

Computational methods for phase retrieval

Non-iterative methods, Ptychography, and Diffractive Shearing Interferometry

Konijnenberg, Sander

DOI

[10.4233/uuid:c8adfe08-43cd-4b8a-9436-54a9a56b4e14](https://doi.org/10.4233/uuid:c8adfe08-43cd-4b8a-9436-54a9a56b4e14)

Publication date

2019

Document Version

Final published version

Citation (APA)

Konijnenberg, S. (2019). *Computational methods for phase retrieval: Non-iterative methods, Ptychography, and Diffractive Shearing Interferometry*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:c8adfe08-43cd-4b8a-9436-54a9a56b4e14>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Computational methods for phase retrieval

Non-iterative methods, Ptychography, and
Diffractive Shearing Interferometry

Computational methods for phase retrieval

Non-iterative methods, Ptychography, and
Diffractive Shearing Interferometry

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 25 oktober 2019 om 10:00 uur

door

Alexander Prasetya KONIJNENBERG

Master of Science, Applied Physics
Technische Universiteit Delft, Delft, Nederland,
geboren te Maastricht, Nederland.

Dit proefschrift is goedgekeurd door de

promotor: prof. dr. W.M.J.M. Coene

promotor: prof. dr. H.P. Urbach

Samenstelling promotiecommissie:

Rector Magnificus,

Prof. dr. W.M.J.M. Coene

Prof. dr. H.P. Urbach

voorzitter

Technische Universiteit Delft

Technische Universiteit Delft

Onafhankelijke leden:

Prof. dr. ir. M. Verhaegen

Prof. dr. J.M. Rodenburg

Prof. dr. H.N. Chapman

Prof. dr. B. Rieger

Dr. S. Witte

Technische Universiteit Delft

University of Sheffield

University of Hamburg

Technische Universiteit Delft

Vrije Universiteit Amsterdam



Keywords: Phase retrieval, ptychography, computational imaging

Printed by: Ipskamp Printing

Front & Back: An engineer's brave attempt at graphic design

Copyright © 2019 by Alexander Prasetya KONIJNENBERG

ISBN 000-00-0000-000-0

An electronic version of this dissertation is available at

<http://repository.tudelft.nl/>.

You know that I write slowly. This is chiefly because I am never satisfied until I have said as much as possible in a few words, and writing briefly takes far more time than writing at length.

Carl Friedrich Gauss

Contents

Summary	xv
Samenvatting	xvii
1 Introduction	1
1.1 Motivation for phase retrieval	2
1.2 Phase retrieval from a single far field intensity pattern	3
1.2.1 Ambiguities and unicity	3
1.2.2 Sampling requirements	4
1.2.3 Coherence requirements	5
1.3 Non-iterative phase retrieval methods	6
1.4 Iterative phase retrieval methods	13
1.4.1 Ptychography	15
1.5 Experimental results.	17
1.6 Conclusion	18
References.	19
2 Non-iterative phase retrieval	27
2.1 Problem statement	28
2.2 Far field intensity and autocorrelation	28
2.3 Robustness to spatial partial coherence	28
2.4 Fourier transform holography	30
2.5 Non-iterative phase retrieval by varying a single parameter.	31
2.5.1 Intuitive line of reasoning.	32
2.5.2 Mathematical derivation.	34
2.5.3 Relation to phase shifting holography.	35
2.5.4 Simulation results	36
2.6 Non-iterative phase retrieval using a star-shaped mask and through-focus scanning.	38
2.6.1 Finding the correct surface of the autocorrelation function	40
2.6.2 Finding the requirements for the object support	41
2.6.3 Experimental results.	44
2.7 Summary	47
References.	48
3 Iterative phase retrieval	51
3.1 Phase retrieval methods using a single intensity pattern	52
3.1.1 Intersection of sets	53
3.1.2 Cost minimization	55

3.1.3	Convex relaxation and rank minimization	56
3.2	Ptychography	56
3.2.1	Basic reconstruction algorithms	57
3.2.2	Global and Sequential updates.	59
3.2.3	Combining HIO and ptychography	63
3.3	Summary	67
	References.	72
4	Noise robust ptychography	77
4.1	Maximum likelihood cost function.	78
4.2	Variance stabilizing transforms	80
4.3	Testing different cost functions for noise robustness.	81
4.4	Noise-robustness by adapting measurement constraints	83
4.4.1	Problems with the maximum likelihood scheme	83
4.4.2	An alternative noise robust scheme	84
4.4.3	Experimental results.	84
4.4.4	Simulations	87
4.5	Summary	96
	References.	98
5	Diffraction Shearing Interferometry	101
5.1	Intersection of sets: applying the constraints	103
5.1.1	Object constraint	104
5.1.2	Measurement constraint	104
5.1.3	The standard DSI algorithm	104
5.2	Alternative solutions	105
5.2.1	Support constraint requirements	107
5.2.2	Introducing synthetic constraints to eliminate alternative solutions	108
5.3	Experimental data	111
5.4	Summary	118
	References.	118
6	3D ptychography	119
6.1	Deriving the scattering integral	121
6.2	Approximating the 3D potential as 2D objects	122
6.2.1	Zeroth order approximation	123
6.2.2	First order approximation.	124
6.3	A real-space derivation	126
6.3.1	Paraxial Helmholtz equation	127
6.3.2	Single scattering approximation	128
6.3.3	Picard Iterations	129
6.3.4	Application to inverse problems: multislice method	131
6.4	Summary	133
	References.	133

7 Conclusion	135
7.1 Future work	136
References	136
Epilogue	139
A Wigner Distribution Deconvolution Method and the autocorrelation function	145
References	148
B Fresnel propagation of the focal field	151
C Reconstruction error and cost functional	153
D Measurement constraint projection for DSI	157
E Scattering in the paraxial approximation	159
References	161
Curriculum Vitæ	163
List of Publications	165

Glossary

- A Optical parameter that can be varied (e.g. defocus). [12](#)
- $B(\mathbf{x})$ HIO feedback function. [64](#)
- E Reconstruction error (object space). [61](#)
- $G(\mathbf{r})$ Green's function. [121](#)
- I Intensity. [6](#)
- $J(\mathbf{x}, \mathbf{y})$ Mutual intensity function. [28](#)
- L Cost functional. [55](#)
- $O(\mathbf{x})$ Object. [11](#)
- $P(\mathbf{x})$ Probe. [11](#)
- $W(\mathbf{x}, \mathbf{K})$ Wigner Distribution Function. [11](#)
- $\mathbf{X}$ 2D real space probe position vector. [11](#)
- β HIO feedback parameter. [14](#)
- $\mathbf{k}$ 2D Fourier space coordinate vector. [2](#)
- $\mathbf{r}$ 3D real space coordinate vector. [120](#)
- $\mathbf{x}$ 2D real space coordinate vector. [2](#)
- $\chi(\mathbf{r})$ Optical contrast function. [120](#)
- $\hat{H}(A)$ Sampling function. [32](#)
- $\hat{f}(\mathbf{k})$ Estimated far field (Fourier transform of estimated exit wave). [13](#)
- $\hat{\psi}(\mathbf{k})$ Far field (Fourier transform of exit wave). [2](#)
- $\mathcal{F}^{-1}$ Inverse Fourier transform. [11](#)
- $\mathcal{F}$ Fourier transform. [11](#)
- $\mathcal{P}$ Projection operator. [53](#)
- $\mathcal{R}$ Reflection operator. [54](#)
- μ Step size. [55](#)

$\psi(\mathbf{x})$ Exit wave (illumination field times reflection/transmission function). 2

$f(\mathbf{x})$ Estimated exit wave. 13

3DBPP 3D Bragg Projection Ptychography. 18

3PIE Three-Dimensional Ptychographic Iterative Engine. 16

AP Alternating Projections. 14

CDI Coherent Diffractive Imaging. 2

CG Conjugate Gradient. 16

CLP Convex Lifted Ptychography. 59

DBFH Double Blind Fourier Holography. 13

DM Difference Map. 14

DSI Diffractive Shearing Interferometry. 101, 102

ePIE Extended Ptychographic Iterative Engine. 16

ER Error Reduction. 13, 101

EUV Extreme Ultraviolet. 2, 102

FOV Field of View. 18

FTH Fourier Transform Holography. 10

FTS Fourier Transform Spectroscopy. 102

GS Gerchberg-Saxton. 13

HERALDO Holography with Extended Reference by Autocorrelation Linear Differential Operator. 13

HHG High Harmonic Generator. 102

HIO Hybrid Input-Output. 14, 101

IC Integrated Circuit. 135

LED Light Emitting Diode. 18

LRP Low Rank Ptychography. 59

- M-HIO** Modified Hybrid Input-Output. [15](#)
- MEP** Modulus Enforced Probe. [16](#)
- NA** Numerical Aperture. [16](#)
- OSS** Oversampling Smoothness. [15](#)
- PIE** Ptychographic Iterative Engine. [15](#)
- PXCT** Ptychographic X-ray Computed Tomography. [17](#)
- R-DSI** Rotational Diffractive Shearing Interferometry. [103](#)
- RAAR** Relaxed Averaged Alternating Projections. [14](#)
- SBP** Space-Bandwidth Product. [18](#)
- SDSI** Synthetic constraints Diffractive Shearing Interferometry. [111](#)
- SLM** Spatial Light Modulator. [32](#)
- SXR** Soft X-Ray. [2](#)
- TIE** Transport of Intensity Equation. [10](#)
- WDDM** Wigner Distribution Deconvolution Method. [11](#)
- WDF** Wigner Distribution Function. [11](#)

Summary

In this thesis, several phase retrieval methods are discussed. Since the focus will mainly be on theory rather than experiment, the structure has been determined by the similarities and differences of the mathematics of these methods. For example, a distinction is made between non-iterative and iterative methods, and between single-shot iterative phase retrieval and multiple-shot iterative phase retrieval (ptychography). However, it must be noted that phase retrieval methods that are mathematically similar, are suitable for fundamentally different experimental setups. For example, one can consider setups for **lensless imaging**, of which an interesting application is metrology using Extreme Ultraviolet (EUV) radiation. In such setups, no focusing optics are used, and one typically computes an image from far-field intensity patterns. On the other hand, there are setups for **aberrated imaging**. In these setups, one does use focusing optics to form images, but by introducing some sort of variations or perturbations, one can generate a set of images from which a complex-valued field can be computed. For example, regular ptychography and Fourier ptychography are mathematically the same, but the former is used for lensless imaging, while the latter is used for aberrated imaging. Mathematically, the only difference between these two ptychographic approaches is that the roles of object space and Fourier space are interchanged.

In Chapter 1 we give a broad overview of the different phase retrieval methods that exist. We discuss several non-iterative methods and iterative methods, and explain how they are related to each other. This provides the context in which the contents of the subsequent chapters can be placed.

In Chapter 2 we discuss a new non-iterative phase retrieval method that can be understood using three-dimensional autocorrelation functions. First, we illustrate as an example how in Fourier Transform Holography (typically used for lensless imaging) one manipulates the sample in such a way that its autocorrelation function allows for a straightforward extraction of the reconstruction. Then, it is explained how one can obtain a three-dimensional data set by varying an optical parameter (which can in principle be used for both aberrated imaging and lensless imaging), and how the reconstruction can be extracted from the corresponding three-dimensional autocorrelation function.

In Chapter 3 the theory behind iterative phase retrieval algorithms is discussed in more detail. For single-shot Coherent Diffractive Imaging, it is explained how the problem can either be tackled using projections or cost-minimization schemes. Then, it is explained how these techniques can be applied to ptychographic phase retrieval, where one aims to reconstruct an image of a sample from multiple mea-

surements. The differences between sequential and global updates are explored, after which new variants of the ptychographic reconstruction algorithm are proposed and tested.

In Chapter 4 it is recalled how ptychography can be modeled as a cost-minimization problem, after which we explore how one can choose the cost functional that is to be minimized. We review how the cost functional can be chosen depending on the assumed noise model. We explain the maximum-likelihood approach, the variance stabilization approach, and how they are related. Then, we test different cost-functionals for their robustness to noise, and conclude that the maximum-likelihood approach does not necessarily yield optimal results. We propose and test through simulations and experiment a new method that aims to improve the noise-robustness by adapting the measurement constraints.

In Chapter 5 we discuss a relatively new phase retrieval problem called Diffractive Shearing Interferometry (DSI). In this problem, the measurement constraint differs from the familiar far-field intensity constraint, and therefore new reconstruction algorithms must be developed. We analyze the already existing DSI reconstruction algorithm and propose a novel algorithm using the theory that was laid out in Chapter 3.

In Chapter 6 a new approach for calculating the exit wave beyond the multiplicative approximation is presented, which may help in developing a ptychographic reconstruction algorithm for samples whose thickness are not negligible. It is shown how the multiplicative approximation (used in regular ptychography) can be derived by applying a zeroth-order approximation to the scattering integral, and how one can derive a more accurate first-order approximation. This improvement is an area for future research.

Samenvatting

In dit proefschrift worden verscheidene faseverkrijgingsmethodes besproken. Aangezien het meer over theorie dan experiment zal gaan, is de structuur bepaald door de overeenkomsten en verschillen van de wiskunde achter deze methodes. Er wordt bijvoorbeeld onderscheid gemaakt tussen niet-iteratieve en iteratieve methodes, en tussen methodes die een enkel intensiteitspatroon gebruiken of meerdere (zoals ptychografie). Het moet echter opgemerkt worden dat faseverkrijgingsmethodes die in wiskundig opzicht op elkaar lijken toepasbaar zijn op fundamenteel verschillende experimentele opstellingen. Men kan bijvoorbeeld opstellingen voor **afbeelden zonder lenzen** beschouwen, waarvan metrologie met Extreme Ultraviolet (EUV) straling een interessante toepassing is. In dergelijke opstellingen wordt geen focuserende optiek gebruikt, en men berekent gebruikelijk een afbeelding uit verre veld intensiteitspatronen. Anderzijds zijn er opstelling voor **geaberreerd afbeelden**. In deze opstellingen gebruikt men focuserende optiek om afbeeldingen te vormen, maar door een bepaald soort variatie of perturbatie te introduceren kan men een set afbeeldingen genereren waaruit een complexwaardig veld berekend kan worden. Normale ptychografie en Fourier ptychografie zijn bijvoorbeeld wiskundig hetzelfde, maar de eerste wordt voor afbeelden zonder lenzen gebruikt, terwijl de tweede wordt gebruikt voor geaberreerd afbeelden. Wiskundig gezien is het enige verschil tussen deze twee ptychografische aanpakken dat de objectruimte en Fourier-ruimte zijn uitgewisseld.

In hoofdstuk 1 wordt een breed overzicht gegeven van bestaande faseverkrijgingsmethodes. We behandelen enkele niet-iteratieve en iteratieve methodes, en leggen uit hoe ze aan elkaar gerelateerd zijn. Dit geeft ons de context waarin de inhoud van de navolgende hoofdstukken geplaatst kan worden.

In hoofdstuk 2 behandelen we een niet-iteratieve faseverkrijgingsmethode die begrepen kan worden met behulp van driedimensionale autocorrelatiefuncties. Eerst gebruiken we Fourier Holografie (wat gebruikt wordt voor afbeelden zonder lenzen) als voorbeeld om te laten zien hoe een monster bewerkt kan worden zodat uit diens autocorrelatiefunctie direct een gereconstrueerd beeld gehaald kan worden. Vervolgens wordt uitgelegd hoe een driedimensionale dataset verkregen kan worden door een optische parameter te variëren (wat in principe gebruikt kan worden voor zowel geaberreerd afbeelden als afbeelden zonder lenzen), en hoe uit de bijbehorende autocorrelatiefunctie de reconstructie gehaald kan worden.

In hoofdstuk 3 wordt de theorie achter faseverkrijgingsalgoritmes in meer detail besproken. Voor het geval van Afbeelden middels Coherente Diffractie (Coherent Diffractive Imaging, CDI) met een enkele meting wordt uitgelegd hoe het probleem

aangepakt kan worden met projecties of met het minimaliseren van een kostenfunctionaal. Daarna wordt uitgelegd hoe deze technieken toegepast kunnen worden op ptychografische faseverkrijging, waarbij men probeert een afbeelding van een monster te reconstrueren uit meerdere metingen. De verschillen tussen sequentiële en globale bijwerkingen worden beschouwd, waarna nieuwe varianten van het ptychografische reconstructiealgoritme worden voorgesteld en getest.

In hoofdstuk 4 wordt herhaald hoe ptychografie beschouwd kan worden als een kostenminimalisatieprobleem, waarna we kijken hoe de te minimaliseren kostenfunctionaal gekozen kan worden. We beschouwen de manieren waarop de kostenfunctionaal gekozen kan worden op basis van het aangenomen ruismodel. We leggen de aanpak van de meest aannemelijke schatter uit, de aanpak van variantiestabilisatie, en hoe deze methodes aan elkaar gerelateerd zijn. Vervolgens testen we verschillende kostenfunctionalen om te kijken hoe ruisbestendig ze zijn, waarna we tot de conclusie komen dat de aanpak van de meest aannemelijke schatter niet per se de beste resultaten oplevert. We stellen een nieuwe methode voor die meer ruisbestendig is doordat het de metingen (die als randvoorwaarden in het algoritme worden gebruikt) aanpast, en we testen deze methode met simulaties en experiment.

In hoofdstuk 5 behandelen we een vrij nieuw faseverkrijgingsprobleem genaamd Diffractieve Shearing Interferometrie (DSI). In dit probleem verschilt de meting die in het algoritme wordt gebruikt van de gebruikelijke intensiteitsmeting, waardoor nieuwe reconstructiealgoritmes ontwikkeld moeten worden. We analyseren het al bestaande DSI reconstructiealgoritme en ontwikkelen een nieuw algoritme, gebruikmakend van de theorie die eerder was behandeld in hoofdstuk 3.

In hoofdstuk 6 wordt een nieuwe methode gepresenteerd voor het berekenen van het getransmitteerde veld dat verder gaat dan de multiplicatieve benadering, wat kan helpen in het ontwikkelen van een ptychografisch reconstructiealgoritme voor monsters met diktes die niet verwaarloosbaar zijn. Het wordt aangetoond dat de multiplicatieve benadering (die gebruikt wordt in de gangbare ptychografie) afgeleid kan worden door een nulde-orde benadering toe te passen op de verstrooiingsintegraal, en hoe een nauwkeurigere benadering afgeleid kan worden door een eerste-orde benadering toe te passen. Deze methode is een onderwerp voor toekomstig onderzoek.

1

Introduction

*Any intelligent fool can make things bigger, more complex, and more violent.
It takes a touch of genius — and a lot of courage to move in the opposite
direction.*

E.F. Schumacher

1.1. Motivation for phase retrieval

A monochromatic scalar optical field in an (x, y) -plane may be described by the complex-valued function $\psi(\mathbf{x})$, where $\mathbf{x} = [x \ y]^T$. However, one can only directly measure its intensity $|\psi(\mathbf{x})|^2$, while its phase information is lost. In phase retrieval, one aims to find the phase of the field from intensity measurements. There are several reasons why one may want to do this:

- **Retrieving aberrations:** wave front aberrations may degrade the quality of an imaging system. One can correct for this using adaptive optics, but in order to correct for the aberrations, one must first characterize them. One way to do this is by using a Shack-Hartmann wave front sensor. By looking at a set of displaced focal spots generated by a lenslet array, one can infer the gradient of the wave front, which one can integrate to find the phase of the field.
- **Aberration correction:** one can compensate for aberrations in the experimental setup using adaptive optics as mentioned previously, but one can also measure an aberrated image and correct for the aberrations afterwards. In that case, in addition to characterizing the aberrations, one must also retrieve the phase of the field in the image plane so that the complex-valued field can be deconvolved using the imaging point-spread function.
- **Imaging phase samples:** highly transparent samples (such as weakly scattering biological samples) tend to only affect the phase of the field while leaving the amplitude unchanged, i.e. their transmission function is of the form $T(\mathbf{x}) = e^{i\phi(\mathbf{x})}$ (which is known as a phase grating). If one were to image $T(\mathbf{x})$ directly, one would only measure $|T(\mathbf{x})|^2 = 1$, while the relevant phase information $\phi(\mathbf{x})$ is lost entirely. One way to measure information about $\phi(\mathbf{x})$ is by using Zernike phase-contrast microscopy, where one phase shifts the unscattered light before forming the image.
- **(Back-)propagating fields:** if one wants to propagate from one plane to another, one needs to know both the amplitude and phase of the field. If one measures a far field intensity $|\hat{\psi}(\mathbf{k})|^2$ (where $\hat{\psi}(\mathbf{k})$ denotes the spatial Fourier transform of $\psi(\mathbf{x})$, and $\mathbf{k}$ denotes the 2D Fourier space coordinate vector), then one requires a method to find the far field phase in order to back-propagate $\hat{\psi}(\mathbf{k})$ to find the near field $\psi(\mathbf{x})$. This is the application that is central to this thesis, and it is commonly known as Coherent Diffractive Imaging (CDI). It is especially useful for wavelengths for which no high-quality focusing optics are available (such as the Extreme Ultraviolet, or EUV, regime, also often referred to as the Soft X-Ray, or SXR, regime), which makes it difficult to form direct images.

A more elaborate list of the possible applications of phase retrieval is given in [1].

1.2. Phase retrieval from a single far field intensity pattern

Given is a scalar monochromatic field $\psi(\mathbf{x})$ of wavelength λ that is restricted by an aperture with radius R . For example, $\psi(\mathbf{x})$ can be the field transmitted by a sample of size R . If we let the field propagate a distance z such that

$$\frac{R^2}{\lambda z} < 1, \quad (1.1)$$

(i.e. if we consider Fraunhofer propagation) then we can calculate the far-field as

$$\psi_z(x, y) = \iint \psi(x', y') e^{-2\pi i(x' \frac{x}{\lambda z} + y' \frac{y}{\lambda z})} dx' dy', \quad (1.2)$$

where we ignored an irrelevant multiplicative constant and a quadratic phase factor which is irrelevant since we only measure intensities. If we write the spatial Fourier transform as

$$\hat{\psi}(\mathbf{k}) = \int \psi(\mathbf{x}) e^{-2\pi i \mathbf{x} \cdot \mathbf{k}} d\mathbf{x}, \quad (1.3)$$

then the far field of $\psi(\mathbf{x})$ is given by $\hat{\psi}(\mathbf{k})$, with $\mathbf{k} = \left[\frac{x}{\lambda z} \quad \frac{y}{\lambda z} \right]^T$. The question is how to retrieve $\psi(\mathbf{x})$ from $|\hat{\psi}(\mathbf{k})|^2$.

1.2.1. Ambiguities and unicity

Let us assume we know a priori that $\psi(\mathbf{x})$ is restricted by a radius R (i.e. we assume we have a support constraint). Is $\psi(\mathbf{x})$ uniquely defined if we know (approximately) its support and $|\hat{\psi}(\mathbf{k})|$? There are three obvious ambiguities:

- **Global phase shift:** if $\psi(\mathbf{x})$ is a solution, then so is $\psi(\mathbf{x})e^{i\theta}$, because then $\hat{\psi}(\mathbf{k})$ is multiplied by a constant phase factor, which does not affect $|\hat{\psi}(\mathbf{k})|$.
- **Translation:** if $\psi(\mathbf{x})$ is a solution, then so is $\psi(\mathbf{x} + \mathbf{x}_0)$ (provided it does not violate the support constraint), because then $\hat{\psi}(\mathbf{k})$ is multiplied by a linear phase function $e^{2\pi i \mathbf{x}_0 \cdot \mathbf{k}}$ which does not affect $|\hat{\psi}(\mathbf{k})|$.
- **Twin image:** if $\psi(\mathbf{x})$ is a solution, then so is its twin image $\psi(-\mathbf{x})^*$ (provided it does not violate the support constraint), because then $\hat{\psi}(\mathbf{k})$ is conjugated (i.e. it becomes $\hat{\psi}(\mathbf{k})^*$), which does not affect $|\hat{\psi}(\mathbf{k})|$.

These are trivial ambiguities since they do not fundamentally affect the relevant features of $\psi(\mathbf{x})$. So aside from these ambiguities, is $\psi(\mathbf{x})$ uniquely defined by $|\hat{\psi}(\mathbf{k})|$ and a support constraint? While this is not true in the 1D case, it has been shown that in the discrete 2D case the solution is unique, except for a set of objects $\psi(\mathbf{x})$ whose Z-transforms are reducible (meaning that the Z-transform can be written as the product of two polynomials), which is a very uncommon property for real-life samples [2–4]. However, it has also been demonstrated that the presence of noise and the existence of near-alternative solutions (i.e. solutions that almost but not

exactly match the measurement constraints) can cause significant problems for the convergence of phase retrieval algorithms [5, 6]. It has been proposed that using randomly phased illumination would eliminate alternative solutions and considerably improve convergence [7, 8]. Moreover, using such illumination would result in less sharply peaked diffraction patterns, which eliminates the problem of dynamic range: using a beam stop or stitching together recordings of different exposure times would become unnecessary [9].

1.2.2. Sampling requirements

In practice, we cannot measure $|\hat{\psi}(\mathbf{k})|^2$ for all $\mathbf{k}$, but rather we sample $\mathbf{k}$ on a discrete grid defined by the pixels of the detector. How finely should we sample $|\hat{\psi}(\mathbf{k})|^2$ in order to be able to retrieve $\psi(\mathbf{x})$ from it? If $\psi(\mathbf{x})$ is restricted to a square aperture of dimension R , then it is sufficient to sample $\hat{\psi}(\mathbf{k})$ on a grid with interval $1/R$, as prescribed by the Nyquist criterion. However, we are not sampling $\hat{\psi}(\mathbf{k})$, but $|\hat{\psi}(\mathbf{k})|^2$, which when inverse Fourier transformed gives the autocorrelation of $\psi(\mathbf{x})$, which has a support of dimension $2R$. Therefore, as was remarked by Sayre in 1952 [10], one should sample the intensity at the Nyquist interval of $|\hat{\psi}(\mathbf{k})|^2$, which is half the Nyquist interval of $\hat{\psi}(\mathbf{k})$, namely $\frac{1}{2R}$, to recover the autocorrelation of $\psi(\mathbf{x})$ without aliasing, which should give the information to recover $\psi(\mathbf{x})$. Given that $\Delta k \leq 1/2R$, and that $k = x/\lambda z$ (see Eq. (1.2)), where x is the spatial coordinate on the detector, we find for the required pixel spacing

$$\Delta x \leq \frac{\lambda z}{2R}. \quad (1.4)$$

We can also look at the issue by using the Discrete Fourier Transform (DFT). The DFT of a discrete 1D array ψ_n with N entries is given by

$$\hat{\psi}_m = \sum_{n=0}^{N-1} \psi_n e^{-2\pi i m n / N}, \quad m = 0, 1, \dots, N-1. \quad (1.5)$$

In this expression, ψ_n and $\hat{\psi}_m$ are just arrays of numbers, unrelated to any physical axes. We can write

$$\begin{aligned} \hat{\psi}_m &= \sum_{n=0}^{N-1} \psi_n e^{-2\pi i \frac{m}{N\Delta x} n\Delta x} \\ &= \sum_{n=0}^{N-1} \psi_n e^{-2\pi i k_m x_n}, \end{aligned} \quad (1.6)$$

where we defined the axes in real and reciprocal space as

$$\begin{aligned} x_n &= n\Delta x, \quad n = 0, 1, \dots, N-1, \\ k_m &= \frac{m}{N\Delta x}, \quad m = 0, 1, \dots, N-1. \end{aligned} \quad (1.7)$$

So we see that:

- If we have a high resolution in x -space (i.e. Δ_x is small), we have a large field of view in k -space (i.e. the maximum k_m is large).
- If we have a large field of view in x -space (i.e. $N\Delta_x$ is large), we have a high resolution in k -space (i.e. $\Delta_k = 1/N\Delta_x$ is small).

So we see once more that the sampling Δ_k in k -space should be sufficiently fine so that the support of the object (or its autocorrelation) fits in the field of view in x -space defined by Δ_k . Moreover, we see that it is convenient to reconstruct a sample $\psi(x)$ from its far-field diffraction pattern, because for large enough distances z the resolution of the reconstruction does not depend on the pixel size of the detector (which is difficult to make smaller), but on the numerical aperture (NA) of the detector: if higher spatial frequencies are captured, the resolution of the reconstruction increases. The farther the field propagates, the more it expands, so the larger the pixels and their spacing may be, as is also seen in Eq. (1.4).

In 1998, Miao et al. [11] presented the following argument with regard to the required oversampling: if we measure

$$|\hat{\psi}_m|^2 = \left| \sum_{n=0}^{N-1} \psi_n e^{-2\pi i m n / N} \right|^2, \quad m = 0, 1, \dots, M-1, \quad (1.8)$$

then we have a set of M equations with $2N$ unknowns: each ψ_n has a real and imaginary part (or there are $2N_1$ unknowns if the global phase is irrelevant). If we sample according to the Nyquist criterion of $\hat{\psi}(k)$ (as opposed to $|\hat{\psi}(bk)|^2$), we have $M = N$ equations, which means that the system of equations is underdetermined by a factor of 2. Note that this is the case regardless of the number of dimensions: for a 1D object we have N equations and $2N$ unknowns, for a 2D object we have N^2 equations and $2N^2$ unknowns, etc. This is a departure from the autocorrelation argument, from which would follow that for a 2D object we should oversample by a factor of 2 in each dimension, giving $4N^2$ equations.

1.2.3. Coherence requirements

Given a maximum allowed distance between adjacent pixels on a detector, one can derive requirements for the spatial and temporal coherence in order for reconstruction algorithms that assume full coherence to work [12].

Spatial coherence

Consider an extended spatially incoherent quasi-monochromatic source. If the source is sufficiently far removed from the sample, each point of the source generates a plane wave with a certain angle in the sample plane. In the Fraunhofer approximation, illuminating the sample at a different angle yields an identical but translated far-field diffraction intensity pattern. To be able to use reconstruction algorithms that assume full coherence, we require that the translation of the diffraction patterns is smaller than the detector pixel spacing, which is given by Eq. (1.4).

Therefore, if the extended source subtends an angle θ_c at the sample, and two adjacent pixels subtend an angle α , we require $\theta_c < \alpha$. See Fig. 1.1a.

Temporal coherence

Consider a normally incident field with wavelengths ranging from λ to $\lambda + \Delta\lambda$. We assume that the sample has a wavelength-independent transmission function $\psi(x)$ (at least within the relevant wavelength range). Each wavelength yields a scaled diffraction pattern

$$I_\lambda(x') = \left| \hat{\psi}\left(\frac{\theta}{\lambda}\right) \right|^2, \quad (1.9)$$

where $\theta = x'/z$ is the diffraction angle, and I denotes intensity. Let us say there is a maximum relevant diffraction angle $\theta_{\max}$, which may be defined by the detector NA or by the noise level. We can consider two detector points $x'_{\max} = z\theta_{\max}$ and $x'_{\max,\Delta} = z\theta_{\max} \frac{\lambda + \Delta\lambda}{\lambda}$ which for different wavelengths λ and $\lambda + \Delta\lambda$ correspond to the same intensity value in the diffraction pattern: $I_\lambda(x'_{\max}) = I_{\lambda + \Delta\lambda}(x'_{\max,\Delta})$. If we want to assume full coherence, we require that $x'_{\max}$ and $x'_{\max,\Delta}$ differ by less than the pixel spacing $\frac{\lambda z}{2R}$

$$z\theta_{\max} \frac{\lambda + \Delta\lambda}{\lambda} - z\theta_{\max} \leq \frac{\lambda z}{2R}. \quad (1.10)$$

Rearranging this expression and defining the coherence length as $L_c = \lambda^2/\Delta\lambda$, we find

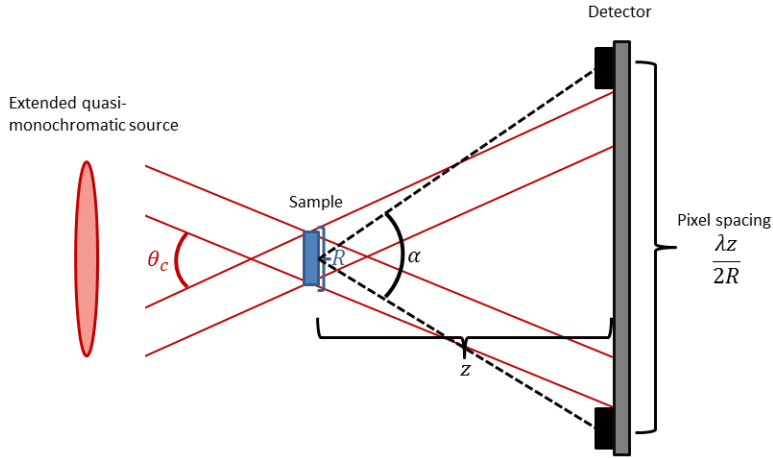
$$2\theta_{\max}R \leq L_c. \quad (1.11)$$

See Fig. 1.1b.

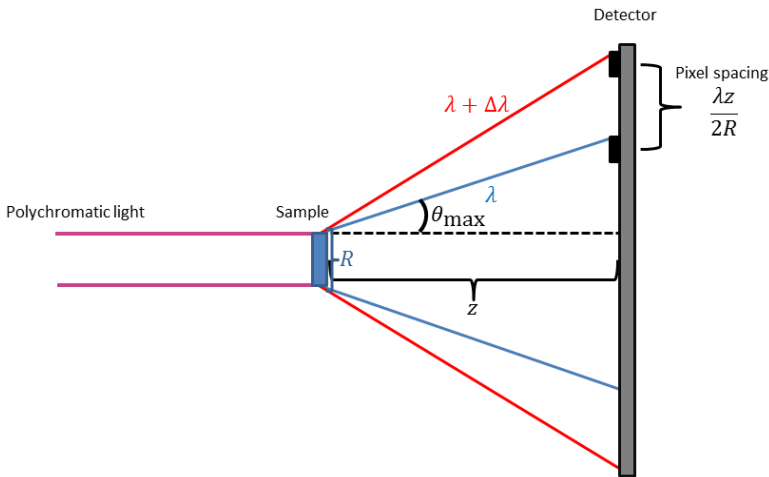
1.3. Non-iterative phase retrieval methods

- **Phase contrast microscopy:** In the early 1930's, Frits Zernike invented the phase contrast microscope for which he would later receive the Nobel Prize in 1953 [13, 14]. The purpose of such a microscope is to image weakly scattering pure phase objects, i.e. objects with transmission functions of the form $\psi(x) = e^{i\phi(x)}$ with small real-valued phase $\phi(x)$. If one would image such a sample directly, one would only measure the intensity $|\psi(x)|^2 = 1$, and none of the relevant features defined by the phase function $\phi(x)$ would be observed. They could be revealed in a slightly out-of-focus plane, but then the resolution would be poor. One common solution was to stain the sample, which converts the phase sample to an amplitude sample that can be imaged directly. Zernike's phase contrast method on the other hand could reveal $\phi(x)$ by phase-shifting the scattered field by 90 degrees with respect to the unscattered field (or vice versa) before the two fields interfere in the image plane. Mathematically it can be described as follows. If we assume $\phi(x)$ is so small that $\phi(x)^2$ is negligible, we can approximate the transmission function as

$$\psi(x) \approx 1 + i\phi(x). \quad (1.12)$$



(a) Spatial coherence



(b) Temporal coherence

Figure 1.1: Requirements for spatial and temporal coherence. The coherence should be such that the blur of the diffraction pattern is smaller than the pixel spacing as defined by the Nyquist criterion.

Here, the first term '1' corresponds to the unscattered light, and the second term ' $i\phi(\mathbf{x})$ ' corresponds (mostly) the scattered light. When the sample is illuminated by a normally incident plane wave, one would observe in the image plane

$$\begin{aligned} I(\mathbf{x}) &\approx |1 + i\phi(\mathbf{x})|^2 \\ &= 1 + |\phi(\mathbf{x})|^2. \end{aligned} \quad (1.13)$$

Because we assumed that $\phi(\mathbf{x})^2$ is negligible, no phase information is observed in the image (of course technically $I(\mathbf{x})$ is exactly equal to 1, and no approximation needs to be made in order to conclude no phase information is observed). In the phase contrast microscope on the other hand, we shift the unscattered light by 90 degrees (which can for example be done in the back focal plane of the lens where the scattered and unscattered light are spatially separated). We then observe in the image plane

$$\begin{aligned} I(\mathbf{x}) &= |i + i\phi(\mathbf{x})|^2 \\ &= 1 + \phi(\mathbf{x})^2 + 2\phi(\mathbf{x}). \end{aligned} \quad (1.14)$$

The term $2\phi(\mathbf{x})$ is not negligible compared to 1, which demonstrates that phase information is observed in the image. A more exact treatment will be discussed in the next point on quantitative phase contrast microscopy.

- **Quantitative phase contrast microscopy:** With the advent of digital cameras, one can perform quantitative measurements that can be used to computationally reconstruct a sample. Thus, one can extend Zernike's phase contrast method by applying multiple phase shifts θ , and computing the transmission function from the corresponding measurements $I_\theta(\mathbf{x})$ [15]. Given is an arbitrary transmission function $\psi(\mathbf{x})$, with spatial Fourier transform $\hat{\psi}(\mathbf{k})$. We can phase-shift the unscattered light (given by $\hat{\psi}(0)$) by θ

$$\hat{\psi}(\mathbf{k}) \rightarrow \hat{\psi}(\mathbf{k}) + \delta(\mathbf{k})\hat{\psi}(0)(e^{i\theta} - 1). \quad (1.15)$$

In the image plane we then measure

$$\begin{aligned} I_\theta(\mathbf{x}) &= |\psi(\mathbf{x}) + \hat{\psi}(0)(e^{i\theta} - 1)|^2 \\ &= |\psi(\mathbf{x})|^2 + 2\text{Re}\{\psi(\mathbf{x})\hat{\psi}(0)^*(e^{-i\theta} - 1)\} + |\hat{\psi}(0)(e^{i\theta} - 1)|^2. \end{aligned} \quad (1.16)$$

For $\psi(\mathbf{x}) = e^{i\phi(\mathbf{x})}$ and $\theta = \pi/2$ one can obtain the more exact expression for the measured intensity for Zernike phase contrast microscopy (compare with Eq. (1.14)). Without loss of generality we can assume that $\hat{\psi}(0)$ is real-valued, which allows us to write

$$I_\theta(\mathbf{x}) = f_1(\mathbf{x}) + c_{2,\theta}f_2(\mathbf{x}) + c_{3,\theta}f_3(\mathbf{x}), \quad (1.17)$$

where

$$\begin{aligned} f_1(\mathbf{x}) &= \text{Re}\{\psi(\mathbf{x})\}^2 + \text{Im}\{\psi(\mathbf{x})\}^2, \\ f_2(\mathbf{x}) &= \hat{\psi}(0)\text{Re}\{\psi(\mathbf{x})\} - \hat{\psi}(0)^2 \\ f_3(\mathbf{x}) &= \hat{\psi}(0)\text{Im}\{\psi(\mathbf{x})\}, \\ c_{2,\theta} &= 2(\cos \theta - 1), \quad c_{3,\theta} = 2 \sin \theta. \end{aligned} \quad (1.18)$$

By choosing three θ one constructs a system of three equations with which one can solve for the three unknowns $f_{1,2,3}(\mathbf{x})$. One can choose the values of θ such that the inversion of the system is most stable (i.e. the matrix of coefficients defined by $c_{2,3}$ has the highest possible determinant) and therefore is most robust against noise. Numerical calculations suggest the optimal values are $\theta = 0, 2\pi/3, 4\pi/3$, which makes sense intuitively. Given $f_{1,2,3}(\mathbf{x})$, one can solve for $\hat{\psi}(0)^2$ using the equation

$$\hat{\psi}(0)^4 + [2f_2(\mathbf{x}) - f_1(\mathbf{x})] \hat{\psi}(0)^2 + f_2(\mathbf{x})^2 + f_3(\mathbf{x})^2 = 0, \quad (1.19)$$

after which one can solve for $\text{Re}\{\psi(\mathbf{x})\}$ and $\text{Im}\{\psi(\mathbf{x})\}$.

Note that in principle this scheme is not limited to a conventional imaging setup, but it can be applied to a CDI setup as well. If we interpret $\hat{\psi}(\mathbf{k})$ as the sample's transmission function, and let $I_\theta(\mathbf{x})$ denote the far-field intensity measurements, the same scheme can be performed to find $\hat{\psi}(\mathbf{k})$.

- **Holography and interferometry:** In interferometry, one lets a field of interest $\psi(\mathbf{x})$ interfere with a reference field $\psi_0(\mathbf{x})$, and record the intensity of the sum

$$\begin{aligned} I(\mathbf{x}) &= |\psi(\mathbf{x}) + \psi_0(\mathbf{x})|^2 \\ &= |\psi(\mathbf{x})|^2 + |\psi_0(\mathbf{x})|^2 + 2\text{Re}\{\psi(\mathbf{x})^* \psi_0(\mathbf{x})\}. \end{aligned} \quad (1.20)$$

The interference term $\text{Re}\{\psi(\mathbf{x})^* \psi_0(\mathbf{x})\}$ gives information about the phase of $\psi(\mathbf{x})$ which would be inaccessible if we measure $|\psi(\mathbf{x})|^2$ directly. The reference field should be a simple and known field, typically a plane wave. If we choose the reference field to be an on-axis plane wave to which we can introduce phase shifts θ , i.e. $\psi_0(\mathbf{x}) = e^{i\theta}$, then Eq. (1.20) becomes

$$I_\theta(\mathbf{x}) = |\psi(\mathbf{x})|^2 + 1 + \text{Re}\{\psi(\mathbf{x})\} \cos \theta + \text{Im}\{\psi(\mathbf{x})\} \sin \theta. \quad (1.21)$$

By choosing three phase shifts θ , one can obtain three equations that are linear in the three unknowns $|\psi(\mathbf{x})|^2$, $\text{Re}\{\psi(\mathbf{x})\}$, and $\text{Im}\{\psi(\mathbf{x})\}$ (though these unknowns are obviously not independent), and solve for them. This is known as **phase-shifting interferometry**, and is very similar to the quantitative phase contrast scheme explained previously. Indeed, if we modulate the phase in a single point $\hat{\psi}(0)$ and then apply a Fourier transform to go to the image plane, we are basically shifting the phase of a reference plane wave that interferes with another field.

Another method is **off-axis holography**. If we choose the reference wave to be an off-axis plane wave $\psi_0(\mathbf{x}) = e^{i2\pi\mathbf{q}\cdot\mathbf{x}}$, then Eq. (1.20) becomes

$$I(\mathbf{x}) = |\psi(\mathbf{x})|^2 + 1 + \psi(\mathbf{x})e^{-i2\pi\mathbf{q}\cdot\mathbf{x}} + \psi(\mathbf{x})^*e^{i2\pi\mathbf{q}\cdot\mathbf{x}}. \quad (1.22)$$

If one Fourier transforms this intensity pattern, one identifies three terms:

1. The Fourier transform of $|\psi(\mathbf{x})|^2 + 1$, which is called the central band of the hologram, and tends to be fairly localized since $|\psi(\mathbf{x})|^2 + 1$ has no phase variation.
2. The Fourier transform of $\psi(\mathbf{x})e^{-i2\pi\mathbf{q}\cdot\mathbf{x}}$, which gives the sideband $\hat{\psi}(\mathbf{k} + \mathbf{q})$. This is the Fourier transform of $\psi(\mathbf{x})$, $\hat{\psi}(\mathbf{k})$, shifted by $-\mathbf{q}$. If this term is isolated, shifted back to the origin, and inverse Fourier transformed, one finds $\psi(\mathbf{x})$.
3. The Fourier transform of $\psi(\mathbf{x})^*e^{i2\pi\mathbf{q}\cdot\mathbf{x}}$, which gives the sideband $\hat{\psi}^*(-(\mathbf{k} - \mathbf{q}))$. This is the twin image of $\hat{\psi}(\mathbf{k})$ shifted by $\mathbf{q}$. If this term is isolated, shifted back to the origin, and inverse Fourier transformed, one finds $\psi(\mathbf{x})^*$. However, note that one does not gain any extra information or achieve a higher signal to noise ratio by using both sidebands as opposed to using only one.

We previously found that phase-shifting interferometry is analogous to quantitative phase contrast imaging, except that in the latter the reference plane wave is generated by modulating the field in one point in a Fourier plane. Similarly, off-axis holography is analogous to **Fourier Transform Holography (FTH)**, except that in FTH the off-axis reference plane wave is introduced by introducing a point source in a Fourier plane. In FTH, one has a sample $\psi(\mathbf{x})$ with a finite size. One can introduce a pinhole sufficiently far removed from the sample, which gives the transmission function $\psi(\mathbf{x}) + \delta(\mathbf{x} + \mathbf{q})$. One can measure the far-field intensity

$$I(\mathbf{k}) = |\hat{\psi}(\mathbf{k}) + e^{2\pi i\mathbf{q}\cdot\mathbf{k}}|^2, \quad (1.23)$$

which will give the same expression as Eq. (1.22), but this time in Fourier space. Thus, one can reconstruct $\psi(\mathbf{x})$ by measuring the far field intensity pattern, inverse Fourier transforming it, and isolating one of the two sidebands which correspond to $\psi(\mathbf{x})$ and its twin image $\psi(-\mathbf{x})^*$.

- **Transport of Intensity Equation (TIE)**: A monochromatic optical field $\psi(\mathbf{x})$ in vacuum satisfies the Helmholtz equation

$$(\nabla^2 + k^2)\psi(\mathbf{x}) = 0, \quad (1.24)$$

where ∇^2 denotes the Laplacian, and k is the wave number $\frac{2\pi}{\lambda}$, where λ is the wavelength. In the paraxial approximation with the z -axis as the optical axis, one can derive the paraxial Helmholtz equation

$$\nabla_{\perp}^2\psi(\mathbf{x}) + 2ik\frac{\partial\psi(\mathbf{x})}{\partial z} + 2k^2\psi(\mathbf{x}) = 0, \quad (1.25)$$

where

$$\nabla_{\perp}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \quad (1.26)$$

Solving the paraxial Helmholtz equation gives the Fresnel diffraction integral. If one writes the field in terms of intensity $I(\mathbf{x})$ and phase $\phi(\mathbf{x})$

$$\psi(\mathbf{x}) = \sqrt{I(\mathbf{x})} e^{i\phi(\mathbf{x})}, \quad (1.27)$$

then one can derive the Transport of Intensity Equation [16]

$$k \frac{\partial}{\partial z} I(\mathbf{x}) = -\nabla \cdot [I(\mathbf{x}) \nabla \phi(\mathbf{x})]. \quad (1.28)$$

By measuring the intensity in two planes z and $z + \Delta z$, one can approximate $\frac{\partial}{\partial z} I(\mathbf{x})$, and solve the differential equation for the phase $\phi(\mathbf{x})$. Uniqueness theorems have been proven for this problem, but they break down when zeros are present in the intensity distributions [17]. The effects of noise, the defocus distance, and using multiple defocus planes have been investigated in [18]. Instead of measuring the intensity at different planes that are related by free-space propagation, one can also measure the far field intensity while varying the phase curvature of the illumination that is incident on the sample [19, 20].

- **Wigner distribution deconvolution method (WDDM)**: Consider an object $O(\mathbf{x})$ that is illuminated by a probe function $P(\mathbf{x})$. One can shift the probe to different positions $\mathbf{X}$, and for each $\mathbf{X}$ one can record the far field intensity pattern¹

$$I(\mathbf{k}, \mathbf{X}) = |\mathcal{F} \{O(\mathbf{x})P(\mathbf{x} - \mathbf{X})\}(\mathbf{k})|^2, \quad (1.29)$$

where $\mathcal{F}$ denotes the forward Fourier transform. The four-dimensional data set $I(\mathbf{k}, \mathbf{X})$ can be inverse Fourier transformed with respect to $\mathbf{k}$, and Fourier transformed with respect to $\mathbf{X}$. Let us use $\mathbf{K}$ to denote the reciprocal space coordinate of $\mathbf{X}$. It can be shown that the result is the product of the Wigner distribution functions (WDFs, denoted $W(\mathbf{x}, \mathbf{K})$) of the object and the probe [21, 22]

$$\mathcal{F}_{\mathbf{k}}^{-1} \{ \mathcal{F}_{\mathbf{X}} \{ I(\mathbf{k}, \mathbf{X}) \} \}(\mathbf{x}, \mathbf{K}) = W_P(\mathbf{x}, -\mathbf{K}) W_O(\mathbf{x}, \mathbf{K}), \quad (1.30)$$

where

$$W_f(\mathbf{x}, \mathbf{K}) = \int f(\mathbf{y} + \mathbf{x}) f(\mathbf{y})^* e^{2\pi i \mathbf{y} \cdot \mathbf{K}} d\mathbf{y}, \quad (1.31)$$

and $\mathcal{F}^{-1}$ denotes the inverse Fourier transform. Assuming the illumination function $P(\mathbf{x})$ is known, we can divide out $W_P(\mathbf{x}, -\mathbf{K})$ to find $W_O(\mathbf{x}, \mathbf{K})$, which

¹Note that in Refs. [21, 22] the authors define the forward Fourier transform using the positive exponential: $\hat{f}(k) = \int f(x) e^{+2\pi i x k} dx$.

basically means applying a deconvolution. Then, we can find the object $O(\mathbf{x})$ by (inverse) Fourier transforming $W_O(\mathbf{x}, \mathbf{K})$ to $\mathbf{x}$ (or $\mathbf{K}$)

$$\begin{aligned}\mathcal{F}_{\mathbf{x}}\{W_O(\mathbf{x}, \mathbf{K})\}(\mathbf{k}, \mathbf{K}) &= \hat{O}(\mathbf{k})\hat{O}(\mathbf{k} - \mathbf{K})^*, \\ \mathcal{F}_{\mathbf{K}}^{-1}\{W_O(\mathbf{x}, \mathbf{K})\}(\mathbf{x}, \mathbf{X}) &= O(\mathbf{X} + \mathbf{x})O(\mathbf{X})^*,\end{aligned}\quad (1.32)$$

where $\hat{O}(\mathbf{k})$ denotes the Fourier transform of $O(\mathbf{x})$. Note that from Eq. (1.30) it follows that $W_O(\mathbf{x}, \mathbf{K})$ can only be reconstructed in the region where $W_p(\mathbf{x}, -\mathbf{K})$ is nonzero. If the probe has a sharp cut-off, $W_p(\mathbf{x}, -\mathbf{K})$ has a finite support in the $\mathbf{x}$ direction, so the retrieved $W_O(\mathbf{x}, -\mathbf{K})$ is cut-off in the $\mathbf{x}$ direction as well, so one would reconstruct $O(\mathbf{x})$ by Fourier transforming $W_O(\mathbf{x}, \mathbf{K})$ in the $\mathbf{K}$ direction (second option in Eq. (1.32)). However, if the *spectrum* of the probe has a sharp cut-off, $W_p(\mathbf{x}, -\mathbf{K})$ has a finite support in the $\mathbf{K}$ direction, so the retrieved $W_O(\mathbf{x}, -\mathbf{K})$ is cut-off in the $\mathbf{K}$ direction as well, so one would reconstruct $\hat{O}(\mathbf{k})$ by Fourier transforming $W_O(\mathbf{x}, \mathbf{K})$ in the $\mathbf{x}$ direction (first option in Eq. (1.32)).

- **Focus variation, the parabola method:** Consider an exit wave $\psi(\mathbf{x})$ with Fourier transform $\hat{\psi}(\mathbf{k})$ which is sharply peaked at $\mathbf{k} = \mathbf{0}$, i.e. $|\hat{\psi}(\mathbf{0})| \gg |\hat{\psi}(\mathbf{k})|$ for $\mathbf{k} \neq \mathbf{0}$. We can take a set of images $I(\mathbf{x}, A)$ for different values of the paraxial defocus parameter A

$$I(\mathbf{x}, A) = \left| \mathcal{F} \left\{ \hat{\psi}(\mathbf{k}) e^{-2\pi i A |\mathbf{k}|^2} \right\} (\mathbf{x}) \right|^2. \quad (1.33)$$

Inverse Fourier transforming this with respect to $\mathbf{x}$ gives the autocorrelation of $\hat{\psi}(\mathbf{k}) e^{-2\pi i A |\mathbf{k}|^2}$

$$\mathcal{F}^{-1} \{I(\mathbf{x}, A)\}(\mathbf{k}) = \int \hat{\psi}(\mathbf{k}')^* \hat{\psi}(\mathbf{k} + \mathbf{k}') e^{2\pi i A |\mathbf{k}'|^2} e^{-2\pi i A |\mathbf{k}' + \mathbf{k}|^2} d\mathbf{k}'. \quad (1.34)$$

Because of the assumption that $|\hat{\psi}(\mathbf{0})| \gg |\hat{\psi}(\mathbf{k})|$ for $\mathbf{k} \neq \mathbf{0}$, we can approximate the autocorrelation integral by considering only the terms involving $\hat{\psi}(\mathbf{0})$

$$\mathcal{F}^{-1} \{I(\mathbf{x}, A)\}(\mathbf{k}) \approx \hat{\psi}(\mathbf{0})^* \hat{\psi}(\mathbf{k}) e^{-2\pi i A |\mathbf{k}|^2} + \hat{\psi}(-\mathbf{k})^* \hat{\psi}(\mathbf{0}) e^{2\pi i A |\mathbf{k}|^2}. \quad (1.35)$$

In principle two different A are sufficient to generate a system of equations that can be solved, but for a better signal-to-noise ratio, more values for A are needed. We can retrieve approximations of $\hat{\psi}(\mathbf{k})$ and its twin image $\hat{\psi}(\mathbf{k})^*$ by applying a Fourier transform with respect to A

$$\begin{aligned}\hat{\psi}(\mathbf{0})^* \hat{\psi}(\mathbf{k}) &\approx \frac{1}{N} \sum_A \mathcal{F}^{-1} \{I(\mathbf{x}, A)\}(\mathbf{k}) e^{2\pi i A |\mathbf{k}|^2}, \\ \hat{\psi}(-\mathbf{k})^* \hat{\psi}(\mathbf{0}) &\approx \frac{1}{N} \sum_A \mathcal{F}^{-1} \{I(\mathbf{x}, A)\}(\mathbf{k}) e^{-2\pi i A |\mathbf{k}|^2},\end{aligned}\quad (1.36)$$

where N denotes the number of images taken, i.e. the number of different A used. This is the simplest reconstruction formula, which applied in the limit $N \rightarrow \infty$, but more sophisticated restoring filters have also been developed [23–26].

- **Extraction from autocorrelation functions:** If we have a field $\psi(\mathbf{x})$ and measure its far field intensity pattern $I(\mathbf{k}) = |\hat{\psi}(\mathbf{k})|^2$, we can inverse Fourier transform $I(\mathbf{k})$ to find the autocorrelation of $\psi(\mathbf{x})$

$$\begin{aligned}\mathcal{F}^{-1}\{I(\mathbf{k})\}(\mathbf{x}) &= \psi(\mathbf{x}) \otimes \psi(\mathbf{x}) \\ &= \int \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) d\mathbf{y},\end{aligned}\tag{1.37}$$

where $\otimes$ denotes the autocorrelation operator. Thus, to reconstruct $\psi(\mathbf{x})$ from $I(\mathbf{k})$ non-iteratively, one needs to manipulate $\psi(\mathbf{x})$ such that its autocorrelation function allows for an easy extraction of the sample's transmission function. FTH does this by adding a point source $\delta(\mathbf{x} + \mathbf{q})$, but it is possible to extend this principle as is done in for example **HERALDO** (Holography with Extended Reference by Autocorrelation Linear Differential Operator) [27], or as in [28–30], or by using Double Blind Fourier Holography (**DBFH**) [31, 32]. We will go into more detail in Chapter 2.

- Other non-iterative phase retrieval methods include the use of **shifting Gaussian filters** [33, 34], **scanning a slit aperture** [35, 36], or using an **aperture array** [37–39].

1.4. Iterative phase retrieval methods

The first iterative computational phase retrieval algorithm was introduced by Gerchberg and Saxton in 1972 [40]. In this method, one measures the field amplitudes $|\psi(\mathbf{x})|$, $|\hat{\psi}(\mathbf{k})|$ at two planes in the image plane and Fourier plane respectively. The **Gerchberg-Saxton algorithm** (**GS**) works by defining an estimated field $f(\mathbf{x})$, and propagating it back and forth between the two planes, each time setting the amplitude equal to the measured amplitude while keeping the phase of the estimated far field $\hat{f}(\mathbf{k})$. Mathematically, the n^{th} iteration is applied as follows:

$$\begin{aligned}\hat{f}_n(\mathbf{k}) &= \mathcal{F}\{f_n(\mathbf{x})\}(\mathbf{k}), \\ f_n^{\text{upd}}(\mathbf{x}) &= \mathcal{F}^{-1}\left\{\frac{\hat{f}_n(\mathbf{k})}{|\hat{f}_n(\mathbf{k})|}|\hat{\psi}(\mathbf{k})|\right\}(\mathbf{x}), \\ f_{n+1}(\mathbf{x}) &= |\psi(\mathbf{x})|\frac{f_n^{\text{upd}}(\mathbf{x})}{|f_n^{\text{upd}}(\mathbf{x})|}.\end{aligned}\tag{1.38}$$

In 1978, Fienup proposed a reconstruction method that uses as constraints the far field amplitude $|\hat{\psi}(\mathbf{k})|$ and the object support (as opposed to the object amplitude $|\psi(\mathbf{x})|$ like in the GS algorithm) [41]. The support constraint states that we know a priori that there is a region γ outside which $\psi(\mathbf{x}) = 0$. Several methods to find an estimate for the object support from its autocorrelation function are discussed in [42]. Other possible object-space constraints include the non-negativity constraint, realness constraint, histogram constraint, and atomicity constraint [43]. Fienup proposed the **Error Reduction algorithm** (**ER**), which alternately applies

the measurement constraint and the support constraint

$$f_{n+1} = \begin{cases} f_n^{\text{upd}}(\mathbf{x}) & \text{if } \mathbf{x} \in \gamma, \\ 0 & \text{if } \mathbf{x} \notin \gamma, \end{cases} \quad (1.39)$$

where $f_n^{\text{upd}}(\mathbf{x})$ is defined in Eq. (1.38). However, this algorithm tends to suffer from stagnation, so in order to speed up convergence, Fienup proposed the **Basic Input-Output Algorithm**, the **Output-Output Algorithm**, and most importantly the **Hybrid Input-Output algorithm (HIO)** [41, 44]

$$f_{n+1} = \begin{cases} f_n^{\text{upd}}(\mathbf{x}) & \text{if } \mathbf{x} \in \gamma, \\ f_n(\mathbf{x}) - \beta f_n^{\text{upd}}(\mathbf{x}) & \text{if } \mathbf{x} \notin \gamma. \end{cases} \quad (1.40)$$

The principle that is used to update the object estimate outside its support is similar to that of negative feedback: if the output f_n^{upd} violates the support constraint, the input is compensated by $-\beta f_n^{\text{upd}}$, where β is the feedback parameter, and it is typically chosen to be around 0.9 [45]. One can explain to some degree why HIO outperforms ER by observing that ER suffers from stagnation, and that the negative feedback used by HIO avoids such stagnation. However, one might still wonder why the HIO algorithm should have the specific form that it does. In 'Phase retrieval algorithms: a personal tour' (2012) [1], Fienup remarks the following about finding the HIO algorithm:

"I [...] tried mixing and matching different operations from different approaches to handling the values where the output image either satisfies or violates the constraints. This was not the beautiful mathematics of an Einstein that predicted what would happen long before an experiment was performed; this was the trial and error approach that Edison used to invent a practical light bulb: keep trying different things (guided by physics, mathematics, and intuition) until you find something that works; and then refine that. There is beautiful mathematics surrounding the phase retrieval problem, and it is centered around the zeros of the Fourier transform analytically extended to the complex plane; but that beautiful mathematics had yielded no practical phase retrieval algorithms."

In 1978, it was explained by Youla how the GS and ER algorithms can be interpreted in terms of **alternating projections (AP)** onto constraints sets [46] (we will go into more detail in Chapter 3). In 1984 it was pointed out how certain properties of the constraint sets, such as their non-convexity, can lead to stagnation [47]. In 2002, Bauschke identified the HIO-algorithm as the Douglas-Rachford algorithm, and explained how it can be interpreted in terms of projection and reflection operators [48]. Since then many other projection- and reflection-based algorithms have been developed, such as the **Difference Map (DM)** algorithm [43] or the **Relaxed Averaged Alternating Reflections (RAAR)** algorithm [49]. They have been compared in 2007 by Marchesini [50]. Alternatively, the ER algorithm can

be interpreted as the **minimization of a cost functional using a steepest descent scheme**, as demonstrated in 1982 by Fienup [44]. Several other variants of the phase retrieval algorithms have been proposed. For example, the **Shrinkwrap** algorithm updates the support constraint estimate during the reconstruction [45], and the **Oversampling Smoothness (OSS) method** [51] and the **Modified HIO (M-HIO) algorithm** [52] aim to improve the noise-robustness of the reconstruction algorithm.

An entirely different approach to the phase retrieval problem was presented by Candès et al. in 2012. They introduce a method called **PhaseLift**, in which one applies a convex relaxation by lifting the problem in a higher dimension, and which is solved through rank minimization [53, 54]. We go into a bit more detail in Sections 3.1.3 and 3.2.1, but otherwise this method does not play a major role in this thesis.

1.4.1. Ptychography

The history of ptychography goes back to as early as 1969 when Hoppe first proposed the method [55, 56] for crystallography using transmission electron microscopy. However, the foundations for ptychography in the form that is popular today were laid by Faulkner and Rodenburg in 2004 [57]. For a more detailed historical overview of the developments prior to 2004, we refer the reader to a review article by Rodenburg [58].

We have seen that one could reconstruct an object $O(\mathbf{x})$ non-iteratively from far-field intensity patterns obtained by illuminating it with a shifting probe $P(\mathbf{x} - \mathbf{X})$ using the Wigner Distribution Deconvolution Method (as explained in Section 1.3). We have also seen that one could reconstruct a wave function $\psi(\mathbf{x})$ iteratively from only a single far-field intensity pattern and a support constraint using algorithms such as HIO. One could wonder whether these two methods can be combined: in that case $O(\mathbf{x})$ could be reconstructed using a smaller data set than what would be required for WDDM, and the reconstruction algorithm would be more robust to noise than the single-intensity algorithms [58].

In 2004, Rodenburg and Faulkner proposed such an algorithm [59], which was later termed the **Ptychographic Iterative Engine (PIE)** [60]. In this algorithm, it is assumed that the illumination probe $P(\mathbf{x})$ is known, and is shifted to different positions $\mathbf{X}_j$. The exit waves are given by $\psi_j(\mathbf{x}) = P(\mathbf{x} - \mathbf{X}_j)O(\mathbf{x})$, and the measured far-field intensities are given by $I_j(\mathbf{k}) = |\hat{\psi}_j(\mathbf{k})|^2$. The probes at adjacent positions should overlap: the interconnections between the reconstructed exit waves is what makes this reconstruction algorithm robust. Briefly put, the algorithm works as follows: given an estimated object, one calculates for a certain position $\mathbf{X}_j$ the estimated exit wave $f_j(\mathbf{x})$, calculates the updated exit wave $f_j^{\text{upd}}(\mathbf{x})$ by applying the intensity constraint (just like in the GS, ER, and HIO algorithms, see Eq. (1.38)), and updates the object by factoring the probe $P(\mathbf{x})$ out of $f_j^{\text{upd}}(\mathbf{x})$. Doing this for all probe positions completes an iteration of the algorithm.

Just like the ER algorithm could be interpreted as an alternating projections scheme or a cost functional minimization scheme, so too is the case for PIE. In 2008, Guizar-Sicairos and Fienup demonstrated that PIE is closely related to a steepest-descent cost functional minimization scheme [61], and in 2009 Thibault et al. formulated a reconstruction algorithm in terms of projections [62]. One advantage is that these frameworks allow one to improve the reconstruction algorithm using other well-known methods such as the Conjugate Gradient (CG) scheme, or the HIO and DM algorithms [61–63]. Another major advantage is that one now can formulate reconstruction schemes that reconstruct not only $O(\mathbf{x})$, but also $P(\mathbf{x})$ (one particular algorithm being the **extended PIE**, or **ePIE** [64]) and X_j [61, 62, 64]. The ability to correct for aberrations in the illumination and uncertainties in the probe positions have made ptychography a particularly robust reconstruction method. Reconstruction schemes have also been developed to deal with partially coherent illumination [65, 66] or thick samples (which is known as **3PIE**) [67].

For single-intensity phase retrieval, we defined three ambiguities: global phase shift, translation, and twin image (see Section 1.2.1). In the case of ptychography where both $O(\mathbf{x})$ and $P(\mathbf{x})$ are reconstructed, one also encounters global phase shift and translation ambiguities, but no twin image ambiguity. Another ambiguity that is present in the reconstruction is the **raster grid pathology** [62]. If $O(\mathbf{x})$ and $P(\mathbf{x})$ are functions such that the exit waves $\psi_j(\mathbf{x}) = P(\mathbf{x} - \mathbf{X}_j)O(\mathbf{x})$ yield the desired far-field intensity patterns, then so are $f(\mathbf{x})O(\mathbf{x})$ and $\frac{P(\mathbf{x})}{f(\mathbf{x})}$, as long as $f(\mathbf{x}) = f(\mathbf{x} - \mathbf{X}_j)$ for all probe positions $\mathbf{X}_j$. So if the probe positions are chosen to be on a regular, periodic grid, then $f(\mathbf{x})$ can be any function with the same period, which means periodic artifacts may arise in the reconstruction. One can reduce the effects of this ambiguity by using a non-regular grid of probe positions [68], or one can eliminate this ambiguity by having an additional measurement constraint $|\hat{P}(\mathbf{k})|^2$, which is known as the **Modulus Enforced Probe (MEP)** method [69].

In Section 1.2.2 we discussed several sampling requirements for the far-field intensity patterns for single-intensity phase retrieval. For ptychography, the analysis becomes more complicated. It has been demonstrated that if there is a sufficient amount of overlap between adjacent probes, object reconstruction is still possible even if the far-field intensity patterns are undersampled according to the criteria of Section 1.2.2 [70, 71]. Moreover, it has been demonstrated that the far-field intensity patterns can be extrapolated if they are significantly cut off by the detector, thus achieving a higher resolution than what would be imposed by the detector **NA** [72]. Also, just like in single-intensity phase retrieval, it has been observed that one tends to obtain better reconstructions if the illumination is ‘wilder’ (i.e. contains more spatial frequencies), leading to more diffuse far-field intensity patterns [9, 73, 74].

1.5. Experimental results

So far, we have discussed the developments of the theory behind various phase retrieval methods. In the following, we briefly highlight several achievements in the experimental implementation of these methods to give an impression of what these methods are capable of in practice (without the ambition of providing an exhaustive list).

- **Wigner-distribution deconvolution in a scanning transmission X-ray microscope (STXM):** In 1996, Chapman used X-rays with a wavelength of $\lambda = 3.1\text{nm}$ to reconstruct images of a 1D grating with 146nm pitch, a $0.5\mu\text{m}$ diameter latex sphere, and a sample consisting of 5 such latex spheres [75]. For the grating, the data set was obtained by shifting the illumination 128 times in one direction with an interval of 22nm. For the sphere, the illumination was shifted across a 32×32 grid with 45nm intervals. For the set of five spheres, 64×64 illumination positions were used. In all these cases, images were successfully reconstructed using Wigner-distribution deconvolution. This was the first deterministic and non-interferometric recovery of phase information in X-ray imaging [76].
- **Single shot CDI for Soft X-Ray imaging using synchrotron radiation:** The first experimental demonstration of single-shot CDI for Soft X-ray imaging of a non-periodic microscopic object was performed by Miao et al. in 1999 [77]. Using monochromatic X-rays with wavelength $\lambda = 1.7\text{nm}$ generated by the National Synchrotron Light Source, they demonstrated imaging at 75nm resolution of a sample that consisted of a collection of gold dots, each 100nm in diameter and 80nm thick, which formed a set of six letters.
- **High-resolution ptychography with a table-top EUV source:** In 2015, Zhang et al. used a table top 30nm high harmonic source to image a reflective sample with 40nm by 80nm lateral resolution. The phase of the reconstructed image was used to determine the height profile of the sample with a resolution of $6\text{\AA}$ [78]. In 2017, Gardner et al. used tabletop 13.5 nm high-harmonic beams to image an extended, nearly periodic sample at a subwavelength resolution of 12.6 nm. They achieved this using the Modulus Enforced Probe (MEP) method [69]. Also in 2017, Porter et al. demonstrated reflection-mode lensless imaging using a 13 nm tabletop source [79].
- **Ptychographic tomography: 3D reconstructing an integrated circuit using hard X-rays:** In 2017, Holler et al. used 6.2 keV ($\lambda = 2\text{\AA}$) X-ray radiation generated by the SAXS beamline of the Swiss Light Source to create three-dimensional images of an integrated circuit with a lateral resolution in all directions down to 14.6 nm [80]. The reconstruction was performed using a mixed real-space/reciprocal-space imaging technique, called ptychographic X-ray computed tomography (PXCT), which had been developed by Dierolf et al. in 2010 to create a 3D image of a bone sample [81]. For the tomography, 1200 projections were acquired. The experimental challenges for ptychography and

tomography and possible ways to deal with them were outlined by Odstrcil in 2018 [82].

- **Bragg ptychography using hard X-rays:** In 2016, Hruszkewycz et al. introduced 3D Bragg projection ptychography (3DBPP), a method to obtain three-dimensional images of nanoscale lattice behaviour and strain fields in crystalline materials [83]. Using 9 keV energy X-rays with wavelength $\lambda = 0.137\text{nm}$ generated by the Hard X-ray Nanoprobe beamline, they imaged periodic embedded SiGe crystals with a resolution of 12, 29, and 35 nm respectively in the x, y, z dimensions from 707 coherent Bragg diffraction patterns.
- **Fourier ptychography using an optical microscope:** In 2013, Zheng et al. introduced Fourier ptychography, a method to reconstruct high-resolution, wide field-of-view (FOV), complex-valued images using a regular low-NA optical microscope, without requiring any mechanically moving components [84]. Using an LED array, a microscopic sample is illuminated at different angles, each illumination yielding a low-resolution intensity image, which are stitched together using the ptychographic algorithm. A conventional optical microscope (using a $\times 2$ objective, 0.08 NA) was transformed into a high-resolution ($0.78\mu\text{m}$ half-pitch resolution, 0.5 NA), wide-FOV ($\sim 120\text{ mm}^2$) microscope with a final space-bandwidth product (SBP) of 0.23 gigapixels. For comparison, a standard $\times 20$ microscope objective has a resolution of $0.8\mu\text{m}$ and a 1.1-mm-diameter FOV, corresponding to an SBP of ~ 7 megapixels.
- **Single-shot ptychography:** In 2015, Sidorenko and Cohen performed single-shot ptychography, where all diffraction patterns are recorded at once using a single detector. They illuminated a pinhole array with a plane wave ($\lambda = 405\text{nm}$), and focused the transmitted field with a lens. This creates an array of overlapping probes in the defocus plane of the lens, where they put the sample. With a second lens, they generated an array of diffraction patterns, which are all recorded on a single detector. A ptychographic reconstruction was performed successfully from this data. This method exploits the robustness of ptychography, while avoiding long acquisition times. Several other single-shot ptychography schemes are also proposed [85].
- **High resolution scanning transmission electron (STEM) ptychography:** In 2018, Jiang et al. used ptychography to achieve an Abbe diffraction-limited resolution of $0.39\text{\AA}$ with an 80 keV (4.2pm wavelength) electron beam [86]. To achieve the improved resolution and dose efficiency, an electron microscope pixel-array detector (EMPAD) with a high quantum efficiency and dynamic range was designed, which allowed for acquisition of the full 4D ptychographic data set in typically a minute. The method was demonstrated by imaging single-atom defects in MoS_2 .

1.6. Conclusion

In this chapter, we gave a broad outline of the theory of phase retrieval, and mentioned a few important experimental milestones in this field. Unicity, sampling

requirements, and coherence requirements for phase retrieval were discussed, and theory and interconnections of several phase retrieval methods were explained. These discussions provide the context in which the contents of the following chapters can be placed.

References

- [1] J. R. Fienup, *Phase retrieval algorithms: a personal tour [invited]*, *Applied Optics* **52**, 45 (2012).
- [2] Y. Bruck and L. Sodin, *On the ambiguity of the image reconstruction problem*, *Optics Communications* **30**, 304 (1979).
- [3] R. Bates, *Fourier phase problems are uniquely solvable in more than one dimension. i: Underlying theory*, *Optik (Stuttgart)* **61**, 247 (1982).
- [4] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, *Phase retrieval with application to optical imaging: a contemporary overview*, *IEEE signal processing magazine* **32**, 87 (2015).
- [5] A. M. J. Huizer and P. van Toorn, *Ambiguity of the phase-reconstruction problem*, *Optics Letters* **5**, 499 (1980).
- [6] J. H. Seldin and J. R. Fienup, *Numerical investigation of the uniqueness of phase retrieval*, *Journal of the Optical Society of America A* **7**, 412 (1990).
- [7] A. Fannjiang, *Absolute uniqueness of phase retrieval with random illumination*, *Inverse Problems* **28**, 075008 (2012).
- [8] A. Fannjiang and W. Liao, *Phase retrieval with random phase illumination*, *Journal of the Optical Society of America A* **29**, 1847 (2012).
- [9] P. Li, D. J. Batey, T. B. Edo, A. D. Parsons, C. Rau, and J. M. Rodenburg, *Multiple mode x-ray ptychography using a lens and a fixed diffuser optic*, *Journal of Optics* **18**, 054008 (2016).
- [10] D. Sayre, *Some implications of a theorem due to shannon*, *Acta Crystallographica* **5**, 843 (1952).
- [11] J. Miao, D. Sayre, and H. Chapman, *Phase retrieval from the magnitude of the fourier transforms of nonperiodic objects*, *JOSA A* **15**, 1662 (1998).
- [12] J. Spence, U. Weierstall, and M. Howells, *Coherence and sampling requirements for diffractive imaging*, *Ultramicroscopy* **101**, 149 (2004).
- [13] F. Zernike, *Phase contrast, a new method for the microscopic observation of transparent objects part ii*, *Physica* **9**, 974 (1942).
- [14] F. Zernike, *How i discovered phase contrast*, *Science* **121**, 345 (1955).

- [15] P. Gao, B. Yao, I. Harder, N. Lindlein, and F. Torcal-Milla, *Phase-shifting zernike phase contrast microscopy for quantitative phase measurement*, *Optics Letters* **36**, 4305 (2011).
- [16] M. R. Teague, *Deterministic phase retrieval: a green's function solution*, *Journal of the Optical Society of America* **73**, 1434 (1983).
- [17] T. Gureyev, A. Roberts, and K. Nugent, *Partially coherent fields, the transport-of-intensity equation, and phase uniqueness*, *JOSA A* **12**, 1942 (1995).
- [18] D. Paganin, A. Barty, P. J. McMahon, and K. A. Nugent, *Quantitative phase-amplitude microscopy. III. the effects of noise*, *Journal of Microscopy* **214**, 51 (2004).
- [19] K. Nugent, A. Peele, H. Chapman, and A. Mancuso, *Unique phase recovery for nonperiodic objects*, *Physical review letters* **91**, 203902 (2003).
- [20] K. Nugent, A. Peele, H. Quiney, and H. Chapman, *Diffraction with wavefront curvature: a path to unique phase recovery*, *Acta Crystallographica Section A: Foundations of Crystallography* **61**, 373 (2005).
- [21] R. Bates and J. Rodenburg, *Sub-ångström transmission microscopy: A fourier transform algorithm for microdiffraction plane intensity information*, *Ultramicroscopy* **31**, 303 (1989).
- [22] P. Li, T. B. Edo, and J. M. Rodenburg, *Ptychographic inversion via wigner distribution deconvolution: Noise suppression and probe design*, *Ultramicroscopy* **147**, 106 (2014).
- [23] P. Schiske, *Zur frage der bildrekonstruktion durch fokusreihen*, in *Proc. Eur. Conf. Electron Microsc.* (1968) pp. 145–146.
- [24] M. O. de Beeck, D. van Dyck, and W. Coene, *Wave function reconstruction in HRTEM: the parabola method*, *Ultramicroscopy* **64**, 167 (1996).
- [25] W. Saxton, *What is the focus variation method? is it new? is it direct?* *Ultramicroscopy* **55**, 171 (1994).
- [26] W. Saxton, *Computer techniques for image processing in electron microscopy, section 9.7*, Vol. 10 (Academic Press, 2013).
- [27] M. Guizar-Sicairos and J. Fienup, *Direct image reconstruction from a fourier intensity pattern using HERALDO*, *Optics Letters* **33**, 2668 (2008).
- [28] A. Martin and L. Allen, *Direct retrieval of a complex wave from its diffraction pattern*, *Optics Communications* **281**, 5114 (2008).
- [29] A. Martin, A. Bishop, D. Paganin, and L. Allen, *Practical implementation of a direct method for coherent diffractive imaging*, *Ultramicroscopy* **111**, 777 (2011).

- [30] A. J. Morgan, A. J. D'Alfonso, A. V. Martin, A. I. Bishop, H. M. Quiney, and L. J. Allen, *High-fidelity direct coherent diffractive imaging of condensed matter*, *Physical Review B* **84** (2011), 10.1103/physrevb.84.144122.
- [31] O. Raz, B. Leshem, J. Miao, B. Nadler, D. Oron, and N. Dudovich, *Direct phase retrieval in double blind fourier holography*, *Optics express* **22**, 24935 (2014).
- [32] B. Leshem, R. Xu, Y. Dallal, J. Miao, B. Nadler, D. Oron, N. Dudovich, and O. Raz, *Direct single-shot phase retrieval from the diffraction pattern of separated objects*, *Nature communications* **7**, 10820 (2016).
- [33] N. Nakajima, *Phase retrieval from fresnel zone intensity measurements by use of gaussian filtering*, *Applied optics* **37**, 6219 (1998).
- [34] N. Nakajima, *Phase-retrieval system using a shifted gaussian filter*, *JOSA A* **15**, 402 (1998).
- [35] N. Nakajima, *Phase retrieval from diffraction intensities by use of a scanning slit aperture*, *Applied optics* **44**, 6228 (2005).
- [36] N. Nakajima, *Improvement of resolution for phase retrieval by use of a scanning slit*, *Applied optics* **45**, 5976 (2006).
- [37] N. Nakajima, *Noniterative phase retrieval from a single diffraction intensity pattern by use of an aperture array*, *Physical review letters* **98**, 223901 (2007).
- [38] N. Nakajima, *Lensless coherent imaging by a deterministic phase retrieval method with an aperture-array filter*, *JOSA A* **25**, 742 (2008).
- [39] N. Nakajima, *Phase retrieval from a high-numerical-aperture intensity distribution by use of an aperture-array filter*, *JOSA A* **26**, 2172 (2009).
- [40] R. W. Gerchberg, *A practical algorithm for the determination of phase from image and diffraction plane pictures*, *Optik* **35**, 237 (1972).
- [41] J. Fienup, *Reconstruction of an object from the modulus of its fourier transform*, *Optics Letters* **3**, 27 (1978).
- [42] J. R. Fienup, *Phase retrieval algorithms: a comparison*, *Applied Optics* **21**, 2758 (1982).
- [43] V. Elser, *Phase retrieval by iterated projections*, *Journal of the Optical Society of America A* **20**, 40 (2003).
- [44] J. R. Fienup, *Phase retrieval algorithms: a comparison*, *Applied optics* **21**, 2758 (1982).
- [45] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, *X-ray image reconstruction from a diffraction pattern alone*, *Physical Review B* **68** (2003), 10.1103/physrevb.68.140101.

- [46] D. Youla, *Generalized image restoration by the method of alternating orthogonal projections*, *IEEE Transactions on Circuits and Systems* **25**, 694 (1978).
- [47] A. Levi and H. Stark, *Image restoration by the method of generalized projections with application to restoration from magnitude*, *Journal of the Optical Society of America A* **1**, 932 (1984).
- [48] H. Bauschke, P. Combettes, and D. Luke, *Phase retrieval, error reduction algorithm, and fienup variants: a view from convex optimization*, *Journal of the Optical Society of America A* **19**, 1334 (2002).
- [49] D. Luke, *Relaxed averaged alternating reflections for diffraction imaging*, *Inverse Problems* **21**, 37 (2004).
- [50] S. Marchesini, *Invited article: A unified evaluation of iterative projection algorithms for phase retrieval*, *Review of Scientific Instruments* **78**, 011301 (2007).
- [51] J. A. Rodriguez, R. Xu, C.-C. Chen, Y. Zou, and J. Miao, *Oversampling smoothness: an effective algorithm for phase retrieval of noisy diffraction intensities*, *Journal of Applied Crystallography* **46**, 312 (2013).
- [52] A. V. Martin, F. Wang, N. D. Loh, T. Ekeberg, F. R. N. C. Maia, M. Hanke, G. van der Schot, C. Y. Hampton, R. G. Sierra, A. Aquila, S. Bajt, M. Barthelmess, C. Bostedt, J. D. Bozek, N. Coppola, S. W. Epp, B. Erk, H. Fleckenstein, L. Foucar, M. Frank, H. Graafsma, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Liang, L. Lomb, S. Marchesini, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, J. Schulz, R. L. Shoeman, H. Soltau, D. Starodub, J. Steinbrener, F. Stellato, L. Strüder, J. Ullrich, G. Weidenspointner, T. A. White, C. B. Wunderer, A. Barty, I. Schlichting, M. J. Bogan, and H. N. Chapman, *Noise-robust coherent diffractive imaging with a single diffraction pattern*, *Optics Express* **20**, 16650 (2012).
- [53] 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**, 1241 (2012).
- [54] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, *SIAM Journal on Imaging Sciences* **6**, 199 (2013).
- [55] W. Hoppe and G. Strube, *Beugung in inhomogenen primärstrahlenwellenfeld. II. lichtoptische analogieversuche zur phasenmessung von gitterinterferenzen*, *Acta Crystallographica Section A* **25**, 502 (1969).
- [56] W. Hoppe, *Beugung im inhomogenen primärstrahlwellenfeld. III. amplituden- und phasenbestimmung bei unperiodischen objekten*, *Acta Crystallographica Section A* **25**, 508 (1969).

- [57] H. M. L. Faulkner and J. M. Rodenburg, *Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm*, *Physical Review Letters* **93** (2004), 10.1103/physrevlett.93.023903.
- [58] J. M. Rodenburg, *Ptychography and related diffractive imaging methods*, *Advances in Imaging and Electron Physics* **150**, 87 (2008).
- [59] J. Rodenburg and H. Faulkner, *A phase retrieval algorithm for shifting illumination*, *Applied physics letters* **85**, 4795 (2004).
- [60] J. Rodenburg, A. Hurst, and A. Cullis, *Transmission microscopy without lenses for objects of unlimited size*, *Ultramicroscopy* **107**, 227 (2007).
- [61] M. Guizar-Sicairos and J. Fienup, *Phase retrieval with transverse translation diversity: a nonlinear optimization approach*, *Optics Express* **16**, 7264 (2008).
- [62] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, *Probe retrieval in ptychographic coherent diffractive imaging*, *Ultramicroscopy* **109**, 338 (2009).
- [63] J. Qian, C. Yang, A. Schirotzek, F. Maia, and S. Marchesini, *Efficient algorithms for ptychographic phase retrieval*, *Inverse Problems and Applications, Contemp. Math* **615**, 261 (2014).
- [64] A. Maiden and J. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, *Ultramicroscopy* **109**, 1256 (2009).
- [65] P. Thibault and A. Menzel, *Reconstructing state mixtures from diffraction measurements*, *Nature* **494**, 68 (2013).
- [66] N. Burdet, X. Shi, D. Parks, J. N. Clark, X. Huang, S. D. Kevan, and I. K. Robinson, *Evaluation of partial coherence correction in x-ray ptychography*, *Optics express* **23**, 5452 (2015).
- [67] A. M. Maiden, M. J. Humphry, and J. Rodenburg, *Ptychographic transmission microscopy in three dimensions using a multi-slice approach*, *JOSA A* **29**, 1606 (2012).
- [68] X. Huang, H. Yan, R. Harder, Y. Hwu, I. Robinson, and Y. Chu, *Optimization of overlap uniformness for ptychography*, *Opt. Express* **22**, 12634 (2014).
- [69] D. F. Gardner, M. Tanksalvala, E. R. Shanblatt, X. Zhang, B. R. Galloway, C. L. Porter, R. Karl Jr, C. Bevis, D. E. Adams, H. C. Kapteyn, et al., *Subwavelength coherent imaging of periodic samples using a 13.5 nm tabletop high-harmonic light source*, *Nature Photonics* **11**, 259 (2017).
- [70] T. Edo, D. Batey, A. Maiden, C. Rau, U. Wagner, Z. Pešić, T. Waigh, and J. Rodenburg, *Sampling in x-ray ptychography*, *Physical Review A* **87**, 053850 (2013).
- [71] J. C. da Silva and A. Menzel, *Elementary signals in ptychography*, *Optics express* **23**, 33812 (2015).

- [72] A. M. Maiden, M. J. Humphry, F. Zhang, and J. M. Rodenburg, *Superresolution imaging via ptychography*, *JOSA A* **28**, 604 (2011).
- [73] A. Maiden, G. Morrison, B. Kaulich, A. Gianoncelli, and J. Rodenburg, *Soft x-ray spectromicroscopy using ptychography with randomly phased illumination*, *Nature Communications* **4** (2013), [10.1038/ncomms2640](https://doi.org/10.1038/ncomms2640).
- [74] M. Guizar-Sicairos, M. Holler, A. Diaz, J. Vila-Comamala, O. Bunk, and A. Menzel, *Role of the illumination spatial-frequency spectrum for ptychography*, *Physical Review B* **86** (2012), [10.1103/physrevb.86.100103](https://doi.org/10.1103/physrevb.86.100103).
- [75] H. N. Chapman, *Phase-retrieval x-ray microscopy by wigner-distribution deconvolution*, *Ultramicroscopy* **66**, 153 (1996).
- [76] K. A. Nugent, *Coherent methods in the x-ray sciences*, *Advances in Physics* **59**, 1 (2010).
- [77] 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**, 342 (1999).
- [78] B. Zhang, D. F. Gardner, M. D. Seaberg, E. R. Shanblatt, H. C. Kapteyn, M. M. Murnane, and D. E. Adams, *High contrast 3d imaging of surfaces near the wavelength limit using tabletop EUV ptychography*, *Ultramicroscopy* **158**, 98 (2015).
- [79] C. L. Porter, M. Tanksalvala, M. Gerrity, G. Miley, X. Zhang, C. Bevis, E. Shanblatt, R. Karl, M. M. Murnane, D. E. Adams, and H. C. Kapteyn, *General-purpose, wide field-of-view reflection imaging with a tabletop 13 nm light source*, *Optica* **4**, 1552 (2017).
- [80] M. Holler, M. Guizar-Sicairos, E. H. R. Tsai, R. Dinapoli, E. Müller, O. Bunk, J. Raabe, and G. Aepli, *High-resolution non-destructive three-dimensional imaging of integrated circuits*, *Nature* **543**, 402 (2017).
- [81] 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**, 436 (2010).
- [82] M. Odstrcil, M. Holler, J. Raabe, and M. Guizar-Sicairos, *High resolution 3d imaging of integrated circuits by x-ray ptychography*, in *Image Sensing Technologies: Materials, Devices, Systems, and Applications V*, edited by N. K. Dhar and A. K. Dutta (SPIE, 2018).
- [83] S. O. Hruszkewycz, M. Allain, M. V. Holt, C. E. Murray, J. R. Holt, P. H. Fuoss, and V. Chamard, *High-resolution three-dimensional structural microscopy by single-angle bragg ptychography*, *Nature Materials* **16**, 244 (2016).
- [84] G. Zheng, R. Horstmeyer, and C. Yang, *Wide-field, high-resolution fourier ptychographic microscopy*, *Nature Photonics* **7**, 739 (2013).

- [85] P. Sidorenko and O. Cohen, *Single-shot ptychography*, *Optica* **3**, 9 (2015).
- [86] Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, *et al.*, *Electron ptychography of 2d materials to deep sub-ångström resolution*, *Nature* **559**, 343 (2018).

2

Non-iterative phase retrieval

Several non-iterative methods for reconstructing a transmitted near field from far-field intensity patterns are discussed (though these methods can also be applied in an imaging setup). We review methods such as Fourier Transform Holography, where one reconstructs the transmitted near field from a single two-dimensional intensity pattern. We then discuss non-iterative methods based on three-dimensional data sets obtained through single-parameter modulation.

2.1. Problem statement

Let us denote the transmitted near field of an object as $\psi(\mathbf{x})$, where $\mathbf{x}$ is a two-dimensional position vector. The far field is given by its Fourier transform, which we denote as $\hat{\psi}(\mathbf{k})$. What we measure experimentally in the ideal noise-free case is the far field intensity

$$I(\mathbf{k}) = |\hat{\psi}(\mathbf{k})|^2. \quad (2.1)$$

The problem is to find $\psi(\mathbf{x})$ from $I(\mathbf{k})$.

2.2. Far field intensity and autocorrelation

One important observation that we will heavily rely on in this chapter, is that the inverse Fourier transform of $I(\mathbf{k})$ gives the autocorrelation of $\psi(\mathbf{x})$

$$\mathcal{F}^{-1}\{I(\mathbf{k})\}(\mathbf{x}) = \int \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) d\mathbf{y}. \quad (2.2)$$

The fact that we want to reconstruct $\psi(\mathbf{x})$ from $I(\mathbf{k})$ non-iteratively is equivalent to requiring that we should be able to extract $\psi(\mathbf{x})$ directly from its autocorrelation function [3]. As an example, we will discuss Fourier Transform Holography, where an object is extracted directly from a **two-dimensional** autocorrelation function.

If we generate a **three-dimensional** data set $I(\mathbf{k}, A)$ by varying an optical parameter A , we want to be able to extract $\psi(\mathbf{x})$ from a three-dimensional autocorrelation function of a fictitious mathematical object that is constructed using $\psi(\mathbf{x})$. The trick is to shape $\psi(\mathbf{x})$ and/or define the optical parameter A in such a way that $\psi(\mathbf{x})$ is straightforwardly encoded in the autocorrelation function. This is what we aim to do in this chapter. Typically, one reconstructs the object by isolating a single contribution to the autocorrelation integral: for example $\psi(\mathbf{P})^* \psi(\mathbf{x} + \mathbf{P})$ (so only the contribution $\mathbf{y} = \mathbf{P}$) or $\psi(\mathbf{P} - \mathbf{x})^* \psi(\mathbf{P})$ (so only the contribution $\mathbf{x} + \mathbf{y} = \mathbf{P}$), where $\mathbf{P}$ is some fixed reference point.

An example of a non-iterative reconstruction method using a **four-dimensional** autocorrelation function (of a fictitious mathematical object that is constructed using the object $\psi(\mathbf{x})$ and probe $P(\mathbf{x})$, see Appendix A) is the Wigner Distribution Deconvolution Method (WDDM), which is closely related to ptychography and pre-dates the iterative ptychographic reconstruction algorithms [4, 5].

2.3. Robustness to spatial partial coherence

The obvious advantages of explicit non-iterative reconstruction methods (such as Fourier Transform Holography) over iterative reconstruction methods (such as Ptychography) or implicit non-iterative reconstruction methods (such as the Transport of Intensity Equation) are the significantly lower computational expenses and the guarantee of a unique solution. A less obvious advantage is its robustness to spatial partial coherence [6–8]. A spatially partially coherent field in a plane is described by a four-dimensional mutual intensity function $J(\mathbf{x}, \mathbf{y})$ which gives the correlation

between the fields at any two points $\mathbf{x}, \mathbf{y}$. This function can be decomposed in coherent modes $\psi_n(\mathbf{x})$ [9]

$$J(\mathbf{x}, \mathbf{y}) = \sum_n \psi_n(\mathbf{x}) \psi_n(\mathbf{y})^*. \quad (2.3)$$

Here, the modes $\psi_n(\mathbf{x})$ are unnormalized. One could also represent the decomposition using normalized modes with some weight per mode. The less coherent the field is, the more modes are required. This is a useful representation, because one can now straightforwardly compute the far field intensity by propagating each mode coherently, and summing their intensities incoherently

$$I(\mathbf{k}) = \sum_n |\hat{\psi}_n(\mathbf{k})|^2. \quad (2.4)$$

Therefore, inverse Fourier transforming a spatially partially coherent far field intensity will give the sum of autocorrelations of all modes. Assuming that in the coherent case extracting $\psi(\mathbf{P})^* \psi(\mathbf{x} + \mathbf{P})$ or $\psi(\mathbf{P}) \psi(\mathbf{P} - \mathbf{x})^*$ from the autocorrelation is a linear operation, we extract in the spatially partially coherent case from the sum of autocorrelations

$$\begin{aligned} \sum_n \psi_n(\mathbf{x} + \mathbf{P}) \psi_n(\mathbf{P})^* &= J(\mathbf{x} + \mathbf{P}, \mathbf{P}), \\ \sum_n \psi_n(\mathbf{P}) \psi_n(\mathbf{P} - \mathbf{x})^* &= J(\mathbf{P}, \mathbf{P} - \mathbf{x}). \end{aligned} \quad (2.5)$$

Thus, using non-iterative reconstruction methods one will find a two-dimensional cross-section of the four-dimensional mutual intensity function. So if a sample with transmission function $\psi(\mathbf{x})$ is illuminated with a field with constant intensity and a Gaussian correlation structure

$$J(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x}) \psi(\mathbf{y})^* e^{-\frac{|\mathbf{x} - \mathbf{y}|^2}{2\sigma^2}}, \quad (2.6)$$

one would reconstruct

$$J(\mathbf{x} + \mathbf{P}, \mathbf{P}) = \psi(\mathbf{x} + \mathbf{P}) \psi(\mathbf{P})^* e^{-\frac{|\mathbf{x}|^2}{2\sigma^2}}, \quad (2.7)$$

which is $\psi(\mathbf{x})$ shifted to $-\mathbf{P}$, multiplied by a Gaussian function centered around $\mathbf{0}$ (one can intuitively imagine the mutual intensity function to be a spotlight that illuminates the object around the reference point $\mathbf{P}$, which has been shifted to the origin). Thus, without making any assumptions on the number of required modes $\psi_n(\mathbf{x})$, one still straightforwardly obtains an accurate impression of the sample, though with a limited field of view that depends on the degree of coherence. This is in stark contrast with iterative methods, where one needs to guess how many more modes need to be introduced as additional degrees of freedom, which increases the computational requirements significantly, and complicates the convergence behaviour of the reconstruction algorithm.

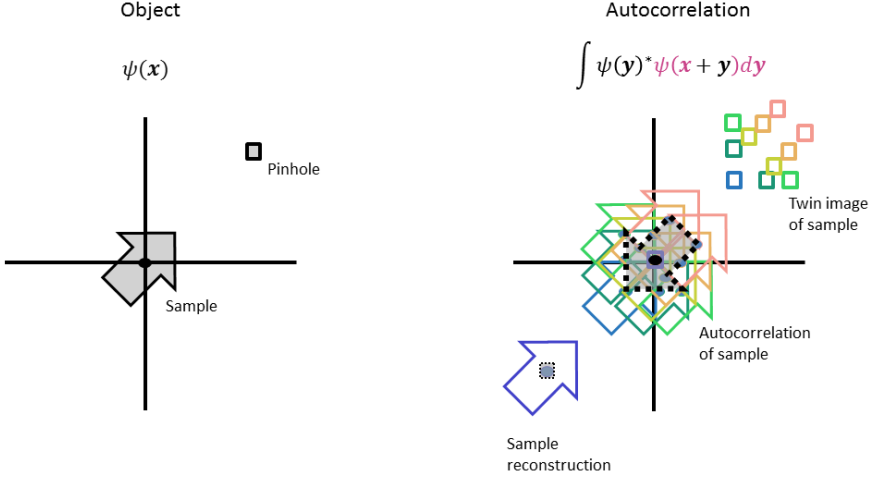


Figure 2.1: Pictorial explanation of Fourier Transform Holography. By introducing a pinhole sufficiently far away from a sample with finite size, one can straightforwardly extract an image of the sample from its autocorrelation function, which is found by inverse Fourier transforming the measured far field intensity pattern.

2.4. Fourier transform holography

As an example of a non-iterative retrieval scheme, we consider Fourier Transform Holography (FTH). In this method, one assumes that the sample $\psi(x)$ has a finite support, and that we can introduce a pinhole at a point P sufficiently far away from the sample. The total transmitted field reads $\psi(x) + a\delta(x - P)$. The autocorrelation of this field (which we assume to be fully coherent) is given by

$$\begin{aligned} \mathcal{F}^{-1}\{I(k)\}(x) &= \int [\psi(y)^* + a^* \delta(y - P)^*] [\psi(x + y) + a\delta(x + y - P)] dy \\ &= |a|^2 \delta(x) + a^* \psi(x + P) + a \psi(P - x)^* + \int \psi(y)^* \psi(x + y) dy. \end{aligned} \quad (2.8)$$

The last term represents the autocorrelation of $\psi(x)$, which one would obtain if the pinhole at P were absent. The middle two terms represent the reconstructed object shifted by $-P$, and its twin image $\psi(-x)^*$ shifted by P . If the shift P is sufficiently large, then the retrieved object and its twin image do not overlap with the autocorrelation of $\psi(x)$, so it can be extracted straightforwardly. See Fig. 2.1.

Note that this raises a problem in the case of spatial partial coherence. Recall that the mutual intensity function acts as a 'spotlight' that we shine at the reference point P (in this case the pinhole)¹, but the object of interest $\psi(x)$ has to be

¹It should be noted that the coherence function is independent of the illumination breadth. A more precise way of formulating this would be to say that the coherence width determines the field of view around the reference point.

sufficiently far removed from $\mathbf{P}$ in order for FTH to work. This means that our retrieved object has to be far away from the 'spotlight', which becomes a problem if the coherence width (i.e. the size of the spot light) is small. Quantitatively speaking, the coherence width should be at least four times the radius of the object.

One can move the reference point closer to the sample, in which case the terms $\psi(\mathbf{x} + \mathbf{P})$ and $\psi(\mathbf{P} - \mathbf{x})^*$ will overlap with the autocorrelation term, and additional intensity patterns should be recorded to generate a set of equations from which the reconstruction terms can be solved [6]. In particular, if $\mathbf{P}$ lies inside the sample, it does not any longer represent the position of a pinhole, but rather a point where we can modulate the phase of $\psi(\mathbf{x})$. By shifting the phase three times we can separate the terms $\psi(\mathbf{x} + \mathbf{P})$ and $\psi(\mathbf{P} - \mathbf{x})^*$ from the autocorrelation term. This method is closely related to phase shifting holography, phase shifting interferometry, and quantitative Zernike phase-contrast microscopy, which are explained in more detail in Section 1.3.

2.5. Non-iterative phase retrieval by varying a single parameter²

Having discussed how one can retrieve $\psi(\mathbf{x})$ under certain conditions using a single two-dimensional intensity measurement, let us now introduce an optical parameter A with which we can define a three-dimensional data set $I(\mathbf{k}, A)$ [1]. Let us modulate the transmitted exit wave $\psi(\mathbf{x})$ by a phase function $e^{-2\pi i A f(\mathbf{x})}$, where A is a parameter which we can vary freely, and $f(\mathbf{x})$ is a to be defined modulation function. We can calculate the Fourier transform as

$$\begin{aligned}\hat{\psi}_A(k_x, k_y) &= \int \psi(x, y) e^{-2\pi i A f(x, y)} e^{-2\pi i (k_x x + k_y y)} dx dy \\ &= \int \psi(x, y) \delta(z - f(x, y)) e^{-2\pi i (k_x x + k_y y + A z)} dx dy dz \\ &= \mathcal{F}_{x,y,z} \{ \psi(x, y) \delta(z - f(x, y)) \} (k_x, k_y, A).\end{aligned}\tag{2.9}$$

What we have demonstrated in the final expression is that we can find the three-dimensional data set $I(\mathbf{k}, A)$ by taking the squared modulus of the 3D-Fourier transform of a fictitious three-dimensional function $\psi(x, y) \delta(z - f(x, y))$. The advantage of this is that we can now formulate an intuitive line of reasoning of what it means to perform a non-iterative reconstruction with a data set $I(\mathbf{k}, A)$. Just like in FTH we manipulated our object $\psi(\mathbf{x})$ such that it can be easily extracted from its autocorrelation function, in the case of single-parameter variation we must choose our modulation function $f(x, y)$ such that $\psi(x, y)$ can be easily extracted from the three-dimensional autocorrelation function of $\psi_{3D}(x, y, z) = \psi(x, y) \delta(z - f(x, y))$. Moreover, using Fourier sampling theorems we can get an intuitive picture of how the sampling of A will affect the quality of the retrieved object.

²The following is original work by the candidate, which has been published in Ultramicroscopy **174**, 70-78 (2018) [1]

It is important to emphasize though, that while it may be helpful to think of the retrieval procedure as evaluating the autocorrelation function on some surface $s(\mathbf{x})$ (the choice of which will be explained in the next section) from a 3D autocorrelation function, it would be unnecessarily cumbersome and computationally expensive in practice to define a 3D array containing all the data $I(\mathbf{k}, A)$, 3D inverse Fourier transforming it, and extracting the desired data on the selected surface out of it. Rather, it is much more practical to compute the autocorrelation on a surface $s(\mathbf{x})$ by summing 2D inverse Fourier transforms of $I(\mathbf{k}, A)$ with respect to $\mathbf{k}$

$$R(\mathbf{x}) \propto \sum_A \mathcal{F}_k^{-1} \{I(\mathbf{k}, A)\}(\mathbf{x}) e^{2\pi i A s(\mathbf{x})}, \quad (2.10)$$

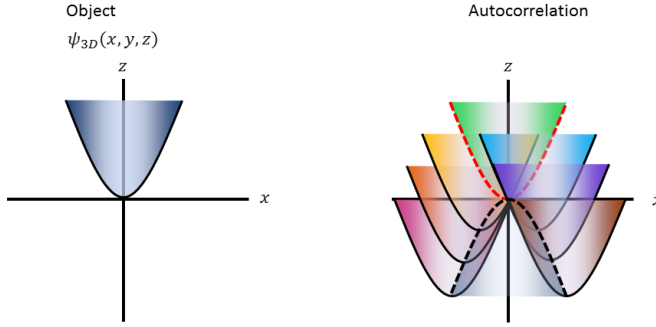
or if one wants to use a sampling window $\hat{H}(A)$ to mitigate the effects of a finite sampling range of A

$$R_H(\mathbf{x}) \propto \sum_A \mathcal{F}_k^{-1} \{I(\mathbf{k}, A)\}(\mathbf{x}) e^{2\pi i A s(\mathbf{x})} \hat{H}(A). \quad (2.11)$$

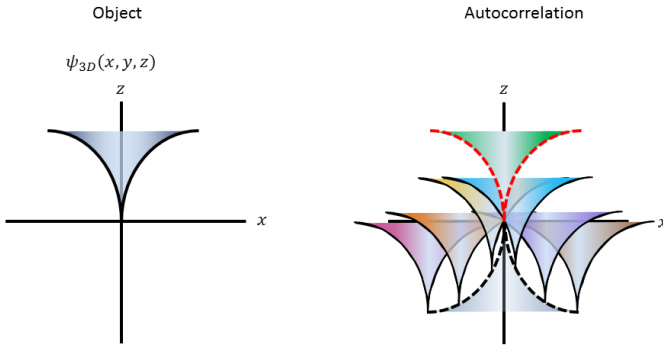
2.5.1. Intuitive line of reasoning

To get a feeling for which choices of $f(\mathbf{x})$ may be suitable and which ones are not, consider $f_2(\mathbf{x}) = |\mathbf{x}|^2$ and $f_{0.5}(\mathbf{x}) = \sqrt{|\mathbf{x}|}$. The subscripts 2 and 0.5 refer to the power that $|\mathbf{x}|$ is raised to. Note that $f_2(\mathbf{x})$ corresponds to paraxial focus variation. $f_{0.5}(\mathbf{x}) = \sqrt{|\mathbf{x}|}$ describes a less natural sort of phase modulation, which could for example be implemented using a Spatial Light Modulator (SLM). In the case of $f_2(\mathbf{x})$, $\psi(\mathbf{x})$ is stretched out onto a parabola, which can be interpreted as McCutchen's generalized pupil (which is conceptually related to the Ewald sphere) in the paraxial approximation [10–12]. When sketching the autocorrelation function, we find that the shifted copies of $\psi_{3D}(x, y, z)$ intersect each other in lots of places (as illustrated in Fig. 2.2c), which makes extracting $\psi(\mathbf{x})$ difficult, except if $|\psi(0)|$ is very large (as explained in Section 1.3: focus variation with the parabola method), or if $\psi(\mathbf{x})$ has a convenient support shape, as we will see in Section 2.6. On the other hand, in the case of $f_{0.5}(\mathbf{x})$, $\psi(\mathbf{x})$ is stretched out onto a surface with a sharp cusp. When sketching the autocorrelation function, we find that we can identify surfaces $z = \pm f_{0.5}(\mathbf{x})$ where the copies of $\psi_{3D}(x, y, z)$ do not interfere, allowing for an easy extraction of $\psi(\mathbf{x})$. See Fig. 2.2.

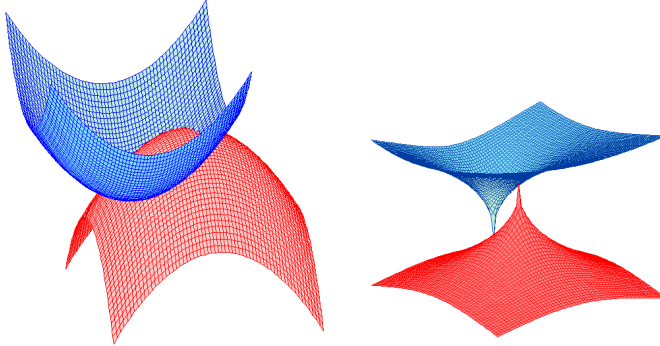
Moreover, we know from the Fourier convolution theorem that if the range along which we sample A is smaller, the more blurred the autocorrelation functions becomes in the z direction (it gets convolved with a sinc-function). If we choose a function with a sharper cusp, e.g. $f_{0.1} = |\mathbf{x}|^{0.1}$, we see that the reconstruction becomes more robust to this blurring. In order to mitigate the blurring effect due to a finite sampling range of A , one might introduce a sampling window $\hat{H}(A)$. In the next section, we will derive more rigorously the results we found here intuitively, and we will discuss how this reconstruction method is related to the phase shifting holography method that was mentioned previously.



(a) $\psi_{3D}(x, y, z)$ and its autocorrelation function for $f_2(x) = |x|^2$. Note that the illustration of $\psi_{3D}(x)$ should be interpreted as a surface, not a solid.



(b) $\psi_{3D}(x, y, z)$ and its autocorrelation function for $f_{0.5}(k) = \sqrt{|x|}$. Note that the illustration of $\psi_{3D}(x)$ should be interpreted as a surface, not a solid.



(c) Intersecting copies in the autocorrelation function for $f_2(k) = |x|^2$. (d) Intersecting copies in the autocorrelation function for $f_{0.5}(k) = \sqrt{|x|}$.

Figure 2.2: Sketches of the autocorrelation functions of $\psi_{3D}(x, y, z)$ for $f_{0.5}(x)$ and $f_2(x)$. In the case of $f_{0.5}(x)$, we can obtain direct reconstructions of $\psi_{3D}(x, y, z)$ on the red and black dotted surfaces (note that the other copies intersect the green copy only at $(x, y, z) = (0, 0, 0)$, as shown in Fig. 2.2d). In the case of $f_2(x)$, the red and black dotted surfaces contain lots of other contributions that impede a direct extraction of $\psi_{3D}(x, y, z)$.

2.5.2. Mathematical derivation

To find the retrieved object $R(\mathbf{x})$ of the actual object $\psi(\mathbf{x})$, we evaluate the 3D autocorrelation of $\psi_{3D}(x, y, z)$ on the surface $z = f(\mathbf{x})$, which according to Fig. 2.2 is the correct surface if $f(0) = 0$ (if $f(0) \neq 0$, then we can consider the function $f(\mathbf{x}) - f(0)$ which yields the same data set $I(\mathbf{k}, A)$):

$$R(\mathbf{x}) = \iint I(\mathbf{k}, A) e^{2\pi i \mathbf{x} \cdot \mathbf{k}} e^{2\pi i A f(\mathbf{x})} d\mathbf{k} dA. \quad (2.12)$$

We know $\iint I(\mathbf{k}, A) e^{2\pi i \mathbf{x} \cdot \mathbf{k}} d\mathbf{k}$ gives the autocorrelation of $\psi(\mathbf{x}) e^{-2\pi i A f(\mathbf{x})}$, so we can write

$$R(\mathbf{x}) = \iint \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) e^{2\pi i A [f(\mathbf{x}) + f(\mathbf{y}) - f(\mathbf{x} + \mathbf{y})]} d\mathbf{y} dA. \quad (2.13)$$

If we apply a window function $\hat{H}(A)$, we get

$$\begin{aligned} R_H(\mathbf{x}) &= \iint \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) e^{2\pi i A [f(\mathbf{x}) + f(\mathbf{y}) - f(\mathbf{x} + \mathbf{y})]} \hat{H}(A) d\mathbf{y} dA \\ &= \iint \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) H[f(\mathbf{x}) + f(\mathbf{y}) - f(\mathbf{x} + \mathbf{y})] d\mathbf{y} \end{aligned} \quad (2.14)$$

where H denotes the inverse Fourier transform of $\hat{H}$. In the case of ideal sampling, so $\hat{H}(A) = 1$, H reduces to a delta function (which equivalently follows from the integration over A in Eq. (2.13))

$$R(\mathbf{x}) = \iint \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) \delta[f(\mathbf{x}) + f(\mathbf{y}) - f(\mathbf{x} + \mathbf{y})] d\mathbf{y}. \quad (2.15)$$

In order to have a direct retrieval of $\psi(\mathbf{x})$ we require that the argument of the delta function vanishes for only one value of $\mathbf{y}$. Note that this requires $\mathbf{x} \neq 0$, because if $\mathbf{x} = 0$, all $\mathbf{y}$ will contribute to the integral (since we assumed $f(0) = 0$). So henceforth we only consider $\mathbf{x} \neq 0$. We can require that $f(\mathbf{x}) + f(\mathbf{y}) - f(\mathbf{x} + \mathbf{y})$ vanishes only for $\mathbf{y} = 0$. In that case $R(\mathbf{x}) = \psi(0)^* \psi(\mathbf{x})$. Had we chosen to evaluate the surface $z = -f(\mathbf{x})$ instead, we would require that the argument of the delta function only vanishes for $\mathbf{x} + \mathbf{y} = 0$, in which case the reconstruction would contain the twin image of $\psi(\mathbf{x})$: $R(\mathbf{x}) = \psi(-\mathbf{x})^* \psi(0)$.

For brevity, let us define

$$g(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) - f(\mathbf{x} + \mathbf{y}). \quad (2.16)$$

How do we choose $f(\mathbf{x})$ such that $g(\mathbf{x}, \mathbf{y})$ vanishes only for $\mathbf{y} = 0$? We demonstrate that a function of the following form satisfies the requirement:

$$f(\mathbf{x}) = h(n(\mathbf{x})), \quad (2.17)$$

where $n(\mathbf{x})$ is a vector norm (e.g. the Euclidean norm $|\mathbf{x}| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$), and $h(a)$ is a monotonically increasing subadditive function, i.e.

$$h(a + b) \leq h(a) + h(b), \quad (2.18)$$

where equality only holds when $a = 0$ or $b = 0$. An example of such a function is $h(a) = a^K$ with $K \in (0, 1)$. To see why a function $f(x)$ of the form $h(n(x))$ works, consider the inequality

$$f(x+y) = h(n(x+y)) \leq h(n(x)+n(y)) \leq h(n(x))+h(n(y)) = f(x)+f(y). \quad (2.19)$$

The first inequality holds because of the triangle inequality (which holds by definition of a vector norm) and because $h(a)$ is a monotonically increasing function. The second inequality holds because $h(a)$ is a subadditive function. Note that equality only holds when $n(x) = 0$ or $n(y) = 0$, which by definition of a vector norm holds only when $x = 0$ or $y = 0$. Thus, $g(x, y)$ only vanishes when $x = 0$ (which we chose not to consider) or $y = 0$, which is what we required. Therefore, phase modulation functions of the form $f_K(x) = |x|^K$ with $K \in (0, 1)$ allow for a non-iterative retrieval of $\psi(x)$.

2.5.3. Relation to phase shifting holography

Previously, we observed that if the cusp of $f(x)$ is sharper, the reconstruction method suffers less from the finite sampling range of A . So let us consider what happens in the extreme case $K \rightarrow 0$ where we modulate the phase everywhere except in a single point (or equivalently: where we modulate the phase in only a single point)

$$f(x) = \begin{cases} 0 & \text{if } |x| = 0, \\ 1 & \text{if } |x| \neq 0. \end{cases} \quad (2.20)$$

In this case, if $x \neq 0$, we get

$$g(x, y) = \begin{cases} 0 & \text{if } y = 0, \\ 2 & \text{if } y = -x, \\ 1 & \text{otherwise.} \end{cases} \quad (2.21)$$

We want $H(g(x, y))$ to give a contribution to the integral of Eq. (2.14) only if $y = 0$, but since $g(x, y)$ can only take three values, it is sufficient to require

$$H(0) = 1 \quad H(1) = 0 \quad H(2) = 0. \quad (2.22)$$

A solution would be

$$H(a) = \frac{1}{3} \left(1 + e^{2\pi i \frac{1}{3}a} + e^{2\pi i \frac{2}{3}a} \right), \quad (2.23)$$

which would correspond to a sampling function of

$$\hat{H}(A) \propto \delta(A) + \delta\left(A - \frac{1}{3}\right) + \delta\left(A - \frac{2}{3}\right), \quad (2.24)$$

which means one only has to take three measurements, and for each measurement shift the phase of a single pixel, which is what is done in quantitative Zernike phase contrast microscopy [13]. Realizing that the single pixel in which the field is modulated is a reference plane wave when taking the Fourier transform, one can also find analogies with phase-shifting interferometry and phase-shifting holography, as explained in Section 1.3.

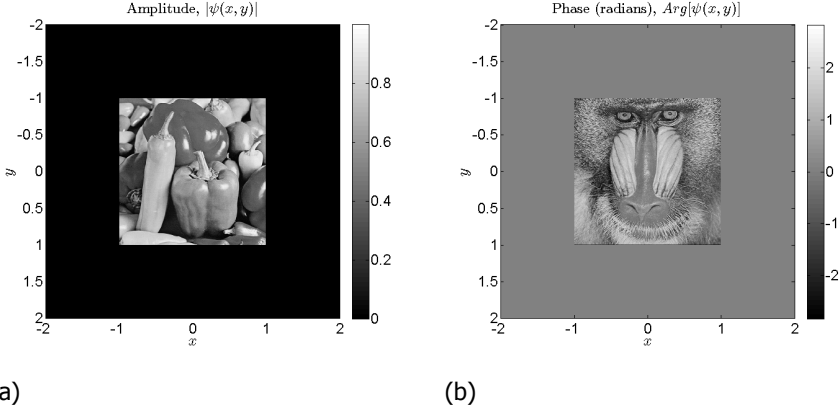


Figure 2.3: Object $\psi(x)$ used for the simulation to test the non-iterative phase retrieval method using phase modulation of the form $f(x) = |x|^K$, $K \in (0, 1)$.

2.5.4. Simulation results

In this section, we simulate the proposed method for a lensless imaging setup. This means that we apply the phase modulations $e^{-2\pi i A f(x)}$ to $\psi(x)$ (i.e. in object space), and measure the intensity patterns in Fourier space (i.e. in a far field plane). However, it is also possible to simulate the method for an aberrated imaging setup (as is done in [1]), which means that the phase modulations $e^{-2\pi i A f(k)}$ are applied to $\hat{\psi}(k)$ (the Fourier transform of the sample), and the measurements are performed in image space. In that case, fewer measurements are needed to get a good reconstruction of $\psi(x)$, since $|\hat{\psi}(k)|$ tends to be sharply peaked at $k = 0$ (the reason why this is beneficial has been explained in Section 1.3: Focus variation, the parabola method).

The simulated object is shown in Fig. 2.3. The actual sample consists of 256×256 pixels, but the array has been padded to a size of 512×512 pixels to prevent aliasing when computing the sample's autocorrelation function, as explained in Section 1.2.2. The units of x and y are chosen such that the sample has width 2, and the entire array has width 4. The (x, y) -grid must be carefully chosen such that the point $(x, y) = (0, 0)$ falls exactly on a pixel, rather than in between two pixels, which could for example happen if you define your axis symmetrically (i.e. going from $-A_{\max}$ to $A_{\max}$) while having an even number of pixels. The reason is that in the reconstruction, there will be a large peak at $(x, y) = (0, 0)$, and if this point does not correspond to a certain pixel, its energy will be spread out over the rest of the reconstruction, making the reconstruction completely unrecognizable.

In Fig. 2.4 the effect of the sampling window $\hat{H}(A)$ is investigated. The modulation function $f_K(x) = |x|^K$ is fixed at $K = 0.1$, the sampling grid $-A_{\max} : \Delta_A : A_{\max}$ is fixed at sampling range $A_{\max} = 20$ and sampling interval $\Delta_A = 0.65$, giving a total of $N = 62$ intensity measurements. The sampling window $\hat{H}(A)$ is chosen to be a

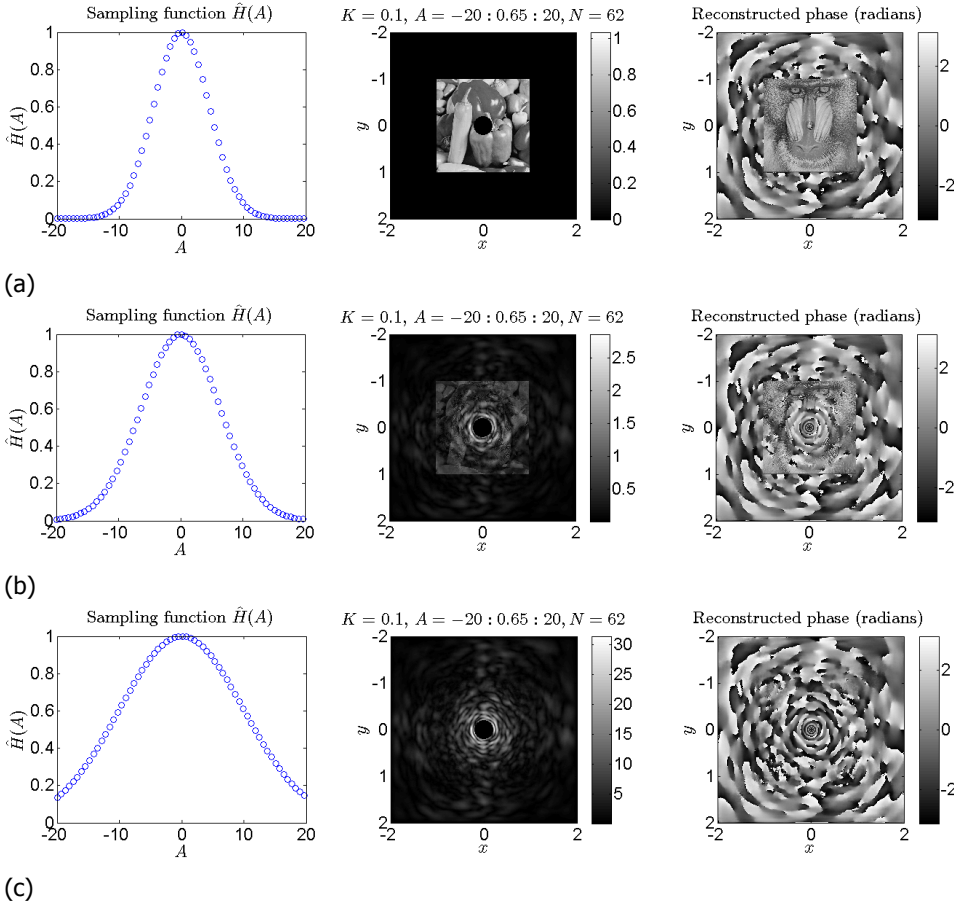


Figure 2.4: Effect of the sampling function $\hat{H}(A)$. The amplitude of the central region of the reconstruction is set to 0 to get rid of the peak at $x = 0$.

Gaussian, of which we vary the width. We observe that the role of the sampling window is significant.

In Fig. 2.5 the effect of the sampling range $A_{\max}$ is investigated. We see that as the sampling range is reduced, the reconstruction gets corrupted at the corners first. This can be understood by looking at the 3D autocorrelation function in Fig. 2.2a. The inverse of Δ_A determines the distance between aliases of the autocorrelation function in the z -direction, while the sampling range $A_{\max}$ determines the degree of blurring in the z -direction. In Fig. 2.5c, the sampling is chosen optimally, so that the aliases of the autocorrelation function almost touch each other at the corners. If the range $A_{\max}$ is reduced, and the autocorrelation gets blurred in the z -direction, the aliases will affect the reconstruction first at the corners. In Fig. 2.6, where the sampling interval Δ_A is increased, the same effect can be observed.

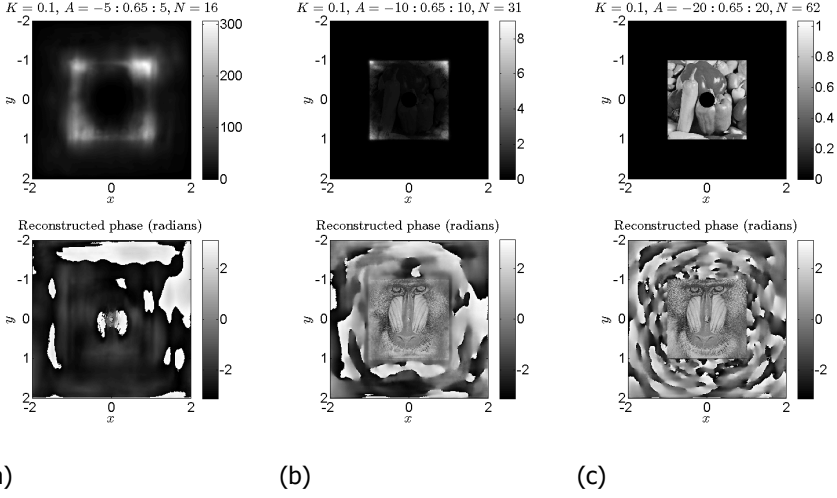


Figure 2.5: Effect of the sampling range $A_{\max}$

In Fig. 2.7 the effect of the modulation function $f_K(\mathbf{x}) = |\mathbf{x}|^K$ is investigated. We see that if K is increased, the autocorrelation function gets stretched out in the z -direction, causing the aliases to overlap.

2.6. Non-iterative phase retrieval using a star-shaped mask and through-focus scanning³

By examining the 3D autocorrelation function for $f_2(\mathbf{x}) = |\mathbf{x}|^2$, we found that if $\psi(\mathbf{x})$ has a suitable support (which will be described more precisely in Section 2.6.2), it should be possible to extract $\psi(\mathbf{x})$ straightforwardly from the boundary of the autocorrelation function [2]. The reason why it is attractive to consider $f_2(\mathbf{x}) = |\mathbf{x}|^2$ is because it corresponds to focus variation in the paraxial limit. As opposed to phase modulation of the sort $f_K(\mathbf{x}) = |\mathbf{x}|^K$ with $K \in (0, 1)$, which would require a spatial light modulator to implement, focus variation can be achieved in a natural way. For example, one can move the detector through focus, or vary the strength of a liquid lens (when using visible light) or an electromagnetic lens (when using electrons). Note that in order to assign negative values to the defocus parameter A , one needs to scan the field at both sides of the focal plane. If $\psi(\mathbf{x})$ is real-valued, the focal field is symmetric in the axial direction, so it is sufficient to consider only positive A . In that case, one can vary A by considering different free-space propagation distances in the Fresnel approximation, and no focusing optics are required in principle.

³The following is original work by the candidate, which has been published in Optics Express **26:7**, 9332-9343 (2018) [2]

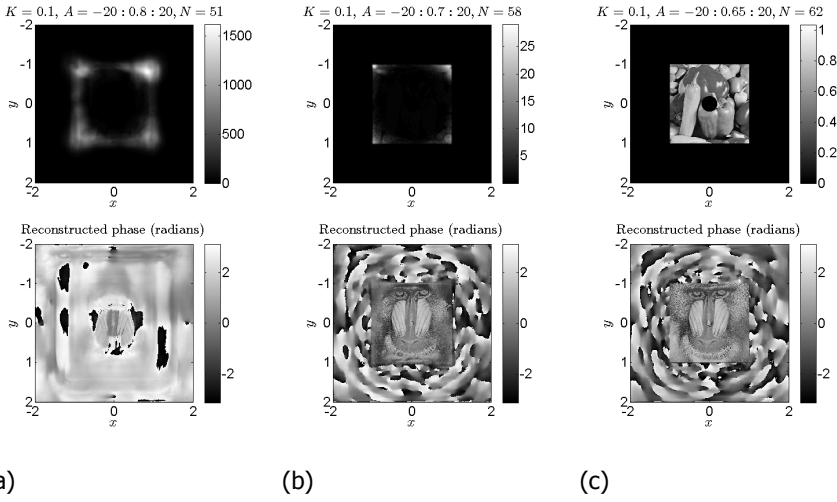


Figure 2.6: Effect of the sampling interval Δ_A

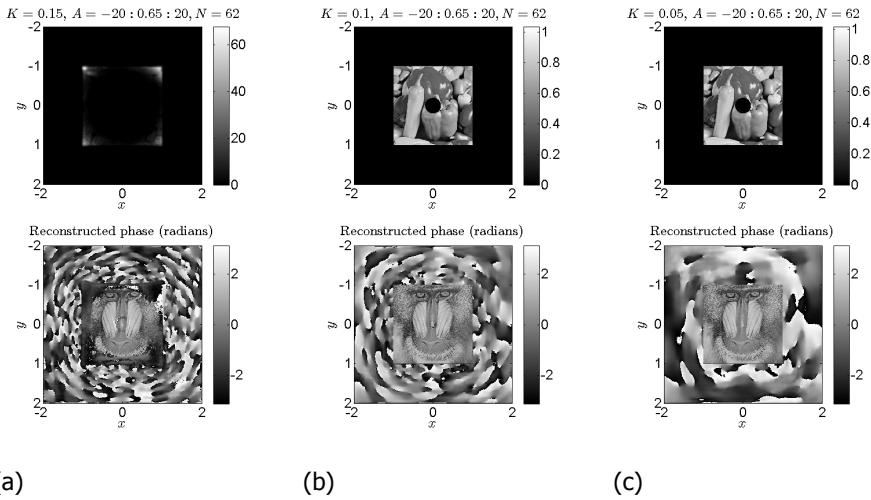


Figure 2.7: Effect of K .

2.6.1. Finding the correct surface of the autocorrelation function

To find the reconstruction $R(\mathbf{x})$ of $\psi(\mathbf{x})$, we evaluate the 3D autocorrelation of $\psi_{3D}(x, y, z)$ on some surface $z = s(\mathbf{x})$

$$\begin{aligned} R(\mathbf{x}) &= \iint I(\mathbf{k}, A) e^{2\pi i \mathbf{x} \cdot \mathbf{k}} e^{2\pi i A s(\mathbf{x})} d\mathbf{k} dA \\ &= \iint \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) \delta[s(\mathbf{x}) + f(\mathbf{y}) - f(\mathbf{x} + \mathbf{y})] d\mathbf{y} \\ &= \iint \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) \delta[s(\mathbf{x}) - |\mathbf{x}|^2 - 2\mathbf{x} \cdot \mathbf{y}] d\mathbf{y}. \end{aligned} \quad (2.25)$$

From our illustration of the autocorrelation function in Fig. 2.8 (which is a more annotated version of Fig. 2.2a), we see that if we want to evaluate the copy $\psi_{3D}(x, y, z)$ that is shifted to the edge of the autocorrelation function, we must choose a reference point $\mathbf{P}$ that lies on the edge of the support of $\psi(\mathbf{x})$, and choose the surface

$$\begin{aligned} z &= -f(\mathbf{P} - \mathbf{x}) + f(\mathbf{P}) \\ &= -|\mathbf{x}|^2 + 2\mathbf{P} \cdot \mathbf{x}, \end{aligned} \quad (2.26)$$

or its mirrored surface

$$\begin{aligned} z &= f(\mathbf{x} + \mathbf{P}) - f(\mathbf{P}) \\ &= |\mathbf{x}|^2 + 2\mathbf{P} \cdot \mathbf{x}, \end{aligned} \quad (2.27)$$

We can also derive mathematically why $s(\mathbf{x}) = -|\mathbf{x}|^2 + 2\mathbf{P} \cdot \mathbf{x}$ and $s(\mathbf{x}) = |\mathbf{x}|^2 + 2\mathbf{P} \cdot \mathbf{x}$ are the surfaces on which we should evaluate the autocorrelation function: we want to choose $s(\mathbf{x})$ such that the argument of the delta function is zero when $\mathbf{y} = \mathbf{P}$ (in the next section we will make sure the argument of the delta function is zero **only** when $\mathbf{y} = \mathbf{P}$). If we define

$$\mathbf{v} = \mu \mathbf{x} - 2\mathbf{y}, \quad \mu = \frac{s(\mathbf{x})}{|\mathbf{x}|^2} - 1, \quad (2.28)$$

we can write the argument of the delta function as $\mathbf{x} \cdot \mathbf{v}$. So the argument of the delta function is zero when $\mathbf{x} \cdot \mathbf{v} = 0$, which means

$$\mathbf{v} = \lambda \begin{bmatrix} y \\ -x \end{bmatrix}, \quad (2.29)$$

where $\mathbf{x} = [x, y]^T$, and for any constant λ . Combining this with the definition of $\mathbf{v}$ for $\mathbf{y} = \mathbf{P}$, we can equate the two expressions for $\mathbf{v}$, yielding the matrix equation

$$\begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} \mu \\ \lambda \end{bmatrix} = 2 \begin{bmatrix} P_x \\ P_y \end{bmatrix}, \quad (2.30)$$

where $\mathbf{P} = [P_x, P_y]^T$. Solving this equation gives $\mu = 2 \frac{\mathbf{P} \cdot \mathbf{x}}{|\mathbf{x}|^2}$, and solving for $s(\mathbf{x})$ from the definition of μ gives $s(\mathbf{x}) = |\mathbf{x}|^2 + 2\mathbf{P} \cdot \mathbf{x}$, which is the desired result. If

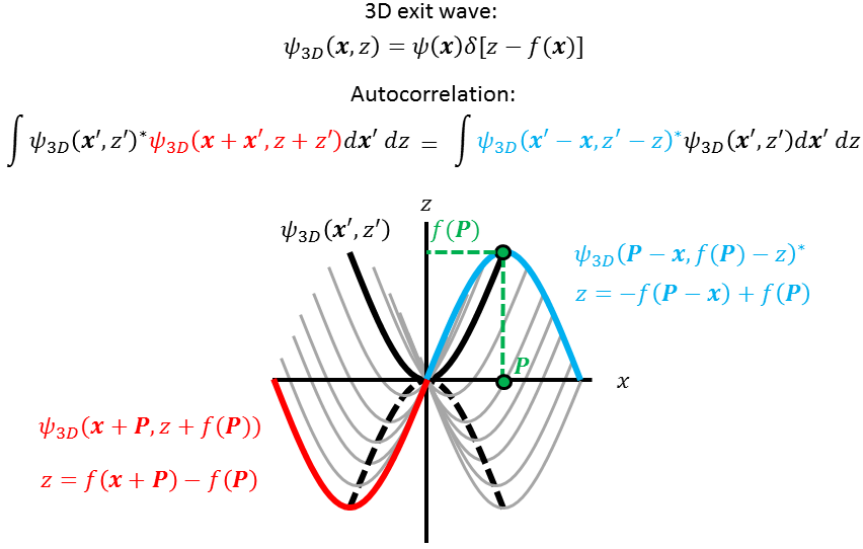


Figure 2.8: Illustration of the 3D autocorrelation function for focus variation, indicating on which surfaces one can find a direct reconstruction of $\psi(\mathbf{x})$, given that $\psi(\mathbf{x})$ has an appropriate support with a reference point $\mathbf{P}$, as explained in Fig. 2.9.

we had required instead that the argument of the delta function is zero only when $\mathbf{y} + \mathbf{x} = \mathbf{P}$, which would correspond to finding the twin image of the object, we would have found the other surface $s(\mathbf{x}) = -|\mathbf{x}|^2 + 2\mathbf{P} \cdot \mathbf{x}$.

2.6.2. Finding the requirements for the object support

We have chosen the surface of evaluation $s(\mathbf{x})$ such that the argument of the delta function vanishes for $\mathbf{y} = \mathbf{P}$. For $s(\mathbf{x}) = |\mathbf{x}|^2 + 2\mathbf{P} \cdot \mathbf{x}$, Eq. (2.25) reduces to

$$R(\mathbf{x}) = \iint \psi(\mathbf{y})^* \psi(\mathbf{x} + \mathbf{y}) \delta[2\mathbf{x} \cdot (\mathbf{P} - \mathbf{y})] d\mathbf{y}. \quad (2.31)$$

Now we must shape the support of the object $\psi(\mathbf{x})$ such that the integrand contributes to the integral **only** for $\mathbf{y} = \mathbf{P}$. For this we require

1. $\psi(\mathbf{P})^* \neq 0$ (and preferably $|\psi(\mathbf{P})|$ is large, to guarantee robustness to noise)
2. $\psi(\mathbf{x} + \mathbf{P}) \neq 0$
3. $\psi(\mathbf{y})^* = 0$ or $\psi(\mathbf{x} + \mathbf{y}) = 0$ when $\mathbf{P} - \mathbf{y}$ is perpendicular to $\mathbf{x}$ and $\mathbf{y} \neq \mathbf{P}$.

We can make sure that these requirements hold by imposing a certain support on $\psi(\mathbf{x})$ (which one can do physically by introducing a mask). We can check these requirements visually as follows:

- Sketch $\psi(\mathbf{x})$ and check that $\psi(\mathbf{P})^* \neq 0$, i.e. $\mathbf{P}$ should lie within the support of $\psi(\mathbf{x})$. This fulfills requirement 1.

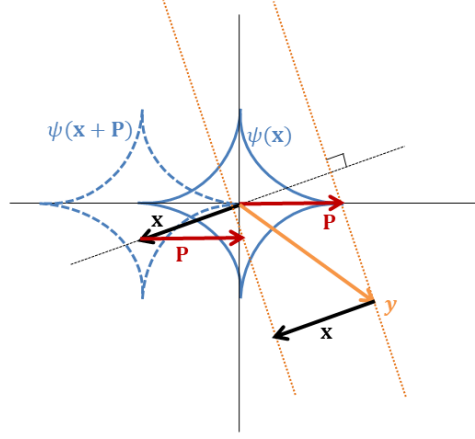


Figure 2.9: Illustration showing that a star-shaped mask (blue) satisfies the three requirements for $\psi(x)$: P (red) is such that $\psi(P)^* \neq 0$; the relevant x are those for which $\psi(x+P) \neq 0$, i.e. all x which lie in the dotted blue star; we require for all relevant x that $\psi(y)^* = 0$ or $\psi(x+y) = 0$ when $P-y$ is perpendicular to x (orange dotted line) and $y \neq P$. Since for all x in the dotted blue star, the line that goes through P and is perpendicular to x intersects the blue star only in P , the requirements are met.

- Sketch $\psi(x+P)$ and see for which x it holds that $\psi(x+P) \neq 0$. These are the x which are relevant for the reconstruction. This fulfills requirement 2.
- For all the directions of x that are relevant for the reconstruction, draw lines perpendicular to x through P , and lines perpendicular to x through $x+P$. These lines make up the collection of y and $x+y$ respectively for which $x \cdot (P-y) = 0$. Verify that there is no y except for $y = P$ for which both $\psi(x+y)$ and $\psi(y)^*$ are nonzero. This fulfills requirement 3.

Through this procedure one can demonstrate that a star-shaped mask is suitable for non-iterative focus-variation reconstruction, see Fig. 2.9. Because of the sharp protrusions of the star, the line that is perpendicular to any relevant x and goes through P will intersect $\psi(x)$ only in a small region, even if the line has a finite thickness due to the finite sampling range of A . Therefore, only a small region of y around P will contribute to the integral in Eq. (2.31). Just like in Eq. (2.14), it may be convenient to use a sampling function $\hat{H}(A)$ (such as a Hamming window) to make sure that in case of a finite sampling range of A , the delta function in Eq. (2.31) does not change to a sinc-function with significant side lobes,

$$R(x) = \iint \psi(y)^* \psi(x+y) W \text{sinc}[2\pi W x \cdot (P-y)] dy, \quad (2.32)$$

(where W denotes the sampling range), but rather to a slightly broader function with fewer side lobes, as shown in Fig. 2.10.

In this proposed method, we require a mask and a through focus scan to obtain a direct reconstruction of $\psi(x)$. In Fourier Transform Holography, we also require a

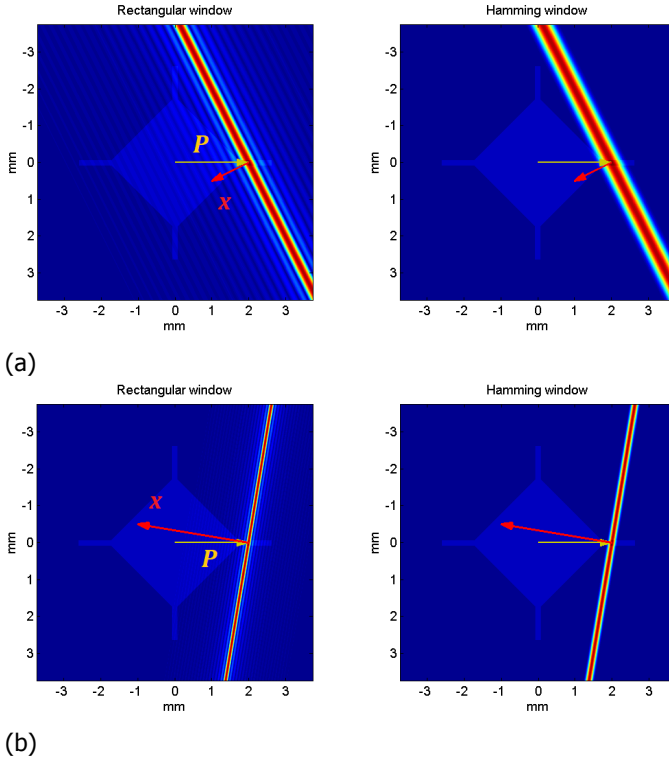


Figure 2.10: Illustration of how the side lobes of the sinc-function in Eq. (2.32) can be suppressed using an appropriate sampling function $\hat{H}(A)$, which is in this case a Hamming window. Note that according to Eq. (2.32), the sinc-function becomes narrower as $|x|$ increases. The dimensions on the axes correspond to the physical dimensions of the pattern assigned to the SLM in the experiment, as shown in Fig. 2.12.

mask, but we only need to take a single measurement to perform a reconstruction. So what is the advantage of the proposed method? For FTH, it was pointed out in Section 2.3 that one encounters problems when using spatially partially coherent illumination, because the reference point P (i.e. the pinhole) needs to be far removed from the sample, which means the sample may lie outside the field of view defined by the coherence width of the illumination. In the proposed method however, P can be on the edge of the sample, and one can define multiple reference points P (i.e. the different points of the star shaped mask) using the same data set $I(k, A)$, hence allowing for a large field of view even for narrow coherence widths. Because of these considerations, the proposed method was tested experimentally using spatially partially coherent light, as is discussed in the following section.

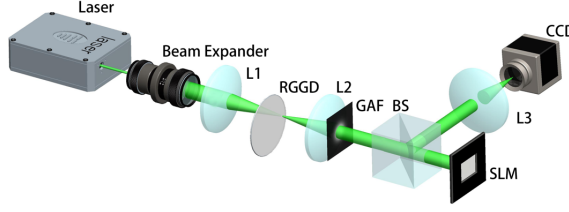


Figure 2.11: Illustration of the experimental setup. RGGD=Rotating Ground Glass Disk, GAF=Gaussian Amplitude Filter, BS=Beam Splitter, SLM=Spatial Light Modulator.

2.6.3. Experimental results

To test the method in a proof-of-principle experiment⁴, the setup as shown in Fig. 2.11 was used. In this setup, we use green laser light ($\lambda = 532\text{nm}$) which is expanded in the beam expander. The beam is made partially coherent by focusing it on a rotating ground-glass disk, after which the beam is approximately collimated again with a second lens. By varying the position of the first lens, we can vary the spot size on the rotating disk, thereby changing the degree of coherence. The resulting beam has a Gaussian correlation structure. The partially coherent beam is incident on a reflective liquid crystal phase-only SLM (HOLOEYE-GAEA-VIS, 3840×2160 resolution, $3.74 \mu\text{m}$ pixel size), on which a pattern is assigned that serves as the phase object $\psi(\mathbf{x})$ that is to be reconstructed. The radius of the object on the SLM is $R = 2.62\text{mm}$, see Fig. 2.12. Using a third lens with focal length $f = 15\text{cm}$, the modulated light is focused, and in the back focal plane of the lens, the intensity pattern is recorded with a CCD camera. In order to generate the through-focus data set, quadratic phase factors are added to the pattern assigned to the SLM. The defocus parameter $A = \frac{z_f R^2}{2\lambda f^2}$ (see Eq. (B.5)) is varied from -7 to 7 in 100 steps, which for the given parameters is equal to physically scanning through the focal field from $z_f = -2.45\text{cm}$ to $z_f = 2.45\text{cm}$.

With this experimental setup, we take two data sets each consisting of 100 through-focus images, where for each data set a different degree of coherence has been used. The coherence widths of the Gaussian-correlated beam with constant amplitude at the SLM-plane are $\sigma = 2.3\text{mm}$ (high coherence) and $\sigma = 0.5\text{mm}$ (low coherence). In Fig. 2.13 the non-iterative reconstructions are plotted, and compared to the simulated amplitudes which are obtained by multiplying the amplitude of $\psi(\mathbf{x} + \mathbf{P})$ with the Gaussian correlation function with respect to the reference point $\mathbf{P}$:

$$\begin{aligned} J(\mathbf{x} + \mathbf{P}, \mathbf{P}) &= \psi(\mathbf{x} + \mathbf{P})\psi(\mathbf{P})^* e^{-|\mathbf{x}+\mathbf{P}-\mathbf{P}|^2/2\sigma^2} \\ &\propto \psi(\mathbf{x} + \mathbf{P}) e^{-|\mathbf{x}|^2/2\sigma^2} \end{aligned} \quad (2.33)$$

It is thus shown that the coherence structure of the illuminating field limits the

⁴The experiment described in this section was performed by Xingyuan Lu at Soochow University under supervision of Chengliang Zhao. The candidate first performed a similar experiment in Delft to verify the method works for coherent illumination. The setup in Soochow was used to introduce partial coherence.

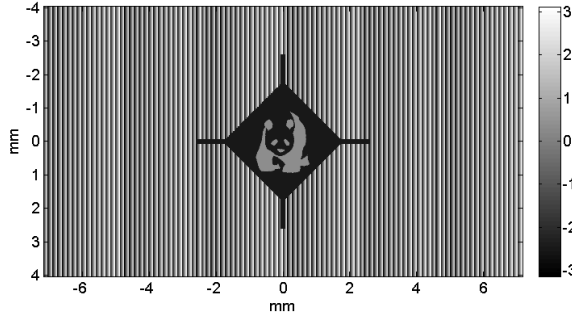


Figure 2.12: The image that is assigned to the SLM, which serves as the object that is to be reconstructed. The grayscale values denote the phase shift in radians.

field of view of the reconstruction (as explained in Section 2.3), and conversely, the degree of coherence can be inferred directly from the non-iterative reconstruction.

For each $\mathbf{P}$, a coherence function $J_0(\mathbf{x}) = J(0, \mathbf{x})$ ‘illuminates’ a different part of the object $\psi(\mathbf{x})$. In the example of Eq. (2.33), we have $J_0(\mathbf{x}) = e^{-|\mathbf{x}|^2/2\sigma^2}$. So by shifting $\mathbf{P}$ around, we ‘illuminate’ different parts of the object, thereby extending the field of view, which shares similarities with ptychography. In Fig. 2.14 it is shown how the field of view can be extended by considering multiple reference points $\mathbf{P}$. In order to synthesize an object reconstruction $\psi(\mathbf{x})$ with an extended field of view from the set of reconstructions for different $\mathbf{P}$, we use the factorization method as is explained in Eqs. (3.23) and (3.24), where we factorize the ‘exit waves’

$$\Psi_j(\mathbf{x}) = \psi(\mathbf{x})J_0(\mathbf{x} - \mathbf{P}_j) \quad (2.34)$$

into $\psi(\mathbf{x})$ and $J_0(\mathbf{x})$. These $\Psi_j(\mathbf{x})$ are the reconstructions we obtain using the non-iterative method as shown in Fig. 2.14, except that we have shifted them by $\mathbf{P}$ to the center. Also, we have multiplied the reconstructions with a window function to eliminate the artifacts of the non-iterative reconstruction, as shown in the top two rows of Fig. 2.15. Note that in the coherence function J_0 , the values of the prefactors $\psi(\mathbf{P})^*$ (see Eq. (2.33)) are absorbed. In this case this is possible because the ‘scanned positions’ do not overlap (see Fig. 2.15, bottom row, third figure from the left), so the factorization is not unique. To still get a plausible factorization, one needs a sensible initial guess when applying the factorization method of Eqs. (3.23) and (3.24).

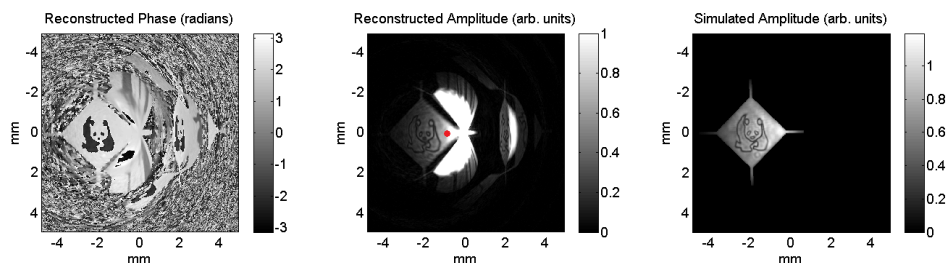
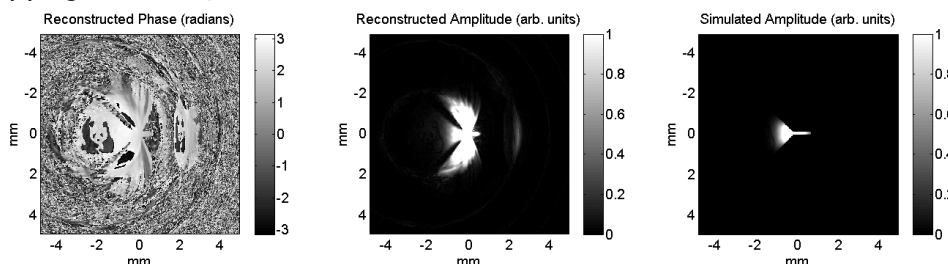
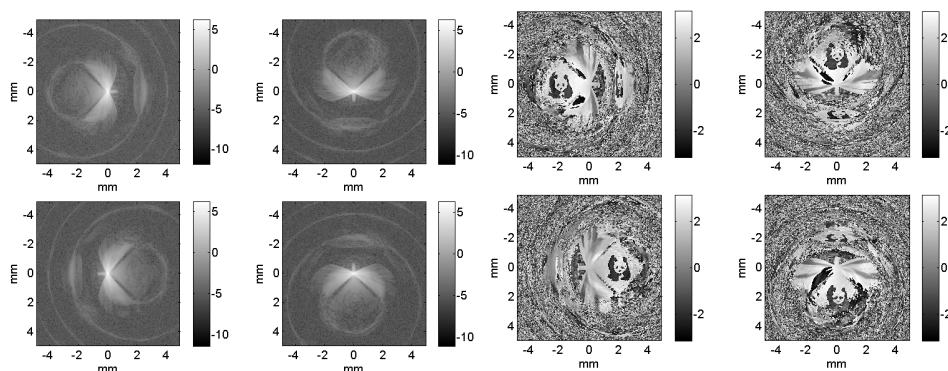
(a) High coherence, $\sigma = 2.3 \text{ mm}$.(b) Low coherence, $\sigma = 0.5 \text{ mm}$.

Figure 2.13: Non-iterative reconstructions for different degrees of coherence. The simulated amplitudes are obtained by multiplying the shifted object amplitude with a Gaussian function $e^{-|x|^2/2\sigma^2}$. Because the amplitude of the non-iterative reconstruction blows up for small $|x|$, their colorbars have been truncated. To ensure a fair comparison, the cut-off value for each plot is determined by its value at the position of the red dot as indicated in the top image.



(a) Amplitude (logarithm) in arbitrary units (b) Phase in radians

Figure 2.14: Non-iterative reconstructions for low coherence for various choices of the reference point P . It is illustrated how the coherence width affects the field of view, and how using multiple reference points P can help in creating a more complete picture of the object.

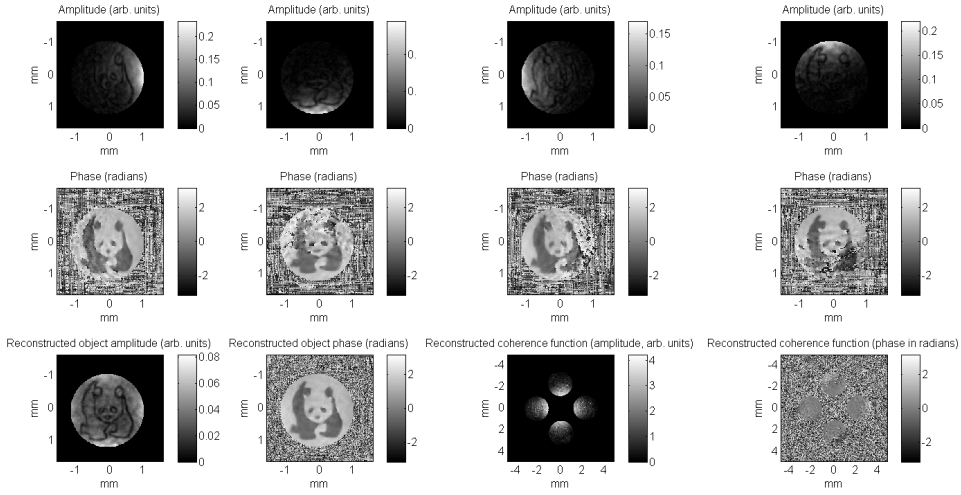


Figure 2.15: Demonstration of how reconstructions for different reference points P_j can be synthesized into a reconstruction of $\psi(x)$ with an extended field of view. In the top two rows are the amplitude and phase of the non-iterative reconstructions $\Psi_j(x)$ for four different P_j that correspond to the four different protrusions of the star-shaped mask as shown in Fig. 2.14. In the bottom row are the amplitude and phase of the reconstructed object $\psi(x)$ and coherence function $J_0(x)$. Note that the coherence function is reconstructed only in the regions that are covered by $\psi(x + P_j)$.

2.7. Summary

- We observed that measuring the far field intensity pattern $I(k)$ of a field $\psi(x)$ is equivalent to measuring its **autocorrelation function**. Therefore, in order to non-iteratively reconstruct $\psi(x)$ from $I(k)$, we must somehow make sure that the autocorrelation is shaped in such a way that it allows for straightforward extraction of $\psi(x)$.
- For example, in **Fourier Transform Holography** a pinhole is introduced so that the autocorrelation function contains an isolated copy of the original object $\psi(x)$, as well as its twin image $\psi(-x)^*$.
- We noted that if the extraction of $\psi(x)$ from the autocorrelation function is a linear operation, the reconstruction method is robust to **spatial partial coherence**, and the degree of coherence affects the field of view of the reconstruction.
- This idea is extended to **three-dimensional autocorrelation functions**: by modulating $\psi(x)$ with a phase function $e^{-2\pi i A f(x)}$, where A is an optical parameter that one can vary freely, we can obtain a three-dimensional data set $I(k, A)$, which in turn gives a three-dimensional autocorrelation function. If the modulation function $f(x)$ is chosen properly, one can identify a surface in the 3D autocorrelation function from which $\psi(x)$ or its twin image can be extracted directly.

- We have found that $f(\mathbf{x}) = |\mathbf{x}|^K$ **with** $K \in (0, 1)$ **is a suitable modulation function**. The reconstruction formula is given by

$$R(\mathbf{x}) = \sum_A \mathcal{F}_k^{-1} \{I(\mathbf{k}, A)\}(\mathbf{x}) e^{2\pi i A f(\mathbf{x})} \hat{H}(A), \quad (2.35)$$

where $\hat{H}(A)$ is a sampling window that mitigates the effects of a finite sampling range of A .

- In the limit of $K \rightarrow 0$, and with an appropriately chosen sampling window $\hat{H}(A)$, this method reduces to **phase shifting holography** or **quantitative Zernike phase-contrast microscopy**.
- We also found that a non-iterative reconstruction is possible using **focus variation** ($f(\mathbf{x}) = |\mathbf{x}|^2$) **and a star-shaped mask**. In this case, the reconstruction formula is given by

$$R(\mathbf{x}) = \sum_A \mathcal{F}_k^{-1} \{I(\mathbf{k}, A)\}(\mathbf{x}) e^{2\pi i A s(\mathbf{x})} \hat{H}(A), \quad (2.36)$$

where $s(\mathbf{x}) = |\mathbf{x}|^2 + 2\mathbf{P} \cdot \mathbf{x}$, where $\mathbf{P}$ is a reference point that must lie on one of the protrusions of the star-shaped mask.

References

- [1] A. Konijnenberg, W. Coene, and H. Urbach, *Non-iterative phase retrieval by phase modulation through a single parameter*, [Ultramicroscopy](#) **174**, 70 (2017).
- [2] A. Konijnenberg, X. Lu, L. Liu, W. M. J. Coene, C. Zhao, and H. P. Urbach, *Non-iterative method for phase retrieval and coherence characterization by focus variation using a fixed star-shaped mask*, [Optics Express](#) **26**, 9332 (2018).
- [3] A. J. D'Alfonso, A. J. Morgan, A. V. Martin, H. M. Quiney, and L. J. Allen, *Fast deterministic approach to exit-wave reconstruction*, [Physical Review A](#) **85** (2012), 10.1103/physreva.85.013816.
- [4] R. Bates and J. Rodenburg, *Sub-ångström transmission microscopy: A fourier transform algorithm for microdiffraction plane intensity information*, [Ultramicroscopy](#) **31**, 303 (1989).
- [5] P. Li, T. B. Edo, and J. M. Rodenburg, *Ptychographic inversion via wigner distribution deconvolution: Noise suppression and probe design*, [Ultramicroscopy](#) **147**, 106 (2014).
- [6] Y. Shao, X. Lu, A. Konijnenberg, C. Zhao, Y. Cai, and H. Urbach, *Spatial coherence measurement and partially coherent diffractive imaging using self-referencing holography*, [Optics Express](#) **26**, 4479 (2018).

- [7] M. Lurie, *Fourier-transform holograms with partially coherent light: Holographic measurement of spatial coherence*, [Journal of the Optical Society of America](#) **58**, 614 (1968).
- [8] J. M. Rodenburg and R. H. T. Bates, *The theory of super-resolution electron microscopy via wigner-distribution deconvolution*, [Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences](#) **339**, 521 (1992).
- [9] E. Wolf, *New theory of partial coherence in the space–frequency domain part i: spectra and cross spectra of steady-state sources*, [Journal of the Optical Society of America](#) **72**, 343 (1982).
- [10] C. W. McCutchen, *Generalized aperture and the three-dimensional diffraction image*, [Journal of the Optical Society of America](#) **54**, 240 (1964).
- [11] C. J. R. Sheppard, S. S. Kou, and J. Lin, *The green-function transform and wave propagation*, [Frontiers in Physics](#) **2** (2014), 10.3389/fphy.2014.00067.
- [12] G. Shabtay, *Three-dimensional beam forming and ewald's surfaces*, [Optics Communications](#) **226**, 33 (2003).
- [13] P. Gao, B. Yao, I. Harder, N. Lindlein, and F. Torcal-Milla, *Phase-shifting zernike phase contrast microscopy for quantitative phase measurement*, [Optics Letters](#) **36**, 4305 (2011).

3

Iterative phase retrieval

We discuss iterative algorithms for phase retrieval. Starting from basic Coherent Diffractive Imaging (CDI) methods which use a single intensity constraint and a support constraint, we review the different ways in which the phase retrieval problem can be formulated. Then we consider the case of ptychography, where multiple intensity constraints are used. We will see how we can apply CDI methods to ptychography, and what variations in algorithmic approaches we can construct. In addition to examining the large collection of already existing ptychographic algorithms, we propose new variants which help in accelerating the convergence, and improve the noise-robustness of simultaneous probe and object retrieval.

We have seen in Chapter 2 how one can reconstruct an image of a sample non-iteratively from one or multiple diffraction intensity patterns. The advantages of non-iterative reconstruction methods over iterative methods are that the non-iterative reconstruction tends to be faster because it is less computationally expensive, the reconstruction is unambiguous (an iterative algorithm might converge to different solutions for different initial guesses), and under certain conditions (as explained in Section 2.3) it is robust to spatial partial coherence.

Nevertheless, there are also disadvantages to the non-iterative methods which may motivate the study of iterative methods. One disadvantage is that for non-iterative methods one requires certain assumptions to hold, but these assumptions tend to be only true in approximation. For example, in the case of Fourier Transform Holography, one can reconstruct the object perfectly in case the pinhole can be modeled as a delta function, but in practice, the pinhole must always be sufficiently large in order to transmit enough light. The resolution of the non-iterative reconstruction will therefore be limited by the size of the pinhole and accurate knowledge of the shape of the pinhole, but it can be improved by applying iterative reconstruction methods that do not rely on the assumption that the pinhole is small or that the shape of the pinhole is accurately known [3]. Generally speaking, iterative algorithms allow for more flexibility. For example, in the case of ptychography, there are many variations of the basic reconstruction algorithm for object reconstruction, allowing for probe reconstruction [4–6], probe position correction [4, 7–11], reconstruction under partially coherent illumination [12, 13], reconstruction of thick samples [14], interpolation [15] or extrapolation [16] of diffraction patterns, etc.

Another disadvantage is that because the non-iterative methods require more special conditions to hold than iterative methods, non-iterative methods tend to require more complicated experimental setups, involving e.g. specially designed masks or phase modulators. The iterative methods bring more computational complexity, but this is a trade-off that is tempting to make, given the increasing computational power that is at our disposal.

3.1. Phase retrieval methods using a single intensity pattern

Just like in the non-iterative case, we aim to reconstruct a transmission function (or reflection function) $\psi(\mathbf{x})$ from its far field intensity $I(\mathbf{k}) = |\hat{\psi}(\mathbf{k})|^2$. Clearly, the intensity constraint $I(\mathbf{k})$ is not sufficient to uniquely define $\psi(\mathbf{x})$, since a trivial reconstruction $f(\mathbf{x})$ that would always satisfy the intensity constraint is given by $f(\mathbf{x}) = \mathcal{F}^{-1}\{\sqrt{I(\mathbf{k})}\}(\mathbf{x})$. This means we need an additional constraint to solve the phase retrieval problem. Typically, in case we use only a single intensity constraint, this would be a support constraint [17]: we assume a priori that there is a region γ outside which $\psi(\mathbf{x}) = 0$. One can define the following function indicating the

support constraint

$$s(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \gamma \\ 0 & \text{if } \mathbf{x} \notin \gamma. \end{cases} \quad (3.1)$$

In the following sections, we will discuss three different ways to formulate the phase retrieval problem, which will lead to different insights and different ways to solve it.

3.1.1. Intersection of sets

We can define two sets corresponding to the two constraints. Let S denote the set containing all functions $f(\mathbf{x})$ that satisfy the support constraint

$$S = \{f(\mathbf{x}) : s(\mathbf{x})f(\mathbf{x}) = f(\mathbf{x})\}. \quad (3.2)$$

Let M denote the set containing all functions $f(\mathbf{x})$ that satisfy the intensity constraint (also called the modulus constraint)

$$M = \{f(\mathbf{x}) : |\hat{f}(\mathbf{k})| = |\hat{\psi}(\mathbf{k})|\}, \quad (3.3)$$

where $\hat{f}(\mathbf{k})$ denotes the Fourier transform of $f(\mathbf{x})$. Since $\psi(\mathbf{x})$ satisfies both constraints, we can formulate the phase retrieval problem as follows:

$$\text{Find } f(\mathbf{x}) \in S \cap M. \quad (3.4)$$

The most straightforward method to find the intersection is to start with some initial guess $f_0(\mathbf{x})$, and repeatedly applying the two constraints. Applying a constraint can be described mathematically by introducing a projection operator $\mathcal{P}$. The projection operator for the support constraint is

$$\mathcal{P}_S : f(\mathbf{x}) \rightarrow s(\mathbf{x})f(\mathbf{x}). \quad (3.5)$$

The projection operator for the modulus constraint is

$$\mathcal{P}_M : f(\mathbf{x}) \rightarrow \mathcal{F}^{-1} \left\{ \frac{\hat{f}(\mathbf{k})}{|\hat{f}(\mathbf{k})|} |\hat{\psi}(\mathbf{k})| \right\}(\mathbf{x}). \quad (3.6)$$

The most straightforward algorithm, called the **Alternating Projections** (AP) algorithm, can be formulated as follows:

$$f_{n+1}(\mathbf{k}) = \mathcal{P}_S \mathcal{P}_M f_n(\mathbf{k}). \quad (3.7)$$

This can be visualized as shown in Fig. 3.1a. This iterative scheme would converge to the correct solution if the two constraint sets S and M were convex. The support constraint set S is convex: if $f_1(\mathbf{x}) \in S$ and $f_2(\mathbf{x}) \in S$, then $\lambda f_1(\mathbf{x}) + (1-\lambda)f_2(\mathbf{x}) \in S$, where $\lambda \in [0, 1]$. However, the modulus constraint set M is not convex: for example, if $|\hat{\psi}| = 1$, then $\hat{f}_1 = 1$ and $\hat{f}_2 = -1$ both lie in M , but the convex combination $\frac{1}{2}(\hat{f}_1 + \hat{f}_2) = 0$ does not. This can be visualized as shown in Fig. 3.1b. One can see that the algorithm can stagnate at a wrong $f(\mathbf{x})$. Therefore, we need more sophisticated algorithms that avoid stagnation. Examples of such algorithms are [18, 19]:

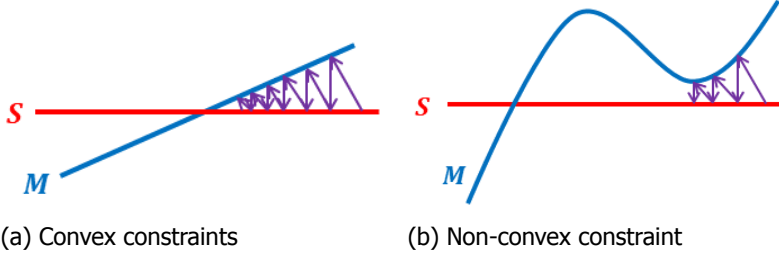


Figure 3.1: Illustration of the convergence behaviour of the Alternating Projections algorithm for convex and non-convex constraints.

- the **Hybrid Input-Output** (HIO) algorithm [20]

$$f_{n+1}(\mathbf{k}) = [\mathcal{P}_S \mathcal{P}_M + (I - \mathcal{P}_S)(I - \beta \mathcal{P}_M)]f_n(\mathbf{k}), \quad (3.8)$$

- the **Difference Map** (DM) algorithm [21]

$$f_{n+1}(\mathbf{k}) = \left[I + \beta \mathcal{P}_S \left(\left[1 + \frac{1}{\beta} \right] \mathcal{P}_M - \frac{1}{\beta} I \right) - \mathcal{P}_M \left(\left[1 - \frac{1}{\beta} \right] \mathcal{P}_S + \frac{1}{\beta} I \right) \right] f_n(\mathbf{k}), \quad (3.9)$$

- the **Relaxed Averaged Alternating Reflections** (RAAR) algorithm [22]

$$f_{n+1}(\mathbf{k}) = [2\beta \mathcal{P}_S \mathcal{P}_M + (1 - 2\beta) \mathcal{P}_M + \beta(I - \mathcal{P}_S)]f_n(\mathbf{k}). \quad (3.10)$$

Here, I denotes the identity operator, and β is a feedback parameter that is typically chosen to be around 0.9 for HIO [23], but one can try to find per application which value gives the best result. Note that for $\beta = 1$ all the algorithms reduce to the same update scheme

$$f_{n+1}(\mathbf{k}) = \frac{\mathcal{R}_S \mathcal{R}_M + I}{2} f_n(\mathbf{k}), \quad (3.11)$$

where $\mathcal{R}$ denotes the reflection operator

$$\begin{aligned} \mathcal{R} &= \mathcal{P} + (\mathcal{P} - I) \\ &= 2\mathcal{P} - I. \end{aligned} \quad (3.12)$$

So intuitively, one can say that these algorithms avoid stagnation by reflecting around the constraints instead of projecting onto them. To prevent the algorithm from becoming too unstable, one averages the reflected function $\mathcal{R}_S \mathcal{R}_M f$ with the current estimate of the function f (as in Eq. (3.11)), and one can introduce further relaxations through the parameter β . Typically, stagnation-avoiding algorithms such as HIO, DM, or RAAR are used for the initial convergence of the algorithm, but for the final refinement of the reconstruction, several iterations of AP are applied.

3.1.2. Cost minimization

If $f(\mathbf{x})$ is an estimate of the object, then we can define the estimated far field as

$$\hat{f}(\mathbf{k}) = \sum_{\mathbf{x}} f(\mathbf{x}) s(\mathbf{x}) e^{-2\pi i \mathbf{x} \cdot \mathbf{k}}. \quad (3.13)$$

We use a sum rather than an integral because in practice the measurement and the reconstruction will always be pixelated. We want to find that $f(\mathbf{x})$ for which the estimated far field amplitude is most similar to the measured amplitude $|\hat{\psi}(\mathbf{k})|$. Mathematically, we can introduce a cost functional

$$L[f(\mathbf{x})] = \sum_{\mathbf{k}} (|\hat{f}(\mathbf{k})| - |\hat{\psi}(\mathbf{k})|)^2, \quad (3.14)$$

which we want to minimize (an explanation of why this particular cost function is chosen is given in Chapter 4). The most straightforward method to do this is by using the **gradient descent** or **steepest descent** method. One can demonstrate that if $f(\mathbf{x})$ is a complex-valued optimization variable, and L is a real-valued cost-function, the direction of steepest ascent is given by the Wirtinger derivative $\frac{\partial L}{\partial f(\mathbf{x})^*}$ [24]. Therefore, for a certain step size μ the update scheme is given by

$$f_{n+1}(\mathbf{x}) = f_n(\mathbf{x}) - \mu \frac{\partial L}{\partial f_n(\mathbf{x})^*}. \quad (3.15)$$

One can calculate from Eqs. (3.13) and (3.14) that the Wirtinger derivative is given by

$$\frac{\partial L}{\partial f(\mathbf{x})^*} = s(\mathbf{x}) \left(f_n(\mathbf{x}) - \mathcal{F}^{-1} \left\{ \frac{\hat{f}(\mathbf{k})}{|\hat{f}(\mathbf{k})|} |\hat{\psi}(\mathbf{k})| \right\}(\mathbf{x}) \right). \quad (3.16)$$

Assuming that $f_n(\mathbf{x}) = s(\mathbf{x}) f_n(\mathbf{x})$, we see that for a step size of $\mu = 1$ this update scheme is identical to the previously discussed Alternating Projection scheme [20]. Because in each iteration the Fourier-space reconstruction error (i.e. the value of the cost functional) reduces, this scheme is also referred to as the Error Reduction (ER) algorithm.

The advantages of formulating the phase retrieval problem in this manner are that it gives insight in what it means to vary the step size μ (which can be useful for accelerating the rate of convergence, or for refining the reconstruction in case noise is present in the measurement, as will be discussed in Section 3.2.2), it allows us to experiment with different kinds of cost functionals (for example a maximum-likelihood cost function in case noise is present, as will be discussed in Section 4.1), and it allows us to experiment with different minimization schemes [19] (such as the Conjugate Gradient method).

In Fig. 3.1 we illustrated using the projection-based view how the non-convex modulus constraint can cause the Alternating Projection algorithm to stagnate. The

stagnation of the Steepest Descent method (which is equivalent to the Alternating Projections method) can be explained using the cost-functional-based view by stating that there are local minima in the cost functional landscape, in which the Steepest Descent algorithm can easily get trapped. Using this picture, one can intuitively imagine how such stagnation can be prevented by introducing ‘momentum’. A ball rolling down a hill at high velocity does not get stuck in shallow pits due to its momentum. Similarly, if we imagine our object estimate $f_n(\mathbf{x})$ as ‘rolling down’ the cost functional landscape when we are performing updates, we can add ‘momentum’ to the updates to prevent stagnation [25]. This idea is applied to Ptychographic phase retrieval in [26].

3.1.3. Convex relaxation and rank minimization

What makes the phase retrieval problem difficult is the fact that it is non-convex: if $\psi_1(\mathbf{x})$ and $\psi_2(\mathbf{x})$ give the same intensity $I(\mathbf{k})$, then the convex combination $\lambda\psi_1(\mathbf{x}) + (1 - \lambda)\psi_2(\mathbf{x})$, $\lambda \in (0, 1)$, generally does not. However, we can write

$$\begin{aligned} I(\mathbf{k}) &= \left| \int \psi(\mathbf{x}) e^{-2\pi i \mathbf{x} \cdot \mathbf{k}} d\mathbf{x} \right|^2 \\ &= \iint \psi(\mathbf{x}) \psi(\mathbf{y})^* e^{-2\pi i (\mathbf{x} - \mathbf{y}) \cdot \mathbf{k}} d\mathbf{x} d\mathbf{y}, \\ &= \iint J(\mathbf{x}, \mathbf{y}) e^{-2\pi i (\mathbf{x} - \mathbf{y}) \cdot \mathbf{k}} d\mathbf{x} d\mathbf{y}, \end{aligned} \quad (3.17)$$

where $J(\mathbf{x}, \mathbf{y}) = \psi(\mathbf{x})\psi(\mathbf{y})^*$. Note that $J(\mathbf{x}, \mathbf{y})$ is a four-dimensional function that corresponds to the mutual intensity function of a coherent field (see Eq. (2.3)). If our optimization variable is $J(\mathbf{x}, \mathbf{y})$ instead of $\psi(\mathbf{x})$, the problem becomes convex, because if $J_1(\mathbf{x}, \mathbf{y})$ and $J_2(\mathbf{x}, \mathbf{y})$ give the same $I(\mathbf{k})$, then the convex combination $\lambda J_1(\mathbf{x}, \mathbf{y}) + (1 - \lambda)J_2(\mathbf{x}, \mathbf{y})$ does so as well. However, this convex combination in general does not correspond to a coherent field, but rather to a partially coherent field. In order to obtain a coherent field, one must minimize the rank of $J(\mathbf{x}, \mathbf{y})$, which means minimizing the number of modes $J(\mathbf{x}, \mathbf{y})$ is expanded in as according to Eq. (2.3). So mathematically speaking, we applied a convex relaxation by lifting the problem in a higher dimension by going from $\psi(\mathbf{x})$ to $J(\mathbf{x}, \mathbf{y})$, and the phase retrieval problem becomes a rank minimization problem [27, 28]. Physically speaking, we have extended the solution space from all coherent solutions to all partially coherent solutions, and we aim to find the most coherent solution within this space. However, in this thesis this method will be of less importance than the projections method and the cost-functional minimization method.

3.2. Ptychography

While methods such as HIO help to prevent stagnation, these phase retrieval methods still have significant limitations: they only work if a sample with a known finite support (although methods such as Shrinkwrap try to alleviate this requirement [23]) is illuminated with a flat wave front. Moreover, while HIO can give good results for very low noise levels, it tends to diverge for moderate noise levels. Therefore,

we require a more robust method, which is ptychography, where multiple intensity patterns are recorded by shifting the illumination, and the set of intensity patterns is used to robustly reconstruct an image of the sample. Another advantage is that ptychography can extend the field of view of the reconstruction without sacrificing resolution, although in principle that could also be achieved by applying HIO separately multiple times if the illumination is a step function.

In ptychography, we illuminate an object $O(\mathbf{x})$ with a probe function $P(\mathbf{x})$ which is shifted to different positions $\mathbf{X}_j$ across the sample. These probe positions are chosen such that the probes for adjacent positions overlap. For each probe position we measure an intensity pattern $I_j(\mathbf{k})$. Assuming the sample is thin, the exit wave $\psi_j(\mathbf{x})$ is given by

$$\psi_j(\mathbf{x}) = O(\mathbf{x})P(\mathbf{x} - \mathbf{X}_j), \quad (3.18)$$

although experimentally it might be more convenient to keep the illumination fixed and shift the sample. However, if we assume far-field propagation this does not affect the intensity patterns, which are given by

$$I_j(\mathbf{k}) = |\hat{\psi}_j(\mathbf{k})|^2. \quad (3.19)$$

In the first ptychographic reconstruction algorithm, called the Ptychographic Iterative Engine (PIE) [29, 30], this set of intensity measurements is used to reconstruct the object $O(\mathbf{x})$, assuming the probe $P(\mathbf{x})$ and the probe positions $\mathbf{X}_j$ are known. Later, the algorithm was extended to Extended PIE (ePIE), which reconstructs both the object and the probe simultaneously [4–6]. Other extensions aim to deal with correcting uncertainties in the probe positions $\mathbf{X}_j$ [4, 7–11, 31, 32], partially coherent illumination [12, 13], or thick samples [14].

3.2.1. Basic reconstruction algorithms

Having seen how we can solve the single-intensity pattern phase retrieval problem using different approaches, we will now see how we can use these approaches to solve the ptychographic phase retrieval problem.

Intersection of sets

Let us denote the collection of estimated exit waves $f_j(\mathbf{x})$ as $\mathbf{f}$, and let $|\hat{\psi}_j(\mathbf{k})|^2$ be the measured (noise-free) intensity patterns. We can define two sets corresponding to the two constraints that we have. Let S denote the set of all $\mathbf{f}$ that can be factorized into an object and a probe

$$S = \{\mathbf{f} : \text{there exists an } O(\mathbf{x}) \text{ such that } f_j(\mathbf{x}) = O(\mathbf{x})P(\mathbf{x} - \mathbf{X}_j) \text{ for all } j\}. \quad (3.20)$$

Note that we assume here that $P(\mathbf{x})$ and $\mathbf{X}_j$ are known. Let M denote the set of $\mathbf{f}$ whose far field intensities match the measurements

$$M = \{\mathbf{f} : |\hat{f}_j(\mathbf{k})| = |\hat{\psi}_j(\mathbf{k})| \text{ for all } j\}. \quad (3.21)$$

We can define the projection operators $\mathcal{P}_M$ and $\mathcal{P}_S$, that project f onto M and S respectively [6]

$$\begin{aligned} (\mathcal{P}_S f)_j(\mathbf{x}) &= O_f(\mathbf{x})P(\mathbf{x} - \mathbf{X}_j), \\ (\mathcal{P}_M f)_j(\mathbf{x}) &= \mathcal{F}^{-1} \left\{ \frac{\hat{f}_j(\mathbf{k})}{|\hat{f}_j(\mathbf{k})|} |\hat{\psi}_j(\mathbf{k})| \right\}(\mathbf{x}), \end{aligned} \quad (3.22)$$

where

$$O_f(\mathbf{x}) = \frac{\sum_{j'} f_{j'}(\mathbf{x})P(\mathbf{x} - \mathbf{X}_{j'})^*}{\alpha + \sum_{j'} |P(\mathbf{x} - \mathbf{X}_{j'})|^2}, \quad (3.23)$$

where α is a small parameter to prevent division by 0. In case the probe $P(\mathbf{x})$ is unknown, we can estimate the probe from the exit waves f and the object $O(\mathbf{x})$ as follows

$$P_f(\mathbf{x}) = \frac{\sum_{j'} f_{j'}(\mathbf{x} + \mathbf{X}_{j'})O(\mathbf{x} + \mathbf{X}_{j'})^*}{\alpha + \sum_{j'} |O(\mathbf{x} + \mathbf{X}_{j'})|^2}. \quad (3.24)$$

If both the object and the probe are unknown, they can both be updated using the updated exit waves f by alternately applying Eqs. (3.23) and (3.24) several times, which would constitute a single application of the projection operator $\mathcal{P}_S$. $P_f(\mathbf{x})$ and $O_f(\mathbf{x})$ are computed using the estimates $O_f(\mathbf{x})$ and $P_f(\mathbf{x})$ from the previous iteration respectively. With these projection operators, one can define algorithms such as AP, HIO, DM, and RAAR using Eqs. (3.7), (3.8), (3.9), (3.10).

Cost functional minimization

We can define a cost functional $L[O_n(\mathbf{x}), P_n(\mathbf{x})]$ that for the n^{th} object estimate $O_n(\mathbf{x})$ and probe estimate $P_n(\mathbf{x})$ quantifies the difference between the estimated far field amplitudes $|\hat{f}_j(\mathbf{k})|$ and the measured far field amplitudes $|\hat{\psi}_j(\mathbf{k})|$

$$L[O_n(\mathbf{x}), P_n(\mathbf{x})] = \sum_j \sum_{\mathbf{k}} (|\hat{f}_{n,j}(\mathbf{k})| - |\hat{\psi}_j(\mathbf{k})|)^2, \quad (3.25)$$

where

$$\hat{f}_{n,j}(\mathbf{k}) = \sum_{\mathbf{x}} O_n(\mathbf{x})P_n(\mathbf{x} - \mathbf{X}_j)e^{-2\pi i \mathbf{x} \cdot \mathbf{k}}. \quad (3.26)$$

If we want to minimize L using the gradient descent scheme, the updates of $O_n(\mathbf{x})$ and $P_n(\mathbf{x})$ (if $P(\mathbf{x})$ is unknown) are defined as

$$\begin{aligned} O_{n+1}(\mathbf{x}) &= O_n(\mathbf{x}) - \mu_O \frac{\partial L}{\partial O_n(\mathbf{x})^*} \\ &= O_n(\mathbf{x}) - \mu_O \sum_j P(\mathbf{x} - \mathbf{X}_j)^* (f_{n,j}(\mathbf{x}) - f_{n,j}^{\text{upd}}(\mathbf{x})) \\ P_{n+1}(\mathbf{x}) &= P_n(\mathbf{x}) - \mu_P \frac{\partial L}{\partial P_n(\mathbf{x})^*} \\ &= P_n(\mathbf{x}) - \mu_P \sum_j O(\mathbf{x} + \mathbf{X}_j)^* (f_{n,j}(\mathbf{x} + \mathbf{X}_j) - f_{n,j}^{\text{upd}}(\mathbf{x} + \mathbf{X}_j)), \end{aligned} \quad (3.27)$$

where

$$f_{n,j}^{\text{upd}}(\mathbf{x}) = \mathcal{F}^{-1} \left\{ \frac{\hat{f}_{n,j}(\mathbf{k})}{|\hat{f}_{n,j}(\mathbf{k})|} |\hat{\psi}_j(\mathbf{k})| \right\}(\mathbf{x}). \quad (3.28)$$

μ is the step size of the respective updates. To prevent divergence, it is recommended to take

$$\begin{aligned} \mu_O &\leq \frac{1}{\max_{\mathbf{x}} \sum_j |P(\mathbf{x} - \mathbf{X}_j)|^2}, \\ \mu_P &\leq \frac{1}{\max_{\mathbf{x}} \sum_j |O(\mathbf{x} + \mathbf{X}_j)|^2}. \end{aligned} \quad (3.29)$$

Just like in the case of single-intensity pattern phase retrieval, one can choose different cost functions, and different minimization schemes [4, 19]. One can also obtain different update functions by choosing certain preconditioners or regularization schemes [19, 26], e.g.

$$O_{n+1}(\mathbf{x}) = O_n(\mathbf{x}) - \mu \frac{\sum_j P(\mathbf{x} - \mathbf{X}_j)^* (f_{n,j}(\mathbf{x}) - f_{n,j}^{\text{upd}}(\mathbf{x}))}{\alpha + \sum_j |P(\mathbf{x} - \mathbf{X}_j)|^2}, \quad (3.30)$$

where α is a small constant to prevent division by 0. This makes the update more similar to the alternating projections scheme, as one can see from the similarities with Eq. (3.23). In Fig. 3.2 the difference between the global alternating projections scheme and the global steepest descent scheme has been investigated, which is equivalent to investigating the effect of introducing a preconditioner as in Eq. (3.30). One finds that the final reconstruction error is the same for both approaches, but the alternating projections scheme converges faster.

Rank minimization

In [33], ptychography is performed using a convex relaxation, which is referred to as Convex Lifted Ptychography (CLP). To reduce the computational expense, another method is proposed where the solution space does not include all partially coherent fields, but rather only the partially coherent fields with a low rank, i.e. fairly coherent fields. This method is termed Low-Rank Ptychography (LRP). While in this method the problem is not fully convex, the relaxation does yield improved results at only moderately higher computational requirements. In this thesis, we will not focus on this method.

3.2.2. Global and Sequential updates

The ptychographic algorithms that were derived previously using projection-based methods or cost-functional minimization methods all involved global updates: in each update, all the measurements are used simultaneously to update the object estimate (and probe estimate, if the probe is unknown). However, in the first ptychographic reconstruction algorithm, PIE, the updates were sequential: the object

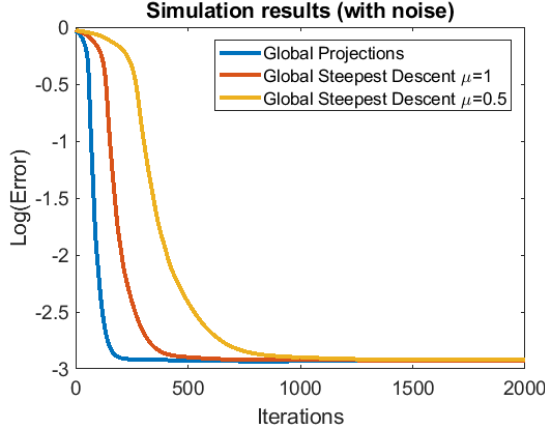


Figure 3.2: Comparison of the steepest descent and alternating projections global algorithms for noisy simulated data, as shown in Fig. 3.3. The global steepest descent algorithm has been tested for step sizes $\mu = \frac{1}{\max_{\mathbf{x}} \sum_j |P(\mathbf{x} - \mathbf{X}_j)|^2}$ and $\mu = 0.5 \frac{1}{\max_{\mathbf{x}} \sum_j |P(\mathbf{x} - \mathbf{X}_j)|^2}$.

is updated using the measurement of a single probe position $\mathbf{X}_j$, before updating the object again at the next probe position $\mathbf{X}_{j+1}$

$$O_n(\mathbf{x}) \rightarrow O_n(\mathbf{x}) + g(\mathbf{x}, \mathbf{X}_j) \mu P(\mathbf{x} - \mathbf{X}_j)^* (f_{n,j}^{\text{upd}}(\mathbf{x}) - f_{n,j}(\mathbf{x})), \quad (3.31)$$

where $n \rightarrow n+1$ when an update has been applied for all probe positions j . μ is the step size, and $g(\mathbf{x}, \mathbf{X}_j)$ is a function that can take different forms. In the original PIE update scheme [30], it is chosen to be

$$g(\mathbf{x}, \mathbf{X}_j) = \frac{|P(\mathbf{x} - \mathbf{X}_j)|}{\max_{\mathbf{x}} |P(\mathbf{x} - \mathbf{X}_j)|} \frac{1}{|P(\mathbf{x} - \mathbf{X}_j)|^2 + \alpha}, \quad (3.32)$$

where α is a small parameter to prevent division by 0. The reasoning behind this factor is that $\frac{1}{|P(\mathbf{x} - \mathbf{X}_j)|^2 + \alpha}$ divides out the $|P(\mathbf{x} - \mathbf{X}_j)|^2$ that is in $P(\mathbf{x} - \mathbf{X}_j)^* f_{n,j}(\mathbf{x})$ and $P(\mathbf{x} - \mathbf{X}_j)^* f_{n,j}^{\text{upd}}(\mathbf{x})$, while $\frac{|P(\mathbf{x} - \mathbf{X}_j)|}{\max_{\mathbf{x}} |P(\mathbf{x} - \mathbf{X}_j)|}$ makes sure that the area that is illuminated the most gets updated the most. However, later it was argued that $g(\mathbf{x}, \mathbf{X}_j) = \frac{1}{\max_{\mathbf{x}} |P(\mathbf{x} - \mathbf{X}_j)|^2}$ may be a more convenient choice [5, 34], which may have to do with the fact that it is more closely related to the steepest descent update for a single probe position (also see Eq. (3.27)). Therefore, this is the choice we will use here (or equivalently, we choose $g(\mathbf{x}, \mathbf{X}_j) = 1$ assuming that $P(\mathbf{x})$ is defined such that it is normalized to 1). A more detailed study of the role of the weighting function is found in [26]. We will look at the difference between the sequential and global update schemes, and introduce other possible variations of the ptychographic reconstruction algorithm. To quantify the reconstruction errors

E in our simulations, we use the quantity

$$E[O_n(\mathbf{x})] = \frac{\sum_{\mathbf{x}} |cO_n(\mathbf{x}) - O(\mathbf{x})|^2}{\sum_{\mathbf{x}} |O(\mathbf{x})|^2}, \quad (3.33)$$

where $O_n(\mathbf{x})$ denotes the n^{th} estimated object, $O(\mathbf{x})$ denotes the reference object, and c is a complex-valued constant that minimizes E . c is necessary to compensate for any global phase offsets in the reconstruction, i.e. $O_n(\mathbf{x}) = O(\mathbf{x})e^{i\theta}$ should still give $E = 0$. c is found by solving $\frac{dE}{dc} = 0$, which gives

$$c^* = \frac{\sum_{\mathbf{x}} O(\mathbf{x})^* O_n(\mathbf{x})}{\sum_{\mathbf{x}} |O_n(\mathbf{x})|^2}. \quad (3.34)$$

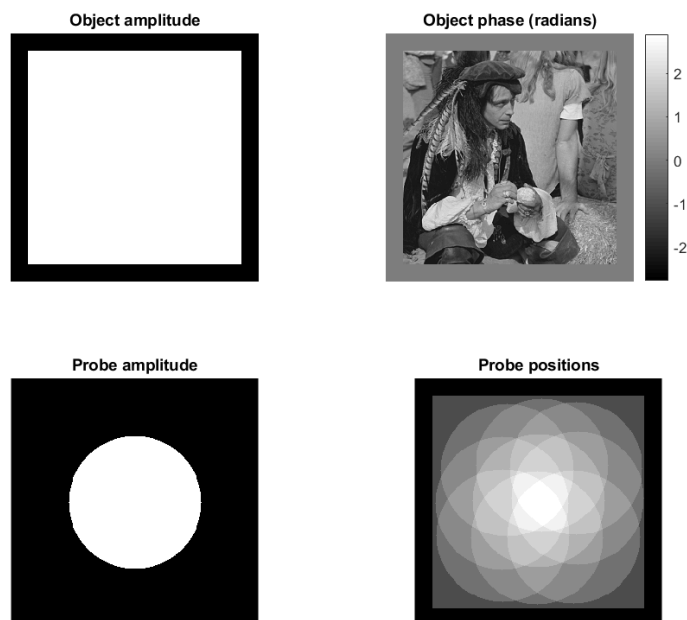
If c is only to correct for global phase offsets, one would require $|c| = 1$. However, since multiplying $O_n(\mathbf{x})$ with any non-zero global constant does not give a fundamentally different object reconstruction (since the same features will still be visible), it seems redundant to enforce any additional constraints on c .

Convergence rate and noise robustness of sequential and global updates

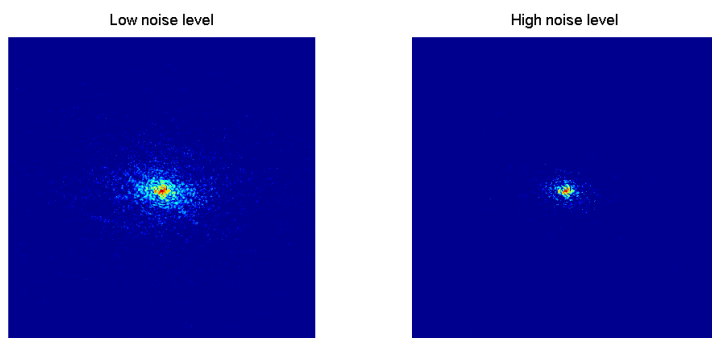
Using sequential updates (see Eq. (3.31)) or global updates (see Eqs. (3.7) and (3.22), or equivalently, Eq. (3.30) with $\mu = 1$) affects how quickly the algorithm converges and how sensitive it is to noise. If there is no noise, using a sequential update tends to converge faster because in one iteration (which means going through all probe positions once) a single point may be updated multiple times depending on how many probes cover that point, whereas in a global update scheme each point is only updated once per iteration. This is demonstrated in the simulation results shown in Fig. 3.4a, using simulated data shown in Fig. 3.3¹. However, if a significant amount of noise is present, a global update scheme tends to give better results because when updating a certain point, the noise from multiple measurements is averaged out [4], which is demonstrated in Fig. 3.4b. One way to achieve both fast convergence and robustness to noise is to use a sequential update scheme with an adaptive step size [35]: by using a large step size initially one achieves rapid convergence, and by reducing the step size in later iterations one achieves a better final reconstruction quality if noise is present. This is also shown in Fig. 3.4b.

One should keep in mind though that the convergence speed cannot be characterized completely just by counting the required number of iterations. Rather, one also has to consider the computation time per iteration. A global update scheme has the advantage that the updates for all exit waves can be computed in parallel, whereas in a sequential scheme one has to first update probe position j before one can update probe position $j + 1$ (although one can somewhat parallelize the sequential update scheme by simultaneously updating the fields at non-overlapping probe

¹The object and probe were chosen to be pure phase functions because it was more convenient to use the same object and probe that were used in the proof-of-principle experiments that were performed with a phase-only SLM (see Chapter 4). There is no further significance behind this particular choice.



(a) Object, probe, and probe positions. 3×3 probe positions are used (note that in many practical situations typically more probe positions are used).



(b) Two different levels of Poissonian noise (i.e. shot noise), plotted in logarithmic scale for visibility. In this section, only the data with a low noise level is used. In Section 4.3 both noise levels are used.

Figure 3.3: Simulated data that is used to compare the noise robustness and convergence speed of global and sequential algorithms.

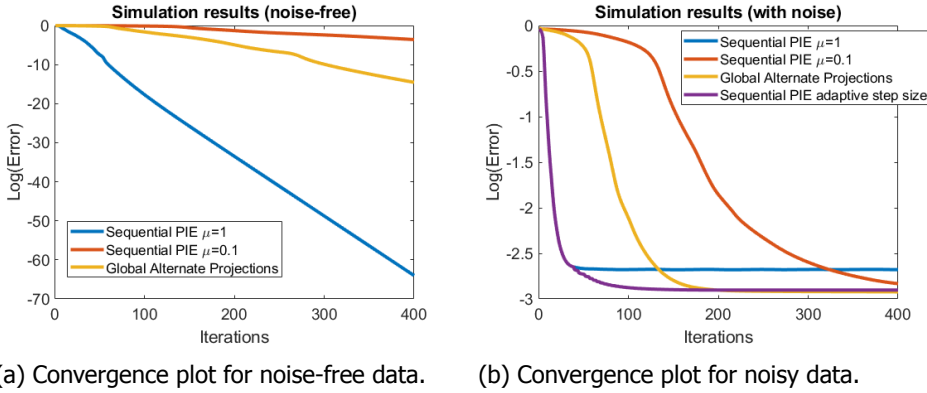


Figure 3.4: Comparison of sequential and global algorithms for noise-free and noisy simulated data, as shown in Fig. 3.3

positions). Global update schemes for HIO, DM, and RAAR have the disadvantage that they require more memory: one needs to store all the exit waves f_j , whereas in a sequential or global AP update scheme one in principle only needs to store one exit wave at a time (if one does not use parallel computation). The reason why a global AP update scheme does not need to store all the exit waves, is because one can compute sums such as $\sum_j P(\mathbf{x} - \mathbf{X}_j)^* (f_{n,j}(\mathbf{x}) - f_{n,j}^{\text{upd}}(\mathbf{x}))$ (which are required for the update) by adding one term at a time: i.e. once $f_{n,j}(\mathbf{x})$ and $f_{n,j}^{\text{upd}}(\mathbf{x})$ have been computed, they can be added to the sum and then immediately discarded from memory.

3.2.3. Combining HIO and ptychography²

We have seen that for single-shot phase retrieval, HIO outperforms the steepest descent and alternating projections algorithms, since it avoids stagnation in local minima. However, HIO is rather sensitive to noise, so M-HIO has been developed to make it more robust [36]. M-HIO works by setting a certain threshold level t , below which the HIO feedback function is set to 0.

Alternatively, one may instead introduce a ptychographic reconstruction scheme which is more reliable and robust. The ptychographic algorithms we discussed once more relied on the steepest descent and alternating projection schemes. One may wonder whether it is possible to improve ptychography by combining it with HIO. One straightforward way to do this, is to take the expression for the HIO update of Eq. (3.8), and substitute the projection operators with their ptychographic counterparts as defined in Eqs. (3.23) and (3.24), which yields a global ptychographic HIO algorithm [19]. Yet, there are still two problems:

²The following is original work by the candidate, part of which has been published in Ultramicroscopy **171**, 43 (2016) [2].

1. We have seen that sequential ptychographic algorithms tend to converge faster than global ptychographic algorithms. It is not obvious how the global ptychographic HIO algorithm can be adapted to become a sequential ptychographic HIO algorithm.
2. M-HIO is a convenient way to make HIO more noise robust, while retaining its stagnation-avoiding properties. It is not obvious how the global ptychographic HIO algorithm can be adapted to become a ptychographic M-HIO algorithm.

In this section, we propose a way to define sequential ptychographic (M)-HIO and global ptychographic M-HIO. The key point is that if the probe has a practically finite support (i.e. it has a negligibly low value outside a certain region), we can define a feedback function for each probe position. When applying the factorization constraint (i.e. decomposing the estimated exit waves in an object and a probe using either a global or a sequential update), we only update the exit waves, while leaving the feedback functions unaltered.

Sequential ptychographic (M)-HIO for accelerated convergence

Since sequential update schemes tend to converge faster (i.e. using fewer iterations) than global update schemes, and HIO tends to converge faster than AP (assuming there is not too much noise), it would be interesting to see whether one can design a sequential HIO update, as opposed to the global HIO update of Eq. (3.8) where the projection operators are substituted with their ptychographic counterparts as defined in Eqs. (3.23) and (3.24). To do this, let us reconsider the problem of single-intensity pattern phase retrieval with a support constraint, and rewrite Eq. (3.8) in terms of an updated exit wave $f_n^{\text{upd}}(\mathbf{x})$ inside the support γ (which for a probe may be defined as the region where $\frac{|P(\mathbf{x}-\mathbf{X})|}{\max_{\mathbf{x}}|P(\mathbf{x}-\mathbf{X})|}$ is higher than a certain threshold value), and a feedback function $B(\mathbf{x})$ outside the support [2]

$$f_{n+1} = \begin{cases} f_n^{\text{upd}}(\mathbf{x}) & \text{if } \mathbf{x} \in \gamma, \\ B(\mathbf{x}) & \text{if } \mathbf{x} \notin \gamma, \end{cases} \quad (3.35)$$

where $f_n^{\text{upd}}(\mathbf{x})$ is defined in Eq. (3.28), and

$$B(\mathbf{x}) = f_n(\mathbf{x}) - \beta f_n^{\text{upd}}(\mathbf{x}). \quad (3.36)$$

If we consider multiple probe positions $\mathbf{X}_j$, and we assume the probe $P(\mathbf{x})$ has a (practically) finite and known support, then we can define for each position an exit wave estimate $F_j(\mathbf{x})$ and a feedback function $B_j(\mathbf{x})$, which are updated as

$$\begin{aligned} F_{n,j}^{\text{upd}}(\mathbf{x}) &= f_{n,j}^{\text{upd}}(\mathbf{x})s(\mathbf{x} - \mathbf{X}_j), \\ B_{n+1,j}(\mathbf{x}) &= [B_{n,j}(\mathbf{x}) - \beta f_{n,j}^{\text{upd}}(\mathbf{x})][1 - s(\mathbf{x} - \mathbf{X}_j)], \end{aligned} \quad (3.37)$$

where $s(\mathbf{x})$ is defined in Eq. (3.1), and

$$f_{n,j}(\mathbf{x}) = F_{n,j}(\mathbf{x}) + B_{n,j}(\mathbf{x}). \quad (3.38)$$

One can find $F_{n+1,j}(\mathbf{x})$ from $F_{n,j}^{\text{upd}}(\mathbf{x})$ either through a global update like in Eqs. (3.23) and (3.24), i.e. apply several times

$$\begin{aligned} O_n(\mathbf{x}) &\rightarrow \frac{\sum_{j'} F_{n,j'}(\mathbf{x}) P_n(\mathbf{x} - \mathbf{X}_{j'})^*}{\alpha + \sum_{j'} |P_n(\mathbf{x} - \mathbf{X}_{j'})|^2}, \\ P_n(\mathbf{x}) &\rightarrow \frac{\sum_{j'} F_{n,j'}(\mathbf{x} + \mathbf{X}_{j'}) O_n(\mathbf{x} + \mathbf{X}_{j'})^*}{\alpha + \sum_{j'} |O_n(\mathbf{x} + \mathbf{X}_{j'})|^2}, \end{aligned} \quad (3.39)$$

and calculate the updated exit wave as

$$F_{n+1,j}(\mathbf{x}) = P_{n+1}(\mathbf{x} - \mathbf{X}_j) O_{n+1}(\mathbf{x}), \quad (3.40)$$

or through a sequential update where $O_n(\mathbf{x})$ and $P_n(\mathbf{x})$ are updated for each probe position

$$\begin{aligned} O'_n(\mathbf{x}) &= O_n(\mathbf{x}) + P_n(\mathbf{x} - \mathbf{X}_j)^* [F_{n,j}^{\text{upd}}(\mathbf{x}) - F_{n,j}(\mathbf{x})], \\ P'_n(\mathbf{x}) &= P_n(\mathbf{x}) + O_n(\mathbf{x} + \mathbf{X}_j)^* [F_{n,j}^{\text{upd}}(\mathbf{x} + \mathbf{X}_j) - F_{n,j}(\mathbf{x} + \mathbf{X}_j)], \end{aligned} \quad (3.41)$$

where the prime ' indicates that this update is an intermediate step in an iteration, and that the iteration is only completed (i.e. $n \rightarrow n + 1$) when all probe positions $\mathbf{X}_j$ have been evaluated. In either case, note how the factorization constraint only affects the estimated exit waves $F_{n,j}(\mathbf{x})$ through Eq. (3.40), while leaving the feedback functions $B_{n,j}(\mathbf{x})$ untouched, as opposed to the global variant of HIO as defined in Eq. (3.8) where the projection operators are substituted with their ptychographic counterparts as defined in Eqs. (3.23) and (3.24).

We can also combine Modified HIO (M-HIO) with ptychography. M-HIO is a noise-robust variant of HIO [36], which works by setting a certain threshold level t , below which the feedback function is set to 0

$$B'(\mathbf{x}) = \begin{cases} B(\mathbf{x}) & \text{if } |B(\mathbf{x})| > t, \\ 0 & \text{if } |B(\mathbf{x})| \leq t. \end{cases} \quad (3.42)$$

If the noise level is higher, t should be increased as well. Note that if $t = 0$, M-HIO reduces to HIO, and if $t \rightarrow \infty$, M-HIO reduces to ER/AP. This adaptation can be implemented straightforwardly in the update function of Eq. (3.37). Here, we choose to express t in terms of a parameter $\alpha \in [0, 1]$

$$t = \alpha \cdot \max_{\mathbf{x}} |B(\mathbf{x})|, \quad (3.43)$$

where M-HIO reduces to HIO if $\alpha = 0$, and to ER/AP if $\alpha = 1$. Note that for the global variant of HIO as defined in Eq. (3.8), it is not obvious how it can be altered to yield a noise-robust ptychographic analogue of M-HIO.

In Fig. 3.5a we see that in the noise-free case, sequential HIO converges significantly faster than sequential PIE in the first several iterations. By the time PIE

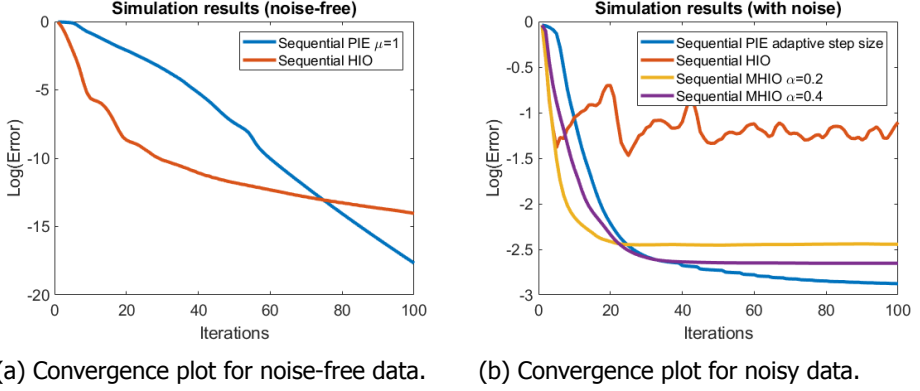


Figure 3.5: Comparison of sequential MHIO and sequential PIE for noise-free and noisy simulated data, as shown in Fig. 3.3

overtakes M-HIO, the reconstruction error is already negligibly small. In Fig. 3.5b we see that if noise is present, sequential HIO fails to converge, but M-HIO with $\alpha = 0.2$ does converge, and it does so faster than sequential PIE with adaptive step size. However, the final reconstruction error for PIE is lower than for M-HIO, so it is recommended to use sequential M-HIO for initial convergence (especially if few probe positions are used), and sequential PIE with reduced step size (or global AP) for final refinement.

Global ptychographic M-HIO for noise-robust simultaneous object and probe retrieval

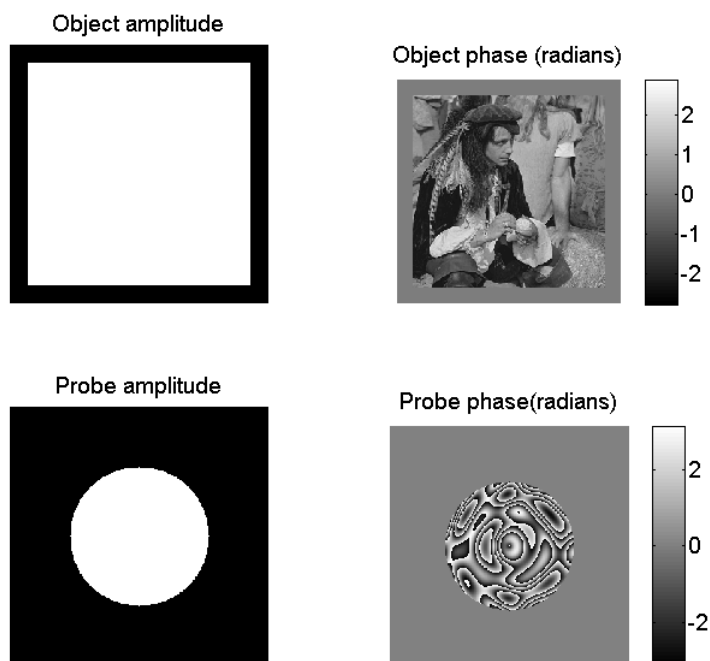
So far, we have assumed in the simulations that the probe $P(x)$ is known, and we only want to reconstruct the object $O(x)$. However, if we want to reconstruct $P(x)$ and $O(x)$ simultaneously, the problem becomes more difficult. Especially if the probe has a complicated structure (which may be desirable [37–39]) and no good initial guess is available, the ePIE algorithm [5] can stagnate. The global Difference Map (DM) algorithm [6] has the ability to escape local minima, but such stagnation-avoiding algorithms tend to become unstable if noise is present. We have seen that combining M-HIO with ptychography can provide a compromise between stagnation prevention and noise robustness, so here we investigate how M-HIO ptychography compares to DM (as defined in Eq. (3.9) with $\beta = 1$) when simultaneously retrieving the object and probe if noise is present. However, while making the comparison one must keep in mind that M-HIO requires a support constraint for the probe, while DM requires no such assumption or prior knowledge.

In Fig. 3.6 the data that was used for the simulations are shown. In Fig. 3.7 the DM reconstructions are shown. We see that DM is successful only for low noise levels. In Fig. 3.8 successful MHIO-reconstructions are shown. For these reconstructions, α has been chosen carefully such that the reconstruction is successful. In Fig. 3.9 it is shown how sensitive the reconstruction is to the choice of α . For the highest

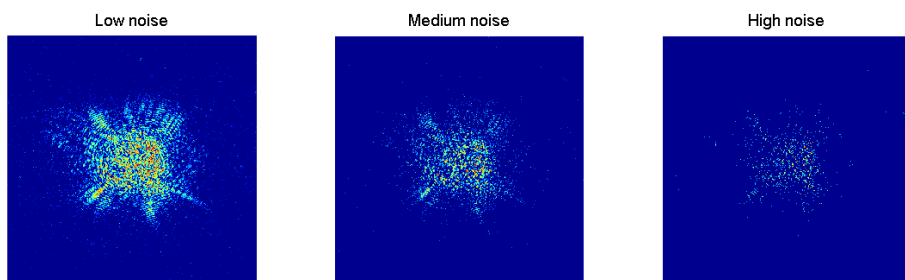
noise level, even a small change in α can cause the reconstruction to fail. One can find the correct α in a trial-and-error fashion by evaluating the reconstructions: if the reconstruction appears too chaotic, α is too low and it should be increased, while if the reconstruction appears to have stagnated at a local minimum, α is too high and it should be decreased.

3.3. Summary

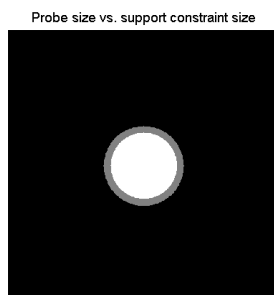
- In the basic phase retrieval problem, one has two constraints with which one tries to reconstruct the object: an **intensity (or modulus) constraint** which is found by measuring the diffracted far-field intensity pattern, and a **support constraint** which is known a priori.
- There are three main ways to solve this problem
 1. **Cost functional minimization:** one defines a cost functional that quantifies the difference between the estimated far field amplitude and the measured amplitude. The reconstruction is found by minimizing the value of this cost functional.
 2. **Finding the intersection of sets:** one defines two sets corresponding to the two constraints. Each set contains all functions satisfying that constraint. Since the reconstruction should satisfy both constraints, a reconstruction is found by finding the intersection of these sets.
 3. **Rank minimization:** in order to make the problem convex, one 'lifts' the problem in a higher dimension, which physically corresponds to considering all partially coherent solutions. One solves the problem by minimizing the rank of the reconstruction, which physically corresponds to finding the most coherent solution.
- The most relevant phase retrieval algorithms that have been designed are
 - **Error Reduction (ER)**, also called Alternating Projections (AP), gradient descent, or steepest descent: this is the most straightforward algorithm. It tends to suffer from stagnation because the phase retrieval problem is non-convex, but it is suitable for final refinement of the reconstruction.
 - **Hybrid Input-Output (HIO), Difference Map (DM), Relaxed Averaged Alternating Reflections (RAAR):** these different algorithms are the same for a certain parameter choice. They avoid stagnation, which makes them more effective, but it also makes their convergence behaviour less predictable. They are typically used for initial convergence, after which ER is used for final refinement.
- A more robust method for phase retrieval is **ptychography**. In this method, we do not use a support constraint on the object, but rather we record multiple far field diffraction patterns by shifting the illumination, which is called the probe function. Because the probes at different positions overlap with each



(a) Object and probe used for the simulations. 5×5 probe positions are used.



(b) Different levels of Poissonian noise (i.e. shot noise), displayed in logarithmic scale for visibility.



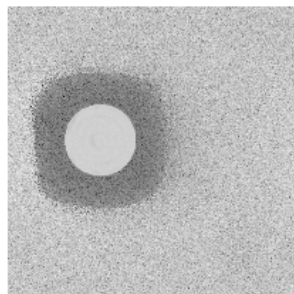
(c) Probe size (small circle) and support constraint size (large circle) for MHIO.

Figure 3.6: Simulated data that is used to compare the DM algorithm with the proposed global MHIO algorithm for simultaneous probe and object retrieval.

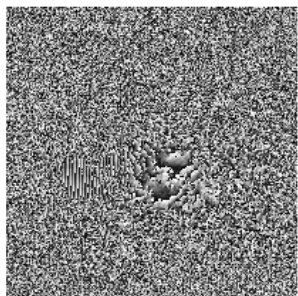
DM, Low noise, Object phase



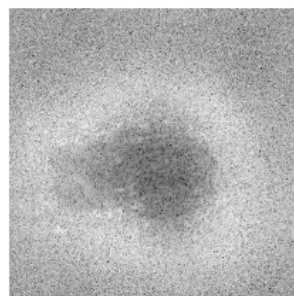
DM, Low noise, Probe amplitude (log)



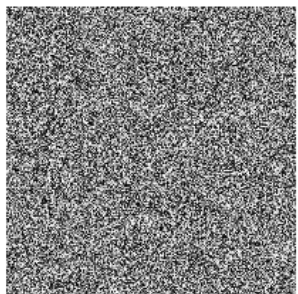
DM, Medium noise, Object phase



DM, Medium noise, Probe amplitude (log)



DM, High noise, Object phase



DM, High noise, Probe amplitude (log)

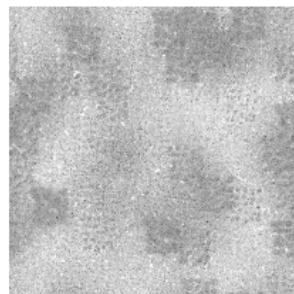


Figure 3.7: DM reconstructions for the different noise levels. The reconstructions were refined using global AP, and phase ramps that were present due to the raster grid pathology have been manually removed.

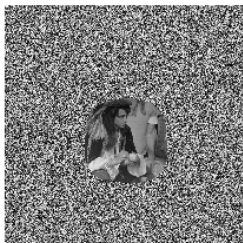
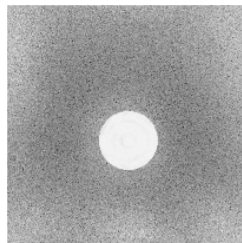
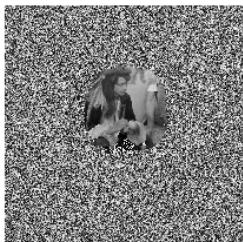
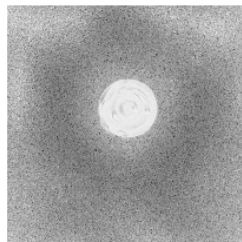
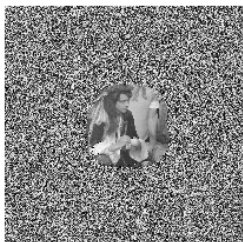
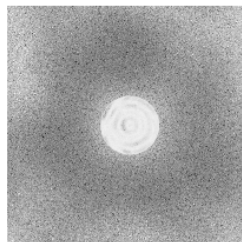
MHIO $\alpha=0$, Low noise, Object phaseMHIO $\alpha=0$, Low noise, Probe amplitude (log)MHIO $\alpha=0.1$, Medium noise, Object phaseMHIO $\alpha=0.1$, Medium noise, Probe amplitude (log)MHIO $\alpha=0.19$, High noise, Object phaseMHIO $\alpha=0.19$, High noise, Probe amplitude (log)

Figure 3.8: Successful MHIO reconstructions for the different noise levels. The reconstructions were refined using global AP, hence no probe support constraint is visible. Phase ramps that were present due to the raster grid pathology have been manually removed.

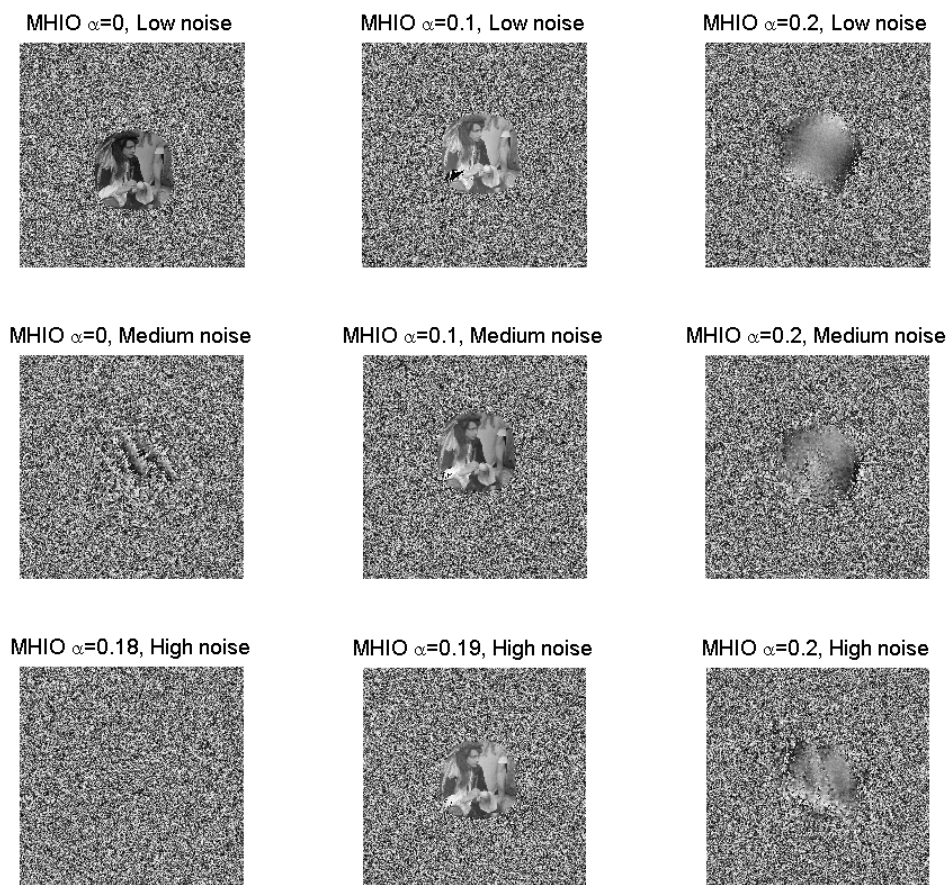


Figure 3.9: MHIO reconstructions for the different noise levels, and for different MHIO parameters α . The reconstructions were refined using global AP, hence no probe support constraint is visible. Phase ramps that were present due to the raster grid pathology have been manually removed.

other, the phase retrieval problem becomes significantly more constrained, which allows for a more robust reconstruction.

- One can apply ptychographic reconstruction algorithms using **global updates** or **sequential updates**. Global updates tend to be more robust to noise, while sequential updates tend to converge faster, especially if little noise is present.
- We designed a way to apply **HIO updates sequentially in ptychography**, and to **combine M-HIO with ptychography** (M-HIO is a more noise robust variant of HIO). In these methods, the key point is that if the probe has a practically finite support, we can define a feedback function for each probe position. When applying the factorization constraint (i.e. decomposing the estimated exit waves in an object and a probe using either a global or a sequential update), we only update the exit waves, while leaving the feedback functions unaltered. Sequential ptychographic M-HIO can be used for accelerating convergence in the first several iterations, while global ptychographic M-HIO can be used to increase the noise robustness of simultaneous object and probe retrieval.

References

- [1] A. Konijnenberg, *An introduction to the theory of ptychographic phase retrieval methods*, *Advanced Optical Technologies* **6** (2017), 10.1515/aot-2017-0049.
- [2] A. Konijnenberg, W. Coene, S. Pereira, and H. Urbach, *Combining ptychographical algorithms with the hybrid input-output (HIO) algorithm*, *Ultramicroscopy* **171**, 43 (2016).
- [3] P. Hessian, B. Pfau, E. Guehrs, M. Schneider, L. Shemilt, J. Geilhufe, and S. Eisebitt, *Holography-guided ptychography with soft x-rays*, *Optics Express* **24**, 1840 (2016).
- [4] M. Guizar-Sicairos and J. Fienup, *Phase retrieval with transverse translation diversity: a nonlinear optimization approach*, *Optics Express* **16**, 7264 (2008).
- [5] A. Maiden and J. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, *Ultramicroscopy* **109**, 1256 (2009).
- [6] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, *Probe retrieval in ptychographic coherent diffractive imaging*, *Ultramicroscopy* **109**, 338 (2009).
- [7] A. Hurst, T. Edo, T. Walther, F. Sweeney, and J. Rodenburg, *Probe position recovery for ptychographical imaging*, *Journal of Physics: Conference Series* **241**, 012004 (2010).
- [8] A. Maiden, M. Humphry, M. Sarahan, B. Kraus, and J. Rodenburg, *An annealing algorithm to correct positioning errors in ptychography*, *Ultramicroscopy* **120**, 64 (2012).

- [9] F. Zhang, I. Peterson, J. Vila-Comamala, A. Diaz, F. Berenguer, R. Bean, B. Chen, A. Menzel, I. Robinson, and J. Rodenburg, *Translation position determination in ptychographic coherent diffraction imaging*, *Optics Express* **21**, 13592 (2013).
- [10] A. Tripathi, I. McNulty, and O. Shpyrko, *Ptychographic overlap constraint errors and the limits of their numerical recovery using conjugate gradient descent methods*, *Optics Express* **22**, 1452 (2014).
- [11] P. Dwivedi, A. Konijnenberg, S. Pereira, and H. Urbach, *Lateral position correction in ptychography using the gradient of intensity patterns*, *Ultramicroscopy* **192**, 29 (2018).
- [12] P. Thibault and A. Menzel, *Reconstructing state mixtures from diffraction measurements*, *Nature* **494**, 68 (2013).
- [13] N. Burdet, X. Shi, D. Parks, J. N. Clark, X. Huang, S. D. Kevan, and I. K. Robinson, *Evaluation of partial coherence correction in x-ray ptychography*, *Optics express* **23**, 5452 (2015).
- [14] A. M. Maiden, M. J. Humphry, and J. Rodenburg, *Ptychographic transmission microscopy in three dimensions using a multi-slice approach*, *JOSA A* **29**, 1606 (2012).
- [15] T. Edo, D. Batey, A. Maiden, C. Rau, U. Wagner, Z. Pešić, T. Waigh, and J. Rodenburg, *Sampling in x-ray ptychography*, *Physical Review A* **87**, 053850 (2013).
- [16] A. M. Maiden, M. J. Humphry, F. Zhang, and J. M. Rodenburg, *Superresolution imaging via ptychography*, *JOSA A* **28**, 604 (2011).
- [17] J. Fienup, *Reconstruction of an object from the modulus of its fourier transform*, *Optics Letters* **3**, 27 (1978).
- [18] S. Marchesini, *Invited article: A unified evaluation of iterative projection algorithms for phase retrieval*, *Review of Scientific Instruments* **78**, 011301 (2007).
- [19] J. Qian, C. Yang, A. Schirotzek, F. Maia, and S. Marchesini, *Efficient algorithms for ptychographic phase retrieval*, *Inverse Problems and Applications, Contemp. Math* **615**, 261 (2014).
- [20] J. R. Fienup, *Phase retrieval algorithms: a comparison*, *Applied Optics* **21**, 2758 (1982).
- [21] V. Elser, *Phase retrieval by iterated projections*, *Journal of the Optical Society of America A* **20**, 40 (2003).
- [22] D. Luke, *Relaxed averaged alternating reflections for diffraction imaging*, *Inverse Problems* **21**, 37 (2004).

- [23] S. Marchesini, H. He, H. N. Chapman, S. P. Hau-Riege, A. Noy, M. R. Howells, U. Weierstall, and J. C. H. Spence, *X-ray image reconstruction from a diffraction pattern alone*, *Physical Review B* **68** (2003), 10.1103/physrevb.68.140101.
- [24] R. Hunger, *An introduction to complex differentials and complex differentiability* (Munich University of Technology, Inst. for Circuit Theory and Signal Processing, 2007).
- [25] S. Ruder, *An overview of gradient descent optimization algorithms*, arXiv preprint arXiv:1609.04747 (2016).
- [26] A. Maiden, D. Johnson, and P. Li, *Further improvements to the ptychographical iterative engine*, *Optica* **4**, 736 (2017).
- [27] 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**, 1241 (2012).
- [28] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, *Phase retrieval via matrix completion*, *SIAM Journal on Imaging Sciences* **6**, 199 (2013).
- [29] H. M. L. Faulkner and J. M. Rodenburg, *Movable aperture lensless transmission microscopy: A novel phase retrieval algorithm*, *Physical Review Letters* **93** (2004), 10.1103/physrevlett.93.023903.
- [30] J. Rodenburg and H. Faulkner, *A phase retrieval algorithm for shifting illumination*, *Applied physics letters* **85**, 4795 (2004).
- [31] M. Beckers, T. Senkbeil, T. Gorniak, K. Giewekemeyer, T. Salditt, and A. Rosenhahn, *Drift correction in ptychographic diffractive imaging*, *Ultramicroscopy* **126**, 44 (2013).
- [32] J. Sun, Q. Chen, Y. Zhang, and C. Zuo, *Efficient positional misalignment correction method for fourier ptychographic microscopy*, *Biomedical Optics Express* **7**, 1336 (2016).
- [33] R. Horstmeyer, R. Chen, X. Ou, B. Ames, J. Tropp, and C. Yang, *Solving ptychography with a convex relaxation*, *New Journal of Physics* **17**, 053044 (2015).
- [34] T. B. Edo, *Advances in Ptychography, see Eqs. 2.21-2.23*, Ph.D. thesis, University of Sheffield (2011).
- [35] C. Zuo, J. Sun, and Q. Chen, *Adaptive step-size strategy for noise-robust fourier ptychographic microscopy*, *Optics Express* **24**, 20724 (2016).
- [36] A. V. Martin, F. Wang, N. D. Loh, T. Ekeberg, F. R. N. C. Maia, M. Hanke, G. van der Schot, C. Y. Hampton, R. G. Sierra, A. Aquila, S. Bajt, M. Barthelmeß, C. Bostedt, J. D. Bozek, N. Coppola, S. W. Epp, B. Erk,

- H. Fleckenstein, L. Foucar, M. Frank, H. Graafsma, L. Gumprecht, A. Hartmann, R. Hartmann, G. Hauser, H. Hirsemann, P. Holl, S. Kassemeyer, N. Kimmel, M. Liang, L. Lomb, S. Marchesini, K. Nass, E. Pedersoli, C. Reich, D. Rolles, B. Rudek, A. Rudenko, J. Schulz, R. L. Shoeman, H. Soltau, D. Starodub, J. Steinbrener, F. Stellato, L. Strüder, J. Ullrich, G. Weidenspointner, T. A. White, C. B. Wunderer, A. Barty, I. Schlichting, M. J. Bogan, and H. N. Chapman, *Noise-robust coherent diffractive imaging with a single diffraction pattern*, *Optics Express* **20**, 16650 (2012).
- [37] P. Li, D. J. Batey, T. B. Edo, A. D. Parsons, C. Rau, and J. M. Rodenburg, *Multiple mode x-ray ptychography using a lens and a fixed diffuser optic*, *Journal of Optics* **18**, 054008 (2016).
- [38] A. Maiden, G. Morrison, B. Kaulich, A. Gianoncelli, and J. Rodenburg, *Soft x-ray spectromicroscopy using ptychography with randomly phased illumination*, *Nature Communications* **4** (2013), 10.1038/ncomms2640.
- [39] M. Guizar-Sicairos, M. Holler, A. Diaz, J. Vila-Comamala, O. Bunk, and A. Menzel, *Role of the illumination spatial-frequency spectrum for ptychography*, *Physical Review B* **86** (2012), 10.1103/physrevb.86.100103.

4

Noise robust ptychography

We review and investigate several methods to make the ptychographic algorithm more noise robust, including the maximum likelihood method and variance stabilization. We find that for Poissonian noise, the optimal object reconstruction is not necessarily found with the maximum likelihood method. Therefore, we propose a novel method to improve noise robustness, which involves iteratively updating the intensity constraints.

We saw in Chapter 3 that one can formulate the phase retrieval problem as a cost functional minimization problem. In Eq. (3.14) we gave a cost functional for single-shot phase retrieval, which can be used to reconstruct the exit wave $\psi(\mathbf{x})$ by using the steepest descent method as described in Eq. (3.15). In Eq. (3.25) we gave a ptychographic cost functional that can be minimized using the steepest descent method as described in Eq. (3.27) (or Eq. (3.30) if one uses a preconditioner). By doing so, one finds a reconstruction for the object $O(\mathbf{x})$ for a given probe $P(\mathbf{x})$, or one can reconstruct $O(\mathbf{x})$ and $P(\mathbf{x})$ simultaneously if both are unknown.

In Eqs. (3.14) and (3.25) we defined the cost functional L that quantifies the reconstruction error using the squared difference of the reconstructed far field amplitude and the measured amplitude. However, this is just one of many cost functionals that can be chosen. For example, why not define the cost functional using the estimated and reconstructed far field intensities, as opposed to the amplitudes? In the following we discuss how one might use different cost functionals depending on the noise model that is assumed. In these discussions, it is important to distinguish three different quantities:

- The **noise-free diffraction patterns**

$$\begin{aligned} m_j(\mathbf{k}) &= |\hat{\psi}_j(\mathbf{k})|^2 \\ &= |\mathcal{F}\{O(\mathbf{x})P(\mathbf{x} - \mathbf{X}_j)\}(\mathbf{k})|^2. \end{aligned} \quad (4.1)$$

These are the far-field intensity patterns that would be measured if no noise is present. $O(\mathbf{x})$ denotes the actual, but unknown, object. $P(\mathbf{x})$ denotes the actual probe, which may or may not be known.

- The **measured diffraction patterns** $y_j(\mathbf{x})$. These are the same as $m_j(\mathbf{k})$, except they are corrupted by a certain type of noise.
- The **estimated diffraction patterns**

$$\begin{aligned} z_{n,j}(\mathbf{k}) &= |\hat{f}_{n,j}(\mathbf{k})|^2 \\ &= |\mathcal{F}\{O_{n,j}(\mathbf{x})P_{n,j}(\mathbf{x} - \mathbf{X}_j)\}(\mathbf{k})|^2. \end{aligned} \quad (4.2)$$

These are the far-field intensity patterns that are calculated from the estimated object (and probe, if it is unknown). For a correct reconstruction, one requires $z_j(\mathbf{k}) = m_j(\mathbf{k})$ for all probe positions j .

4.1. Maximum likelihood cost function

Let us say that given a noise-free intensity value m at a certain pixel $\mathbf{k}$, the probability to measure a noise-corrupted value y is given by a probability distribution $P(y|m)$. If we measure an entire data set $\mathbf{y}$ that contains noise, we want to find the object $O(\mathbf{x})$ that would give an estimated noise-free data set $\mathbf{z}$ that is most likely to yield $\mathbf{y}$. Thus, we must express the probability P_{tot} to measure $\mathbf{y}$ as a function of

z (which is a function of $O(x)$)

$$P_{\text{tot}}[z] = \prod_{j,k} P(y_j(k)|z_j(k)), \quad (4.3)$$

where we assumed that noise at different pixels is uncorrelated. We want to find $O(x)$ that maximizes $P_{\text{tot}}[z]$. Since the logarithm preserves order, we can also minimize $-\log P_{\text{tot}}[z]$. This means we can define the cost functional [2-4]

$$L[O(x)] = - \sum_{j,k} \log P(y_j(k)|z_j(k)). \quad (4.4)$$

If we have Gaussian noise

$$P(y|m) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-m)^2}{2\sigma^2}}, \quad (4.5)$$

then we obtain the intensity-based cost functional

$$L[O(x)] = \sum_{j,k} (y_j(k) - z_j(k))^2, \quad (4.6)$$

where we have ignored irrelevant constant additive terms and multiplicative factors.

If we have Poissonian noise, then y (which is a non-negative integer) follows a Poissonian distribution with mean value m (which is a non-negative real number)

$$P(y|m) = \frac{m^y e^{-m}}{y!}. \quad (4.7)$$

The cost functional is then given by

$$L[O(x)] = \sum_{j,k} z_j(k) - y_j(k) \log z_j(k), \quad (4.8)$$

where we have ignored irrelevant constant additive terms. We can find the following Taylor expansion in terms of $\sqrt{z}$ around the point $\sqrt{z} = \sqrt{y}$

$$z - y \log z \approx y - y \log y + 2(\sqrt{z} - \sqrt{y})^2. \quad (4.9)$$

Thus, ignoring irrelevant additive and multiplicative constants, we can define the cost functional [4]

$$L[O(x)] = \sum_{j,k} \left(\sqrt{y_j(k)} - \sqrt{z_j(k)} \right)^2. \quad (4.10)$$

This is the amplitude-based cost functional that is used to derive regular PIE and ePIE.

4.2. Variance stabilizing transforms

Another way to look at the problem is from the perspective of variance stabilization. For example, consider the intensity-based cost function of Eq. (4.8). If we assume that the noise is Poissonian, then for larger measured intensities y we expect a larger error $(y - m)^2$. We therefore find that the intensity-based cost weighs the pixels with high intensity more strongly than pixels with low intensity (assuming that the algorithm converges to a solution $z \approx m$). If we want each pixel to have equal weight regardless of its intensity, we need to apply a transformation so that the expected error is independent of the measured intensity, which is a variance-stabilizing transform [5]. For Poisson noise, this transformation is taking the square-root, which leads to the amplitude-based cost function of Eq. (4.10). It has been proposed to use other variance-stabilizing transforms, such as the Anscombe transform [6].

Indeed, we can show that the maximum likelihood approach as outlined in the previous section is in fact also a method to make the variance the same for each pixel regardless of its measured intensity y . If $P(y|m)$ denotes the probability we measure a noisy value y given a noise-free value m , then we can use Bayes' rule to calculate the probability $P(m|y)$ that m is the noise-free value if we measure a noisy value y

$$P(m|y) = \frac{P(y|m)P(m)}{P(y)}. \quad (4.11)$$

We now want to find a transformation T_y such that $T_y(m)$ is normally distributed with mean 0 and standard deviation 1, independently of the measured intensity value y

$$P(m|y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{T_y(m)^2}{2}}. \quad (4.12)$$

The cost function we then want to minimize is

$$L[\mathcal{O}(\mathbf{x})] = \sum_{j,k} T_{y_j(\mathbf{k})}(z_j(\mathbf{k}))^2. \quad (4.13)$$

We can use Eq. (4.12) solve for $T_y(m)^2$

$$\begin{aligned} T_y(m)^2 &= -2 \left(\log \sqrt{2\pi} + \log P(m|y) \right) \\ &= -2 \left(\log \sqrt{2\pi} + \log P(y|m) + \log P(m) - \log P(y) \right). \end{aligned} \quad (4.14)$$

We can plug this expression into Eq. (4.13). If we assume we have no prior information about m (so $P(m)$ is independent of m), and we ignore all the irrelevant additive constants that are independent of z , and we ignore global multiplicative constants, we find once more the maximum likelihood cost function of Eq. (4.4).

4.3. Testing different cost functions for noise robustness

We have seen that according to the maximum-likelihood approach, an amplitude-based cost functional should give optimal results if the measurements are corrupted by Poisson noise, and an intensity-based cost function should give optimal results if there is Gaussian noise. In order to further investigate these predictions, we introduce the cost functional

$$\begin{aligned} L_\alpha[f(\mathbf{x})] &= \sum_{\mathbf{k}} (z(\mathbf{k})^\alpha - y(\mathbf{k})^\alpha)^2 \\ &= \sum_{\mathbf{k}} (|\hat{f}(\mathbf{k})|^{2\alpha} - y(\mathbf{k})^\alpha)^2 \end{aligned} \quad (4.15)$$

where for simplicity we consider only a single probe position $\mathbf{X}_j$. For $\alpha = 0.5$ we obtain the amplitude-based cost functional, and for $\alpha = 1$ we obtain the intensity-based cost functional. For $\alpha = 0.5$, we find for the Wirtinger derivative

$$\frac{\partial L_{0.5}}{\partial f(\mathbf{x})^*} = \mathcal{F}^{-1} \left\{ \left(|\hat{f}(\mathbf{k})| - \sqrt{y(\mathbf{k})} \right) \frac{\hat{f}(\mathbf{k})}{|\hat{f}(\mathbf{k})|} \right\}(\mathbf{x}). \quad (4.16)$$

As noted in Chapter 3, applying the gradient descent scheme using this cost functional gives the same update scheme as the method of Alternate Projections, which is stable. For arbitrary α , the derivative is

$$\begin{aligned} \frac{\partial L_\alpha}{\partial f(\mathbf{x})^*} &= \mathcal{F}^{-1} \left\{ 2\alpha \left(|\hat{f}(\mathbf{k})|^{2\alpha} - \sqrt{y(\mathbf{k})}^{2\alpha} \right) |\hat{f}(\mathbf{k})|^{2\alpha-2} \hat{f}(\mathbf{k}) \right\}(\mathbf{x}) \\ &= \mathcal{F}^{-1} \left\{ \left(|\hat{f}(\mathbf{k})| - \sqrt{y(\mathbf{k})} \frac{\sqrt{y(\mathbf{k})}^{2\alpha-1}}{|\hat{f}(\mathbf{k})|^{2\alpha-1}} \right) \frac{\hat{f}(\mathbf{k})}{|\hat{f}(\mathbf{k})|} (2\alpha |\hat{f}(\mathbf{k})|^{4\alpha-2}) \right\}(\mathbf{x}). \end{aligned} \quad (4.17)$$

The expression in the last line highlights that compared to the case of $\alpha = 0.5$, the derivative is approximately multiplied by $2\alpha |\hat{f}(\mathbf{k})|^{4\alpha-2}$ (we assume that the impact of the scaling factor with which $\sqrt{y(\mathbf{k})}$ is multiplied is less relevant, since it should be close to 1 if the estimated amplitude $|\hat{f}(\mathbf{k})|$ approximately matches the measured amplitude $\sqrt{y(\mathbf{k})}$). Since we know that for $\alpha = 0.5$ the global steepest descent update as defined in Eq. (3.30) is stable for a step size $\mu = 1$ (reducing the step size is relevant for sequential updates, not so much for global updates), we define for different α the steepest descent update as Eq. (3.30), except we substitute $f_{n,j}(\mathbf{x}) - f_{n,j}^{\text{upd}}(\mathbf{x})$ with

$$\frac{1}{\max_{\mathbf{k}} (2\alpha |\hat{f}_{n,j}(\mathbf{k})|^{4\alpha-2})} \frac{\partial L_{\alpha,j}}{\partial f_{n,j}(\mathbf{x})^*}. \quad (4.18)$$

We assume the probe $P(\mathbf{x})$ is known, so only $O(\mathbf{x})$ needs to be updated.

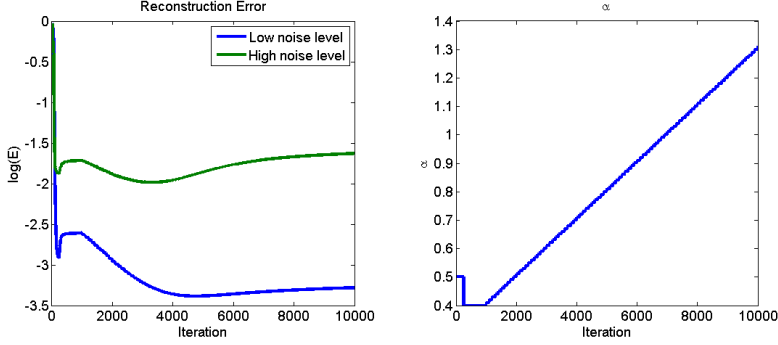


Figure 4.1: Reconstruction error for different cost functions L_α as defined in Eq. (4.17).

4

In simulations it has been tested how the object reconstruction error as defined in Eq. (3.33) varies as a function of α . The simulated data that are used for testing different L_α are shown in Fig. 3.3. To compare the different L_α , we first run the standard global ptychographic algorithm (i.e. $\alpha = 0.5$) to ensure we have a good object estimate, to which only final refinement needs to be applied. Then, we apply several updates using $\alpha = 0.4$ to make sure we have minimized $L_{0.4}$. Then, we incrementally increase α by 0.01 and run the algorithm for 100 iterations to make sure the algorithm has converged. By repeating this process every 100 iterations, we find how the reconstruction error changes as a function of α . The results are shown in Fig. 4.1. The minimal reconstruction error is not found for $\alpha = 0.5$, even though we simulated Poisson noise¹. Moreover, the optimal α is different for different noise levels (around $\alpha = 0.7$ for the low noise level, and 0.6 for the high noise level). Fundamentally, the problem is that we need to find a cost functional $L[f(x)]$ which, when minimized, yields a minimized reconstruction error $E[f(x)]$ as defined in Eq. (3.33). Apparently, finding the reconstruction for which the measured data set is most likely to occur given a certain noise model (by using a maximum-likelihood approach), does not generally give a reconstruction with the lowest reconstruction error. In Appendix C this problem is explored in more detail.

¹Gaussian noise has not been tested, because it gives problems when the intensity is low. For Gaussian noise with a certain standard deviation, there should be a probability to measure negative intensities at points where the noise-free intensity is close to zero. Since negative intensities are physically impossible, Gaussian noise cannot be modelled in such a way that the claims of maximum-likelihood refinement can be tested fairly. One could test the hypothesis for exponentially distributed noise, but that noise model has never been seriously considered for ptychography, so testing it would be of lesser practical value.

4.4. Noise-robustness by adapting measurement constraints²

4.4.1. Problems with the maximum likelihood scheme

We have seen that the maximum-likelihood cost functionals do not necessarily yield the optimal reconstruction, which means that there is some room for improvement. Another indication that one might be able to improve the algorithm, is the fact that we typically do not achieve $L = 0$ if noise is present: it means that the reconstruction does not match the constraints, which means we know something is wrong with the reconstruction or the constraints, but we do not use this information to improve the reconstruction. Moreover, all these algorithms push the solution towards $z_j(\mathbf{k}) = y_j(\mathbf{k})$, which we almost certainly know to be wrong, because if $y_j(\mathbf{k})$ is randomly distributed, the probability that $y_j(\mathbf{k}) = m_j(\mathbf{k})$ for all $j, \mathbf{k}$ is extremely small.

One argument for why the maximum likelihood method will not give the optimal object reconstruction, is that the wrong quantity is maximized. In Eq. (4.4), we defined the cost functional such that when it is minimized, the quantity $P(\mathbf{y}|\mathbf{z})$ is maximized. In other words, we try to find the object for which it is most likely to measure $\mathbf{y}$. However, it would make more sense to maximize $P(\mathbf{z}|\mathbf{y})$: given our measurements $\mathbf{y}$, we want to find the object that yields diffraction patterns $\mathbf{z}$ that are most likely implied by $\mathbf{y}$. Note that $P(\mathbf{y}|\mathbf{z})$ and $P(\mathbf{z}|\mathbf{y})$ are very dissimilar. $P(\mathbf{y}|\mathbf{z})$ is simply given by the product of probabilities for all pixels if the noise at different pixels is uncorrelated (see Eq. (4.3)). On the other hand, $P(\mathbf{z}|\mathbf{y})$ cannot be expressed as such a straightforward product, because of the redundancy of information in $\mathbf{y}$ due to the overlap between probes. If $y_1(\mathbf{k})$ and $y_2(\mathbf{k})$ denote the measured intensity patterns for two adjacent probe positions, then $P(z_1, z_2 | y_1, y_2) \neq P(z_1 | y_1)P(z_2 | y_2)$, because $P(z_1 | y_2)$ and $P(z_2 | y_1)$ are nontrivial functions: the measurement for probe position 1 contains information about the diffraction pattern for probe position 2, and vice versa.

Another way to think about it is that in Eq. (4.14) we assumed that we have no prior information about the noise-free intensity value $m(\mathbf{k})$, so $P(m(\mathbf{k}))$ would be independent of $m(\mathbf{k})$. However, given the overlap constraints in ptychography, one might argue that due to the redundancy in the data, there is in fact some prior information available if we only consider one single intensity pattern at a time. For example, consider a single probe position $\mathbf{X}_{j_0}$ and all its adjacent probe positions. If we measure the noisy intensity patterns $y_j(\mathbf{k})$ for all adjacent probe positions $\mathbf{X}_j$, then the noise-free pattern $m_{j_0}(\mathbf{k})$ for probe position $\mathbf{X}_{j_0}$ is not arbitrary. More specifically, one could reconstruct the object using all probe positions except $\mathbf{X}_{j_0}$, and use the reconstruction to compute a fairly accurate estimate of $m_{j_0}(\mathbf{x})$, even though we did not actually measure an intensity pattern for $\mathbf{X}_{j_0}$: therefore, one could argue that in some sense we have prior information that is not used in the

²The following is original work by the candidate, which has been published in Optics Express, **26:5** 5857-5874 (2018) [1]

variance stabilization method.

4.4.2. An alternative noise robust scheme

In order to improve the reconstruction quality in the presence of noise, we use the reconstructed object to update the estimate of the noise-free intensity patterns, which we in turn use to update the object reconstruction. We can come up with the following scheme [1]:

1. Choose the initial estimated noise-free intensity patterns to be $m_{j,\text{est},0}(\mathbf{k}) = y_j(\mathbf{k})$ or some denoised version of $y_j(\mathbf{k})$.
2. Run the reconstruction algorithm (PIE, ePIE, or some other variant) with a small step size for a certain number of iterations, using $m_{j,\text{est},n}(\mathbf{k})$ as the intensity constraints. We denote the resulting estimated diffraction patterns as $z_{j,n}(\mathbf{k})$.
3. Set

$$m_{j,\text{est},n+1}(\mathbf{k}) = \mu z_{j,n}(\mathbf{k}) + (1 - \mu)m_{j,\text{est},n}(\mathbf{k}), \quad (4.19)$$

where $\mu > 0$ is a small step size (i.e. much smaller than 1).

4. Repeat steps 2 and 3 until $z_{j,n}(\mathbf{k}) \approx m_{j,\text{est},n}(\mathbf{k})$.

Note that in case $\mu = 0$ the algorithm reduces to regular (e)PIE, and generally speaking the algorithm fails to achieve $L = 0$ if noise is present. If $\mu = 1$ the algorithm achieves $L = 0$ and stagnates immediately when Eq. (4.19) is applied, making this scheme redundant. If we choose μ to be small but nonzero, the intensity constraints are gradually updated until the estimated intensity patterns match the constraints, and $L = 0$ is achieved. Whether this scheme actually improves the reconstruction quality depends on the noise level of the measurements $y_j(\mathbf{k})$: if the measurements are too noisy, the reconstruction is too poor to update the constraints reliably, and this scheme will fail to improve the reconstruction quality. However, if the noise level is low enough to yield a decent object reconstruction, this scheme can be used as final refinement to slightly improve the reconstruction. Regardless of whether the improvement is significant enough to be of major practical value, this method may shed light on the possible limits of noise-robustness of ptychographic reconstruction algorithms. Moreover, this proposed method is also in line with the observation made in Appendix C, that the cost functional should be adapted according to the specific realization of noise, rather than the general noise statistics.

4.4.3. Experimental results

³To create and shift a phase object on which to test our proposed algorithm experimentally, we use a phase-only Spatial Light Modulator (SLM) to which we assign an image that we want to reconstruct. The SLM in which the object is created is

³The experiment was performed by the candidate, using an experimental setup that was built by MSc. student Ruben Biesheuvel.

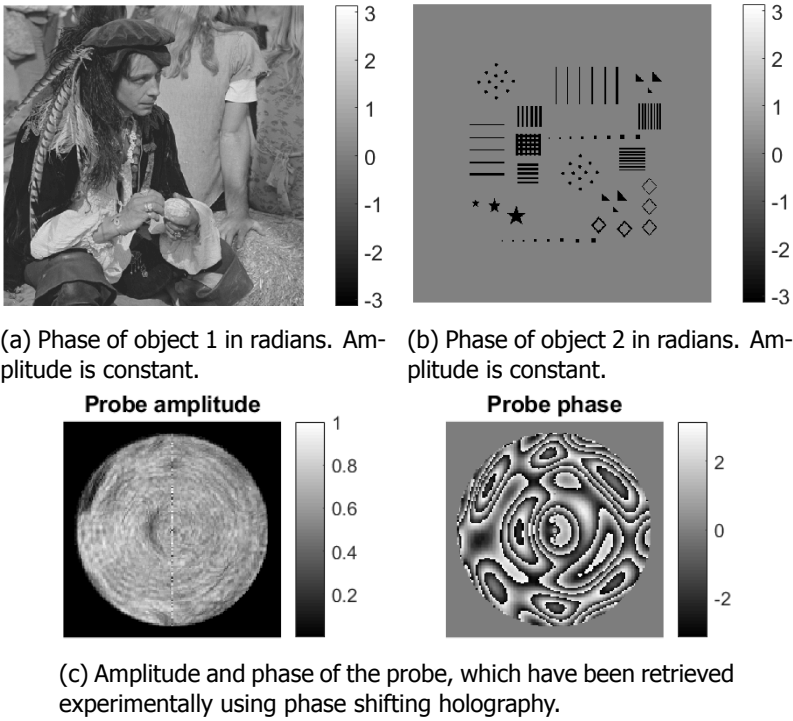


Figure 4.2: The phase objects and the illuminating probe used for the ptychography experiment.

a reflective liquid crystal phase-only PLUTO SLM by Holoeye, with a resolution of 1920×1080 pixels, and a pixel pitch of $8.0\mu\text{m}$. With a lens with a focal length of 15cm we create the Fourier transform of the field that is reflected by the SLM. To reduce the dynamic range of the diffraction patterns [7–9], a fixed rapidly spatially varying phase pattern as shown in Fig. 4.2(c) is added with the SLM on top of the shifted object, which more or less defines the probe $P(\mathbf{x})$. The illuminated area of the SLM (which corresponds to the size of the probe) is a circle with a radius of 250 pixels. The object, which is 800×800 pixels, is shifted along a 7×7 square grid with a period of 50 pixels with some random offsets to reduce the raster grid pathology in the reconstruction [10]. The images are recorded with an 8-bit SVS-VISTEK eco204MVGE CCD camera with a resolution of 1024×768 pixels and a pixel size of $4.65\mu\text{m} \times 4.65\mu\text{m}$. For each intensity measurement, we take the average of 50 pictures. The object is then reconstructed using ePIE⁴ [11] and a probe position correction scheme [12]. Another reconstruction is then performed which uses, in addition to ePIE and the probe position correction scheme, the proposed noise-robust scheme where the intensity constraints are updated as in (4.19). The experimental setup is shown in Fig. 4.3.

⁴Since this study is about the final refinement of the reconstruction, it is sufficient to use a local minimizer such as ePIE as opposed to DM or RAAR.

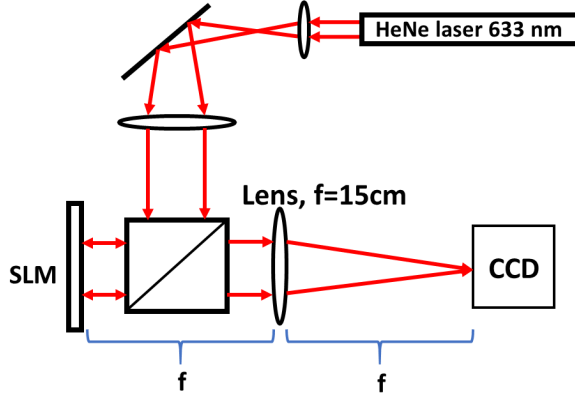


Figure 4.3: Experimental setup to generate ptychographic data sets to test the proposed algorithm on.

In Fig. 4.2 the phase objects that are applied to the SLM are shown. The object should in principle have a constant amplitude, but the contours of the phase pattern become apparent in the amplitude if the object is low-pass filtered. In the experiment, low-pass filtering of the object is inevitable since the information of the higher spatial frequencies that is present in the diffraction patterns is lost due to the finite size and limited dynamic range of the detector. Also, in Fig. 4.2 the probe with which the reconstruction was performed is shown. This probe was reconstructed using phase-shifting holography and refined using ePIE [11].

Before performing the reconstruction on the noisy data set using ePIE and the proposed algorithm, the intensity measurements were denoised. To do this, measurements that were taken in the dark (i.e. when the laser is completely blocked) were subtracted from the ptychographic measurements. Then, a rectangular region in the measurements was selected where the images should be practically zero, as shown in Fig. 4.4. In this region, the mean and standard deviation of the noise level were determined. The mean value was subtracted from the entire image, and everything below three times the standard deviation was set to zero. This denoising method as well as others are described in [13]. We run two reconstruction algorithms: ePIE with a reduced step size combined with a probe position correction scheme, and the proposed adaption of this algorithm. For the first algorithm we eventually use a step size of 0.1, although in the initial iterations a larger step size can be used. For the position correction update scheme [12], we use a step size of 0.1. For the proposed noise-robust algorithm we do the same, but after 150 iterations, we update the intensity constraints according to (4.19) each 10 iterations with a step size of $\mu = 0.05$. In Fig. 4.5 the reconstruction for the object 1 is shown. For object 1 we observe in some regions an increase in contrast. In Fig. 4.6 it is shown how adapting the intensity measurements causes a stronger presence of higher spatial frequencies, thus indeed leading to an increased resolution in the object reconstruction. In Fig. 4.7 it is shown that for the amplitude-based cost function L as defined in Eq. (4.10) L cannot converge to 0 when using regular

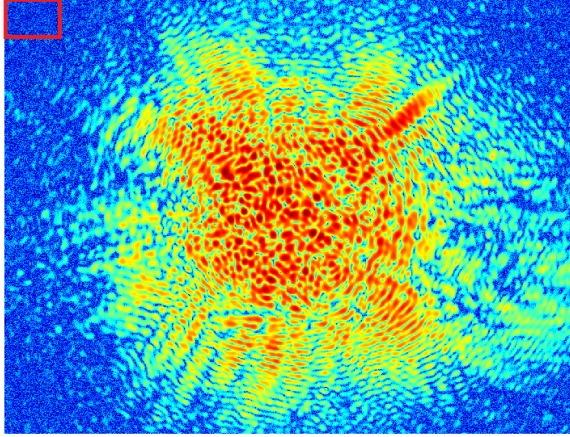


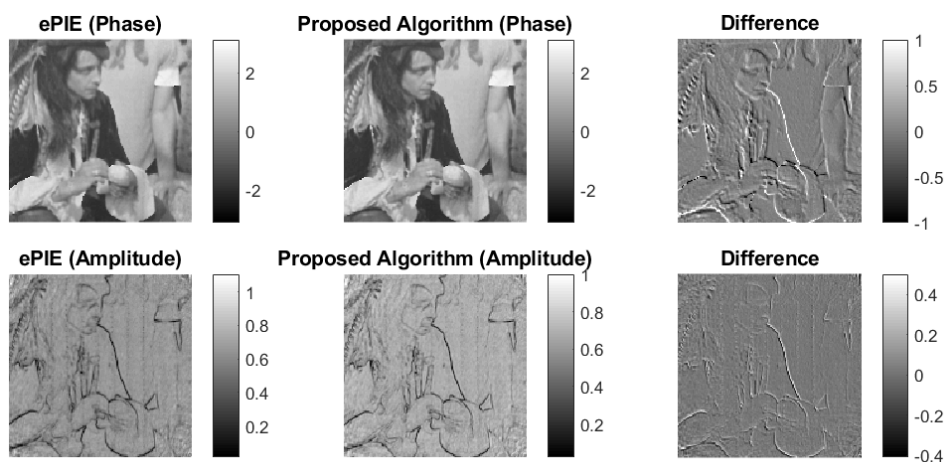
Figure 4.4: A measured intensity pattern (log scale) where a region is selected (red rectangle in the top left corner) to calculate the mean and the standard deviation of the noise.

ePIE, whereas when we adapt the amplitude constraints the value of L gradually approaches 0. For object 2, which is a binary image, we see in Fig. 4.8 that the smaller structures become more clearly visible by applying the proposed algorithm. The structured nature of object 2 causes the differences in the reconstructions to be more easily visible, so we choose this object to perform further tests of the algorithm. In Fig. 4.9 only a single measurement is taken per probe position, meaning that the signal-to-noise ratio is significantly lower than in Fig. 4.8. Additionally, in Fig. 4.10 we use incorrect initial probe guesses (with errors ranging from -5 pixels to +5 pixels in both the x and y dimensions) and use probe position correction. One can observe that in Fig. 4.10 the reconstruction quality is significantly lower than in 4.9, which is because the algorithm fails to find the correct probe positions due to the high noise levels. Nonetheless, in both cases the noise-robust extension appears to improve the reconstruction quality in some points, as evidenced by the zoomed-in images where the phase along the vertical lines are more uniform when the proposed algorithm is used⁵.

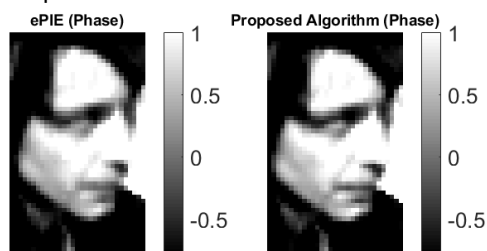
4.4.4. Simulations

To quantify the benefit of the proposed algorithm more accurately, we performed reconstructions with simulated data that closely resembled the measured data, i.e. we used the same object and probe as the ones that were assigned in the SLM in the experiment, and we used an oversampling rate that approximately matches the

⁵Unfortunately, it is difficult to perform further quantitative analysis on these reconstructions. Even though we know the test-object that has been assigned to the SLM, there are still small uncertainties in the magnification factor between the assigned image and the reconstructed image, the rotation angle between the SLM and the detector which causes the reconstructed image to be slightly rotated, and the way the SLM converts assigned grayscale values into phase shifts. Even though the uncertainties may be small, the differences between the reconstruction errors for the two algorithms are small as well, so it is difficult to use the test-object that is assigned to the SLM for quantitative analysis.



(a) Reconstruction results for the ePIE algorithm and the proposed algorithm where the intensity constraints are updated.



(b) Zoom with reduced phase range to increase contrast. Increased contrast can be observed between and below the lips.

Figure 4.5: Reconstructions of object 1.

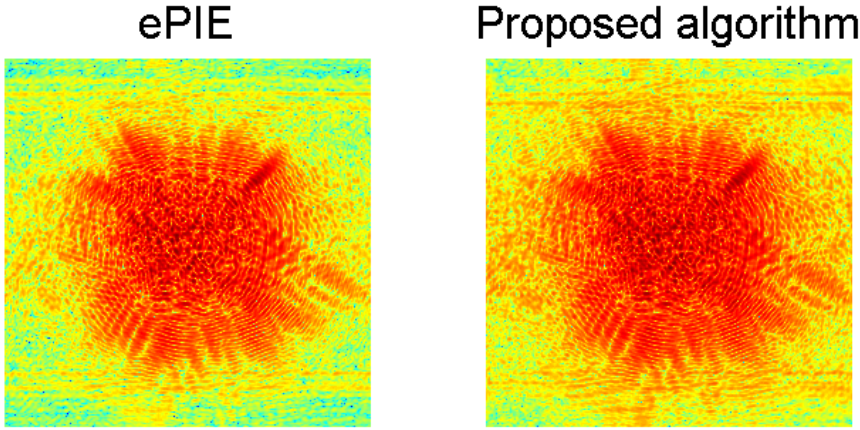


Figure 4.6: A comparison of the estimated diffraction patterns (log scale) of object 1 using standard ePIE and the proposed algorithm. One can observe that in the proposed algorithm the higher spatial frequencies have a stronger presence, thus causing the reconstructed object to have a higher resolution.

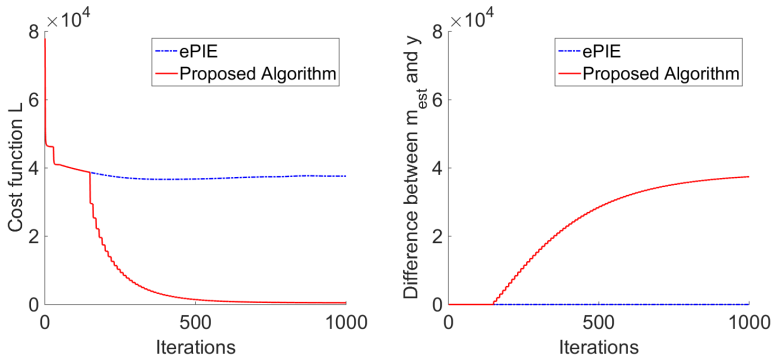
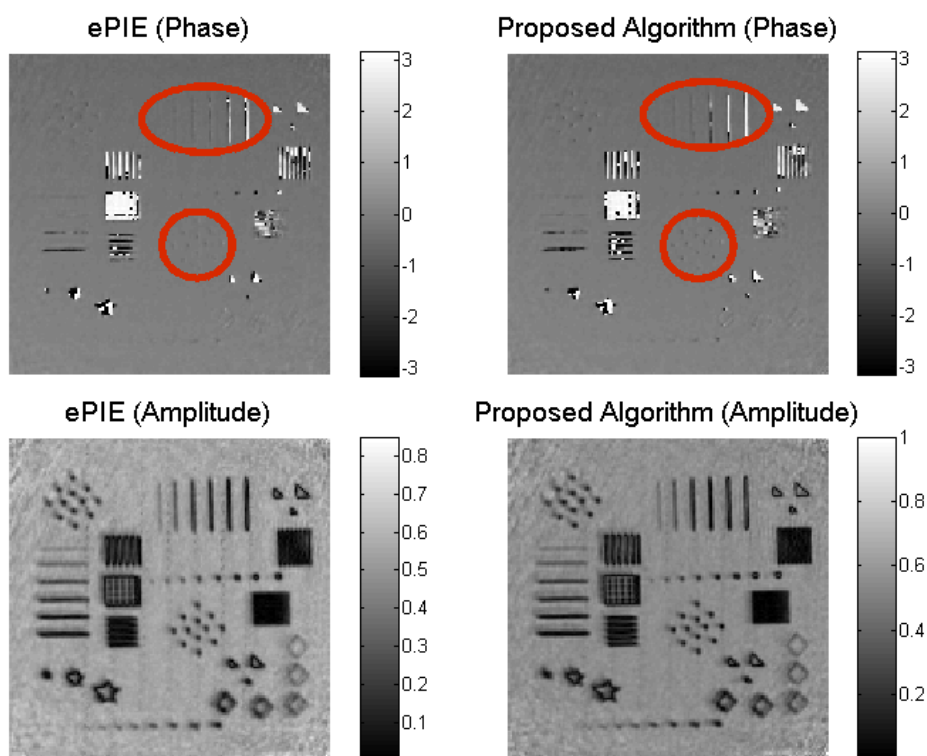
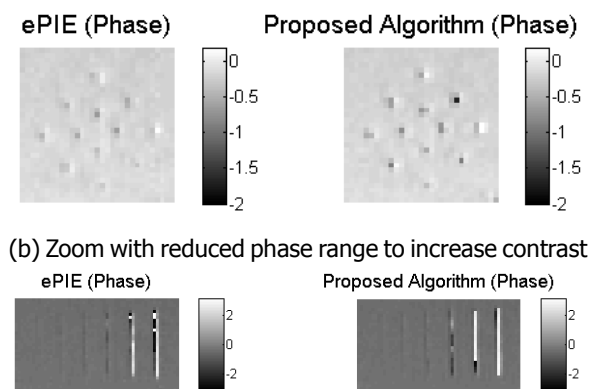


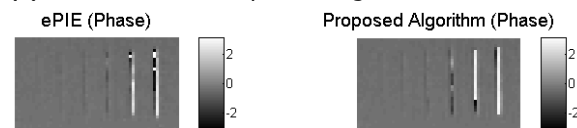
Figure 4.7: Plot of the amplitude-based cost function L as a function of the number of iterations. For the regular ePIE algorithm, the estimated amplitudes are compared with the measured amplitudes $y_{\mathbf{x}}(\mathbf{k})$, while for the proposed algorithm the estimated amplitudes are compared with the adapted amplitude constraints $m_{\mathbf{x},\text{est}}(\mathbf{k})$.



(a) Reconstruction results

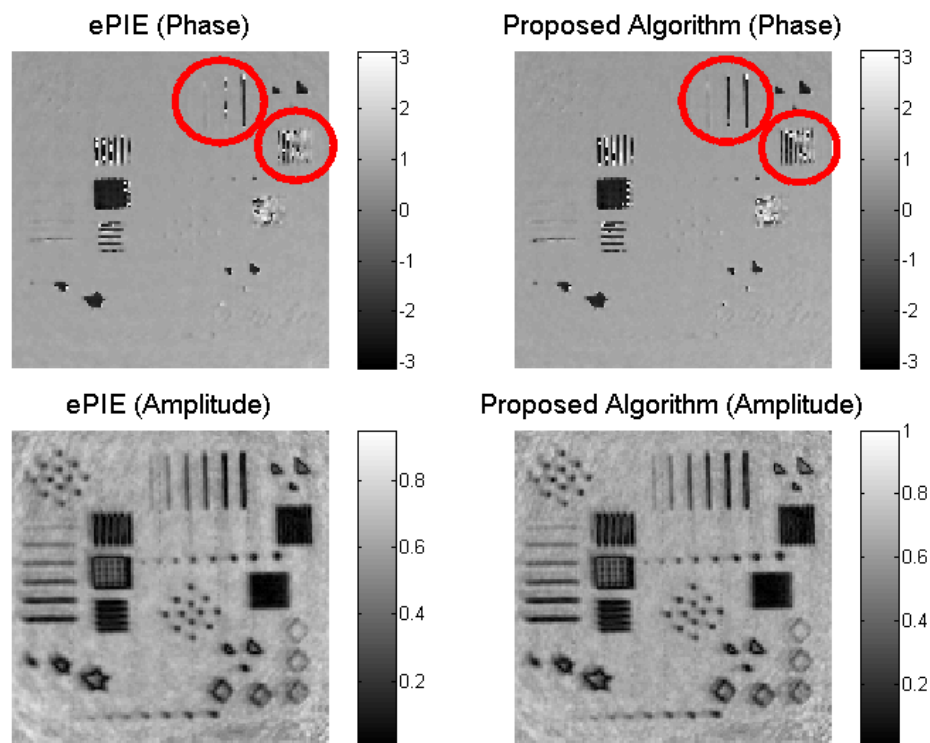


(b) Zoom with reduced phase range to increase contrast

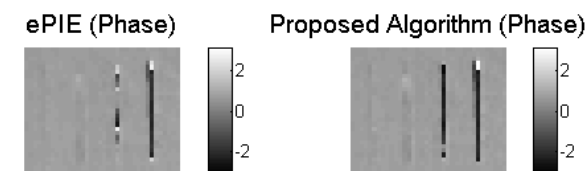


(c) Zoom

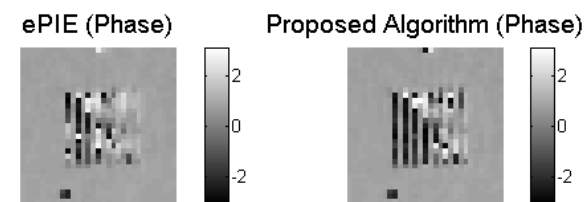
Figure 4.8: Reconstructions of object 2. For each probe position, 50 measurements were averaged. Position correction was used.



(a) Reconstruction results

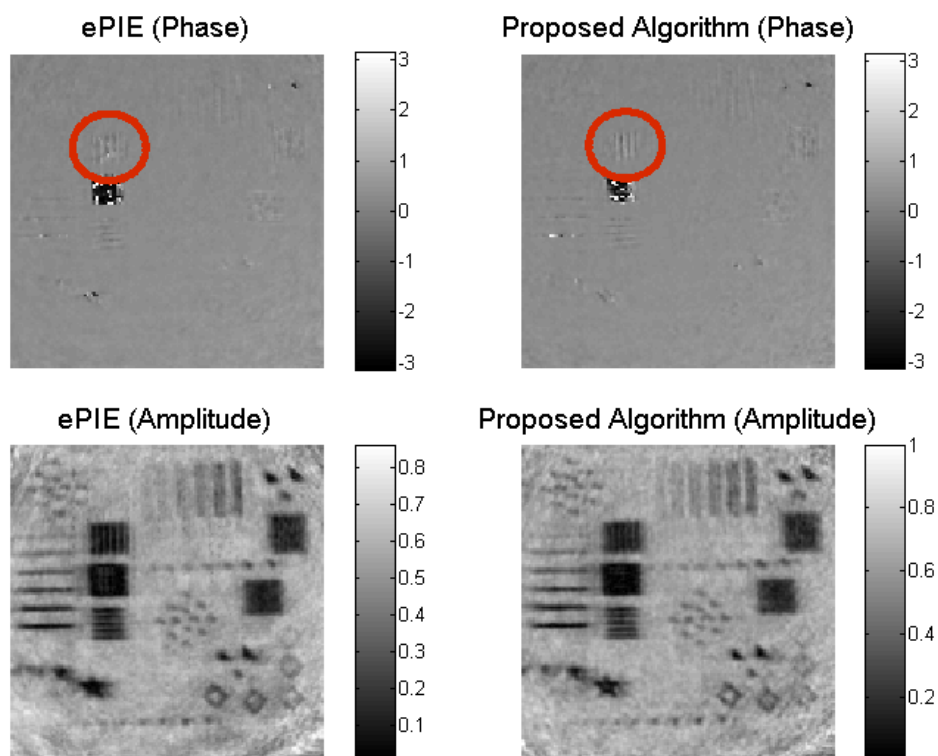


(b) Zoom

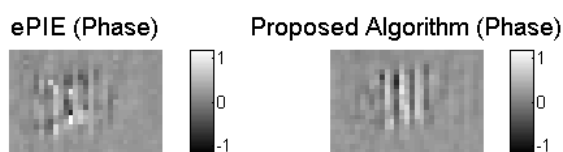


(c) Zoom

Figure 4.9: Reconstructions of object 2. For each probe position, a single measurement was taken. Correct probe positions were used in the reconstruction and no position correction was used.



(a) Reconstruction results



(b) Zoom with reduced phase range to increase contrast

Figure 4.10: Reconstructions of object 2. For each probe position, a single measurement was taken. Position correction was used.

one from the experiment. We multiplied the simulated noise-free intensity patterns with a certain value to change the photon count⁶. Then, Poisson noise was introduced to simulate shot noise. This process was performed for different values of the photon count, and for each value we compared the PIE reconstruction with the reconstruction obtained with the proposed algorithm (we assumed no uncertainty in the probe or probe positions). The reconstruction error is calculated using the error metric as defined in Eq. (3.33).

The results are shown in Fig. 4.11. We see that for higher photon counts (higher than 10^4) the proposed algorithm gives the same reconstruction error as using the regular PIE algorithm at a photon count that is $10^{0.5} \approx 3.2$ times higher. In practice this would mean that one could reduce the radiation dose to which the sample is exposed by a factor of 3 without sacrificing reconstruction quality, which would be desirable in X-ray imaging where the radiation can damage the sample. The reason why the benefit of the proposed algorithm is lower for higher noise levels is because it uses the reconstructed object to update the intensity constraints. If the noise levels are higher, the reconstructed object is worse, and the intensity constraints are updated in a less reliable manner. While this may appear as a major drawback for the proposed scheme (it works worse when it is needed more), one could argue that this trend is inevitable for any denoising scheme that exploits the redundancy of information in ptychographic data sets. If the scheme exploits the redundancy of information, then naturally the scheme has to perform worse if that information is more corrupted by noise.

With these simulations it was also investigated how many PIE updates should be applied each time after the intensity constraints have been updated according to Eq. (4.19). For an update parameter of $\mu = 0.05$, the reconstruction error has been plotted for different numbers of PIE updates in Fig. 4.12. One can observe that for this situation, applying just one PIE update after the intensity constraints are updated gives the same reconstruction error as applying 20 PIE updates, and significantly fewer iterations are needed.

To demonstrate that the proposed method is not restricted to one specific noise model, simulations with different amounts of additive Gaussian noise were performed. After normalizing the set of intensity patterns such that the overall maximum value is 1, Gaussian noise with standard deviations of $10^{-3.5}$, 10^{-3} , $10^{-2.5}$ was added. The resulting simulated intensity patterns were then denoised using the same procedure as described in Fig. 4.4. The reconstruction results shown in Fig. 4.13 indicate that the proposed method also works for additive Gaussian noise, provided that the noise level is not too high.

Lastly, we consider Fourier ptychography. We have seen that the maximum-likelihood schemes and variance-stabilizing schemes aim to construct a cost function that gives equal weight to pixels with low intensity and pixels with high intensity.

⁶‘Photon count’ means the photon count per diffraction pattern. Each noise-free diffraction pattern $m_j(\mathbf{k})$ is multiplied with the same gain factor G so that $\sum_{\mathbf{k}} G m_j(\mathbf{k})$ yields approximately the defined photon count. Per pixel $\mathbf{k}$, the noisy intensity value $y_j(\mathbf{k})$ is found by drawing from a Poisson distribution with mean $G m_j(\mathbf{k})$.

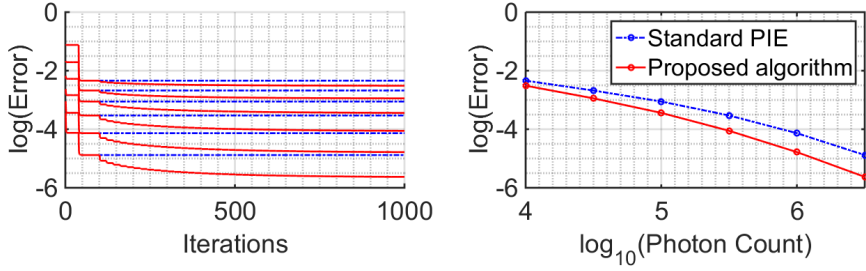


Figure 4.11: Error plots of simulated ptychographic reconstructions of object 1. Left: the reconstruction errors as a function of the number iterations for different photon counts which are indicated on the horizontal axis of the plot on the right. The dotted blue curves correspond to the standard PIE algorithm, and the solid red curves correspond to the proposed algorithm. Right: the reconstruction errors after 1000 iterations for different photon counts.

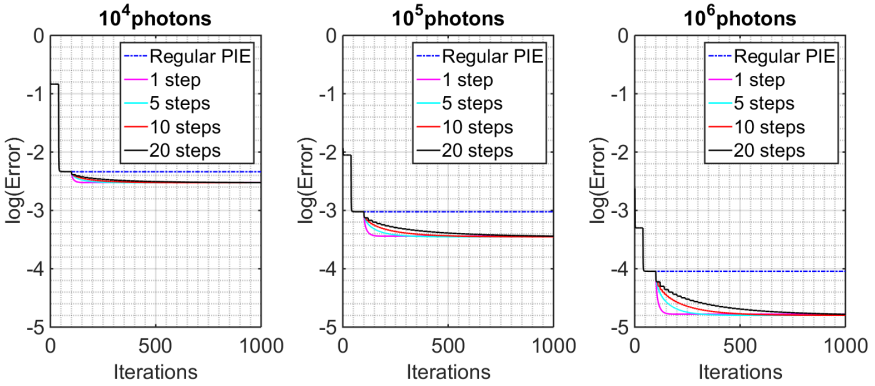
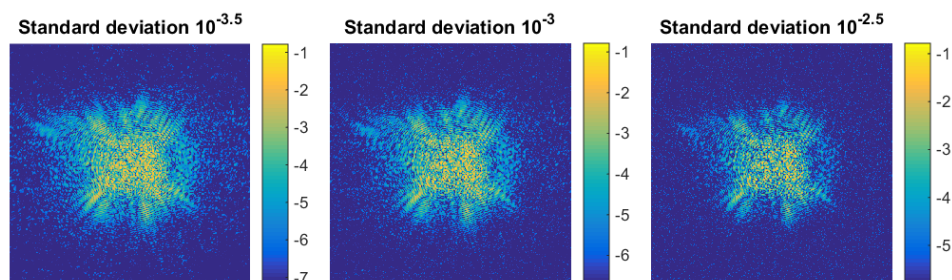
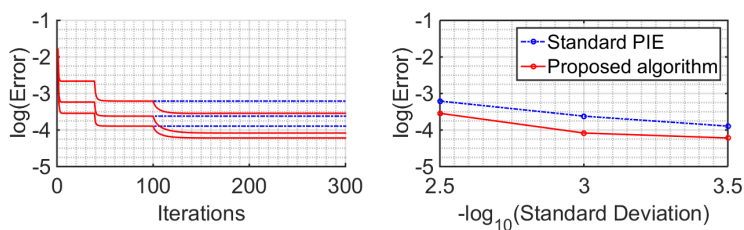


Figure 4.12: Plots showing the reconstruction errors for the proposed method for different levels of Poisson noise when different numbers of PIE iterations are applied after the intensity constraint is updated.

It has been demonstrated that this is especially important in Fourier ptychography because the dark-field images which carry information about the high spatial frequencies of the object have a very low intensity compared to the bright-field images [2]. However, in regular ptychography such a distinction is absent (although there may be a large dynamic range in each single measurement, depending on what illuminating probe function one uses), and it has even been suggested to give pixels with a higher signal-to-noise ratio (which for Poisson noise are the pixels with higher intensity) more weight [14]. Therefore, simulations were performed to test if the proposed method also works for Fourier ptychography. As shown in Fig. 4.14, a phase object with constant amplitude was used for the simulations. The simulation results presented in Fig. 4.15 indicate that also in this case the proposed method is capable of improving the reconstruction results. For example, one can visually observe that by using the proposed extension, the reconstructed amplitude becomes more constant, as it should be since the object is a pure phase object.



(a) Logarithm of the simulated intensity patterns with different amounts of additive Gaussian noise.



(b) Plots of the reconstruction error

Figure 4.13: Simulation results for ptychography with Gaussian noise. Left: the reconstruction errors as a function of the number iterations for different standard deviations of the noise which are indicated on the horizontal axis of the plot on the right. The dotted blue curves correspond to the standard PIE algorithm, and the solid red curves correspond to the proposed algorithm. Right: the reconstruction errors after 300 iterations for different standard deviations.

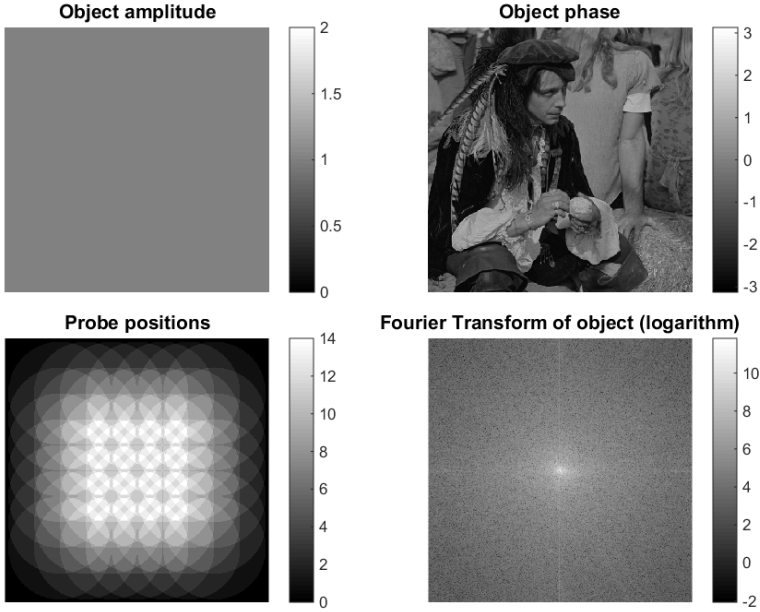


Figure 4.14: Figures showing the amplitude and phase of the object, as well as the probe positions that were used to simulate the Fourier ptychography dataset.

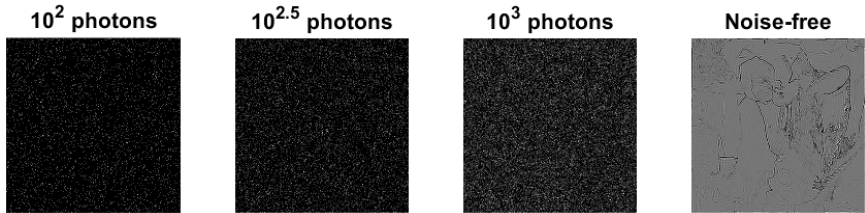
4.5. Summary

- Ptychography can be formulated as a cost-minimization problem. In the presence of noise that obeys a certain noise model, one can use the **maximum-likelihood method** or **variance stabilizing transforms** to find a cost functional that is optimized for that noise model. We showed that these two approaches are equivalent.
- Using these approaches, one finds that **an intensity-based cost functional is most suited for Gaussian noise**, while **an amplitude-based cost functional is most suited for Poisson noise**.
- We tested the accuracy of these claims by introducing a **parametrized cost functional** L_α , where $\alpha = 0.5$ corresponds to an amplitude-based cost functional, and $\alpha = 1$ corresponds to an intensity-based cost functional

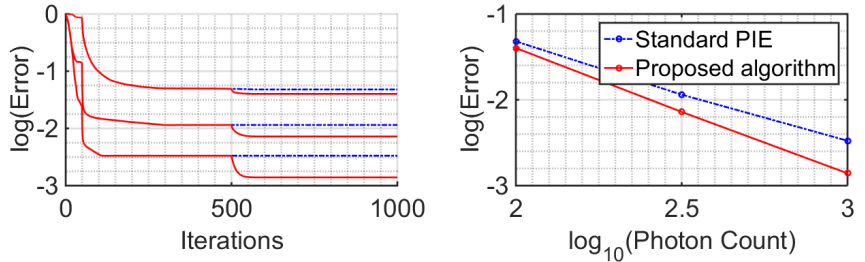
$$L_\alpha = \sum_{j,k} (z_j(k)^\alpha - y_j(k)^\alpha)^2. \quad (4.20)$$

For each α , we compared the reconstructed object to the reference object. We found that **the optimal cost-functionals are not necessarily the ones predicted by the maximum-likelihood method**.

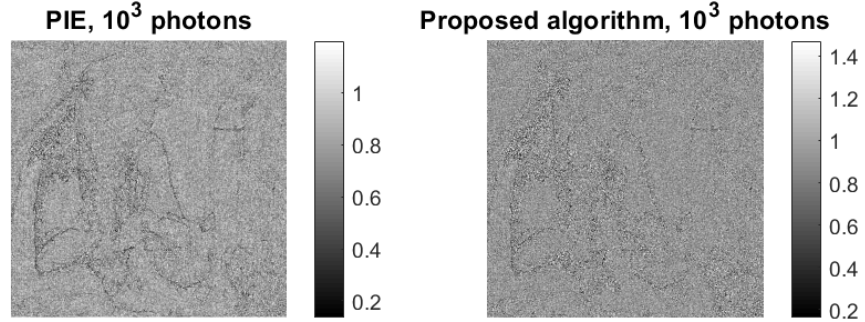
- We proposed a noise-robust method that works by **using the reconstruction to update the measurement constraints**. We verified that this



(a) Simulated intensity patterns with different amounts of Poisson noise.



(b) Plots of the reconstruction error



(c) Amplitude reconstruction

Figure 4.15: Simulation results for Fourier ptychography. One can observe a visible improvement in the amplitude of the reconstruction: since the object is a pure phase object, the amplitude should show as few features as possible.

method works using both simulations and experiment. Since the method exploits the redundancy of information in the ptychographic data set, the performance becomes worse when this information is more corrupted by noise.

References

- [1] A. Konijnenberg, W. M. J. Coene, and H. P. Urbach, *Model-independent noise-robust extension of ptychography*, *Optics Express* **26**, 5857 (2018).
- [2] L. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller, *Experimental robustness of fourier ptychography phase retrieval algorithms*, *Optics Express* **23**, 33214 (2015).
- [3] L. Bian, J. Suo, J. Chung, X. Ou, C. Yang, F. Chen, and Q. Dai, *Fourier ptychographic reconstruction using poisson maximum likelihood and truncated wirtinger gradient*, *Scientific Reports* **6** (2016), 10.1038/srep27384.
- [4] P. Thibault and M. Guizar-Sicairos, *Maximum-likelihood refinement for coherent diffractive imaging*, *New Journal of Physics* **14**, 063004 (2012).
- [5] S. Marchesini, H. Krishnan, B. Daurer, D. Shapiro, T. Perciano, J. Sethian, and F. Maia, *Sharp: a distributed gpu-based ptychographic solver*, *Journal of Applied Crystallography* **49**, 1245 (2016).
- [6] Y. Zhang, P. Song, and Q. Dai, *Fourier ptychographic microscopy using a generalized anscombe transform approximation of the mixed poisson-gaussian likelihood*, *Optics Express* **25**, 168 (2017).
- [7] P. Li, D. J. Batey, T. B. Edo, A. D. Parsons, C. Rau, and J. M. Rodenburg, *Multiple mode x-ray ptychography using a lens and a fixed diffuser optic*, *Journal of Optics* **18**, 054008 (2016).
- [8] A. Maiden, G. Morrison, B. Kaulich, A. Gianoncelli, and J. Rodenburg, *Soft x-ray spectromicroscopy using ptychography with randomly phased illumination*, *Nature Communications* **4** (2013), 10.1038/ncomms2640.
- [9] M. Guizar-Sicairos, M. Holler, A. Diaz, J. Vila-Comamala, O. Bunk, and A. Menzel, *Role of the illumination spatial-frequency spectrum for ptychography*, *Physical Review B* **86** (2012), 10.1103/physrevb.86.100103.
- [10] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, *Probe retrieval in ptychographic coherent diffractive imaging*, *Ultramicroscopy* **109**, 338 (2009).
- [11] A. Maiden and J. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, *Ultramicroscopy* **109**, 1256 (2009).
- [12] P. Dwivedi, A. Konijnenberg, S. Pereira, and H. Urbach, *Lateral position correction in ptychography using the gradient of intensity patterns*, *Ultramicroscopy* **192**, 29 (2018).

- [13] C. Wang, Z. Xu, H. Liu, Y. Wang, J. Wang, and R. Tai, *Background noise removal in x-ray ptychography*, [Applied Optics](#) **56**, 2099 (2017).
- [14] M. Guizar-Sicairos and J. Fienup, *Phase retrieval with transverse translation diversity: a nonlinear optimization approach*, [Optics Express](#) **16**, 7264 (2008).

5

Diffraction Shearing Interferometry

We discuss the problem of Diffraction Shearing Interferometry (DSI), where the measurement constraint consists of an interference term of two shifted far fields. In case the two far fields overlap, the problem reduces to regular Coherent Diffraction Imaging (CDI). We investigate how we can apply the tools and insights of CDI to solve and understand the DSI problem. We investigate the requirements for the support constraint in the case of single-shot CDI, and as an addition to the standard DSI algorithm (which is more similar to reflection-based CDI algorithms such as HIO), we propose a new DSI algorithm which is based on projections (more similar to ER).

In conventional CDI, one uses monochromatic illumination to illuminate the sample, and records the intensity of the far field, which is given by a straightforward Fourier transform

$$I(\mathbf{k}) = |\hat{\psi}(\mathbf{k})|^2. \quad (5.1)$$

However, when using Extreme Ultraviolet (EUV) light that is generated by a High Harmonic Generator (HHG), the illumination is by nature not monochromatic (though it can be made so using a monochromator, if one chooses), but rather it exists of several harmonic frequencies ν . As a result, the far field diffraction pattern is an incoherent sum of monochromatic far-field intensities

$$I(\mathbf{k}) = \sum_{\nu} |\hat{\psi}_{\nu}(\mathbf{k})|^2. \quad (5.2)$$

To separate the different frequencies, one can apply Fourier Transform Spectroscopy (FTS) [1]. In FTS, one splits the illuminating beam in two, and introduces a variable path length difference between the two beams which results in a time delay τ . These two beams illuminate the sample at slightly different angles (though it is not a fundamental requirement for FTS), thereby creating two shifted far field wavefronts that interfere with each other

$$\begin{aligned} I(\mathbf{k}) &= \sum_{\nu} \left| \hat{\psi}_{\nu} \left(\mathbf{k} + \frac{\Delta \mathbf{k}}{2} \right) + e^{2\pi i \nu \tau} \hat{\psi}_{\nu} \left(\mathbf{k} - \frac{\Delta \mathbf{k}}{2} \right) \right|^2 \\ &= \sum_{\nu} \left| \hat{\psi}_{\nu} \left(\mathbf{k} + \frac{\Delta \mathbf{k}}{2} \right) \right|^2 + \left| \hat{\psi}_{\nu} \left(\mathbf{k} - \frac{\Delta \mathbf{k}}{2} \right) \right|^2 \\ &\quad + \hat{\psi}_{\nu} \left(\mathbf{k} + \frac{\Delta \mathbf{k}}{2} \right) \hat{\psi}_{\nu} \left(\mathbf{k} - \frac{\Delta \mathbf{k}}{2} \right)^* e^{-2\pi i \nu \tau} \\ &\quad + \hat{\psi}_{\nu} \left(\mathbf{k} + \frac{\Delta \mathbf{k}}{2} \right)^* \hat{\psi}_{\nu} \left(\mathbf{k} - \frac{\Delta \mathbf{k}}{2} \right) e^{2\pi i \nu \tau}. \end{aligned} \quad (5.3)$$

By taking the inverse Fourier transform with respect to τ , one can extract for different frequencies ν the monochromatic measurement constraint [2]

$$M_{\nu}(\mathbf{k}) = \hat{\psi}_{\nu} \left(\mathbf{k} + \frac{\Delta \mathbf{k}}{2} \right) \hat{\psi}_{\nu} \left(\mathbf{k} - \frac{\Delta \mathbf{k}}{2} \right)^*. \quad (5.4)$$

Reconstructing $\psi(\mathbf{x})$ from this constraint has been termed Diffractive Shearing Interferometry (DSI). Note that if the shear vector $\Delta \mathbf{k}$ goes to 0, one finds once more the regular CDI intensity constraint of Eq. (5.1). One important difference between the CDI constraint and the DSI constraint is that the DSI constraint is a complex-valued function that contains phase information of $\hat{\psi}(\mathbf{k})$, while the CDI constraint is a real-valued function. In the following sections, we examine how we can use insights from CDI and ptychography as described in Chapter 3 to understand and solve the DSI problem. In particular, we consider the following two cases:

1. We have a measurement constraint for only one shear $\Delta \mathbf{k}$. In this case we need a support constraint to reconstruct the object. In the direction perpendicular to the shear, we have a similar situation as in regular CDI, so we assume a tight support constraint. In the direction parallel to the shear, a support constraint is needed to eliminate alternative solutions. We investigate how the required tightness of the support constraint parallel to the shear direction depends on the shear size $|\Delta \mathbf{k}|$.
2. We have multiple shears $\Delta \mathbf{k}_j$ with the same size but different directions. This case is referred to as Rotational DSI (**R-DSI**). We investigate whether it is possible to reconstruct the object from the measurement constraints without any additional constraints obtained from prior knowledge.

As a result of our investigations, a new algorithm for phase-retrieval in DSI will be derived in the next section.

5.1. Intersection of sets: applying the constraints¹

We have seen that we can solve the CDI problem by defining two constraint sets, and using an update scheme that uses projections onto these sets. In single-intensity pattern CDI these two constraints were the support constraint and intensity constraint. In ptychography these two constraints were the factorization or overlap constraint (all exit waves $f_j(\mathbf{x})$ can be factorized as $O(\mathbf{x})P(\mathbf{x} - \mathbf{X}_j)$) and the intensity constraint.

In DSI, if we consider multiple shears $\Delta \mathbf{k}_j$, we have the measurement constraint

$$M_j(\mathbf{k}) = \hat{\psi}\left(\mathbf{k} + \frac{\Delta \mathbf{k}_j}{2}\right) \hat{\psi}\left(\mathbf{k} - \frac{\Delta \mathbf{k}_j}{2}\right)^*, \quad (5.5)$$

which we can write more briefly as

$$M_j(\mathbf{k}) = \hat{\psi}_{j,+}(\mathbf{k}) \hat{\psi}_{j,-}(\mathbf{k})^*. \quad (5.6)$$

Therefore, we can define as optimization variables $\hat{f}_{j,\pm}(\mathbf{k})$, and these have to satisfy the measurement constraint

$$M_j = \hat{f}_{j,+}(\mathbf{k}) \hat{f}_{j,-}(\mathbf{k})^*, \quad (5.7)$$

and they have to satisfy the object constraint

$$\mathcal{F}^{-1}\left\{\hat{f}_{j,+}\left(\mathbf{k} - \frac{\Delta \mathbf{k}_j}{2}\right)\right\}(\mathbf{x}) = \mathcal{F}^{-1}\left\{\hat{f}_{i,-}\left(\mathbf{k} + \frac{\Delta \mathbf{k}_i}{2}\right)\right\}(\mathbf{x}) \quad \forall i, j, \quad (5.8)$$

or equivalently

$$f_{j,+}(\mathbf{x}) e^{\pi i \Delta \mathbf{k}_j \cdot \mathbf{x}} = f_{i,-}(\mathbf{x}) e^{-\pi i \Delta \mathbf{k}_i \cdot \mathbf{x}} \quad \forall i, j. \quad (5.9)$$

We will now see how we can apply the constraints.

¹The following is original work by the candidate.

5.1.1. Object constraint

One can straightforwardly estimate the exit wave by averaging all the $f_{j,\pm}(\mathbf{x})$ after a proper linear phase correction

$$f_{\text{est}}(\mathbf{x}) = \frac{1}{2J} \sum_{j=1}^J f_{j,+}(\mathbf{x}) e^{\pi i \Delta \mathbf{k}_j \cdot \mathbf{x}} + f_{j,-}(\mathbf{x}) e^{-\pi i \Delta \mathbf{k}_j \cdot \mathbf{x}}. \quad (5.10)$$

This estimate is then used to calculate the updated $\hat{f}_{j,\pm}(\mathbf{k})$

$$\hat{f}_{j,\pm}^{\text{upd}}(\mathbf{k}) = \mathcal{F} \{ f_{\text{est},j,\pm}(\mathbf{x}) e^{\mp \pi i \Delta \mathbf{k}_j \cdot \mathbf{x}} \}(\mathbf{k}). \quad (5.11)$$

5.1.2. Measurement constraint

To apply the measurement constraint, we need to find $\hat{f}_{\pm}^{\text{upd}}(\mathbf{k})$ such that

$$\hat{f}_{j,+}^{\text{upd}}(\mathbf{k}) \hat{f}_{j,-}^{\text{upd}}(\mathbf{k})^* = M_j(\mathbf{k}), \quad (5.12)$$

and

$$L(\mathbf{k}) = \left| \hat{f}_{j,+}^{\text{upd}}(\mathbf{k}) - \hat{f}_{j,+}(\mathbf{k}) \right|^2 + \left| \hat{f}_{j,-}^{\text{upd}}(\mathbf{k}) - \hat{f}_{j,-}(\mathbf{k}) \right|^2 \quad (5.13)$$

is minimized. One can calculate (see [Appendix D](#)) that if the difference between $\hat{f}_{j,\pm}^{\text{upd}}(\mathbf{k})$ and $\hat{f}_{j,\mp}(\mathbf{k})$ is small, the update reads approximately

$$\begin{aligned} \hat{f}_{j,+}^{\text{upd}}(\mathbf{k}) &= \hat{f}_{j,+}(\mathbf{k}) + \frac{(M_j(\mathbf{k}) - M_{\text{est},j}(\mathbf{k})) \hat{f}_{j,-}(\mathbf{k})}{|\hat{f}_{j,+}(\mathbf{k})|^2 + |\hat{f}_{j,-}(\mathbf{k})|^2}, \\ \hat{f}_{j,-}^{\text{upd}}(\mathbf{k}) &= \hat{f}_{j,-}(\mathbf{k}) + \frac{(M_j(\mathbf{k}) - M_{\text{est},j}(\mathbf{k}))^* \hat{f}_{j,+}(\mathbf{k})}{|\hat{f}_{j,+}(\mathbf{k})|^2 + |\hat{f}_{j,-}(\mathbf{k})|^2}, \end{aligned} \quad (5.14)$$

where $M_{\text{est},j}(\mathbf{k}) = \hat{f}_{j,+}(\mathbf{k}) \hat{f}_{j,-}(\mathbf{k})^*$. One can add a small constant to the denominators in order to prevent division by 0.

5.1.3. The standard DSI algorithm

The formula we found for applying the measurement constraint is only valid when the correction is small. Therefore, this formula is only suitable for final refinement of the reconstruction, but not for initial convergence. So what update do we use for the initial convergence of the algorithm? It turns out that the standard DSI algorithm [2] is surprisingly suitable for this.

In standard DSI the measurement constraint is applied as outlined in [2]

$$\begin{aligned} \hat{f}_{j,+}^{\text{upd}}(\mathbf{k}) &= \frac{M_j(\mathbf{k})}{\hat{f}_{j,-}(\mathbf{k})^*}, \\ \hat{f}_{j,-}^{\text{upd}}(\mathbf{k})^* &= \frac{M_j(\mathbf{k})}{\hat{f}_{j,+}(\mathbf{k})}. \end{aligned} \quad (5.15)$$

Normally, regularization is applied to prevent division by 0, but for conceptual clarity this is omitted in the above expression. Notice that we are not actually applying a projection on the measurement constraint in the sense that $\hat{f}_{j,+}^{\text{upd}}(\mathbf{k})\hat{f}_{j,-}^{\text{upd}}(\mathbf{k})^* = M_j(\mathbf{k})$. Rather

$$\begin{aligned}\hat{f}_{j,+}^{\text{upd}}(\mathbf{k})\hat{f}_{j,-}^{\text{upd}}(\mathbf{k})^* &= \frac{M_j(\mathbf{k})^2}{\hat{f}_{j,+}(\mathbf{k})\hat{f}_{j,-}(\mathbf{k})^*} \\ &= \frac{M_j(\mathbf{k})^2}{M_{j,\text{est}}(\mathbf{k})}.\end{aligned}\tag{5.16}$$

We are only applying the correct constraint when we already have the correct solution, i.e. when $M_{j,\text{est}}(\mathbf{k}) = M_j(\mathbf{k})$. When this is not the case, we are overcorrecting: if $M_{j,\text{est}}(\mathbf{k})$ is too large, we apply a constraint that is smaller than the actual $M_j(\mathbf{k})$ and vice versa. It is reminiscent of phase retrieval algorithms that use reflection operators in addition to projection operators in order to prevent the algorithm from getting stuck in local minima: a reflection around a constraint C reduces to a projection only if the optimization variable x already lies in C . If x does not lie in C , you are overcorrecting. From the simulation results shown in Fig. 5.4 one can see how standard DSI can be used for initial convergence, while an Alternating Projections scheme as defined by Eq. (5.14) is used for final refinement, analogously to how in CDI one typically uses HIO for initial convergence and ER/AP for final refinement. Alternatively, a gradient-based method has been proposed in [3].

5

5.2. Alternative solutions

We will start by considering the ideal case where $\mathbf{k}$ is a continuous variable with an infinite range. However, in practice $\mathbf{k}$ is pixelated and has a finite range due to the finite detector size. This has consequences for our analysis: for example, while in the continuous case we can choose different $\Delta\mathbf{k}$ with arbitrarily small differences, it is not obvious what it means in the discrete case to have differences in $\Delta\mathbf{k}$ that are smaller than the pixel size.

Let $\hat{\psi}(\mathbf{k})$ be the solution that satisfies Eq. (5.5). Then $\hat{A}(\mathbf{k})\hat{\psi}(\mathbf{k})$ is an alternative solution as long as

$$\hat{A}\left(\mathbf{k} + \frac{\Delta\mathbf{k}_j}{2}\right)\hat{A}\left(\mathbf{k} - \frac{\Delta\mathbf{k}_j}{2}\right)^* = 1 \quad \text{for all } j, \text{ for all } \mathbf{k} \text{ where } M_j(\mathbf{k}) \neq 0.\tag{5.17}$$

Let us assume that $M_j(\mathbf{k})$ is nonzero everywhere. If we write

$$\hat{A}(\mathbf{k}) = |\hat{A}(\mathbf{k})|e^{i\theta(\mathbf{k})},\tag{5.18}$$

then we find that the requirement in Eq. (5.17) becomes

$$\begin{aligned}\theta\left(\mathbf{k} - \frac{\Delta\mathbf{k}_j}{2}\right) - \theta\left(\mathbf{k} + \frac{\Delta\mathbf{k}_j}{2}\right) &= 0 \quad \forall \mathbf{k} \\ \left|\hat{A}\left(\mathbf{k} - \frac{\Delta\mathbf{k}_j}{2}\right)\right| \left|\hat{A}\left(\mathbf{k} + \frac{\Delta\mathbf{k}_j}{2}\right)\right| &= 1 \quad \forall \mathbf{k}.\end{aligned}\tag{5.19}$$

From this it follows for $\theta(\mathbf{k})$ that

$$\theta(\mathbf{k}) = \theta(\mathbf{k} + \Delta\mathbf{k}_j) \quad \forall \mathbf{k}. \quad (5.20)$$

It follows for $|\hat{A}(\mathbf{k})|$ that

$$\begin{aligned} |\hat{A}(\mathbf{k})| &= \frac{1}{|\hat{A}(\mathbf{k} + \Delta\mathbf{k}_j)|}, \\ |\hat{A}(\mathbf{k} + \Delta\mathbf{k}_j)| &= \frac{1}{|\hat{A}(\mathbf{k} + 2\Delta\mathbf{k}_j)|}, \end{aligned} \quad (5.21)$$

so after substituting the second equation in the first one, we get

$$|\hat{A}(\mathbf{k})| = |\hat{A}(\mathbf{k} + 2\Delta\mathbf{k}_j)|, \quad (5.22)$$

Combining the results for $\theta(\mathbf{k})$ and $|\hat{A}(\mathbf{k})|$, we can conclude that

$$\hat{A}(\mathbf{k}) = \hat{A}(\mathbf{k} + 2\Delta\mathbf{k}_j). \quad (5.23)$$

Note that this statement is weaker than the combination of Eqs. (5.20) and (5.21). By applying these equalities repeatedly for different $\Delta\mathbf{k}_j$, we can write for any set of integers N_j

$$\hat{A}(\mathbf{k}) = \hat{A}\left(\mathbf{k} + \sum_j N_j 2\Delta\mathbf{k}_j\right). \quad (5.24)$$

However, keep in mind that this argument only holds when $M_j(\mathbf{k})$ is non-zero everywhere. If $M_j(\mathbf{k})$ is zero somewhere, one cannot continue to use this recursive argument. Since in practice the detector size is finite, this will always be an issue, which complicates the search for alternative solutions. However, also near-alternative solutions (i.e. estimates that almost match the constraints) are of great interest since they can manifest themselves as local minima in which the algorithm can get trapped. Especially in the presence of noise, it may be difficult to distinguish such local minima from the actual global minimum.

Eq. (5.24) indicates that there is a grid of points set up by the $\Delta\mathbf{k}_j$ where $\hat{A}(\mathbf{k})$ should be equal to each other. If this grid covers all $\mathbf{k}$, $\hat{A}(\mathbf{k})$ is constant (and by Eq. (5.21) this constant equals 1), so any alternative solution $\hat{\psi}(\mathbf{k})\hat{A}(\mathbf{k})$ would be trivial, which means there is a non-ambiguous solution. However, because we aim to use a limited number of shears $\Delta\mathbf{k}_j$, and we ignored the finite pixel and detector size, the grid does not necessarily cover all $\mathbf{k}$. Even if they technically do, near-alternative solutions may also pose problems in the reconstruction. Since $\hat{A}(\mathbf{k})$ is expected to be periodic (with a periodicity defined by the grid points), $A(\mathbf{x})$ is expected to consist of a periodic array of delta functions, so the alternative solutions $A(\mathbf{x}) \otimes \psi(\mathbf{x})$ are expected to be a sum of a series of copies of the actual solution. In Fig. 5.4 one indeed sees multiple copies of the actual object in the reconstruction. The reason why these lie along the -45° direction, is because the shears are in the

0°, 45°, and 90° directions, so the grid described by Eq. (5.24) consists of lines along the 45° direction. Therefore, the inverse Fourier transform of $\hat{A}(\mathbf{k})$ consists of delta peaks along the -45° direction.

5.2.1. Support constraint requirements

Let us consider the case of having a measurement constraint for a single shear Δk which is in the k_x -direction, i.e.

$$\Delta \mathbf{k} = \begin{bmatrix} \Delta k_x \\ 0 \end{bmatrix}. \quad (5.25)$$

We want to find how the required width of the support constraint S_x (to eliminate alternative solutions) depends on the shear size Δk_x . From Eqs. (5.20) and (5.22) it follows that $\hat{A}(k_x, k_y)$ can be written as

$$\hat{A}(k_x, k_y) = \text{III}_{2\Delta k_x}(k_x) \otimes \hat{g}(k_x, k_y), \quad (5.26)$$

where $\otimes$ denotes a convolution with respect to k_x , $\text{III}_{2\Delta k_x}(k_x)$ denotes a Delta comb with period $2\Delta k_x$, and $\hat{g}(k_x, k_y)$ should be such that the requirements of Eq. (5.19) are met. Inverse Fourier transforming $\hat{A}(\mathbf{k})$ gives

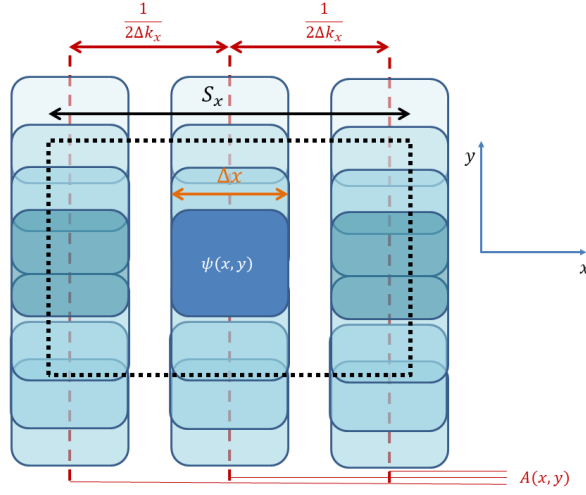
$$A(x, y) = \text{III}_{\frac{1}{2\Delta k_x}}(x)g(x, y), \quad (5.27)$$

where $g(x, y)$ is the inverse Fourier transform of $\hat{g}(k_x, k_y)$. This means that $A(\mathbf{x})$ can consist of streaks in the y -direction, spaced at intervals of $\frac{1}{2\Delta k_x}$. Now let us assume that $\psi(\mathbf{x})$ has a finite support with width Δx . Let S_x be the width of the support constraint that we impose during the reconstruction (so not the actual support of $\psi(\mathbf{x})$). How large does S_x have to be in order to rule out all alternative solutions $A(\mathbf{x}) \otimes \psi(\mathbf{x})$?

For all $A(x, y) = \text{III}_{\frac{1}{2\Delta k_x}}(x)g(x, y)$ that have an infinitely large support, any finite support constraint should in principle be sufficient to rule out the corresponding alternative solutions. However, let us assume that there are alternative solutions for which $A(x, y) = \text{III}_{\frac{1}{2\Delta k_x}}(x)g(x, y)$ has non-negligible values only for $x = 0$ and $x = \pm \frac{1}{2\Delta k}$. We assume that the values of $A(x, y)$ for $x = \pm n \frac{1}{2\Delta k}$ with $|n| > 1$ are very small (compared to the noise level of the measurement). Under this assumption, we require that S_x is tight enough so that the presence of these copies would violate the support constraint. In particular it means the following:

- If $\frac{1}{2\Delta k_x} > \Delta x$, then the copies in the x -direction do not overlap, so there is no real need to eliminate those using a support constraint (see Fig. 5.1).
- If $\frac{1}{2\Delta k_x} < \Delta x$, the width of the support constraint should satisfy

$$\Delta x \leq S_x < \frac{1}{\Delta k_x} + \Delta x, \quad (5.28)$$



5

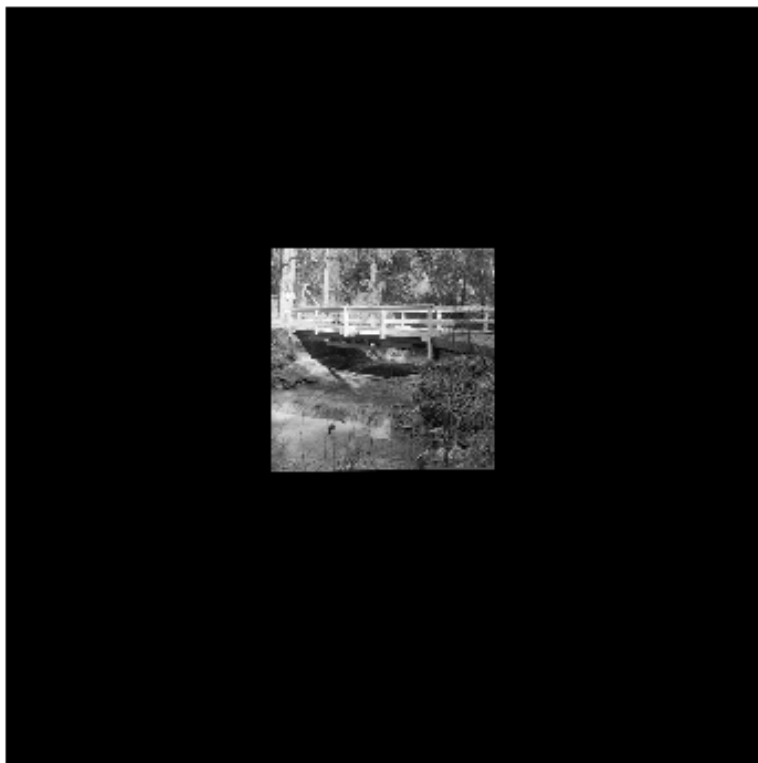
Figure 5.1: Support constraint requirements for single-constraint DSI. The sample $\psi(\mathbf{x})$ has support with width Δx in the shearing direction. The width of the support constraint S_x that is applied during the reconstruction must be such that the alternative solutions $A(\mathbf{x}) \otimes \psi(\mathbf{x})$ are excluded by the constraint. The alternative solutions consist of copies of $A(\mathbf{x})$, spaced $\frac{1}{2\Delta k_x}$ apart. Assuming that alternative solutions exist for which there are only the nearest copies, S_x should be tight enough so that the presence of these copies would violate the support constraint.

so as the shear Δk_x becomes larger, the support constraint S_x should become tighter. This has been verified with simulations, as shown in Figs. 5.2 and 5.3.

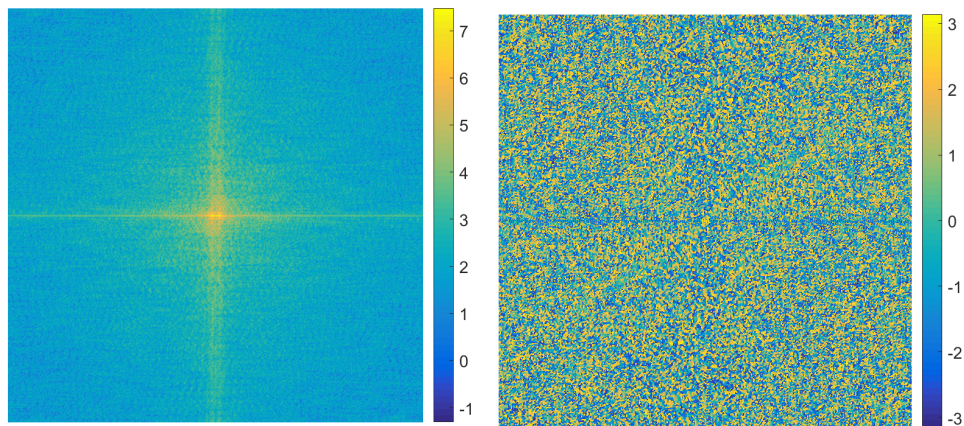
5.2.2. Introducing synthetic constraints to eliminate alternative solutions

We saw that alternative solutions are of the form $\hat{A}(\mathbf{k})\hat{\psi}(\mathbf{k})$, where $\hat{A}(\mathbf{k})$ has a small period that is defined by the shears Δk_j . Therefore, a non-trivial alternative solution should have rapid fluctuations, especially in the phase. So in order to eliminate alternative solutions, we can introduce an additional synthetic constraint, that forces the phase of $\hat{f}(\mathbf{k})$ to be smooth. We can do this as follows: if we write the constraint $M_{\Delta \mathbf{k}}(\mathbf{k})$ as

$$\begin{aligned}
 M_{\Delta \mathbf{k}}(\mathbf{k}) &= \hat{\psi}\left(\mathbf{k} + \frac{\Delta \mathbf{k}}{2}\right) \hat{\psi}\left(\mathbf{k} - \frac{\Delta \mathbf{k}}{2}\right)^* \\
 &= \left| \hat{\psi}\left(\mathbf{k} + \frac{\Delta \mathbf{k}}{2}\right) \right| \left| \hat{\psi}\left(\mathbf{k} - \frac{\Delta \mathbf{k}}{2}\right) \right| e^{i\left[\theta\left(\mathbf{k} + \frac{\Delta \mathbf{k}}{2}\right) - \theta\left(\mathbf{k} - \frac{\Delta \mathbf{k}}{2}\right)\right]},
 \end{aligned} \tag{5.29}$$



(a) Reference object (purely amplitude, 680×680 pixels).



(b) Amplitude (log scale) of the simulated measurement constraint as defined in Eq. (5.5) (shear size of 12 pixels).

(c) Phase of the simulated measurement constraint as defined in Eq. (5.5) (shear size of 12 pixels).

Figure 5.2: The simulated data that was used to test the DSI algorithm using a single measurement constraint and a support constraint. (see Fig. 5.3)

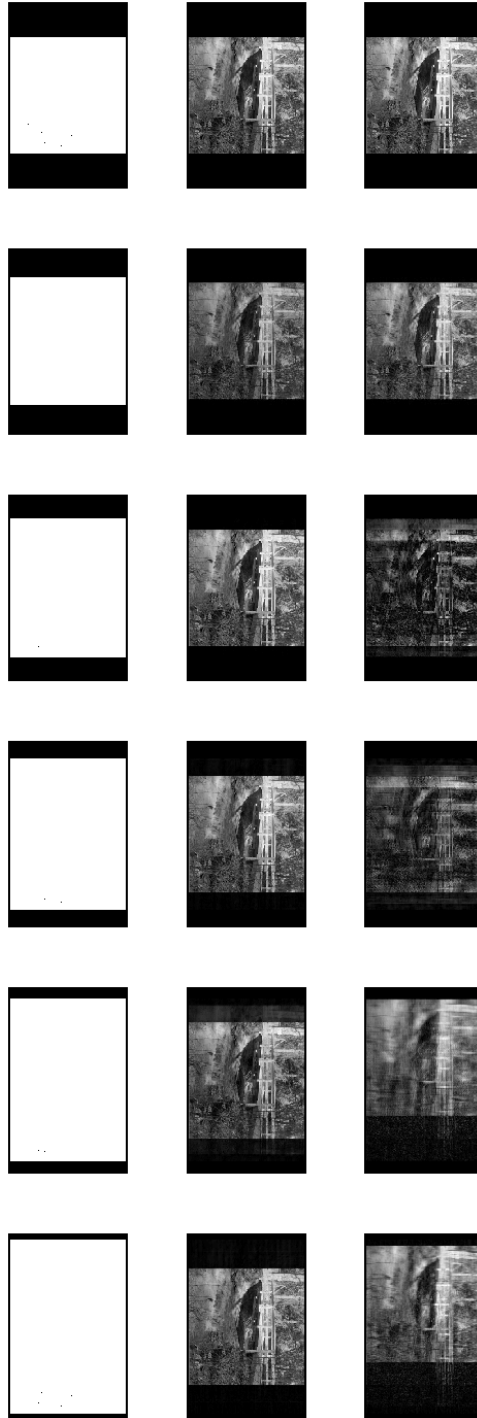


Figure 5.3: (Figure is 90° rotated) Top row: DSI reconstructions of Fig. 5.2 with a shear size of 18 pixels. Middle row: DSI reconstructions with a shear size of 12 pixels. Bottom row: support constraints enforced in the reconstruction. The results show how the support constraint should become tighter as the shear size becomes larger, as predicted by Eq. (5.28). Because for larger shears the alternative solutions lie closer together (see Fig. 5.1), one can see that in the top row the reconstruction can shift more freely around without violating the enforced support constraint.

then if $\hat{\psi}(k)$ is smooth, and α is a number close to 1, we estimate $M_{\alpha\Delta k}(k)$ to be

$$\begin{aligned} M_{\alpha\Delta k}(k) &\approx \left| \hat{\psi}\left(k + \frac{\Delta k}{2}\right) \right| \left| \hat{\psi}\left(k - \frac{\Delta k}{2}\right) \right| e^{i\alpha[\theta(k + \frac{\Delta k}{2}) - \theta(k - \frac{\Delta k}{2})]} \\ &= |M_{\Delta k}(k)| e^{i\alpha\theta_M(k)}, \end{aligned} \quad (5.30)$$

where $\theta_M(k)$ is the phase of $M_{\Delta k}(k)$. So without performing any additional measurements, we can generate an additional synthetic constraint $M_{\alpha\Delta k}(k)$ that enforces the smoothness of $\hat{f}(k)$, thereby eliminating alternative solutions $\hat{A}(k)\hat{\psi}(k)$ that have rapid oscillations. In Fig. 5.5 it is shown how the use of synthetic constraints eliminates alternative solutions.

5.3. Experimental data

The algorithms that have been explained in this chapter have been tested on experimental data² obtained using an EUV high-harmonic generator [4, 5]. More detailed information on the experiment can be found in [2]. The frequency that has been picked out using FTS is 8.7 PHz, which corresponds to a wavelength of 34 nm. The shearing angles of Δk are 0, 10, 15, 80, and 90 degrees, and the corresponding shear sizes vary from 4.3 to 5 pixels on the detector. An electron microscope (SEM) image of the sample is shown in Fig. 5.6. The sample acts as a purely absorbing amplitude sample, but this prior knowledge has not been used in the reconstruction algorithm.

Let us denote the three algorithms as

- **DSI**: The standard DSI algorithm as explained in Section 5.1.3.
- **AP**: The alternating projections scheme as explained in Section 5.1.2.
- **SDSI**: The standard DSI algorithm using additional synthetic constraints as explained in Section 5.2.2.

To test these algorithms, we performed several reconstructions using 20,000 iterations which are distributed over the three algorithms. The following cases have been tested:

1. Only DSI: 20,000 **DSI**.
2. DSI and AP final refinement: 18,000 **DSI** + 2,000 **AP**.
3. DSI, alternately DSI and AP, and AP final refinement: 2,000 **DSI** + 80×(100 **AP** + 100 **DSI**) + 2,000 **AP**.
4. 2,000 **SDSI** + 18,000 **DSI**.
5. 2,000 **SDSI** + 16,000 **DSI** + 2,000 **AP**.

²The experiment was performed by Anne de Beurs at the VU in Amsterdam under supervision of Stefan Witte.

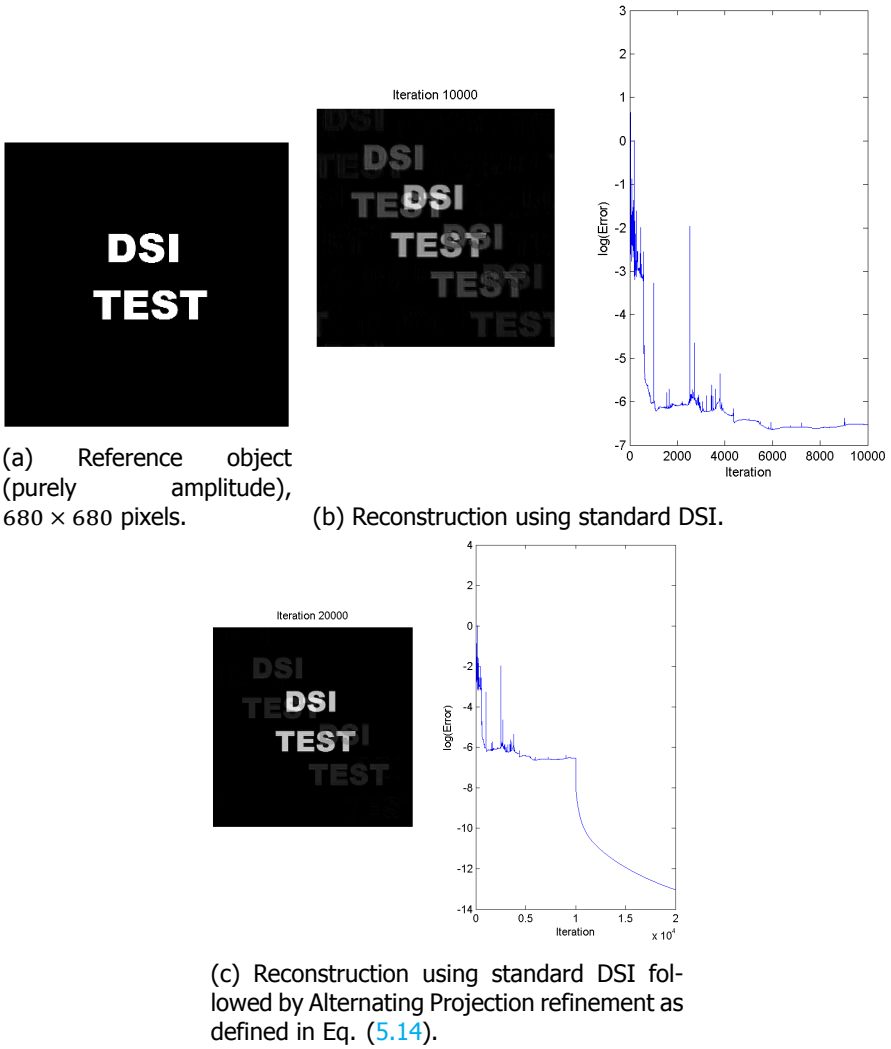
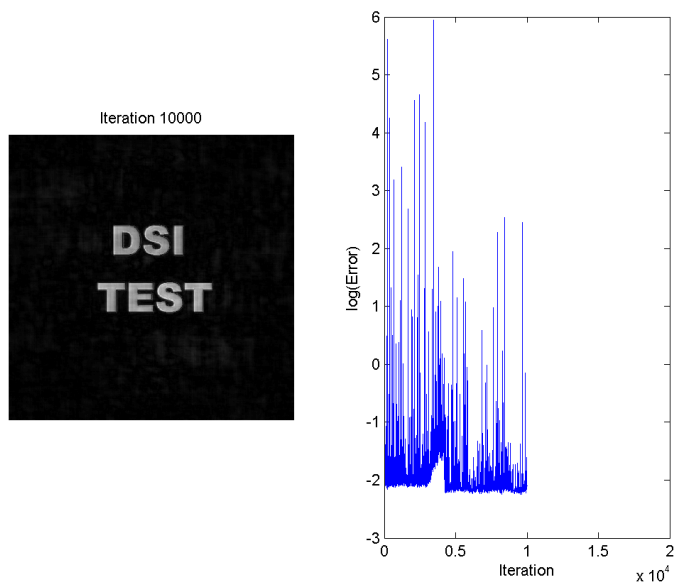
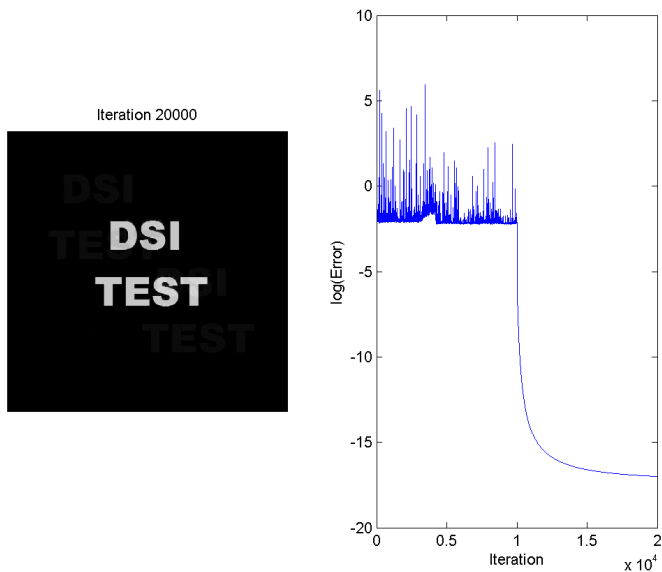


Figure 5.4: Demonstration using simulated noise-free data of how standard DSI can be used for initial convergence, and alternating projections for final refinement. The reconstruction error is defined in Fourier space (i.e. it compares the estimated $M_{\text{est}}(\mathbf{k})$ with the measured $M(\mathbf{k})$). The shear size is 6 pixels, and the shear angles are 0, 45, and 90 degrees.



(a) Reconstruction with DSI and synthetic constraints.



(b) Reconstruction with DSI and synthetic constraints, followed by alternating projections (as defined in Eq. (5.14)) without synthetic constraints.

Figure 5.5: Demonstration of how synthetic constraints can push the algorithm towards a better solution (compare to Fig. 5.4). DSI with synthetic constraints serves to find the correct local minimum, while the final refinement with alternating projections serves to achieve the actual minimal value of that local minimum.

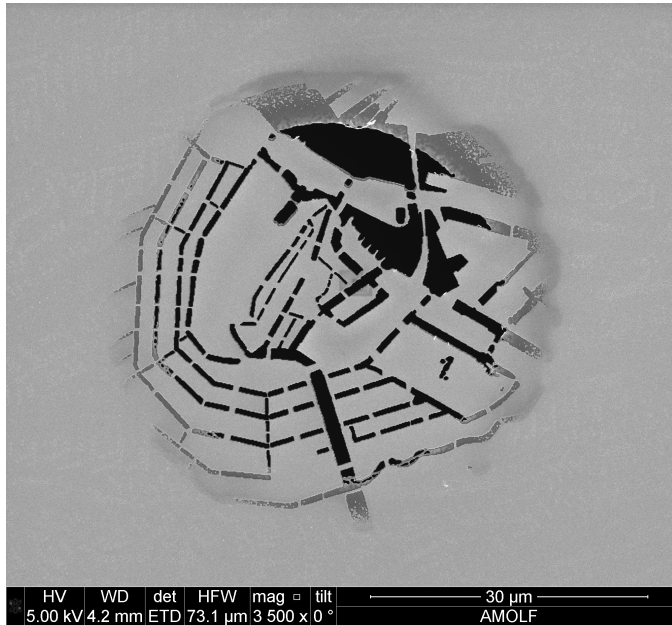


Figure 5.6: Electron microscope (SEM) image of the sample. The sample fades at the edges because it was fabricated with the intention of it having different transmission amplitudes, but it turned out only the central part transmitted light.

6. 1,000 **SDSI** + 80×(100 **DSI** + 100 **AP**) + 3,000 **AP**.

Case 1 serves as a reference, with which the results of the proposed AP and SDSI can be compared. Case 2 is used to test the added value of the AP refinement, while Case 3 tests whether there is any added value to alternately using DSI and AP (since analogously in regular CDI it is common to alternate between HIO and ER). In the reconstruction results in Fig. 5.7 one can see that in Cases 1-3 (and Case 3 especially) the algorithm tries to reconstruct the sample at two different locations. In Cases 4-6 it is tested whether SDSI can improve the reconstruction reconstruction result by eliminating this ambiguity.

The reconstructed amplitudes for these cases are shown in Figs. 5.7, 5.8, and the reconstruction errors are shown in Fig. 5.9. From the Case 3 reconstruction, one clearly sees in Fig. 5.7 that there are two solutions between which the algorithm must choose. In the Case 1 and Case 2 reconstruction, traces of the alternative solution are quite visible, whereas in Cases 4, 5, and 6 (where SDSI is used) such traces are less obvious. By comparing Case 1 with Case 2, and Case 4 with Cases 5 and 6, one sees that the AP refinement reduces the amplitude surrounding the reconstructed object. In the error plots of Fig. 5.9, it is seen that the AP refinement indeed reduces the reconstruction error. However, from Fig. 5.8 it is difficult to conclude whether AP refinement gives a visually better reconstruction.

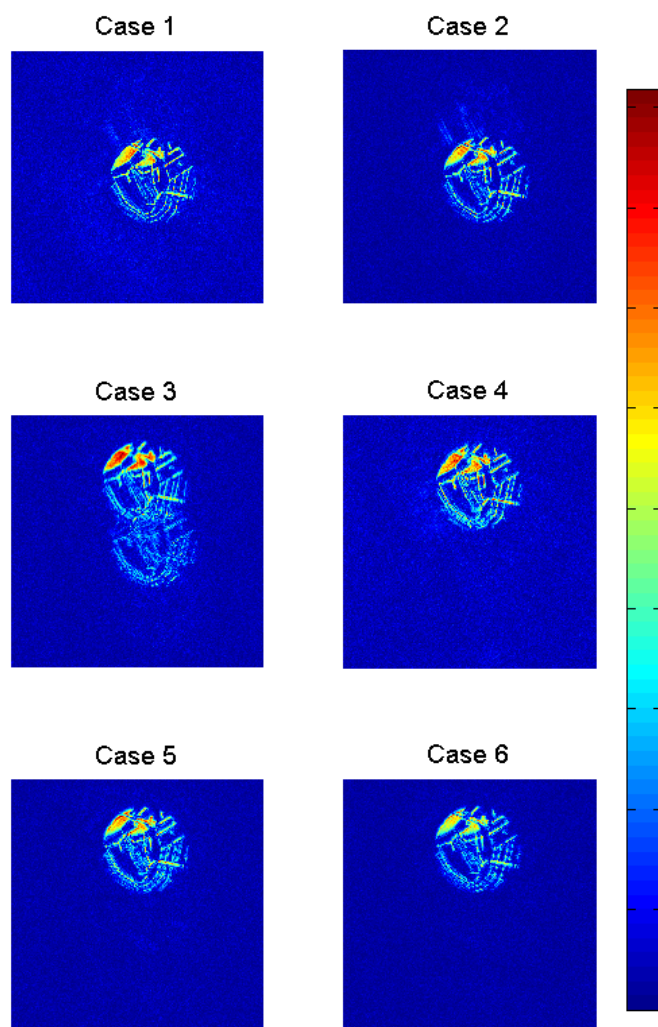


Figure 5.7: Reconstructed amplitudes obtained from the experimental DSI data using different combinations of algorithms.

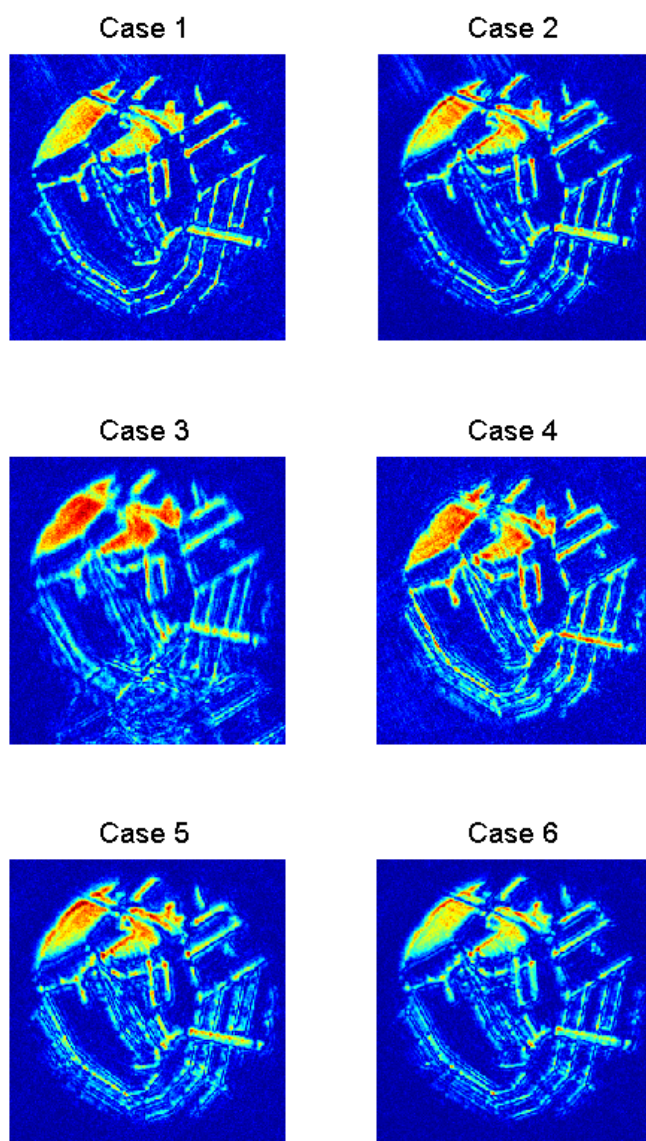
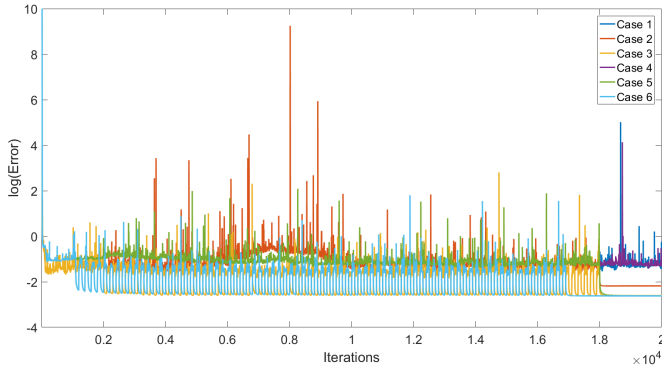
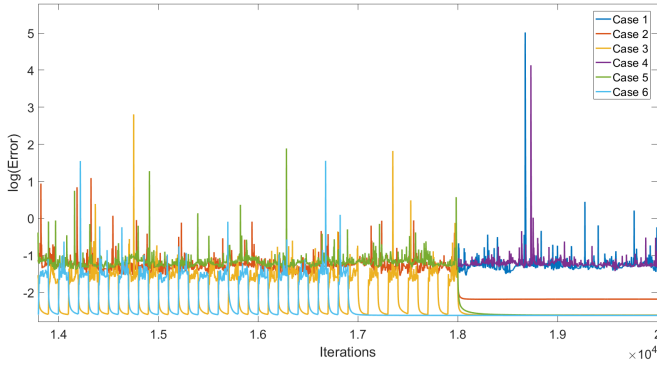


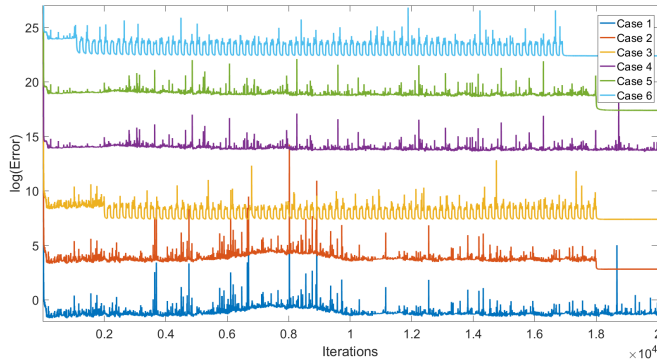
Figure 5.8: Zoom of the reconstructed amplitudes of Fig. 5.7.



(a) Reconstruction errors.



(b) Reconstruction errors (zoom).



(c) Reconstruction errors (different y-offsets).

Figure 5.9: Reconstruction errors for the DSI reconstructions. The reconstruction error is defined in Fourier space (i.e. it compares the estimated $M_{\text{est}}(\mathbf{k})$ with the measured $M(\mathbf{k})$).

5.4. Summary

- The **DSI problem** uses as a measurement constraint the product of two shifted far field wavefronts. In case the shift (or shear) is equal 0, the problem reduces to the regular phase retrieval (CDI) problem.
- Like in CDI, we can tackle the problem by defining **two constraint sets**, and finding the corresponding **projection operators**. We found the formula for a projection onto the measurement constraint set, under the approximation that the difference between the initial estimate and its projection is small. Therefore, this projection is suitable for **final refinement**.
- For **initial convergence**, the **standard DSI algorithm** turns out to be very suitable since it avoids getting trapped in local minima. This is because the measurement constraint is applied in such a way that it is **overcorrecting**, which is similar to the reflection operations that one uses in regular CDI to avoid stagnation.
- We investigated the **alternative solutions** to the DSI problems. These alternative solutions consist of multiple displaced copies of the actual object, the locations of which depend on the shear size and direction. In the case of a single DSI-constraint these alternative solutions can be eliminated by enforcing a **sufficiently tight support constraint**. In the case of Rotational DSI, one can attempt to eliminate the alternative solution by taking measurements for more shears, or by introducing **synthetic constraints** (that do not require additional measurements) which enforce the phase of the far-field to be smooth.

References

- [1] G. S. M. Jansen, D. Rudolf, L. Freisem, K. S. E. Eikema, and S. Witte, *Spatially resolved fourier transform spectroscopy in the extreme ultraviolet*, *Optica* **3**, 1122 (2016).
- [2] G. S. M. Jansen, A. de Beurs, X. Liu, K. S. E. Eikema, and S. Witte, *Diffraction shear interferometry for extreme ultraviolet high-resolution lensless imaging*, *Optics Express* **26**, 12479 (2018).
- [3] C. Falldorf, C. von Kopylow, and R. B. Bergmann, *Wave field sensing by means of computational shear interferometry*, *JOSA A* **30**, 1905 (2013).
- [4] A. McPherson, G. Gibson, H. Jara, U. Johann, T. S. Luk, I. A. McIntyre, K. Boyer, and C. K. Rhodes, *Studies of multiphoton production of vacuum-ultraviolet radiation in the rare gases*, *Journal of the Optical Society of America B* **4**, 595 (1987).
- [5] X. F. Li, A. L'Huillier, M. Ferray, L. A. Lompré, and G. Mainfray, *Multiple-harmonic generation in rare gases at high laser intensity*, *Physical Review A* **39**, 5751 (1989).

6

3D ptychography

Having investigated phase retrieval for the reconstruction of two-dimensional transmission and reflection functions, in this chapter we turn to the case of three-dimensional samples. After briefly reviewing the reconstruction methods in the literature that reconstruct three-dimensional samples, we propose a method that adds a correction term for the two-dimensional reconstruction of samples whose thickness is not negligible.

So far, we have only looked at thin samples for which the transmitted field $\psi(\mathbf{x})$ is found by simply multiplying the incident field $P(\mathbf{x})$ with a complex-valued transmission function $O(\mathbf{x})$. For samples for which this multiplicative approximation does not hold one needs a more sophisticated forward model. One such model is the multislice model (for a transmission setup), where a sample is modeled as a sequence of thin samples $O_s(\mathbf{x})$ spaced a distance Δz apart, and for each slice the multiplicative approximation holds [1]. The advantage of performing ptychography using this model is that multiple scattering is taken into consideration, because the incident field at one slice is given by the propagated scattered field of the previous slice. A disadvantage is that it is not immediately clear what the appropriate distance Δz between the slices should be. If Δz is too large then it is not appropriate to use the multiplicative approximation. If Δz is too small then the multiplicative approximation holds for two subsequent slices together, which means that if $O_s(\mathbf{x})$ and $O_{s+1}(\mathbf{x})$ are the actual transmission functions of the two slices, then $\frac{1}{f(\mathbf{x})}O_s(\mathbf{x})$ and $f(\mathbf{x})O_{s+1}(\mathbf{x})$ are also solutions since either way the total transmission function of the two slices is given by $O_s(\mathbf{x})O_{s+1}(\mathbf{x})$. The problem of finding the right slice separation has been further explored in [2], where an algorithm is proposed that updates the slice separations during the reconstruction. Another ambiguity in the reconstruction, regardless of the choice of Δz , is that each slice s can be multiplied by a constant c_s such that $\prod_s c_s = 1$, because in that case the exit wave after the sample is still the same. Moreover, it is not obvious how this method can be extended to reflective samples.

In this chapter, we propose a different method for dealing with samples whose thickness is not negligible. The method is based on the scattering integral under the single-scattering approximation (i.e. Born approximation). This integral describes how a three-dimensional optical contrast function $\chi(\mathbf{r})$ (where $\mathbf{r}$ denotes the 3D real space coordinate vector) scatters light when illuminated by a three-dimensional incident field $P(\mathbf{r})$, assuming that multiple scattering is negligible. If the sample is thin (i.e. the extent of $\chi(\mathbf{r})$ is limited in the z -direction), then one can apply a zeroth-order approximation under which the integral reduces to the familiar product $P(\mathbf{x})O(\mathbf{x})$, where $P(\mathbf{x})$ and $O(\mathbf{x})$ are two-dimensional functions. In our newly proposed method we make a first-order approximation, rather than a zeroth-order approximation, which leads to a second transmission function $\tilde{O}(\mathbf{x})$ that needs to be taken into account in the forward calculation. In doing so, we aim to reduce the error as a consequence of approximating a thick sample as a two-dimensional function. It is a subject for future research, outside the scope of this PhD thesis, to investigate how significant the contribution of the correction term is, and how it could be implemented in a reconstruction algorithm.

6.1. Deriving the scattering integral

Let us define $\mathbf{r} = [x \ y \ z]^T$ as a 3D position vector, and let $U(\mathbf{r})$ be the total field. The field should obey the Helmholtz equation

$$(\nabla^2 + k^2 n(\mathbf{r})^2) U(\mathbf{r}) = 0, \quad (6.1)$$

where $k = \frac{2\pi}{\lambda}$ is the wave number in vacuum, and $n(\mathbf{r})$ is the refractive index that varies in space. We can write $k^2 n(\mathbf{r})^2$ as a known quantity $k^2 n_b(\mathbf{r})^2$ that depends on the known background refractive index $n_b(\mathbf{r})$, minus an unknown contrast $\chi(\mathbf{r})$

$$k^2 n(\mathbf{r})^2 = k^2 n_b(\mathbf{r})^2 - \chi(\mathbf{r}). \quad (6.2)$$

This allows us to rewrite Eq. (6.1) as

$$(\nabla^2 + k^2 n_b(\mathbf{r})^2) U(\mathbf{r}) = \chi(\mathbf{r}) U(\mathbf{r}). \quad (6.3)$$

The Green's function $G(\mathbf{r})$ is defined as the solution to

$$(\nabla^2 + k^2 n_b(\mathbf{r})^2) G(\mathbf{r}) = \delta(\mathbf{r}). \quad (6.4)$$

The Green's function $G(\mathbf{r} - \mathbf{r}')$ can be interpreted as the impulse response to an excitation at the point $\mathbf{r}'$. Eq. (6.3) says that at each point $\mathbf{r}$ there is an excitation with amplitude $\chi(\mathbf{r})U(\mathbf{r})$. Summing all these impulse responses together gives an expression for the scattered field $U_s(\mathbf{r})$

$$\begin{aligned} U_s(\mathbf{r}) &= \int G(\mathbf{r} - \mathbf{r}') \chi(\mathbf{r}') U(\mathbf{r}') d\mathbf{r}' \\ &= \int G(\mathbf{r} - \mathbf{r}') \chi(\mathbf{r}') [P(\mathbf{r}') + U_s(\mathbf{r}')] d\mathbf{r}', \end{aligned} \quad (6.5)$$

where in the last line we wrote the total field $U(\mathbf{r})$ as the incident field $P(\mathbf{r})$ plus the scattered field $U_s(\mathbf{r})$. The result is a complicated implicit equation for the scattered field $U_s(\mathbf{r})$. If we apply the first Born approximation (or the single-scattering approximation), we assume that only the incident field is scattered by $\chi(\mathbf{r})$, while the scattered field itself is not scattered again by $\chi(\mathbf{r})$ (i.e. we ignore multiple scattering), which results in the explicit equation for $U_s(\mathbf{r})$

$$U_s(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}') \chi(\mathbf{r}') P(\mathbf{r}') d\mathbf{r}'. \quad (6.6)$$

If we assume the background medium to be vacuum, i.e. $n(\mathbf{r}) = 1$, then the Green's function is a spherical wave

$$G(\mathbf{r}) = \frac{e^{ik|\mathbf{r}|}}{4\pi|\mathbf{r}|}. \quad (6.7)$$

We can approximate the argument of $G(\mathbf{r} - \mathbf{r}')$ in the far field (i.e. $|\mathbf{r}| \gg |\mathbf{r}'|$) using the approximation

$$\begin{aligned}
 |\mathbf{r} - \mathbf{r}'| &= \sqrt{|\mathbf{r}|^2 + |\mathbf{r}'|^2 - 2\mathbf{r} \cdot \mathbf{r}'} \\
 &= |\mathbf{r}| \sqrt{1 + \frac{|\mathbf{r}'|^2 - 2\mathbf{r} \cdot \mathbf{r}'}{|\mathbf{r}|^2}} \\
 &\approx |\mathbf{r}| \left(1 + \frac{|\mathbf{r}'|^2 - 2\mathbf{r} \cdot \mathbf{r}'}{2|\mathbf{r}|^2} \right) \\
 &= |\mathbf{r}| + \frac{|\mathbf{r}'|^2 - 2\mathbf{r} \cdot \mathbf{r}'}{2|\mathbf{r}|} \quad \text{Fresnel} \\
 &\approx |\mathbf{r}| - \frac{\mathbf{r} \cdot \mathbf{r}'}{|\mathbf{r}|} \quad \text{Fraunhofer.}
 \end{aligned} \tag{6.8}$$

In the Fraunhofer approximation we can thus write (apart from a factor $e^{ik|\mathbf{r}|}$)

$$G(\mathbf{r} - \mathbf{r}') \approx \frac{e^{i\mathbf{K} \cdot \mathbf{r}}}{4\pi|\mathbf{r}|} e^{-i\mathbf{K} \cdot \mathbf{r}'}, \tag{6.9}$$

where $\mathbf{K}$ indicates the observation direction

$$\mathbf{K} = k \frac{\mathbf{r}}{|\mathbf{r}|}. \tag{6.10}$$

Since $|\mathbf{K}| = k$, we can write

$$\mathbf{K} = \begin{bmatrix} K_x & K_y & \sqrt{k^2 - K_x^2 - K_y^2} \end{bmatrix}^T, \tag{6.11}$$

where we assumed $K_z \geq 0$. Plugging Eq. (6.9) into Eq. (6.6) yields [3]

$$U_s(\mathbf{r}) \approx \frac{e^{i\mathbf{K} \cdot \mathbf{r}}}{4\pi|\mathbf{r}|} \int e^{-i\mathbf{K} \cdot \mathbf{r}'} \chi(\mathbf{r}') P(\mathbf{r}') d\mathbf{r}', \tag{6.12}$$

which says that under the single-scattering approximation and the far-field approximation, the scattered field is given by the 3D Fourier transform of the contrast function $\chi(\mathbf{r})$ evaluated on a sphere with radius $k = \frac{2\pi}{\lambda}$, namely the Ewald sphere.

6.2. Approximating the 3D potential as 2D objects

We can write for the incident field

$$P(x, y, z) = \int \hat{P}(k_x, k_y) e^{i(k_x x + k_y y + k_z z)} dk_x dk_y, \tag{6.13}$$

where $\hat{P}(k_x, k_y)$ denotes the Angular Spectrum of $P(x, y, 0)$, and $k_z = -\sqrt{k^2 - k_x^2 - k_y^2}$ if we assume that the field propagates in the negative z -direction. If we shift the probe laterally by

$$\mathbf{X} = [X \quad Y \quad 0]^T, \tag{6.14}$$

we can write

$$P(\mathbf{r} - \mathbf{X}) = \int \hat{P}(k_x, k_y) e^{i(k_x(x-X) + k_y(y-Y) + k_z z)} dk_x dk_y. \quad (6.15)$$

If we write the total field for a certain shift $\mathbf{X}$ as the sum of the incident field and scattered field

$$U_{\mathbf{X}}(\mathbf{r}) = P(\mathbf{r} - \mathbf{X}) + U_{s,\mathbf{X}}(\mathbf{r}), \quad (6.16)$$

we can write Eq. (6.12) (writing $U_s(\mathbf{r})$ as a function of $\mathbf{K} = k \frac{\mathbf{r}}{|\mathbf{r}|}$) as

$$\begin{aligned} U_{s,\mathbf{X}}(\mathbf{K}) &= \\ &= \frac{e^{i\mathbf{K} \cdot \mathbf{r}}}{4\pi|\mathbf{r}|} \iint e^{-i(K_x x' + K_y y' + K_z z')} \chi(x', y', z') \hat{P}(k_x, k_y) e^{i(k_x(x'-X) + k_y(y'-Y) + k_z z')} dx' dy' dz' dk_x dk_y \\ &= \\ &= \frac{e^{i\mathbf{K} \cdot \mathbf{r}}}{4\pi|\mathbf{r}|} \int V(K_x, K_y, k_x, k_y, x', y') e^{-i((K_x - k_x)x' + (K_y - k_y)y')} \hat{P}(k_x, k_y) e^{-i(k_x X + k_y Y)} dx' dy' dk_x dk_y, \end{aligned} \quad (6.17)$$

where

$$V(K_x, K_y, k_x, k_y, x', y') = \int \chi(x', y', z') e^{-i(K_z - k_z)z'} dz'. \quad (6.18)$$

Recall that

- K_x, K_y denote the observation angles,
- k_x, k_y denote the incident angles of the illumination $P(\mathbf{r})$,
- x', y', z' denote the coordinates in the scattering potential $\chi(\mathbf{r})$.

We want to make certain approximations so that $V(K_x, K_y, k_x, k_y, x', y')$ can be simplified.

6.2.1. Zeroth order approximation

Suppose the main direction of incidence is given by

$$\mathbf{k}_0 = \begin{bmatrix} k_{0x} & k_{0y} & -\sqrt{k^2 - k_{0x}^2 - k_{0y}^2} \end{bmatrix}^T, \quad (6.19)$$

and the main scattering direction is given by

$$\mathbf{K}_0 = \begin{bmatrix} K_{0x} & K_{0y} & \sqrt{k^2 - K_{0x}^2 - K_{0y}^2} \end{bmatrix}^T, \quad (6.20)$$

and the angle between the two vectors is 2θ . The sign difference between k_z and K_z indicates we are assuming a reflective sample. If we assume the spreads of

k_x, k_y and K_x, K_y are small enough, we can assume $k_z(k_x, k_y)$ and $K_z(K_x, K_y)$ to be approximately constant around k_0 and K_0 , and in particular one can make the zeroth-order approximation

$$K_z - k_z \approx K_{0z} - k_{0z} = 2k \cos(\theta). \quad (6.21)$$

In this case, $V(K_x, K_y, k_x, k_y, x', y')$ will lose its dependence on K_x, K_y , and can be reduced to a two-dimensional function $O(x', y')$ which we define as

$$O(x', y') = \int \chi(x', y', z') e^{-i2k \cos(\theta)z'} dz', \quad (6.22)$$

and the scattered field as described by Eq. (6.17) can be found using the familiar multiplicative approximation

$$U'_{s,X}(K_x, K_y) \approx \frac{e^{i\mathbf{K} \cdot \mathbf{r}}}{4\pi|\mathbf{r}|} \mathcal{F}\{O(x, y)P(x - X, y - Y)\}(K_x, K_y). \quad (6.23)$$

6

However, one must note the differences with the conventions used in ptychography:

- In ptychography, $O(x, y)$ is a dimensionless transmission function, but $O(x, y)$ as defined in Eq. (6.22) is not dimensionless. This is compensated for by the fact that the prefactor $\frac{1}{|\mathbf{r}|}$ in Eq. (6.23) differs in dimension from the prefactor $\frac{1}{\lambda z}$ for Fraunhofer propagation, which is used in ptychography.
- The result in Eq. (6.23) describes only the scattered field, whereas in ptychography $\mathcal{F}\{O(x, y)P(x - X, y - Y)\}(K_x, K_y)$ describes the total propagated field, which includes both the scattered and the unscattered field.

6.2.2. First order approximation¹

If we want a more accurate approximation, we can expand $K_z(K_x, K_y) - k_z(k_x, k_y)$ around $(k_{0x}, k_{0y}), (K_{0x}, K_{0y})$ by defining

$$\begin{aligned} k_x &= k_{0x} + k'_x & k_y &= k_{0y} + k'_y, \\ K_x &= K_{0x} + K'_x & K_y &= K_{0y} + K'_y, \end{aligned} \quad (6.24)$$

¹The following is original work by the candidate.

which gives

$$\begin{aligned}
 K_z - k_z &= \sqrt{k - K_{0x}^2 - K_{0y}^2 - K_x'^2 - K_y'^2 - 2K_x'K_{x0} - 2K_y'K_{y0}} \\
 &\quad + \sqrt{k - k_{0x}^2 - k_{0y}^2 - k_x'^2 - k_y'^2 - 2k_x'k_{x0} - 2k_y'k_{y0}} \\
 &= K_{0z} \sqrt{1 - \frac{K_x'^2 + K_y'^2 + 2K_x'K_{x0} + 2K_y'K_{y0}}{K_{0z}^2}} - k_{0z} \sqrt{1 - \frac{k_x'^2 + k_y'^2 + 2k_x'k_{x0} + 2k_y'k_{y0}}{k_{0z}^2}} \\
 &\approx K_{0z} - k_{0z} - \frac{K_x'^2 + K_y'^2 + 2K_x'K_{x0} + 2K_y'K_{y0}}{2K_{0z}} + \frac{k_x'^2 + k_y'^2 + 2k_x'k_{x0} + 2k_y'k_{y0}}{2k_{0z}} \\
 &= K_{0z} - k_{0z} - f(\mathbf{K}) + g(\mathbf{k}),
 \end{aligned} \tag{6.25}$$

where we can use the definitions of Eq. (6.24) to define

$$\begin{aligned}
 f(\mathbf{K}) &= \frac{K_x^2 + K_y^2 - K_{0x}^2 - K_{0y}^2}{2K_{0z}}, \\
 g(\mathbf{k}) &= \frac{k_x^2 + k_y^2 - k_{0x}^2 - k_{0y}^2}{2k_{0z}}.
 \end{aligned} \tag{6.26}$$

Thus we can approximate

$$\begin{aligned}
 e^{-i(K_z - k_z)z'} &\approx e^{-i(K_{0z} - k_{0z})z'} e^{-i(g(\mathbf{k}) - f(\mathbf{K}))z'} \\
 &\approx e^{-i(K_{0z} - k_{0z})z'} (1 - i(g(\mathbf{k}) - f(\mathbf{K}))z').
 \end{aligned} \tag{6.27}$$

Defining

$$\tilde{O}(x, y) = \int \chi(x, y, z') e^{-i2k \cos(\theta)z'} z' dz', \tag{6.28}$$

we can approximate $V(K_x, K_y, k_x, k_y, x, y)$ as

$$V(K_x, K_y, k_x, k_y, x, y) \approx O(x) - i(g(\mathbf{k}) - f(\mathbf{K}))\tilde{O}(x) \tag{6.29}$$

We saw that the term $O(x)$ led to the familiar multiplicative approximation. For the second term of V we find we can approximate Eq. (6.17) as

$$\begin{aligned}
 if(\mathbf{K}) &\int \tilde{O}(x', y') e^{-i((K_x - k_x)x' + (K_y - k_y)y')} \hat{P}(k_x, k_y) e^{-i(k_x X + k_y Y)} dx' dy' dk_x dk_y \\
 &\quad - i \int \tilde{O}(x', y') g(\mathbf{k}) e^{-i((K_x - k_x)x' + (K_y - k_y)y')} \hat{P}(k_x, k_y) e^{-i(k_x X + k_y Y)} dx' dy' dk_x dk_y \\
 &= if(\mathbf{K}) \mathcal{F}\{\tilde{O}(x, y) P(x - X, y - Y)\}(K_x, K_y) \\
 &\quad - i \int \tilde{O}(x', y') e^{-i((K_x - k_x)x' + (K_y - k_y)y')} \hat{P}(k_x, k_y) g(\mathbf{k}) e^{-i(k_x X + k_y Y)} dx' dy' dk_x dk_y.
 \end{aligned} \tag{6.30}$$

Observing that from Eq. (6.26) it follows

$$g(\mathbf{k})e^{-i(k_x X + k_y Y)} = -\frac{1}{2k_{0z}} \left(\frac{d^2}{dX^2} + \frac{d^2}{dY^2} \right) e^{-i(k_x X + k_y Y)} - \frac{k_{0x}^2 + k_{0y}^2}{2k_{0z}} e^{-i(k_x X + k_y Y)}, \quad (6.31)$$

we can write (apart from a factor $\frac{e^{i\mathbf{K} \cdot \mathbf{r}}}{4\pi|\mathbf{r}|}$)

$$U'_{s,\mathbf{X}}(\mathbf{K}) \approx \mathcal{F}\{O(x,y)P(x-X,y-Y)\}(K_x,K_y) + i \left(f(\mathbf{K}) + \frac{k_{0x}^2 + k_{0y}^2}{2k_{0z}} + \frac{1}{2k_{0z}} \frac{d^2}{dX^2} + \frac{1}{2k_{0z}} \frac{d^2}{dY^2} \right) \mathcal{F}\{\tilde{O}(x,y)P(x-X,y-Y)\}(K_x,K_y). \quad (6.32)$$

Note that to compute the total far field, one must add the unscattered field $\mathcal{F}\{P(x-X,y-Y)\}(K_x,K_y)$. Thus, we see that under the approximations

- the spread of incident angles $\mathbf{k}$ is small,
- the spread of scattered angles $\mathbf{K}$ is small,
- the integration range of z' is small,
- the effects of multiple scattering are small,

we can reconstruct a 3D object by reconstructing a 2D object $O(x,y)$ (which is a zeroth order approximation), the phase of which contains information about the height 3D structure [4]. It is important to observe that the same assumptions that make the 2D approximation hold (small spread of incident and scattered angles) also limit the resolution with which the object can be reconstructed. As an improvement to that approximation, $\tilde{O}(x,y)$ can be introduced.

6.3. A real-space derivation

In Eq. (6.32) we presented the approximate result for the scattered far field assuming single scattering. This result was derived in Fourier space (i.e. $\mathbf{K}$ -space). In this section, we present another derivation for the case of a transmissive sample that is performed in real space (i.e. $\mathbf{r}$ -space). To do this, let us first make some simplifying assumptions that allow us to rewrite Eq. (6.32):

- We assume a transmissive sample, so that Eqs. (6.22) and (6.28) become

$$\begin{aligned} O(x,y) &= \int \chi(x,y,z') dz', \\ \tilde{O}(x,y) &= \int \chi(x,y,z') z' dz'. \end{aligned} \quad (6.33)$$

- We assume that the incident field propagates in the z -direction so that $k_{0x} = k_{0y} = K_{0x} = K_{0y} = 0$ and $k_{0z} = K_{0z} = k$.

- We consider only a single probe position $\mathbf{X} = 0$.
- We define $K_{\perp}^2 = K_x^2 + K_y^2$.
- We define $\bar{\Delta} = \frac{i}{2k} \left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} \right)$

Under these assumptions, we can rewrite Eq. (6.32) as

$$U'_s(\mathbf{K}) \approx \mathcal{F}\{O(\mathbf{x})P(\mathbf{x})\}(\mathbf{K}) + \left(i \frac{K_{\perp}^2}{2k} \mathcal{F}\{\tilde{O}(\mathbf{x})P(\mathbf{x})\}(\mathbf{K}) + \mathcal{F}\{\tilde{O}(\mathbf{x})\bar{\Delta}P(\mathbf{x})\}(\mathbf{K}) \right). \quad (6.34)$$

Note that in the final term, $\bar{\Delta}$ acts only on P because the second derivatives are taken with respect to the probe position X, Y in Eq. (6.32). Taking the inverse Fourier transform of this expression gives the transmitted exit wave

$$\Psi(\mathbf{x}) \approx O(\mathbf{x})P(\mathbf{x}) - \bar{\Delta} [\tilde{O}(\mathbf{x})P(\mathbf{x})] + \tilde{O}(\mathbf{x})\bar{\Delta}P(\mathbf{x}). \quad (6.35)$$

We will reproduce this expression using a real-space derivation. The derivation consists of finding an approximate solution to the inhomogeneous paraxial Helmholtz equation for a scattering potential $\chi(\mathbf{r})$. We will compare two different methods to solve the equation: one is by explicitly applying the single-scattering approximation, the other is by Picard iterations, which is the same method as used in [5].

6

6.3.1. Paraxial Helmholtz equation

We assume that an optical monochromatic scalar field $\psi(\mathbf{r})$ is propagating mainly in the z -direction, so that we can write it as a modulation $\phi(\mathbf{r})$ to the incident plane wave e^{ikz}

$$\psi(\mathbf{r}) = \phi(\mathbf{r})e^{ikz}, \quad (6.36)$$

where $\mathbf{r} = [x \ y \ z]^T$, and where k denotes the wave number $k = 2\pi/\lambda$, where λ is the wavelength. Under the paraxial approximation and in the absence of a scattering potential, the field obeys the free-space paraxial Helmholtz equation [6]

$$\begin{aligned} \Delta\psi + 2ik\frac{\partial\psi}{\partial z} + 2k^2\psi &= 0, \\ \Delta\phi + 2ik\frac{\partial\phi}{\partial z} &= 0, \end{aligned} \quad (6.37)$$

where Δ denotes the Laplacian in the transverse coordinates

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \quad (6.38)$$

This equation can be derived from the regular Helmholtz equation $(\Delta + k^2)\psi = 0$ by substituting Eq. (6.36) and neglecting the term $\frac{\partial^2\phi}{\partial z^2}$. The Fresnel diffraction integral is an exact solution to the paraxial Helmholtz equation [6].

In case we do not assume free-space propagation, but rather propagation through a potential (or optical contrast) $\chi(\mathbf{r})$, the equation becomes

$$\Delta\phi + 2ik\frac{\partial\phi}{\partial z} - \chi\phi = 0. \quad (6.39)$$

We can rewrite this as

$$\frac{\partial\phi}{\partial z} = (\bar{\Delta} + \bar{\chi})\phi, \quad (6.40)$$

where

$$\begin{aligned} \bar{\Delta} &= \frac{i}{2k}\Delta, \\ \bar{\chi} &= -\frac{i}{2k}\chi. \end{aligned} \quad (6.41)$$

Free-space propagation is described by

$$\phi(x, y, z) = e^{\bar{\Delta}z}\phi(x, y, 0). \quad (6.42)$$

Propagation through a potential χ is given by

$$\phi(x, y, z) = e^{\bar{\Delta}z + \int_0^z \bar{\chi}(x, y, z') dz'} \phi(x, y, 0). \quad (6.43)$$

6.3.2. Single scattering approximation²

Let us assume that we have an incident field $\phi(\mathbf{r})$ propagating in the z -direction. In the plane $z = 0$, the field is described by $\phi(x, y, 0) = P(\mathbf{x})$, where $\mathbf{x} = [x \ y]^T$. The field is incident on a slice with thickness ϵ that has a scattering potential $\chi(x, y, z)$, as shown in Fig. 6.1. To calculate the scattered exit wave in the single-scattering approximation, we must do the following:

1. Propagate the incident field to a plane z assuming free-space propagation. This yields $e^{\bar{\Delta}z}\phi(x, y, 0)$.
2. Multiply the propagated field by the potential in that plane $\bar{\chi}_z(x, y) = \bar{\chi}(x, y, z)$ (the reason why we multiply with $\bar{\chi}$ instead of χ is explained in [Appendix E](#)). This yields $\bar{\chi}_z(x, y)e^{\bar{\Delta}z}\phi(x, y, 0)$.
3. Propagate the multiplied field by the remaining distance $\epsilon - z$ to the end of the slice assuming free-space propagation. This yields $e^{\bar{\Delta}(\epsilon-z)}\bar{\chi}_z(x, y)e^{\bar{\Delta}z}\phi(x, y, 0)$.
4. Do steps 1-3 for all z in $[0, \epsilon]$ and add the results together. The resulting exit wave in the Fresnel approximation and single-scattering approximation is

$$\phi(x, y, \epsilon) = \int_0^\epsilon e^{\bar{\Delta}(\epsilon-z)}\bar{\chi}_z(x, y)e^{\bar{\Delta}z} dz \phi(x, y, 0). \quad (6.44)$$

²The following is original work by the candidate.

One can approximate for small z and $\epsilon - z$

$$\begin{aligned} e^{\bar{\Delta}z} &\approx 1 + \bar{\Delta}z, \\ e^{\bar{\Delta}(\epsilon-z)} &\approx 1 + \bar{\Delta}\epsilon - \bar{\Delta}z. \end{aligned} \quad (6.45)$$

We can define $O(x)$ and $\tilde{O}(x)$ as in Eq. (6.33) (apart from a factor $\frac{i}{2k}$), and addition we can define $\tilde{\tilde{O}}(x)$ as

$$\begin{aligned} O(x) &= \int_0^\epsilon \bar{\chi}(x, y, z') dz', \\ \tilde{O}(x) &= \int_0^\epsilon \bar{\chi}(x, y, z') z' dz', \\ \tilde{\tilde{O}}(x) &= \int_0^\epsilon \bar{\chi}(x, y, z') z'^2 dz'. \end{aligned} \quad (6.46)$$

Using the fact that $\phi(x, y, 0) = P(x)$, we can approximate Eq. (6.44) to find

$$\phi(x, y, \epsilon) \approx (1 + \epsilon \bar{\Delta}) [O(x)P(x)] - \bar{\Delta} [\tilde{O}(x)P(x)] + (1 + \epsilon \bar{\Delta}) [\tilde{O}(x)\bar{\Delta}P(x)] - \bar{\Delta} [\tilde{\tilde{O}}(x)\bar{\Delta}P(x)]. \quad (6.47)$$

If we ignore the terms with ϵ and $\tilde{\tilde{O}}(x)$, we find

$$\phi(x, y, \epsilon) \approx O(x)P(x) - \bar{\Delta} [\tilde{O}(x)P(x)] + \tilde{O}(x)\bar{\Delta}P(x), \quad (6.48)$$

which is the same as Eq. (6.35), apart from a factor $\frac{i}{2k}$ that is absorbed in the definitions of $O, \tilde{O}, \tilde{\tilde{O}}$.

6.3.3. Picard Iterations

Another method to solve Eq. (6.40) is by using Picard Iterations [5], which is the same method that is used to derive the Born series. In this method, one rewrites Eq. (6.40) as an integral equation

$$\phi(x, y, z) = \phi(x, y, 0) + \int_0^z (\bar{\Delta} + \bar{\chi}(x, y, z')) \phi(x, y, z') dz', \quad (6.49)$$

(where one can substitute $\phi(x, y, 0)$ with $P(x, y)$) and iteratively compute estimates of $\phi(x, y, z)$ as follows

$$\phi_{n+1}(x, y, z) = P(x, y) + \int_0^z (\bar{\Delta} + \bar{\chi}(x, y, z')) \phi_n(x, y, z') dz'. \quad (6.50)$$

We choose

$$\phi_0(x, y, z) = P(x, y), \quad (6.51)$$

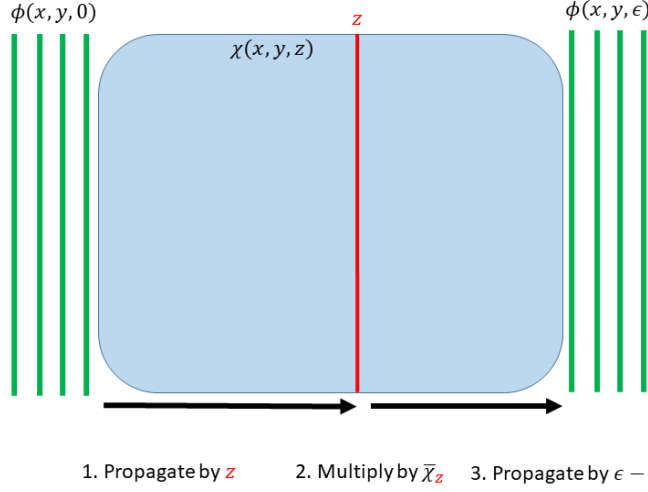


Figure 6.1: Illustration of how to calculate the transmitted field through a slice of width ϵ in the single-scattering approximation.

6

and we introduce the following shorthand notation

$$\begin{aligned}
 \phi_n(x, y, z) &\rightarrow \phi, \\
 P(x, y) &\rightarrow P, \\
 \bar{\chi}(x, y, z) &\rightarrow \bar{\chi}_z, \\
 O_z &= \int_0^z \bar{\chi}(x, y, z') dz', \\
 \tilde{O}_z &= \int_0^z \bar{\chi}(x, y, z') z' dz'.
 \end{aligned} \tag{6.52}$$

Note that here O_z and $\tilde{O}_z$ are functions of z , as opposed to the definitions introduced in Eq. (6.46). We find the following sequence

$$\phi_0 = P, \tag{6.53}$$

$$\phi_1 = P + \int_0^z (\bar{\Delta} + \bar{\chi}_{z'}) P dz' \tag{6.54}$$

$$= (1 + z\bar{\Delta})P + O_z P \tag{6.55}$$

$$\phi_2 = P + \int_0^z (\bar{\Delta} + \bar{\chi}_{z'}) [(1 + z'\bar{\Delta})P + O_{z'} P] dz' \tag{6.56}$$

$$= \left(1 + z\bar{\Delta} + \frac{(z\bar{\Delta})^2}{2}\right) P + (1 + \bar{\Delta}z) O_z P + \tilde{O}_z \bar{\Delta} P - \bar{\Delta}(\tilde{O}_z P) + \frac{1}{2} O_z^2 P. \tag{6.57}$$

To obtain the final line, we have used the following results [5]

$$\begin{aligned} \int_0^z O_{z'} dz' &= zO_z - \tilde{O}_z, \\ \int_0^z \tilde{\chi}_{z'} O_{z'} dz' &= \frac{1}{2} O_z^2. \end{aligned} \quad (6.58)$$

In the final expression for ϕ_z , one recognizes in the first term the start of the Taylor expansion of $e^{\Delta z P}$, which corresponds to the unscattered propagated field. The following three terms are also found in Eq. (6.47), and can be associated with single-scattering effects since they depend linearly on χ . The final term depends quadratically on $\tilde{\chi}$, and can therefore be associated with second-order multiple scattering effects.

6.3.4. Application to inverse problems: multislice method³

Let us assume we have a sample with thickness T , i.e. $\chi(x, y, z)$ is nonzero from $z = 0$ to $z = T$. By illuminating it with a field $P(x, y)$ and measuring the scattered field, we want to reconstruct χ . However, according to Eq. (6.43), the field in the plane immediately behind the sample is given by

$$\begin{aligned} \phi(x, y, T) &= e^{\tilde{\Delta}T + \int_0^T \tilde{\chi}(x, y, z') dz'} P(x, y) \\ &= e^{\tilde{\Delta}T + O_T} P(x, y). \end{aligned} \quad (6.59)$$

Since the propagation operator depends only on O_T , which is the integral of $\tilde{\chi}(x, y, z)$ over the total thickness $0 < z < T$, one can only hope to reconstruct O_T but not $\tilde{\chi}(x, y, z)$, no matter how many different $P(x, y)$ are used. Nevertheless, many successful ptychographic reconstructions using the multislice method have been reported in the literature [1, 7–9]. How can this be the case?

A possible reason could be that the multislice model makes an assumption that in fact imposes an additional constraint (i.e. prior knowledge) that eliminates alternative solutions. The slice width ϵ has to be chosen carefully so that the correct constraint is enforced. To explain this argument, consider the following expression

$$e^{\tilde{\Delta}T + O_T} \approx e^{\int_{T-\epsilon}^T \tilde{\chi}_z dz} e^{\tilde{\Delta}\epsilon} e^{\int_{T-2\epsilon}^{T-\epsilon} \tilde{\chi}_z dz} e^{\tilde{\Delta}\epsilon} \dots e^{\tilde{\Delta}\epsilon} e^{\int_{\epsilon}^{2\epsilon} \tilde{\chi}_z dz} e^{\tilde{\Delta}\epsilon} e^{\int_0^{\epsilon} \tilde{\chi}_z dz}. \quad (6.60)$$

This expression formulates the multislice model: it says that to a good approximation one can propagate the field through the sample by alternately applying free space propagation over a distance ϵ (which is done by applying $e^{\tilde{\Delta}\epsilon}$), and multiplying the field with a transmission function (which is done by applying $e^{\int_{z_0}^{z_0+\epsilon} \tilde{\chi}_z dz}$). In this expression lies the assumption that propagation over a distance ϵ through the potential $\tilde{\chi}$ can be approximated as

$$e^{\tilde{\Delta}\epsilon + O_\epsilon} \approx e^{\tilde{\Delta}\epsilon} e^{O_\epsilon}. \quad (6.61)$$

³The following is original work by the candidate.

However, the Baker-Campbell-Hausdorff (BCH) formula states that

$$e^{\bar{\Delta}\epsilon} e^{O_\epsilon} = e^{\bar{\Delta}\epsilon + O_\epsilon + \frac{\epsilon}{2}[\bar{\Delta}, O_\epsilon] + \dots}, \quad (6.62)$$

where the straight brackets $[\cdot]$ denote the commutator. Thus, the multislice model implicitly enforces the constraint that $\frac{\epsilon}{2}[\bar{\Delta}, O_\epsilon]$ is negligible. If ϵ is chosen too large, then this constraint is invalid, so it eliminates the actual solution from the set of allowed solutions. If ϵ is chosen too small, then the constraint becomes trivial, so it fails to eliminate alternative solutions. For example, the Lie product formula states

$$e^{\bar{\Delta}T + O_T} = \lim_{N \rightarrow \infty} \left(e^{\bar{\Delta} \frac{T}{N}} e^{\frac{O_T}{N}} \right)^N, \quad (6.63)$$

so if one chooses $\epsilon = \frac{T}{N}$, with $N \rightarrow \infty$, then we can write

$$\begin{aligned} e^{\bar{\Delta}T + O_T} &= \lim_{\epsilon \rightarrow 0} \left(e^{\bar{\Delta}\epsilon} e^{\epsilon \frac{O_T}{T}} \right)^{\frac{T}{\epsilon}} \\ &= \lim_{\epsilon \rightarrow 0} \left(e^{\bar{\Delta}\epsilon} e^{\int_{z_0}^{z_0+\epsilon} \frac{O_T}{T} dz} \right)^{\frac{T}{\epsilon}}, \end{aligned} \quad (6.64)$$

from which we see that $\bar{\chi}_z = \frac{O_T}{T}$ is a trivial alternative solution.

In [5], an improved multislice model is proposed

$$e^{\bar{\Delta}\epsilon + O_\epsilon} \approx e^{\tilde{O}_\epsilon/\epsilon} e^{\bar{\Delta}\epsilon} e^{O_\epsilon - \tilde{O}_\epsilon/\epsilon}. \quad (6.65)$$

While this scheme may make the forward calculation more accurate or less computationally expensive, it remains to be seen whether such a scheme will be beneficial for solving the inverse problem. After all, we just argued that an exact forward model allows for many alternative solutions, and that an approximate forward model must be used in order to enforce a constraint that eliminates those alternative solutions. In that light, one may question the added value of an improved forward model (such as reducing the slice thickness ϵ).

One way to eliminate alternative solutions is by performing measurements at different angles. In [3] this was done in a Fourier ptychographic setup to reconstruct a thick sample in the single scattering approximation, and in [10, 11] the multislice method was combined with a tilt-series measurement. By performing measurements at different angles, one can in principle reconstruct $\int \bar{\chi}_z dz$ for different z -axes, and essentially perform limited-angle tomography using these projections to reconstruct $\bar{\chi}$. If the projections $\int \bar{\chi}_z dz$ are reconstructed with the multislice method, one can take multiple scattering into account. In such a case, where one does not completely rely on the multislice approximation to eliminate alternative solutions, an improved multislice model may help in improving the reconstruction quality or reducing the computation time.

6.4. Summary

- In regular ptychography, the transmitted exit wave is calculated using the **multiplicative approximation** $\psi(\mathbf{x}) = P(\mathbf{x} - \mathbf{X})O(\mathbf{x})$, which holds for thin samples. In this chapter, we have looked at **higher-order approximations** which may be useful for samples whose thickness is not negligible. All the results are derived using the **single-scattering (i.e. first Born) approximation**.
- The first derivation is a **Fourier-space derivation**. The scattered far field is expressed as a function of the scattering potential and the incident field. By **Taylor expanding the z -components of the wave vectors of incident and scattered fields** in terms of the x - and y -components, we can simplify the expression. A zeroth order expansion results in the familiar multiplicative approximation, while a first order expansion yields additional correction terms. The derivation applies to both reflective and transmissive samples.
- The second and third derivations are **real-space derivations** for transmissive samples. It consists of finding an approximate solution to the **paraxial inhomogeneous Helmholtz equation**.
- One way to approximately solve the paraxial inhomogeneous Helmholtz equation is by explicitly applying the **single-scattering approximation**. The solution is found by propagating the incident field to a plane z inside the sample, multiplying it with the scattering potential, and propagating the result to the exit plane of the sample. By adding the results for all z inside the sample, one finds the scattered exit wave.
- Another way to approximately solve the paraxial inhomogeneous Helmholtz equation is by using **Picard iterations**. By rewriting the differential equation as an integral equation, one can iteratively improve the solution through a recursion relation.
- It has been demonstrated that **the three methods give very similar results**, which is evidence for their validity.
- It has been briefly discussed how the derived results may play a role in solving inverse problems. Some remarks have been made regarding the multislice method (in particular about how **the multislice approximation imposes a constraint that eliminates alternative solutions** if the slice thickness is chosen appropriately), but it remains a topic of further research whether and how the results presented in this chapter can be implemented in a practical manner.

References

- [1] A. M. Maiden, M. J. Humphry, and J. Rodenburg, *Ptychographic transmission microscopy in three dimensions using a multi-slice approach*, JOSA A **29**, 1606 (2012).

- [2] E. H. Tsai, I. Usov, A. Diaz, A. Menzel, and M. Guizar-Sicairos, *X-ray ptychography with extended depth of field*, *Optics express* **24**, 29089 (2016).
- [3] R. Horstmeyer, J. Chung, X. Ou, G. Zheng, and C. Yang, *Diffraction tomography with fourier ptychography*, *Optica* **3**, 827 (2016).
- [4] B. Zhang, D. F. Gardner, M. D. Seaberg, E. R. Shanblatt, H. C. Kapteyn, M. M. Murnane, and D. E. Adams, *High contrast 3d imaging of surfaces near the wavelength limit using tabletop EUV ptychography*, *Ultramicroscopy* **158**, 98 (2015).
- [5] D. v. Dyck, *Fast computational procedures for the simulation of structure images in complex or disordered crystals: a new approach*, *Journal of Microscopy* **119**, 141 (1980).
- [6] R. Grella, *Fresnel propagation and diffraction and paraxial wave equation*, *Journal of Optics* **13**, 367 (1982).
- [7] T. M. Godden, R. Suman, M. J. Humphry, J. M. Rodenburg, and A. M. Maiden, *Ptychographic microscope for three-dimensional imaging*, *Optics Express* **22**, 12513 (2014).
- [8] S. Gao, P. Wang, F. Zhang, G. T. Martinez, P. D. Nellist, X. Pan, and A. I. Kirkland, *Electron ptychographic microscopy for three-dimensional imaging*, *Nature Communications* **8** (2017), 10.1038/s41467-017-00150-1.
- [9] H. Öztürk, H. Yan, Y. He, M. Ge, Z. Dong, M. Lin, E. Nazaretski, I. K. Robinson, Y. S. Chu, and X. Huang, *Multi-slice ptychography with large numerical aperture multilayer laue lenses*, *Optica* **5**, 601 (2018).
- [10] K. Shimomura, A. Suzuki, M. Hirose, and Y. Takahashi, *Precession x-ray ptychography with multislice approach*, *Physical Review B* **91** (2015), 10.1103/physrevb.91.214114.
- [11] K. Shimomura, M. Hirose, and Y. Takahashi, *Multislice imaging of integrated circuits by precession x-ray ptychography*, *Acta Crystallographica Section A Foundations and Advances* **74**, 66 (2018).

7

Conclusion

In this thesis, we have investigated different aspects of phase retrieval. We gave an overview of several non-iterative and iterative phase retrieval methods, explained their interconnections, and their advantages and drawbacks. We examined several methods in more detail, laid out some new insights and proposed some new methods, namely

- a non-iterative phase retrieval method by variation of a single optical parameter (Section 2.5);
- a non-iterative phase retrieval method using focus variation and a star-shaped mask (Section 2.6);
- a method that combines (M)-HIO with ptychography, either through sequential or global updates (Section 3.2.3);
- a method to make ptychographic phase retrieval more noise robust, that works by adapting the measurement constraints (Section 4.4.2);
- an alternate projections algorithm for DSI that is suitable for final refinement (Section 5.1.2);
- an algorithm for DSI that uses synthetic constraints to eliminate alternative solutions (Section 5.2.2);
- a method to more accurately compute the field scattered by a thick sample (Section 6.2.2).

These may help us in better understanding and improving phase retrieval methods, which are used in various imaging applications. In particular, it may be useful for lensless imaging using EUV radiation, which can be applied in the inspection of semiconductor samples. Thus, the researched topic may for example help in producing better computer chips more efficiently, in line with Moore's famous law of IC manufacturing.

7.1. Future work

The major limitation of the methods discussed so far, is that they are diffraction-limited, i.e. the imaging resolution one can achieve is limited by the wavelength of the illumination and the acceptance angle of the detector in the far field. This is because if one uses the scalar model of light fields, and models the transmission through a sample using the multiplicative approximation, the high-resolution information of the sample is encoded in the evanescent field that does not propagate to the far field. There are two ways to achieve super resolution from far field measurements.

- **Super resolution through sparsity constraints:** One can compensate for the loss of high-resolution information by introducing prior information about the sample, which can be formulated as a sparsity constraint of some sort. For example, one might know that the object can be parametrized in a certain manner using few coefficients [1], or otherwise significantly limit the solution space [2]. When inspecting semiconductor samples, one typically knows what the sample should ideally look like according to its design, and only needs to check for a certain set of deviations from that design. Therefore, lots of prior information is available that can be incorporated in the form of a sparsity constraint.
- **Super resolution through multiple scattering:** By making the single-scattering assumption (which is traditionally made in ptychography when making the multiplicative approximation that assumes the exit wave can be computed by multiplying the probe with the object's transmission or reflection function), one assumes that all the high-resolution information of the samples is encoded in the evanescent field, which does not propagate to the far field. However, if one takes into account that the evanescent scattered field may interact once more with the sample and be converted to a propagating field, one realizes that there may in fact be super resolution information available in the far field. Therefore, by taking multiple scattering into account, one might achieve super resolution from far-field measurements. [3, 4].

References

- [1] A. Szameit, Y. Shechtman, E. Osherovich, E. Bulkich, P. Sidorenko, H. Dana, S. Steiner, E. B. Kley, S. Gazit, T. Cohen-Hyams, S. Shoham, M. Zibulevsky, I. Yavneh, Y. C. Eldar, O. Cohen, and M. Segev, *Sparsity-based single-shot subwavelength coherent diffractive imaging*, [Nature Materials](#) **11**, 455 (2012).
- [2] T. Zhang, C. Godavarthi, P. C. Chaumet, G. Maire, H. Giovannini, A. Talneau, M. Allain, K. Belkebir, and A. Sentenac, *Far-field diffraction microscopy at $\lambda/10$ resolution*, [Optica](#) **3**, 609 (2016).
- [3] F. Simonetti, *Multiple scattering: The key to unravel the subwavelength world from the far-field pattern of a scattered wave*, [Physical Review E](#) **73** (2006), [10.1103/physreve.73.036619](#).

- [4] F. Simonetti, M. Fleming, and E. A. Marengo, *Illustration of the role of multiple scattering in subwavelength imaging from far-field measurements*, [Journal of the Optical Society of America A](#) **25**, 292 (2008).

Epilogue

Western civilization, it seems to me, stands by two great heritages. One is the scientific spirit of adventure — the adventure into the unknown, an unknown which must be recognized as being unknown in order to be explored; the demand that the unanswerable mysteries of the universe remain unanswered; the attitude that all is uncertain; to summarize it — the humility of the intellect. The other great heritage is Christian ethics — the basis of action on love, the brotherhood of all men, the value of the individual — the humility of the spirit.

Richard Feynman

It is said that where there is a will, there is a way. This phrase serves to remind us that if one is to put a significant amount of effort in some endeavor, one ought to be able to answer, at least to themselves, the question: ‘why?’ So having spent all this effort on doing a PhD, let us now ask: why?

Reasons for doing science

More broadly we can ask: why should one do science? From my experience, the reasons people tend to have can be divided into four categories:

1. *Idealism*: Being a good scientist is about following certain ideals: examine all claims critically, draw conclusions only from the facts, not from what you want to be true. We do science because there is something inherently good about following scientific ideals.
2. *Productivity*: Science has greatly improved productivity and human well-being on a global scale. We do science because it is an excellent tool to improve the world.
3. *Earning a living*: Our modern economy relies heavily on technology, so if you want to have a good job, you should be scientifically literate. We do science because we need to make money.
4. *Status*: Personal achievements in scientific research demonstrate a superior work ethic and intellect, which elevates one as an individual. We do science to earn the respect of ourselves and of others.

The scientific enterprise started from the ideals as posed in Reason 1. Over time, by following the scientific method we discovered that the universe can be explained in

purely materialistic terms. The universe is indifferent, there is no higher purpose, and there are no 'true' ideals. There are only ideals we choose to follow if we happen to like them, and ideals that are enforced if sufficiently many people with power happen to like them. Therefore, the result of adhering to scientific ideals is that the belief that scientific ideals are a fundamental truth has become obsolete.

Rather, it would be more appropriate to speak of a scientific toolkit. As inhabitants of a meaningless, indifferent universe, we are free to choose what our goals are (especially in Western liberal, individualistic societies), and it turns out that science is an effective tool to achieve many of them. So if you choose your goal to be to change the world, make money, or attain status, then Reasons 2, 3, 4 would be rational reasons to do science.

So it seems that if I should have any reason to do science, it should not be Reason 1. It is hopelessly self-contradictory, as is (possibly unintentionally) made explicit by the words of Carl Sagan: *'The sacred truth of science is that there are no sacred truths'*. One might even consider these ideals to be a self-righteous excuse for self-indulgence, as if one would say: 'I want to do science because I like it. I do not want to take any responsibility for practical matters, but I still want to feel morally superior'. Another might regard holding on to these ideals as a sign of weakness: it is naive wishful thinking by those who cannot handle the idea of an indifferent universe.

Without expecting to perfectly resolve all these issues, let us further explore the question: what are the reasons to stick to ideals that transcend the individual?

The role of transcendent ideals

7

The success of scientific ideals allowed for criticizing scientific ideals. An analogy can be made with Christianity, which is all about following the highest ideal, namely God. The Judeo-Christian foundations on which Western civilization allowed Enlightenment values to develop (though this claim could spark an entire discussion by itself), which were subsequently used to heavily criticize religion. In *Democracy in America* (1835), Alexis De Tocqueville states:

'By the side of these religious men I discern others whose looks are turned to the earth more than to Heaven; they are the partisans of liberty, not only as the source of the noblest virtues, but more especially as the root of all solid advantages; and they sincerely desire to extend its sway, and to impart its blessings to mankind. It is natural that they should hasten to invoke the assistance of religion, for they must know that liberty cannot be established without morality, nor morality without faith; but they have seen religion in the ranks of their adversaries, and they inquire no further; some of them attack it openly, and the remainder are afraid to defend it.'

Later, in 1983 Aleksandr Solzhenitsyn commented on the atrocities that had transpired in the communist Soviet Union:

'If I were asked today to formulate as concisely as possible the main cause of

the ruinous Revolution that swallowed up some sixty million of our people, I could not put it more accurately than to repeat: Men have forgotten God; that's why all this has happened.'

The point is that while traditional ideals may appear obsolete or even backward at first sight, and a more rational alternative seems available, one abandons the former at one's own peril, and one ought not to underestimate the magnitude of the potential consequences. So is there a danger lurking when we make the scientific enterprise increasingly more pragmatic and abandon traditional ideals? Is there a danger to maintaining that there is nothing intrinsically meaningful to doing science, and that it only serves as a tool to achieve individual goals?

The risks of doing science without reverence for higher ideals

Suppose we maintain that there is nothing fundamentally valuable about doing science, but that it is simply an effective means to an end. The steps to reaching that end seem straightforward:

1. Define the goal with some quantifiable metric.
2. Incentivize researchers to reach that goal.

Therefore, we define the quality of research output by metrics such as the number of publications, the number of citations, and the impact factors of the journals one publishes in. We incentivize researchers through rankings, awards, and honours. It seems simple enough, but experience with optimization methods teaches us that the most straightforward optimization algorithm can be far from the best one. Especially the issue of motivation and incentives is tricky.

Social media have aimed to harness the power of motivation. Through 'likes' and similar rewards they try to keep the consumer hooked. That 'motivation' has now been more properly identified as 'addiction' and a cause of stress and anxiety. But when researchers try to rack up their number of publications and citations, how different is that from racking up the number of 'likes' on your social media accounts? Where lies the boundary between motivating someone and encouraging an unhealthy obsession?

Moreover, relying on one's research output to find motivation is inherently unreliable. The point of doing research is exploring the unknown, which by definition means it is uncertain whether the outcome will be positive or negative, regardless of the competence of the researcher. If the competence of a researcher is determined by their ability to generate positive research results, researchers gain an incentive to neglect scientific integrity, as scientific integrity only makes it more difficult to obtain positive results. Of course scientific integrity can be enforced through rules and regulations, but it should concern one when integrity is seen as a burden rather than a goal. The guarantee for positive results and the demand for scientific integrity are incompatible.

If, however, one maintains that adhering to scientific ideals has intrinsic value, one does not have to rely on addictive rewards generated by positive research out-

put to find motivation, and scientific integrity becomes less of a burden and more of a reward by itself. Paradoxically, moving the focus away from producing output results in a resilience to failure, while the serving of a greater ideal will help generate better output in the long term. Victor Frankl worded this excellently:

'Don't aim at success. The more you aim at it and make it a target, the more you are going to miss it. For success, like happiness, cannot be pursued; it must ensue, and it only does so as the unintended side effect of one's personal dedication to a cause greater than oneself or as the by-product of one's surrender to a person other than oneself. [...] In the long-run, [...] success will follow you precisely because you had forgotten to think about it.'

This is ancient wisdom, as it was already articulated in the New Testament in Matthew 6:31-33

'So do not worry, saying, 'What shall we eat?' or 'What shall we drink?' or 'What shall we wear?' [...] Seek first his kingdom and his righteousness, and all these things will be given to you as well.'

Indeed, one finds hints of this dedication to a greater cause in some of the greatest scientists. Johannes Kepler said: *'Truth is the daughter of time, and I feel no shame in being her midwife.'* Isaac Newton said: *'Plato is my friend — Aristotle is my friend — but my greatest friend is truth.'* However, in an individualistic, secular society, where quality metrics, awards, and honors are routinely used to incentivize researchers, it appears that researchers are not encouraged to serve scientific ideals, but rather the remnants of scientific ideals are made to serve the glorification of individual researchers.

7

Why did I do a PhD?

So with all that in mind, let us return to the original question: why did I spend so much effort on getting a PhD?

I try not to be too motivated by 'likes': the publications that I can put on my CV, or the title that I can put in front of my name (though I am not too naive to recognize the practical utility of these). I do not wish to take pride in 'being smart' or 'working hard'. These qualities are not commendable in and of themselves, for they are worthless if they do not serve the right purpose. Personally, I feel proudest of conducting research and educational activities in a way that is mostly in line with my ideals.

Even so, one must not overestimate the value of following scientific ideals. There is something beautiful about doing science, but if it in some way causes human misery (e.g. by insisting on moral relativism and existential nihilism ad nauseam), one should have one's priorities in order. Carl Friedrich Gauss put it as follows:

'There are problems to whose solution I would attach an infinitely greater impor-

tance than to those of mathematics, for example touching ethics, or our relation to God, or concerning our destiny and our future; but their solution lies wholly beyond us and completely outside the province of science.'

Feynman articulated it more succinctly:

'Tell your son to stop trying to fill your head with science - for to fill your heart with love is enough!'

I am grateful to those who inspired me to examine my own beliefs, values, and motivations more carefully, whether they did it by helping me, challenging me, or otherwise. Perhaps more importantly than contributing to scientific research, I think doing a PhD has helped me to develop more as an individual. I hope that others who do a PhD find the opportunity to experience the same.

We do not ask for what useful purpose the birds do sing, for song is their pleasure, since they were created for singing. Similarly, we ought not to ask why the human mind troubles to fathom the secrets of the heavens.

Johannes Kepler as quoted by Carl Sagan



Wigner Distribution Deconvolution Method and the autocorrelation function

Consider an object $O(\mathbf{x})$ that is illuminated by a shifted probe function $P(\mathbf{x} - \mathbf{X})$. We denote the transmitted exit wave as

$$\psi(\mathbf{x}, \mathbf{X}) = O(\mathbf{x})P(\mathbf{x} - \mathbf{X}). \quad (\text{A.1})$$

The far field is given by $\hat{\psi}(\mathbf{k}, \mathbf{X})$, where $\hat{\psi}$ denotes the Fourier transform of $\psi(\mathbf{x}, \mathbf{X})$ with respect to $\mathbf{x}$, and $\mathbf{k}$ is the corresponding reciprocal space coordinate. By measuring the far field intensity for different probe positions $\mathbf{X}$, we obtain the four-dimensional data set

$$I(\mathbf{k}, \mathbf{X}) = |\hat{\psi}(\mathbf{k}, \mathbf{X})|^2. \quad (\text{A.2})$$

We follow the line of reasoning as outlined in [1] to write $\hat{\psi}(\mathbf{k}, \mathbf{X})$ as a four-dimensional Fourier transform of a four-dimensional function $f(\mathbf{x}, \mathbf{K})$, so that the four-dimensional inverse Fourier transform of $I(\mathbf{k}, \mathbf{X})$ can be interpreted as the autocorrelation of $f(\mathbf{x}, \mathbf{K})$.

We can write $\hat{\psi}(\mathbf{k}, \mathbf{X})$ explicitly as

$$\hat{\psi}(\mathbf{k}, \mathbf{X}) = \int O(\mathbf{x})P(\mathbf{x} - \mathbf{X})e^{-2\pi i \mathbf{k} \cdot \mathbf{x}} d\mathbf{x}. \quad (\text{A.3})$$

We can rewrite $P(\mathbf{x} - \mathbf{X})$ as

$$P(\mathbf{x} - \mathbf{X}) = \int \hat{P}(\mathbf{K})e^{2\pi i(\mathbf{x} - \mathbf{X}) \cdot \mathbf{K}} d\mathbf{K}, \quad (\text{A.4})$$

where $\hat{P}(\mathbf{K})$ denotes the Fourier transform of $P(\mathbf{x})$. Plugging this expression in Eq. (A.3) gives

$$\begin{aligned}\hat{\psi}(\mathbf{k}, \mathbf{X}) &= \iint O(\mathbf{x}) \hat{P}(\mathbf{K}) e^{-2\pi i \mathbf{k} \cdot \mathbf{x}} e^{2\pi i (\mathbf{x} - \mathbf{X}) \cdot \mathbf{K}} d\mathbf{x} d\mathbf{K} \\ &= \iint O(\mathbf{x}) \hat{P}(\mathbf{K}) e^{2\pi i \mathbf{x} \cdot \mathbf{K}} e^{-2\pi i \mathbf{k} \cdot \mathbf{x}} e^{-2\pi i \mathbf{X} \cdot \mathbf{K}} d\mathbf{x} d\mathbf{K} \\ &= \iint f(\mathbf{x}, \mathbf{K}) e^{-2\pi i \mathbf{k} \cdot \mathbf{x}} e^{-2\pi i \mathbf{X} \cdot \mathbf{K}} d\mathbf{x} d\mathbf{K},\end{aligned}\quad (\text{A.5})$$

where we defined

$$f(\mathbf{x}, \mathbf{K}) = O(\mathbf{x}) \hat{P}(\mathbf{K}) e^{2\pi i \mathbf{x} \cdot \mathbf{K}}. \quad (\text{A.6})$$

So we see that $\hat{\psi}(\mathbf{k}, \mathbf{X})$ can be written as the four-dimensional Fourier transform of the four-dimensional function $f(\mathbf{x}, \mathbf{K})$. Therefore, inverse Fourier transforming $I(\mathbf{k}, \mathbf{X})$ with respect to both $\mathbf{k}$ and $\mathbf{X}$ yields the four-dimensional autocorrelation function of $f(\mathbf{x}, \mathbf{K})$

$$\begin{aligned}\mathcal{F}_{\mathbf{k}, \mathbf{X}}^{-1} \{I(\mathbf{k}, \mathbf{X})\}(\mathbf{x}, \mathbf{K}) &= \iint f(\mathbf{x}', \mathbf{K}')^* f(\mathbf{x} + \mathbf{x}', \mathbf{K} + \mathbf{K}') d\mathbf{x}' d\mathbf{K}' \\ &= e^{2\pi i \mathbf{x} \cdot \mathbf{K}} \int O(\mathbf{x}')^* O(\mathbf{x} + \mathbf{x}') e^{2\pi i \mathbf{K} \cdot \mathbf{x}'} d\mathbf{x}' \int \hat{P}(\mathbf{K}')^* \hat{P}(\mathbf{K} + \mathbf{K}') e^{2\pi i \mathbf{K}' \cdot \mathbf{x}} d\mathbf{K}' \\ &= e^{2\pi i \mathbf{x} \cdot \mathbf{K}} W_O(\mathbf{x}, \mathbf{K}) W_{\hat{P}}(\mathbf{K}, \mathbf{x}),\end{aligned}\quad (\text{A.7})$$

where $W_g(\mathbf{x}, \mathbf{K})$ denotes the Wigner distribution function

$$W_g(\mathbf{x}, \mathbf{K}) = \int g(\mathbf{x}')^* g(\mathbf{x} + \mathbf{x}') e^{2\pi i \mathbf{x}' \cdot \mathbf{K}} d\mathbf{x}'. \quad (\text{A.8})$$

We see that the four-dimensional autocorrelation function (which is the 4D inverse Fourier transform of the 4D measured data set) separates the contributions from $O(\mathbf{x})$ and $\hat{P}(\mathbf{K})$, so if the illumination $P(\mathbf{x})$ is known, one can divide out $W_{\hat{P}}(\mathbf{K}, \mathbf{x})$ to find the Wigner distribution $W_O(\mathbf{x}, \mathbf{K})$ of the object $O(\mathbf{x})$, which can be inverted to find $O(\mathbf{x})$.

The final expression of Eq. (A.7) differs slightly from Eq. (1.30), because there we defined the forward Fourier transform using the positive exponential $\hat{f}(k) = \int f(x) e^{+2\pi i x k} dx$ (while here we use the negative exponential), and because in Eq. (1.30) we take one forward and one inverse Fourier transform (while here we apply two inverse Fourier transforms), and because in Eq. (1.32) the result is written in terms of the Wigner distribution function of $P(\mathbf{x})$ (while here we use the Wigner distribution function of $\hat{P}(\mathbf{K})$). Nevertheless, the fundamental result that the measurements $I(\mathbf{k}, \mathbf{X})$ allow for a factorization of the Wigner distributions of the object and the probe remains apparent.

One reason why it is useful to think about the WDDM in terms of an autocorrelation function of $O(\mathbf{x}) \hat{P}(\mathbf{K}) e^{2\pi i \mathbf{x} \cdot \mathbf{K}}$ is that it allows one to apply general arguments

that hold for any direct reconstruction method based on autocorrelation functions, such as the one presented in Section 2.3. Another reason is that one can intuitively understand how the choice of the illumination function $P(\mathbf{x})$ and sampling of the probe positions $\mathbf{X}$ affect the reconstruction quality of $O(\mathbf{x})$. To see how this works, let us consider a 1D object and probe so that $f(x, K) = O(x)\hat{P}(K)e^{2\pi i x K}$ becomes a 2D function which can easily be visualized in an x, K -plot, just like its autocorrelation function $\text{Aut}\{f\}(x, K)$ as shown in Fig. A.1.

If the probe has a high spatial frequency content, $\hat{P}(K)$, $W_{\hat{P}}(K, x)$, $f(x, K)$, and $\text{Aut}\{f\}(x, K)$ will be elongated in the K -direction. This is beneficial for the reconstruction [2], because the Wigner distribution function of $O(x)$ is given by

$$W_O(x, K) = \frac{\text{Aut}\{f\}(x, K)}{W_{\hat{P}}(K, x)} e^{-2\pi i x K}, \quad (\text{A.9})$$

and $O(x)$ is found by inverse Fourier transforming $W_O(x, K)$ along K (or the Fourier transform of $O(x)$ is found by inverse Fourier transforming $W_O(x, K)$ along x and evaluating it in K). Thus, $W_O(x, K)$ needs to be well defined along K , which means we need to avoid $\frac{\text{Aut}\{f\}(x, K)}{W_{\hat{P}}(K, x)} = \frac{0}{0}$, which is done by giving the probe a sufficiently high spatial frequency content.

To understand the effect of sampling the probe position X along a discrete grid with interval ΔX , one must note that this corresponds to multiplying $I(k, X)$ with a delta comb $\text{III}_{\Delta X}(X)$, which means that $\text{Aut}\{f\}(x, K)$ is convolved with a delta comb $\text{III}_{1/\Delta X}(K)$. Thus, aliases of $\text{Aut}\{f\}(x, K)$ will appear along the K -direction with interval $\frac{1}{\Delta X}$. If $P(\mathbf{x})$ has a high spatial frequency content, ΔX must be smaller to prevent overlap between the aliases. This can be interpreted intuitively: if $P(\mathbf{x})$ is a focal spot, it has a high spatial frequency content, and it must be scanned with a very small step size ΔX across the sample to ensure overlap between the adjacent probes.

One might ask what these observations about WDDM (non-iterative ptychography) can tell us about the limitations of iterative ptychography. The constraint that one uses in iterative ptychography that is not used in WDDM, is that $\frac{\text{Aut}\{f\}(\mathbf{x}, \mathbf{K})}{W_{\hat{P}}(\mathbf{K}, \mathbf{x})} e^{-2\pi i \mathbf{x} \cdot \mathbf{K}}$ must be a Wigner distribution: in WDDM it is merely assumed that this is the case, but it is not enforced as a separate constraint. Iterative ptychographic algorithms have demonstrated that by enforcing such a constraint, one can reconstruct $O(\mathbf{x})$ even for large (but not too large) ΔX [3] (which means one can correct for aliasing of $\text{Aut}\{f\}(\mathbf{x}, \mathbf{K})$ in the $\mathbf{K}$ -direction), one can reconstruct $O(\mathbf{x})$ and $P(\mathbf{x})$ simultaneously [4, 5], and one can correct for errors in the probe positions $\mathbf{X}$ [6]. However, iterative schemes cannot compensate for the loss of information due to $\frac{\text{Aut}\{f\}(\mathbf{x}, \mathbf{K})}{W_{\hat{P}}(\mathbf{K}, \mathbf{x})} = \frac{0}{0}$. For example, in the extreme case that $P(\mathbf{x})$ is a plane wave, then its spatial frequency content is minimal, and one is unable to find a reconstruction using WDDM. If one uses an iterative ptychographic algorithm, it is still not possible

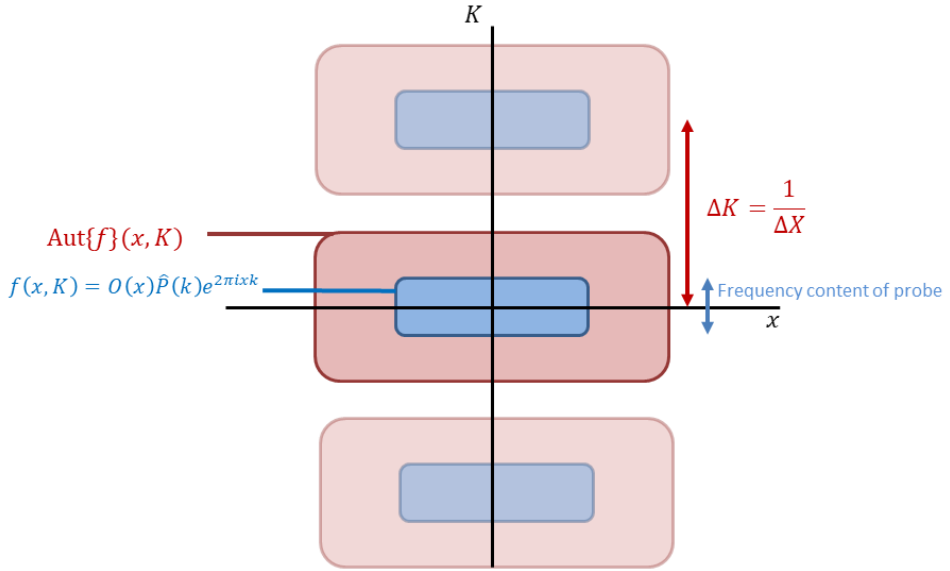


Figure A.1: Illustration of $f(x, K) = O(x)\hat{P}(K)e^{2\pi i x K}$ and its autocorrelation function which is used in the WDDM.

to reconstruct $O(x)$ because one measures the same intensity pattern for all probe positions X .

References

- [1] R. Bates and J. Rodenburg, *Sub-ångström transmission microscopy: A fourier transform algorithm for microdiffraction plane intensity information*, *Ultramicroscopy* **31**, 303 (1989).
- [2] M. Guizar-Sicairos, M. Holler, A. Diaz, J. Vila-Comamala, O. Bunk, and A. Menzel, *Role of the illumination spatial-frequency spectrum for ptychography*, *Physical Review B* **86** (2012), 10.1103/physrevb.86.100103.
- [3] O. Bunk, M. Dierolf, S. Kynde, I. Johnson, O. Marti, and F. Pfeiffer, *Influence of the overlap parameter on the convergence of the ptychographical iterative engine*, *Ultramicroscopy* **108**, 481 (2008).
- [4] P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, *Probe retrieval in ptychographic coherent diffractive imaging*, *Ultramicroscopy* **109**, 338 (2009).
- [5] A. Maiden and J. Rodenburg, *An improved ptychographical phase retrieval algorithm for diffractive imaging*, *Ultramicroscopy* **109**, 1256 (2009).
- [6] A. Maiden, M. Humphry, M. Sarahan, B. Kraus, and J. Rodenburg, *An annealing*

algorithm to correct positioning errors in ptychography, [Ultramicroscopy](#) **120**, 64 (2012).

B

Fresnel propagation of the focal field

The focal field of a coherent pupil field $\psi(x, y)$ is given by

$$\hat{\psi}(x_f, y_f) = \mathcal{F}\{\psi(x, y)\}\left(\frac{x_f}{\lambda f}, \frac{y_f}{\lambda f}\right), \quad (\text{B.1})$$

where (x_f, y_f) denote the coordinates in the back focal plane, λ is the wavelength, f is the focal length of the lens, and (x, y) are the coordinates in the front focal plane. We propagate this field by z_f using the Angular Spectrum method. The spectrum of $\hat{\psi}(x_f, y_f)$ is

$$\begin{aligned} \mathcal{F}\{\hat{\psi}(x_f, y_f)\}(f_x, f_y) &= \iint \psi(x, y) e^{-2\pi i \left(x \frac{x_f}{\lambda f} + y \frac{y_f}{\lambda f}\right)} e^{-2\pi i (x_f f_x + y_f f_y)} dx dy dx_f dy_f \\ &= (\lambda f)^2 \psi(-\lambda f f_x, -\lambda f f_y). \end{aligned} \quad (\text{B.2})$$

In the Fresnel approximation the propagator is given by

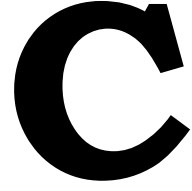
$$\begin{aligned} e^{iz_f \sqrt{k^2 - k_x^2 - k_y^2}} &\approx e^{iz_f k} e^{-iz_f \frac{k_x^2 + k_y^2}{2k}}, \\ &= e^{iz_f k} e^{-iz_f \frac{(2\pi)^2}{2k} (f_x^2 + f_y^2)} \\ &= e^{iz_f k} e^{-iz_f \pi \lambda (f_x^2 + f_y^2)}. \end{aligned} \quad (\text{B.3})$$

Thus, for the propagated field we get

$$\begin{aligned}
 \hat{\psi}(x_f, y_f, z_f) &\approx e^{iz_fk} (\lambda f)^2 \iint \psi(-\lambda f f_x, -\lambda f f_y) e^{-iz_f \pi \lambda (f_x^2 + f_y^2)} e^{2\pi i (f_x x_f + f_y y_f)} df_x df_y \\
 &= e^{iz_fk} \iint \psi(x', y') e^{-i \frac{z_f \pi}{\lambda f^2} (x'^2 + y'^2)} e^{-2\pi i \left(x' \frac{x_f}{\lambda f} + y' \frac{y_f}{\lambda f} \right)} dx' dy' \\
 &= \mathcal{F} \left\{ \psi(x, y) e^{-i \frac{z_f \pi}{\lambda f^2} (x^2 + y^2)} \right\} \left(\frac{x_f}{\lambda f}, \frac{y_f}{\lambda f} \right).
 \end{aligned} \tag{B.4}$$

If we define R to be the size of the object, so that $(X, Y) = \left(\frac{x}{R}, \frac{y}{R} \right)$ are normalized coordinates, we can write

$$\hat{\psi}(x_f, y_f, z_f) \approx R^2 \mathcal{F} \left\{ \psi(X, Y) e^{-i \frac{z_f \pi}{\lambda} \left(\frac{R}{f} \right)^2 (X^2 + Y^2)} \right\} \left(\frac{x_f}{\lambda f}, \frac{y_f}{\lambda f} \right). \tag{B.5}$$



Reconstruction error and cost functional

The question that we consider is the following: if the measured far-field intensity patterns are corrupted by noise, what cost functional L should we choose such that when minimized, the object reconstruction error E is minimized? Let $f(\mathbf{x})$ denote the estimated object, and $\psi(\mathbf{x})$ the actual object. We define the reconstruction error as

$$E[f(\mathbf{x})] = \sum_{\mathbf{x}} |f(\mathbf{x}) - \psi(\mathbf{x})|^2. \quad (\text{C.1})$$

For simplicity we omit the normalization and compensation for global phase shifts that we used in Eq. (3.33). Using Parseval's theorem, we can write E in terms of the diffracted fields $\hat{f}(\mathbf{k})$ and $\hat{\psi}(\mathbf{k})$

$$\begin{aligned} E[f(\mathbf{x})] &= \sum_{\mathbf{k}} |\hat{f}(\mathbf{k}) - \hat{\psi}(\mathbf{k})|^2 \\ &= \sum_{\mathbf{k}} |\hat{f}(\mathbf{k})|^2 + |\hat{\psi}(\mathbf{k})|^2 - 2|\hat{f}(\mathbf{k})||\hat{\psi}(\mathbf{k})| \cos \theta(\mathbf{k}), \end{aligned} \quad (\text{C.2})$$

where $\theta(\mathbf{k})$ denotes the phase difference between the estimated diffracted field $\hat{f}(\mathbf{k})$ and the actual diffracted field $\hat{\psi}(\mathbf{k})$. As we saw in Section 4.3, the specific choice of the cost functional $L[f(\mathbf{x})]$ is only used for the final refinement of the estimated object $f(\mathbf{x})$. Therefore, we only need to consider $E[f(\mathbf{x})]$ for $f(\mathbf{x})$ close to $\psi(\mathbf{x})$. If $f(\mathbf{x}) \approx \psi(\mathbf{x})$, then we can expect $\theta(\mathbf{k})$ to be very small, since the phase of the far field $\hat{f}(\mathbf{k})$ affects $f(\mathbf{x})$ far more than its amplitude (θ can be large when $|\hat{\psi}|$ is small, but then the term containing θ is negligible anyway). For small $\theta(\mathbf{k})$, we can apply the first-order Taylor expansion to $\cos \theta(\mathbf{k})$, which gives $\cos \theta(\mathbf{k}) \approx 1$,

which allows us to approximate $E[f(\mathbf{x})]$ by

$$\begin{aligned} E[f(\mathbf{x})] &\approx \sum_{\mathbf{k}} |\hat{f}(\mathbf{k})|^2 + |\hat{\psi}(\mathbf{k})|^2 - 2|\hat{f}(\mathbf{k})||\hat{\psi}(\mathbf{k})| \\ &= \sum_{\mathbf{k}} (|\hat{f}(\mathbf{k})| - |\hat{\psi}(\mathbf{k})|)^2, \end{aligned} \quad (\text{C.3})$$

which is reminiscent of the amplitude-based cost functional in Eq. (4.10). However, we do not measure $|\hat{\psi}(\mathbf{x})|$, but a noise-corrupted version of it $|\hat{\psi}(\mathbf{k})| + n(\mathbf{k})$, where $n(\mathbf{k}) \geq -|\hat{\psi}(\mathbf{k})|$. Therefore, we can define the cost functional $L[f(\mathbf{x})]$ as

$$\begin{aligned} L[f(\mathbf{x})] &= \sum_{\mathbf{k}} (|\hat{f}(\mathbf{k})| - (|\hat{\psi}(\mathbf{k})| + n(\mathbf{k})))^2 \\ &= \sum_{\mathbf{k}} |\hat{f}(\mathbf{k})|^2 + |\hat{\psi}(\mathbf{k})|^2 - 2|\hat{f}(\mathbf{k})||\hat{\psi}(\mathbf{k})| + |n(\mathbf{k})|^2 + 2|\hat{\psi}(\mathbf{k})||n(\mathbf{k})| - 2|\hat{f}(\mathbf{k})|n(\mathbf{k}) \\ &\approx E[f(\mathbf{x})] + \sum_{\mathbf{k}} |n(\mathbf{k})|^2 + 2|\hat{\psi}(\mathbf{k})||n(\mathbf{k})| - 2|\hat{f}(\mathbf{k})|n(\mathbf{k}). \end{aligned} \quad (\text{C.4})$$

Since $|n(\mathbf{k})|^2 + 2|\hat{\psi}(\mathbf{k})||n(\mathbf{k})|$ is independent of $f(\mathbf{x})$, one can equivalently define the cost functional

$$L[f(\mathbf{x})] = E[f(\mathbf{x})] - \sum_{\mathbf{k}} 2|\hat{f}(\mathbf{k})|n(\mathbf{k}). \quad (\text{C.5})$$

We wanted to know how to define $L[f(\mathbf{x})]$ such that when minimized, $E[f(\mathbf{x})]$ is minimized. We see that for our current choice of $L[f(\mathbf{x})]$ (the amplitude-based cost functional), there is an extra term $-\sum_{\mathbf{k}} 2|\hat{f}(\mathbf{k})|n(\mathbf{k})$. The problem arises when $L[f(\mathbf{x})]$ can be made smaller by making $-\sum_{\mathbf{k}} 2|\hat{f}(\mathbf{k})|n(\mathbf{k})$ smaller at the expense of making $E[f(\mathbf{x})]$ bigger: in this case minimizing $L[f(\mathbf{x})]$ does not minimize $E[f(\mathbf{x})]$. In order to find an appropriate $L[f(\mathbf{x})]$, one must know something about the particular realization of $n(\mathbf{k})$, rather than its statistics (which is in a sense a trivial remark: if you know $n(\mathbf{k})$ you can simply denoise the measurement and reconstruct $\psi(\mathbf{x})$ perfectly). While the statistics of $n(\mathbf{k})$ may allow one to find a $f(\mathbf{x})$ for which measuring $|\hat{\psi}(\mathbf{k})| + n(\mathbf{k})$ is most likely (using a maximum likelihood cost functional), it is not obvious how such an $f(\mathbf{x})$ affects $E[f(\mathbf{x})]$, which is ultimately the quantity of interest.

This analysis also explains why it makes sense to use the amplitude-based cost functional when applying the noise-robust scheme of Section 4.4: not because such a cost function implicitly assumes Poissonian noise (it was demonstrated in Fig. 4.1 that the amplitude-based cost function does not necessarily yield the optimal result), but because Eq. (C.5) reveals that compared to other cost functionals, the amplitude-based cost functional is most directly related to the object reconstruction error, which is what matters in the end. The fact that the amplitude-based

cost functional is closely related to the Alternating Projections method, and the fact that the steepest descent method may diverge (if a poor initial guess is chosen) for other cost functionals even if no noise is present, also indicate that there is more to the amplitude-based cost functional than it just being the maximum-likelihood cost functional for Poissonian noise.

D

Measurement constraint projection for DSI

Suppose you have two complex-valued optimization variables x, y that should satisfy the constraint

$$xy^* = C. \quad (\text{D.1})$$

Given initial estimates x_0, y_0 , how do we project them onto the constraint C ? We should find x, y that satisfy the constraint, such that the distance between x, y and x_0, y_0 is minimized. So we should find δ, ϵ such that

$$(x_0 + \delta)(y_0 + \epsilon)^* = C \quad (\text{D.2})$$

and

$$L = |\delta|^2 + |\epsilon|^2 \quad (\text{D.3})$$

is minimized. Expanding Eq. (D.2) gives

$$x_0 y_0^* + \delta y_0^* + \epsilon^* x_0 + \delta \epsilon^* = C. \quad (\text{D.4})$$

Let us say the correction is small, so that we can neglect the term $\delta \epsilon^*$, and get

$$\delta y_0^* + \epsilon^* x_0 = D, \quad (\text{D.5})$$

where $D = C - x_0 y_0^*$. We find

$$\delta = \frac{D - \epsilon^* x_0}{y_0^*}. \quad (\text{D.6})$$

Substituting this into Eq. (D.3) gives

$$\begin{aligned} L &= |\epsilon|^2 + \frac{|D - \epsilon^* x_0|^2}{|y_0^*|^2} \\ &= |\epsilon|^2 \left(1 + \frac{|x_0|^2}{|y_0|^2} \right) - 2 \frac{\text{Re}\{D \epsilon x_0^*\}}{|y_0|^2} + \frac{|D|^2}{|y_0|^2}. \end{aligned} \quad (\text{D.7})$$

To minimize L , we should choose the phase of ϵ such that $\text{Re}\{D\epsilon x_0^*\}$ is as large as possible, which means

$$D\epsilon x_0^* = |D\epsilon x_0^*|, \quad (\text{D.8})$$

so

$$\epsilon = \frac{|D\epsilon x_0^*|}{Dx_0^*}. \quad (\text{D.9})$$

We should now minimize with respect to $|\epsilon|$

$$L = |\epsilon|^2 \left(1 + \frac{|x_0|^2}{|y_0|^2} \right) - 2 \frac{|D||x_0|}{|y_0|^2} |\epsilon| + \frac{|D|^2}{|y_0|^2}. \quad (\text{D.10})$$

Setting $\frac{dL}{d|\epsilon|} = 0$ gives

$$\begin{aligned} |\epsilon| &= \frac{\frac{|D||x_0|}{|y_0|^2}}{1 + \frac{|x_0|^2}{|y_0|^2}} \\ &= \frac{|D||x_0|}{|y_0|^2 + |x_0|^2}. \end{aligned} \quad (\text{D.11})$$

Plugging this into Eq. (D.9) gives

$$\epsilon = \frac{D^* x_0}{|y_0|^2 + |x_0|^2}. \quad (\text{D.12})$$

For δ we find

$$\delta = \frac{Dy_0}{|y_0|^2 + |x_0|^2}. \quad (\text{D.13})$$



Scattering in the paraxial approximation

If we apply the single-scattering approximation to Eq. (6.5), the scattered field is found by multiplying the incident field with the scattering potential χ . However, in Eq. (6.44), where we worked with the paraxial Helmholtz equation, the scattered field is computed by multiplying the incident field with $\tilde{\chi} = \frac{i}{2k}\chi$. In the following, we explain why there is an additional factor of $\frac{i}{2k}$ present.

In Eq. (6.5) we stated that for a scattering potential $\chi(\mathbf{r})$, the scattered field $U_s(\mathbf{r})$ is given by

$$U_s(\mathbf{r}) = \int G(\mathbf{r} - \mathbf{r}')\chi(\mathbf{r}') (U_s(\mathbf{r}') + P(\mathbf{r}')) d\mathbf{r}', \quad (\text{E.1})$$

where $P(\mathbf{r})$ is the incident field, and the Green's function $G(\mathbf{r})$ is given by

$$G(\mathbf{r}) = \frac{e^{ik|\mathbf{r}|}}{4\pi|\mathbf{r}|}. \quad (\text{E.2})$$

In the single-scattering approximation, the expression becomes

$$U_s(x, y, z) = \iiint G(x - x', y - y', z - z')\chi(x', y', z')P(x', y', z') dx' dy' dz'. \quad (\text{E.3})$$

Now we assume the paraxial approximation holds, which means that freely propagating fields satisfy the paraxial Helmholtz equation as given by Eq. (6.37)

$$\Delta\phi(x, y, z) + 2ik\frac{\partial\phi(x, y, z)}{\partial z} = 0, \quad (\text{E.4})$$

where ϕ is related to the actual field $P(x, y, z)$ as the modulation of the plane wave e^{ikz}

$$P(x, y, z) = \phi(x, y, z)e^{ikz}, \quad (\text{E.5})$$

and Δ denotes the Laplacian in the transverse coordinates

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \quad (\text{E.6})$$

It can be shown that the paraxial Helmholtz equation is satisfied by fields that are propagated using the Fresnel diffraction integral [1]

$$\phi(x, y, z) = \iint \phi(x', y', z') F(x - x', y - y', z - z') dx' dy', \quad (\text{E.7})$$

where

$$F(x, y, z) = \frac{1}{i\lambda z} e^{\frac{ik}{2z}(x^2 + y^2)}. \quad (\text{E.8})$$

Note that for $z^2 \gg x^2 + y^2$

$$G(x, y, z) \approx \frac{i}{2k} e^{ikz} F(x, y, z). \quad (\text{E.9})$$

Substituting this in Eq. (E.3) gives

$$\begin{aligned} U_s(x, y, z) &\approx \iiint \frac{i}{2k} e^{ik(z-z')} F(x - x', y - y', z - z') \chi(x', y', z') \phi(x', y', z') e^{ikz'} dx' dy' dz' \\ &= e^{ikz} \iiint F(x - x', y - y', z - z') \frac{i}{2k} \chi(x', y', z') \phi(x', y', z') dx' dy' dz' \end{aligned} \quad (\text{E.10})$$

The integration over x' and y' corresponds to propagating the field $\frac{i}{2k} \chi(x', y', z') \phi(x', y', z')$ from plane z' to plane z . Thus, to find the scattered field in a plane z , one must multiply the incident field ϕ in each plane z' with χ , propagate the product from plane z' to plane z , and sum over all z' .

Note that convolution with $F(x, y, z)$ is equivalent to applying the operator $e^{\frac{i}{2k} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) z}$, as in Eq. (6.42). This can be seen by observing that both expressions are solutions to the paraxial Helmholtz equation, or more explicitly by noting that

$$\mathcal{F} \left\{ \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \phi(x, y) \right\} (f_x, f_y) = -(f_x^2 + f_y^2) \mathcal{F} \{ \phi(x, y) \} (f_x, f_y), \quad (\text{E.11})$$

so

$$\mathcal{F} \left\{ e^{\frac{i}{2k} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) z} \phi(x, y) \right\} (f_x, f_y) = e^{-\frac{i}{2k} (f_x^2 + f_y^2) z} \mathcal{F} \{ \phi(x, y) \} (f_x, f_y), \quad (\text{E.12})$$

which can be inverse Fourier transformed to give

$$\begin{aligned} e^{\frac{i}{2k} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) z} \phi(x, y) &= \mathcal{F}^{-1} \left\{ e^{-\frac{i}{2k} (f_x^2 + f_y^2) z} \right\} (x, y) \otimes \phi(x, y) \\ &\propto F(x, y, z) \otimes \phi(x, y), \end{aligned} \quad (\text{E.13})$$

where $\otimes$ denotes a convolution.

References

- [1] R. Grella, *Fresnel propagation and diffraction and paraxial wave equation*, *Journal of Optics* **13**, 367 (1982).

Curriculum Vitæ

Alexander Prasetya KONIJNENBERG

01-12-1992 Born in Maastricht, The Netherlands.

Education

2004–2010 Secondary School
RSG 't Rijks, Bergen op Zoom

2010–2013 BSc in Applied Mathematics & Applied Physics
Delft University of Technology

2013–2015 MSc in Applied Physics
Delft University of Technology

2015–2019 PhD. Physics
Delft University of Technology
Thesis: Computational methods for phase retrieval
Promotors: Prof. dr. W.M.J.M. Coene, Prof. dr. H.P. Urbach

List of Publications

Journal publications

14. **A.P. Konijnenberg**, L. Wei, N. Kumar, L. Couto Correa Pinto Filho, L. Cisotto, S.F. Pereira, and H.P. Urbach, *Demonstration of an optimised focal field with long focal depth and high transmission obtained with the Extended Nijboer-Zernike theory*, [Optics Express Vol. 22, Issue 1, pp. 311-324 \(2014\)](#).
13. **A.P. Konijnenberg**, S.F. Pereira, *Pupil engineering to create sheets, lines, and multiple spots at the focal region*, [Journal of Optics, Volume 17, Number 12 \(2015\)](#)
12. **A.P. Konijnenberg**, W.M.J. Coene, H.P. Urbach, *Combining ptychographical algorithms with the Hybrid Input-Output (HIO) algorithm*, [Ultramicroscopy, Volume 171, Pages 43-54 \(2016\)](#)
11. **A.P. Konijnenberg**, W.M.J. Coene, H.P. Urbach, *Non-iterative phase retrieval by phase modulation through a single parameter*, [Ultramicroscopy, Volume 174, Pages 70-78 \(2017\)](#)
10. Xiaosi Xu, **A.P. Konijnenberg**, S. F. Pereira, and H. P. Urbach, *Phase retrieval of the full vectorial field applied to coherent Fourier scatterometry*, [Optics Express Vol. 25, Issue 24, pp. 29574-29586 \(2017\)](#)
9. **A.P. Konijnenberg**, *An introduction to the theory of ptychographic phase retrieval methods*, [Advanced Optical Technologies, Volume 6, Issue 6 \(2017\)](#)
8. **A.P. Konijnenberg***, Xingyuan Lu*, Leixin Liu, W.M.J. Coene, Chengliang Zhao, and H.P. Urbach, (*: equal contribution) *Non-iterative method for phase retrieval and coherence characterization by focus variation using a fixed star-shaped mask*, [Optics Express Vol. 26, Issue 7, pp. 9332-9343 \(2018\)](#)
7. Yifeng Shao, Xingyuan Lu, **Sander Konijnenberg**, Chengliang Zhao, Yangjian Cai, and H. Paul Urbach, *Spatial coherence measurement and partially coherent diffractive imaging using self-referencing holography*, [Optics Express Vol. 26, Issue 4, pp. 4479-4490 \(2018\)](#)
6. **A.P. Konijnenberg**, W.M.J. Coene, and H.P. Urbach, *Model-independent noise-robust extension of ptychography*, [Optics Express Vol. 26, Issue 5, pp. 5857-5874 \(2018\)](#)
5. P. Dwivedi, **A.P. Konijnenberg**, S.F. Pereira, H.P. Urbach, *Lateral position correction in ptychography using the gradient of intensity patterns*, [Ultramicroscopy, Volume 192, Pages 29-36 \(2018\)](#)
4. Xingyuan Lu, Yifeng Shao, Chengliang Zhao, **Sander Konijnenberg**, Xinlei Zhu, Ying Tang, Yangjian Cai, H. Paul Urbach, *Noniterative spatially partially coherent diffractive imaging using pinhole array mask*, [Adv. Photon. 1\(1\) 016005 \(2019\)](#)

3. P. Dwivedi, **A.P. Konijnenberg**, S.F. Pereira, J. Meisner, and H.P. Urbach, *Position correction in ptychography using hybrid input-output (HIO) and cross-correlation*, [Journal of Optics](#), 21(3), p.035604 (2019).
2. X. Lu, C. Zhao, Y. Shao, J. Zeng, **S. Konijnenberg**, X. Zhu, S. Popov, H. P. Urbach, and Y. Cai, *Phase detection of coherence singularities and determination of the topological charge of a partially coherent vortex beam*, [Appl. Phys. Lett.](#) 114, 201106 (2019)
1. A.C.C. de Beurs, X. Liu, G.S.M. Jansen, **A.P. Konijnenberg**, W.M.J. Coene, K.S.E. Eikema, S. Witte, *Extreme ultraviolet lensless imaging without object support through interferometric probe rotation*, expected publication in 2019

Conference proceedings

3. **A.P. Konijnenberg**, W.M.J. Coene, S.F. Pereira, H.P. Urbach, *Ptychographic phase retrieval by applying hybrid input-output (HIO) iterations sequentially*, [Proceedings Volume 10335, Digital Optical Technologies 2017; 103351I](#) (2017), Event: SPIE Digital Optical Technologies, 2017, Munich, Germany
2. Matthias Strauch, **Sander Konijnenberg**, Yifeng Shao, H. Paul Urbach, *Wavefront shaping with an electrowetting liquid lens using surface harmonics (Conference Presentation)*, [Proceedings Volume 10073, Adaptive Optics and Wavefront Control for Biological Systems III; 1007304](#) (2017), SPIE BiOS, 2017, San Francisco, California, United States
1. P. Dwivedi, **A. P. Konijnenberg**, S. F. Pereira, H. P. Urbach, *An alternative method to correct translation positions in ptychography*, [Proceedings Volume 10677, Unconventional Optical Imaging; 106772A](#) (2018), Event: SPIE Photonics Europe, 2018, Strasbourg, France

Patent applications

1. W.M.J.M. Coene, **A.P. Konijnenberg**, *Inspection apparatus and method, lithographic apparatus, method of manufacturing devices and computer program*, [WO2017153133A1](#)