar, F. Förster, A. A. Hyman, J. M. Plitzko, and W. Baumeister, "Visualizing the molecular sociology at the HeLa cell nuclear periphery," *Science*, vol. 351, no. 6276, pp. 969–972, 2016.

[23] J. Frank, *Three-dimensional electron microscopy of macromolecular assemblies: Visualization of biological molecules in their native state*. Oxford University Press, 2006.

[24] Y. Cheng, N. Grigorieff, P. A. Penczek, and T. Walz, "A primer to single-particle cryo-electron microscopy," *Cell*, vol. 161, no. 3, pp. 438–449, 2015.

[25] R. Fernandez-Leiro and S. H. Scheres, "Unravelling biological macromolecules with cryo-electron microscopy," *Nature*, vol. 537, no. 7620, p. 339, 2016.

[26] B. Trus, A. Steven, A. McDowall, M. Unser, J. Dubochet, and R. Podolsky, "Interactions between actin and myosin filaments in skeletal muscle visualized in frozen-hydrated thin sections," *Biophysical journal*, vol. 55, no. 4, pp. 713–724, 1989.

[27] A. Amzallag, C. Vaillant, M. Jacob, M. Unser, J. Bednar, J. D. Kahn, J. Dubochet, A. Stasiak, and J. H. Maddocks, "3D reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy," *Nucleic acids research*, vol. 34, no. 18, pp. e125–e125, 2006.

[28] M. Zhao, S. Wu, Q. Zhou, S. Vivona, D. J. Cipriano, Y. Cheng, and A. T. Brunger, "Mechanistic insights into the recycling machine of the snare complex," *Nature*, vol. 518, no. 7537, pp. 61–67, 2015.

[29] C. Sorzano, R. Marabini, J. Vargas, J. Otón, J. Cuenca-Alba, A. Quintana, J. de la Rosa-Trevín, and J. Carazo, "Interchanging geometry conventions in 3DEM: mathematical context for the development of standards," in *Computational Methods for Three-Dimensional Microscopy Reconstruction*, pp. 7–42, Springer, 2014.

[30] D. Swinehart, "The beer-lambert law," *Journal of chemical education*, vol. 39, no. 7, p. 333, 1962.

[31] F. Natterer, *The mathematics of computerized tomography*. Society for Industrial and Applied Mathematics, jan 2001.

[32] P. Penczek, J. Zhu, R. Schröder, and J. Frank, "Three dimensional reconstruction with contrast transfer compensation from defocus series," *Scanning Microscopy*, vol. 11, pp. 147–154, 1997.

[33] C. Sorzano, L. De La Fraga, R. Clackdoyle, and J. Carazo, "Normalizing projection images: A study of image normalizing procedures for single particle three-dimensional electron microscopy," *Ultramicroscopy*, vol. 101, no. 2-4, pp. 129–138, 2004.

[34] H. Shigematsu and F. Sigworth, "Noise models and cryo-EM drift correction with a direct-electron camera," *Ultramicroscopy*, vol. 131, pp. 61–69, 2013.

[35] W. T. Baxter, R. A. Grassucci, H. Gao, and J. Frank, "Determination of signal-to-noise ratios and spectral SNRs in cryo-em low-dose imaging of molecules," *Journal of structural biology*, vol. 166, no. 2, pp. 126–132, 2009.

[36] N. Kawase, M. Kato, H. Nishioka, and H. Jinnai, "Transmission electron microtomography without the "missing wedge" for quantitative structural analysis," *Ultramicroscopy*, vol. 107, no. 1, pp. 8–15, 2007.

[37] M. F. Hohmann-Marriott, A. A. Sousa, A. A. Azari, S. Glushakova, G. Zhang, J. Zimmerberg, and R. D. Leapman, "Nanoscale 3D cellular imaging by axial scanning transmission electron tomography," *Nat. Methods*, vol. 6, pp. 729–731, oct 2009.

[38] K. Aoyama, T. Takagi, A. Hirase, and A. Miyazawa, "STEM tomography for thick biological specimens," *Ultramicroscopy*, vol. 109, no. 1, pp. 70–80, 2008.

[39] A. Yakushevska, M. Lebbink, W. Geerts, L. Spek, E. van Donselaar, K. Jansen, B. Humbel, J. Post, A. Verkleij, and A. Koster, "STEM tomography in cell biology," *J. Struct. Biol.*, vol. 159, no. 3, pp. 381–391, 2007.

[40] Z. Saghi, M. Benning, R. Leary, M. Macias-Montero, A. Borras, and P. A. Midgley, "Reduced-dose and high-speed acquisition strategies for multi-dimensional electron microscopy," *Adv. Struct. Chem. Imaging*, vol. 1, p. 7, dec 2015.

[41] K. Batenburg, S. Bals, J. Sijbers, C. Kübel, P. Midgley, J. Hernandez, U. Kaiser, E. Encina, E. Coronado, and G. Van Tendeloo, "3D imaging of nanomaterials by discrete tomography," *Ultramicroscopy*, vol. 109, no. 6, pp. 730–740, 2009.

[42] S. Van Aert, K. J. Batenburg, M. D. Rossell, R. Erni, and G. Van Tendeloo, "Three-dimensional atomic imaging of crystalline nanoparticles," *Nature*, vol. 470, pp. 374–377, feb 2011.

[43] R. Leary, Z. Saghi, P. A. Midgley, and D. J. Holland, "Compressed sensing electron tomography," *Ultramicroscopy*, vol. 131, pp. 70–91, 2013.

[44] Z. Saghi, D. J. Holland, R. Leary, A. Falqui, G. Bertoni, A. J. Sederman, L. F. Gladden, and P. A. Midgley, "Three-dimensional morphology of iron oxide nanoparticles with reactive concave surfaces. a compressed sensing-electron tomography (CS-ET) approach," *Nano Lett.*, vol. 11, pp. 4666–4673, nov 2011.

[45] O. Nicoletti, F. de la Peña, R. K. Leary, D. J. Holland, C. Ducati, and P. A. Midgley, "Three-dimensional imaging of localized surface plasmon resonances of metal nanoparticles," *Nature*, vol. 502, pp. 80–84, oct 2013.

[46] B. Goris, S. Bals, W. Van den Broek, E. Carbó-Argibay, S. Gómez-Graña, L. M. Liz-Marzán, and G. Van Tendeloo, "Atomic-scale determination of surface facets in gold nanorods," *Nat. Mater.*, vol. 11, pp. 930–935, oct 2012.

[47] J. P. Buban, Q. Ramasse, B. Gipson, N. D. Browning, and H. Stahlberg, "High-resolution low-dose scanning transmission electron microscopy.," *J. Electron Microsc. (Tokyo).*, vol. 59, no. 2, pp. 103–12, 2010.

[48] "The potential for bayesian compressive sensing to significantly reduce electron dose in high-resolution STEM images," *Microscopy*, vol. 63, no. 1, pp. 41–51, 2014.

[49] E. Nogales, "The development of cryo-EM into a mainstream structural biology technique," *Nature methods*, vol. 13, no. 1, pp. 24–27, 2016.

[50] A. Bartesaghi, A. Merk, S. Banerjee, D. Matthies, X. Wu, J. L. Milne, and S. Subramaniam, "2.2 å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor," *Science*, vol. 348, no. 6239, pp. 1147–1151, 2015.

[51] E. Pettersen, T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin, "UCSF Chimera-A visualization system for exploratory research and analysis," *Journal of computational chemistry*, vol. 25, no. 13, pp. 1605–12, 2004.

[52] J. Frank, "New opportunities created by single-particle cryo-EM: the mapping of conformational space," 2018.

[53] J. Frank and T. Wagenknecht, "Automatic selection of molecular images from electron micrographs," *Ultramicroscopy*, vol. 12, no. 3, pp. 169–175, 1983.

[54] J. Z. Chen and N. Grigorieff, "SIGNATURE: a single-particle selection system for molecular electron microscopy," *Journal of structural biology*, vol. 157, no. 1, pp. 168–173, 2007.

[55] Z. Ripstein and J. Rubinstein, "Processing of cryo-EM movie data," in *Methods in enzymology*, vol. 579, pp. 103–124, Elsevier, 2016.

[56] C. V. Sindelar and N. Grigorieff, "An adaptation of the wiener filter suitable for analyzing images of isolated single particles," *Journal of structural biology*, vol. 176, no. 1, pp. 60–74, 2011.

[57] S. H. Scheres, "Semi-automated selection of cryo-EM particles in RELION-1.3," *Journal of structural biology*, vol. 189, no. 2, pp. 114–122, 2015.

[58] J. Frank, B. Shimkin, and H. Dowse, "SPIDER—a modular software system for electron image processing," *Ultramicroscopy*, vol. 6, no. 4, pp. 343–357, 1981.

[59] N. Grigorieff, "FREALIGN: High-resolution refinement of single particle structures," *Journal of Structural Biology*, vol. 157, no. 1, pp. 117–125, 2007.

[60] M. Hohn, G. Tang, G. Goodyear, P. R. Baldwin, Z. Huang, P. A. Penczek, C. Yang, R. M. Glaeser, P. D. Adams, and S. J. Ludtke, "SPARX, a new environment for cryo-EM image processing," *Journal of Structural Biology*, vol. 157, no. 1, pp. 47–55, 2007.

[61] J. de la Rosa-Trevín, A. Quintana, L. del Cano, A. Zaldívar, I. Foche, J. Gutiérrez, J. Gómez-Blanco, J. Burguet-Castell, J. Cuenca-Alba, V. Abrishami, J. Vargas, J. Otón, G. Sharov, J. Vilas, J. Navas, P. Conesa, M. Kazemi, R. Marabini, C. Sorzano, and J. Carazo, "Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy," *Journal of Structural Biology*, vol. 195, no. 1, pp. 93 – 99, 2016.

[62] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination," *Nature Methods*, vol. 14, pp. 290–296, Feb. 2017.

[63] M. Van Heel, "Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction," *Ultramicroscopy*, vol. 21, no. 2, pp. 111–123, 1987.

[64] A. Singer, R. R. Coifman, F. J. Sigworth, D. W. Chester, and Y. Shkolnisky, "Detecting consistent common lines in cryo-EM by voting," *Journal of structural biology*, vol. 169, no. 3, pp. 312–322, 2010.

[65] L. Wang, A. Singer, and Z. Wen, "Orientation determination of cryo-EM images using least unsquared deviations," *SIAM journal on imaging sciences*, vol. 6, no. 4, pp. 2450–2483, 2013.

[66] I. Greenberg and Y. Shkolnisky, "Common lines modeling for reference free ab-initio reconstruction in cryo-EM," *Journal of structural biology*, vol. 200, no. 2, pp. 106–117, 2017.

[67] G. Pragier and Y. Shkolnisky, "A common lines approach for ab initio modeling of cyclically symmetric molecules," *Inverse Problems*, vol. 35, no. 12, p. 124005, 2019.

[68] P. A. Penczek, R. A. Grassucci, and J. Frank, "The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles," *Ultramicroscopy*, vol. 53, no. 3, pp. 251–270, 1994.

[69] T. Baker and R. Cheng, "A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy," *Journal of Structural Biology*, vol. 116, no. 1, pp. 120–130, 1996.

[70] J. Carazo, C. Sorzano, J. Otón, R. Marabini, and J. Vargas, "Three-dimensional reconstruction methods in single particle analysis from transmission electron microscopy data," *Archives of Biochemistry and Biophysics*, vol. 581, pp. 39–48, 2015.

[71] "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[72] F. J. Sigworth, "A maximum-likelihood approach to single-particle image refinement," *Journal of structural biology*, vol. 122, no. 3, pp. 328–339, 1998.

[73] S. Scheres, "A bayesian view on cryo-EM structure determination," *Journal of molecular biology*, vol. 415, no. 2, pp. 406–418, 2012.

[74] A. Punjani, M. Brubaker, and D. Fleet, "Building proteins in a day: Efficient 3D molecular structure estimation with electron cryo microscopy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 706–718, 2017.

[75] C. O. S. Sorzano, R. Marabini, A. Pascual-Montano, S. H. Scheres, and J. M. Carazo, "Optimization problems in electron microscopy of single particles," *Annals of Operations Research*, vol. 148, no. 1, pp. 133–165, 2006.

[76] R. Henderson, A. Sali, M. L. Baker, B. Carragher, B. Devkota, K. H. Downing, E. H. Egelman, Z. Feng, J. Frank, N. Grigorieff, W. Jiang, S. J. Ludtke, O. Medalia, P. A. Penczek, P. B. Rosenthal, M. G. Rossmann, M. F. Schmid, G. F. Schröder, A. C. Steven, D. L. Stokes,

J. D. Westbrook, W. Wriggers, H. Yang, J. Young, H. M. Berman, W. Chiu, G. J. Kleywegt, and C. L. Lawson, "Outcome of the first electron microscopy validation task force meeting," *Structure*, vol. 20, no. 2, pp. 205–214, 2012.

[77] T. Bendory, A. Bartesaghi, and A. Singer, "Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities," *IEEE Signal Processing Magazine*, vol. 37, no. 2, pp. 58–76, 2020.

[78] P. A. Penczek, R. Renka, and H. Schomberg, "Gridding-based direct fourier inversion of the three-dimensional ray transform," *Journal of the Optical Society of America A*, vol. 21, no. 4, pp. 499–509, 2004.

[79] V. Abrishami, J. Bilbao-Castro, J. Vargas, R. Marabini, J. Carazo, and C. Sorzano, "A fast iterative convolution weighting approach for gridding-based direct fourier three-dimensional reconstruction with correction for the contrast transfer function," *Ultramicroscopy*, vol. 157, pp. 79–87, 2015.

[80] C. O. S. Sorzano, J. Vargas, J. Otón, J. de la Rosa-Trevín, J. Vilas, M. Kazemi, R. Melero, L. Del Caño, J. Cuenca, P. Conesa, and J. Gomez-Blanco, "A survey of the use of iterative reconstruction algorithms in electron microscopy," *BioMed research international*, vol. 2017, 2017.

[81] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *Journal of Theoretical Biology*, vol. 29, no. 3, pp. 471–481, 1970.

[82] R. Marabini, G. T. Herman, and J. M. Carazo, "3D reconstruction in electron microscopy using art with smooth spherically symmetric volume elements (blobs)," *Ultramicroscopy*, vol. 72, no. 1-2, pp. 53–65, 1998.

[83] F. J. Sigworth, P. C. Doerschuk, J.-M. Carazo, and S. H. Scheres, "An introduction to maximum-likelihood methods in cryo-EM," in *Methods in Enzymology*, vol. 482, pp. 263–294, Elsevier, 2010.

[84] M. Li, G. Xu, C. O. Sorzano, F. Sun, and C. L. Bajaj, "Single-particle reconstruction using L2-gradient flow," *Journal of structural biology*, vol. 176, no. 3, pp. 259–267, 2011.

[85] J. Zhu, P. A. Penczek, R. Schröder, and J. Frank, "Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: Procedure and application to the 70s escherichiacoli ribosome," *Journal of Structural Biology*, vol. 118, no. 3, pp. 197–219, 1997.

[86] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[87] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine*, vol. 34, pp. 85–95, Nov. 2017.

[88] G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica*, vol. 6, no. 8, pp. 921–943, 2019.

[89] T. Bepler, A. J. Noble, and B. Berger, "Topaz-Denoise: general deep denoising models for cryoem," *bioRxiv*, p. 838920, 2019.

[90] F. Wang, H. Gong, G. Liu, M. Li, C. Yan, T. Xia, X. Li, and J. Zeng, "DeepPicker: a deep learning approach for fully automated particle picking in cryo-EM," *Journal of structural biology*, vol. 195, no. 3, pp. 325–336, 2016.

[91] Y. Zhu, Q. Ouyang, and Y. Mao, "A deep convolutional neural network approach to single-particle recognition in cryo-electron microscopy," *BMC bioinformatics*, vol. 18, no. 1, p. 348, 2017.

[92] D. Tegunov and P. Cramer, "Real-time cryo-EM data pre-processing with Warp," *BioRxiv*, p. 338558, 2018.

[93] T. Wagner, F. Merino, M. Stabrin, T. Moriya, C. Antoni, A. Apelbaum, P. Hagel, O. Sitsel, T. Raisch, D. Prumbaum, *et al.*, "SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM," *Communications Biology*, vol. 2, no. 1, p. 218, 2019.

[94] T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble, and B. Berger, "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs," *Nature methods*, pp. 1–8, 2019.

[95] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. description of system," *The British journal of radiology*, vol. 46, no. 552, pp. 1016–1022, 1973.

[96] D. De Rosier and A. Klug, "Reconstruction of three dimensional structures from electron micrographs," *Nature*, vol. 217, no. 5124, pp. 130–134, 1968.

[97] G. Ramachandran and A. Lakshminarayanan, "Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of fourier transforms," *Proceedings of the National Academy of Sciences*, vol. 68, no. 9, pp. 2236–2240, 1971.

[98] M. Unser, "Sampling—50 years after Shannon," *Proceedings IEEE*, vol. 88, pp. 569–587, apr 2000.

[99] R. M. Lewitt, "Multidimensional digital image representations using generalized kaiser–bessel window functions," *Journal of the Optical Society of America A*, vol. 7, no. 10, pp. 1834–1846, 1990.

[100] M. Nilchian, J. P. Ward, C. Vonesch, and M. Unser, "Optimized kaiser–bessel window functions for computed tomography," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3826–3833, 2015.

[101] I. Daubechies, *Ten lectures on wavelets*, vol. 61. Siam, 1992.

[102] M. Nilchian, "High performance reconstruction framework for straight ray tomography," tech. rep., EPFL, 2015.

[103] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Princeton university bulletin*, pp. 49–52, 1902.

[104] A. Tarantola, *Inverse problem theory and methods for model parameter estimation*, vol. 89. siam, 2005.

[105] M. Unser and P. D. Tafti, *An introduction to sparse stochastic processes*. Cambridge University Press, 2014.

[106] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, p. 969, 2007.

[107] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, vol. 87. Springer Science & Business Media, 2013.

[108] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[109] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Foundations and Trends® in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.

[110] L. Condat, "A generic proximal algorithm for convex optimization—application to total variation minimization," *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 985–989, 2014.

[111] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. Vũ, "A forward-backward view of some primal-dual optimization methods in image recovery," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4141–4145, IEEE, 2014.

[112] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[113] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[114] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical x-ray ct reconstruction," *IEEE Trans. Med. Imaging*, vol. 31, pp. 677–688, mar 2012.

[115] P. Combettes and J. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse Problems*, vol. 24, no. 6, p. 065014, 2008.

[116] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212, Springer, 2011.

[117] M. S. Almeida and M. Figueiredo, "Deconvolving images with unknown boundaries using the alternating direction method of multipliers," *IEEE Transactions on Image processing*, vol. 22, no. 8, pp. 3074–3086, 2013.

[118] M. Unser, E. Soubies, F. Soulez, M. McCann, and L. Donati, "GlobalBioIm: A unifying computational framework for solving inverse problems," in *Computational Optical Sensing and Imaging*, pp. CTu1B–1, Optical Society of America, 2017.

[119] E. Soubies, F. Soulez, M. McCann, T.-A. Pham, L. Donati, T. Debarre, D. Sage, and M. Unser, "Pocket guide to solve inverse problems with globalbioim," *Inverse Problems*, 2019.

[120] L. Donati, E. Soubies, and M. Unser, "User-friendly building of reconstruction algorithms with globalbioim," in *Quantitative BioImaging Conference 2020*, 2020.

[121] T. F. Chan and J. Shen, "Nontexture inpainting by curvature-driven diffusions," *Journal of Visual Communication and Image Representation*, vol. 12, no. 4, pp. 436–449, 2001.

[122] S. G. Wolf, L. Houben, and M. Elbaum, "Cryo-scanning transmission electron tomography of vitrified cells," *Nature methods*, vol. 11, no. 4, p. 423, 2014.

[123] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, pp. 21–30, mar 2008.

[124] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging.," *Magn. Reson. Med.*, vol. 58, pp. 1182–95, dec 2007.

[125] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, "Faster STORM using compressed sensing," *Nat. Methods*, vol. 9, pp. 721–723, apr 2012.

[126] R. M. Willett, R. F. Marcia, and J. M. Nichols, "Compressed sensing for practical optical imaging systems: a tutorial," *Optical Engineering*, vol. 50, no. 7, pp. 072601–072601–13, 2011.

[127] A. Pacureanu, M. Langer, E. Boller, P. Tafforeau, and F. Peyrin, "Nanoscale imaging of the bone cell network with synchrotron x-ray tomography: optimization of acquisition setup," *Medical Physics*, vol. 39, no. 4, pp. 2229–2238, 2012.

[128] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[129] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[130] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[131] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[132] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, *A l1-Unified Variational Framework for Image Restoration*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[133] D. S. Taubman and M. W. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2001.

[134] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Med. Phys.*, vol. 35, no. 2, p. 660, 2008.

[135] J. Provost and F. Lesage, "The Application of Compressed Sensing for Photo-Acoustic Tomography," *IEEE Trans. Med. Imaging*, vol. 28, pp. 585–594, apr 2009.

[136] H. S. Anderson, J. Ilic-Helms, B. Rohrer, J. Wheeler, and K. Larson, "Sparse imaging for fast electron microscopy," *Proc. SPIE-IS&T Electron. Imaging*, vol. 8657, p. 86570C, 2013.

[137] K. Song, L. R. Comolli, and M. Horowitz, "Removing high contrast artifacts via digital inpainting in cryo-electron tomography: An application of compressed sensing," *J. Struct. Biol.*, vol. 178, no. 2, pp. 108–120, 2012.

[138] A. Stoschek and R. Hegerl, "Denoising of electron tomographic reconstructions using multiscale transformations," *J. Struct. Biol.*, vol. 120, no. 3, pp. 257–265, 1997.

[139] C. Sorzano, E. Ortiz, M. López, and J. Rodrigo, "Improved Bayesian image denoising based on wavelets with applications to electron microscopy," *Pattern Recognit.*, vol. 39, no. 6, pp. 1205–1213, 2006.

[140] C. Vonesch, L. Wang, Y. Shkolnisky, and A. Singer, "Fast wavelet-based single-particle reconstruction in Cryo-EM," in *2011 IEEE Int. Symp. Biomed. Imaging From Nano to Macro*, pp. 1950–1953, IEEE, mar 2011.

[141] M. Abramoff, P. J. Magalhães, and S. J. Ram, "Image processing with ImageJ," *Biophotonics Int.*, vol. 11, no. 7, pp. 36–42, 2004.

[142] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[143] R. Coifman and D. Donoho, "Wavelets and statistics," *Lecture notes in Statistics. In: Translation-invariant de-noising. Antoniadis, A. and Oppenheim, G.(eds.)*, pp. 125–150, 1995.

[144] M. A. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.

[145] M. Guerquin-Kern, M. Haberlin, K. P. Pruessmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE transactions on medical imaging*, vol. 30, no. 9, pp. 1649–1660, 2011.

[146] U. Kamilov, E. Bostan, and M. Unser, "Wavelet shrinkage with consistent cycle spinning generalizes total variation denoising," *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 187–190, 2012.

[147] M. Unser and P. D. Tafti, *An introduction to sparse stochastic processes*. Cambridge University Press, 2014.

[148] N. Parikh, S. Boyd, *et al.*, "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[149] S. Trepout, C. Messaoudi, S. Perrot, P. Bastin, and S. Marco, "Scanning transmission electron microscopy through-focal tilt-series on biological specimens.," *Micron*, vol. 77, pp. 9–15, oct 2015.

[150] A. Wirgin, "The inverse crime," *arXiv preprint math-ph/0401050*, 2004.

[151] C. E. Chavez, F. Alonzo-Atienza, and D. Alvarez, "Avoiding the inverse crime in the inverse problem of electrocardiography: estimating the shape and location of cardiac ischemia," in *Computing in Cardiology Conference (CinC), 2013*, pp. 687–690, IEEE, 2013.

[152] M. Van Heel and M. Schatz, "Fourier shell correlation threshold criteria," *Journal of structural biology*, vol. 151, no. 3, pp. 250–262, 2005.

[153] R. Brun and Schönenberger, "Cultivation and in vitro cloning or procyclic culture forms of Trypanosoma brucei in a semi-defined medium. Short communication.," *Acta Trop.*, vol. 36, pp. 289–92, sep 1979.

[154] W. Saxton, W. Baumeister, and M. Hahn, "Three-dimensional reconstruction of imperfect two-dimensional crystals," *Ultramicroscopy*, vol. 13, pp. 57–70, jan 1984.

[155] M. Unser and A. Aldroubi, "A review of wavelets in biomedical applications," *Proceedings of the IEEE*, vol. 84, no. 4, pp. 626–638, 1996.

[156] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Engineer*, vol. 29, no. 6, pp. 33–41, 1984.

[157] P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Transactions on Image Processing*, vol. 7, no. 1, pp. 27–41, 1998.

[158] J. Dengler, "A multi-resolution approach to the 3D reconstruction from an electron microscope tilt series solving the alignment problem without gold particles," *Ultramicroscopy*, vol. 30, no. 3, pp. 337–348, 1989.

[159] M. Desco, J. Hernandez, A. Santos, and M. Brammer, "Multiresolution analysis in fmri: Sensitivity and specificity in the detection of brain activation," *Human Brain Mapping*, vol. 14, no. 1, pp. 16–27, 2001.

[160] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," in *ACM Transactions on Graphics (TOG)*, vol. 25, pp. 787–794, ACM, 2006.

[161] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind image deconvolution: A review," *Digital Signal Processing*, vol. 47, pp. 116–127, 2015.

[162] A. Saad and W. Chiu, "Hierarchical wavelets projection matching for orientation determination of low contrast electron cryomicroscopic images of icosahedral virus particles," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, vol. 4, pp. 2270–2273, IEEE, 2000.

[163] C. Sorzano, S. Jonić, C. El-Bez, J. Carazo, S. De Carlo, P. Thévenaz, and M. Unser, "A multiresolution approach to orientation assignment in 3D electron microscopy of single particles," *Journal of Structural Biology*, vol. 146, no. 3, pp. 381–392, 2004.

[164] R. Henderson, S. Chen, J. Z. Chen, N. Grigorieff, L. A. Passmore, L. Ciccarelli, J. L. Rubinstein, R. A. Crowther, P. L. Stewart, and P. B. Rosenthal, "Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy," *Journal of Molecular Biology*, vol. 413, no. 5, pp. 1028–1046, 2011.

[165] S. H. Scheres and S. Chen, "Prevention of overfitting in cryo-EM structure determination," *Nature Methods*, vol. 9, no. 9, p. 853, 2012.

[166] M. T. McCann, M. Nilchian, M. Stampanoni, and M. Unser, "Fast 3D reconstruction method for differential phase contrast X-ray CT," *Optics Express*, vol. 24, no. 13, pp. 14564–14581, 2016.

[167] A. Delaney and Y. Bresler, "A fast and accurate Fourier algorithm for iterative parallel-beam tomography," *IEEE Transactions on Image Processing*, vol. 5, no. 5, pp. 740–753, 1996.

[168] C. Vonesch, L. Wang, Y. Shkolnisky, and A. Singer, "Fast wavelet-based single-particle reconstruction in cryo-EM," in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 1950–1953, March 2011.

[169] L. Wang, Y. Shkolnisky, and A. Singer, "A Fourier-based approach for iterative 3D reconstruction from cryo-EM images," *arXiv preprint arXiv:1307.5824*, 2013.

[170] A. Bartesaghi, D. Matthies, S. Banerjee, A. Merk, and S. Subramaniam, "Structure of β-galactosidase at 2.2Å resolution obtained by cryo-electron microscopy," *Proceedings of the National Academy of Sciences*, vol. 111, no. 32, pp. 11709–11714, 2014.

[171] M. G. Campbell, D. Veesler, A. Cheng, C. S. Potter, and B. Carragher, "2.8 å resolution reconstruction of the thermoplasma acidophilum 20s proteasome using cryo-electron microscopy," *Elife*, vol. 4, p. e06380, 2015.

[172] C. Sorzano, J. Bilbao-Castro, Y. Shkolnisky, M. Alcorlo, R. Melero, G. Caffarena-Fernández, M. Li, G. Xu, R. Marabini, and J. Carazo, "A clustering approach to multireference alignment of single-particle projections in electron microscopy," *Journal of structural biology*, vol. 171, no. 2, pp. 197–206, 2010.

[173] C. Sorzano, J. Vargas, J. de la Rosa-Trevin, J. Oton, A. Alvarez-Cabrera, V. Abrishami, E. Sesmero, R. Marabini, and J. Carazo, "A statistical approach to the initial volume problem in single particle analysis by electron microscopy," *Journal of Structural Biology*, vol. 189, no. 3, pp. 213–219, 2015.

[174] C. Sorzano, J. Vargas, J. de la Rosa-Trevín, A. Jiménez, D. Maluenda, R. Melero, M. Martínez, E. Ramírez-Aportela, P. Conesa, J. Vilas, *et al.*, "A new algorithm for high-resolution reconstruction of single particles by electron microscopy," *Journal of structural biology*, vol. 204, no. 2, pp. 329–337, 2018.

[175] S. H. Scheres, "RELION: Implementation of a Bayesian approach to cryo-EM structure determination," *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, 2012.

[176] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination," *Nature Methods*, vol. 14, no. 3, p. 290, 2017.

[177] J. de la Rosa-Trevín, J. Otón, R. Marabini, A. Zaldívar, J. Vargas, J. Carazo, and C. Sorzano, "Xmipp 3.0: An improved software suite for image processing in electron microscopy," *Journal of Structural Biology*, vol. 184, no. 2, pp. 321–328, 2013.

[178] N. Grigorieff, "FREALIGN: High-resolution refinement of single particle structures," *Journal of Structural Biology*, vol. 157, no. 1, pp. 117–125, 2007.

[179] F. Sigworth, "Principles of cryo-EM single-particle image processing.," *Microscopy*, vol. 65, no. 1, pp. 57–67, 2016.

[180] E. Soubies, F. Soulez, M. McCann, T.-A. Pham, L. Donati, T. Debarre, D. Sage, and M. Unser, "Pocket guide to solve inverse problems with GlobalBioIm," *Inverse Problems*, 2019.

[181] L. Armijo, "Minimization of functions having Lipschitz continuous first partial derivatives.," *Pacific Journal of Mathematics*, vol. 16, no. 1, pp. 1–3, 1966.

[182] J. Nocedal and S. Wright, *Numerical Optimization*. New York, NY, USA: Springer, second ed., 2006.

[183] Z. Yang and P. A. Penczek, "Cryo-EM image alignment based on nonuniform fast fourier transform," *Ultramicroscopy*, vol. 108, no. 9, pp. 959 – 969, 2008.

[184] J. Lee, P. C. Doerschuk, and J. E. Johnson, "Exact reduced-complexity maximum likelihood reconstruction of multiple 3-d objects from unlabeled unoriented 2-d projections and electron microscopy of viruses," *IEEE Transactions on Image Processing*, vol. 16, pp. 2865–2878, Dec 2007.

[185] Z. Zhao and A. Singer, "Rotationally invariant image representation for viewing direction classification in cryo-EM," *Journal of Structural Biology*, vol. 186, no. 1, pp. 153 – 166, 2014.

[186] G. Laxmikanthan, C. Xu, A. F. Brilot, D. Warren, L. Steele, N. Seah, W. Tong, N. Grigorieff, A. Landy, and G. D. Van Duyne, "Structure of a holliday junction complex reveals mechanisms governing a highly regulated dna transaction," *Elife*, vol. 5, p. e14313, 2016.

[187] X. Qi, P. Schmiege, E. Coutavas, J. Wang, and X. Li, "Structures of human patched and its complex with native palmitoylated sonic hedgehog," *Nature*, vol. 560, no. 7716, pp. 128–132, 2018.

[188] R. Fernandez-Leiro and S. Scheres, "A pipeline approach to single-particle processing in Relion," *Acta Crystallographica Section D*, vol. 73, no. 6, pp. 496–502, 2017.

[189] Y. Zhu, "An augmented ADMM algorithm with application to the generalized lasso problem," *Journal of Computational and Graphical Statistics*, vol. 26, no. 1, pp. 195–204, 2017.

[190] I. Tosic and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.

[191] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[192] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[193] J. B. Kruskal, *Multidimensional scaling*. No. 11, Sage, 1978.

[194] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[195] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[196] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: essential theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, 2015.

[197] D. Koks, *Explorations in mathematical physics: the concepts behind an elegant language*. Springer Science & Business Media, 2006.

[198] B. A. Rosenfeld, *A History of Non-Euclidean Geometry: Evolution of the Concept of a Geometric Space*. Springer New York, Sept. 1988. Google-Books-ID: DRLpAFZM7uwC.

[199] D. Q. Huynh, "Metrics for 3D rotations: Comparison and analysis," *Journal of Mathematical Imaging and Vision*, vol. 35, no. 2, pp. 155–164, 2009.

[200] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, "Ucsf chimera—a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.

[201] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.

[202] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265–283, 2016.

[203] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 539–546, IEEE, 2005.

[204] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *2014 22nd International Conference on Pattern Recognition*, pp. 34–39, IEEE, 2014.

[205] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[206] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, pp. 2672–2680, 2014.

[207] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, pp. 8024–8035, 2019.

[208] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*, pp. 214–223, 2017.

[209] C. Villani, *Optimal transport: old and new*, vol. 338. Springer Science & Business Media, 2008.

[210] G. Peyré, M. Cuturi, *et al.*, "Computational optimal transport," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[211] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in neural information processing systems*, pp. 5767–5777, 2017.

[212] N. Sharon, J. Kileel, Y. Khoo, B. Landa, and A. Singer, "Method of moments for 3D single particle ab initio modeling with non-uniform distribution of viewing angles," *Inverse Problems*, 2019.

[213] M. Vulović, R. B. Ravelli, L. J. van Vliet, A. J. Koster, I. Lazić, U. Lücken, H. Rullgård, O. Öktem, and B. Rieger, "Image formation modeling in cryo-electron microscopy," *Journal of structural biology*, vol. 183, no. 1, pp. 19–32, 2013.

[214] H. Rullgård, L.-G. Öfverstedt, S. Masich, B. Daneholt, and O. Öktem, "Simulation of transmission electron microscope images of biological specimens," *Journal of microscopy*, vol. 243, no. 3, pp. 234–256, 2011.

[215] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," *arXiv:1710.10196*, 2017.

[216] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4401–4410, 2019.

[217] R. Lewitt, "Multidimensional digital image representations using generalized Kaiser–Bessel window functions," *Journal of the Optical Society of America A*, vol. 7, no. 10, pp. 1834–1846, 1990.

[218] V. M. Panaretos *et al.*, "On random tomography with unobservable projection angles," *The Annals of Statistics*, vol. 37, no. 6A, pp. 3272–3306, 2009.

[219] S. Helgason, *The radon transform*, vol. 2. Springer, 1980.

LAURÈNE DONATI

@ laurene.donati@ep?.ch +41 79 283 93 55 Rte des Fontanettes 4 | 3968 Veyras | Switzerland 19 Sep 1991

EDUCATION

PhD in Electrical Engineering

Biomedical Imaging Group, EPFL

Jan 2016 – June 2020 (ongoing) Lausanne, Switzerland

Reconstruction Methods for Cryo-Electron Microscopy:
From Model-based to Data-driven
Thesis Director: Prof. Michael Unser, EPFL

Master of Science (MSc) in Bio-Engineering

School of Life Sciences, EPFL

Sept 2013 – June 2015 Lausanne, Switzerland

Specialization in Imaging & Optical Devices
One-year Master Project at Columbia University, New York City, USA
Topic: « Detection of Gliomas on Brain MRI Images»

Bachelor of Science (BSc) in Life Sciences & Technology

School of Life Sciences, EPFL

Sept 2010 – June 2013 Lausanne, Switzerland

EXPERIENCE

Coordination & Support

Imaging@EPFL Initiative, EPFL

January 2019 – March 2020 Lausanne, Switzerland

- Coordination and support for the Imaging@EPFL Steering Committee during the launching year of the initiative.
- Initiation and organization of several projects and events.
- Responsible of the internal and external communication.

Deputy Parliament Member

State Parliament of Valais

March 2013 – March 2017 Sion, Switzerland

- The State Parliement of Valais is the legislature of the Canton of Valais.
- From 2013 to 2017, I represented voters from the Sierre district on a variety of subjects, with a special focus on health, education, and social services.

Head of Volunteering Sector

Tohu-Bohu Music Festival

Sept. 2012 – Sept. 2016 Veyras, Switzerland.

- Coordination of the on-site activities of 300+ volunteers during the event.

Head of « Teams & Staff » Sector

International Physicists Tournament

Sept. 2013 – Sept. 2014 Lausanne, Switzerland.

- Overseeing the journey and stay of the 9 international visiting teams.
- Head of staff, coordination of volunteers during the tournament.

EXPERTISE

- Imaging Science

Image Analysis | Electron Microscopy | Biomedical Imaging | Machine Learning | Signal Processing | Inverse Problems | Optimization | Mathematical Imaging

- Project Management

Monitoring of Projects and Events | Interdisciplinary Collaborations | Scienti?c Coordination | Academic Cooperation

- Communication

Webmastering | Social Media | Scienti?c Communication

AWARD

2014 WISH Community Career Investment Award

By the EPFL Foundation « Women in Sciences and Humanities »

Lausanne, April 2014

"For demonstrating academic and professional excellence"

SOFT SKILLS

Organization | Ef?ciency | Initiative | Self-Management | Professionalism | Responsibility | Time Management | Multi-Tasking | Flexibility | Teamwork | Communication | Synthetic Mind

TECHNICAL SKILLS

Microsoft Of?ce, WordPress
TeX, Git, Matlab, Python, Java, ImageJ
Social Media Management (incl. Twitter)

LANGUAGES

French (Fluent) / English (Fluent)
German (Basic) / Italian (Basic)

PUBLICATIONS

Journal Papers

- M. Zehni, **L. Donati**, E. Soubies, Z.J. Zhao, M. Unser. *Joint Angular Re□nement and Reconstruction for Single-Particle Cryo-EM*, IEEE Transactions in Image Processing, 2020
- E. Soubies, F. Soulez, M.T. McCann, T.A. Pham, **L. Donati**, T. Debarre, D. Sage, M. Unser. *Pocket Guide to Solve Inverse Problems with GlobalBioIm*, Inverse Problems, 2019
- **L. Donati**, M. Nilchian, C.O.S. Sorzano, M. Unser. *Fast Multiscale Reconstruction for Cryo-EM*, Journal of Structural Biology, 2018
- **L. Donati**, M. Nilchian, S. Trépout, C. Messaoudi, S. Marco, M. Unser. *Compressed sensing for STEM tomography*, Ultramicroscopy, 2017
- D. Sage, **L. Donati**, F. Soulez, D. Fortun, G. Schmit, A. Seitz, R. Guiet, C. Vonesch, M. Unser. *DeconvolutionLab2: An open-source software for deconvolution microscopy*, Methods, 2017

Journal Papers (preprints)

- H. Gupta, M.T. McCann, **L. Donati**, M. Unser. *CryoGAN: A New Reconstruction Paradigm for Single-particle Cryo-EM Via Deep Adversarial Learning*, Preprint Available on BiorXiv, 2020
- E. Gómez-de-Mariscal, C. García-López-de-Haro, **L. Donati**, M. Unser, A. Muñoz-Barrutia, D. Sage. *DeepImageJ: A user-friendly plugin to run deep learning models in ImageJ*, Preprint Available on BiorXiv, 2020

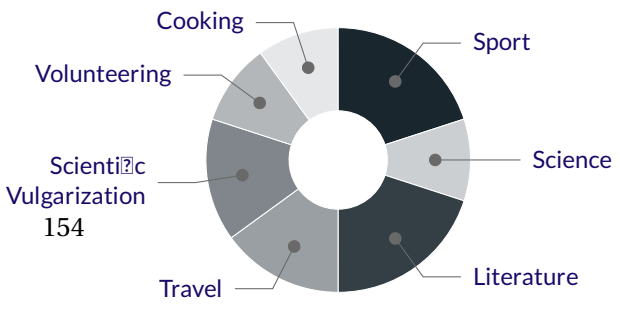
Conference Papers

- **L. Donati**, E. Soubies, M. Unser. *Inner-loop-Free ADMM for Cryo-EM*, IEEE International Symposium on Biomedical Imaging, 2019
- M. Unser, E. Soubies, F. Soulez, M. McCann, **L. Donati**. *GlobalBioIm: A unifying computational framework for solving inverse problems*, Computational Optical Sensing and Imaging, 2017
- **L. Donati**, M. Nilchian, S. Trépout, C. Messaoudi, S. Marco, M. Unser. *Compressed sensing for dose reduction in STEM tomography*, IEEE International Symposium on Biomedical Imaging, 2017

Conference Abstracts

- **L. Donati**, E. Soubies, M. Unser. *User-friendly Building of Reconstruction Algorithms with GlobalBioIm*, Quantitative Bioimaging Conference, 2020
- **L. Donati**, D. Fortun, D. Sage, M. Unser, *Higher-order Total Variation Reconstruction for Structured Illumination Microscopy*. Zeiss Symposium on Digital Optics, 2016
- N. Chu, R. Nussbaumer, **L. Donati**, D. Fortun, D. Sage, M. Unser. *A Robust Reconstruction and Pattern Calibration Framework for Structured Illumination Microscopy*, Quantitative Biomedical Imaging, 2016

PERSONAL INTERESTS



TEACHING

Occasional Lecturer

Signals & Systems I-II, Prof. Michael Unser

2019-2020 EPFL

Sections of Life Sciences and Micro-Engineering, 3rd Year Bachelor

Teaching Assistant

Signals & Systems I-II, Prof. Michael Unser & Prof. Pierre Vandergheynst

2016-2020 EPFL

Sections of Life Sciences and Micro-Engineering, 3rd Year Bachelor

Student Teaching Assistant

Analysis I-II, Prof. Hans-Jorg Ruppen
Physics I, Prof. Rolf Gruetter

2012-2014 EPFL

Section of Life Sciences, 1st Year Bachelor

SUPERVISION

Master Theses

Octave Martin, Bio-Engineering

Fall 2019 EPFL

Fast Iterative Tomographic Reconstruction Algorithm based on the Slant Stack Radon Transform. Co-supervision with Carl Zeiss, Jena, Germany.

Matthieu Broisin, Micro-Engineering

Fall 2018 EPFL

Segmentation of Nucleus Images using a Deep Learning Approach. Co-supervision with the Broad Institute, Harvard/MIT, Boston, USA.

Master Semester Project

Jelena Banjac, Computer Science

Fall 2019 EPFL

Deep Learning for Angle Estimation in Cryo-EM.

Bachelor Semester Projects

Stefania Konstantinidi, Micro-Engineering

Fall 2018 EPFL

Learning Protein Structure with Dictionary Learning.

Martial Bernard-Michel, Micro-Engineering

Spring 2018 EPFL

Optimal Con□gurations for PBT.

Matthieu Broisin, Micro-Engineering

Fall 2017 EPFL

PBT in the GlobalBioIm Library.