New Algorithms in Computational Microscopy

Pham, Minh

2020

UNIVERSITY OF CALIFORNIA
Los Angeles

New Algorithms in Computational Microscopy

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Minh Pham

2020

ABSTRACT OF THE DISSERTATION

New Algorithms in Computational Microscopy

by

Minh Pham
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2020
Professor Stanley Osher, Chair

Microscopy plays an important role in providing tools to microscopically observe objects and their surrounding areas with much higher resolution ranging from the scale between molecular machineries (angstrom) and individual cells (micrometer). Under microscopes, illumination, such as visible light and electron-magnetic radiation/electron beam, interacts with samples, then they are scattered to a plane and are recorded. Computational microscopy corresponds to image reconstruction from these measurements as well as improving quality of the images. Along with the evolution of microscopy, new studies are discovered and algorithms need development not only to provide high-resolution imaging but also to decipher new and advanced researches.

In this dissertation, we focus on algorithm development for inverse problems in microscopy, specifically phase retrieval and tomography, and the application of these techniques to machine learning. The four studies in this dissertation demonstrates the use of optimization and calculus of variation in imaging science and other different disciplines.

Study 1 focuses on coherent diffractive imaging (CDI) or phase retrieval, a non-linear inverse problem that aims to recover 2D image from it Fourier transforms in modulus taking into account that extra information provided by oversampling as a second constraint. To solve this two-constraint minimization, we proceed from Hamilton-Jacobi partial differential equation (HJ-PDE) and Hopf-Lax formula. Introducing generalized Bregman distance to the HJ-PDE and applying Legendre transform, we derive our generalized proximal smoothing (GPS) algorithm under the form of primal-dual hybrid gradient (PDHG). While the reflection operator, known as extrapolating momentum, helps overcome local minima, the smoothing by the generalized Bregman distance is adjusted to improve convergence and consistency of phase retrieval.

Study 2 focuses on electron tomography, 3D image reconstruction from a set of 2D projections obtained from a transmission electron microscope (TEM) or X-ray microscope. Notice that current tomography algorithms limit to a single tilt axis and fail to work with fully or partially missing data. In the light of calculus of variations and Fourier slice theorem (FST), we develop a highly accurate tomography iterative algorithm that can provide higher resolution imaging and work with missing data as well as has capability to perform multiple-tilt-axis tomography. The algorithm is further developed to work with non-isolated objects and partially-blocked projections which have become more popular in experiment. The success of real space iterative reconstruction engine (RESIRE) opens a new era to the study of tomography in material science and magnetic structures (vector Tomography).

Study 3 and 4 are applications of our algorithms to machine learning. Study 3 develops a backward Euler method in a stochastic manner to solve K-mean clustering, a well-known non-convex optimization problem. The algorithm has been shown to improve minimums and consistency, providing a new powerful tool to the class of classification techniques. Study 4 is a direct application of GPS to deep learning gradient descent algorithms. Linearizing the Hopf-Lax formula derived in GPS, we derive our method Laplacian smoothing gradient descent (LSGD), simply known as gradient smoothing. Our experiment shows that LSGD has the ability to search for better and flatter minimums, reduce variation, and obtain higher accuracy and consistency.

The dissertation of Minh Pham is approved.

Wotao Yin

Luminita Vese

Christopher Anderson

Stanley Osher, Committee Chair

University of California, Los Angeles

2020

To my mother . . .

for her inexhausted support to my education—

And to my punctilious wife Tram

for her unconditional love and ten years waiting.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to sincerely thank my advisor, professor Stanley Osher, for his mentorship and support over the past four years, and for showing me the value of research. I always remember his kind of taking me as his mentee when my former advisor left, and all the opportunities and advice that he provided me. And thanks to his referral, I have a great chance to work with professor Jianwei (John) Miao, and his Coherent Imaging Group, to whom I would also give my most thanks as my unofficial advisor. From John, I have learned the qualities to become a true researcher, and I am so grateful for his guidance as showing me the value of passion and persistence. I would like to thank all members of the Coherent Imaging Group and Osher Lab group, especially Penghang Yin and Arjun Rana, for their guidance when I first did research. Without these help, my journey would have never been successful and enjoyable. Finally I would like to thank my committee for their support and consideration of this thesis.

Chapter 2 is reprinted with permission from Minh Pham, Penghang Yin, Arjun Rana, Stanley Osher, and Jianwei Miao, "GPS: Generalized proximal smoothing for phase retrieval." Opt. Express 27, 2792-2808 (2019) © The Optical Society. I thank and acknowledge the contributions form the co-authors, Penghang Yin, Arjun Rana, Stanley Osher, and Jianwei Miao.

Chapter 3 is reprinted from Minh Pham, Yakun Yuan, Arjun Rana, Jianwei Miao, and Stanley Osher. "RESIRE: real space iterative reconstruction engine for Tomography." arXiv preprint arXiv:2004.10445 (2020). I thank and acknowledge the contributions from the co-authors Yakun Yang, Arjun Rana, Stanley Osher, and Jianwei Miao.

Chapter 4 is reprinted by permission from Springer Nature Customer Service Centre GmbH: Nature Springer, Journal of Scientific Computing, "Stochastic backward Euler: an implicit gradient descent algorithm for k-means clustering." Penghang Yin, Minh Pham, Adam Oberman and Stanley Osher. 77, 1133–1146 © (2018). I thank and acknowledge the contributions from the co-authors Penghang Yin, Adam Oberman and Stanley Osher.

Chapter 5 is reprinted from Stanley Osher, Bao Wang, Penghang Yin, Xiyang Luo,

Farzin Barekat, Minh Pham, and Alex Lin. “Laplacian smoothing gradient descent.” arXiv preprint arXiv:1806.06317 (2018). I thank and acknowledge the contributions from the co-authors Stanley Osher, Bao Wang, Penghang Yin, Xiyang Luo, Farzin Barekat, and Alex Lin.

Lastly, I acknowledge support by National Science Foundation Science and Technology Center under Grant No. DMS-1554564 (STROBE), DMR 1548924, and DMS-1924935, U.S Department of Energy DOE (DE-SC0010378), Air Force Research Laboratory under Grant No. FA9550-18-0167 and MURI FA9550-18-1-0502, the Office of Naval Research under Grant No. N00014-18-1-2527.

VITA

2014–2020	Graduate Student Researcher Department of Mathematics University of California, Los Angeles Los Angeles, CA
2015–2017	Teaching Assistant Department of Mathematics University of California, Los Angeles Los Angeles, CA
2011–2014	B.S. Mathematics and B.S. Computer Science University of Georgia Athens, GA

PUBLICATIONS

M. Pham, P. Yin, A. Rana, S. Osher, and J. Miao. Generalized proximal smoothing (GPS) for phase retrieval. *Optics Express* **27**, 2792-2808 (2019).

M. Pham, A. Rana, J. Miao, and S. Osher. Semi-implicit relaxed Douglas-Rachford algorithm (sDR) for ptychography. *Optics Express* **27**, 31246-31260 (2019)

P. Yin, **M. Pham**, A. Oberman, and S. Osher. Stochastic Backward Euler: An Implicit Gradient Descent Algorithm for k-Means Clustering. *Journal of Scientific Computing* **77**, 1133–1146 (2018).

S. Osher, B. Wang,P. Yin, X. Luo, F. Barekat, **M. Pham**, and A. Lin. Laplacian smoothing gradient descent. *arXiv preprint arXiv:1806.06317* (2018).

A. Rana, J. Zhang, **M. Pham**, A. Yuan, Y. H. Lo, H. Jiang, and J. Miao. Potential of attosecond coherent diffractive imaging. *arXiv preprint arXiv:1907.00411* (2018).

S. Osher, B. Wang, P. Yin, X. Luo, F. Barekat, **M. Pham**, and A. Lin. Laplacian smoothing gradient descent. *arXiv preprint arXiv:1806.06317* (2018).

CHAPTER 1

Introduction

1.1 Background

1.1.1 History of microscopy

Mankind history sees the evolution of microscopy and technological innovation has enabled humans to explore and visualize the microworld microscopically. Under a light microscope, humans saw cells and bacteria for the first time, and hence they learned about the basic units of living organisms. With the latest breakthrough, humans can further see the complex structure of samples with resolutions in the tens of nanometers, and observe its change in time (4D imaging). Localization of individual molecules or atom positions can be specified with high accuracy under fluorescent and electron microscopes, respectively.

Along with the development of microscopes, humans also see the evolution of computer technology helps scientists to store and process extremely large data with high performance and efficiency. Last but not least, the development of computational mathematics has significantly benefit scientists in deciphering those biological structures and inorganic materials. Microscopy provides us images of samples that; however, have been modified by the optical system while recovering the original images is considered as inverse problems. Depending on microscope setting, reconstruction can be mathematically designed to decode images with various options according to the demand of scientists.

To understand these inverse problems, we can take a look at the history of microscopy. The principle of microscopy is based on the diffraction theory of light which has been developed for the last couple of centuries. The earliest “wave theory of light” proposed by

Christian Huygens in 1678, stated that “each point on a wavefront can be considered as a new source of a secondary source.” The theory was verified by the famous double slit interference experiment by Thomas Young in 1804. During the gap between these two landmarks, Isaac Newton proposed the “corpuscular theory” of light, stating that light is made up of small discrete particles called ”corpuscles” (little particles) which travel in a straight line with a finite velocity and possess impetus.

In 1860, Maxwell formed a set of partial differential equations that, together with the Lorentz force law, set the foundation of classical electromagnetism and optics. Even though Maxwell’s equations later become more famous in mathematics and electromagnetism, Maxwell first used those equations to propose that light is an electromagnetic phenomenon.

However in early 1818, Fresnel presented the calculation of diffraction patterns that, with highly accuracy, was later mathematically formulated by Kirchhoff in 1882 and further developed by Rayleigh and Sommerfeld. Kirchhoff’s diffraction formula (also known as Fresnel–Kirchhoff diffraction formula) expresses the wave disturbance and its propagation in space when a monochromatic spherical wave passes through an opening in an opaque screen. From the formula were derived Fraunhofer and Fresnel diffraction which established the fundamentals of modern microscopy.

1.1.2 Optical microscopy & phase retrieval

1.1.2.1 Optical microscopy

Microscopy is the technical field of using microscopes to view objects and areas of objects with higher resolution ranging from a couple of micrometers to angstroms [2]. There are three well-known branches of microscopy: optical, electron, and scanning probe microscopy, along with the emerging field of X-ray microscopy.

Optical and electron microscopy use visible lights or electron beams to interact with specimens, and then collect the radiation that is scattered from the beam under diffraction, reflection and refraction. This process may be carried out by wide-field irradiation of the sample (for example standard light microscopy and transmission electron microscopy) or by scanning a fine beam over the sample (for example confocal laser scanning microscopy and scanning electron microscopy).

In optical microscopy, a single lens or multiple lenses allows visible light to transmit through or reflect from the sample and magnify its view. The resulting image can be detected directly by the eye, or captured digitally by a camera (CCD). There are diverse techniques in optical microscopes according to interest in research. Bright field microscopy is the standard and simplest optical-microscopy technique, widely used in most microscopy labs. However, to improve specimen contrast or highlight certain structures in a sample, special techniques must be used. On the other hand, dark field microscopy can improve the contrast of unstained, transparent specimens by trying to minimize the quantity of directly transmitted (unscattered) light entering the image plane, and collect only the light scattered by the sample. When one would like to highlight a certain chemical-compound characteristic, fluorescent dyes can be used to stain these structures or chemical compounds and this introduces fluorescence microscopy. Another important technology is confocal that can prevent out of focus and help improve resolution slightly.

There are more microscopy technologies to enhance visualization of samples such as phase contrast, cross-polarized light, etc; however, the fundamental of resolution improvement is using light sources with smaller wavelengths. For example, electron beams can replace farther resolution and see structures of couple angstrom. Even though less common, X-ray with resolution lying in between that of light and electron microscopy has also been developed since the late 1940s.

1.1.2.2 Phase retrieval

Fresnel diffraction formula, approximating the Kirchhoff–Fresnel diffraction in the near field, implies that the electromagnetic field at a distance away from the diffraction object plan is proportional to its Fourier transform [12]. The diffraction patterns obtained at camera show their intensities, which are the squared amplitude of the electromagnetic field, i.e. diffraction pattern can be approximated from Fourier transform. Recovering the image from diffraction patterns is an inverse problem, and is equivalent to retrieving the phase.

Early iterative algorithms that were developed to solve the phase retrieval problem such as error reduction, and hybrid input-output fundamentally work based on alternating projection between real space and Fourier space. Important applications of phase retrieval include X-ray crystallography, transmission electron microscopy and coherent diffractive imaging where the last one is a "lensless" technique and was first demonstrated in 1999 by Miao.

The phase retrieval problem encountered difficulties such as "shrink wrap" [52] where the reconstructed image stagnates and produces features of both object and its conjugate. Note that both image $f(x)$ and its conjugate $f^*(x)$ have the same Fourier modulus; hence inadequate measurement can cause stagnation in the reconstruction. Thanks to the advanced technology in digital imaging, the invention of high-resolution CCD cameras allows us to collect measurements with a finer grid, i.e. oversampling. This experimental technique provides useful information for the development of phase retrieval algorithms. Being combined with reflection operator as extrapolating momentum, HIO [51] has been practically shown to obtain good reconstructions with oversampling ratio in the range $[3, 4]$. However, consistency and robustness of the algorithm are still big question marks. Optimization is supposed to help fill in the gap and improve the reconstruction.

Other approaches have been exhaustively studied in the past few decades such as phaselift [19] and Wirtinger flow [20] but those methods are not widely used in practices. They are both expensive, under-performance, and only attract interest from the theoretical and analysis community.

We finish this part by stating that the success of CDI opened up X-ray microscopy to the realistic possibility of imaging non-crystalline specimens in a truly diffraction and dose-limited manner while development in instrumentation can push spatial resolution better than 10 nm [130]. As coherent diffractive X-ray microscopy enters the mainstream, it is now a still-growing yet proven technique.

1.1.3 Electron microscopy & tomography

1.1.3.1 Electron microscopy

Electron microscopes use a beam of accelerated electrons as a source of illumination. With a far shorter wavelength than that of visible light (about 100,000 times), electron microscopy has a higher resolving power than light microscopes and can reveal the structure of smaller objects such as atoms and particles. There are two main techniques under electron microscopy: transmission electron microscopy (TEM) that sends an electron beam through a very thin slice of specimen, and scanning electron microscopy (SEM) that can visualize details on surfaces of specimens.

In 2005, TEM resolution was shown to push up to 0.5Å [45] and obtain atomic structures such as atoms and particles. Different from optical microscopy, TEM can also work with image mode when the objective aperture is inserted in a back focal plane (BFP) of the objective lens (where diffraction spots are formed). Depending on the use of objective aperture, bright field image (BF image) or dark field image (DF image) is obtained. With the help of Intermediate and Projector lenses, the selected signal is magnified and projected on a screen (or camera), forming an image of the sample.

TEM and its modification scanning transmission electron microscope (STEM) allow us to perform tomography by collecting images at different tilt angles [83]. This extension of traditional TEM is known as electron tomography, opens a new study to high-resolution 3D imaging of physical and biological samples, including atomic electron tomography (AET) that studies 3D atomic structures of point and crystal defects such as dislocation, chemical order/disorder, grain and anti-phase boundaries.

1.1.3.2 Tomography

The first X-ray tomography experiment was first demonstrated by Oldendorf in 1963, but the mathematical theory had been built much earlier in 1917 by Johann Radon [115, 116]. The Austrian mathematician invented Radon transform and mathematically showed that a function can be reassembled from an infinite set of its projections. In 1937, Polish mathematician Stefan Kaczmarz [74] developed a method, along with Allan McLeod Cormack's theoretical and experimental work, to find an approximate solution to a large system of linear equations, and laid the foundation for algebraic reconstruction technique ART [31, 60].

Tomography algorithms have been developed for decades and mainly fall into two categories: one-step method filtered back projection (FBP) [42] and iterative reconstruction (IR) [64, 72]. FBP performs well with adequate measurements, and shows efficiency in computational cost and time. On the other hand, IR, despite more computational expense, refines better images relative to FBP, especially when measurements are limited or missing.

Since tomography becomes popular in both commercial use and research, numerical methods are developed to obtain fast and efficient reconstructions [140]. Another important trend of research is to improve performance and generalize Tomography to the multiple-tilt-axis case, for example generalized Fourier iterative reconstruction GENFIRE [112].

To finish this part, we report that TEM has found diverse applications in cancer research, virology, and materials science as well as pollution, nanotechnology and semiconductor research, but also in other fields such as paleontology and palynology. Together with that build-out, algorithm development in Tomography has been shown to be very important to improving high-resolution 3D imaging.

1.2 Contribution

To that end, the main objective of this dissertation is to develop advanced algorithms for inverse problems: phase retrieval and tomography, in microscopy. To proceed with these objectives, I apply optimization techniques and calculus of variation. Carefully reformulating these problems, I mathematically derive a better algorithm and design a suitable numerical scheme for each case. This dissertation is organized into three aims. The first and second aim contain one chapter for each, focusing on algorithm development for phase retrieval and tomography. The third aim contains two chapters, showing applications of these techniques in machine learning, especially in k-means clustering and deep learning algorithms.

1.2.1 Aim 1: Algorithm development for phase retrieval

Project 1: Develop a high-performance phase retrieval algorithm. The objective goals are composed of two following parts. First, the algorithm must have the ability to overcome local minima and achieve lower objective values. Second, the method needs to improve consistency and robustness to noise.

Chapter 2: Generalized proximal smoothing (GPS) for phase retrieval

We recall that oversampling technique provides useful information and acts as a second constraint in the phase retrieval problem. Thanks to oversampling, the minimization then becomes reconstructing a 2D image that satisfies two constraints: Fourier modulus constraint and "real, positive and support" constraint. I apply generalized proximal smoothing, where a generalized Bregman metric replaces the usual Euclidean distance, to relax these constraints into two generalized "sum of squared errors" fidelity terms. The relaxation, interpreted as a solution to an unusual Hamilton-Jacobi partial differential equation, smooths out the non-convex & non-smooth function in the sense of bringing down maxima, lifting up bad minima, and widening good minima. The smoothing improves capability of escaping bad local minima (sharp & narrow) and robustness to noise. Finally, primal-dual hybrid gradient (PDHG) [22] algorithm is borrowed to solve this bi-objective function.

1.2.2 Aim 2: Algorithm development for tomography

Project 2: Develop a highly accurate tomographic iterative algorithm. The method needs not only providing a high-qualitative imaging but also robust to noise, especially must work with missing-wedge data. In addition, it must have capability to work with generalized multiple-tilt-axis case and partially-blocked projection.

Chapter 3: RESIRE: real space iterative reconstruction engine

Considering calculus of variations, I reformulate tomography as l_2 minimization problem and derive the corresponding Euler-Lagrange equation. Simple gradient descent can be applied to iteratively refine reconstructions. The significant impact is that the new formulation mathematically allows working with partial missing data, i.e. only part of projection images is given. This capability benefits the reconstruction where projections of samples are partially blocked, for example by a cap. Equivalently, this functionality can be extended to work with non-isolated objects or thin film where the size of sample is larger than its field of view (FOV). Along with deriving the new formulation, I also implement an effective and efficient iterative method while considering Fourier slice theorem to improve accuracy of forward projection. The success of this algorithm opens a new study to tensor tomography where scientists are interested in exploring vector fields caused by magnetic material.

1.2.3 Aim 3: Application to machine learning

1.2.3.1 K-mean clustering

Project 3: Develop an iterative algorithm to improve the performance of K-mean clustering considering a combination of stochastic gradient descent and backward Euler method.

Chapter 4: Stochastic backward Euler: an implicit gradient descent algorithm for K-means clustering

Motivated from "deep relaxation" [26], an optimization method in deep learning, I simplify the algorithm and apply it to the K-means problem, minimizing a non-convex "sum of squared errors" function $f(x)$. The simplified method is a coupling of stochastic gradient descent and backward Euler where the later one can be processed by a few iterations of fixed-point method. This is an interpretation of Hopf-Lax formula, solution under implicit form to a viscous Hamilton-Jacobi partial differential equation (HJ-PDE). The aim is to minimize a smoothing of $f(x)$, termed $f_{HJ}(x)$, obtained from the viscous HJ-PDE. The combination provides a "better descent" direction than the usual GD method since it goes toward the descent direction of $f_{HJ}(x)$ instead of $f(x)$. Through simulation and practical experiments, we show that stochastic backward Euler (SBE) provides more faithful clustering with lower objective values and higher consistency. The probability of escaping from local minima is improved since SBE can always avoid those traps by descending in a better direction.

1.2.3.2 Laplacian smoothing gradient descent

Project 4: Develop an iterative method for deep learning to improve convergence and consistency by gradient smoothing.

Chapter 5: Laplacian smoothing gradient descent

Taking a different motivation than the method in “deep relaxation” [26], we aim to correct the descent direction in the GD method by applying smoothing on the gradient. In this new approach, we instead focus on the non-viscous HJ-PDE where we further modify the Hamiltonian by a generalized metric. This modification yields a backward Euler (BE) equation with a modified gradient. When simplifying the BE and smoothing is taken into account, we obtain the Laplacian smoothing gradient descent (LSGD). Experimental testing shows that LSGD acquires better minima with more consistency.

1.3 Summary to dissertation

In summary, the four studies presented in this dissertation combine three main topics: algorithm development in (1) diffraction imaging, (2) 3D reconstruction, and (3) applications to machine learning. In terms of optimization, this collection of work provides advanced techniques to inverse problems in microscopy. Mathematical optimization is the key to their success, and open new studies in microscopy, for example material science and magnetic structure. The far-reaching goal of this work is application of these algorithms in a plethora of imaging modalities across disciplines. As technology continues to develop, more microscopy problems will emerge and algorithm development always plays a crucial part in imaging science.

CHAPTER 2

Generalized proximal smoothing (GPS) for phase retrieval

In this study, we report the development of the generalized proximal smoothing (GPS) algorithm for phase retrieval of noisy data. GPS is a optimization-based algorithm, in which we relax both the Fourier magnitudes and support constraint. We relax the support constraint by incorporating the Moreau-Yosida regularization and heat kernel smoothing, and derive the associated proximal mapping. We also relax the magnitude constraint into a least squares fidelity term, whose proximal mapping is available as well. GPS alternatively iterates between the two proximal mappings in primal and dual spaces, respectively. Using both numerical simulation and experimental data, we show that GPS algorithm consistently outperforms the classical phase retrieval algorithms such as hybrid input-output (HIO) and oversampling smoothness (OSS), in terms of the convergence speed, consistency of the phase retrieval, and robustness to noise.

2.1 Introduction

Phase retrieval has been fundamental to several disciplines, ranging from imaging, microscopy, crystallography and optics to astronomy [102, 100, 131, 90]. It aims to recover an object only from its Fourier magnitudes. Without the Fourier phases, the recovery can be achieved via iterative algorithms when the Fourier magnitudes are sampled at a frequency sufficiently finer than the Nyquist interval [94]. In 1972, Gerchberg and Saxton developed an iterative algorithm for phase retrieval, utilizing the magnitude of an image and the Fourier magnitudes as constraints [55]. In 1982, Fienup generalized the Gerchberg-Saxton algorithm by developing two iterative algorithms: error reduction (ER) and hybrid input-output (HIO), which use a support and non-negativity as constraints in real space and the Fourier magnitudes in reciprocal space [50]. In 1998, Miao, Sayre and Chapman proposed, when the number of independently measured Fourier magnitudes is larger than the number of unknown variables associated with a sample, the phases are in principle encoded in the Fourier magnitudes and can be retrieved by iterative algorithms [94]. These developments finally led to the first experimental demonstration of coherent diffractive imaging (CDI) by Miao and collaborators in 1999 [95], which has stimulated wide spread research activities in phase retrieval, CDI, and their applications in the physical and biological sciences ever since [100, 25, 120].

For a finite object, when its Fourier transform is sampled at a frequency finer than the Nyquist interval (i.e. oversampled), mathematically it is equivalent to padding zeros to the object in real space. In another words, when the magnitude of the Fourier transform is oversampled, the correct phases correspond to the zero-density region surrounding the object, which is known as the oversampling theorem [94, 93]. The phase retrieval algorithms iterate between real and reciprocal space using zero-density region and the Fourier magnitudes as dual-space constraints. A support is typically defined to separate the zero-density region from the object. The support constraint, which is also known as the non-negativity constraint, is applied to the density inside the support. In the ER algorithm, the no-density region outside the support and the negative density inside the support are set to zero in each iteration [50]. The HIO algorithm relaxes the ER in the sense that it gradually reduces the densities that violate the support constraint instead of directly forcing them to zero [50]. This relaxation often leads to good reconstructions from noise-free patterns. However, in real experiments, the diffraction intensities, which are proportional to the square of Fourier magnitudes, are corrupted by a combination of Gaussian and Poisson noise and missing data. In the presence of experimental noise and missing data, phase retrieval becomes much more challenging, and the ER and HIO algorithms may only converge to sub-optimal solutions. Simply combining ER and HIO still suffers from stagnation and the iterations can get trapped at local minima [51]. To alleviate these problems, more advanced phase retrieval algorithms have been developed such as the shrink-wrap algorithm and guided HIO (gHIO) [28, 90]. In 2010, a smoothness constraint in real space was first introduced to improve the phase retrieval of noisy data [117]. Later, a noise robust framework was implemented for enhancing the performance of existing algorithms [91]. Recently, Rodriguez et al. proposed to impose the smoothness constraint on the no-density region outside the support by applying Gaussian filters [123]. The resulting oversampling smoothness (OSS) algorithm successfully reduces oscillations in the reconstructed image, and is more robust to noisy data than the existing algorithms.

Since phase retrieval can be cast as a non-convex minimization problem, many efforts have been made to study phase retrieval algorithms from the viewpoint of optimization. For example, Bauschke et al. [11] related HIO to a particular relaxation of the Douglas-Rachford algorithm [39] and introduced the hybrid projection reflection algorithm [10, 11]. Using similar ideas, researchers further proposed several projection algorithms such as iterated difference map [44] and relaxed averaged alternation reflection [88]. In [30], Chen and Fannjiang analyzed a Fourier-domain Douglas-Rachford algorithm for phase retrieval. By taking noise into account, the Wirtinger Flow [20] replaces the amplitude constraint by the intensity constraint which is then relaxed into a least square fidelity term that measures the misfit of measured Fourier intensity data, to which gradient descent is applied. Other methods in this line include alternating direction methods [23, 24, 141] that have been widely used in image processing, as well as lifting approaches [21] such as PhaseLift [18, 19] by Candès et al. and its variants [139, 146].

In this paper, we propose an optimization-based phase retrieval method, termed generalized proximal smoothing (GPS), which effectively addresses the noise in both real and Fourier spaces. Motivated by the success of OSS [123], GPS incorporates the idea of Moreau-Yosida [103, 148] regularization with heat kernel smoothing, to relax the support constraint into a continuous penalty term. We further relax the magnitude constraint into a least squares fidelity term, for de-noising in Fourier space. To minimize the resulting primal-dual formulation, GPS iterates back and forth between efficient proximal mappings of the two relaxed functions, respectively. Our experimental results using noisy experimental data of biological and inorganic specimens demonstrate that GPS consistently outperforms the state-of-the-art algorithms HIO and OSS in terms of both speed and robustness. We also refer readers to the recent paper [147] about training quantized neural networks, which shows another success of using Moreau-Yosida regularization to relax the hard constraint.

Notations. Let us fix some notations. For any complex-valued vectors $u, v \in \mathbb{C}^n$, $\overline{u}$ is the complex conjugate of u, whereas $u^* := \overline{u}^\top$ is the Hermitian transpose. $\mathrm{Re}(u)$ and $\mathrm{Im}(u)$ are the real and imaginary parts of u, respectively.

$\langle u, v\rangle := u^* v = \sum_{i=1}^n \overline{u_i} v_i$ is the Hermitian inner product of u and v. $u \odot v$ is the element-wise (Hadamard) product of u and v given by $(u \odot v)_i = u_i v_i$. $\|u\| := \sqrt{\langle u, u\rangle}$ denotes the ℓ_2 norm of u. Given any Hermitian positive semi-definite matrix $K \in \mathbb{C}^{n\times n}$, we define $\|u\|_K := \sqrt{\langle u, K\,u\rangle}$.

$\arg(u) := \tan^{-1}\left(\dfrac{\mathrm{Im}(u)}{\mathrm{Re}(u)}\right)$ denotes the argument (or phase) of a non complex-valued vector $u = \mathrm{Re}(u) + \mathbf{i}\cdot \mathrm{Im}(u)$. For convenience we set $\arg(0) = 1$.

$\mathcal{I}_{\mathcal{X}}$ is the indicator function of a closed set $\mathcal{X} \subset \mathbb{C}^n$ given by $\mathcal{I}_{\mathcal{X}}(x) = 0$ if $x \in \mathcal{X}$ and $\mathcal{I}_{\mathcal{X}}(x) = \infty$ otherwise

$\mathrm{proj}_{\mathcal{X}}(u) := \arg\min_{v\in\mathcal{X}} \|v - u\|$ is the projection of u onto $\mathcal{X}$,

$\mathrm{prox}_f(u) := \arg\min_v \left\{f(v) + \frac{1}{2}\|v-u\|^2\right\}$ is the proximal mapping of the function f.

$h^*(y) := \sup\limits_{u\in\mathcal{C}} \left\{\mathrm{Re}\langle y, u\rangle - h(u)\right\}$ is the extended Legendre transformation to the complex plane of a real-valued function h.

2.2 Proposed model

We first fix some settings of the phase retrieval problem. Consider the reconstruction of a 2D image u defined on a discrete lattice $\Omega := \{(i,j) : 1 \le i \le n_1, 1 \le j \le n_2\}$.

For simplicity, we represent u in terms of a vector in $\mathbb{R}^n$ by the lexicographical order with $n = n_1 \times n_2$. Then u_i represents the density of image at the i-th pixel. Due to oversampling, the object densities reside in a sub-domain $S \subset \Omega$ known as the support, and u is supposed to be zero outside S. Throughout the paper, we assume that the support S is centered around the domain Ω. The support constraint is $\mathcal{S} := \{u \in \mathbb{R}^n : u_i \geq 0 \text{ if } i \in S,\ u_i = 0 \text{ otherwise}\}$.

The Fourier magnitude data is obtained as $b = |\mathcal{F}u|$, where $\mathcal{F} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{C}^{n_1 \times n_2}$ is the discrete Fourier transform (DFT). We denote the magnitude constraint by $\mathcal{T} := \{u \in \mathbb{R}^n : |\mathcal{F}u| = b\}$.

In the absence of noise, phase retrieval (PR) problem is simply to

$$\text{find } u \in \mathbb{R}^n, \quad \text{such that } u \in \mathcal{S} \cap \mathcal{T}. \tag{2.1}$$

This amounts to the following composite minimization problem

$$\min_{u \in \mathbb{R}^n} \ f(u) + g(\mathcal{F}u), \tag{2.2}$$

where $f(u) := \mathcal{I}_{\mathcal{S}}(u)$ and $g(z) := \mathcal{I}_{|z|=b}(z)$ are two indicator functions that enforce the object and Fourier magnitudes constraints, respectively. Note that f is a closed and convex function while g is closed but non-convex. This gives rise to the non-convex optimization problem of (2.2).

2.2.1 A new noise-removal model

In real experiments, the Fourier data are contaminated by experimental noise. Moreover, the densities outside the support are not exactly equal to zero. In the noisy case, the image to be reconstructed no longer fulfills either the Fourier magnitudes or the support constraint. The ER algorithm (a.k.a. alternating projection), which performs alternatively projects onto these two constraints, apparently does not take care of the noise. The HIO "relaxes" the support constraint on densities wherever it is violated. This relaxation only helps in the noiseless case. In the presence of noise, the feasible set $\mathcal{S} \cap \mathcal{T}$ can be empty, and alternating projection method and its relaxation, like ER and HIO, may fail to converge and keep oscillating. The OSS [123] improves the HIO by applying extra Gaussian filters to smooth the densities outside the support at different stages of the iterative process. None of them, however, seems to properly address the corruption of the Fourier magnitudes.

Introducing the splitting variable $z = \mathcal{F}u \in \mathbb{C}^n$, we reformulate (2.2) as

$$\min_{u,z\in\mathbb{C}^n} \; f(u) + g(z) \quad \text{subject to} \quad z = \mathcal{F}u. \tag{2.3}$$

In the presence of noise, we seek to relax the indicator functions f and g that enforce hard constraints into soft constraints. To this end, we first relax g into a least square fidelity, which is the sum of squared errors as follows

$$g_\sigma(z) = \frac{1}{2\sigma}\||z| - b\|^2. \tag{2.4}$$

This fidelity term has been considered in the literature by assuming the measurements being corrupted by i.i.d. Gaussian noise; see [23] for example. In practice, we observe that it works well even with a combination of Gaussian and Poisson noise. This relaxation is related to the Moreau-Yoshida regularization [103, 148], which will be explained in detail in the appendix.

Following this line, we further relax $f = \mathcal{I}_\mathcal{S}$ into

$$f_G(u) := \inf_v \; \left\{ f(v) + \frac{1}{2}\|v - u\|^2_{G^{-1}} \right\} = \inf_{v\in\mathcal{S}} \; \frac{1}{2}\|v - u\|^2_{G^{-1}} \tag{2.5}$$

for some Hermitian positive definite matrix G. The choice of G here is tricky, and will be discussed later in section 3. The relaxation of both constraints thus leads to the proposed noise-removal model

$$\min_{u,z\in\mathbb{C}^n} \; f_G(u) + g_\sigma(z) \quad \text{subject to} \quad z = \mathcal{F}u. \tag{2.6}$$

This constraint optimization can be solved by a splitting method, such as alternating direction method of multipliers (ADMM) [57, 16, 144] or primal dual hybrid gradient (PDHG) algorithm [22, 46, 58, 107].

2.2.2 A primal-dual formulation

We solve the constraint optimization problem (2.6) by introducing the Lagrangian

$$\mathcal{L}(u, z; y) = f_G(u) + g_\sigma(z) + \mathrm{Re}\langle y, \mathcal{F}^* z - u\rangle, \tag{2.7}$$

where $\mathcal{F}^* = \mathcal{F}^{-1}$ is the adjoint of $\mathcal{F}$ or the inverse DFT. The corresponding Karush-Kuhn-Tucker (KKT) condition is

$$y \in \partial f_G(u),\ -\mathcal{F}y \in \partial g_\sigma(z) \iff \mathcal{F}^* z = u \in \partial f_G^*(y),\ -\mathcal{F}y \in \partial g_\sigma(z), \tag{2.8}$$

where the right hand side equivalence is obtained by the convex conjugate property and is exactly the KKT condition of the following min-max saddle point problem

$$\min_z \max_y\ g_\sigma(z) - f_G^*(y) + \mathrm{Re}\langle z, \mathcal{F}y\rangle, \tag{2.9}$$

with $f_G^*(y)$ is the generalized Legendre-Fenchel transformation of f_G^* and has an explicit form thanks to Moreau-Yoshida decomposition

$$f_G^*(y) = f^*(y) + \frac{1}{2}\|y\|_G^2, \tag{2.10}$$

where f^* is the Legendre-Fenchel transformation of f defined as

$$f^*(y) = \sup_{v \in \mathcal{S}}\ \mathrm{Re}\langle y, v\rangle = \begin{cases} 0 & \text{if } \mathrm{Re}(y) \le 0 \text{ on } S, \\ \infty & \text{otherwise.} \end{cases} \tag{2.11}$$

For notational convenience, we denote $\mathcal{S}^* := \{y \in \mathcal{S}^n : \mathrm{Re}(y) \le 0 \text{ on } S\}$ the dual constraint set of $\mathcal{S}$. Our finalized min-max problem is thus given by

$$\min_z \max_{y \in \mathcal{S}^*}\ \frac{1}{2\sigma}\||z| - b\|^2 - \frac{1}{2}\|y\|_G^2 + \mathrm{Re}\langle z, \mathcal{F}y\rangle. \tag{2.12}$$

2.3 Generalized proximal smoothing (GPS) algorithm

We carry out the minimization of the saddle point problem (2.9) or (2.12) by a generalized primal dual hybrid gradient (PDHG) algorithm [22, 46, 58, 107], which iterates

$$\begin{cases} z^{k+1} = \mathrm{prox}_{tg_\sigma}\big(z^k - t\mathcal{F}y^k\big) \\ y^{k+1} = \mathrm{prox}_{sf_G^*}\big(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k)\big) \end{cases} \tag{2.13}$$

for some step sizes $s, t > 0$. The update of z^{k+1} calls for computing the proximal mapping of tg_σ [23], whose analytic solution is given by

$$\mathrm{prox}_{tg_\sigma}(z) = \arg\min_{v\in\mathbb{C}^n} \frac{1}{2\sigma}\||v| - b\|^2 + \frac{1}{2t}\|v - z\|^2 = \frac{b \odot \exp(\mathbf{i}\cdot \arg(z)) + (\sigma/t)z}{1+\sigma/t}, \quad (2.14)$$

which is essentially a weighted average between z and its projection onto the magnitude constraint $\{z \in \mathbb{C}^n : |z| = b\}$.

Moreover, we need to find the proximal mapping of sf_G^* for updating y^{k+1}, which reduces to

$$\begin{aligned}\mathrm{prox}_{sf_G^*}(y) &= \arg\min_{v\in\mathbb{C}^n} f_G^*(v) + \frac{1}{2s}\|v-y\|^2 = \arg\min_{v\in\mathbb{C}^n} f^*(v) + \frac{1}{2}\|v\|_G^2 + \frac{1}{2s}\|v-y\|^2 \\ &= \arg\min_{v\in\mathcal{S}^*} \frac{1}{2}\|v\|_G^2 + \frac{1}{2s}\|v-y\|^2. \qquad (2.15)\end{aligned}$$

Eq. (2.15) has closed-form solution only when G is a diagonal matrix. In the other case, we provide an approximation.

We devise two versions of GPS algorithm based on different choices of G.

2.3.1 Real space smoothing

One choice of G is $G = \gamma\, D^\top D$. Here D is the discrete gradient operator, and then $D^\top D$ is the negative of discrete Laplacian. In this case,

$$f_G^*(y) = f^*(y) + \frac{\gamma}{2}\|Dy\|^2, \quad (2.16)$$

which we shall refer as the real space smoothing. Since G is not diagonal, the closed-form solution to Eq. (2.15) is not available. For small γ, we approximate the solution by 2 steps: projection and smoothing,

$$\mathrm{prox}_{sf_G^*}(y) \approx (I + s\gamma D^\top D)^{-1}\mathrm{proj}_{\mathcal{S}^*}(y). \quad (2.17)$$

The linear inversion is in fact the Laplacian smoothing that follows the projection to ensure the smoothness of the reconstructed image after each iteration. The real space smoothing is related to the diffusion process and can be approximated by a low-pass filter. Recall

$\mathcal{G}_t(x) := \frac{1}{(4\pi t)^{n/2}} \exp\big(-\frac{\|x\|^2}{4t}\big)$ is a (heat) Gaussian kernel and its Fourier transform $W_t(\xi) = \exp(-4\pi^2 t|\xi|^2)$ is a normalized Gaussian function. Then replacing the linear inversion by the Gaussian kernel and convolution give a fast approximated implementation of Eq. (2.17) for small γ

$$y^{k+1} = \mathcal{G}_{s\gamma} * \mathrm{proj}_{\mathcal{S}^*}\big(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k)\big), \tag{2.18}$$

The projection on $\mathcal{S}^*$ can be computed as follows

$$\mathrm{proj}_{\mathcal{S}^*}(y)_i = \begin{cases} \mathrm{Re}(y_i)^- + \mathbf{i} \cdot \mathrm{Im}(y_i) & \text{if } i \in S, \\ y_i & \text{otherwise.} \end{cases} \tag{2.19}$$

Here $x^- := \min(0, x)$ and $x^+ := \max(0, x)$ for $x \in \mathbb{R}$. The remaining convolution can be done via the efficient DFT by computing

$$\mathcal{G}_\gamma * u = \mathrm{Re}\Big\{\mathcal{F}^{-1}\big(W_\gamma \odot \mathcal{F}u\big)\Big\}. \tag{2.20}$$

Using Eq. (2.14), Eq. (2.19), and Eq. (2.20), we arrive at our first algorithm

Algorithm 3: GPS-R features low-pass filters for smoothing. Here we abuse notation γ to imply the filter. Inspired by OSS [123], we choose a sequence of increasing spatial frequency filters $\{\gamma_l\}$ (a sequence of finer filters). In our experiments, we do 1000 iterations with a total of 10 stages. Each stage contains 100 iterations, in which the filter frequency is held constant. We monitor the R-factor (relative error) during the iterative process, which is defined as

$$R_F(u^k) = \frac{\sum_i \big||\mathcal{F}u^k|_i - b_i\big|}{\sum_i b_i}. \tag{2.21}$$

The reconstruction with minimal R_F at each stage is fed into the next stage. By applying the smoothing on the entire domain, GPS-R can remove noise in real space and obtain the spatial resolution with fine features.

2.3.2 Fourier space smoothing

Another simple choice of G is the diagonal matrix $G = \gamma\,\mathrm{Diag}(r \odot r)$ where $r \in \mathbb{R}^n$ and r_i is the distance in the original 2D lattice between the i-th pixel and the center of image. Note

Algorithm 1 GPS-R: GPS with smoothing in real space.

```
Input: measurements $b$, regularization parameters $\{\gamma_l\}_{l=1}^{10}$, step sizes $s$, $t > 0$
Initialize: $z^0$, $y^0$.
$R_F^{best} = 1$, $z_{best} = z^0$
  for $l = 1, \dots, 10$ do
    $y^0 = y_{best}$, $z^0 = z_{best}$
    for $k = 1, \dots, 100$ do
      $z^{k+1} = \text{prox}_{tg_\sigma}\big(z^k - t\mathcal{F}y^k\big)$
      $y^{k+1/2} = \text{proj}_{\mathcal{S}^*}\big(y^k + s\mathcal{F}^{-1}\big(2z^{k+1} - z^k\big)\big)$
      $y^{k+1} = \mathcal{G}_{\gamma_l} * y^{k+1/2}$
      if $R_F^k < R_F^{best}$, then $R_F^{best} = R_F^k$, $z_{best} = z^k$, end if
    end for
  end for
Output: $\text{Re}\big(\mathcal{F}^{-1}z_{best}\big)^+$
```

that G is not invertible since $r_i = 0$ for the pixel at the center. By Eq. (A.10),

$$f_G(u) = \begin{cases} \sum_i \frac{\left|u - \text{proj}_{\mathcal{S}}(u)\right|_i^2}{2\gamma\, r_i^2} & \text{if } u_i \geq 0 \text{ at the center,} \\ \infty & \text{otherwise.} \end{cases} \tag{2.22}$$

So $f_G(u)$ is a weighted sum of squares penalty on u. The weight is inversely proportional to the squared radius, which is infinity for density in the center. The further the density off the center, the smaller the penalty for the support constraint being violated.

By Parseval's identity, for a square-integrable function u, we have

$$\int \left|\frac{d}{d\xi}\hat{u}(\xi)\right|^2 d\xi = \int |x\, u(x)|^2 dx, \tag{2.23}$$

where $\hat{u}$ is the Fourier transform of u. In the discrete setting, this amounts to

$$\|D\,\mathcal{F}u\|^2 = \|r \odot u\|^2, \tag{2.24}$$

Therefore, by Eq. (2.10),

$$f_G^*(y) = f^*(y) + \frac{\gamma}{2}\|r \odot y\|^2 = f^*(y) + \frac{\gamma}{2}\|D\,\mathcal{F}y\|^2. \tag{2.25}$$

This means that we are smoothing $f^*(y)$ by regularizing with the ℓ_2 gradient of Fourier coefficients of y. We thus refer to it as Fourier space smoothing.

Since G is diagonal, Eq. (2.15) has the closed-form solution

$$\operatorname{prox}_{sf_G^*}(y) = \frac{1}{1 + s\gamma r^2} \odot \operatorname{proj}_{\mathcal{S}^*}(y) \approx \exp(-s\gamma r^2) \odot \operatorname{proj}_{\mathcal{S}^*}(y). \tag{2.26}$$

Define $W_{\frac{s\gamma}{4\pi^2}} := \exp(-s\gamma r^2)$. The above quantity can be approximated by a direct multiplication for small γ. We update y^{k+1} as

$$y^{k+1} = W_{\frac{s\gamma}{4\pi^2}} \odot \operatorname{proj}_{\mathcal{S}^*}\big(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k)\big). \tag{2.27}$$

GPS with smoothing in Fourier space (GPS-F) is summarized in Algorithm 2.

Algorithm 2 GPS-F: GPS with smoothing in Fourier space.

```
Input: measurements $b$, regularization parameters $\{\gamma_l\}_{l=1}^{10}$, step sizes $s, t > 0$.
Initialize: $z^0$, $y^0$.
$R_F^{best} = 1$, $z_{best} = z^0$
  for $l = 1, \dots, 10$ do
    $y^0 = y_{best}$, $z^0 = z_{best}$
    for $k = 1, \dots, 100$ do
      $z^{k+1} = \operatorname{prox}_{tg_\sigma}\big(z^k - t\mathcal{F}y^k\big)$
      $y^{k+1/2} = \operatorname{proj}_{\mathcal{S}^*}\big(y^k + s\mathcal{F}^{-1}(2z^{k+1} - z^k)\big)$
      $y^{k+1} = W_{\frac{s\gamma_l}{4\pi^2}} \odot y^{k+1/2}$
      if $R_F^k < R_F^{best}$, then $R_F^{best} = R_F^k$, $z_{best} = z^k$, end if
    end for
  end for
Output: $\operatorname{Re}\big(\mathcal{F}^{-1}z_{best}\big)^+$
```

We illustrate the core of GPS algorithm as a flowchart in Fig. 2.1

2.3.3 Incomplete measurements

In practice, not all diffraction intensities can be experimentally measured. For example, to prevent a detector from being damaged by an intense X-ray beam, either a beamstop has

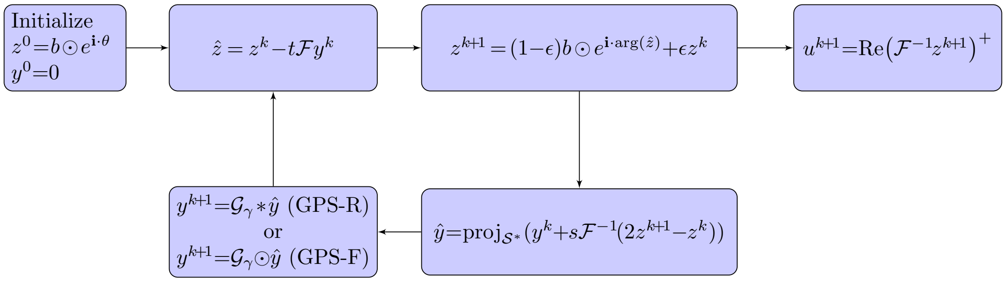


Figure 2.1: Flowchart of GPS algorithm for stage 1. For simplicity, we denote $\epsilon = \frac{\sigma}{\sigma+t}$ in Eq. (2.14). The input z^0 is initialized with random phases and magnitudes are equal the measurements, while y^0 is a zero vector. The $\mathrm{proj}_{\mathcal{S}^*}$ and the convolution $*$ follows Eq. (2.19) and Eq. (2.20) respectively. z^k and y^k with the smallest R_F will be fed as an input into the next stage where a Gaussian kernel $\mathcal{G}_\gamma$ with larger γ is applied.

to be placed in front of the detector to block the direct beam or a hole has to be made at the center of the detector, resulting in missing data at the center [98]. Furthermore, missing data may also be present in the form of gaps between detector panels. For incomplete data, the alternating projection algorithms skip the projection onto the magnitude constraint in this region. Similarly, we only apply the relaxation $g_\sigma = \frac{1}{2\sigma}\sum_i \big||z_i| - b_i\big|^2$ on the known data for GPS. A simple exercise shows that

$$z_i^{k+1} = \begin{cases} \Big(\mathrm{prox}_{tg_\sigma}\big(z^k - t\mathcal{F}y^k\big)\Big)_i & \text{if } b_i \text{ is known,} \\ \big(z^k - t\mathcal{F}y^k\big)_i & \text{otherwise.} \end{cases} \tag{2.28}$$

2.4 Experimental results

2.4.1 Reconstruction from simulated data

We expect GPS to be a reliable PR algorithm in the reconstruction of the images of weakly scattering objects, in particular biological specimens, which have become more popular [96]. Since OSS has been shown to consistently outperform ER, HIO, ER-HIO, NR-HIO [123], we perform both quantitative and qualitative comparisons between GPS and OSS.

Due to the discrete nature of photon counting, experimentally measured data inherently

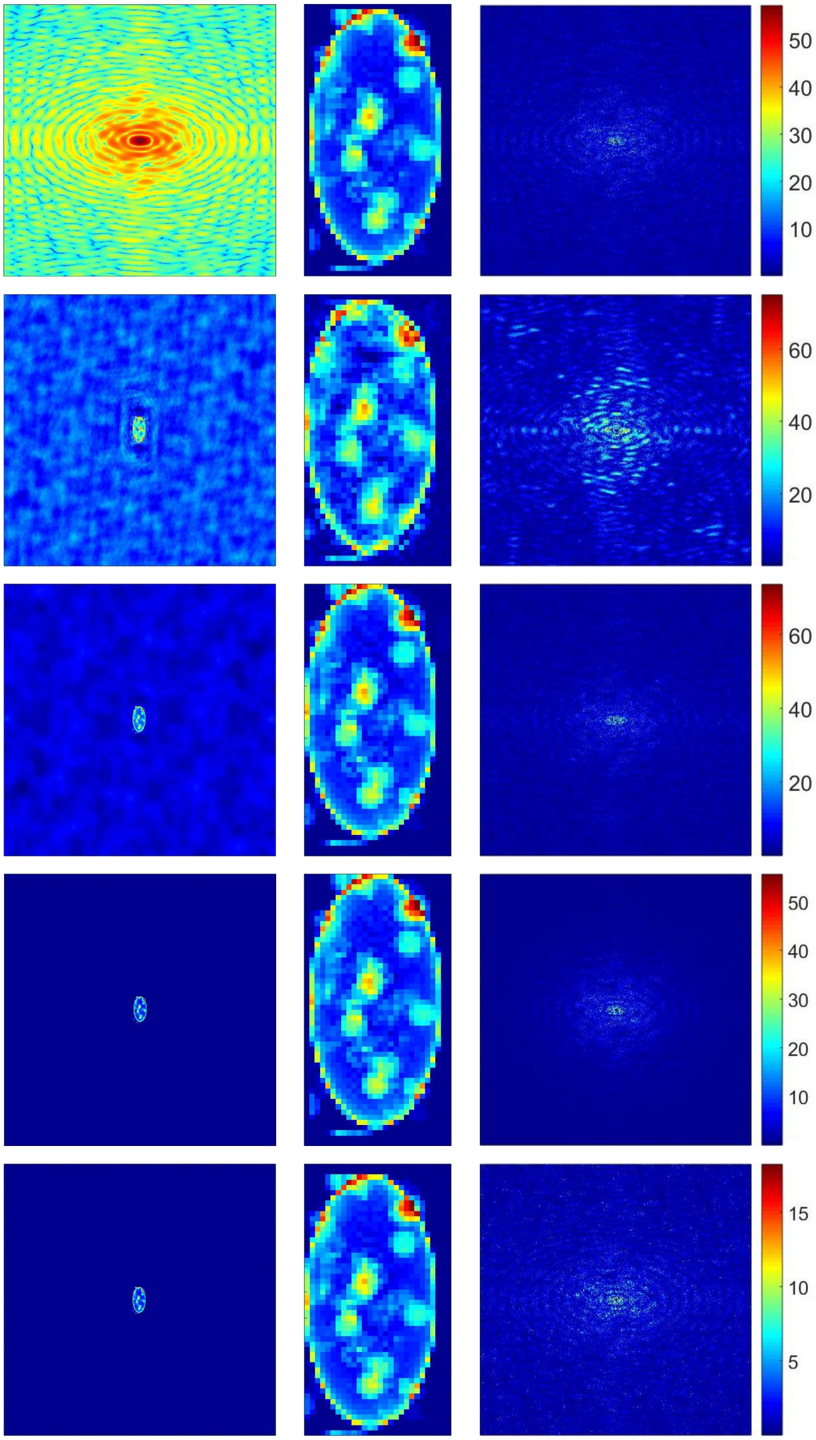


Figure 2.2: First row: vesicle model with log-scaled diffraction pattern (left) , zoom-in image (center) and residual (right). Second row: HIO: $R_F = 12.87\%$, $R_{real} = 21.14\%$. Third row: OSS: $R_F = 6.08\%$, $R_{real} = 3.59\%$. Fourth row: GPS-R $R_F = 5.90\%$, $R_{real} = 2.85\%$. Fifth row: GPS-F $R_F = 5.89\%$ and $R_{real} = 0.7\%$. Third column: residual.

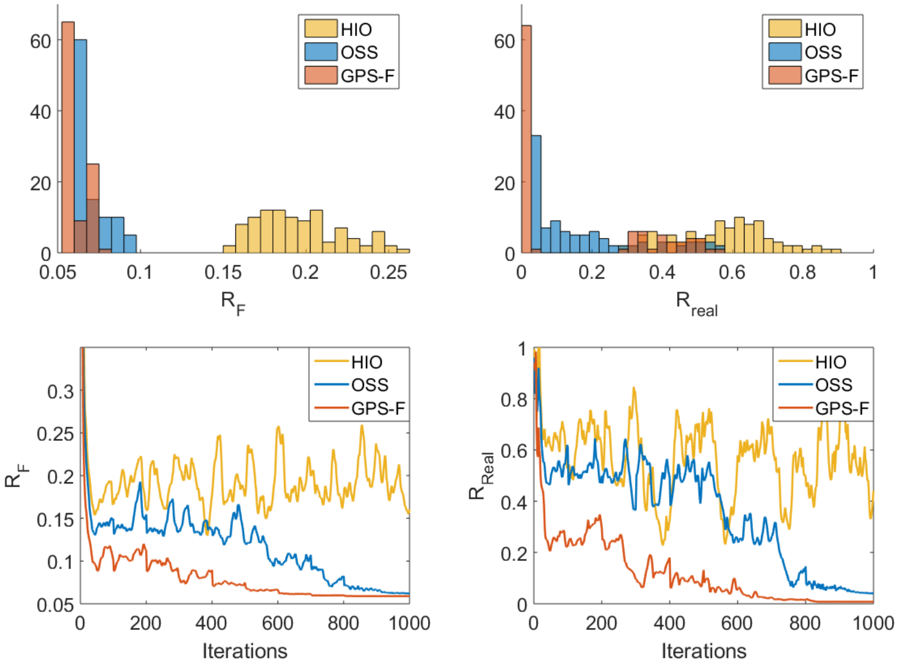


Figure 2.3: Histogram (first row), and convergence (second row) of R_F and R_{real} on Vesicle data using HIO, OSS, GPS. GPS consistently produces smaller R_F and R_{real} than HIO or OSS. Moreover, GPS converges fastest with the fewest oscillations.

contain Poisson noise that is directly related to the incident photon flux on the object. In addition to Poisson noise, the data is also corrupted with zero-mean Gaussian noise to simulate readout from a CCD. Any resulting negative values are set to zero. Therefore, an accurate simulation of b_i can be calculated as

$$b_i = \sqrt{\text{Poisson}\left(\frac{|\mathcal{F}u^o|_i^2 \cdot \text{flux}}{\|\mathcal{F}u^o\|^2}\right) + \mathcal{N}(0,\sigma)}, \tag{2.29}$$

where u^o is the noise-free model, and b are noisy Fourier magnitudes. σ is proportional to the readout noise. [133]. We use R_{noise} to quantify the relative error with respect to the noise-free Fourier measurements

$$R_{noise} = \frac{\sum_i \left|b_i - |\mathcal{F}u^o|_i\right|}{\sum_i |\mathcal{F}u^o|_i}, \tag{2.30}$$

For simulation, the Fourier magnitudes of the vesicle model are corrupted with Poisson and Gaussian noises by Eq. (2.29). The corresponding relative error is $R_{noise} \approx 6\%$.

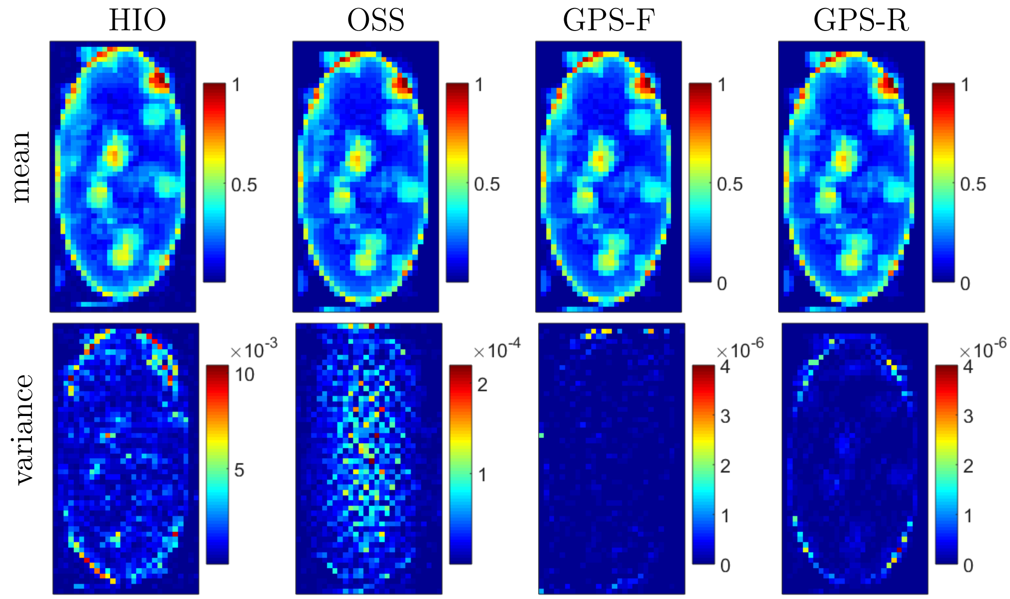


Figure 2.4: The (scaled) means and variances of the best five of 500 independent reconstructions by HIO, OSS, GPS-F and GPS-R

In some cases, the reconstructed image yields a small relative error R_F but has low quality. This is the issue of over-fitting, an example of which can be seen in certain reconstructions using ER-HIO [123]. Smoothing is a technique to avoid data over-fitting. To validate results and show that our algorithm does not develop over-fitting, we measure the difference between the reconstructed image and the model by

$$R_{real} = \frac{\sum_i \left| u_i^k - u_i^o \right|}{\sum_i u_i^o}. \tag{2.31}$$

In addition, we also look at the residual $\text{Res} = \| |\mathcal{F}u| - b \|$ which measures the difference between the Fourier magnitudes of the reconstructed image and the experimental measurements. The residual can validate the noise-removal model, telling how much noise is removed. Fig. 3.1 shows the reconstruction of vesicle model from simulated noisy data using HIO, OSS, GPS-R, and GPS-F. GPS-F and GPS-R obtain lower R_F and R_{real} than OSS. Moreover, GPS-F can get very close to the ground truth with $R_F = 5.90\%$ and $R_{real} = 0.7\%$. In addition to lower R values, GPS-R and GPS-F converge to zero outside the support. They both obtain lower residuals than OSS, specifically GPS-F produces the least residual. If we choose larger parameter for the ℓ_2 gradient regularizer in Fourier space, we will get a smoother residual. Overall for realistic Gaussian and Poisson noise in Fourier space, GPS-F is a suitable noise-removal model.

Fig. 2.3 shows the histogram and the convergence of R_F and R_{real} on 100 independent, randomly seeded runs using HIO, OSS and GPS on the simulated vesicle data. The histogram shows that GPS is more consistent and robust than OSS. It has a higher chance to converge to good minima with lower R_F and R_{real} than OSS. Furthermore, R_F and R_{real} of OSS scatter widely due to the initial value dependency. In contrast, GPS is more selective and less dependent on initial values. R_F and R_{real} of GPS are seen at a lower mean minimum with less variance.

Similar to HIO and ER, OSS keeps oscillating until a finer low-pass filter is applied. In contrast, GPS converges faster and is less oscillatory than OSS. In the presence of noise, alternating projection methods (ER, HIO, OSS) keep oscillating but do not converge. Applying smoothing and replacing the measurement constraint by the least squares fidelity term $g_\sigma(z) = \frac{1}{2\sigma}\||z| - b\|^2$ helps to reduce the oscillations; hence, the method can converge to a stationary point. Note that larger σ reduces more oscillations, but also decreases the chance to escape from local minima. Alternating projection methods have $\sigma = 0$ since they impose measurement constraints. GPS obtains both smaller R_F, R_{real}, and lower variance. Even though R_F are close to each other, R_{real} of GPS is much smaller than OSS. This means GPS recovers the vesicle cell with higher quality than OSS.

Next, we test the consistency of GPS algorithm for vesicle model. Since OSS has been shown to outperform HIO, ER-HIO and HR-HIO [123], it suffices to carry out the comparison among HIO, OSS and GPS. We look for the mean and variance of the best five of 500 independent reconstructions by HIO, OSS, GPS-F and GPS-R (lowest R_F). Fig. 2.4 shows that GPS has much smaller variance than OSS and HIO. These tests shows that GPS is more reliable and consistent than OSS and HIO.

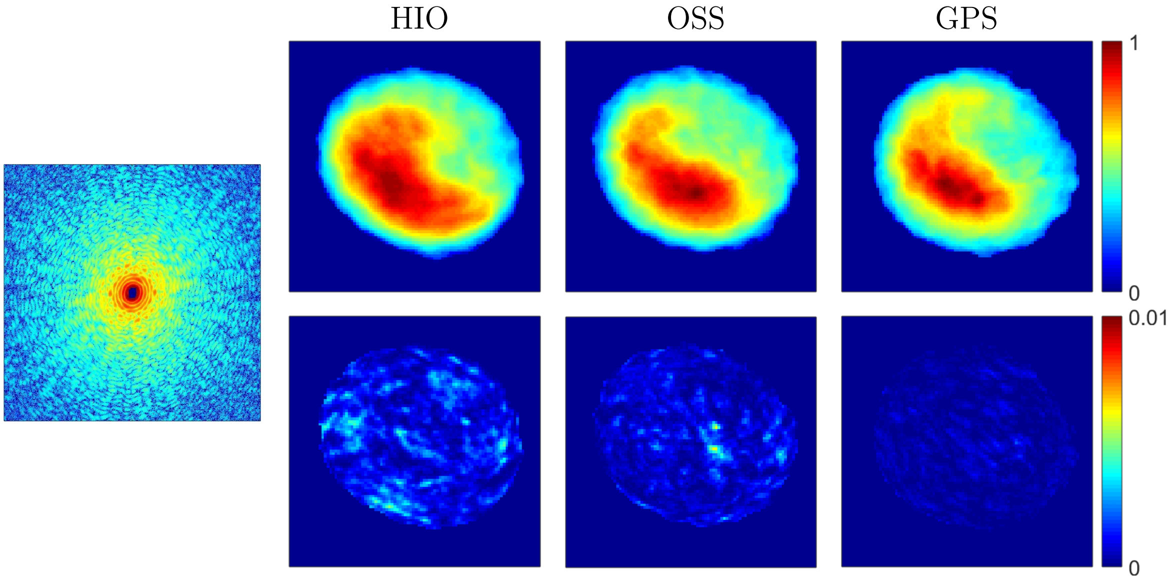


Figure 2.5: S. *pombe* yeast spore log-scaled diffraction pattern, size 500 × 500 (top). Means (first row) and variance (second row) are computed from the best 5 of 500 independent reconstructions. HIO: $R_F = 15.697\% \pm 0.526\%$, OSS: $R_F = 9.775\% \pm 0.202\%$, and GPS: $R_F = 8.672\% \pm 0.025\%$.

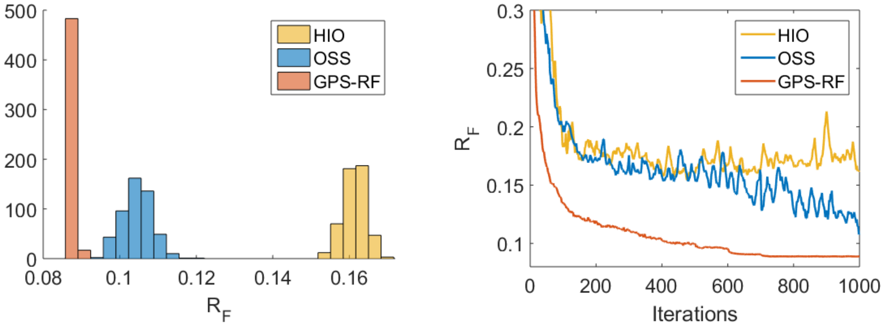


Figure 2.6: Top: histogram of R_F in 500 independent runs (top). Bottom: the convergence curves of a singe construction of R_F on S. *pombe* yeast spore by GPS-RF, OSS, and HIO.

2.4.2 Reconstructions from experimental data

2.4.2.1 S. *pombe* yeast spore

To demonstrate the applicability of GPS to biological samples, we do phase retrieval on the diffraction pattern in Fig. 2.5 taken of a S. *Pombe* yeast spore from an experiment done using beamline BL29 at SPring-8 [68]. We do 500 independent, randomly seeded reconstructions with each algorithm and record R_F, excluding the missing center. We choose default parameters for these experiments: $t = 1$, $s = 0.9$. The sequence of low-pass filters are chosen to be the same as in OSS [123]. For the first 400 iterations, $\sigma = 0.01$, then is increased to $\sigma = 0.1$ for the remaining 600 iterations. The left column of Fig. 2.5 is the mean of the best 5 reconstructions obtained by the respective algorithm. The right column shows the variance of the same 5 reconstructions. It is evident from the variance that GPS achieves more consistent reconstructions. Figure 2.6 shows the histogram and convergence of R_F. We can conclude that not only are GPS-R results the most consistent, but also the most faithful to the experimental data.

2.4.2.2 Nanorice

To demonstrate the generality of GPS, we also do testing with experimental data of inorganic samples. The diffraction patterns shown in the top row of Fig 2.7 from ellipsoidal iron oxide nanoparticles (referred to as 'nanorice1'and 'nanorice2') were taken at the AMO instrument at LCLS at an X-ray energy of 1.2keV [76]. This data is freely available online on the CXIDB [89]. We choose default parameters for these experiments: $t = 1$, $s = 0.9$. The sequence of low-pass filters are chosen to be the same as in OSS [123]. The fidelity parameter σ is chosen small for the first 800 iterations, specifically $\sigma \in [0, 0.01]$, to produce oscillations which is necessary for the algorithm to skip bad local minima. Once the reconstruction arrives at a good local minimum region, we increase σ to reduce oscillations. This later value of σ depends on noise level and data. We test different values of σ and $\sigma = 1$ has been found to

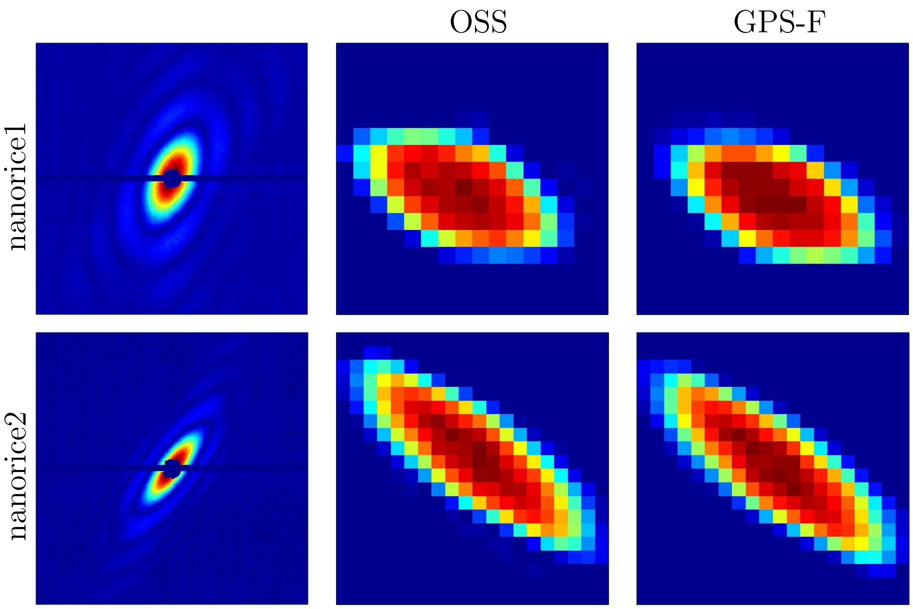


Figure 2.7: Diffraction pattern of nanorice1 and nanorice2 253 × 253 (first column) and reconstructions using OSS: $R_F = 18.23\%, 16.32\%$ and GPS-F: $R_F = 17.40\%, 15.83\%$ respectively. GPS obtains less noise on the boundary and lower relative error R_F.

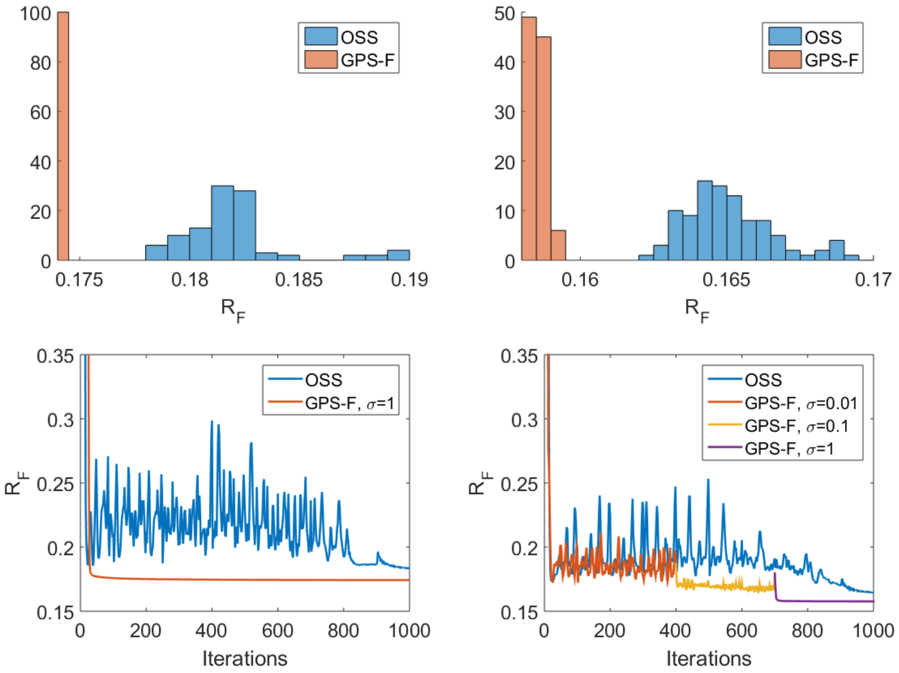


Figure 2.8: Histogram (first row) and convergence (second row) of OSS and GPS-F on nanorice1 (first column) and nanorice2 (second column). The results of HIO are omitted due to lack of convergence.

be the optimal for both nanorice data. Figure 2.7 shows OSS(second row) and GPS-F(third row) reconstructions of the two nanorice particles. Figure 2.8 shows again that GPS obtains more consistent and faithful reconstructions as compared to those obtained by OSS. GPS-F with $\sigma = 1$ converges to lower relative error than OSS at all times. OSS cannot get lower relative error because $\sigma = 0$ does not work for this case. In general, alternating projection methods do not treat noise correctly. For example, in this case, HIO keeps oscillating but does not converge. Therefore, its results are omitted here. The better approach, OSS model, can reduce oscillations by smoothing but this is not enough. In contrast, the least squares $g_\sigma(z) = \frac{1}{2\sigma}\||z|-b\|^2$ of GPS works for noise removal since relaxing the constraints allows GPS to reach lower relative error. The values of σ depend on noise level and type. To optimize the convergence of GPS-F on nanorice2, we apply $\sigma = 0.01$ for the first 400 iterations, $\sigma = 0.1$ for the next 300 iterations, and $\sigma = 1$ for the last 300 iterations. This test shows the effect of σ on the convergence. OSS ($\sigma = 0$) oscillates the most. GPS with $\sigma = 0.01, 0.1, 1$ oscillates less and less. As σ increases, GPS also gets to lower R_F. The algorithm finally reaches a stable minimum as σ goes up to 1. Continuing to increase σ does not help with R_F. Choosing large σ in the beginning may reduce oscillations but also limit the mobility to skip local minima. We recommend start with small σ and then gradually increase it until the iterative process reaches a stable minimum.

2.5 Conclusion

In this work, we have developed a fast and robust phase retrieval algorithm GPS for the reconstruction of images from noisy diffraction intensities. Similar to [147], the Moreau-Yosida regularization was used to relax the hard constraints considered in the noiseless model. GPS utilizes a primal-dual algorithm and a noise-removal technique, in which the ℓ_2 gradient smoothing is effectively applied on either real or Fourier space of the dual variable. GPS shows more reliable and consistent results than OSS, HIO for the reconstruction of weakly scattered objects such as biological specimens. Looking forward, we aim to explore the role of dual variables in non-convex optimization. Smoothing the dual variable, which is equivalent to smoothing the gradient of convex conjugate, represents a new and effective technique that can in principle be applied to other non-smooth, non-convex problems.

In the next chapter, we will explore tomography, another problem in image processing. In contrast to phase retrieval, tomography is linear and convex. With the development in electron tomography (ET), resolution can reach further to the scale of micrometer and Angstrom. However, missing data in ET causes the problem under-determination and makes a challenge to reconstruction algorithms. We will use l_2 minimization approach with extension to continuous tomography in order to improve "forward & back projection" operators.

CHAPTER 3

RESIRE: real space iterative reconstruction engine for Tomography

Tomography has made a revolutionary impact on diverse fields, ranging from macro-/mesoscopic scale studies in biology, radiology, plasma physics to the characterization of 3D atomic structure in material science. The fundamental of tomography is to reconstruct a 3D object from a set of 2D projections. To solve the tomography problem, many algorithms have been developed. Among them are methods using transformation technique such as computed tomography (CT) based on Radon transform and Generalized Fourier iterative reconstruction (GENFIRE) based on Fourier slice theorem (FST), and direct methods such as Simultaneous Iterative Reconstruction Technique (SIRT) and Simultaneous Algebraic Reconstruction Technique (SART) using gradient descent and algebra technique. In this paper, we propose a hybrid gradient descent to solve the tomography problem by combining Fourier slice theorem and algebraic technique. By using simulated and experimental data, we show that the state-of-art RESIRE can produce more superior results than previous methods; the reconstructed objects have higher quality and smaller relative errors. More importantly, RESIRE can deal with partially blocked projections rigorously where only part of projection information is provided while other methods fail. We anticipate RESIRE will not only improve the reconstruction quality in all existing tomographic applications, but also expand the tomography method to a broad class of functional thin films. To facilitate the use of RESIRE, the MATLAB code of RESIRE algorithm is freely accessible on our website.

3.1 Introduction

Tomography has widespread applications in physical, biological and medical sciences. Electron tomography lately has become a revolution in high-resolution 3D imaging of physical and biological samples. In the physical sciences, atomic electron tomography (AET) has been developed to determine the 3D atomic structure of crystal defects such as grain boundaries, anti-phase boundaries, stacking faults, dislocations, chemical order/disorder and point defects, and to precisely localize the 3D coordinates of individual atoms in materials without assuming crystallinity [101, 125, 29, 143, 61]. Using the advanced X-ray sources, coherent diffractive imaging (CDI) methods [95] can be combined with tomographic reconstruction for 3D quantitative imaging of thick samples with resolutions in the tens of nanometers [99, 106, 68, 34, 69, 100, 65].

Many reconstruction algorithms have been developed for tomography in the last decades. Among them is the famous filtered back projection (FBP) [53, 72, 64], a one-step method that can work fast and greatly when there are a large number of projections and no missing data. The fundamental of FBP is based on the relationship between Fourier slice theorem (FST) and Radon transform [17, 85] whose inverse corresponds to the reconstruction of 3D objects. Since working on the polar coordinate, inverse Radon transform requires the knowledge of all tilt angles in order to obtain a good reconstruction. In the cases of low radiation dose and geometric restrictions which produce inadequate data, FBP fails to reconstruct faithful objects and suffers artifacts.

Real space iterative methods such as algebraic reconstruction technique (ART) [60] can partially compensate for the artifacts. Modifications of ART such as simultaneous algebraic reconstruction technique (SART) [4] and simultaneous iterative reconstruction technique (SIRT) [56] show significant improvement compared to FBP. These algorithms are developed based on a least square optimization problem to minimize the difference between measured and calculated projections. Gradient descent, an iterative method, is employed to refine reconstructions. Simplicity, fast running-time, and parallelization are the advantages of these methods.

Lately, Fourier-based iterative algorithms [145, 97, 108], using information in both real and Fourier space as constraints in implementations, dramatically improve the performance. Equal slope tomography (EST) [97], an example of such algorithm, has been successfully applied in AET to reconstruct the 3D atomic structure of crystal defects in materials, which shows successful recovery of diffraction pattern in the missing wedge direction [125, 101, 29, 143]. Additionally, with significantly lower radiation dose, EST can still produce reconstructions comparable to modern medical CT techniques [48, 151, 49]. However, the drawback of EST is the strict requirement that the tilt angles must follow equal slope increments along a single tilt axis, which limits its broader applications.

The EST development is followed by GENFIRE [145, 112], a generalized Fourier iterative reconstruction algorithm, which has been reported to produce high-resolution 3D imaging from a limited number of 2D projections. Using FST, GENFIRE transforms measurement constraint (measured projections) on real space into Fourier constraint on reciprocal space. The problem then becomes finding a 3D image that satisfies spatial constraints (real, nonnegativity and compact support) and Fourier constraint. The alternating projection method, known as GENFIRE, is exploited to solve this classical two-constraint minimization problem. In addition, GENFIRE requires gridding [54, 132] to obtain on-Cartesian-grid Fourier values, and oversampling [94, 93] to improve accuracy of the gridding. These procedures are major drawbacks of GENFIRE since they require more memory (due to oversampling) and cause numerical error (caused by interpolation).

In this research, we motivate ourselves to develop a highly accurate Tomography gradient descent method by solving the least square error (LSE) problem. In the light of calculus of variations [8], we derive the gradient of the “sum of squared errors” (SSE). FST and interpolations are employed to compute the gradient which involves “forward projection” and “back projection,” termed that we borrow from ART [4]. The 3D image is then reconstructed iteratively with gradient descent method. These crucial points institute our proposed real space iterative reconstruction engine (RESIRE).

3.2 Method

3.2.1 RESIRE algorithm

3.2.1.1 Gradient descent

Tomography is formulated as a least square optimization problem, minimizing the sum of squared errors (SSE)

$$\min_{O} \ \varepsilon(O) = \frac{1}{2} \sum_{\theta} \|\Pi_\theta(O) - b_\theta\|_F^2 \tag{3.1}$$

where Π_θ is the linear projection operator with respect to the tilt angle θ and b_θ is the corresponding measure projection. In a simple case where there is one rotation axis (either x,y, or z), tilt angles are just single values. In the case of multiple rotation axes, a tilt angle can be expressed as Euler angles, a set of three angles, to describe the rotation of objects with respect to a fixed coordinate system, and this topic will be discussed in later part. The error metric ε is a net sum of every ε_θ caused by each measure projection b_θ. We get a more explicit form for each error ε_θ

$$\varepsilon_\theta(O) = \frac{1}{2} \sum_{x,y} |\Pi_\theta(O)\{x,y\} - b_\theta\{x,y\}|^2 \tag{3.2}$$

We compute the gradient of the error metric ε w.r.t. the object O in the sense of calculus of variation method [8]. The following derivation is a simplified (discrete) version of a continuous problem. Furthermore, in order for the derivation to be logical, we must assume the object is continuous. At a pixel or spatial variable $\{u, v, w\}$, the gradient reads:

$$\begin{aligned} \frac{\partial \varepsilon_\theta}{\partial O\{u,v,w\}} &= \sum_{x,y} \Big(\Pi_\theta(O)\{x,y\} - b_\theta\{x,y\}\Big) \frac{\partial}{\partial O\{u,v,w\}} \sum_{z} O\Big\{R_\theta \begin{bmatrix} x \\ y \\ z \end{bmatrix}\Big\} \\ &= \Pi_\theta(O)\{x,y\} - b_\theta\{x,y\} \qquad \text{where} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = R_\theta \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad \text{for some } z \end{aligned} \tag{3.3}$$

The first line of Eqn. 3.3 is obtained by vanilla chain rule while the second line uses an assumption that pixels of object are independent from each other, i.e. $\partial O\{x, y, z\}/\partial O\{u, v, w\} =$

1 if $\{x, y, z\} = \{u, v, w\}$ and 0 otherwise. Using the fact that the transpose of a rotation matrix is its inverse, we derive the following transformation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} R_{1,1} & R_{2,1} \\ R_{1,2} & R_{2,2} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} R_{3,1} \\ R_{3,2} \end{bmatrix} w \tag{3.4}$$

where $R_{i,j}$ is the $(i, j)^{th}$ element of the rotation matrix R_θ. Note that all the pixels $\{u, v, w\}$ are on the Cartesian grid (integer coordinates) while $\{x, y\}$ are not. To understand the formula, we decompose the procedure into two steps and borrow terminology in ART for explanation.

Step 1: computing "forward projection." The projection $\Pi_\theta(O)$ is computed from the current iterative object O via FST. To improve the accuracy, the 3D object is padded with zeros before the Fourier transform. The oversampling ratio of this padding procedure is in range $[2, 4]$ providing a good trade-off between accuracy and efficiency. We now apply FST, i.e. the 2D Fourier transform of a projection w.r.t tilt angle θ represents a 2D plane slicing through the origin of the 3D Fourier transform of that object. Taking the inverse Fourier transform of this 2D slice and cropping the size back to the original projection size, we obtain the desired projection. This process is similar to the "forward projection" step in ART. However, ART presents the "forward projection" as a matrix (vector) multiplication while RESIRE employs FST and interpolation. As a result, accuracy of the "forward projection" is improved with RESIRE.

Step 2: computing "back projection." The measured projections are subtracted from the forward projections to obtain the difference $\Pi_\theta(O) - b_\theta$. Now the task is to compute the gradient $\Pi_\theta^T(\Pi_\theta(O) - b_\theta)$, i.e. apply Π^T on $\Pi_\theta(O) - b_\theta$. Eqn. 3.4 shows how to "back project" a 2D image to a 3D image. Specifically, each (u, v) slice of the gradient (with w varied) is a linear transformation of the difference. Further notice that those (u, v) slices where $w \neq 0$ are translations of the (u, v) slice at the origin ($w = 0$). The amount of translation is exactly the zero order term $[R_{3,1} \;\; R_{3,2}]^T w$ in Eqn. 3.4.

On the other hand, SIRT presents the "back projection" as a row-normalized transpose of the "forward projection"

$$O^{k+1} = O^k - t\sum_i \frac{a_i O^k - b_{\theta,i}}{a_i a_i^T} a_i^T = O^k - t\hat{A}^T\Big(AO^k - b_\theta\Big) \tag{3.5}$$

where A is the projection operator defined by ART, $\{a_i\}_i$ are its row vectors, and $\hat{A}$ is its row-normalized matrix, i.e. the row vector $\hat{a}_i$ of $\hat{A}$ is normalized as $\hat{a}_i = a_i/(a_i^T a_i)$. $b_{\theta,i}$ is the i^{th} element of b_θ. SIRT in fact uses vector multiplication form (the first equality in Eqn. 3.5) rather than matrix multiplication form (the second equality). This strategy allows SIRT to enforce parallel computing on CPU or GPU [109]. However, the downside of SIRT is its limitation to 2D case (single tilt axis). When extending to the multiple-tilt-axis case, computing and storing matrix A is infeasible. Furthermore, A is defined differently than Π_θ as ART does not require the linear transformation as RESIRE does. Instead, ART can choose arbitrary equidistant points for its line integral approximation [4]. The restricted linear transformation in RESIRE is the major difference between these two algorithms and explains why RESIRE outperforms the other.

To finish this part, we recap that RESIRE derives the gradient of the SSE in Eqns. 3.1 & 3.2 through calculus of variation to improve the accuracy of the forward and back projections. Furthermore, interpolation plays an important role in our algorithm as it is required in both steps. It first helps to apply FST by computing 2D slices through the origin of a 3D Fourier transform in "forward projection" and then compute linear transformations in "back projection".

3.2.1.2 Step size analysis

This is the tricky part of the gradient descent method. We hope to approximate the Lipschitz constant L of the gradient such that the following inequality holds

$$\|\nabla\varepsilon(O_1) - \nabla\varepsilon(O_2)\| \leq L\|O_1 - O_2\| \quad \forall\, O_1,\, O_2 \tag{3.6}$$

Hence, we can choose the step size to be $1/L$ for convergence guarantee. Assuming that the sum of intensity is conserved under projection operator and using the fact that l_2 norm is invariant under rotation, we can reduce the step size analysis to the simple case where tilt angle $\theta = 0$. We further simplify the analysis to 2D case under assumption of one tilt axis. Without loss of generality, we can assume the 2D object has dimension $N_z \times N_x$ and the projection is applied along z direction (the first dimension). Hence, the forward projection matrix is just a simple row vector $\Pi = \mathbf{1}^{1\times N_z}$ where all elements are ones. Similarly, the back projection operator $\Pi^T = \mathbf{1}^{N_z\times 1}$ is a column vector of ones, i.e. Π^T/N_z uniformly distributes 1D image back to 2D image. Then the composed operator $\Pi^T\Pi = \mathbf{1}^{N_z\times N_z}$ is an all-one matrix. We can easily check its l_2 norm $\|\Pi^T\Pi\| = N_z$. We now derive the Lipschitz constant

$$\|\nabla\varepsilon_\theta(O_1) - \nabla\varepsilon_\theta(O_2)\| = \|\Pi^T\Pi(O_1 - O_2)\| \leq \|\Pi^T\Pi\|\,\|O_1 - O_2\| = N_z\,\|O_1 - O_2\| \tag{3.7}$$

Since there are n projections that contribute to the gradient, the lipschitz constant increases by a factor of n, i.e. the accumulated Lipschitz constant becomes $L = nN_z$. Then the gradient descent step size can be presented as $t/(nN_z)$ where $t \in (0, 1]$ is normalized step size. When objects are sparse, larger t can work. Our experimental results show that the algorithm still converges with $t = 2$. This value of t is also selected for all experiments in this paper. We finalize our gradient descent method by showing the governing formula. At the $(k+1)^{th}$ iteration, the update reads

$$O^{k+1}\{u,v,w\} = O^k\{u,v,w\} - \frac{t}{nN_z}\sum_\theta \Big(\Pi_\theta(O^k)\{x,y\} - b_\theta\{x,y\}\Big) \tag{3.8}$$

As mentioned above, tilt angles are single values in the case of one rotation axis. However, multiple-rotation-axis Tomography appears more often in practice. In this case, Euler angles, a set of three angles, are used to describe the orientation of a 3D object with respect to a fixed coordinate system. Hence a rotation can be presented as a sequence of rotations about standard x,y and z axes. We prefer the Euler convention ZYX to demonstrate object rotations, i.e. each tilt angle is expressed as a triple set (ϕ, θ, ψ) which corresponds to rotations about z, y, and x axes in this respected order. The composed rotation matrix $R_{\{\phi,\theta,\psi\}} = Z_\phi Y_\theta X_\psi$ is a product of 3 basic rotation matrices Z_ϕ, Y_θ and X_ψ where Y_θ representing a rotation matrix about Y axis by an angle θ. Similar definitions are for Z_ϕ and X_ψ.

3.2.1.3 Pseudocode

The method is summarized in algorithm 3 as a pseudocode. Bi-linear interpolation is used in both "forward projection" and "back projection" due to its simplicity, efficiency, and fast running time.

Algorithm 3 RESIRE algorithm

```
Input: set of N projections $\{b_{\theta_i}\}_{i=1}^N$ and corresponding tilt angles $\{\theta_i\}_{i=1}^N$, number of iterations $K$, step size t
Initialize: $O^0$.
  for $k = 1, \dots, K$ do
    for $i = 1, \dots, N$ do
      compute "forward projection" $\Pi_{\theta_i}(O^k)$ using FST and interpolation.
      compute residual    $R_i(O^k) = \Pi_{\theta_i}(O^k) - b_{\theta_i}$
      compute "back projection" (or gradient) $\nabla\varepsilon_{\theta_i}(O^k)$ from $R_i(O^k)$ by formula 3.3 using interpolation.
    end for
    $O^{k+1} = O^k - \frac{t}{nN}\sum_{i=1}^N \nabla\varepsilon_{\theta_i}(O^k)$
  end for
Output: $O^K$
```

3.3 Experimental results

3.3.1 Reconstruction from a numerical simulation: biological vesicle

A model of a biological vesicle of size $64 \times 64 \times 64$ pixels was used in this test to demonstrate the performance of the RESIRE algorithm. We assume Y is the only rotation axis, then the other two Euler angles ϕ and ψ are identically zeros. Using FST, we calculate 41 projections whose tilt angles θ range from -70^0 to 70^0 with step size 3.5^0. Gaussian noise was added to projections to make simulation data more realistic. Reconstructions are performed using RESIRE, GENFIRE, FBP and SIRT. To monitor convergence, we use R-factor R_F

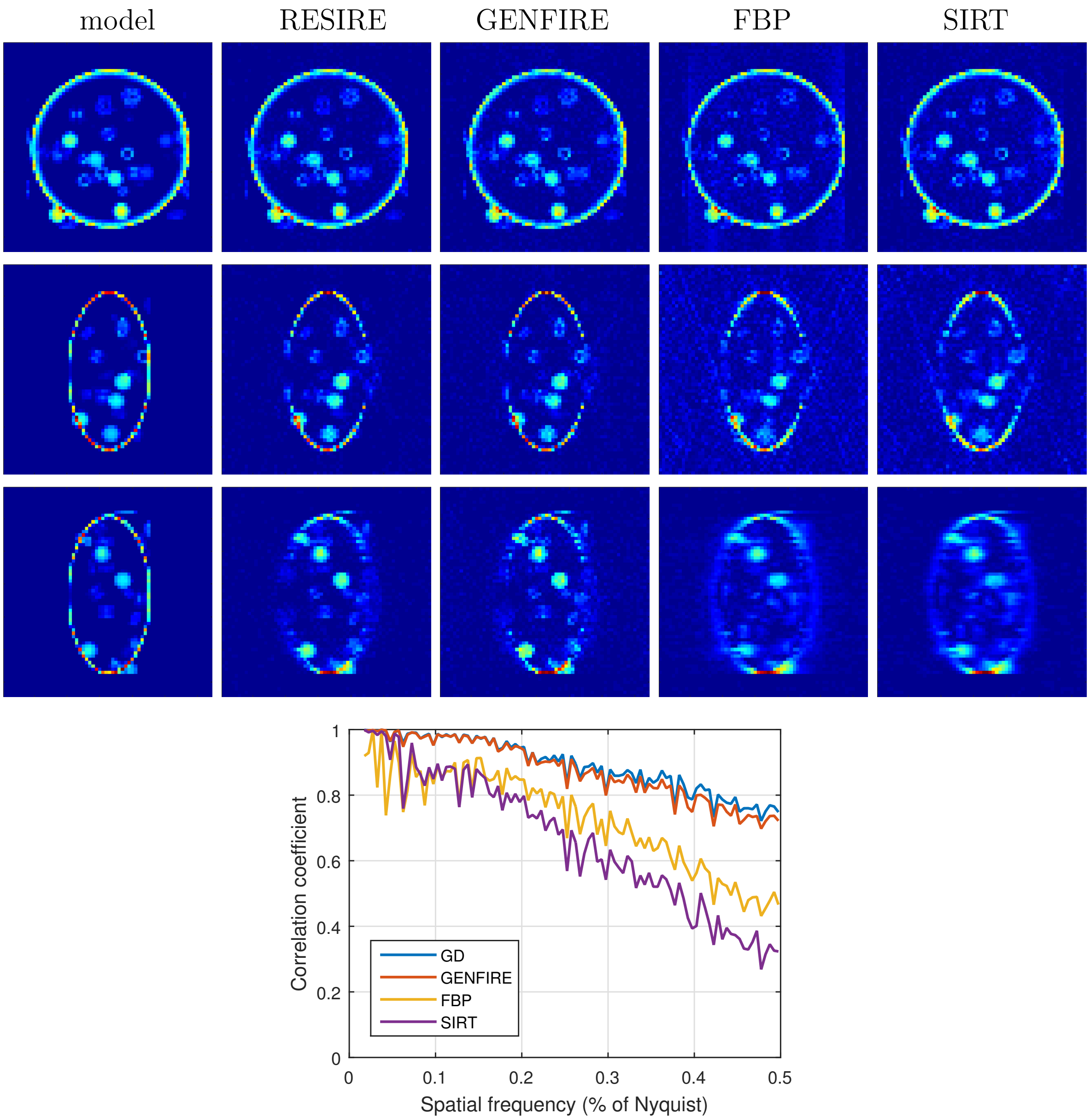


Figure 3.1: A vesicle model of 64 × 64 × 64 pixel size from which 41 noisy projections simulated and 3D reconstructions using RESIRE, GENFIRE, FBP and SIRT. (a-c) Three 10-pixel-thick central slices of the vesicle model in the XY, XZ and YZ planes, respectively. The corresponding three reconstructed slices with RESIRE (d-f), GENFIRE (g-i), FBP (j-l), and SIRT (m-o), where the missing wedge axis is along the z-axis. The FSC between the reconstructions and the model shows that RESIRE produces a more faithful reconstruction than the other algorithms at all spatial frequencies.

(relative error) as a metric to compare the relative difference between forward and measured projections.

$$R_F = \frac{1}{n}\sum_{\theta} \Big[\sum_{x,y} |\Pi_\theta(O)\{x,y\}| - b_\theta\{x,y\}| \Big/ \sum_{x,y} |b_\theta\{x,y\}|\Big] \tag{3.9}$$

All iterative reconstruction algorithms RESIRE, GENFIRE and SIRT perform with 400 iterations. ASTRA Toolbox [138, 137] is employed to achieve SIRT reconstruction while FBP reconstruction is performed using IMOD [92]. The obtained R-factors are 28.59, 38.98, 35.44, 30.46% for RESIRE, GENFIRE, FBP and SIRT, respectively. Since directly minimizing the least square, RESIRE and SIRT obtain lower R_F than the other methods. FBP obtains the largest R_F because it is a one-step method. The R_F of GENFIRE is also high due to the numerical error in the gridding.

Figure 3.1 show 10-pixel-thick central slices of the 3D reconstructions in the XY (first row), XZ (second row) and YZ (third row) planes by RESIRE, GENFIRE, FBP and SIRT respectively, where the z-axis is the missing wedge direction. Because there is no missing data in this direction, the XY central slices from all methods exhibit good agreement with the model. However, along the missing wedge direction z, FBP and SIRT reconstructions show degradation, noise, and missing wedge artifact (shadow and smearing) while features obtained by RESIRE and GENFIRE appear to be finer and more isotropic. GENFIRE reduces but still shows some artificial noise in the XZ and YZ central slices because of the high numerical error in the gridding. On the other hand, RESIRE reduces significant noise, and obtains the finest and clearest images.

To quantify the results, we use Fourier shell correlation (FSC) to measure the normalized cross-correlation coefficient between the model and each reconstruction over corresponding shells in Fourier space (as a function of spatial frequency correlation). The FSC curves in Figure 3.1 confirms that the RESIRE reconstruction is superior at all spatial frequencies compared to the other algorithms.

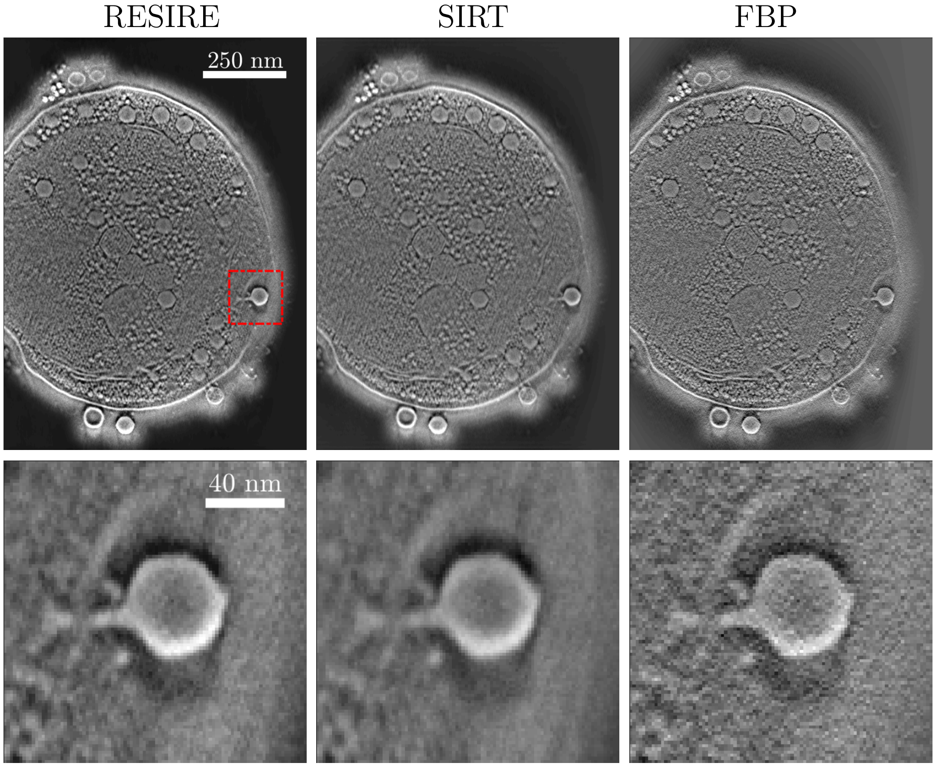


Figure 3.2: Reconstruction of a frozen-hydrated marine cyanobacterium, capturing the penetration of a cyanophage into the cell membrane. (a-c) 5.4-nm-thick (3-pixel-thick) slices of the cell in the XY plane reconstructed by RESIRE, SIRT, and FBP respectively, where the y axis is the tilt axis and the beam direction is along the z axis. (d-e) Magnified views of the penetration of a cyanophage for the RESIRE, SIRT and FBP reconstructions in the XY, respectively. The high quality projections result in good reconstructions. However, since RESIRE and SIRT use iterations, their reconstructions and R-factors are much better than FBP. $R_F = 2.50\%$, 4.39% and 40.63% for RESIRE, SIRT and FBP respectively. RESIRE obtains not only smaller R_F but also better quality and resolution than SIRT and FBP. The membrane surrounding the cyanophage is more visible with RESIRE.

3.3.2 Reconstruction of a frozen hydrated cell

We run the reconstruction of a frozen-hydrated marine cyanobacterium in a late stage of infection by cyanophages. 42 projections were collected from a single tilt series, ranging from -58^o to $+65^o$ with an increment of 3^o. The projections were binned by 4×4 pixels and resulted in images with approximately 1.8×1.8 nm^2 per pixel. The detail of experiment and pre-processing were explained in the GENFIRE paper [112]. In this single tilt axis case, we can perform a comparison among RESIRE, SIRT and FBP algorithms. Both RESIRE and SIRT reconstructions were performed with 150 iterations.

Figure 3.2 shows a 5.4-nm-thick (3-pixel-thick) slice of the 3D reconstruction in the XY plane, capturing the penetration of a cyanophage into the cell membrane during the infection process. The SIRT and RESIRE reconstructions look similar and much better than FBP. The FBP reconstruction suffers peripheral noise through which features could not be seen clearly. This is because FBP is a one-step method; hence, it cannot resolve the feature well with missing data. On the other hand, both RESIRE and SIRT iterate to fit the measurements (measured projections). As a result, features in RESIRE and SIRT are resolved with higher quality. These results are also consistent with R-factors where RESIRE (2.5%) and SIRT (4.39%) obtain much lower values than FBP (40.46%). Because the projections have high quality and the noise level is low, both RESIRE and SIRT achieve good results. Therefore, their difference is marginal, and it is difficult to justify with the naked eye. Under careful comparison, the reconstruction by RESIRE can be verified to be better with sharper and clearer features, especially at the edges. In addition to this reasoning, the lowest R-factor reveals that RESIRE reconstruction fit the measurement better than other methods.

3.3.3 Reconstruction from an experimental data: Seriatopora aculeata coral skeleton

The Seriatopora aculeata sample used in this test is a pencil-thick, short and tapered branch termed nubbin. Experiment was performed to obtain high-angle annular dark-field (HAADF-STEM) tomography data which includes a total of 69 images collected with a tilt

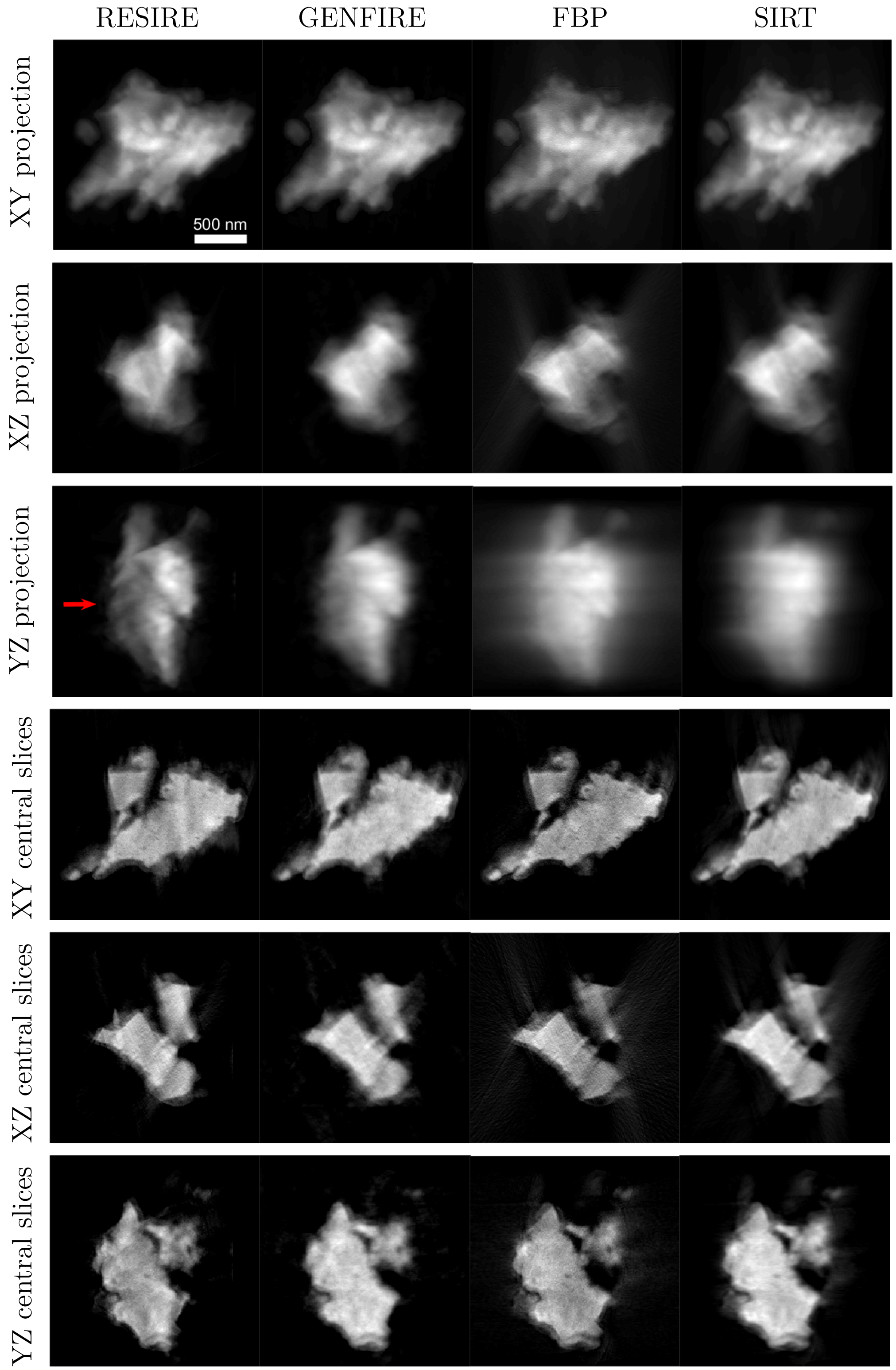


Figure 3.3: Reconstruction of a Seriatopora aculeata coral skeleton by RESIRE, GENFIRE, FBP, and SIRT: projections and 30.3-nm-thick (3-pixel-thick) central slices in the XY, XZ and YZ planes respectively.

range from -60 to 76 degrees in 2 degrees of increment. The projections were pre-processed with alignment, constant background subtraction and normalization. Details of this experiment and pre-processing can be found elsewhere [87].

After the pre-processing, Tomography reconstructions of the S. aculeate sample were performed by RESIRE, GENFIRE, FBP, and SIRT algorithms. Figure 3.3 shows the reconstructed projections and 30.3-nm-thick (3-pixel-thick) central slices in XY, XZ and YZ planes. RESIRE is shown to completely outperform the other algorithms as the features obtained by RESIRE are finer, clearer and more visible in all directions. Especially in the missing wedge direction, the YZ projections reconstructed by FBP and SIRT are corrupted by blurring and smearing. GENFIRE achieves a better reconstruction where significant artifacts are reduced; however, the features are still blurred. RESIRE, in contrast, obtains the finest solution among these four algorithms. The features indicated by the red arrow in the YZ projection can be only obtained clearly and visibly with RESIRE.

3.3.4 Reconstruction of a nickel platinum nanoparticles

The NiPt nanoparticles were synthesized using procedures published elsewhere [66] and deposited on 5-nm-thick silicon nitride membranes. The tomography tilt series was acquired from a NiPt nanoparticle using the TEAM I microscope operated at 300 kV in annular dark-field scanning transmission electron microscopy (ADF-STEM) mode. 56 projections were collected with a tilt range from -72 to 66 degrees in approximated 2.5 degrees of increment. Similar to other experimental data, pre-processing was performed to polish the projections, including denoising, constant background subtraction, alignment, and normalization. The projections then went through angle refinement [112] where reconstruction and angles are simultaneously updated. Along the refinement, R_F improves, and so does the reconstruction. At the final round when the optimal angle set has been found, we perform the final reconstruction on the NiPt sample.

Figure 3.4 shows the reconstruction of 1.40-Å-thick (3-pixel-thick) central slices of NiPt nanoparticle sample in the YZ plane where y is the rotation axis and z is the missing wedge

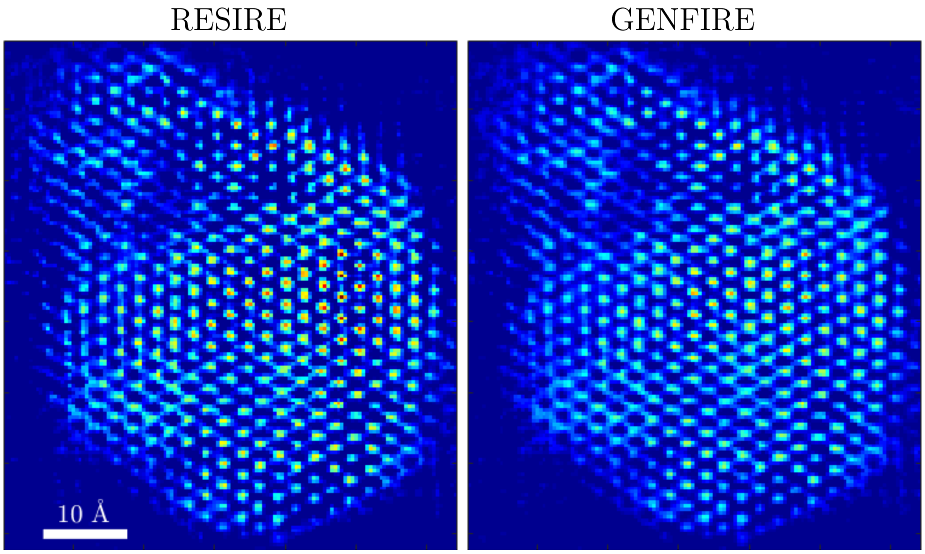


Figure 3.4: Reconstruction of 1.40-Å-thick (3-pixel-thick) central slices of nickel platinum alloy in the YZ plane. In this missing wedge direction z, the atom reconstructed by GENFIRE suffers elongation and blurring because GENFIRE favors over-smoothed images. RESIRE, in contrast, produces sharper and non-blurred images. The atoms are reconstructed more separably, distinctly, and isotropically with RESIRE. The contrast also exhibits better than the one with GENFIRE.

direction. Since GENFIRE refines images based on Fourier constraint, it generates over-smoothed solutions. The particles suffer blurring and cannot be distinguished from each other in some regions. RESIRE, on the other hand, produces sharper and more distinct particles. The Pt particles obtained by RESIRE have higher intensity than the ones by GENFIRE. This result is very helpful for atom-type classification [145] since the atoms' intensities of the two types are more separable from each other.

3.4 Discussion

In this article, we present RESIRE, a Tomography iterative reconstruction algorithm for high-resolution 3D images. The method is designed to work with multiple tilt axes, the general case of Tomography. Through numerical simulation and experimental data of material science and biological specimens, RESIRE has been shown to produce superior reconstructions to several algorithms. RESIRE can work well with limited projections and missing wedge while other methods fail and suffer artifacts such as elongation, blurring, and shearing.

Basically, RESIRE uses gradient descent to solve the least square problem $||\Pi O - b||^2$. Instead of pre-computing and storing the (forward) projection operator Π as a matrix (such as SIRT does), RESIRE uses bi-linear interpolation and FST. FST helps to improve the forward projection while bi-linear interpolation saves enormous memory in the multiple-tilt-axis case where computing and storing operator matrix is extremely expensive. These two techniques are the major points that constitute our "real space iterative reconstruction engine" RESIRE.

From a mathematical point of view, ART, SIRT and SART approximate the forward projection (line integral) as a discrete sum. However, they do not specify how to choose grid points for this finite sum. RESIRE, in contrast, does as Eqn. 3.4 points out the linear transformation; therefore, our method significantly improves the performance compared to SIRT. Moreover, RESIRE can exploit the refining technique (dividing a pixel into 4 sub-pixels) as Radon transform [17, 85] does to further improve the "forward operator."

For global-transform-based methods such as GENFIRE, it converts measurements (projections) on real space to constraint on the reciprocal space via FST. This technique causes numerical error due to the gridding process (interpolation from non-grid points). Furthermore, favoring smoothed solutions, reciprocal constraint prevents GENFIRE from obtaining high resolution but blurred images instead. On the other hand, using a hybrid model (compute forward projection via FST and back projection via bi-linear interpolation), RESIRE can solve the downside issues of these methods. This explains why RESIRE has the ability to work with high noise and produce more faithful reconstruction with higher resolution and finer and sharper features.

Furthermore, the extension of RESIRE to the thinfilm case allows us to work with non-isolated or extending samples which are very popular in material science. Methods using global transform, such as FBP and GENFIRE, fail in this case because they interpret objects that appear in high tilt angles as noise. Through RESIRE, we can simply correct this case.

Tomography is an ill-posed inverse problem and might need regularizer to compensate for the under-determination [32]. The optimal solution can be achieved after some iterations based on a suitable discrepancy principle [82]. This principle tells us a criteria where to stop the iterations according to prior noise information. The refinement will take place until the R-factor hits that certain value. Further iteration will improve the R-factor but not the result. Instead, the solution will move away from the optimal one.

Thanks to the simplicity, efficiency, and flexibility, RESIRE can have intensive applications in 3D image reconstructions, such as cryo-ET and vector tomography [36, 37, 38] where reconstructing the 3D magnetic field of magnetic materials is a big interest. Looking forward, we expect RESIRE can be applied to a plethora of imaging modalities across different disciplines.

Together with the developed algorithm GPS in chapter 2, we have provided two advanced reconstruction techniques in phase retrieval and 3D tomography. Both algorithms are built on mathematics and optimization with detailed derivations and explanations. Those two algorithms can find cooperation in further applications, such as ptychographic-tomography which will be addressed in conclusion.

Moreover, those techniques do not only serve image processing but also find their extension in other disciplines, such as machine learning. In the next two chapters, we will explore how Hopf-Lax formula and Hamilton-Jacobi partial differential equations are applied into K-means clustering and training algorithms in machine learning.

CHAPTER 4

Stochastic Backward Euler: An Implicit Gradient Descent Algorithm for K-Means Clustering

In this paper, we propose an implicit gradient descent algorithm for the classic k-means problem. The implicit gradient step or backward Euler is solved via stochastic fixed-point iteration, in which we randomly sample a mini-batch gradient in every iteration. It is the average of the fixed-point trajectory that is carried over to the next gradient step. We draw connections between the proposed stochastic backward Euler and the recent entropy stochastic gradient descent (Entropy-SGD) for improving the training of deep neural networks. Numerical experiments on various synthetic and real datasets show that the proposed algorithm practically finds the global minimum (or its neighborhood) with high probability, when given the correct number of clusters. Stochastic backward Euler provides better clustering results than the state-of-the-art k-means algorithms in the sense that it goes deeper in minimization and is much more robust to the choice of initial centroids.

4.1 Introduction

K-means is originally a method of vector quantization in signal processing, which now has become popular for clustering analysis in data mining. In the seminal paper [86], Lloyd proposed a two-step alternating algorithm for minimizing quantization error that quickly converges to a local minimum. Lloyd's algorithm is also known as an instance of the more general Expectation-Maximization (EM) algorithm applied to Gaussian mixtures. In [14], Bottou and Bengio cast Lloyd's algorithm as Newton's method, which explains its fast convergence.

Aiming to speed up Lloyd's algorithm, Elkan [43] proposed to keep track of the distances between the computed centroids and data points, and then cleverly leverage the triangle inequality to eliminate unnecessary computations of the distances. Similar techniques can be found in [35]. It is worth noting that these algorithms do not improve the clustering quality of Lloyd's algorithm, but only achieve acceleration. [7] introduced a smart initialization scheme such that the initial centroids are well-separated, which gives more robust clustering than random initialization.

Mini-batch [126, 135] was later introduced to adapt k-means for large scale data with high dimensions. The centroids are updated using a randomly selected mini-batch rather than all of the data. Mini-batch (stochastic) k-means has a flavor of stochastic gradient descent whose benefits are twofold. First, it dramatically reduces the per-iteration cost for updating the centroids and thus is able to handle big data efficiently. Second, similar to its successful application to deep learning [80], mini-batch gradient introduces noise in minimization and may help to bypass some bad local minima. Furthermore, the aforementioned Elkan's technique can be combined with mini-batch k-means for further acceleration [105].

In this paper, we propose a backward Euler based algorithm for k-means clustering. Fixed-point iteration is performed to solve the implicit gradient step. As is done for stochastic mini-batch k-means, we compute the gradient only using a mini-batch of samples instead of the whole data, which enables us to handle massive data. Unlike the standard fixed-point iteration, the resulting stochastic fixed-point iteration outputs an average over its trajectory. Extensive experiments show that, with proper choice of step size for backward Euler, the proposed algorithm empirically locates the neighborhood of global minimum with high probability.

4.2 Stochastic backward Euler

The celebrated proximal point algorithm (PPA) [121] for minimizing some function $f(x)$ is:

$$x^{k+1} = \mathrm{prox}_{\gamma f}(x^k) := \arg\min_x f(x) + \frac{1}{2\gamma}\|x - x^k\|^2. \tag{4.1}$$

PPA has the advantage of being monotonically decreasing, which is guaranteed for any step size $\gamma > 0$. Indeed, by the definition of x^{k+1} in (4.1), we have

$$f(x^{k+1}) \leq f(x^k) - \frac{1}{2\gamma}\|x^{k+1} - x^k\|^2.$$

When $\gamma \in [c, \frac{1}{L(\nabla f)})$ for any $c > 0$ with $L(\nabla f)$ being the Lipschitz constant of ∇f, the (subsequential) convergence to a stationary point is established in [73]. If f is differentiable at x^{k+1}, it is easy to check that the following optimality condition to (4.1) holds

$$\nabla f(x^{k+1}) + \frac{1}{\gamma}(x^{k+1} - x^k) = 0.$$

By rearranging the terms, we arrive at implicit gradient descent or the so-called backward Euler:

$$x^{k+1} = x^k - \gamma \nabla f(x^{k+1}). \tag{4.2}$$

Fixed point iteration is a viable option for solving (5.4) if ∇f has Lipschitz constant $L(\nabla f)$ and $\gamma < \frac{1}{L(\nabla f)}$.

Proposition 1. *If $\gamma < \frac{1}{L(\nabla f)}$, then we have*

(a) $f(x) + \frac{1}{2\gamma}\|x - x^k\|^2$ is strongly convex, and the proximal problem (4.1) has a unique solution y^*.

(b) The fixed point iteration

$$y^{l+1} = x^k - \gamma \nabla f(y^l) \tag{4.3}$$

generates a sequence $\{y^l\}$ converging to y^* at least linearly.

This result is standard, and we omit the proof here.

Let us consider k-means clustering for a set of data points $\{p_i\}_{i=1}^N$ in $\mathbb{R}^d$ with K centroids $\{x_j\}_{j=1}^K$. We assume each cluster contains the same number of points. Denoting $x = [x_1, \ldots, x_K]^\top \in \mathbb{R}^{Kd}$, we seek to minimize

$$\min_{x \in \mathbb{R}^{Kd}} \ \phi(x) := \frac{1}{2N} \sum_{i=1}^{N} \min_{1 \leq j \leq K} \|x_j - p_i\|^2. \tag{4.4}$$

Note that ϕ is non-differentiable at x if there exist p_i and $j_1 \neq j_2$ such that

$$j_1, j_2 \in \arg \min_{1 \leq j \leq K} \|x_j - p_i\|^2.$$

This means that there is a data point p_i which has two or more distinct nearest centroids x_{j_1} and x_{j_2}. The same situation may happen in the assignment step of Lloyd's algorithm. In this case, we simply assign p_i to one of the nearest centroids. With that said, ϕ is basically piecewise differentiable. By abuse of notation, we can define the 'gradient' of ϕ at any point x by

$$\nabla \phi(x) = \frac{1}{N}[\sum_{i \in \mathbb{C}_1}(x_1 - p_i), \ldots, \sum_{i \in \mathbb{C}_K}(x_K - p_i)]^\top, \tag{4.5}$$

where $\mathbb{C}_j$ denotes the index set of the points that are assigned to the centroid x_j. Similarly, we can compute the 'Hessian' of ϕ as was done in [14]:

$$\nabla^2 \phi(x) = \frac{1}{N}\mathrm{Diag}\big(|\mathbb{C}_1|\mathbf{1}_{(|\mathbb{C}_1|)}, \ldots, |\mathbb{C}_K|\mathbf{1}_{(|\mathbb{C}_K|)}\big),$$

where $\mathbf{1}_{(n)}$ is an n-D vector of all ones. When the number of points N is large and x_j's are distinct from each other, the jumps at discontinuities of $\nabla\phi$ caused by the ambiguity in the assignment of data points to centroids are very small. Thus with (4.5), one can roughly consider $\nabla\phi$ to be Lipschitz continuous with the Lipschitz constant $L(\nabla\phi) \approx \frac{1}{K}$ by ignoring these tiny jumps. In what follows, we analyze how the fixed point iteration (4.3) works on the piecewise differentiable ϕ with discontinuous $\nabla\phi$.

Definition 1. *g is piecewise Lipschitz continuous on Ω with Lipschitz constant L if Ω can be partitioned into sub-domains Ω_I and g is Lipschitz continuous on each sub-domain Ω_I, i.e., for each Ω_I we have*

$$\|g(x) - g(y)\| \leq L\|x - y\| \quad \forall x, y \in \Omega_I$$

According to the definition, we can see that $\nabla\phi$ is approximately piecewise $\frac{1}{K}$-Lipschitz continuous. But in the extreme case where just one point is assigned to each of the first $K-1$ clusters and $N-K+1$ points to the last cluster, $\nabla\phi$ only has piecewise Lipschitz constant of $\frac{N-K+1}{N}$. Our main result proves the convergence of fixed point iteration on k-means problem.

Theorem 1. *Suppose $\nabla\phi$ is piecewise L-Lipschitz. If $\gamma < 1/L$, then the fixed point iteration for minimizing $h(x) := \phi(x) + \frac{1}{2\gamma}\|x - x^k\|^2$ given by*

$$y^{l+1} = x^k - \gamma\nabla\phi(y^l)$$

with initialization $y^0 = x^k$ satisfies

(a) $h(y^{l+1}) \leq h(y^l) - (\frac{1}{2\gamma} - \frac{L}{2})\|y^{l+1} - y^l\|^2$ and $\|y^{l+1} - y^l\| \to 0$ as $l \to \infty$.

(b) $\{y^l\}$ is bounded. Moreover, if any limit point y^* of a convergent subsequence of $\{y^l\}$ lies in the interior of some sub-domain, then the whole sequence $\{y^l\}$ converges to y^* which is a fixed point obeying

$$y^* = x^k - \gamma\nabla\phi(y^*).$$

Proof. (a) We know that ϕ is piecewise quadratic. Suppose $y^l \in \Omega_I$ (note that y^l could be on the boundary), then ϕ has a uniform expression restricted on Ω_I which is a quadratic function, denoted by ϕ_{Ω_I}. We can extend the domain of ϕ_{Ω_I} from Ω_I to the whole $\mathbb{R}^{Kd}$, and we denote the extended function still by ϕ_{Ω_I}. Since ϕ_{Ω_I} is quadratic, $\nabla\phi_{\Omega_I}$ is L-Lipschitz continuous on $\mathbb{R}^{Kd}$. Then we have the following well-known inequality

$$\begin{aligned}\phi_{\Omega_I}(y^{l+1}) &\leq \phi_{\Omega_I}(y^l) + \langle\nabla\phi_{\Omega_I}(y^l), y^{l+1} - y^l\rangle + \frac{L}{2}\|y^{l+1} - y^l\|^2 \\ &= \phi(y^l) + \langle\nabla\phi(y^l), y^{l+1} - y^l\rangle + \frac{L}{2}\|y^{l+1} - y^l\|^2.\end{aligned}$$

Using the above inequality and the definition of ϕ, we have

$$\begin{aligned}h(y^{l+1}) &= \phi(y^{l+1}) + \frac{1}{2\gamma}\|y^{l+1} - x^k\|^2 \leq \phi_{\Omega_I}(y^{l+1}) + \frac{1}{2\gamma}\|y^{l+1} - x^k\|^2 \\ &\leq \phi(y^l) + \langle\nabla\phi(y^l), y^{l+1} - y^l\rangle + \frac{L}{2}\|y^{l+1} - y^l\|^2 + \frac{1}{2\gamma}\|y^{l+1} - x^k\|^2 \\ &= \phi(y^l) + \langle\nabla\phi(y^l), y^{l+1} - y^l\rangle + (\frac{L}{2} - \frac{1}{2\gamma})\|y^{l+1} - y^l\|^2 \\ &\qquad\qquad + \frac{1}{2\gamma}\|y^l - x^k\|^2 + \frac{1}{\gamma}\langle y^{l+1} - x^k, y^{l+1} - y^l\rangle \\ &= h(y^l) - (\frac{1}{2\gamma} - \frac{L}{2})\|y^{l+1} - y^l\|^2 + \langle\frac{1}{\gamma}(y^{l+1} - x^k) + \nabla\phi(y^l), y^{l+1} - y^l\rangle \\ &= h(y^l) - (\frac{1}{2\gamma} - \frac{L}{2})\|y^{l+1} - y^l\|^2.\end{aligned}$$

In the second equality above, we used the identity

$$\frac{1}{2}\|a-b\|^2 + \langle a, b\rangle = \frac{1}{2}\|a\|^2 + \frac{1}{2}\|b\|^2$$

with $a = y^{l+1} - y^l$ and $b = y^{l+1} - x^k$. Since $\gamma < \frac{1}{L}$, $\{h(y^l)\}$ is monotonically decreasing. Moreover, since h is bounded from below by 0, $\{h(y^l)\}$ converges and thus $\|y^{l+1} - y^l\| \to 0$ as $l \to \infty$.

(b) Since $h(y) \to \infty$ as $y \to \infty$, combining with the fact that $h(y^l) \leq h(y^{l+1})$, we have $\{y^l\} \subseteq \{y \in \mathbb{R}^{Kd} : h(y) \leq h(y^0)\}$ is bounded. Consider a convergent subsequence $\{y^{l_m}\}$ whose limit y^* lies in the interior of some sub-domain. Then for sufficiently large l_m, $\{y^{l_m}\}$ will always remain in the same sub-domain in which y^* lies and thus $\lim_{l_m \to \infty} \nabla\phi(y^{l_m}) = \nabla\phi(y^*)$. Since by (a), $\|y^{l+1} - y^l\| \to 0$, we have $\|\nabla\phi(y^{l+1}) - \nabla\phi(y^l)\| = \frac{1}{\gamma}\|y^l - y^{l-1}\| \to 0$ as $l \to \infty$. Therefore,

$$\begin{aligned} 0 &= \lim_{l_m \to \infty} y^{l_m} - x^k + \gamma\nabla\phi(y^{l_m - 1}) = \lim_{l_m \to \infty} y^{l_m} - x^k + \gamma\nabla\phi(y^{l_m}) \\ &= y^* - x^k + \gamma\phi\nabla(y^*), \end{aligned}$$

which implies y^* is a fixed point. Furthermore, by the piecewise Lipschitz condition,

$$\|y^{l_m+1} - y^*\| = \gamma\|\nabla\phi(y^{l_m}) - \nabla\phi(y^*)\| \leq L\gamma\|y^{l_m} - y^*\|.$$

Since $L\gamma < 1$, when l_m is sufficiently large, y^{l_m+1} is also in the same sub-domain containing y^*. By repeatedly applying the above inequality for $l > l_m$, we conclude that $\{y^l\}$ converges to y^*.

□

4.2.1 Algorithm description

Instead of using the full gradient $\nabla\phi$ in fixed-point iteration, we adopt a randomly sampled mini-batch gradient

$$\nabla_l\phi = \frac{1}{M}[\sum_{i \in \mathbb{C}_1^l}(x_1 - p_i), \ldots, \sum_{i \in \mathbb{C}_K^l}(x_K - p_i)]^\top$$

at the l-th inner iteration. Here, $\mathbb{C}_j^l$ denotes the index set of the points in the l-th mini-batch associated with the centroid x_j obeying $\sum_{j=1}^K |\mathbb{C}_j^l| = M$. The fixed-point iteration outputs a forward looking average over its trajectory. Intuitively averaging greatly stabilizes the noisy mini-batch gradients and thus smooths the descent. We summarize the proposed algorithm in Algorithm 4. Another key ingredient of our algorithm is an aggressive initial step size $\gamma^0 \approx \frac{1}{L(\nabla\phi)} \approx K$, which helps pass bad local minimum at the early stage. Unlike in deterministic backward Euler, diminishing step size is needed to ensure convergence. But γ should decay slowly because large step size is good for a global search.

Algorithm 4 Stochastic backward Euler for k-means.

```
Input: number of clusters $K$, step size $\gamma^0 \approx K$, mini-batch size $M$, averaging parameter $\alpha > 0$, step size decay parameter $\beta \lessapprox 1$.
Initialize: centroid $x^0$.
  for $k = 1, \dots, \text{omaxit}$ do
    $y^0 = x^{k-1}$
    $x^k = y^0$
    for $l = 1, \dots, \text{imaxit}$ do
      Randomly sample a mini-batch gradient $\nabla_l \phi$.
      $y^l = x^{k-1} - \gamma \nabla_l \phi(y^{l-1})$
      $x^k = \alpha x^k + (1-\alpha) y^l$
    end for
    $\gamma^k = \beta \gamma^{k-1}$
  end for
Output: $x^{\text{omaxit}}$
```

4.2.2 Related work

Chaudhari et al. [26] recently proposed the entropy stochastic gradient descent (Entropy-SGD) algorithm to tackle the training of deep neural networks. Relaxation techniques arising in statistical physics were used to change the energy landscape of the original non-convex objective function $f(x)$ yet with the minimizers being preserved, which allows easier minimization to obtain a 'good' minimizer with a better geometry. More precisely, they suggest to replace $f(x)$ with a modified objective function $f_\gamma(x)$ called local entropy [9] as follows

$$f_\gamma(x) := -\frac{1}{\beta}\log\big(G_{\beta^{-1}\gamma} * \exp(-\beta f(x))\big),$$

where $G_\gamma(x) = (2\pi\gamma)^{-d/2}\exp\big(-\frac{|x|^2}{2\gamma}\big)$ is the heat kernel. The connection between Entropy-SGD and nonlinear partial differential equations (PDEs) was later established in [27]. The local entropy function f_γ turns out to be the solution to the following viscous Hamilton-Jacobi (HJ) PDE at $t=\gamma$

$$u_t = -\frac{1}{2}|\nabla u|^2 + \frac{\beta^{-1}}{2}\Delta u \tag{4.6}$$

with the initial condition $u(x,0) = f(x)$. In the limit $\beta^{-1} \to 0$, (4.6) reduces to the non-viscous HJ equation

$$u_t = -\frac{1}{2}|\nabla u|^2,$$

whose solution is closely related to the proximal operator $\text{prox}_{tf}(x)$:

$$u(x,t) = \inf_y \big\{f(y) + \frac{1}{2t}\|y-x\|^2\big\} = \frac{1}{t}\text{prox}_{tf}(x).$$

The gradient descent dynamics for f_γ is obtained by taking the limit of the following system of stochastic differential equation as the homogenization parameter $\varepsilon \to 0$:

$$\begin{aligned} dx(s) &= -\gamma^{-1}(x-y)ds \\ dy(s) &= -\frac{1}{\varepsilon}\big[\nabla f(y) + \frac{y-x}{\gamma}\big]ds + \frac{\beta^{-1/2}}{\sqrt{\varepsilon}}dW(s) \end{aligned} \tag{4.7}$$

where $W(s)$ is the standard Wiener process. Specifically, we have

$$-\nabla f_\gamma(x) = -\gamma^{-1}(x - \langle y\rangle)$$

with $\langle y\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T y(s)ds$ and $y(s)$ being the solution of (4.7) for fixed x. This gives rise to the implementation of Entropy-SGD [27]. We remark that stochastic backward Euler is equivalent to Entropy-SGD with the step size of the gradient flow being equal to γ.

4.3 Experimental results

We show by several experiments that the proposed stochastic backward Euler (SBE) gives superior clustering results compared with the state-of-the-art algorithms for k-means. SBE scales well for large problems. In practice, only a small number of fixed-point iterations are needed in the inner loop, and this seems not to depend on the size of the problem. Specifically, we chose the parameters `imaxit` $= 5$ or 10 and the averaging parameter $\alpha = 0.75$ in all experiments. Moreover, we always set $\gamma^0 = K$.

4.3.1 2-D synthetic Gaussian data

We generated 4000 synthetic data points in 2-D plane by multivariate normal distributions with 1000 points in each cluster. The means and covariance matrices used for Gaussian distributions are as follows:

$$\mu_1 = \begin{bmatrix} -5 \\ -3 \end{bmatrix}, \; \mu_2 = \begin{bmatrix} 5 \\ -3 \end{bmatrix}, \; \mu_3 = \begin{bmatrix} 0.0 \\ 5.0 \end{bmatrix}, \; \mu_4 = \begin{bmatrix} 2.5 \\ 4.0 \end{bmatrix};$$

$$\Sigma_1 = \begin{bmatrix} 0.8 & 0.1 \\ 0.1 & 0.8 \end{bmatrix}, \; \Sigma_2 = \begin{bmatrix} 1.2 & 0.6 \\ 0.6 & 0.7 \end{bmatrix}, \; \Sigma_3 = \begin{bmatrix} 0.5 & 0.05 \\ 0.05 & 1.6 \end{bmatrix}, \; \Sigma_4 = \begin{bmatrix} 1.5 & 0.05 \\ 0.05 & 0.6 \end{bmatrix}.$$

For the initial centroids given below, Lloyd's algorithm (or EM) got stuck at a local minimum; see the left plot of Fig. 4.1.

$$x_1 = \begin{bmatrix} -5.5989 \\ -2.7090 \end{bmatrix}, \; x_2 = \begin{bmatrix} -4.4572 \\ -4.0614 \end{bmatrix}, \; x_3 = \begin{bmatrix} -0.1082 \\ 5.2889 \end{bmatrix}, \; x_4 = \begin{bmatrix} 2.3485 \\ 3.5286 \end{bmatrix}.$$

Starting from where EM got stuck, we can see that SBE managed to jump over the trap of local minimum and arrived at the global minimum; see the right plot of Fig. 4.1.

4.3.2 Iris dataset

The Iris dataset, which contains 150 4-D data samples from 3 clusters, was used for comparisons between SBE and the EM algorithm. 100 runs were realized with the initial centroids randomly selected from the data samples. For the parameters, we chose mini-batch size

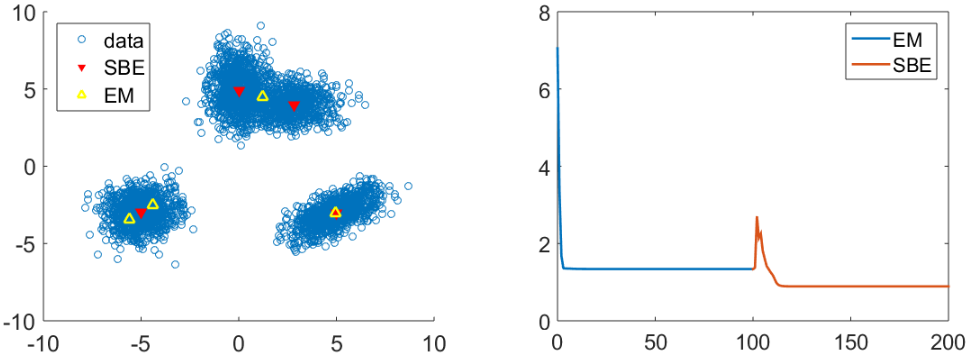


Figure 4.1: Synthetic Gaussian data with 4 centroids. Left: Computed centroids by EM and SBE corresponding to the objective values 1.34 and 0.89, respectively. Right: Plot of objective value v.s. number of iteration. EM converged quickly but got trapped at a local minimum. SBE passed this local minimum and reached the global minimum by jumping over a hill.

$M = 60$, initial step size, `imaxit`$= 40$, `omaxit`$= 10$, and decay parameter $\beta = \frac{1}{1.01}$. The histograms in Fig. 4.2 record the frequency of objective values given by the two algorithms. Clearly there was 29% chance that EM got stuck at a local minimum whose value is about 0.48, whereas ESGD managed to locate the near global minimum region valued at around 0.264 *every time*.

4.3.3 Gaussian data with MNIST centroids

We selected 8 hand-written digit images of dimension $28 \times 28 = 784$ from MNIST dataset shown in Fig. 4.3, and then generated 60,000 images from these 8 centroids by adding Gaussian noise. We compare SBE with both EM and mini-batch EM (mb-EM) [126, 135] on 100 independent realizations with random initial guess. For each method, we recorded the minimum, maximum, mean and variance of the 100 objective values by the computed centroids.

We first compare SBE and EM with the true number of clusters $K = 8$. For SBE, mini-batch size $M = 1000$, maximum number of iterations for backward Euler `omaxit`$=150$, maximum fixed-point iterations `imaxit`$= 10$ for SBE. We set the maximum number of

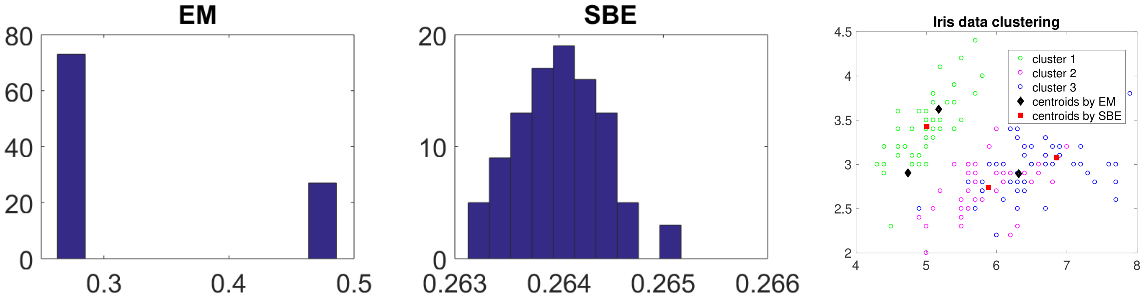


Figure 4.2: The Iris dataset with 3 clusters. Left: histogram of objective values obtained by EM in 100 trials. Middle: histogram of objective values obtained by SBE (proposed) in 100 trials. Right: computed centroids by EM (black) and SBE (red), corresponding to the objective values 0.48 and 0.264, respectively.

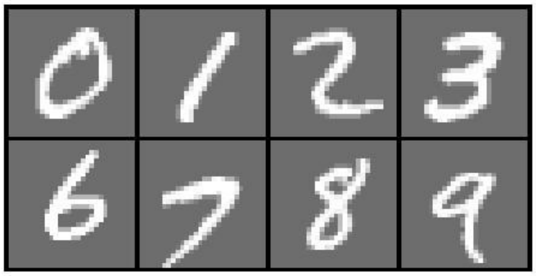


Figure 4.3: 8 selected images from MNIST dataset. 60,000 sample images are generated from these 8 images by adding Gaussian noise.

K	Method	Batch size	Max iter	Min	Max	Mean	Variance
8	EM	7500	50	15.6800	27.2828	20.0203	6.0030
	SBE	1000	(150,10)	15.6808	15.6808	15.6808	1.49×10^{-10}
6	mb-EM	500	100	20.44	23.4721	21.8393	0.67
	SBE	500	(100,5)	20.2989	21.2047	20.4939	0.0439
8	mb-EM	500	100	15.9193	18.5820	16.4009	0.7646
	SBE	500	(100,5)	15.6816	15.6821	15.6820	1.18×10^{-9}
10	mb-EM	500	100	15.9148	18.1848	16.1727	0.4332
	SBE	500	(100,5)	15.6823	15.6825	15.6824	1.5×10^{-9}

Table 4.1: Gaussian data generated from MNIST centroids by adding noise. Ground truth $K = 8$. Clustering results for 100 independent trails with random initialization.

iterations for EM to be 50, which was sufficient for its convergence. The results are listed in the first two rows of Table 4.3.3. We observed that the global minimum was around 15.68 and that SBE always found the global minimum up to a tiny error due to the noise from mini-batch.

In the comparison between SBE and mb-EM, we reduced mini-batch size to $M = 500$, `omaxit`$= 100$, `imaxit`$= 5$ and tested for $K = 6, 8, 10$. Table 4.3.3 shows that with the same mini-batch size, SBE outperforms mb-EM in all three cases, in terms of both mean and variance of the objective values.

4.3.4 Raw MNIST data

In this example, We used the 60,000 images from the MNIST training set for clustering test, with 6000 samples for each digit (cluster) from 0 to 9. The comparison results are shown in Table 4.2. We conclude that SBE consistently performs better than EM and mb-EM. The histograms of objective value by the three algorithms in the case $K = 10$ are plotted in Fig. 4.4.

K	Method	Batch size	Max iter	Min	Max	Mean	Variance
10	EM	6000	50	19.6069	19.8195	19.6725	0.0028
	SBE	1000	(150,10)	19.6087	19.7279	19.6201	5.7×10^{-4}
8	mb-EM	500	100	20.4948	20.7126	20.5958	0.0018
	SBE	500	(100,5)	20.2723	20.4104	20.3090	0.0014
10	mb-EM	500	100	19.9029	20.2347	20.0146	0.0041
	SBE	500	(100,5)	19.6103	19.7293	19.6354	0.0011
12	mb-EM	500	100	19.3978	19.7147	19.5136	0.0042
	SBE	500	(100,5)	19.0492	19.1582	19.0972	6.2×10^{-4}

Table 4.2: Raw MNIST training data. The ground truth number of clusters is $K = 10$. Clustering results for 100 independent trails with random initialization.

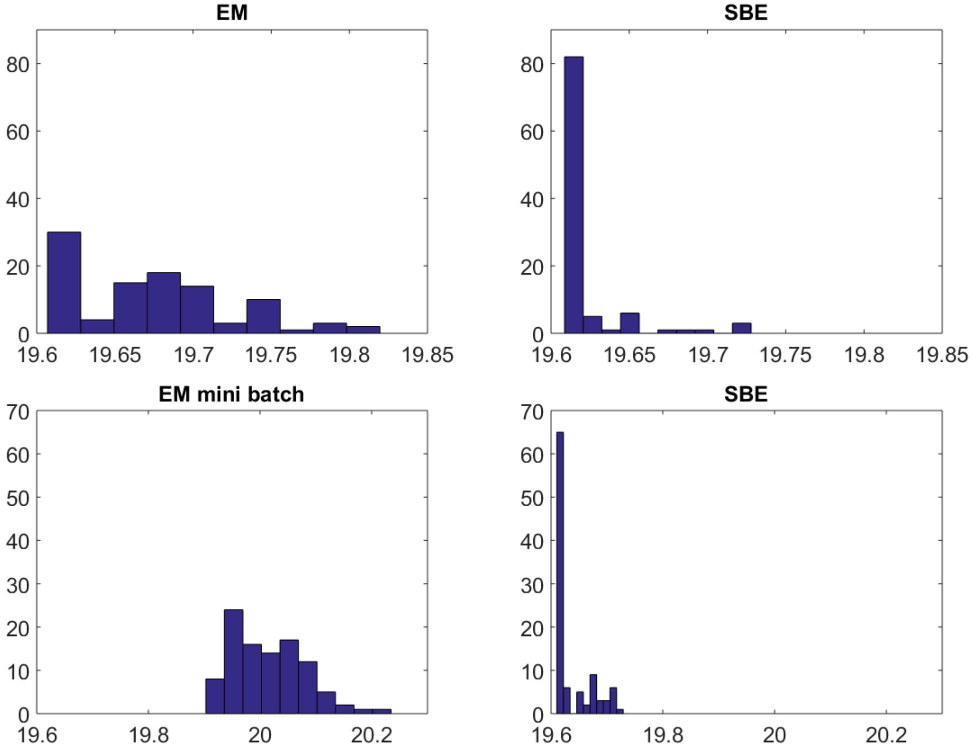


Figure 4.4: Histograms of objective value for MNIST training data with ground truth number of clusters $K = 10$. Top left: EM. Top right: SBE, mini-batch size of 1000. Bottom left: mn-EM, mini-batch size of 500. Bottom right: SBE, mini-batch size of 500.

K	Method	Batch size	Max iter	Min	Max	Mean	Variance
10	EM	6000	50	1.6238	3.0156	2.1406	0.0977
	SBE	1000	(150,10)	1.6238	1.6239	1.6239	2.7×10^{-10}
8	mb EM	500	100	2.3428	3.5972	2.7157	0.0666
	SBE	500	(100,5)	2.2833	2.4311	2.3274	0.0015
10	mb EM	500	100	1.6504	2.6676	2.1391	0.0712
	SBE	500	(100,5)	1.6239	1.6242	1.6240	1.37×10^{-9}
12	mb EM	500	100	1.5815	2.6189	1.7853	0.0661
	SBE	500	(100,5)	1.5326	1.5891	1.5622	9.8×10^{-5}

Table 4.3: MNIST features generated by LeNet-5 network. The ground truth number of clusters is $K = 10$. Clustering results for 100 independent trails with random initialization.

4.3.5 MNSIT features

We extracted the feature vectors of MNIST training data prior to the last layer of LeNet-5 [79]. The feature vectors have dimension 64 and lie in a better manifold compared with the raw data. The results are shown in Table 4.3 and Fig. 4.5 and 4.6.

4.4 Discussions

At the k-th iteration, SBE solves $x = x^k - \gamma\nabla\phi(x)$. Since $\nabla\phi$ is technically only piecewise Lipschitz continuous, the backward Euler may have multiple solutions, and we will obtain these solutions with certain probabilities. For example, in this 1D example, we get two solutions $x_{BE,1}$ in the leftmost valley and $x_{BE,2}$ in the second from the left. x^{k+1} by SBE is the extrapolation of these solutions in expectation. The averaging step helps reduce variance. If $x_{BE,2}$ is far to the right, then x^{k+1} is dragged away from the leftmost valley to the second valley, i.e. x^{k+1} passes the local minimum. During the above process, the objective value $\phi(x^{k+1})$ may increase, which explains the jump of objective value showed in the right plot of Fig. 4.1. In some cases, however, the jump simply does not appear. The reason can be

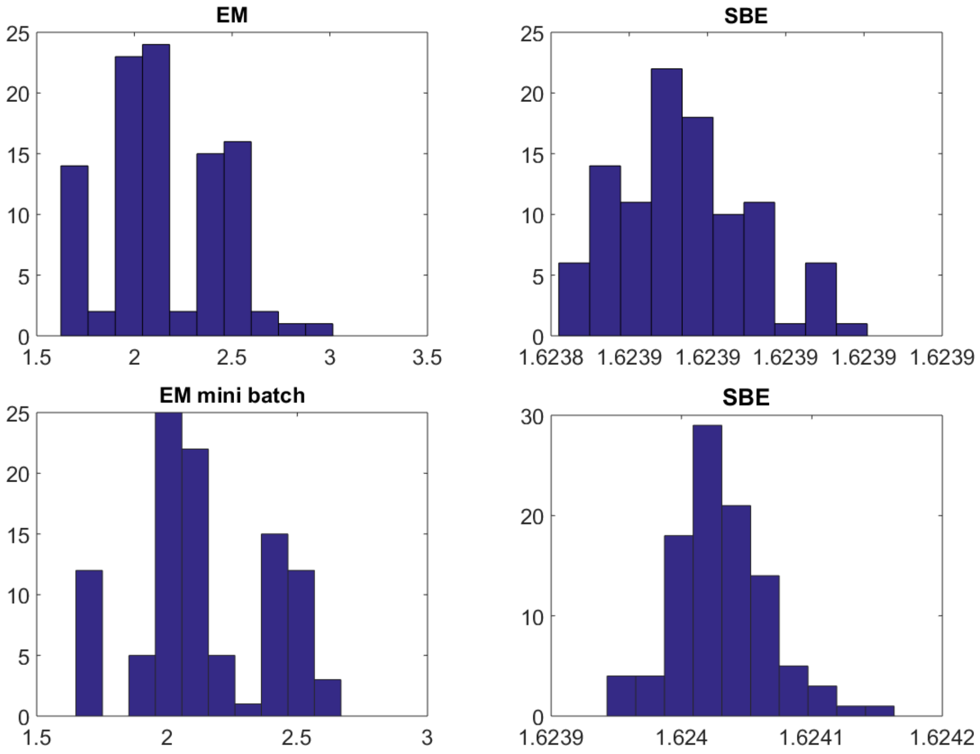


Figure 4.5: Histograms of objective value for MNIST feature data with ground truth number of clusters K=10. Top left: EM. Top right: SBE, mini-batch size of 1000. Bottom left: mn-EM, mini-batch size of 500. Bottom right: SBE, mini-batch size of 500.

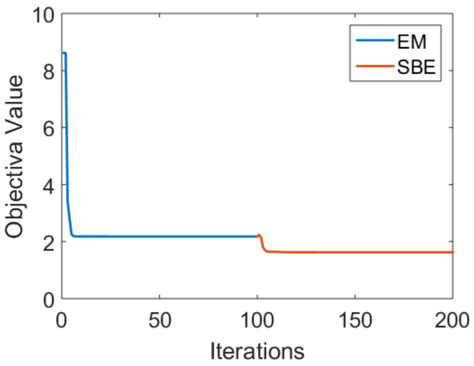


Figure 4.6: Objective Value for MNIST training features with 60, 000 data points of size 64. The ground truth number of clusters is $K = 10$. EM got trapped at local minimum around 2.178. Initializing SBE with this local minimizer, global minimum around 1.623 was found.

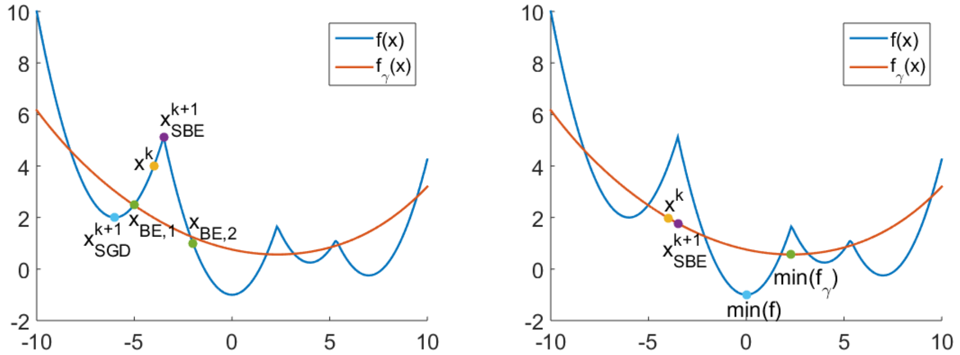


Figure 4.7: Comparison the updates between SGD and SBE. Left: At the $(k+1)$-th update, SGD gives $\mathbb{E}_{SGD}[x^{k+1}] = x^k - \gamma\nabla f(x^k)$ while SBE solves the Backward Euler for two solution $x_{BE,1}$ and $x_{BE,2}$, and updates x^{k+1}_{SBE} as the interpolation between these solutions. Right: SBE converges to the global minimum of local entropy f_γ: the average of solutions to the Backward Euler. Therefore, we let $\gamma \to 0$ in order that SBE converges to the global minimum of f.

that the second valley is wider and deeper, and thus $x_{BE,2}$ is further to the right. The x^{k+1} could be pulled to the second valley with the smaller objective value with some probability.

To finish this chapter, we conclude that the combination of backward Euler and stochasticity provides a "better" direction for gradient-based method. The joining is numerical implementation to Hopf-Lax formula that, the implicit-form solution in time to viscous HJ-PDE , yields to a regularizer on f. This smoothing technique helps the algorithm to avoid bad local minima (sharp & narrow) and seek for better ones (wide). In the next chapter, we will examine another regularizer technique, termed gradient smoothing, by considering a modification of HJ-PDE and apply it to gradient-based training algorithm in deep learning.

CHAPTER 5

Laplacian smoothing gradient descent

We propose a class of very simple modifications of gradient descent and stochastic gradient descent. We show that when applied to a large variety of machine learning problems, ranging from logistic regression to deep neural nets, the proposed surrogates can dramatically reduce the variance, allow to take a larger step size, and improve the generalization accuracy. The methods only involve multiplying the usual (stochastic) gradient by the inverse of a positive definitive matrix (which can be computed efficiently by FFT) with a low condition number coming from a one-dimensional discrete Laplacian or its high order generalizations. It also preserves the mean and increases the smallest component and decreases the largest component. The theory of Hamilton-Jacobi partial differential equations demonstrates that the implicit version of the new algorithm is almost the same as doing gradient descent on a new function which (i) has the same global minima as the original function and (ii) is “more convex”. Moreover, we show that optimization algorithms with these surrogates converge uniformly in the discrete Sobolev H_σ^p sense and reduce the optimality gap for convex optimization problems.

5.1 Introduction

Stochastic gradient descent (SGD) [119] has been the workhorse for solving large-scale machine learning (ML) problems. It gives rise to a family of algorithms that enables efficient training of many ML models including deep neural nets (DNNs). SGD utilizes training data very efficiently at the beginning of the training phase, as it converges much faster than GD and L-BFGS during this period [15, 62]. Moreover, the variance of SGD can help gradient-based optimization algorithms circumvent local minima and saddle points and reach those that generalize well [124, 67]. However, the variance of SGD also slows down the convergence after the first few training epochs. To account for the effect of SGD's variance and to ensure the convergence of SGD, a decaying step size has to be applied which is one of the major bottlenecks for the fast convergence of SGD [13, 129, 128]. Moreover, in training many ML models, typically the stage-wise schedule of learning rate is used in practice [127, 124]. In this scenario, the variance of SGD usually leads to a large optimality gap.

A natural question arises from the above bottlenecks of SGD is: *C*an we improve SGD such that the variance of the stochastic gradient is reduced on-the-fly with negligible extra computational and memory overhead and a larger step size is allowed to train ML models?

We answer the above question affirmatively by applying the discrete one-dimensional Laplacian smoothing (LS) operator to smooth the stochastic gradient vector on-the-fly. The LS operation can be performed efficiently by using the fast Fourier transform (FFT).

Another issue of GD and SGD is that when the Hessian of the objective function has a large condition number, GD performs poorly. In this case, the derivative increases rapidly in one direction, while growing slowly in another. As a by-product, numerically we will show that LS can avoid oscillation along steep directions and help make progress in shallow directions effectively [75]. The implicit version of our proposed approach is linked to an unusual Hamilton-Jacobi partial differential equation (HJ-PDE) whose solution makes the original loss function more convex while retaining its flat (and global) minima, and essentially works on this surrogate function with a much better landscape. See [26] for related work.

5.1.1 Our contribution

In this study, we propose a new modification to the stochastic gradient-based algorithms, which at its core uses the LS operator to reduce the variance of stochastic gradient vector on-the-fly. The fundamental of LS is based on Hamilton-Jacobi (HJ) equation and generalized Bregman metric distance. Our Laplacian smoothing gradient descent algorithm approximates solution to the HJ in a linearized sense. The algorithm can be also generalized to higher order and shows improvement on variance reduction as well.

5.1.2 Related work

There is an extensive volume of research over the past decades for designing algorithms to speed up the convergence. These include using momentum and other heavy-ball methods, reduce the variance of the stochastic gradient, and adaptive the learning rate. We will discuss the related work from these three perspectives.

The first type of idea to accelerate the convergence of GD and SGD is to apply the momentum. Around local optima, the surface curves can be much more steeply in one dimension than in another [134], whence (S)GD oscillates across the slopes of the ravine while only making hesitant progress along the bottom towards the local optimum. Momentum is proposed to accelerate (S)GD in the relevant direction and dampens oscillations [113]. Nesterov accelerated gradient (NAG) is also introduced to slow down the progress before the surface curve slopes up, and it provably converge faster in specific scenarios [104]. There are lots of recent progress in the development of momentum; a relatively complete survey can be found at [3].

Due to the bottleneck of the variance of the stochastic gradient, a natural idea is to reduce the variance of the stochastic gradient. There are several principles in developing variance reduction algorithms [15], including Dynamic sample size methods; Gradient aggregation, control variate type of technique is widely used along this direction, some representative works are SAGA [33], SCSG [81], and SVRG [70]; Iterative averaging methods.

Another category of work tries to speed up the convergence of GD and SGD by using an adaptive step size, which makes use of the historical gradient to adapt the step size. RMSProp [136] and Adagrad [41] adapts the learning rate to the parameters, performing smaller updates (i.e., low learning rates) for parameters associated with frequently occurring features, and more substantial updates (i.e., high learning rates) for parameters associated with infrequent features. Both RMSProp and Adagrad make the learning rate to be historical gradient dependent. Adadelta [150] extends the idea of RMSProp and Adagrad, instead of accumulating all past squared gradients, it restricts the window of accumulated past gradients to some fixed size w. Adam [77] and AdaMax [77] behave like a heavy ball with friction, and they compute the decaying averages of past and past squared gradients to adaptive the learning rate. AMSGrad [118] fix the issue of Adam that may fail to converge to an optimal solution. Adam can be viewed as a combination of RMSprop and momentum: RMSprop contributes the exponentially decaying average of past squared gradients, while momentum accounts for the exponentially decaying average of past gradients. Since NAG is superior to vanilla momentum, Dozat [40] proposed NAdam which combines the idea Adam and NAG.

5.1.3 Notations

In this chapter, we use boldface upper-case letters $\boldsymbol{A}$, $\boldsymbol{B}$ to denote matrices and boldface lower-case letters $\boldsymbol{w}$, $\boldsymbol{u}$ to denote vectors. For vectors, we use $\|\cdot\|$ to denote the ℓ_2-norm for vectors and spectral norm for matrices, respectively. And we use $\lambda_{max}(\boldsymbol{A})$, $\lambda_{min}(\boldsymbol{A})$, and $\lambda_i(\boldsymbol{A})$ to denote the largest, smallest, and the i-th largest eigenvalues, respectively. For a function $f:\mathbb{R}^n\rightarrow\mathbb{R}$, we use ∇f and $\nabla^2 f$ to denote its gradient and Hessian, and f^* to denote a local minimum of f. For a positive definite matrix $\boldsymbol{A}$, we define the vector induced norm by as $\|\boldsymbol{w}\|_{\boldsymbol{A}} := \sqrt{\langle \boldsymbol{w}, \boldsymbol{A}\boldsymbol{w}\rangle}$. List $\{1, 2, \cdots, n\}$ is denoted by $[n]$.

5.2 Laplacian Smoothing (Stochastic) Gradient Descent

5.2.1 Hamilton-Jacobi PDEs

The motivation for the proposed LS-SGD comes from the Hamilton-Jacobi PDE (HJ-PDE). Consider the following unusual HJ-PDE with the empirical risk function, $f(\boldsymbol{w})$, as initial condition

$$\begin{cases} u_t + \frac{1}{2}\left\langle \nabla_{\boldsymbol{w}} u, \boldsymbol{A}_\sigma^{-1}\nabla_{\boldsymbol{w}} u\right\rangle = 0, & (\boldsymbol{w}, t) \in \Omega\times[0,\infty) \\ u(\boldsymbol{w}, 0) = f(\boldsymbol{w}), & \boldsymbol{w}\in\Omega \end{cases} \tag{5.1}$$

where $\boldsymbol{A}_\sigma$ is an elliptic operator, i.e. positive definite. By the Hopf-Lax formula [47], the unique viscosity solution to (5.1) is represented by

$$u(\boldsymbol{w}, t) = \inf_{\boldsymbol{v}} \left\{ f(\boldsymbol{v}) + \frac{1}{2t} \big\langle \boldsymbol{v} - \boldsymbol{w}, \boldsymbol{A}_\sigma (\boldsymbol{v} - \boldsymbol{w}) \big\rangle \right\}. \tag{5.2}$$

This viscosity solution $u(\boldsymbol{w}, t)$ makes $f(\boldsymbol{w})$ “more convex”, an intuitive definition and theoretical explanation of ”more convex” can be found in [26], by bringing down the local maxima while retaining and widening local minima. The viscous solution $u(\boldsymbol{w}, t)$ is explained to be the minimum of f plus an penalty around $\boldsymbol{w}$. According to the HJ-PDE, we can take the penalty as generalized Bregman distance metric. Solving the problem 5.2 is equivalent to finding solution to the following backward Euler

$$\boldsymbol{v} = \boldsymbol{w} - t\boldsymbol{A}_\sigma^{-1} f(\boldsymbol{v}) \tag{5.3}$$

Note that the viscous solution $u(\boldsymbol{w}, t)$ can be interpreted as a generalized convex envelope of $f(\boldsymbol{w})$, i.e. both the value and location of some local extrema of the viscous solution are unchanged, while the regions of local extrema are widened. Hence, if we apply this rule iteratively, we can solve for the minimum of f and its location but with smoother convergence. The update reads backward Euler equation as follows

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - t\boldsymbol{A}_\sigma^{-1} f(\boldsymbol{w}^{k+1}) \tag{5.4}$$

Solving backward Euler equation is expensive and inefficient, especially in deep learning problems. Instead, we approximate the solution by fixed point iteration method which requires a couple of inner-iterations for each update. In the most simple case where only one inner iteration is taken, we obtain the gradient descent, or forward Euler

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - t\boldsymbol{A}_\sigma^{-1} f(\boldsymbol{w}^k) \tag{5.5}$$

Since we replace backward Euler by gradient descent, the chosen step-size t needs to be small for better approximation. The tasks now become defining smoothing operator $\boldsymbol{A}_\sigma$ and designing an efficient implementation to Eqn. 5.5.

5.2.2 Smoothing Operator

The smoothing operator $\boldsymbol{A}_\sigma$ applied on gradients is a circulant matrix

$$\boldsymbol{A}_\sigma := \begin{bmatrix} 1+2\sigma & -\sigma & 0 & \dots & 0 & -\sigma \\ -\sigma & 1+2\sigma & -\sigma & \dots & 0 & 0 \\ 0 & -\sigma & 1+2\sigma & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -\sigma & 0 & 0 & \dots & -\sigma & 1+2\sigma \end{bmatrix} \tag{5.6}$$

for some positive constant $\sigma \geq 0$. In fact, we can write $\boldsymbol{A}_\sigma = \boldsymbol{I} - \sigma \boldsymbol{L}$, where $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{L}$ is the discrete one-dimensional Laplacian which acts on indices. If we define the (periodic) forward finite difference matrix as

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ 0 & 0 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 0 & -1 \end{bmatrix}.$$

Then, we have $\boldsymbol{A}_\sigma = \boldsymbol{I} - \sigma \boldsymbol{D}_- \boldsymbol{D}_+$, where $\boldsymbol{D}_- = -\boldsymbol{D}_+^T$ is the backward finite difference. We summarize the benefits of this simple LS operation below:

- It reduces the variance of stochastic gradient on-the-fly, and reduces the optimality gap when constant step size is used.
- It avoids local sharp minima empirically.
- It is more robust to the noisy gradient.

Moreover, as a straightforward extension, we generalize the LS to high-order smoothing operators, e.g., biharmonic smoothing.

5.2.3 Laplacian Smoothing Gradient Descent

We present our algorithm for SGD in the finite-sum setting. The GD and other settings follow straightforwardly. Consider the following finite-sum optimization

$$\min_{\boldsymbol{w}} F(\boldsymbol{w}) := \frac{1}{n}\sum_{i=1}^{n} f_i(\boldsymbol{w}), \tag{5.7}$$

where $f_i(\boldsymbol{w}) \doteq f(\boldsymbol{w}, \boldsymbol{x}_i, y_i)$ is the loss of a given ML model on the training data $\{\boldsymbol{x}_i, y_i\}$. This finite-sum formalism is an abstract of training many ML models mentioned above. To resolve the optimization problem (5.7), starting from some initial guess $\boldsymbol{w}^0$, the $(k+1)$-th iteration of SGD reads

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - \eta_k \nabla f_{i_k}(\boldsymbol{w}^k), \tag{5.8}$$

where η_k is the step size, i_k is a random sample with replacement from $[n]$.

We propose to replace the stochastic gradient $\nabla f_{i_k}(\boldsymbol{w}^k)$ by the Laplacian smoothed surrogate, and we call the resulting algorithm LSSGD, which is written as

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - \eta_k \boldsymbol{A}_\sigma^{-1} \nabla f_{i_k}(\boldsymbol{w}^k). \tag{5.9}$$

Intuitively, compared to the standard GD, this scheme smooths the gradient on-the-fly by an elliptic smoothing operator while preserving the mean of the entries of the gradient. We adopt fast Fourier transform (FFT) to compute $\boldsymbol{A}_\sigma^{-1}\nabla f(\boldsymbol{w}^k)$, which is available in both PyTorch [110] and TensorFlow [1]. Given a vector $\boldsymbol{g}$, a smoothed vector $\boldsymbol{d}$ can be obtained by computing $\boldsymbol{d} = \boldsymbol{A}_\sigma^{-1}\boldsymbol{g}$. This is equivalent to $\boldsymbol{g} = \boldsymbol{d} - \sigma \boldsymbol{v} * \boldsymbol{d}$, where $\boldsymbol{v} = [-2, 1, 0, \cdots, 0, 1]^T$ and $*$ is the convolution operator. Therefore

$$\boldsymbol{d} = \mathrm{ifft}\left(\frac{\mathrm{fft}(\boldsymbol{g})}{\boldsymbol{1} - \sigma \cdot \mathrm{fft}(\boldsymbol{v})}\right),$$

where we use component-wise division (here, fft and ifft are the FFT and inverse FFT, respectively). Hence, the gradient smoothing can be done in quasilinear time. This additional time complexity is almost the same as performing a one step update on the weights vector $\boldsymbol{w}$. For many machine learning models, we may need to concatenate the parameters into a vector. This reshaping might lead to some ambiguity, nevertheless, based on our tests, both row and column majored reshaping work for the LS-GD algorithm. Moreover, in deep learning cases, the weights in different layers might have different physical meanings. For these cases, we perform layer-wise gradient smoothing, instead. We summarize the LSSGD for solving the finite-sum optimization (5.7) in Algorithm 5.

In image processing and elsewhere, the Sobolev gradient [71] uses a multi-dimensional Laplacian operator that operates on $\boldsymbol{w}$, and is different from the one-dimensional discrete Laplacian operator employed in our LS-GD scheme that operates on indices.

5.2.4 Generalized smoothing gradient descent

We can generalize $\boldsymbol{A}_\sigma$ to the n-th order discrete hyper-diffusion operator as follows

$$\boldsymbol{I} + (-1)^n \sigma \boldsymbol{L}^n \doteq \boldsymbol{A}_\sigma^n.$$

Each row of the discrete Laplacian operator $\boldsymbol{L}$ consists of an appropriate arrangement of weights in central finite difference approximation to the 2^{nd} order derivative. Similarly, each row of $\boldsymbol{L}^n$ is an arrangement of the weights in the central finite difference approximation to the $2n^{th}$ order derivative.

Remark: The n-th order smoothing operator $\boldsymbol{I} + (-1)^n \sigma \boldsymbol{L}^n$ can only be applied to the problem with dimension at least $2n+1$. Otherwise, we need to add dummy variables.

Again, we apply FFT to compute the smoothed gradient vector. For a given gradient vector $\boldsymbol{g}$, the smoothed surrogate, $(\boldsymbol{A}_\sigma^n)^{-1}\boldsymbol{g} \doteq \boldsymbol{d}$, can be obtained by solving $\boldsymbol{g} = \boldsymbol{d} + (-1)^n \sigma \boldsymbol{v}_n * \boldsymbol{d}$, where $\boldsymbol{v}_n = (c_{n+1}^n, c_{n+2}^n, \cdots, c_{2n+1}^n, 0, \cdots, 0, c_1^n, c_2^n, \cdots, c_{n-1}^n, c_n^n)$ is a vector of the same dimension as the gradient to be smoothed. And the coefficient vector $\boldsymbol{c}^n = (c_1^n, c_2^n, \cdots, c_{2n+1}^n)$ can be obtained recursively by the following formula

$$\boldsymbol{c}^1 = (1, -2, 1), \quad c_i^n = \begin{cases} 1 & i = 1, 2n+1 \\ -2c_1^{n-1} + c_2^{n-1} & i = 2, 2n \\ c_{i-1}^{n-1} - 2c_i^{n-1} + c_{i+1}^{n-1} & \text{otherwise.} \end{cases}$$

The computational complexities for different order smoothing schemes are the same when the FFT is utilized for computing the surrogate gradient.

Algorithm 5 LSSGD

```
Input: $f_i(\boldsymbol{w})$ for $i = 1, 2, \cdots, n$.
$\boldsymbol{w}^0$: initial guess of $\boldsymbol{w}$, $T$: the total number of iterations, and $\eta_k$, $k = 0, 1, \cdots, T$: the scheduled step size.
Output: The optimized weights $\boldsymbol{w}^{opt}$.
for $k = 0, 1, \cdots, T$ do
    $\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - \eta_k \boldsymbol{A}_\sigma^{-1}\left(\nabla f_{i_k}(\boldsymbol{w}^k)\right)$.
end for
return $\boldsymbol{w}^T$
```

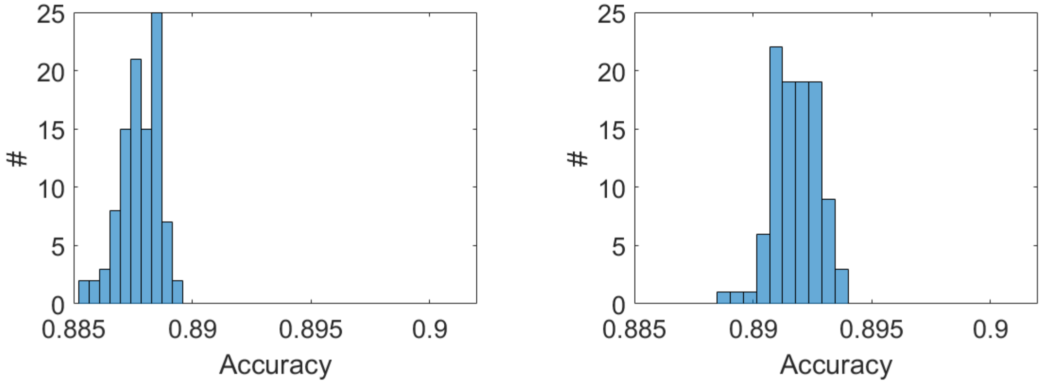
5.3 Experimental results

5.3.1 Multi-class Logistic regression

Consider applying the proposed optimization schemes to train the multi-class Logistic regression model. We run 200 epochs of SGD and different order smoothing algorithms to maximize the likelihood of multi-class Logistic regression with batch size 100. And we apply the exponentially decaying learning rate with initial value 0.5 and decay 10 times after every 50 epochs. We train the model with only 10 % randomly selected MNIST training data and test the trained model on the entire testing images. We further compare with SVRG under the same setting. Figure. 5.1 shows the histograms of generalization accuracy of the model trained by SGD (a); SVRG (b); LS-SGD (order 1) (c); LS-SGD (oder 2) (d). It is seen that SVRG somewhat improves the generalization with higher averaged accuracy. However, the first and the second order LSSGD type algorithms lift the averaged generalization accuracy by more than 1% and reduce the variance of the generalization accuracy over 100 independent trials remarkably.

Next, we show the evolution of the loss in training the multi-class Logistic regression model by SGD, SVRG, LSGD with first and second order smoothing, respectively. As illustrated in Figure 5.2. At each iteration, among 100 independent experiments, SGD has the largest variance, SGD with first order smoothed gradient significantly reduces the

(a) SGD (b) SVRG



(c) LS-GD: Order 1 (d) LS-GD: Order 2

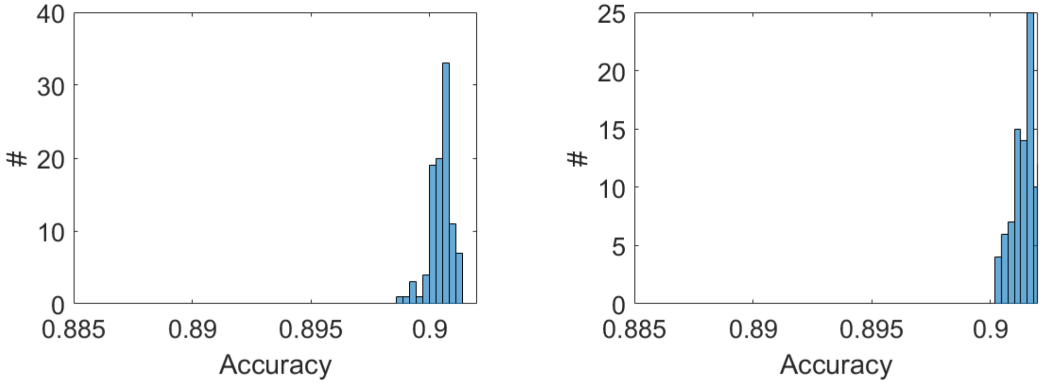
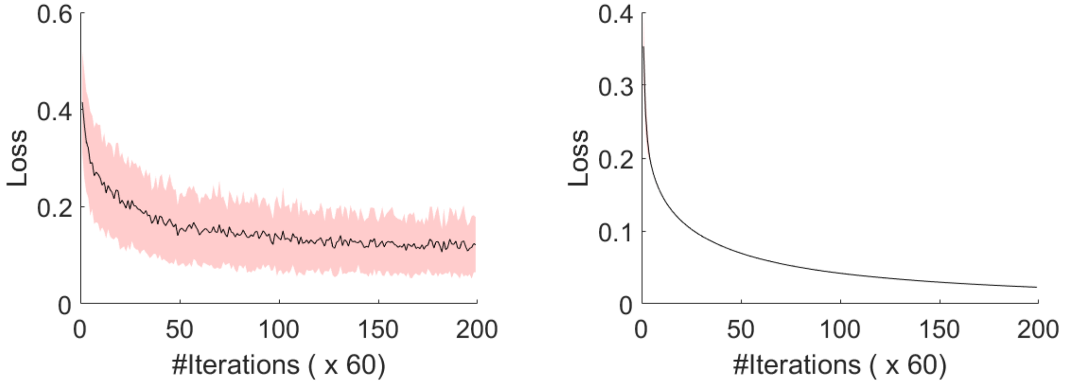


Figure 5.1: Histogram of testing accuracy over 100 independent experiments of the multi-class Logistic regression model trained on randomly selected 10% MNIST data by different algorithms.

variance of loss among different experiments. The second order smoothing can further reduce the variance. The variance of loss in each iteration among 100 experiments is minimized when SVRG is used to train the multi-class Logistic model. However, the generalization performance of the model trained by SVRG is not as good as the ones by LS-SGD, or higher order smoothed gradient descent (Figure 5.1 (b)).


(a) SGD (b) SVRG

(c) LS-GD: Order 1 (d) LS-GD: Order 2

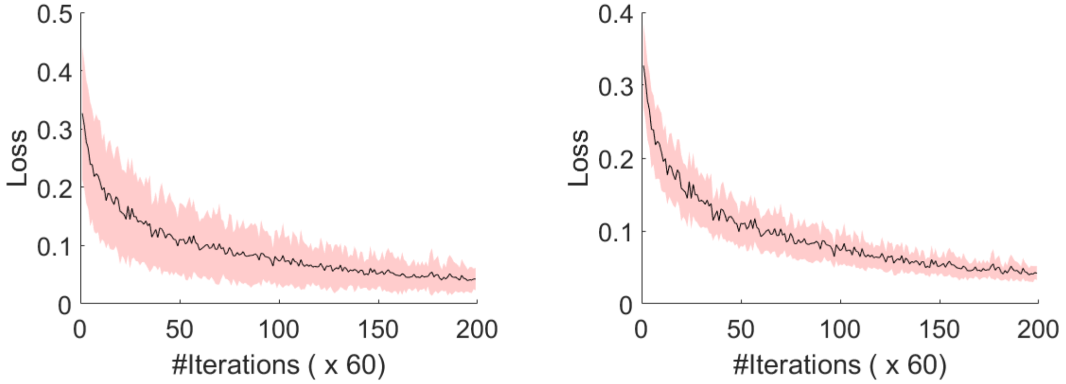


Figure 5.2: Iterations v.s. loss for SGD, SVRG, and LS-SGD with order 1 and order 2 gradient smoothing for training the multi-class Logistic regression model.

5.3.2 LeNet with small batch size

Many advanced artificial intelligence tasks make high demands on training neural nets with extremely small batch sizes. The milestone technique for this is group normalization [142]. In this section, we show that LS-SGD successfully trains DNN with extremely small batch size. We consider LeNet-5 [79] for MNIST classification. Our network architecture is as follows

$$\text{LeNet-5: } input_{28\times 28} \to conv_{20,5,2} \to conv_{50,5,2} \to \text{fc}_{512} \to softmax.$$

The notation $conv_{c,k,m}$ denotes a 2D convolutional layer with c output channels, each of which is the sum of a channel-wise convolution operation on the input using a learnable kernel of size $k \times k$, it further adds ReLU nonlinearity and max pooling with stride size m. fc_{512} is an affine transformation that transforms the input to a vector of dimension 512. Finally, the tensors are activated by a multi-class Logistic function. The MNIST data is first passed to the layer $input_{28\times 28}$, and further processed by this hierarchical structure. We run 100 epochs of both SGD and LS-SGD with initial learning rate 0.01 and divide by 5 after 50 epochs, and use a weight decay of 0.0001 and momentum of 0.9. Figure. 5.3(a) plots the generalization accuracy on the test set with the LeNet5 trained with different batch sizes. For each batch size, LS-SGD with $\sigma = 1.0$ keeps the testing accuracy more than 99.4%, SGD reduce the accuracy to 97% when batch size 4 is used. The classification become just a random guess, when the model is trained by SGD with batch size 2. Small batch size leads to large noise in the gradient, which may make the noisy gradient not along the decent direction; however, Lapacian smoothing rescues this by decreasing the noise.

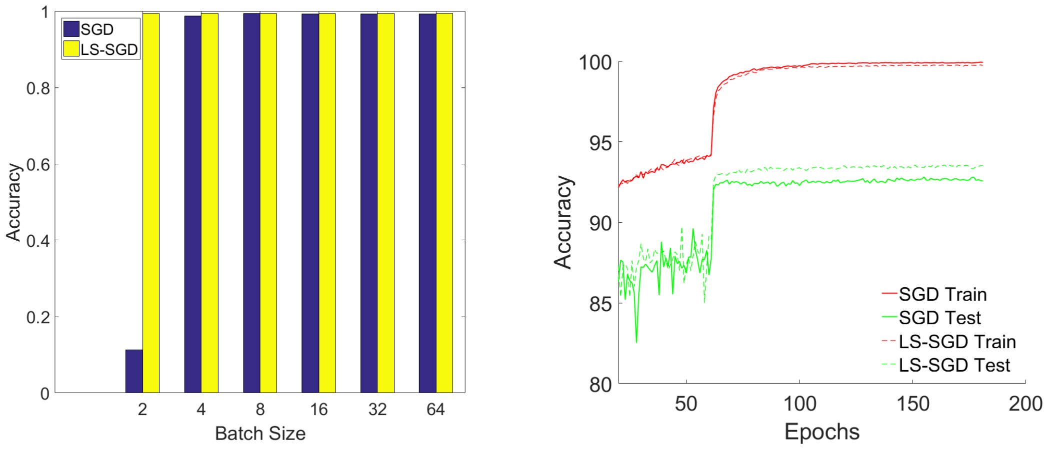


Figure 5.3: (a). Testing accuracy of LeNet5 trained by SGD/LS-SGD on MNIST with various batch sizes. (b). The evolution of the pre-activated ResNet56's training and generalization accuracy by SGD and LS-SGD. (Start from the 20-th epoch.)

5.3.3 Improve generalization accuracy

The skip connections in ResNet smooth the landscape of the loss function of the classical CNN [63, 84]. This means that ResNet has fewer sharp minima. On Cifar10 [78], we compare the performance of LS-SGD and SGD on ResNet with the pre-activated ResNet56 as an illustration. We take the same training strategy as that used in [63], except that we run 200 epochs with the learning rate decaying by a factor of 5 after every 40 epochs. For ResNet, instead of applying LS-SGD for all epochs, we only use LS-SGD in the first 40 epochs, and the remaining training is carried out by SGD (this will save the extra computational cost due to LS, and we noticed that the performance is similar to the case when LS is used for the whole training process). The parameter σ is set to 1.0. Figure 5.3(b) depicts one path of the training and generalization accuracy of the neural nets trained by SGD and LS-SGD, respectively. It is seen that, even though the training accuracy obtained by SGD is higher than that by LS-SGD, the generalization is however inferior to that of LS-SGD. We conjecture that this is due to the fact that SGD gets trapped into some sharp but deeper minimum, which fits better than a flat minimum but generalizes worse. We carry out 25 replicas of this experiments, the histograms of the corresponding accuracy are shown in Figure 5.4.

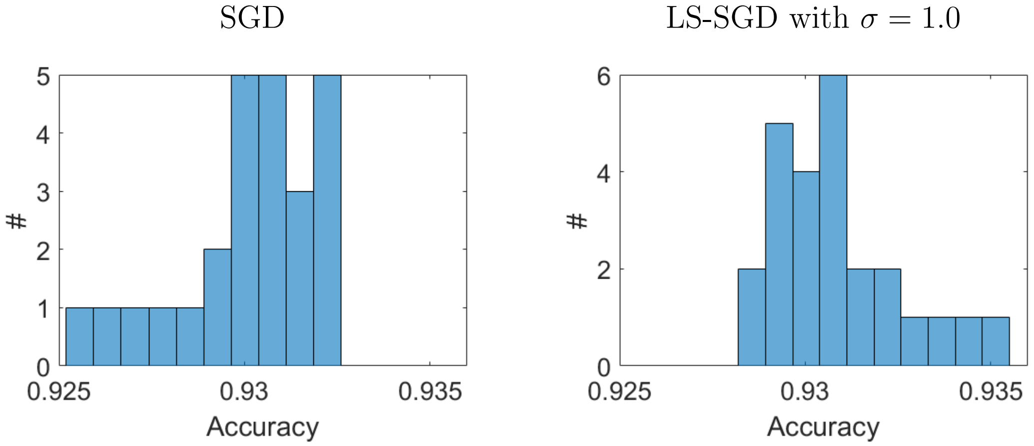


Figure 5.4: The histogram of the generalization accuracy of the pre-activated ResNet56 on Cifar10 trained by SGD and LS-SGD over 25 independent experiments.

5.3.4 Training Wassersterin GAN

Generative Adversarial Networks (GANs) [59] are notoriously delicate and unstable to train [5]. In [6], Wasserstein-GANs (WGANs) are introduced to combat the instability in the training GANs. In addition to being more robust in training parameters and network architecture, WGANs provide a reliable estimate of the Earth Mover (EM) metric which correlates well with the quality of the generated samples. Nonetheless, WGANs training becomes unstable with a large learning rate or when used with a momentum based optimizer [6]. In this section, we demonstrate that the gradient smoothing technique in this paper alleviates the instability in the training, and improves the quality of generated samples. Since WGANs with weight clipping are typically trained with RMSProp [136], we propose replacing the gradient g by a smoothed version $g_\sigma = \boldsymbol{A}_\sigma^{-1} g$, and also update the running averages using g_σ instead of g. We name this algorithm LS-RMSProp.

To accentuate the instability in training and demonstrate the effects of gradient smoothing, we deliberately use a large learning rate for training the generator. We compare the regular RMSProp with the LS-RMSProp. The learning rate for the critic is kept small and trained approximately to convergence so that the critic loss is still an effective approximation to the Wasserstein distance. To control the number of unknowns in the experiment and make a meaningful comparison using the critic loss, we use the classical RMSProp for the critic, and only apply LS-RMSProp to the generator.

We train the WGANs on the MNIST dataset using the DCGAN [114] for both the critic and generator. In Figure 5.5 (left), we observe the loss for RMSProp trained with a large learning rate has multiple sharp spikes, indicating instability in the training process. The samples generated are also lower in quality, containing noisy spots as shown in Figure 5.6 (a). In contrast, the curve of training loss for LS-RMSProp is smoother and exhibits fewer spikes. The generated samples as shown in Figure 5.6 (b) are also of better quality and visibly less noisy. The generated characters shown in Figure 5.6 (b) are more realistic compared to the ones shown in Figure 5.6 (a). The effects are less pronounced with a small learning rate, but still result in a modest improvement in sample quality as shown in Figure 5.6 (c) and

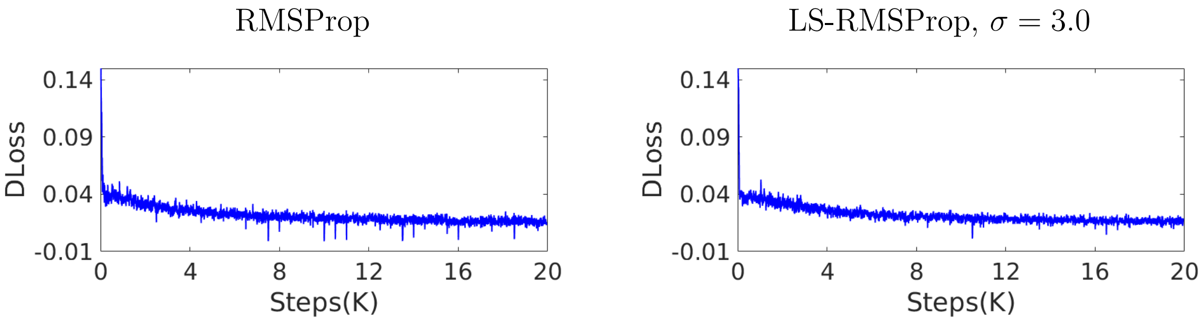
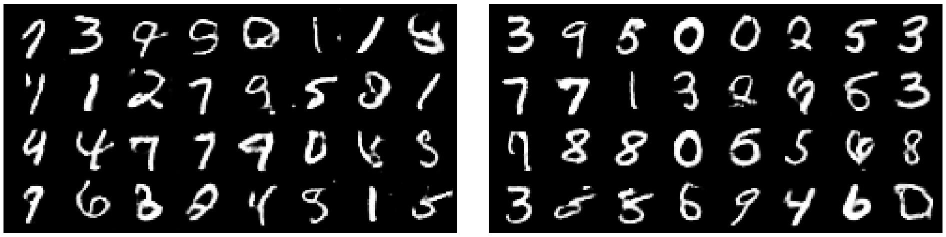


Figure 5.5: Critic loss with learning rate $lrD = 0.0001$, $lrG = 0.005$ for RMSProp (left) and LS-RMSProp (right), trained for 20K iterations. We apply a mean filter of window size 13 for better visualization. The loss from LS-RMSProp is visibly less noisy.

(a) RMSProp (b) LS-RMSProp, $\sigma = 3.0$



(c) RMSProp (d) LS-RMSProp, $\sigma = 3.0$

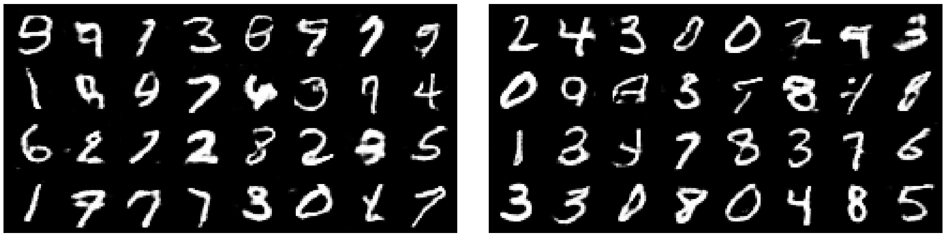


Figure 5.6: Samples from WGANs trained with RMSProp (a, c) and LS-RMSProp (b, d). The learning rate is set to $lrD = 0.0001$, $lrG = 0.005$ for both RMSProp and LS-RMSProp in (a) and (b). And $lrD = 0.0001$, $lrG = 0.0001$ are used for both RMSProp and LS-RMSProp in (c) and (d). The critic is trained for 5 iterations per step of the generator, and 200 iterations per every 500 steps of the generator.

(d).We also apply LS-RMSProp for training the critic, but do not see a clear improvement in the quality. This may be because the critic is already trained near optimality during each iteration, and does not benefit much from gradient smoothing.

5.4 Discussion & Conclusion

5.4.1 Connection to Hamilton-Jacobi PDEs

We take an example to Hamilton-Jacobi PDEs and demonstrate its solution as a function of time. Figure 5.7 shows that the viscous solution $u(\boldsymbol{w}, t)$ to Eqn. 5.1 for a given time t acts as a surrogate to $f(\boldsymbol{w})$ as it brings down the local maxima and shrink their concave regions while retaining and widening the local minima.

If we could solve the backward Euler in Eqn. 5.4 with proper step size on the function $u(\boldsymbol{w}, t)$, then we can reach the global or at least a flatter minima of the original nonconvex function $f(\boldsymbol{w})$.

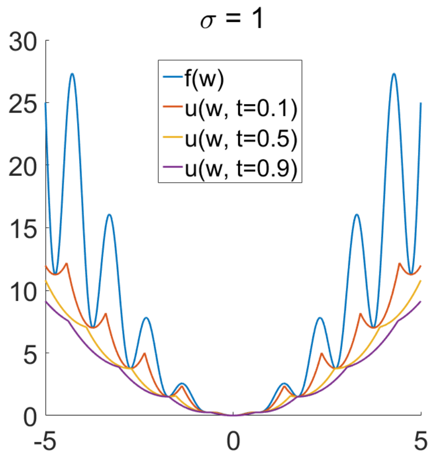


Figure 5.7: $f(\boldsymbol{w}) = \|\boldsymbol{w}\|^2\big(1 + \frac{1}{2}\sin(2\pi\|\boldsymbol{w}\|)\big)$ is made more convex by solving (5.1). The plot shows the cross section of the 5D problem with $\sigma = 1$ and different t values.

Notice that the viscous solution $u(\boldsymbol{w}, t)$ is flattened more as t increases. Since we replace the backward Euler by the gradient descent, small step-size t must be taken

5.4.2 Conclusion

Motivated by the theory of Hamilton-Jacobi partial differential equations, we proposed Laplacian smoothing gradient descent and its high order generalizations. This simple modification dramatically reduces the variance and optimality gap in stochastic gradient descent, allows us to take a larger step size, and helps to find better minima. Extensive numerical examples ranging from toy cases and shallow and deep neural nets to generative adversarial networks and deep reinforcement learning, all demonstrate the advantage of the proposed smoothed gradient.

This result shows an application of our HJ-PDE technique in machine learning and ends the thesis of new algorithms in computational microscopy. We give summary to our finding and discuss future research in the conclusion chapter.

CHAPTER 6

Conclusion

6.1 Summary

My thesis demonstrated algorithm developments in microscopy inverse problems via optimization techniques. Two main topics that were addressed are phase retrieval in coherent diffracting imaging and 3D tomography reconstruction in electron microscopy. The optimization techniques used in these researches were also found applications in machine learning, specifically K-mean clustering and gradient-based algorithm for training process in deep learning. The thesis is summarized into 3 aims with 4 studies as follows.

Aim 1: Algorithm development for phase retrieval

Study 1 reported algorithm development for phase retrieval in coherent diffractive imaging. The method was developed from the motivation of Hamilton-Jacobi partial differential equations (HJ-PDE), generalized Bregman distance, and Tikhonov regularizers. The phase retrieval objective function, which is non-convex & non-smooth, was relaxed by generalized proximal smoothing and resulted in a "better" smoothed function. Since we apply smoothing/Tikhonov regularizer on objective function instead of object, minimizers barely change when value γ of the regularizer is small enough. Therefore, we incorporated our algorithm generalized proximal smoothing (GPS), derived from primal-dual hybrid gradient (PDHG), with the increasing in γ along the iterations to improve the generalization of solutions.

Through simulation and experimental data, we demonstrated that the algorithm improves image processing fidelity, minimums and consistency relative to hybrid input-output (HIO) and oversampling smoothness (OSS). GPS was shown to have higher capability to overcome bad local minima (sharp & narrow) and stay in good ones (deep & wide). Therefore, GPS can obtain lower objective values and achieve minimizers that have higher generalization. The Tikhonov regularizer used in GPS helps to improve reliability and robustness when working with noise. Moreover, the regularizer was also shown to obtain smoother convergence and less oscillation than other algorithms. GPS now becomes a great asset to the class of phase retrieval methods.

Aim 2: Algorithm development for tomography

Study 2 reported a new algorithm for 3D tomography reconstruction, named "real space iterative reconstruction engine" RESIRE. In this research, we approached the goal via l_2 minimization and took motivation from algebraic technique (ART). In the light of variational calculus and functional derivative, we derived a highly accurate "back projection" operator. Through simulations and experimental data, RESIRE was shown to reconstruct high resolution 3D imaging and superior results to other methods, such as filtered back projection (FBP), Simultaneously iterative reconstruction technique (SIRT), and generalized Fourier iterative reconstruction (GENFIRE). RESIRE provided more faithful images with less missing wedge artifacts, while noise and blur were significantly reduced.

Moreover, RESIRE was further developed to work with non-isolated objects and partially-blocked projections where microscope could only capture parts of samples or samples are blocked or grasped by caps. Those scenarios have become more popular in tomography and challenged reconstruction algorithms. While other methods restrict their reconstructions to only isolated objects, RESIRE provides a flexible but also rigorous technique to decipher these dilemmas. Inductively, RESIRE can choose good regions of projections for reconstructions and masks out bad ones; therefore, maximizes field of view FOV. Additionally, the flexibility allows reconstruction to be performed via partitioning. This functionality permits sub-reconstructions with different depth as controlling the depth is believed to improve reconstruction quality. Last but not least, the partitioning enhances parallel computing and allows reconstructions with larger samples.

Aim 3: Application to machine learning

Study 3 and 4 reported applications of optimization technique used in study 1 to machine learning, including K-means clustering and gradient-based algorithm for deep learning.

Specifically, study 3 developed stochastic backward Euler (SBE) algorithm from Hopf-Lax formula, an implicit-form solution to HJ-PDE, to solve K-mean clustering. This technique was used in study 1 as a smoothing regularizer to a non-convex & non-smooth objective function when an unusual non-viscous HJ-PDE was being used. In this research, we focused on viscous HJ-PDE instead and interpreted its Hopf-Lax formula as backward Euler method in a stochastic sense. This corporation also provided a better "descent" direction for gradient-based algorithms as GPS did. Through simulation and experimental data, SBE was shown to achieve lower minimums with higher consistency. When naive K-mean gradient descent got trapped in bad local minima, SBE was demonstrated to overcome these traps by sampling "better" descent directions. This highlighted the advantage of such a backward Euler scheme and function smoothing in non-convex & non-smooth optimization.

Another application of function smoothing via HJ-PDE was shown in study 4. The derived algorithm, named Laplacian smoothing gradient descent (LSGD), relaxed gradients by Laplacian smoothing to compensate for stochasticity in random sampling. LSGD reduced the randomness in stochastic gradient and provided a smooth descent direction to avoid bad local minima (shape & narrow) and seek for good ones (deep & wide minima). This technique is also considered as function smoothing as in study 1 & 3. Through simulation and experimental tests ranging from deep neural nets to generative adversarial networks and deep reinforcement learning, LSGD was shown to achieve lower minimums with smoother convergence and higher consistency. Especially in classification problems, LSGD was shown to obtain higher accuracy in both training and testing, i.e. higher generalization.

The triumph of this thesis opens more applications and directions in future research which is discussed in the final section.

6.2 Future direction

Vector tomography

Magnetic metalattices are thought to exhibit multiple, topologically distinct magnetic phases which sharply transition as a function of the applied field. This behavior makes the metalattice a candidate for a new class of "spin-engineered" materials. We would like to explore the vector field caused by magnetic metalattices via ptychographic tomography experiments using X-ray microscopy technique. This opens a new area of research as well a new mathematical problem to microscopy community. Being pioneer in this field, we would like to be the first in developing a new algorithm for vector tomography. The principle of this development is based on the reconstruction algorithm developed in study 2. As discussion in chapter 3, the flexibility in RESIRE provides the ability to extend 3D to 4D imaging.

Beside developing the algorithm, we would like to design an optimal experimentation for vector tomography. Through simulation, we analyze the minimization problem, and optimize experimental parameters in order to produce an optimal solutions, i.e. maximize outcome and minimize error. The results represents an optimal strategy for experimentation.

Non-linear tomography with multiple scattering effects

Recent electron tomography assumes linearity in projections but does not take in account the multiple scattering effect between quantum matter waves like electrons and solid material. In order to improve imaging fidelity, this effect should be considered in tomographic reconstruction algorithms; hence non-linear tomography with multiple scattering effect arises together with demand for algorithm development. The multiple scattering effect can be simulated via a numerical non-linear iterative process, known as multislice method. Due to computational expense, the simulation is too difficult to be integrated in reconstruction process. To address this problem, we need to search for a faster and simpler simulation in the trade-off with accuracy. Another possible approach is developing a non-linear tomography model via statistical and probabilistic methods. By data fitting, we wish to obtain a correlation between linear and multislice projections. Basic non-linear operators such as polynomials, exponents, and their combination can be examined as a candidate for the correlation.

Super resolution ptychography

By diffraction limit, resolution of coherent diffractive imaging is defined by scattering angle and wavelength of the illumination used in experiment. To improve the resolution, electromagnetic wave with shorter wavelength or aperture with structure that can obtain larger scattering angle should be exploited. However, for a fixed illumination and aperture, numerical method can help to obtain the resolution beyond diffraction limit. Many research in the last decade revealed that a combination of structured aperture and expanding diffraction pattern can improve the resolution by a factor of 2, and maybe up to 4. However diffraction expansion causes instability in reconstruction process due to the increasing number of variables. As the degrees of freedom increase, the system becomes under-determined. Improving resolution by a factor of 4 only happens in ideal systems. We wish to develop an advanced algorithm with the help of a Tikhonov regularizer to overcome this under-constraint minimization. The regularizer needs not to be biased and should be developed based on prior knowledge of the reconstructed and diffractive images.

APPENDIX A

Chapter 2-5 supplementary materials

A.1 Chapter 2 supplementary materials

Proximal mapping on complex domain

We extend the definition of the Moreau-Yosida regularization [103, 148] to complex domain. Let $G \in \mathbb{C}^{n\times n}$ be a Hermitian positive definite matrix. The Moreau-Yosida regularization of a lower semi-continuous extended-real-valued function $h : \mathbb{C}^n \to (-\infty, \infty]$, associated with G, is defined by

$$h_G(u) := \inf_{v\in\mathbb{C}^n} \Big\{ h(v) + \frac{1}{2}\|v-u\|^2_{G^{-1}} \Big\}. \tag{A.1}$$

We see that h_G converges pointwise to h as $\|G\| \to 0^+$. In the special case where $G = tI$ is a multiple of identity matrix with $t > 0$, h_G reduces to

$$h_t(u) := \inf_{v\in\mathbb{C}^n} \Big\{ h(v) + \frac{1}{2t}\|v-u\|^2 \Big\}. \tag{A.2}$$

For any indicator function $h = \mathcal{I}_{\mathcal{X}}$ of a closed set $\mathcal{X} \subset \mathbb{C}^n$,

$$h_t(u) := \frac{1}{2t} \inf_{v\in\mathcal{X}} \|v-u\|^2 = \frac{1}{2t}\|u - \mathrm{proj}_{\mathcal{X}}(u)\|^2 \tag{A.3}$$

is $\frac{1}{2t}$ of the squared ℓ_2 distance from u to the set $\mathcal{X}$. Similar idea of relaxing an indicator function into a distance function has been successfully applied to the quantization problem of deep neural networks in [147]. Taking $\mathcal{X} = \{z \in \mathbb{R}^n : |z| = b\}$ to be the magnitude constraint set and $\sigma > 0$, we first relax $g = \mathcal{I}_{|z|=b}$ in (2.3) into

$$g_\sigma(z) := \frac{1}{2\sigma} \inf_{|v|=b} \|v-z\|^2. \tag{A.4}$$

Since $\mathrm{proj}_{|z|=b}(z) = b \odot \exp(\mathbf{i} \cdot \arg(z))$ is the projection of z onto the set $\{z \in \mathbb{R}^n : |z| = b\}$, a simple calculation shows that

$$g_\sigma(z) = \frac{1}{2\sigma}\|b \odot \exp(\mathbf{i} \cdot \arg(z)) - z\|^2 = \frac{1}{2\sigma}\|(b - |z|) \odot \exp(\mathbf{i} \cdot \arg(z))\|^2 = \frac{1}{2\sigma}\||z| - b\|^2 \quad \text{(A.5)}$$

is a least squares fidelity, which measures the difference between the observed magnitudes and fitted ones.

Generalized Legendre-Fenchel transformation

We can express any function $h : \mathbb{C}^n \to \mathbb{R}$ as a function $\tilde{h}$ defined on $\mathbb{R}^{2n}$ in the following way

$$h(u) := \tilde{h}\big(\mathrm{Re}(u), \mathrm{Im}(u)\big) = \tilde{h}(\tilde{u}), \tag{A.6}$$

where $\tilde{u} = \Big[\mathrm{Re}(u); \mathrm{Im}(u)\Big] \in \mathbb{R}^{2n}$ and $\tilde{h} : \mathbb{R}^{2n} \to \mathbb{R}$. We define that h is convex, if $\tilde{h}$ is convex on $\mathbb{R}^{2n}$. Note that for any $u, y \in \mathbb{C}^n$,

$$\mathrm{Re}\langle u, y\rangle = \langle \mathrm{Re}(u), \mathrm{Re}(y)\rangle + \langle \mathrm{Im}(u), \mathrm{Im}(y)\rangle = \langle \tilde{u}, \tilde{y}\rangle. \tag{A.7}$$

We propose to generalize the Legendre-Fenchel transformation (a.k.a. convex conjugate) [122] of an extended-real-valued convex function h defined on $\mathbb{C}^n$ as

$$h^*(y) := \sup_{u \in \mathbb{C}^n} \Big\{\mathrm{Re}\langle y, u\rangle - h(u)\Big\}. \tag{A.8}$$

Infimal convolution and Moreau-Yoshida decomposition

By Eq. (2.5), f_G is the infimal convolution [111] between the convex functions f and $\frac{1}{2}\|\cdot\|^2_{G^{-1}}$. The Moreau-Yoshida decomposition decouples the Legendre-Fenchel transformation of f_G

$$f_G^*(y) = f^*(y) + \Big(\frac{1}{2}\|\cdot\|^2_{G^{-1}}\Big)^*(y), \tag{A.9}$$

where $\Big(\frac{1}{2}\|\cdot\|^2_{G^{-1}}\Big)^*(y) = \frac{1}{2}\|y\|^2_G$. Here we remark that G only needs to be positive semi-definite in Eq. (2.5), as f_G can take the extended value ∞. In this case, although G^{-1} does not exist, since f_G is convex and lower semi-continuous, and the strong duality $f_G^{**} = f_G$

holds here, we can re-define f_G via the biconjugate as

$$\begin{aligned} f_G(u) = f_G^{**}(u) = \big(f_G^*(y)\big)^* &= \sup_{y\in\mathbb{C}^n} \Big\{\mathrm{Re}\langle u, y\rangle - f_G^*(y)\Big\} = \sup_{y\in\mathbb{C}^n} \Big\{\mathrm{Re}\langle u, y\rangle - f^*(y) - \frac{1}{2}\|y\|_G^2\Big\} \\ &= \sup_{y\in\mathcal{S}^*} \Big\{\mathrm{Re}\langle u, y\rangle - \frac{1}{2}\|y\|_G^2\Big\}. \end{aligned} \tag{A.10}$$

Bibliography

[1] M. Abadi, A. Agarwal, and et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[2] M. Abramowitz and M. Davidson. Introduction to microscopy. *Molecular Expressions*, pages 08–22, 2007.

[3] Z. Allen-Zhu. Katyusha: The first direct acceleration of stochastic gradient methods. *Journal of Machine Learning Research*, 18(1):8194–8244, 2017.

[4] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1): 81–94, 1984.

[5] M. Arjovsky and L. Bottou. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*, 2017.

[6] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 214–223. PMLR, 2017.

[7] D. Arthur and S. Vassilvitskii. K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, page 1027–1035. Society for Industrial and Applied Mathematics, 2007.

[8] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations*, volume 147. Springer Science & Business Media, 2006.

[9] C. Baldassi, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina. Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses. *Physical Review Letters*, 115(12):128101, 2015.

[10] H. Bauschke, P. Combettes, and D. Luke. Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. *JOSA A*, 19(7):1334–45, 2002.

[11] H. Bauschke, P. Combettes, and D. Luke. Hybrid projection-reflection method for phase retrieval. *JOSA A*, 20(6):1025–34, 2003.

[12] M. Born and E. Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light.* Elsevier, 2013.

[13] L. Bottou. Stochastic gradient descent tricks. *Neural Networks, Tricks of the Trade, Reloaded*, 7700:421–436, 2012.

[14] L. Bottou and Y. Bengio. Convergence properties of the k-means algorithms. In *Proceedings of the 7th International Conference on Neural Information Processing Systems*, pages 585–592. MIT Press, 1995.

[15] L. Bottou, E. F. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018.

[16] 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*, 3(1):1–122, 2011.

[17] R. N. Bracewell. *Two-dimensional imaging*, pages 505–537. Prentice-Hall, Inc., 1995.

[18] E. Candès, Y. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Journal on Imaging Sciences*, 6(1):199–225, 2013.

[19] E. Candès, T. Strohmer, and V. Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[20] E. J. Candes, X. Li, and M. Soltanolkotabi. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

[21] A. Chai, M. Moscoso, and G. Papanicolaou. Array imaging using intensity-only measurements. *Inverse Problems*, 27(1):015005, 2010.

[22] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1): 120–145, 2011.

[23] H. Chang, Y. Lou, M. K. Ng, and T. Zeng. Phase retrieval from incomplete magnitude information via total variation regularization. *SIAM Journal on Scientific Computing*, 38(6):A3672–A3695, 2016.

[24] H. Chang, S. Marchesini, Y. Lou, and T. Zeng. Variational phase retrieval with globally convergent preconditioned proximal algorithm. *SIAM Journal on Imaging Sciences*, 11(1):56–93, 2018.

[25] H. N. Chapman and K. A. Nugent. Coherent lensless X-ray imaging. *Nature Photonics*, 4(12):833–839, 2010.

[26] P. Chaudhari, A. Oberman, S. Osher, S. Soatto, and G. Carlier. Deep relaxation: partial differential equations for optimizing deep neural networks. *Research in the Mathematical Sciences*, 5(3):30, 2018.

[27] P. Chaudhari, A. Choromanska, S. Soatto, Y. LeCun, C. Baldassi, C. Borgs, J. Chayes, L. Sagun, and R. Zecchina. Entropy-sgd: Biasing gradient descent into wide valleys. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124018, 2019.

[28] C. Chen, J. Miao, C. Wang, and T. Lee. Application of optimization technique to noncrystalline X-ray diffraction microscopy: Guided hybrid input-output method. *Physical Review B*, 76(6):064113, 2007.

[29] C.-C. Chen, C. Zhu, E. R. White, C.-Y. Chiu, M. Scott, B. Regan, L. D. Marks, Y. Huang, and J. Miao. Three-dimensional imaging of dislocations in a nanoparticle at atomic resolution. *Nature*, 496(7443):74–77, 2013.

[30] P. Chen and A. Fannjiang. Fourier phase retrieval with a single mask by Douglas-Rachford algorithms. *Applied and Computational Harmonic Analysis*, 44(3):665–699, 2018.

[31] E. K. Chong and S. H. Zak. *An introduction to optimization*, pages 226–230. John Wiley & Sons, 3rd edition, 2008.

[32] C. Clason, B. Kaltenbacher, and E. Resmerita. Regularization of ill-posed problems with non-negative solutions. In *Splitting Algorithms, Modern Operator Theory, and Applications*, pages 113–135. Springer, 2019.

[33] A. Defazio and F. Bach. SAGA: A fast incremental gradient method with support for non-strongly convex composite objectives. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1*, page 1646–1654. MIT Press, 2014.

[34] M. Dierolf, A. Menzel, P. Thibault, P. Schneider, C. M. Kewish, R. Wepf, O. Bunk, and F. Pfeiffer. Ptychographic X-ray computed tomography at the nanoscale. *Nature*, 467(7314):436–439, 2010.

[35] Y. Ding, Y. Zhao, X. Shen, M. Musuvathi, and T. Mytkowicz. Yinyang k-means: A drop-in replacement of the classic k-means with consistent speedup. In *Proceedings of the 32nd International Conference on Machine Learning - Volume 37*, pages 579–587. JMLR.org, 2015.

[36] C. Donnelly. *Hard X-ray tomography of three dimensional magnetic structures*. PhD thesis, ETH Zurich, 2017.

[37] C. Donnelly, S. Gliga, V. Scagnoli, M. Holler, J. Raabe, L. J. Heyderman, and M. Guizar-Sicairos. Tomographic reconstruction of a three-dimensional magnetization vector field. *New Journal of Physics*, 20(8):83009, 2018.

[38] C. Donnelly, M. Guizar-Sicairos, V. Scagnoli, S. Gliga, M. Holler, J. Raabe, and L. J. Heyderman. Hard X-ray magnetic tomography: A new technique for the visualization

of three dimensional magnetic structures. *Microscopy and Microanalysis*, 24(S2):80–81, 2018.

[39] J. Douglas and H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439, 1956.

[40] T. Dozat. Incorporating Nesterov momentum into Adam. In *4th International Conference on Learning Representation Workshop*, 2016.

[41] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, 2011.

[42] D. E. Dudgeon. *Multidimensional digital signal processing.* Prentice-Hall, 1983.

[43] C. Elkan. Using the triangle inequality to accelerate k-means. In *Proceedings of the 20th International Conference on Machine Learning*, pages 147–153. AAAI Press, 2003.

[44] V. Elser. Phase retrieval by iterated projections. *JOSA A*, 20(1):40–55, 2003.

[45] R. Erni, M. D. Rossell, C. Kisielowski, and U. Dahmen. Atomic-resolution imaging with a sub-50-pm electron probe. *Physical Review Letters*, 102(9):096101, 2009.

[46] E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[47] L. Evans. *Partial Differential Equations.* American Mathematical Society, 2nd edition, 2010.

[48] B. P. Fahimian, Y. Mao, P. Cloetens, and J. Miao. Low-dose X-ray phase-contrast and absorption CT using equally sloped tomography. *Physics in Medicine & Biology*, 55(18):5383, 2010.

[49] B. P. Fahimian, Y. Zhao, Z. Huang, R. Fung, Y. Mao, C. Zhu, M. Khatonabadi, J. J. DeMarco, S. J. Osher, M. F. McNitt-Gray, et al. Radiation dose reduction in medical X-ray CT via Fourier-based iterative reconstruction. *Medical physics*, 40(3):31914, 2013.

[50] J. Fienup. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters*, 3(1):27–29, 1978.

[51] J. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, 1982.

[52] J. Fienup and C. Wackerman. Phase-retrieval stagnation problems and solutions. *JOSA A*, 3(11):1897–1907, 1986.

[53] J. Frank. *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state.* Oxford University Press, 2006.

[54] R. Franke. Scattered data interpolation: tests of some methods. *Mathematics of Computation*, 38(157):181–200, 1982.

[55] R. Gerchberg and W. Saxton. A practical algorithm for the determination of the phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

[56] P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36(1):105–117, 1972.

[57] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, 9(R2):41–76, 1975.

[58] T. Goldstein, M. Li, X. Yuan, E. Esser, and R. Baraniuk. Adaptive primal-dual hybrid gradient methods for saddle-point problems. *arXiv preprint arXiv:1305.0546*, 2013.

[59] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, pages 2672–2680. MIT Press, 2014.

[60] 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*, 29(3):471–481, 1970.

[61] G. Haberfehlner, P. Thaler, D. Knez, A. Volk, F. Hofer, W. E. Ernst, and G. Kothleitner. Formation of bimetallic clusters in superfluid helium nanodroplets analysed by atomic resolution electron tomography. *Nature Communications*, 6(1):8779, 2015.

[62] M. Hardt, B. Recht, and Y. Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *Proceedings of the 33rd International Conference on Machine Learning - Volume 48*, page 1225–1234. JMLR.org, 2016.

[63] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[64] G. T. Herman. *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media, 2009.

[65] M. Holler, M. Guizar-Sicairos, E. H. Tsai, R. Dinapoli, E. Müller, O. Bunk, J. Raabe, and G. Aeppli. High-resolution non-destructive three-dimensional imaging of integrated circuits. *Nature*, 543(7645):402–406, 2017.

[66] X. Huang, Z. Zhao, L. Cao, Y. Chen, E. Zhu, Z. Lin, M. Li, A. Yan, A. Zettl, Y. M. Wang, X. Duan, T. Mueller, and Y. Huang. High-performance transition metal–doped pt3ni octahedra for oxygen reduction reaction. *Science*, 348(6240):1230–1234, 2015.

[67] S. Jastrzebski, Z. Kenton, N. Ballas, A. Fischer, Y. Bengio, and A. Storkey. DNN's sharpest directions along the SGD trajectory. *Stat*, 1050:13, 2018.

[68] H. Jiang, C. Song, C.-C. Chen, R. Xu, K. S. Raines, B. P. Fahimian, C.-H. Lu, T.-K. Lee, A. Nakashima, J. Urano, et al. Quantitative 3D imaging of whole, unstained cells by using X-ray diffraction microscopy. *Proceedings of the National Academy of Sciences*, 107(25):11234–11239, 2010.

[69] H. Jiang, R. Xu, C.-C. Chen, W. Yang, J. Fan, X. Tao, C. Song, Y. Kohmura, T. Xiao, Y. Wang, et al. Three-dimensional coherent X-ray diffraction imaging of molten iron in mantle olivine at nanoscale resolution. *Physical Review Letters*, 110(20):205501, 2013.

[70] R. Johoson and T. Zhang. Accelerating stochastic gradient descent using predictive variance reduction. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*, pages 315–323. Curran Associates Inc., 2013.

[71] M. Jung, G. Chung, G. Sundaramoorthi, L. Vese, and A. Yuille. Sobolev gradients and joint variational image segmentation, denoising, and deblurring. In *Computational Imaging VII*, volume 7246, pages 131 – 143. SPIE, 2009.

[72] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001.

[73] A. Kaplan and R. Tichatschke. Proximal point methods and nonconvex optimization. *Journal of Global Optimization*, 13(4):389–406, 1998.

[74] S. Karczmarz. Angenaherte auflosung von systemen linearer glei-chungen. *Bull. Int. Acad. Pol. Sic. Let., Cl. Sci. Math. Nat.*, pages 355–357, 1937.

[75] A. Karpathy et al. Cs231n convolutional neural networks for visual recognition. *Neural Networks*, 1:1, 2016.

[76] S. Kassemeyer, J. Steinbrener, L. Lomb, E. Hartmann, A. Aquila, A. Barty, A. V. Martin, C. Y. Hampton, S. Bajt, M. Barthelmess, et al. Femtosecond free-electron

laser X-ray diffraction data sets for algorithm development. *Optics Express*, 20(4): 4149–4158, 2012.

[77] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[78] A. Krizhevsky. Learning multiple layers of features from tiny images, 2009.

[79] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[80] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[81] L. Lei, C. Ju, J. Chen, and M. Jordan. Nonconvex finite-sum optimization via SCSG methods. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 2345–2355. Curran Associates Inc., 2017.

[82] J. Lellmann, D. A. Lorenz, C. Schonlieb, and T. Valkonen. Imaging with kantorovich–rubinstein discrepancy. *SIAM Journal on Imaging Sciences*, 7(4):2833–2859, 2014.

[83] B. D. Levin, E. Padgett, C.-C. Chen, M. Scott, R. Xu, W. Theis, Y. Jiang, Y. Yang, C. Ophus, H. Zhang, et al. Nanomaterial datasets to advance tomography in scanning transmission electron microscopy. *Scientific Data*, 3:160041, 2016.

[84] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein. Visualizing the loss landscape of neural nets. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 6389–6399. Curran Associates Inc., 2018.

[85] J. S. Lim. *Two-dimensional signal and image processing*, pages 42–45. Prentice-Hall, Inc., 1990.

[86] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[87] Y. H. Lo. *Developments in Mesoscale Correlative Multimodal X-ray Microscopy.* PhD thesis, University of California, Los Angeles, 2019.

[88] D. Luke. Relaxed averaged alternating reflections for diffraction imaging. *Inverse problems*, 21(1):37–50, 2005.

[89] F. R. Maia. The Coherent X-ray Imaging Data Bank. *Nature Methods*, 9(9):854, 2012.

[90] S. Marchesini. Invited article: A unified evaluation of iterative projection algorithms for phase retrieval. *Review of Scientific Instruments*, 78(1):011301, 2007.

[91] A. V. Martin, F. Wang, N.-t. D. Loh, T. Ekeberg, F. R. Maia, M. Hantke, G. van der Schot, C. Y. Hampton, R. G. Sierra, A. Aquila, et al. Noise-robust coherent diffractive imaging with a single diffraction pattern. *Optics Express*, 20(15):16650–16661, 2012.

[92] D. N. Mastronarde. Dual-axis tomography: an approach with alignment methods that preserve resolution. *Journal of Structural Biology*, 120(3):343–352, 1997.

[93] J. Miao and D. Sayre. On possible extensions of X-ray crystallography through diffraction-pattern oversampling. *Acta Crystallographica Section A*, 56(6):596–605, 2000.

[94] J. Miao, D. Sayre, and H. N. Chapman. Phase retrieval from the magnitude of the fourier transforms of nonperiodic objects. *JOSA A*, 15(6):1662–1669, 1998.

[95] J. Miao, P. Charalambous, J. Kirz, and D. Sayre. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. *Nature*, 400(6742):342–344, 1999.

[96] J. Miao, K. O. Hodgson, T. Ishikawa, C. A. Larabell, M. A. LeGros, and Y. Nishino. Imaging whole escherichia coli bacteria by using single-particle X-ray diffraction. *Proceedings of the National Academy of Sciences*, 100(1):110–112, 2003.

[97] J. Miao, F. Förster, and O. Levi. Equally sloped tomography with oversampling reconstruction. *Physical Review B*, 72(5):52103, 2005.

[98] J. Miao, Y. Nishino, Y. Kohmura, B. Johnson, C. Song, S. H. Risbud, and T. Ishikawa. Quantitative image reconstruction of gan quantum dots from oversampled diffraction intensities alone. *Physical Review Letters*, 95(8):085503, 2005.

[99] J. Miao, C.-C. Chen, C. Song, Y. Nishino, Y. Kohmura, T. Ishikawa, D. Ramunno-Johnson, T.-K. Lee, and S. H. Risbud. Three-dimensional GaN-Ga2O3 core shell structure revealed by X-ray diffraction microscopy. *Physical Review Letters*, 97(21): 215503, 2006.

[100] J. Miao, T. Ishikawa, I. K. Robinson, and M. M. Murnane. Beyond crystallography: Diffractive imaging using coherent X-ray light sources. *Science*, 348(6234):530–535, 2015.

[101] J. Miao, P. Ercius, and S. J. L. Billinge. Atomic electron tomography: 3D structures without crystals. *Science*, 353(6306):aaf2157, 2016.

[102] R. P. Millane. Phase retrieval in crystallography and optics. *JOSA A*, 7(3):394–411, 1990.

[103] J.-J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299, 1965.

[104] Y. Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. akad. nauk Sssr*, 269:543–547, 1983.

[105] J. Newling and F. Fleuret. Nested mini-batch k-means. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, pages 1352–1360. Curran Associates, Inc., 2016.

[106] Y. Nishino, Y. Takahashi, N. Imamoto, T. Ishikawa, and K. Maeshima. Three-dimensional visualization of a human chromosome using coherent X-ray diffraction. *Physical Review Letters*, 102(1):18101, 2009.

[107] D. O'Connor and L. Vandenberghe. On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting. *Mathematical Programming*, pages 1–24, 2018.

[108] Y. Z. O'Connor and J. A. Fessler. Fourier-based forward and back-projectors in iterative fan-beam tomographic image reconstruction. *IEEE Transactions on Medical Imaging*, 25(5):582–589, 2006.

[109] W. Palenstijn, K. Batenburg, and J. Sijbers. Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs). *Journal of Structural Biology*, 176(2):250–253, 2011.

[110] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch, 2017.

[111] R. Phelps. *Convex Functions, Monotone Operators and Differentiability.* Springer, 1991.

[112] A. Pryor, Y. Yang, A. Rana, M. Gallagher-Jones, J. Zhou, Y. H. Lo, G. Melinte, W. Chiu, J. A. Rodriguez, and J. Miao. GENFIRE: A generalized Fourier iterative reconstruction algorithm for high-resolution 3D imaging. *Scientific Reports*, 7(1):1–12, 2017.

[113] N. Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151, 1999.

[114] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[115] J. Radon. On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging*, 5(4):170–176, 1986.

[116] J. Radon. 1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Classic papers in modern diagnostic radiology*, 5:21, 2005.

[117] K. S. Raines, S. Salha, R. L. Sandberg, H. Jiang, J. A. Rodríguez, B. P. Fahimian, H. C. Kapteyn, J. Du, and J. Miao. Three-dimensional structure determination from a single view. *Nature*, 463(7278):214–217, 2010.

[118] S. Reddi, S. Kale, and S. Kumar. On the convergence of adam and beyond. In *6th International Conference on Learning Representation*, 2018.

[119] H. Robinds and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–407, 1951.

[120] I. Robinson and R. Harder. Coherent X-ray diffraction imaging of strain at the nanoscale. *Nature Materials*, 8(4):291–298, 2009.

[121] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.

[122] R. T. Rockafellar. *Convex analysis*. Princeton university, 2015.

[123] J. Rodriguez, R. Xu, C. Chen, Y. Zou, and J. Miao. Oversampling smoothness (OSS): an effective algorithm for phase retrieval of noisy diffraction intensities. *Journal of Applied Crystallography*, 46:312–318, 2013.

[124] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[125] M. Scott, C.-C. Chen, M. Mecklenburg, C. Zhu, R. Xu, P. Ercius, U. Dahmen, B. Regan, and J. Miao. Electron tomography at 2.4-ångström resolution. *Nature*, 483(7390): 444–447, 2012.

[126] D. Sculley. Web-scale k-means clustering. In *Proceedings of the 19th International Conference on World wide web*, pages 1177–1178. Association for Computing Machinery, 2010.

[127] A. Senior, G. Heigold, M. Ranzato, and K. Yang. An empirical study of learning rates in deep neural networks for speech recognition. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6724–6728, 2013.

[128] O. Shamir and T. Zhang. Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. In *Proceedings of the 30th International Conference on Machine Learning - Volume 28*, page I–71–I–79. JMLR.org, 2013.

[129] A. Shapiro and Y. Wardi. Convergence analysis of gradient descent stochastic algorithms. *Journal of Optimization Theory and Applications*, 91(2):439–454, 1996.

[130] D. Shapiro, S. Roy, R. Celestre, W. Chao, D. Doering, M. Howells, S. Kevan, D. Kilcoyne, J. Kirz, S. Marchesini, et al. Development of coherent scattering and diffractive imaging and the COSMIC facility at the Advanced Light Source. In *Journal of Physics: Conference Series*, volume 425, page 192011. IOP Publishing, 2013.

[131] Y. Shechtman, Y. Eldar, O. Cohen, H. Chapman, J. Miao, and M. Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, 2015.

[132] D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 23rd ACM national conference*, pages 517–524. Association for Computing Machinery, 1968.

[133] C. Song, K. Tono, J. Park, T. Ebisu, S. Kim, H. Shimada, S. Kim, M. Gallagher-Jones, D. Nam, T. Sato, T. Togashi, K. Ogawa, Y. Joti, T. Kameshima, S. Ono, T. Hatsui, S. Iwata, M. Yabashi, and T. Ishikawa. Multiple application X-ray imaging chamber for single-shot diffraction experiments with femtosecond X-ray laser pulses. *Journal of Applied Crystallography*, 47(1):188–197, 2014.

[134] R. Sutton. Two problems with backpropagation and other steepest-descent learn-

ing procedures for networks. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pages 823–832, 1986.

[135] C. Tang and C. Monteleoni. Convergence rate of stochastic k-means. *arXiv preprint arXiv:1610.04900*, 2016.

[136] T. Tieleman and G. Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, 4(2):26–31, 2012.

[137] 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*, 157:35–47, 2015.

[138] W. Van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers. Fast and flexible X-ray tomography using the ASTRA Toolbox. *Optics Express*, 24(22):25129–25147, 2016.

[139] I. Waldspurger, A. d'Aspremont, and S. Mallat. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1-2):47–81, 2015.

[140] X. Wang, A. Sabne, S. Kisner, A. Raghunathan, C. Bouman, and S. Midkiff. High performance model based image reconstruction. *ACM SIGPLAN Notices*, 51(8):1–12, 2016.

[141] Z. Wen, C. Yang, X. Liu, and S. Marchesini. Alternating direction methods for classical and ptychographic phase retrieval. *Inverse Problems*, 28(11):115010, 2012.

[142] Y. Wu and K. He. Group normalization. *International Journal of Computer Vision*, 128:742–755, 2020.

[143] R. Xu, C.-C. Chen, L. Wu, M. Scott, W. Theis, C. Ophus, M. Bartels, Y. Yang, H. Ramezani-Dakhel, M. R. Sawaya, et al. Three-dimensional coordinates of individual atoms in materials revealed by electron tomography. *Nature Materials*, 14(11):1099–1103, 2015.

[144] M. Yan and W. Yin. Self equivalence of the alternating direction method of multipliers. *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 165–194, 2016.

[145] Y. Yang, C.-C. Chen, M. Scott, C. Ophus, R. Xu, A. Pryor, L. Wu, F. Sun, W. Theis, J. Zhou, et al. Deciphering chemical order/disorder and material properties at the single-atom level. *Nature*, 542(7639):75–79, 2017.

[146] P. Yin and J. Xin. Phaseliftoff: an accurate and stable phase retrieval method based on difference of trace and frobenius norms. *Communications in Mathematical Sciences*, 13:1033–1049, 2015.

[147] P. Yin, S. Zhang, J. Lyu, S. Osher, Y. Qi, and J. Xin. Binaryrelax: A relaxation approach for training deep neural networks with quantized weights. *SIAM Journal on Imaging Sciences*, 11(4):2205–2223, 2018.

[148] K. Yosida. *Functional analysis*. Springer, Berlin, 1964.

[149] Y. Yuan, D. S. Kim, J. Zhou, et al. Direct observation of 3D atomic packing in monatomic amorphous solids. Manuscript submitted for publication, 2020.

[150] M. Zeiler. Adadelta: An adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

[151] Y. Zhao, E. Brun, P. Coan, Z. Huang, A. Sztrókay, P. C. Diemoz, S. Liebhardt, A. Mittone, S. Gasilov, J. Miao, et al. High-resolution, low-dose phase contrast X-ray tomography for 3D diagnosis of human breast cancers. *Proceedings of the National Academy of Sciences*, 109(45):18290–18294, 2012.