

Automated focusing and astigmatism correction in electron microscopy

Citation for published version (APA):

Rudnaya, M. (2011). *Automated focusing and astigmatism correction in electron microscopy*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Mathematics and Computer Science]. Technische Universiteit Eindhoven. <https://doi.org/10.6100/IR716361>

DOI:

[10.6100/IR716361](https://doi.org/10.6100/IR716361)

Document status and date:

Published: 01/01/2011

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

AUTOMATED FOCUSING AND
ASTIGMATISM CORRECTION IN
ELECTRON MICROSCOPY

Maria Rudnaya

Copyright © 2011 by Maria Rudnaya, Eindhoven, The Netherlands

All rights are reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the author.

Cover photo by Masha Ru, model **Justine**, styling **Lara Verheijden**, special thanks to **VICE**
Cover design by **xmas95**

This work has been carried out as a part of the Condor project at FEI Company under the responsibilities of the Embedded Systems Institute (ESI). This project is partially supported by the Dutch Ministry of Economic Affairs under the BSIK program.

A catalogue record is available from the Eindhoven University of Technology Library

ISBN: 978-90-386-2610-9

Automated focusing and astigmatism correction in electron microscopy

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College
voor Promoties in het openbaar te verdedigen
op dinsdag 6 september 2011 om 16.00 uur

door

Maria Rudnaya

geboren te Moskou, Rusland

Dit proefschrift is goedgekeurd door de promotor:

prof.dr. R.M.M. Mattheij

Copromotor:

dr. J.M.L. Maubach

To my parents

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Problem formulation	5
1.3	Outline	8
2	Modelling	11
2.1	Notation	11
2.2	Image formation model	12
2.3	Object function	13
2.4	Point spread function	14
2.4.1	The Lévi stable density function	14
2.4.2	Wave aberration function	15
2.5	Defocus and stigmator control variables	18
2.6	The sharpness function	20
2.7	Discrete images	22
3	Derivative-based approach	25
3.1	Derivative-based sharpness function	25
3.2	General properties	26
3.3	Digital image object	28
3.4	Two-dimensional setting	33
3.4.1	Rotationally symmetric point spread function	33
3.4.2	Non-symmetric point spread function	35
3.5	Discretization	37
3.6	The fast autofocus algorithm	38
3.7	Numerical experiments with STEM images	40
4	Fourier transform-based approach	45
4.1	Discrete Fourier transform	46
4.2	Power spectrum orientation	49
4.3	Power spectrum benchmarks	51
4.4	Orientation identification	53
4.5	The focus series method	54
4.5.1	A Lévi stable density benchmark	55
4.6	A link to the derivative-based approach	57
4.7	Discretization	58
4.8	Numerical experiments	61
4.8.1	A Gaussian benchmark	61

4.8.2	SEM images	62
5	Statistics-based approaches	67
5.1	Autocorrelation-based sharpness function	67
5.1.1	Image autocorrelation	67
5.1.2	The sharpness function	68
5.1.3	Discretization	70
5.2	Variance- and intensity-based sharpness functions	70
5.3	Histogram-based sharpness function	72
6	Assessing sharpness functions	75
6.1	The sharpness functions	76
6.2	The images	77
6.3	Assessment	78
6.4	Ranking results	81
6.5	Numerical results	85
7	Simultaneous defocus and astigmatism correction	91
7.1	Simulations	92
7.1.1	Description	92
7.1.2	Numerical results	97
7.2	Derivative-free optimization	97
8	On-line STEM application	101
8.1	The experimental set-up	101
8.2	The focus series algorithm	103
8.3	Simultaneous defocus and astigmatism correction	103
8.3.1	Gold particle samples	106
8.3.2	Carbon cross grating sample	108
8.3.3	Nelder-Mead vs. Powell	112
8.4	Fast autofocus method	115
9	Future recommendations	119
A	Fast autofocus iterative algorithm	121
B	Power spectrum analytical approximation	123
C	The image variance, robustness to noise	127
	List of frequently used symbols	129
	Bibliography	129
	Index	135
	Summary	139

Contents iii

Acknowledgements **141**

Curriculum Vitae **143**

Chapter 1

Introduction

The *resolution* or *resolving power* of a typical *light microscope* is limited only by the *wavelength* of visible light. The electron microscope uses electrons instead of photons. The wavelength of electrons is much smaller than the wavelength of photons, which makes it possible to achieve much higher *magnifications* than in light microscopy.

Nowadays electron microscopy is a powerful tool for material science, biology, nanotechnology and medical academic research. The electron microscopy images are used for spectroscopic and chemical analysis [5], quantitative analysis of material properties [98] (e.g., average size and distribution of particles). They are highly valued among others in semiconductor industry [31, 91], archeology and paleontology [86], where they are used for production monitoring, control and troubleshooting. Figure 1.1 shows examples of electron microscopy images.

1.1 Motivation

The history of electron microscopy goes back to 1931, when German engineers Ernst Ruska and Max Knoll constructed the prototype electron microscope, capable of only $400\times$ magnification. The simplest *Transmission Electron Microscope (TEM)* is an exact analogue of the light microscope. In Figure 1.2

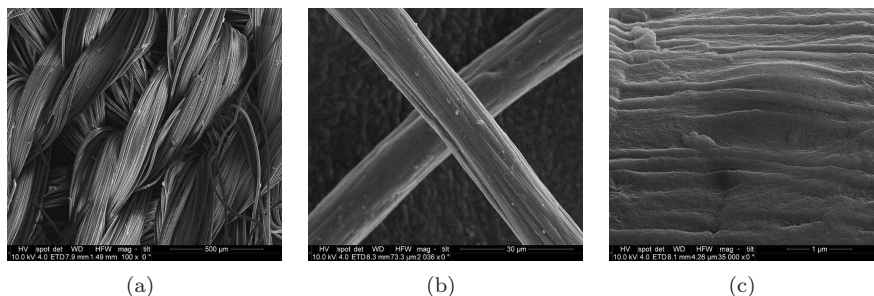


FIGURE 1.1: Electron microscopy images from an industrial application.

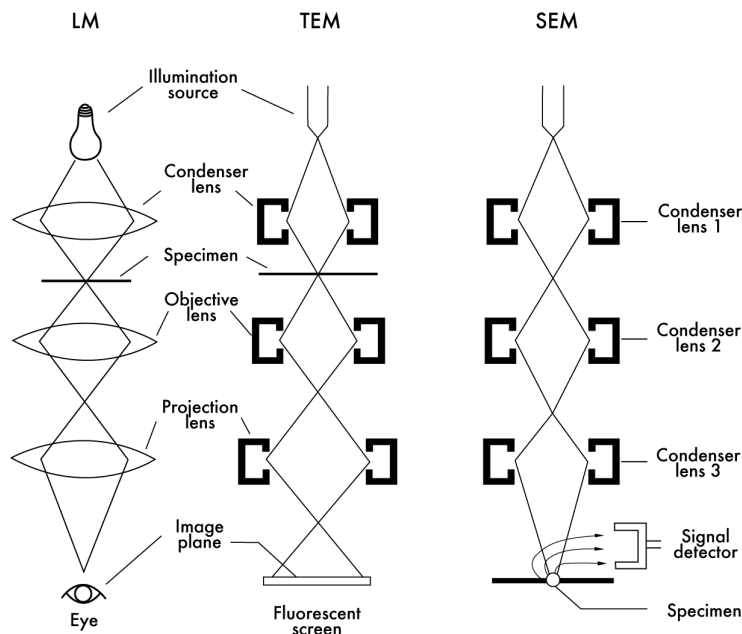


FIGURE 1.2: From left to right: a schematic diagram of the light microscope and the transmission electron microscope; a schematic diagram of the scanning electron microscope (taken from [75]).

schematic diagrams of the light microscope and the TEM are given. The illumination coming from the electron gun is concentrated on a *sample* with the condenser lens. The electrons transmitted through a sample are focused by an objective lens into a magnified intermediate image, which is enlarged by projector lenses and formed on the fluorescent screen or photographic film. The practical realization of TEM *Transmission Electron Microscope (TEM)* is more complex than the diagram in Figure 1.2 suggests: high vacuum, long electron path, highly stabilized electronic supplies for electron lenses are required [75].

The *Scanning Electron Microscope (SEM)* is most widely used of all electron beam instruments [2]. In SEM a fine probe of electrons with energies from a few hundred eV to tens of keV is focused at the surface of a sample and scanned across it (see Figure 1.2). A current of emitted electrons is collected, amplified and used to modulate the brightness of a cathode-ray tube. The number of electrons reflected from each spot indicates the image intensity in a current *pixel*. *Scanning Transmission Electron Microscope (STEM)* is a combination of SEM and TEM. A fine probe of electrons is scanned over a sample and transmitted electrons are being collected to form an image signal [20]. The resolution of STEM achieves 0.05 nm nowadays. *High-resolution* is an imaging mode of the electron microscope that allows the imaging of the crystallographic structure of a sample at an atomic scale [72]. Figure 1.3(a) shows a STEM (FEI

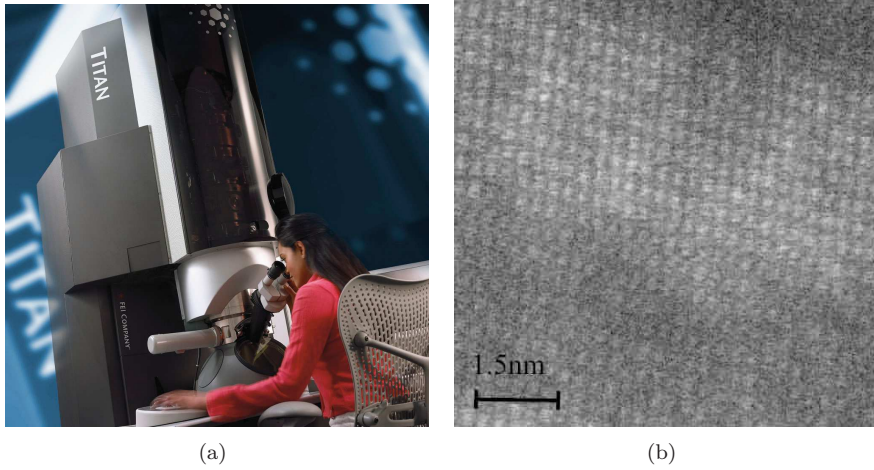


FIGURE 1.3: 1.3(a) FEI scanning transmission electron microscope; 1.3(b) STEM image recorded with in the high-resolution mode (taken from [84]).

company), and Figure 1.3(b) shows a STEM image of a $\text{LaAlO}_3/\text{SrTiO}_3$ sample recorded in the high-resolution mode. The resolution in electron microscopy is limited by *aberrations* of the magnetic lens, but not by the wavelength, as in light microscopy.

The *image recording* and interpretation in electron microscopy is more complicated and challenging than in light microscopy. A number of *aberrations* typical for electron microscopy, such as *spherical aberration*, *chromatic aberration*, *astigmatism* influence the electron beam. The *signal-to-noise ratio* is worse than in light microscopy due to the limited electron dose that can be applied to a sample. In order to obtain a higher signal-to-noise ratio the microscope exposure time has to be increased, which is not always acceptable for the real-world applications that require fast image acquisition. During the image formation process in the electron microscope a sample can be damaged, contaminated or charged. Ferromagnetic *hysteresis* and coupled dynamics in the magnetic lens system of electron microscopes degrade the microscope's performance in terms of steady-state error and transition time [82].

Electron microscopes are operated manually by skilled technicians, who execute complex and repetitive procedures, such as measuring nanoparticles, using mainly visual feedback [78]. Hence, there is a need to automate these procedures. Next generation of microscopes should not only record images but automatically extract information from the samples (chemical analysis, particle size distribution, chemical analysis) [76, 81]. The *Condor project* [25] deals with system performance and evolvability. Performance is defined as high-end image quality and measurement accuracy, productivity (fast response time), ease-of-use, and instrument autonomy (autotuning and calibration). Evolvability is the adaptability to various applications and different (and changing) requirements



FIGURE 1.4: The ideal goal of the *Condor project* is to construct the electron microscope as a modern photocamera: we press the button, we record the image with automatically determined sample characteristics.

during the planned life cycle. Within this project the FEI electron microscopes are used as an industrial reference case. The project seeks to transform the traditional electron microscopes from qualitative imaging instruments into flexible quantitative nanomeasurement tools. Simply speaking the ideal goal of Condor would be to construct a microscope similar to a modern photcamera: one presses the button and the image with automatically determined sample characteristics is recorded (Figure 1.4). This thesis deals with the part of the project, which aims for automated defocus and astigmatism correction in electron microscopy.

1.2 Problem formulation

In electron microscopy as well as in a variety of other optical devices, such as photo cameras and telescopes, *focusing* is defined as an act of making the image as sharp as possible (the image is *in-focus*) by adjusting the objective lens [9]. Astigmatism is the lens aberration that deals with the fact that the lens is not perfectly symmetric, which is present in all modern magnetic lenses. Due to astigmatism the image cannot be totally sharp, and has a different amount of defocus in different directions.

Figure 1.5 shows a photograph of a person and its synthetically generated versions: *out-of-focus* without astigmatism and *out-of-focus* with astigmatism. The *stigmatic image* (Figure 1.5(c)) is not just unsharp. We can observe a stretching in the particular (in this case in the horizontal) direction. Figure 1.6 shows experimental SEM images recorded with and without astigmatism. The two parts of the cross on the stigmatic image have different levels of unsharpness, while on the *out-of-focus* image without astigmatism they are uniformly unsharp.

In a number of practical applications both the *defocus* and the twofold astigmatism have to be corrected regularly during continuous image recording. For instance in *electron tomography*, 50-100 images are recorded at different tilt angles, where each tilting changes the defocus [84]. Other possible reasons for change in defocus and twofold astigmatism are for instance the instabilities of the electron microscope and environment, as well as the magnetic nature of some samples. Nowadays electron microscopy still requires an expert operator to trigger recording of *in-focus* and astigmatism-free images using a visual feedback [74, 78], which is a tedious task.

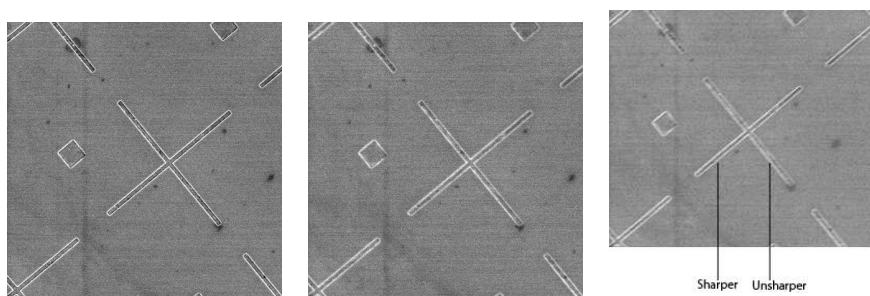
Figure 1.7 shows the scheme of correction loop. The object geometry is generally unknown. The human operator adjusts the microscope controls (*defocus* and *stigmators* in the scope of this thesis). The controls influence the shape of the electron beam, which produces a new final image. The human observes the image with the eyes and adjusts the controls again in order to improve the correction. In future the manual operation has to be automated to improve the speed, the quality and the repeatability of the measurements.

The defocus and twofold astigmatism correction methods were studied for various types of microscopy. Autofocus techniques were investigated for flu-



(a) In-focus image without astigmatism. (b) Out-of-focus image without astigmatism. (c) Out-of-focus image with astigmatism.

FIGURE 1.5: Synthetically generated images.



(a) In-focus image without astigmatism. (b) Out-of-focus image without astigmatism. (c) Out-of-focus image with astigmatism.

FIGURE 1.6: Experimental SEM images.

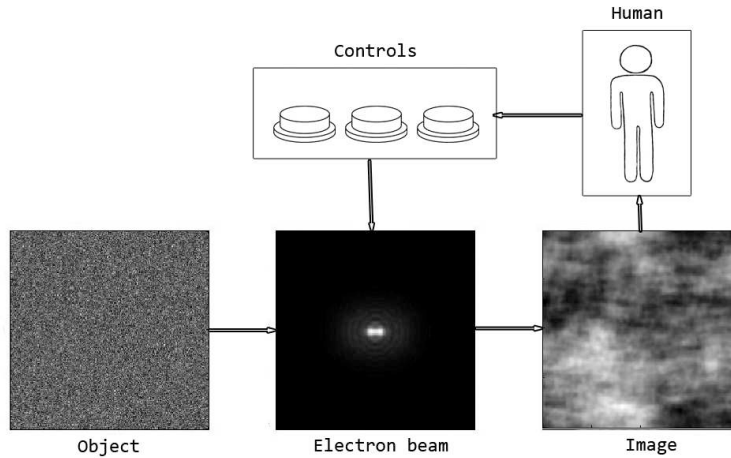


FIGURE 1.7: Defocus and astigmatism correction in electron microscopy.

orescent *light microscopy* [69, 94], non-fluorescence light microscopy [73, 40], scanning electron microscopy [62, 61]. Astigmatism is not important for light microscopy nowadays, thus it was not considered in [69, 94, 73, 40]. Few methods for simultaneous autofocus and astigmatism correction for scanning electron microscopy were proposed [16, 52]. Fourier transform-based, variance-based, autocorrelation-based iterative autofocus techniques were implemented, tested and compared for electron tomography [84], but astigmatism was not taken into account.

A number of methods implemented on aberrated-corrected microscopes are able to correct high and low aberrations, which include defocus and astigmatism.

Some of them are based on *Ronchigrams* (*shadow images*) [41, 70, 13, 32, 11]. They assume particular object geometry during the adjustment, i.e. require an *amorphous* (or structureless) sample region [70, 13, 32] or a *crystalline* sample [41]. Another group of methods is based on the image Fourier transform [22, 4, 95, 33, 96]. These methods as well as the method described in [52], can hardly be used for situations where the image Fourier transform only has a few Fourier components or is strongly influenced by unknown sample's geometry. The mentioned Ronchigram-based and Fourier transform-based methods are non-iterative, they provide the absolute measure of the aberrations. These methods correct for defocus and twofold astigmatism from a small finite number of recorded images (for example, two images in [4], three images in [22, 95, 96]). Unfortunately, these methods are not suitable for applications that require continuous operation since they are not fully automated [77] (a human operator has to point to amorphous area or enter a range of parameters). Besides some of them make use of additional equipment, such as aberration correctors or a camera for Ronchigrams recording, which is not a part of every microscope.

Most of the automatic focusing methods are based on a sharpness function, which delivers a real-valued estimate of an image quality. We study sharpness functions based on image derivative, image Fourier transform, image variance, autocorrelation and histogram. The capacity of the modern processors allows computations of a sharpness function within a negligible amount of time. However, *image recording* might require a noticeable amount of time. In particular in *scanning transmission electron microscope*, one image recording can take 1- to 30 seconds. The development of a method that requires fewer images is therefore important. A new method for rapid automated focusing is developed, based on a quadratic interpolation of the derivative-based sharpness function (*fast autofocus* method). This function has been already used before on heuristic grounds. We give a more solid mathematical foundation for this function and get a better insight into its analytical properties.

Further we consider a *focus series method*, which can act as an extension for an autofocus technique. The method is meant to obtain the astigmatism information from the *through-focus series* of images. The method is based on the moments of the image Fourier transforms. After all the method of simultaneous defocus and astigmatism correction is developed. The method is based on a three-parameter optimization (the *Nelder-Mead simplex method* or the *interpolation-based trust-region method*) of a sharpness function. We have implemented all three methods (fast autofocus method, focus series method and simultaneous defocus and astigmatism correction method) and successfully tested their performance as part of a real-world application in the STEM microscope.

1.3 Outline

In Chapter 2 we derive the models used in the following chapters. In particular a linear image formation model is explained. Models for the sample object and the microscope point spread function are given as well as the general definition of the sharpness function.

In Chapter 3 we introduce the derivative-based sharpness function explicitly and investigate its behaviour with respect to the defocus. In this chapter we show analytically that for the noise-free image formation the L_2 -norm derivative-based sharpness function reaches its optimum for the in-focus image, and does not have any other optima. Moreover, under certain assumptions the function can accurately be approximated by a quadratic polynomial. The error of this approximation can be decreased by controlling the artificial blur variable, which is given as input to the autofocus method. The proposed quadratic polynomial interpolation leads to a new autofocus method that requires recording of three or four images only.

Chapter 4 introduces a method for defocus and astigmatism correction based on the image Fourier transform, more precisely the mathematical moments of the power spectrum. The method is tested with the help of a Gaussian benchmark, as well as with the scanning electron microscopy and scanning transmis-

sion electron microscopy experimental images. The method can be used as a tool to increase the capabilities of defocus and astigmatism correction of a non-experienced scanning electron microscopy user, as well as a basis for automated application.

In Chapter 5 we study autocorrelation and intensity-based and variance-based sharpness functions. Their relation with the derivative-based sharpness function studied in Chapter 3 is discussed. The functions are demonstrated for the experimental data from the SEM.

In Chapter 6 different autofocus techniques are applied to a variety of experimental through-focus series of SEM images with different geometries. The techniques include the approaches described in the previous chapters 3-5 and the histogram-based approach. The procedure of quality *ranking* is described. It is shown that varying an extra parameter can dramatically increase the quality of an autofocus technique.

Chapter 7 explains the method of simultaneous defocus and astigmatism correction based on derivative-free optimization. Numerical simulations show that the variance-based sharpness function reaches its maximum at the *Scherzer defocus* point with zero astigmatism. This is demonstrated for the synthetic amorphous images and the ellipsoid particles image with and without noise. The simulations are based on the wave aberration point spread function discussed in Chapter 2. They show that derivative-free optimization can be beneficial for simultaneous defocus and astigmatism correction in electron microscopy. Two methods of derivative-free optimization are discussed.

The methods described in chapters 3, 5, 7 are implemented and tested online on FEI Tecnai F20 STEM. Chapter 8 describes and discusses results of this test implementation. It will be shown that the method of simultaneous defocus and astigmatism correction successfully finds proper control variable values with time and accuracy compared to a human operator.

Chapter 9 provides future recommendations. The list of frequently used symbols is provided at the end of the thesis.

Chapter 2

Modelling

To simplify our presentation we will sometimes restrict our analysis to one-dimensional images. It will be shown that for rotationally symmetric objective lenses this restriction does not affect the analysis, because the two-dimensional case in image formation is a superposition of the one-dimensional case in an orthogonal direction. The images we are going to analyse are the elements of $\mathbb{L}_2(\mathbb{R}^D)$, where the dimension $D = 1$ or $D = 2$.

2.1 Notation

For further use in the thesis we provide a few definitions below. We define the *spatial coordinate* for one-dimension as x and for two-dimension as $\mathbf{x} := (x, y)^T$; the *frequency coordinate* for one-dimension as ω and for two-dimension as $\mathbf{u} := (u, v)^T$. The *Fourier transform* \hat{f} of a function $f \in \mathbb{L}_2(\mathbb{R}^D)$ plays a fundamental role in our analysis and modelling

$$\mathfrak{F}[f(\mathbf{x})](\mathbf{u}) := \hat{f}(\mathbf{u}) := \iint_{-\infty}^{\infty} f(\mathbf{x}) e^{-i\mathbf{u} \cdot \mathbf{x}} d\mathbf{x},$$

where \cdot denotes the vector inner product. The inverse Fourier transform is defined as

$$\mathfrak{F}^{-1}[\hat{f}(\mathbf{u})](\mathbf{x}) := \frac{1}{2\pi} \iint_{-\infty}^{\infty} \hat{f}(\mathbf{u}) e^{i\mathbf{u} \cdot \mathbf{x}} d\mathbf{u},$$

For a vector $\mathbf{w} := (w_i)_{i=1}^N$ we define $\|\mathbf{w}\| := (\sum_i |w_i|^2)^{1/2}$. We define the *rotation operator* $\mathcal{R}_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ as

$$\mathcal{R}_\theta[f(\mathbf{x})] := f(\mathcal{R}_\theta \mathbf{x}), \quad (2.1)$$

where \mathcal{R}_θ is the *rotation matrix*

$$\mathcal{R}_\theta := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad (2.2)$$

and the *stretching operator* $\mathcal{J}_{\mathbf{w}} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$

$$\mathcal{J}_{\mathbf{w}}[f(\mathbf{x})] := f(\mathcal{J}_{\mathbf{w}} \mathbf{x}), \quad (2.3)$$

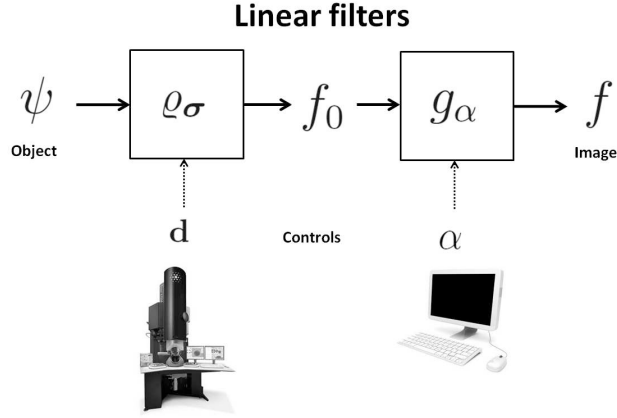


FIGURE 2.1: The image formation model.

where \mathcal{J}_w is the *stretching matrix*

$$\mathcal{J}_w := \begin{pmatrix} w_1 & 0 \\ 0 & w_2 \end{pmatrix}. \quad (2.4)$$

2.2 Image formation model

Images for which our sharpness function will be computed are the output images f of the so-called *image formation model* represented by Figure 2.1. We apply the *linear image formation model*, which is often used for different optical devices [7, 21, 49, 97]. We consider low-to-medium *magnification* of the electron microscope (*resolution* coarser than or equal to 1 nm), thus the image formation can accurately be approximated by the linear image formation model [50]. This implies that the relevant filters are linear and space invariant which easily can be described by means of *convolution* products

$$f_0 := \psi * \varrho_\sigma + \varepsilon, \quad f := f_0 * g_\alpha. \quad (2.5)$$

In (2.5) ε is the *noise function*.

The object's geometry (or the *object function*) is denoted by ψ . The filter ϱ_σ in Figure 2.1 describes the *point spread function* of an optical device. The output of the ϱ_σ filter is denoted by f_0 and is often post-processed, cf. Figure 2.1. In our model we assume that the post-processing is a filtering of the image f_0 by a *Gaussian function*, which is defined for $\mathbf{x} \in \mathbb{R}^D$ as

$$g_\alpha(\mathbf{x}) := \frac{1}{(\sqrt{2\pi\alpha})^D} e^{-\frac{\|\mathbf{x}\|^2}{2\alpha^2}}.$$

If no image post-processing is applied then $\alpha = 0$ and $f = f_0$. This filtering is often applied for denoising purposes and is a simple alternative to advanced

denoising techniques [37, 46, 55, 92]. The control variable α does not only serve to denoise the image f_0 . As explained in Chapter 3, it influences approximation errors.

2.3 Object function

We assume that the object function $\psi \in \mathbb{L}_2(\mathbb{R}^D)$. For real-world applications this is satisfied because the function ψ will have a finite domain, i.e., the object has a finite size. As a consequence ψ is bounded and continuous.

In classical signal analysis a discrete signal ψ is modelled by a finite linear combination of delta functions (cf.[54])

$$\psi(\mathbf{x}) = \sum_{k,l=1}^K a_{k,l} \delta(\mathbf{x} - \boldsymbol{\mu}_{k,l}). \quad (2.6)$$

In our setting, the finite sequence of numbers $a_{k,l}$ are the intensities of ψ (or the *object pixel values*) at $\mathbf{x} = \boldsymbol{\mu}_{k,l}$. We consider an equally distributed set of the object pixels

$$\boldsymbol{\mu}_{k,l} := \tau \mathbf{k}, \quad \tau > 0, \quad \mathbf{k} := (k, l)^T, \quad k, l = 1, \dots, K. \quad (2.7)$$

The parameter τ in (2.7) is often referred to as a *pixel width*. We define the matrix of the object pixel values as

$$\mathbf{A} := (a_{k,l})_{k,l=1}^K. \quad (2.8)$$

Property 2.1. *For the power spectrum of the object function (2.6) we have*

$$|\hat{\psi}(\mathbf{u})|^2 = \sum_{n,m} \rho_{n,m} e^{i\tau \mathbf{n} \cdot \mathbf{u}}, \quad \mathbf{n} := (n, m)^T, \quad (2.9)$$

where

$$\rho_{n,m} := \sum_{k,l} a_{k,l} a_{n+k, m+l}, \quad (2.10)$$

are the autocorrelation coefficients of the object pixel values.

Proof. The Fourier transform of the object function (2.6)

$$\hat{\psi}(\mathbf{u}) = \sum_{k,l} a_{k,l} \int_{-\infty}^{\infty} e^{-i\mathbf{x} \cdot \mathbf{u}} \delta(\mathbf{x} - \tau \mathbf{k}) d\mathbf{x} = \sum_{k,l} a_{k,l} e^{-i\tau \mathbf{k} \cdot \mathbf{u}},$$

is a periodic function with the period $\frac{2\pi}{\tau}$ in both directions. Then its squared modulus $|\hat{\psi}(\mathbf{u})|^2$ is also a periodic function with period $\frac{2\pi}{\tau}$ having the Fourier expansion

$$|\hat{\psi}(\mathbf{u})|^2 = \left(\sum_{k,l=1}^K a_{k,l} e^{-i\tau \mathbf{k} \cdot \mathbf{u}} \right) \left(\sum_{k,l=1}^K a_{k,l} e^{i\tau \mathbf{k} \cdot \mathbf{u}} \right) = \sum_{n,m=-K+1}^{K-1} \rho_{n,m} e^{i\tau \mathbf{n} \cdot \mathbf{u}},$$

where

$$\begin{aligned}\rho_{n,m} &= \frac{\tau}{2\pi} \iint_{-\frac{\pi}{\tau}}^{\frac{\pi}{\tau}} |\hat{\psi}(\mathbf{u})|^2 e^{-i\tau\mathbf{u}\cdot\mathbf{n}} d\mathbf{u} = \\ &= \frac{\tau}{2\pi} \sum_{k,l} a_{k,l} \int_{-\frac{\pi}{\tau}}^{\frac{\pi}{\tau}} \hat{\psi}^*(\mathbf{u}) e^{-i\tau(\mathbf{k}+\mathbf{m})\cdot\mathbf{u}} d\mathbf{u} = \sum_l a_{k,l} a_{k+n,l+m}^* = \sum_l a_{k,l} a_{k+n,l+m}.\end{aligned}$$

□

As a special example of an object function consider one for which the power spectrum corresponds to a Gaussian function. It can be for instance an approximation of a single particle object

$$|\hat{\psi}(\mathbf{u})|^2 = C e^{-\|\mathbf{u}\|^2 \gamma^2}, \quad C > 0, \quad \gamma \geq 0. \quad (2.11)$$

For $\gamma = 0$ in (2.11), $|\hat{\psi}|^2$ is a constant, which approximates the situation when the object is *amorphous* (or structureless).

2.4 Point spread function

In this section we discuss two possibilities of modelling the point spread function: with the Lévi stable density function and with the wave aberration function.

2.4.1 The Lévi stable density function

For a wide class of optical devices the point spread function $\hat{\varrho}_\sigma$ can accurately be approximated by a *Lévi stable density* function [7, 8, 27]. For the optical device parameter $0 < \beta \leq 1$ this function is implicitly defined by its Fourier transform

$$\hat{\varrho}_\sigma(\omega) := e^{-\sigma^{2\beta} \omega^{2\beta}/2}, \quad 0 < \beta \leq 1. \quad (2.12)$$

If $\beta = 1$ in (2.12) then ϱ and $\hat{\varrho}$ are Gaussian functions. A Gaussian function (or a composition of Gaussian functions) is often used as an approximation of the point spread function for different optical devices [46, 49, 97], including electron microscopes [16, 48]. The parameter σ in (2.15) is known as the *width* of the point spread function. For a Gaussian point spread function, the width σ is equal to its *standard deviation*. Due to the physical limitations of the optical device it has a positive lower bound: $\sigma > \sigma_0 > 0$.

For $\beta = \frac{1}{2}$ in (2.12), one obtains the Lorentzian function (or the Cauchy function). When $\beta = 1$, $\hat{\varrho}_\sigma$ has a slim tail and finite variance. When $0 < \beta < 1$, $\hat{\varrho}_\sigma$ has a fat tail and infinite variance.

In a two-dimensional setting, due to the presence of astigmatism, the point spread function is not always rotationally symmetric. Actually it is often taken as a tensor product of two one-dimensional point spread functions in the x and y directions including the possibility of system rotation

$$\hat{\varrho}_\sigma(\mathbf{u}) := \mathcal{R}_\theta \mathcal{J}_\sigma \hat{\varrho}_1, \quad (2.13)$$

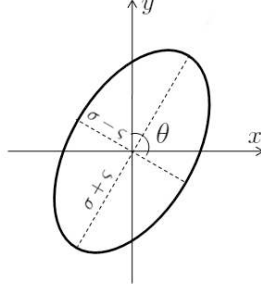


FIGURE 2.2: Asymmetric point spread function, schematic representation.

where

$$\hat{\varrho}_1 := e^{\|\mathbf{u}\|^{2\beta}/2}, \quad (2.14)$$

$$\boldsymbol{\sigma} := (\sigma - \zeta, \sigma + \zeta)^T. \quad (2.15)$$

For the Fourier transform it trivially follows that for all linear operators \mathcal{R}

$$\begin{aligned} \mathfrak{F}[f(\mathcal{R}\mathbf{x})](\mathbf{u}) &:= \iint_{-\infty}^{\infty} f(\mathcal{R}\mathbf{x}) e^{-i\mathbf{u}\cdot\mathbf{x}} d\mathbf{x} \underset{\mathbf{y}=\mathcal{R}\mathbf{x}}{=} \\ &= \frac{1}{|\det \mathcal{R}|} \iint_{-\infty}^{\infty} f(\mathbf{y}) e^{-i(\mathcal{R}^{-T}\mathbf{u})\cdot\mathbf{y}} d\mathbf{y} = \frac{1}{|\det \mathcal{R}|} \hat{f}(\mathcal{R}^{-T}\mathbf{u}). \end{aligned}$$

Since the rotation matrix (2.2) satisfies the properties $\det \mathcal{R}_\theta = 1$, $\mathcal{R}_\theta^{-T} = \mathcal{R}_\theta$, the rotation angle θ of the point spread function in Fourier space is equal to the rotation angle of the point spread function in the real space

$$\mathfrak{F}[\mathcal{R}_\theta \varrho] = \mathcal{R}_\theta \hat{\varrho}. \quad (2.16)$$

Figure 2.2 shows a schematic representation of elliptic ϱ_σ . For $\zeta = 0$ in (2.15) there is no astigmatism, and the point spread function is rotationally symmetric. For $\zeta \neq 0$ and simultaneously $\sigma = 0$ (i.e. the image is *stigmatic* and *in-focus*), ϱ_σ is symmetric with the width ζ , which means that the image is not totally sharp. Parameter θ in (2.13) indicates the unknown characteristic of the optical device.

2.4.2 Wave aberration function

In this subsection we briefly describe a different point spread function model, which takes into account the *spherical aberration* typical for electron microscopy. Figure 2.3 illustrates the ray diagram in one-dimension with the spherical aberration. The portion of the lens furthest from the optical axis brings rays to a focus nearer the lens than does the central portion of the lens. Another way of

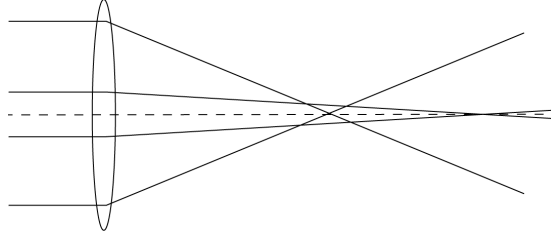


FIGURE 2.3: Ray diagram in one-dimension illustrates *spherical aberration*.

expressing this concept is to say that the optical ray path length from object point to focused image point should always be the same. This naturally implies that the focus for marginal rays is nearer to the lens than the focus for paraxial rays (those which are almost parallel to the axis). Spherical aberration is always present in magnetic lenses [71].

A detailed explanation of the electron microscope *wave aberration function* can be found in [30]. Here we only provide a short overview. In the Fourier space the wave function that enters the *sample* is given by

$$G(\mathbf{u}) = A(\mathbf{u})e^{-i\chi(\mathbf{u})}, \quad (2.17)$$

where the *aperture function* A is

$$A(\mathbf{u}) = \begin{cases} 1, & \text{if } \|\mathbf{u}\| \leq R_A \\ 0, & \text{otherwise,} \end{cases} \quad (2.18)$$

and the wave aberration function χ is defined as in [23, 30, 32]

$$\chi(\mathbf{u}) = \pi\lambda(\|\mathbf{u}\|^2d + \frac{1}{2}\lambda^2\|\mathbf{u}\|^4C_s + C_b(v^2 - u^2) + \frac{1}{2}C_cuv), \quad (2.19)$$

$\lambda, d, C_s, C_b, C_c$ represent the *wavelength*, the *defocus*, the spherical aberration, the twofold astigmatism respectively. The electron wave length λ is related to the electron energy E , the speed of light c , the electron's rest mass m_0 and the Planck's constant \hbar (cf. [30])

$$\lambda = \frac{\hbar c}{\sqrt{E(2m_0c^2 + E)}}. \quad (2.20)$$

The electron energy E can be set to different values within a certain range, which depends on the particular microscope. The defocus and astigmatism variables d, C_b, C_c can be controlled by a human operator. The defocus control variable will be in detail discussed in the next section. The spherical aberration C_s is the characteristics of the microscope.

The aperture radius R_A in (2.18) controls the *convergence semi-angle* η_A of the beam

$$R_A = \frac{\eta_A}{\lambda}. \quad (2.21)$$

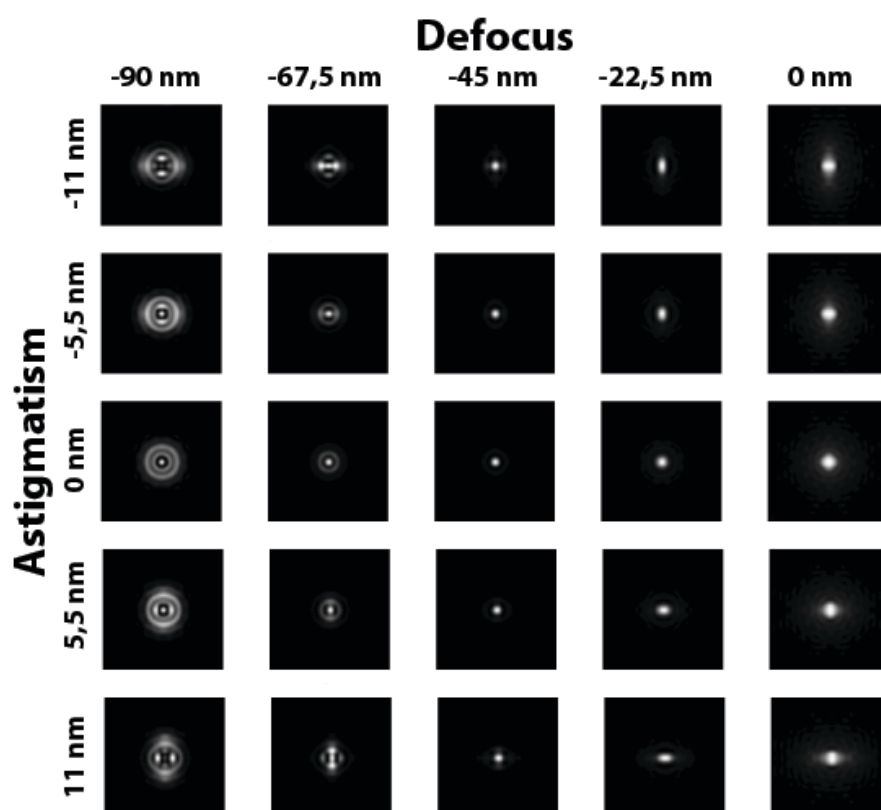


FIGURE 2.4: Simulations of the point spread function (2.22) for different defocus d and astigmatism C_b values. Astigmatism parameter C_c is set to zero.

The point spread function is the intensity of the scanning probe, that is the inverse Fourier transform of the wave function (2.17) [30]

$$h(\mathbf{x}) = C|\mathfrak{F}^{-1}[G]|^2, \quad (2.22)$$

where C is a normalization constant. The microscope defocus can be used to offset the effect of spherical aberration. The ideal control variable values for the wave aberration model are known as *Scherzer conditions* (see [30]). They are expressed through the spherical aberration of the electron microscope, i.e. the *Scherzer defocus* point is defined as

$$d_{Sh} := -(1.5C_s\lambda)^{1/2}, \quad (2.23)$$

and the *Scherzer aperture* is defined

$$R_{A_{Sh}} := \frac{1}{\lambda} \left(\frac{6\lambda}{C_s} \right)^{\frac{1}{4}}. \quad (2.24)$$

Then the *Scherzer convergence semi-angle* can be trivially computed as

$$\eta_{A_{Sh}} := \lambda R_{A_{Sh}}.$$

Figure 2.4.2 shows the simulations of the point spread function based on the model described in this subsection. The Scherzer defocus value in this simulation is equal to 45 nm. We observe that the width of the point spread function in this simulation for 45 nm defocus is smaller than the width for 0 nm defocus.

For analytical observations in this thesis we use the point spread function model that does not take into account the spherical aberration. The wave aberration model explained in this section is used for numerical simulations and experiments in Chapter 7.

2.5 Defocus and stigmator control variables

Astigmatism is a lens aberration caused by rotational asymmetry of the magnetic lens. Figure 2.5(a) shows a ray diagram for the astigmatism-free situation. The lens has one focal point F . The only adjustable parameter is the current through the lens; it changes the focal length of the lens and focuses the magnetic beam on the image plane [52]. The current is controlled by the *defocus* variable d . Astigmatism implies that the rays traveling through a horizontal plane will be focused at a focal point different from the rays traveling through a vertical plane (Figure 2.5(b)). Figures 2.5(a), 2.5(b) show diagrams in two-dimension, which is different from Figure 2.3 that shows ray diagram in one-dimension. This leads to two different focal points F_1 and F_2 of the lens. The image cannot be totally sharp. Due to the presence of astigmatism, the electron beam becomes elliptic.

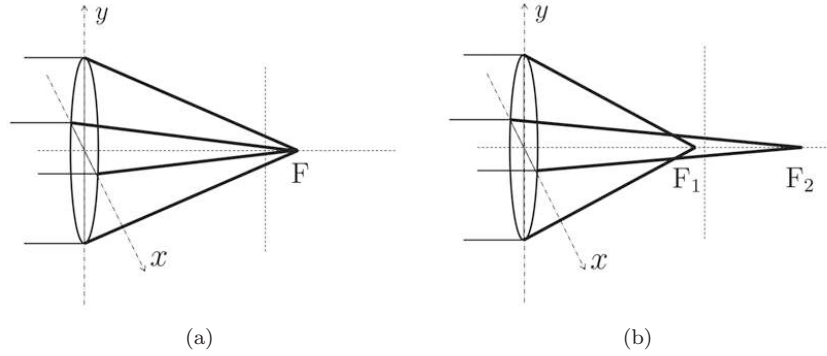


FIGURE 2.5: Ray diagrams in two-dimensions: 2.5(a) for a lens without *astigmatism* with one focal point; 2.5(b) a lens with *astigmatism* with two focal points.

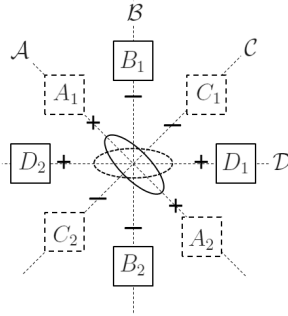


FIGURE 2.6: Typical for the electron microscope configuration of electrostatic *stigmators* [52].

For astigmatism correction in electron microscopy, electrostatic or electromagnetic *stigmators* are used. They produce an electromagnetic field for the correction of the ellipticity of the electron beam [59]. A typical configuration of them is shown in Figure 2.5. The elliptic electron beam is depicted in the middle of the scheme. Currents of magnitude I_1 pass through coils A_1 , A_2 , C_1 , and C_2 , while currents of magnitude I_2 pass through coils B_1 , B_2 , D_1 , and D_2 . The field generated by A_1 , A_2 , C_1 and C_2 influences the stretching of the electron beam along the two orthogonal axes \mathcal{A} and \mathcal{C} . Similarly, the field generated by coils B_1 , B_2 , D_1 and D_2 influences the stretching along the two orthogonal axes \mathcal{C} and \mathcal{D} [52]. The angle between axes \mathcal{A} and \mathcal{B} is always $\frac{\pi}{4}$. Magnitude and direction of the current through the coils A_1 , A_2 , C_1 and C_2 are controlled by the *x-stigmator* control variable d_x , and magnitude and direction of the current through coils B_1 , B_2 , D_1 and D_2 are controlled by the *y-stigmator* control variable d_y .

In this thesis we deal with the vector of three microscope control variables

$$\mathbf{d} := (d, d_x, d_y)^T. \quad (2.25)$$

The vector of the ideal control variable values (the setting when the output image has the highest possible quality) is defined as

$$\mathbf{d}_0 := (d_0, d_{x_0}, d_{y_0})^T. \quad (2.26)$$

The goal of the autofocus procedure is to find the value of d_0 . The goal of the automated astigmatism correction procedure is to find the values of d_{x_0}, d_{y_0} .

We define

$$\begin{aligned} d_h &:= d - d_0, & d_{x_h} &:= d_x - d_{x_0} + 1, & d_{y_h} &:= d_y - d_{y_0} + 1, \\ \mathbf{d}_h &:= (d_h, d_h)^T, & \mathbf{d}_{x_h} &:= (d_{x_h}, \frac{1}{d_{x_h}})^T, & \mathbf{d}_{y_h} &:= (d_{y_h}, \frac{1}{d_{y_h}})^T. \end{aligned}$$

The point spread function can be expressed through the control variables as

$$\hat{\varrho}_\sigma = \mathcal{T}_\mathbf{d} \hat{\varrho}_1,$$

where the operator $\mathcal{T}_\mathbf{d} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is defined as $\mathcal{T}_\mathbf{d}[f(\mathbf{x})] := f(\mathcal{T}_\mathbf{d}\mathbf{x})$ with the transformation matrix

$$\begin{aligned} \mathcal{T}_\mathbf{d} &:= \mathcal{J}_{\mathbf{d}_h} \mathcal{J}_{\mathbf{d}_{x_h}} \mathcal{R}_{\pi/4} \mathcal{J}_{\mathbf{d}_{y_h}} \mathcal{R}_{-\pi/4} = \\ &\frac{d}{2} \begin{pmatrix} d_{x_h} (d_{y_h} + \frac{1}{d_{y_h}}) & d_{x_h} (d_{y_h} - \frac{1}{d_{y_h}}) \\ \frac{1}{d_{x_h}} (d_{y_h} - \frac{1}{d_{y_h}}) & \frac{1}{d_{x_h}} (d_{y_h} + \frac{1}{d_{y_h}}) \end{pmatrix}. \end{aligned}$$

For $d_y = d_{y_0}$ one has

$$\sigma = \frac{d_h}{2} \left(d_{x_h} + \frac{1}{d_{x_h}} \right), \quad \varsigma = \frac{d_h}{2} \left(d_{x_h} - \frac{1}{d_{x_h}} \right), \quad \theta = 0,$$

and for $d_x = d_{x_0}$ and $d_y = d_{y_0}$

$$\sigma = d - d_0, \quad \varsigma = 0, \quad \theta = 0.$$

2.6 The sharpness function

Many existing *autofocus* methods are based on a *sharpness function* $S : \mathbb{L}_2(\mathbb{R}^2) \rightarrow \mathbb{R}$, a real-valued estimate of the image's sharpness. In the literature a number of sharpness functions have been considered and discussed for different optical devices, such as photographic and video cameras [17, 28, 34], telescopes [29, 47], light microscopes [6, 24, 40, 69, 73, 93, 94] and electron microscopes [16, 53, 61, 62, 74, 84]. For a *through-focus series* of images the sharpness

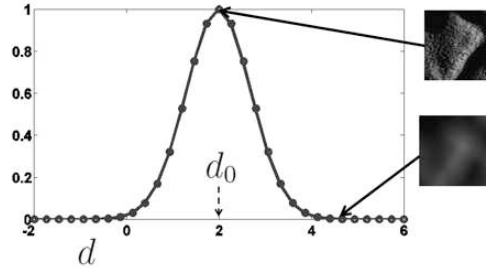


FIGURE 2.7: Sharpness function S reaches its optimum at the in-focus image. The goal of the autofocus procedure is to find the in-focus value d_0 .

function is computed for different values of d given a fixed value of α . A typical sharpness function shape is shown in Figure 2.7. The image at the defocus $d = d_0$ is sharp or *in-focus* when the sharpness function reaches its optimum. An image away from d_0 is called *out-of-focus*. Ideally sharpness functions should have a single optimum (maximum or minimum) at the in-focus image. The sharpness functions are also used for studies of the *hysteresis* in electromagnetic lenses [82, 83] and reconstructions of three-dimensional microscopic objects [36, 49].

In this thesis we will use the following notations: $S[f]$ is the sharpness function value computed for the image f ; $S(\mathbf{d})$ is the sharpness function value computed for the image f , recorded with machine control variables \mathbf{d} ; similarly for only autofocus problem we will use $S(d)$; as the defocus control variable d is closely related to the point spread function width σ it is sometimes convenient to define the sharpness function as $S(\sigma)$, or for the two-dimensional setting $S(\boldsymbol{\sigma})$.

We assume that for our autofocus procedure α is fixed and a finite number, say N , of values for the defocus control d are chosen: d_1, \dots, d_N with $d_1 < d_2 < \dots < d_N$. For each of the corresponding images f_1, f_2, \dots, f_N the value of the sharpness function is computed

$$S_i := S(d_i - d_0), \quad i = 1, \dots, N. \quad (2.27)$$

The problem of *automated focusing* (or autofocus) is to estimate the location d_0 of the optimum of S given the points S_i in (2.27). For simultaneous defocus and astigmatism correction stigmator controls are adjusted as well, and the goal is to estimate \mathbf{d}_0 from the values of the sharpness function computed at different points \mathbf{d} .

An autofocus method can be established in two different ways described below.

- *Static autofocus.* A number of images is taken within a wide defocus range and for each image the sharpness function is computed giving a

discrete set of sharpness function values . Then the optimal image (the in-focus image) is determined as the optimum of this discrete set of data (*course focusing*). Eventually the same procedure is repeated within a smaller defocus range around the optimum, found in the previous step (*fine focusing*).

- *Dynamic autofocus*. Starting out with an initial defocus parameter d , an iterative optimization method is used to find the optimal defocus value d_0 , (for example, the Fibonacci search [40, 94], the *Nelder-Mead simplex method* [68] or the *interpolation-based trust-region method* [60]).

The first approach requires recording of about 20-30 images, which can be time-consuming for real-world applications. The goal of the second approach is to minimize the number of images necessary to perform the autofocus. It usually requires at least 10 images for the autofocus procedure. On the other hand, the first approach is more robust to the local optima in the sharpness function, which often occur in electron microscopy due to the *noise* in the image formation.

In the next chapters we will discuss different types of sharpness functions based on image derivative, Fourier transform, variance, autocorrelation, intensity and histogram, denoted as S^{der} , S^{ft} , S^{var} , S^{ac} , S^{int} , S^{his} respectively.

2.7 Discrete images

In real-world applications the image f is always camera-recorded, and therefore discrete and bounded. Assume for $X \in \mathbb{R}$ the support of f is

$$\mathbb{X} := [0, X]^D,$$

i.e., $f(\mathbf{x}) = 0$ for \mathbf{x} outside of \mathbb{X} . For $i = 1, \dots, N$ we define the grid points $x_i := \frac{\Delta x}{2} + (i-1)\Delta x$, where $\Delta x := \frac{X}{N}$ (for the default $X = 1$, $\Delta x = \frac{1}{N}$).

The microscopy images are discrete images that can be represented by a matrix

$$\mathbf{F} := (f_{i,j})_{i,j=1}^N, \quad (2.28)$$

of the *image pixel values*

$$f_{i,j} := f(x_i, x_j). \quad (2.29)$$

We use the *mid-point rule* for approximation of image integration. Hence the integration of the image with compact support over the image domain in two-dimension is approximated by

$$\int_{\mathbb{X}} f(\mathbf{x}) d\mathbf{x} \doteq (\Delta x)^2 \sum_{i,j} f(x_i, x_j) \stackrel{\Delta x=1/N}{=} \frac{1}{N^2} \sum_{i,j=1}^N f_{i,j}, \quad (2.30)$$

similarly

$$\|f\|_{L_p} \doteq \left(\frac{1}{N^2} \sum_{i,j=1}^N f_{i,j}^p \right)^{1/p}.$$

Similarly for one-dimension

$$\int_{\mathbb{X}} f(x) dx \doteq \frac{1}{N} \sum_{i=1}^N f_i, \quad \|f\|_{L_p} \doteq \left(\frac{1}{N} \sum_i f_i^p \right)^{1/p}.$$

For the given discrete image the sampling period Δx is fixed. Thus considering higher order integration will not decrease the integration error.

Below we discuss the numerical differentiation of the discrete images. By dropping the limit in the definition of the *differential operator*

$$\frac{\partial}{\partial x} f(\mathbf{x}) := \lim_{\epsilon \rightarrow 0} \frac{f(x + \epsilon, y) - f(x, y)}{\epsilon}$$

and keeping ϵ fixed at a distance of $k \in \mathbb{N}$ pixels, we obtain a *finite difference* approximation at (x_i, x_j)

$$\frac{\partial}{\partial x} f(x_i, x_j) \doteq \frac{1}{(k\Delta x)} (f_{i+k,j} - f_{i,j}). \quad (2.31)$$

We refer to k as the *pixel difference* parameter for the discrete image derivatives. The directional derivative of f at \mathbf{x} in the unit direction $\mathbf{w} := (w_x, w_y)^T \in \mathbb{R}^2$ is

$$\mathbf{w}^T \nabla f(\mathbf{x}) := w_x \frac{\partial}{\partial x} f(\mathbf{x}) + w_y \frac{\partial}{\partial y} f(\mathbf{x}).$$

For the polar angle $\pi/4$ it follows that $\mathbf{w}_{\pi/4} := \frac{1}{\sqrt{2}}(1, 1)^T$ and

$$\mathbf{w}_{\pi/4}^T \nabla f(x_i, x_j) \doteq \frac{1}{\sqrt{2}(k\Delta x)} (f_{i+k,j} - 2f_{i,j} + f_{i,j+k}),$$

for the angle $-\pi/4$ it follows that $\mathbf{w}_{-\pi/4} := \frac{1}{\sqrt{2}}(1, -1)^T$ and

$$\mathbf{w}_{-\pi/4}^T \nabla f(x_i, x_j) \doteq \frac{1}{\sqrt{2}(k\Delta x)} (f_{i+k,j} - f_{i,j+k}).$$

Two alternative derivative interpolation solutions appear commonly in the literature: fitting polynomial approximations [3, 90] and smoothing with a filter, for instance a Gaussian function [18, 45]

$$\frac{\partial}{\partial x} f(\mathbf{x}) \doteq \mathcal{D}_x := \frac{\partial}{\partial x} (f * g) = f * \frac{\partial}{\partial x} g. \quad (2.32)$$

Chapter 3

Derivative-based approach

In this chapter we introduce the derivative-based sharpness function explicitly and investigate its behaviour [64, 67]. The advantage of using derivative-based sharpness functions has been shown experimentally for SEM [61, 62] and other optical devices [6, 49]. The use of these functions is heuristic. Usually they are based on the assumption that the *in-focus* image has a larger difference between neighbouring *pixels* than the *out-of-focus* image. In this chapter we show analytically that for the noise-free image formation the L_2 -norm derivative-based sharpness function reaches its optimum for the in-focus image, and does not have any other optima. Moreover, under certain assumptions the function can accurately be approximated by a quadratic polynomial. The error of this approximation can be decreased by controlling the artificial blur variable, which is given as input to the *autofocus* method. The proposed quadratic polynomial interpolation leads to a new autofocus method that requires recording of three or four images only. This provides the speed improvement in comparison with existing approaches, which usually require recording of more than ten images for autofocus.

For the simplification of our analysis in the beginning of this chapter (sections 3.1-3.3) we restrict the theoretical observations to a one-dimensional setting. In the following sections, as well as in our numerical experiments and real-world application two-dimensional images are used. Throughout the chapter we use the notation S instead of S^{der} for the derivative-based sharpness function, to be defined below.

3.1 Derivative-based sharpness function

The derivative-based sharpness function is defined (cf.[6, 34, 40, 94])

$$S := \left\| \frac{\partial^n}{\partial x^n} f \right\|_{L_p}^p, \quad p = 1, 2. \quad (3.1)$$

For $n = 0$ in (3.1) we obtain a so-called *intensity-based sharpness function* S^{int} , which will be discussed in Chapter 5. In different literature sources different norms are applied to the image derivatives for autofocus purposes, i.e. $p = 1$ in [26, 39] or $p = 2$ in [17, 47]. In this chapter we mostly focus on $p = 2$ in (3.1).

It will be explained below that L_2 -norm derivative-based sharpness functions are less sensitive to *noise* than L_1 -norm based. For the *linear image formation model* (2.5), we have therefore

$$S = \left\| \frac{\partial^n}{\partial x^n} (\psi * \varrho_\sigma * g_\alpha) \right\|_{L_2}^2. \quad (3.2)$$

As explained in Section 2.6 the problem of automated focusing is to estimate the optimum location d_0 of the sharpness function from the given points (2.27). In this chapter our aim is to do this using a small number of recorded images, i.e., $N=3$ or $N=4$, while in the literature $N>10$ is usually used [34, 40, 94, 97]. For this purpose we will look for the function shape which can accurately be approximated by a quadratic polynomial. In Section 3.3 error estimates of such an approximation for derivative-based sharpness function are provided.

In some practical applications (cf.[34, 40, 94]) an appropriate power p of the sharpness function, i.e. the function S^p is used as a sharpness function. Here p is usually taken to be $\frac{1}{2}, 1, 2$. The power p does not influence the optimum position of the sharpness functions. However, it influences the function shape, which can simplify the task of finding an optimum in a real-world application.

In the next sections we collect some useful properties of the derivative-based sharpness function. First we deal with general properties of S in case the spread function ϱ_σ is a *Lévi stable density* function. Further we restrict ourselves to the Gaussian point spread functions and study in more detail properties of S for a typical collection of object functions: a Gaussian benchmark and a more general case of a digital image.

3.2 General properties

In this section we discuss basic properties of the derivative-based sharpness function in one-dimensional setting.

Property 3.1. *The sharpness function (3.2) can be expressed as follows*

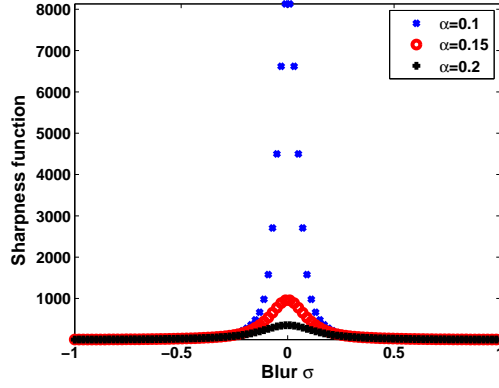
$$S(\sigma) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^{2n} |\hat{\psi}(\omega)|^2 e^{-\sigma^{2\beta} \omega^{2\beta}} e^{-\alpha^2 \omega^2} d\omega. \quad (3.3)$$

Proof. For $\hat{\psi}, \hat{g}, \hat{f}$, the Fourier transforms of ψ, g, f respectively, it holds that $\hat{f} = \hat{\psi} \hat{\varrho}_\sigma \hat{g}_\alpha$. Then from Parseval's identity we find

$$S(\sigma) = \left\| \frac{\partial}{\partial x} f \right\|_{L_2}^2 = \frac{1}{2\pi} \|\omega^n \hat{f}\|_{L_2}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^{2n} |\hat{\psi}(\omega)|^2 |\hat{\varrho}_\sigma(\omega)|^2 |\hat{g}_\alpha(\omega)|^2 d\omega.$$

□

The following corollaries follow directly from Property 3.1.

FIGURE 3.1: Numerically computed sharpness functions S .

Corollary 3.1. *The sharpness function (3.2) is smooth, and is strictly increasing for $\sigma < 0$ and strictly decreasing for $\sigma > 0$.*

Corollary 3.2. *For $\alpha > 0$ the sharpness function (3.2) has a finite maximum at $\sigma = 0$*

$$\max_{\sigma} S(\sigma) = S(0).$$

Figure 3.1 shows the numerically computed sharpness function S for different values of α . From now on we consider a Gaussian point spread function, i.e. $\beta = 1$ in (2.12). We set $n = 1$ in (3.1).

Property 3.2. *For the function (2.11) and the Gaussian point spread function we have*

$$S(\sigma) = \frac{C}{4\sqrt{\pi}(\sigma^2 + \alpha^2 + \gamma^2)^{\frac{3}{2}}}.$$

Proof. By substituting $\eta = \sqrt{\sigma^2 + \alpha^2 + \gamma^2}$ into the identity

$$\int_{-\infty}^{\infty} \omega^2 e^{-\eta^2 \omega^2} d\omega = \frac{\sqrt{\pi}}{2\eta^3}, \quad (3.4)$$

we obtain

$$S(\sigma) = \frac{C}{2\pi} \int_{-\infty}^{\infty} \omega^2 e^{-(\sigma^2 + \alpha^2 + \gamma^2)\omega^2} d\omega = \frac{C}{4\sqrt{\pi}(\sigma^2 + \alpha^2 + \gamma^2)^{\frac{3}{2}}}.$$

□

We also observe that the location d_0 of the maximum of S does not depend on α . This will be true in general. Note that for the object function (2.11) the sharpness function to the power $-2/3$ is a quadratic polynomial

$$S^{-2/3}(d - d_0) = \sqrt[3]{\frac{\pi}{C^2}}((d - d_0)^2 + \alpha^2 + \gamma^2). \quad (3.5)$$

It will be shown that in the general case the function $S^{-2/3}$ can be well approximated by a quadratic polynomial for suitable choices of the blur variable α . The quadratic shape of the sharpness function makes finding its optimum faster and more robust in the real-world applications.

3.3 Digital image object

In this section we consider a digital image object (2.11) with autocorrelation coefficients (2.10).

Property 3.3. *The sharpness function S is expressed by means of the autocorrelation coefficients (2.10) as follows*

$$S(\sigma) = \frac{1}{8\sqrt{\pi}(\alpha^2 + \sigma^2)^{3/2}} \sum_m \rho_m \left(2 - \frac{m^2\tau^2}{\alpha^2 + \sigma^2}\right) e^{-\frac{1}{4}\frac{m^2\tau^2}{\alpha^2 + \sigma^2}}. \quad (3.6)$$

Proof. After we rewrite the sharpness function (3.3) for $\beta = 1$ as

$$S(\sigma) = \frac{1}{2\pi(\sigma^2 + \alpha^2)^{3/2}} \int_{-\infty}^{\infty} \omega^2 |\hat{\psi}\left(\frac{\omega}{\sqrt{\alpha^2 + \sigma^2}}\right)|^2 e^{-\omega^2} d\omega.$$

and substitute the expression for the power spectrum (2.9), we achieve

$$S(\sigma) = \frac{1}{2\pi(\sigma^2 + \alpha^2)^{3/2}} \sum_m \rho_m \int_{-\infty}^{\infty} \omega^2 e^{\frac{im\omega\tau}{\sqrt{\alpha^2 + \sigma^2}}} e^{-\omega^2} d\omega. \quad (3.7)$$

Using the identity

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} e^{i\eta\omega} d\omega = \frac{(2 - \eta^2)e^{-\frac{\eta^2}{4}}}{8\sqrt{\pi}},$$

we obtain (3.6) directly from (3.7). \square

In the two theorems below we approximate the sharpness function S by a function of the type $\frac{C}{(\alpha^2 + \sigma^2)^{3/2}}$ in such a way that S can be written as

$$S(\sigma) = \frac{C}{(\alpha^2 + \sigma^2)^{3/2}} (1 + R(\sigma)), \quad (3.8)$$

where C depends only on the *object pixel values* (2.8), i.e. $C = C(\mathbf{a})$ and a relative error R , which can be small in typical circumstances. This implies that the function (3.5) can be expressed as

$$S^{-2/3}(d - d_0) = \mathcal{P}(d)(1 + \epsilon(d)),$$

where \mathcal{P} is a second order polynomial. For a small error $R(\sigma)$, the relative error $\epsilon(d)$ will be small: $\epsilon(d) \doteq -\frac{2}{3}R(\sigma)$.

In practical applications the value of σ is important in relation to the *pixel width* τ . For instance if $\sigma \gg \tau$, the image is totally out-of-focus (for example, Figure 3.2(e)). It is often the case that $\sigma > \tau$, but not $\sigma \gg \tau$. However, by controlling the blur α , the value $\sqrt{\alpha^2 + \sigma^2}$ can be much larger than τ , which is important for our error analysis in the next theorems.

Theorem 3.1. *The sharpness function can be expressed as follows*

$$S(\sigma) = \frac{C_1}{2\pi(\alpha^2 + \sigma^2)^{3/2}}(1 + R_1(\sigma)), \quad (3.9)$$

where

$$|R_1(\sigma)| \leq K_1 \frac{\tau}{\sqrt{\alpha^2 + \sigma^2}}, \quad (3.10)$$

and C_1, K_1 depend only on the object pixel values, i.e. $a_{k,l}$ in (2.8).

Proof. Splitting $e^{\frac{im\tau\omega}{\sqrt{\alpha^2 + \sigma^2}}}$ into $(e^{\frac{im\tau\omega}{\sqrt{\alpha^2 + \sigma^2}}} - 1) + 1$ in (3.7), one obtains

$$S(\sigma) = \frac{1}{2\pi(\sigma^2 + \alpha^2)^{3/2}} \left(\underbrace{\int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} d\omega \sum_m \rho_m}_{C_1} + \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} \sum_m \rho_m (e^{\frac{im\tau\omega}{\sqrt{\alpha^2 + \sigma^2}}} - 1) d\omega \right). \quad (3.11)$$

Applying (3.4) for $\eta = 1$, one obtains

$$C_1 = \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} d\omega \sum_m \rho_m = \frac{\sqrt{\pi}}{2} \|\mathbf{a}\|_1.$$

To estimate R_1 observe that

$$|e^{i\eta} - 1| = 2 \left| \sin \frac{\eta}{2} \right| \leq |\eta|, \quad \eta \in \mathbb{R}, \quad (3.12)$$

for $\eta = \frac{m\tau\omega}{\sqrt{\alpha^2 + \sigma^2}}$, and consequently

$$\left| \sum_m \rho_m (e^{\frac{im\tau\omega}{\sqrt{\alpha^2 + \sigma^2}}} - 1) \right| \leq \left(\sum_m |m| \rho_m \right) \frac{|\omega| \tau}{\sqrt{\alpha^2 + \sigma^2}}. \quad (3.13)$$

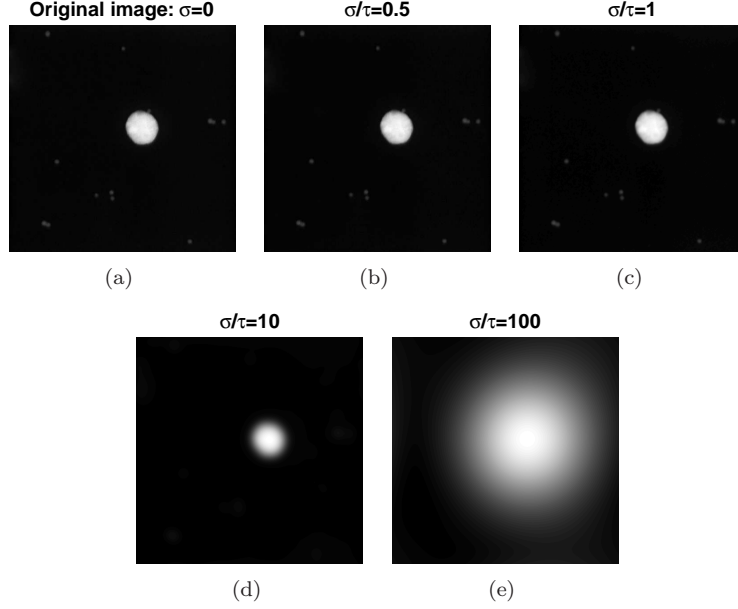


FIGURE 3.2: Artificially blurred images of a gold particle with different values of σ/τ .

From the estimate (3.13) and $\int_{-\infty}^{\infty} |\omega|^3 e^{-\omega^2} d\omega = 1$ it follows that

$$\left| \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} \sum_m \rho_m \left(e^{\frac{im\tau\omega}{\sqrt{\alpha^2 + \sigma^2}}} - 1 \right) d\omega \right| \leq \left(\sum_m |m| \rho_m \right) \frac{\tau}{\sqrt{\alpha^2 + \sigma^2}}.$$

Then the statement of the theorem follows directly with

$$K_1 = \frac{2}{\sqrt{\pi}} \frac{\sum_m |m| \rho_m}{\sum_m \rho_m}$$

in (3.9). □

It follows from the theorem that the function (3.5) can be approximated by a quadratic polynomial at any accuracy by increasing the value of the blur α .

Now let $\sigma \leq \tau$. This means that the image is almost in-focus and might be only slightly unsharp. Figures 3.2(a)-3.2(c) show examples of artificially blurred images. From left to right: original image, blurred image with $\sigma/\tau = 0.5$, blurred image with $\sigma/\tau = 1$. We hardly detect differences between the original and blurred images. However, if we zoom into the details (figures 3.3(a)-3.3(c)) the difference is visible. This corresponds to the *fine focusing*, which is considered in the theorem below.

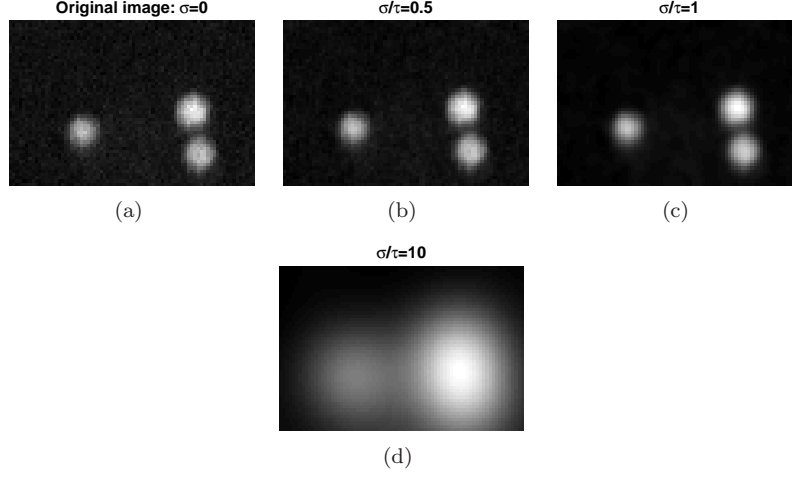


FIGURE 3.3: Artificially blurred images of a gold particle with different values of σ/τ . The images are the magnified versions of those shown in Figure 3.2. Only if we zoom into the small particles we observe the difference in the image quality for the small values of σ .

Theorem 3.2. *The sharpness function can be expressed as follows*

$$S(\sigma) = \frac{C_2}{2\pi(\alpha^2 + \sigma^2)^{3/2}}(1 + R_2(\sigma)), \quad (3.14)$$

where

$$|R_2(\sigma)| \leq K_2 \frac{\alpha^2 + \sigma^2}{\tau^2}, \quad (3.15)$$

and C_2, K_2 depend only on the object pixel values, i.e. $a_{k,l}$ in (2.8).

Proof. Splitting $\sum_m \rho_m$ into $\rho_0 + \sum_{m \neq 0} \rho_m$ in (3.7) one obtains

$$S(\sigma) = \frac{1}{2\pi(\sigma^2 + \alpha^2)^{3/2}} \left(\underbrace{\rho_0 \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} d\omega}_{C_2} + \sum_{m \neq 0} \rho_m \int_{-\infty}^{\infty} \omega^2 e^{-\frac{im\tau\omega}{\sqrt{\alpha^2 + \sigma^2}}} e^{-\omega^2} d\omega \right),$$

$$C_2 = \rho_0 \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} d\omega = \frac{\sqrt{\pi}}{2} \|\mathbf{a}\|_2.$$

To estimate R_2 observe that

$$\left| \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} e^{i\eta\omega} d\omega \right| = \left| \frac{\sqrt{\pi}}{4} (2 - \eta^2) e^{-\frac{\eta^2}{4}} \right| \leq \frac{4}{\eta^2}, \quad (3.16)$$

i.e. substitute $\eta = \frac{m\tau}{\sqrt{\alpha^2 + \sigma^2}}$

$$\left| \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} e^{\frac{im\tau\omega}{\sqrt{\alpha^2 + \sigma^2}}} d\omega \right| \leq 4 \frac{\alpha^2 + \sigma^2}{m^2 \tau^2}.$$

Then the statement of the theorem follows with

$$K_2 = \frac{8}{\sqrt{\pi}} \frac{\sum_{m \neq 0} \frac{\rho_m}{m^2}}{\rho_0}$$

in (3.15). \square

Theorem 3.2 considers the situation of a very fine focusing, which is different from Theorem 3.1, where a more general case is considered. However, it is shown that in both situations the function $S^{-2/3}$ can be approximated by a quadratic polynomial with a given accuracy. This coincides with findings of Property 3.2 for the benchmark object (2.11).

Further we provide one more representation of the sharpness function with a different error estimates, which is controlled by $\frac{\sigma}{\alpha}$.

Theorem 3.3. *The sharpness function can be expressed as follows*

$$S(\sigma) = \frac{1}{2\pi(\sigma^2 + \alpha^2)^{3/2}} \left(\int_{-\infty}^{\infty} \omega^2 |\hat{\psi}(\frac{\omega}{\alpha})|^2 e^{-\omega^2} d\omega + R_3(\sigma) \right), \quad (3.17)$$

where

$$|R_3(\sigma)| \leq \left(\sum_m |m| \rho_m \right) \frac{\tau}{\alpha} \left(\frac{\sigma}{\alpha} \right)^2.$$

Proof. It is clear that in (3.17) we have

$$\begin{aligned} R_3 &= \int_{-\infty}^{\infty} \omega^2 \left(\left| \hat{\psi}\left(\frac{\omega}{\sqrt{\alpha^2 + \sigma^2}}\right) \right|^2 - \left| \hat{\psi}\left(\frac{\omega}{\alpha}\right) \right|^2 \right) e^{-\omega^2} d\omega = \\ &= \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} \sum_m \rho_m \left(e^{im\tau \frac{\omega}{\sqrt{\alpha^2 + \sigma^2}}} - e^{im\tau \frac{\omega}{\alpha}} \right) d\omega. \end{aligned}$$

Using (3.12), we obtain

$$|R_3| \leq 2 \int_{-\infty}^{\infty} \omega^2 e^{-\omega^2} \sum_m \rho_m \left| \sin \frac{m\tau}{2} \omega \left(\frac{1}{\alpha} - \frac{1}{\sqrt{\alpha^2 + \sigma^2}} \right) \right| d\omega.$$

Moreover, we have

$$\frac{1}{\alpha} - \frac{1}{\sqrt{\alpha^2 + \sigma^2}} = \frac{1}{\alpha \sqrt{1 + (\frac{\sigma}{\alpha})^2}} \frac{(\frac{\sigma}{\alpha})^2}{1 + \sqrt{1 + (\frac{\sigma}{\alpha})^2}} \leq \frac{\sigma^2}{\alpha^3}.$$

Therefore,

$$|R_3| \leq \frac{\sigma^2 \tau}{\alpha^3} \int_{-\infty}^{\infty} \omega^3 e^{-\omega^2} \sum_m |m| \rho_m d\omega = \left(\sum_m |m| \rho_m \right) \frac{\tau}{\alpha} \left(\frac{\sigma}{\alpha} \right)^2.$$

\square

In the next section it will be shown that the role of the artificial blur control variable α in a higher dimension becomes more important due to the possible presence of astigmatism. If we only do the autofocus the proper choice of α helps to avoid the multiple optima of the sharpness function. If we perform automated simultaneous defocus and astigmatism correction, by controlling α we can improve the shapes of the sharpness function and increase the speed of optimization.

3.4 Two-dimensional setting

In this section we provide the general properties of the derivative-based sharpness function in two-dimensional setting

$$S_{n,m} := \left\| \frac{\partial^n}{\partial x^n} \frac{\partial^m}{\partial y^m} f \right\|_{L^p}^p, \quad p = 1, 2. \quad (3.18)$$

Below we consider the particular case

$$S := \left\| |\nabla f| \right\|_{L^2}^2 = S_{1,0} + S_{0,1}. \quad (3.19)$$

Property 3.4. *If f is given by (2.5) with the point spread function (2.13), then the sharpness function (3.19) can be written as follows*

$$S(\sigma) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\mathbf{u}\|^2 |\hat{\psi}(\mathbf{u})|^2 e^{-\|(\mathcal{J}_\sigma \mathcal{R}_\theta \mathbf{u})\|^{2\beta}} e^{-\|\mathbf{u}\|^2 \alpha^2} d\mathbf{u}. \quad (3.20)$$

Proof. Because of Parseval's identity we have

$$S(\sigma) = \frac{1}{2\pi} \|u\hat{f}\|_{L^2}^2 + \frac{1}{2\pi} \|v\hat{f}\|_{L^2}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\mathbf{u}\|^2 |\hat{\psi}(\mathbf{u})|^2 |\hat{\rho}_\sigma(\mathbf{u})|^2 |\hat{g}_\alpha(\mathbf{u})|^2 d\mathbf{u}.$$

□

3.4.1 Rotationally symmetric point spread function

In this subsection we consider the rotationally symmetric point spread function, i.e. $\varsigma = 0$ in (2.15). The three corollaries below follow directly from Property 3.4.

Corollary 3.3. *The sharpness function (3.19) can be expressed as*

$$S(\sigma) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \|\mathbf{u}\|^2 |\hat{\psi}(\mathbf{u})|^2 e^{-\sigma^{2\beta} \|\mathbf{u}\|^{2\beta}} e^{-\|\mathbf{u}\|^2 \alpha^2} d\mathbf{u}. \quad (3.21)$$

Corollary 3.4. *The sharpness function (3.19) is smooth, and is strictly increasing for $\sigma < 0$ and strictly decreasing for $\sigma > 0$.*

Corollary 3.5. *The sharpness function (3.19) has a finite maximum at $\sigma = 0$ for $\alpha > 0$; in particular*

$$\max_{\sigma} S(\sigma) = S(0).$$

It follows that the basic properties of the derivative-based sharpness function in two-dimension are similar to the properties in one-dimension, if we assume a rotationally symmetric point spread function: for the noise-free image formation the function has a unique optimum at the in-focus image. Further we consider the Gaussian point spread function ($\beta = 1$ in (3.21)).

Property 3.5. *The sharpness function S can be expressed by means of the autocorrelation coefficients (2.10) as follows*

$$S(\sigma) = \frac{1}{8\sqrt{\pi}(\alpha^2 + \sigma^2)^2} \sum_{n,m} \rho_{n,m} \left(4 - \frac{\|\mathbf{n}\|^2 \tau^2}{\alpha^2 + \sigma^2}\right) e^{-\frac{1}{4} \frac{\|\mathbf{n}\|^2 \tau^2}{\alpha^2 + \sigma^2}}. \quad (3.22)$$

Proof. The proof is the analogue of the proof for Property 3.3. \square

Theorem 3.4. *The sharpness function can be expressed as*

$$S(\sigma) = \frac{C_4}{2\pi(\alpha^2 + \sigma^2)^2} (1 + R_4(\sigma)), \quad (3.23)$$

where

$$|R_4(\sigma)| \leq K_4 \frac{\tau}{\sqrt{\alpha^2 + \sigma^2}}, \quad (3.24)$$

and C_4, K_4 depend only on the object pixel values, i.e. $a_{k,l}$ in (2.8).

Proof. The proof is the analogue to the proof of Theorem 3.1 with

$$C_4 = \iint_{-\infty}^{\infty} |\mathbf{u}|^2 e^{-|\mathbf{u}|^2} d\mathbf{u} \sum_{n,m} \rho_{n,m} = \pi \sum_{n,m} \rho_{n,m}.$$

and

$$K_4 = \frac{3}{2\sqrt{\pi}} \frac{\sum_{n,m} (|n| + |m|) \rho_{n,m}}{\sum_{n,m} \rho_{n,m}}.$$

\square

It follows from (3.23) that the function $S^{-1/2}$ can be approximated with any accuracy by a quadratic polynomial by increasing the value of the control variable α . This also corresponds to the findings made for the one-dimensional setting before. The only difference is the power of the sharpness function to be taken for a quadratic approximation. Below we examine a more general case of a non-symmetric point spread function.

3.4.2 Non-symmetric point spread function

Property 3.6. *For the sharpness function value, the point spread function rotation is equivalent to the object rotation*

$$S[\psi * (\mathcal{R}_\theta \varrho_\sigma) * g_\alpha] = S[(\mathcal{R}_\theta \psi) * \varrho_\sigma * g_\alpha].$$

Proof. It follows from (2.16) that

$$\begin{aligned} S[\psi * \mathcal{R}_\theta \varrho_\sigma * g_\alpha] &= \frac{1}{2\pi} \iint_{-\infty}^{\infty} |\mathbf{u}|^2 |\hat{\psi}|^2 |\mathcal{R}_\theta \hat{\varrho}_\sigma|^2 |\hat{g}_\alpha|^2 d\mathbf{u} = \\ &= \frac{1}{2\pi \det \mathcal{R}_\theta} \iint_{-\infty}^{\infty} |\mathcal{R}_\theta^{-T} \mathbf{u}|^2 |\mathcal{R}_\theta^{-T} \hat{\psi}|^2 |\hat{\varrho}_\sigma|^2 |\mathcal{R}_\theta^{-T} \hat{g}_\alpha|^2 d\mathbf{u} = \\ &= \frac{1}{2\pi} \iint_{-\infty}^{\infty} |\mathbf{u}|^2 |\mathcal{R}_\theta \hat{\psi}|^2 |\hat{\varrho}_\sigma|^2 |\hat{g}_\alpha|^2 d\mathbf{u} = S[(\mathcal{R}_\theta \psi) * \varrho_\sigma * g_\alpha]. \end{aligned}$$

□

Corollary 3.6. *For a rotationally invariant object ($\mathcal{R}_\theta \psi = \psi$), the point spread function rotation does not influence the sharpness function.*

Such objects often occur in practice. The benchmark (2.11) satisfies this property as well. For further simplification of our analysis we make therefore an assumption $\theta = 0$ in (3.20). In this case the adjustment of the y -stigmator d_y is not necessary. Neglecting the point spread function rotation angle does not limit the theoretical observations. However, in real-world applications defocus and astigmatism correction still remain a three-parameter problem. It has not been possible so far to implement point spread function rotation directly in the hardware; thus its elliptic form can be adjusted only by a combination of the two stigmator control variables.

Property 3.7. *For the object function (2.11) and the Gaussian point spread function the sharpness function (3.19) is given by*

$$S(\sigma) = \frac{C(\varsigma^2 + \sigma^2 + \alpha^2 + \gamma^2)}{2\left((\varsigma^2 + \sigma^2 + \alpha^2 + \gamma^2)^2 - 4\varsigma^2\sigma^2\right)^{3/2}}. \quad (3.25)$$

Proof. By definition

$$\left\| \frac{\partial}{\partial x} f \right\|_{L_2}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} u^2 e^{-((\varsigma - \sigma)^2 + \alpha^2 + \gamma^2)u^2} du \int_{-\infty}^{\infty} e^{-((\varsigma + \sigma)^2 + \alpha^2 + \gamma^2)v^2} dv,$$

or

$$S_{1,0}(\sigma) = \frac{C}{4} \left((\varsigma - \sigma)^2 + \alpha^2 + \gamma^2 \right)^{-3/2} \left((\varsigma + \sigma)^2 + \alpha^2 + \gamma^2 \right)^{-1/2}.$$

Similarly we compute $\left\| \frac{\partial}{\partial y} f \right\|_{L_2}^2$. Then the statement of the property is straightforward. □

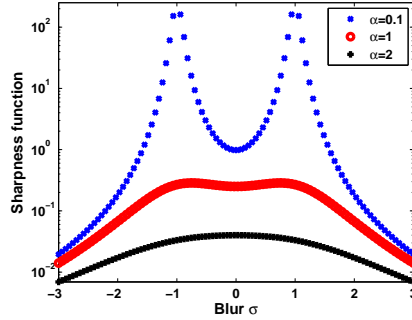


FIGURE 3.4: Sharpness functions S shape for a through-focus series with non-symmetric point spread function. When the value of the blur α increases the local optima disappear.

By analysing the derivative of the sharpness function (3.25)

$$S'(\sigma) = \frac{2C\sigma \left((\alpha^2 + \gamma^2 - \sigma^2) + \zeta^2(2\zeta^2 + \sigma^2 - \alpha^2 - \gamma^2) \right)}{\left((\zeta^2 + \sigma^2 + \alpha^2 + \gamma^2)^2 - 4\zeta^2\sigma^2 \right)^{3/2}}$$

we find that for $\sqrt{\alpha^2 + \gamma^2} < \sqrt{2}\zeta$ the sharpness function has three optima: a minimum at $\sigma_0 = 0$ and a maximum at σ_1 and σ_2 , where

$$\sigma_{1,2} = \pm \frac{1}{\sqrt{2}} \sqrt{\zeta \sqrt{8\alpha^2 + 8\gamma^2 + 9\zeta^2} - 2\alpha^2 - 2\gamma^2 - \zeta^2}.$$

If $\sqrt{\alpha^2 + \gamma^2} \geq \sqrt{2}\zeta$ the sharpness function has a maximum at $\sigma = 0$ and does not have any other optima. Figure 3.4 shows functions (3.25) computed for $\zeta = 1, \gamma = 0$ for different values of α . For a small value ($\alpha = 0.1$) the function has two local maxima. For a larger value ($\alpha = 1$) the distance between the optima decreases, and their amplitudes are smaller. In both cases the sharpness function has a minimum instead of a maximum at the in-focus position ($\sigma = 0$). For $\alpha = 2 > \sqrt{2}\zeta$ the function has a unique optimum at $\sigma = 0$. This benchmark example is important, because it shows that due to the presence of astigmatism a standard autofocus procedure might fail. However, the proper choice of the artificial blur control variable α might help to deal with it.

Property 3.8. *For the object function (2.11) and the Gaussian point spread function the sharpness function in the two-parameter space $S = S(\sigma)$ has a maximum at $\sigma = (0, 0)^T$ and does not have any other optimum for any value of the artificial blur α .*

Proof. It is straightforward that partial derivatives of the function (3.25)

$$\frac{\partial}{\partial \sigma} S(0, 0) = \frac{\partial}{\partial \zeta} S(0, 0) = 0.$$

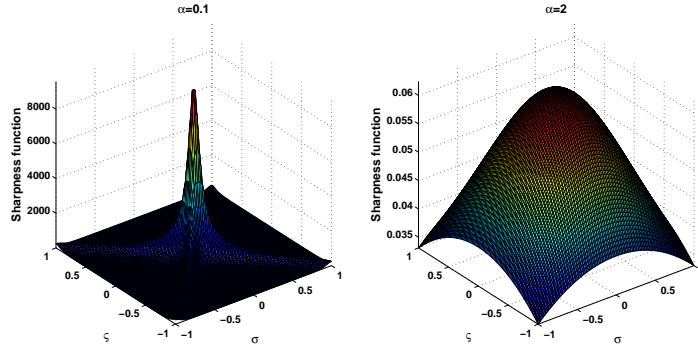


FIGURE 3.5: Sharpness functions S shape in a two-parameter space.

Further it is clear that for $\zeta \geq 0$

$$\frac{\partial}{\partial \sigma} S(\sigma_k, \zeta) \neq 0, \quad \frac{\partial}{\partial \zeta} S(\sigma_k, \zeta) \neq 0, \quad k = 1, 2.$$

□

Figure 3.5 shows the sharpness function shape in a two-parameter space computed for $\alpha = 0.1$ and $\alpha = 2$. In both cases the sharpness function has a maximum at $\sigma = (0, 0)^T$ and does not have any other optima. This is convenient for simultaneous defocus and astigmatism correction, which could be done by optimizing the sharpness function in two-parameter space [68, 60]: the local optima that the sharpness function obtains in one-dimension are not optima anymore in higher dimensions. Still, tuning the artificial blur α makes the shape of the sharpness function closer to convex, which might increase the speed of optimization.

The corollary below follows directly from Property 3.4.

Corollary 3.7. *For the benchmark object (2.11) and the symmetric Gaussian point spread function ($\zeta = 0$) the sharpness function (3.19) is given by*

$$S(\sigma) = \frac{C}{2(\sigma^2 + \alpha^2 + \gamma^2)^2}. \quad (3.26)$$

In this case the sharpness function to the power $-1/2$ is a quadratic polynomial

$$S^{-1/2}(d - d_0) = \sqrt{\frac{\pi}{C}}((d - d_0)^2 + \alpha^2 + \gamma^2).$$

3.5 Discretization

Using discrete integration (2.30) and discrete differentiation (2.31) for the image matrix (2.28), we trivially obtain a discrete version of the sharpness function

(3.1) for $n = 1$

$$S_x \doteq s_x^{\text{der}} := \frac{1}{N^{2+p}k^p} \sum_{i,j} |f_{i,j} - f_{i,j+k}|^p, \quad p = 1, 2, \quad k \in \mathbb{N}. \quad (3.27)$$

where k (the *pixel difference*) adjusts the sensitivity of the sharpness function to the noisy images. It is clear that for $n = 2$ in (3.27) larger differences between pixels are stronger weighted than smaller ones. This leads to the suppression of the contribution made by noise [88]. To improve the robustness to noise a threshold Θ is often applied to the difference between pixels, which is taken into account [39]

$$s_{x,\Theta}^{\text{der}} := \frac{1}{N^{2+n}k^n} \sum_{i,j} |f_{i,j} - f_{i,j+k}|^n, \quad |f_{i,j} - f_{i,j+k}|^n > \Theta, \quad \Theta > 0. \quad (3.28)$$

The threshold Θ is determined experimentally [88]. In SEM and STEM often the difference between only the pixels in horizontal direction is taken into account, because the SEM scanning is performed in horizontal direction and therefore noise is correlated there. This sharpness function can fail for certain image geometries (for example, a number or uniform horizontal stripes). Let $s_{y,\Theta}^{\text{der}}$ be the function that computes the norm of the pixel difference in vertical direction. Then the form that generalizes derivative-based sharpness function is

$$s_{\Theta}^{\text{der,c}} := s_{x,\Theta}^{\text{der}} + \nu s_{y,\Theta}^{\text{der}}, \quad \nu = \{0, 1\}. \quad (3.29)$$

Usually in applications only pixel difference parameter values $k = 1, 2$ are used [6], [69]. In Chapter 6 we experimentally show that the larger values of k often provide better results.

If we consider derivative interpolation by a convolution with a Gaussian derivative kernel (2.32), we obtain

$$s_{\Theta}^{\text{der,c}} = \sum_{i,j} ((\mathbf{F} * \mathbf{G}_1)_{i,j}^2 + (\mathbf{F} * \mathbf{G}_2)_{i,j}^2), \quad ((\mathbf{F} * \mathbf{G}_1)_{i,j}^2 + (\mathbf{F} * \mathbf{G}_2)_{i,j}^2) > \Theta, \quad (3.30)$$

where the Gaussian derivative kernels $\mathbf{G}_1, \mathbf{G}_2$ could be for instance defined as

$$\mathbf{G}_1 = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \quad \mathbf{G}_2 = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}. \quad (3.31)$$

The form of Gaussian kernels (3.31) is known in application literature as *Sobel operators* [69].

3.6 The fast autofocus algorithm

As mentioned in Section 1.2 the image recording in electron microscopy might require a noticeable amount of time, thus the function evaluations in our problem are very expensive. Quadratic interpolation is therefore a convenient approach for computing a quadratic polynomial approximation of the sharpness

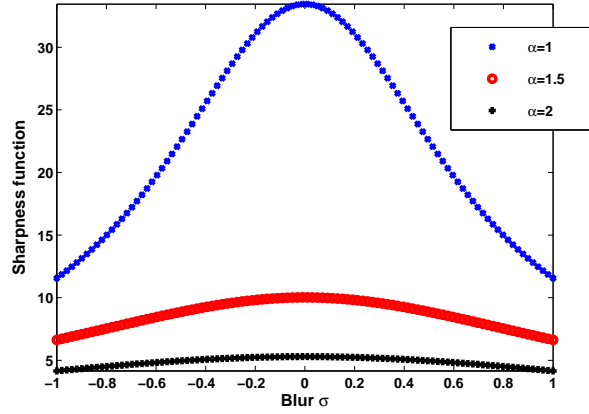


FIGURE 3.6: Sharpness function S computed for different values of the blur α .

function. In our autofocus method we take the minimum of the polynomial as the minimum of the sharpness function. For the given data points $S_k := S(d_k)$, $k = 1, 2, 3$ we interpolate the function $S^{-1/2}$ by a polynomial $\mathcal{P}(d) := c_0 + c_1d + c_2d^2$. So one has

$$S^{-1/2}(d) = \mathcal{P}(d)(1 + \varepsilon(d)),$$

where $P(d_k) = S_k^{-1/2}$, $k = 1, 2, 3$.

From Theorem 3.4 we conclude that the error $\varepsilon(d)$ can be decreased by increasing α . Theoretically the error of this approximation can be made as small as needed by dramatically increasing the value α . However, if $\alpha \rightarrow \infty$ then $S(d) \rightarrow 0$ and all its derivatives, which may cause numerical errors and can make it difficult to find the optimum of the function. Figure 3.6 shows three sharpness functions computed for different α -values. In the next section it will be shown how large values of α influence the shape of the sharpness function computed for experimental *through-focus series*.

The above observations lead to the following algorithm.

Algorithm 3.1. *Autofocus*

1. Let d_2 be the current defocus control value of the optical device. Choose a Δd , then $d_1 := d_2 - \Delta d$, $d_3 := d_2 + \Delta d$.
2. Record three images at d_1, d_2, d_3 and compute $S_1^{-1/2}, S_2^{-1/2}, S_3^{-1/2}$.
3. Interpolate three points with a quadratic polynomial. Estimate the sharpness function optimum

$$d_{\text{opt}} = -\frac{c_1}{2c_2}$$

as the optimum of the polynomial.

TABLE 3.1: Overview of carbon cross grating experimental through-focus series.

\mathcal{N}	Magnification	Pixel width τ [nm]	Defocus range $(d_N - d_1)$ [nm]	Defocus step Δd [nm]	Image number N
1	10 000 \times	42	36000	2000	19
2	10 000 \times	42	10000	500	21
3	200 000 \times	2.1	20000	1000	21
4	200 000 \times	2.1	10000	500	21
5	400 000 \times	1.05	900	50	19

The main goal of the autofocus method described in this chapter is to try to estimate the in-focus image position from three or four recorded images. For a more precise autofocus an iterative algorithm could be used (see Appendix A). The convergence properties of such an algorithm should be a topic of the future research.

3.7 Numerical experiments with STEM images

Ten experimental *through-focus series* are obtained with the FEI STEM microscope. Two different samples are used: a *carbon cross grating* sample and a *gold particles* sample. Carbon cross grating is the standard sample for STEM calibration. The gold particle sample is a typical image example, which could be used for investigation by *particle analysis algorithms*. The size of each image in the series is 512×512 pixels. The series are recorded at different magnifications and with different defocus steps. Figures 3.7-3.8 show the first image in the series, the in-focus image, and the computed sharpness function values plotted versus the values of defocus control. Every figure represents five series, described in the tables 3.1-3.2 (carbon cross grating sample and gold particles sample respectively). The line numbers in the tables $\mathcal{N}=1,2,3,4,5$ correspond to the columns of figures 3.7-3.8 (from left to right). For each series two functions are computed: with $\alpha = 0$ (dotted line) and with $\alpha > 0$ (dashed line). The values of both functions are scaled between 0 and 1. Computed derivative-based sharpness functions with $\alpha > 0$ can accurately be approximated by a quadratic polynomial.

The series shown in the second columns of figures 3.7-3.8 are recorded with a small defocus step. The qualities of the first image in the series and the in-focus image do not differ so much: we can see the details on the first images from the series, only the edges are a bit unsharp. It is shown in tables 3.1-3.2 ($\mathcal{N}=2$) that these series have relatively small defocus ranges and defocus steps for particular magnification. For these cases the sharpness function has a shape nearly quadratic even with $\alpha = 0$, as follows from Theorem 3.2. The sharpness function shape is different in a broader defocus range for the same sample at the same magnification (Figure 3.7, first column and Figure 3.8, first

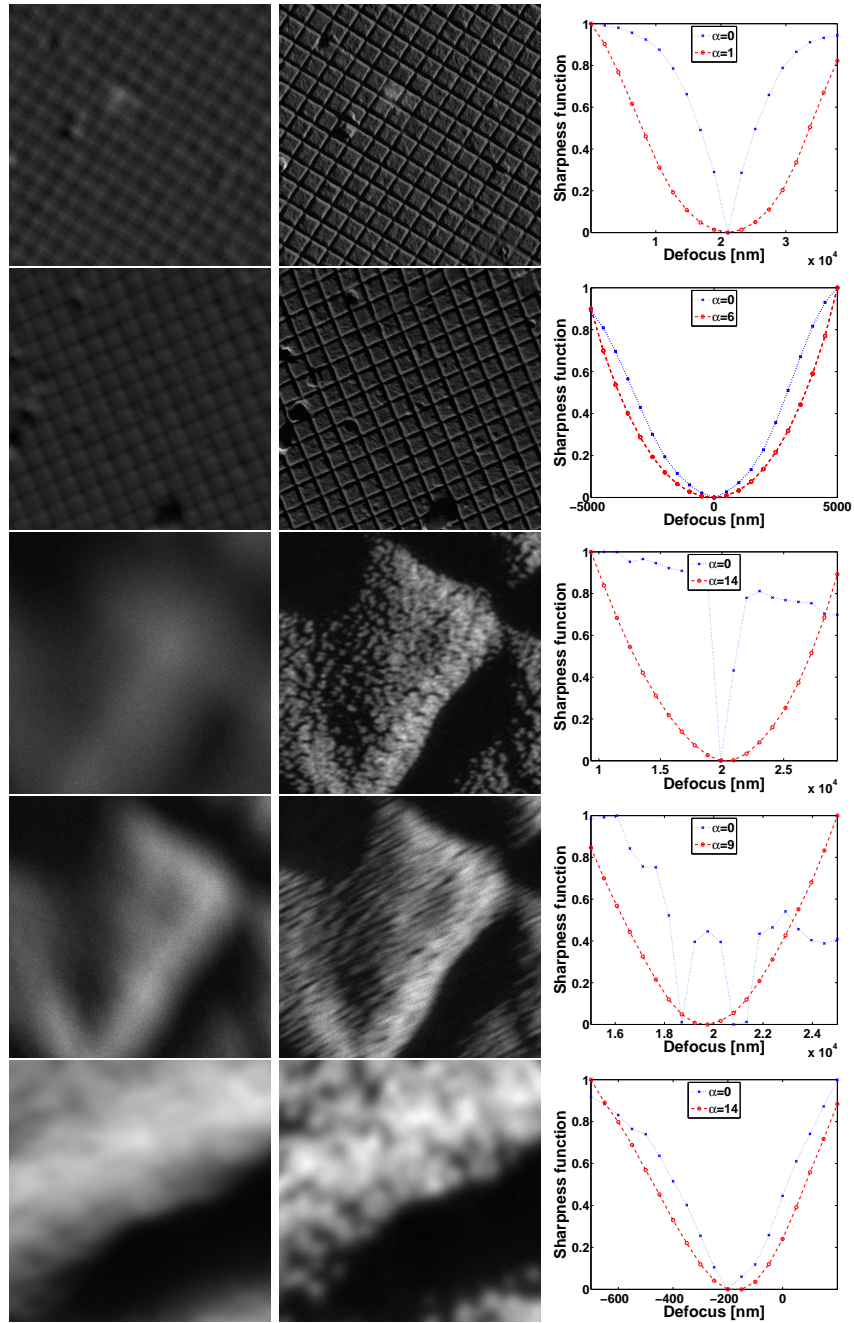


FIGURE 3.7: Sharpness functions $S^{-1/2}$ computed for experimental STEM through-focus series of a *carbon cross grating* sample. From left to right: the first image in the series, in-focus image from the series, sharpness functions with and without artificial blur plotted versus defocus. From top to bottom: five different experimental through-focus series.

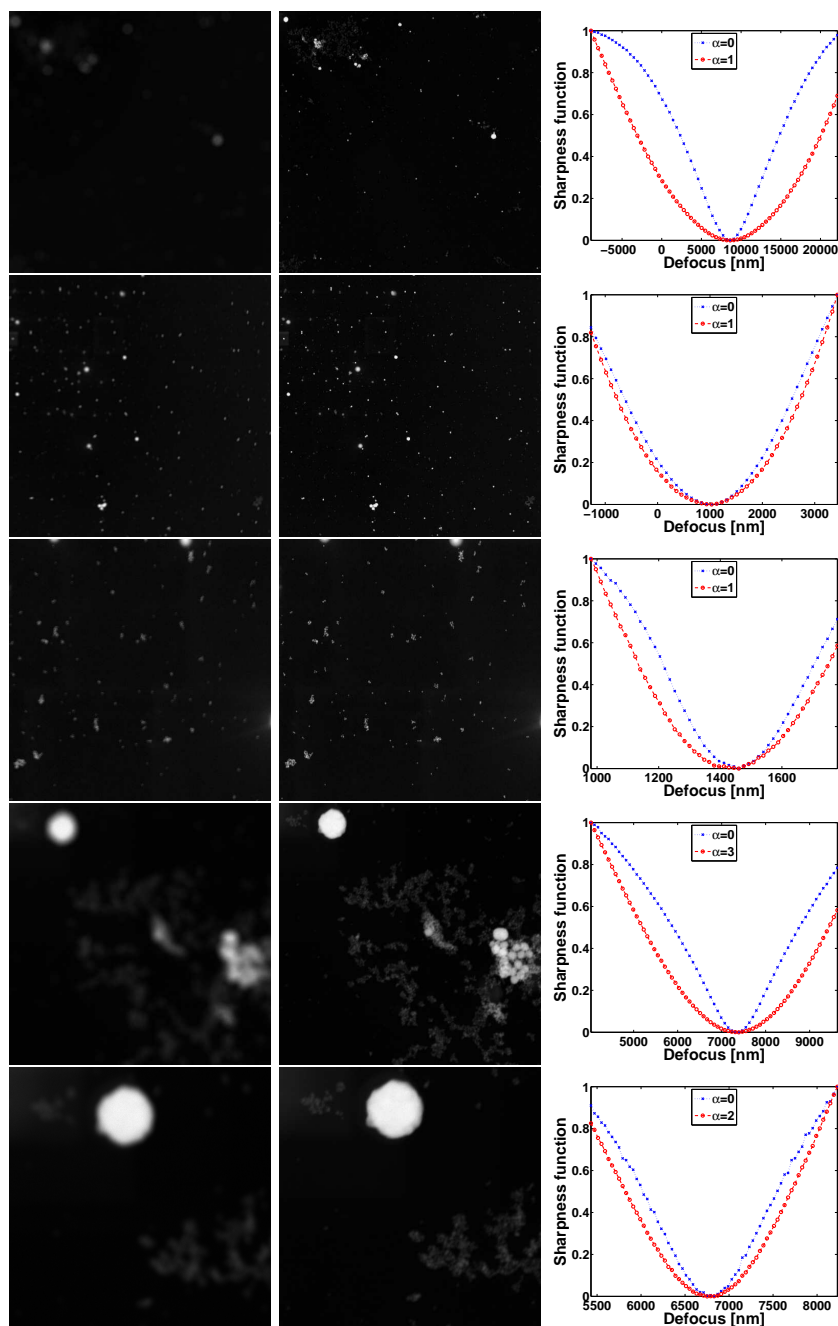
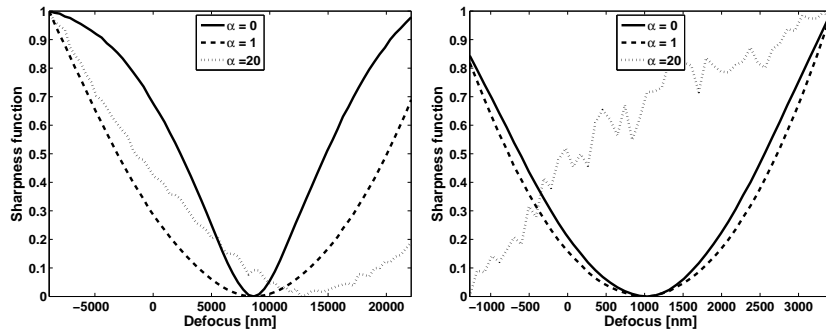


FIGURE 3.8: Sharpness functions $S^{-1/2}$ computed for experimental STEM through-focus series of a *gold particles* sample. From left to right: the first image in the series, in-focus image from the series, sharpness functions with and without artificial blur plotted versus defocus. From top to bottom: five different experimental through-focus series.

TABLE 3.2: Overview of gold particles experimental through-focus series.

\mathcal{N}	Magnification	Pixel width τ [nm]	Defocus range ($d_N - d_1$) [nm]	Defocus step Δd [nm]	Image number N
1	10 000 \times	42	31500	450	70
2	10 000 \times	42	4704	96	50
3	56 000 \times	7.5	800	16	51
4	56 000 \times	7.5	5600	80	71
5	115 000 \times	3.75	2800	40	71

FIGURE 3.9: Sharpness function $S^{-1/2}$ computed for different values of the blur α for experimental through-focus series of gold particles ($\mathcal{N}=1,2$), Table 3.2.

column). The functions with $\alpha = 0$ have shapes similar to a Gaussian, but not a quadratic polynomial. In this case the functions have a nearly quadratic shape after applying the blur α to the images.

The fourth series of carbon cross grating (Figure 3.7, fourth column) is the only experimental series recorded with the presence of astigmatism. For other experimental series astigmatism of the magnetic lens is corrected before the recording. The lens with astigmatism is not perfectly symmetric and as a consequence has more than one focal point [53], which results in the asymmetry of the point spread function. Consequently the recorded image cannot be totally sharp. The sharpness function might have local optima due to the presence of astigmatism [16, 68]. Two local minima can be seen in the plot. They disappear after applying the artificial blur.

In the last experiments (figures 3.7-3.8, fifth column) the magnification of the microscope is higher and as a consequence the influence of noise on the image quality increases. We can see that in these cases the blur α helps to deal with noise in the sharpness function.

Figure 3.9 shows the sharpness function $S^{-1/2}$ computed for different values of the blur α for experimental through-focus series of gold particles ($\mathcal{N}=1,2$). For the large $\alpha = 20$ the function becomes noisy and does not provide useful information anymore. It follows that for the proper performance of the method we have to make a proper choice of the value α that should not be too large

neither too small. This choice might depend on the sample geometry as well as on the defocus range (how far away we are actually from the ideal d_0). The choice does not have to be made every time, but once for a particular application, where we deal with the class of geometrical objects.

Chapter 4

Fourier transform-based approach

The Fourier transform is an important mathematical tool for image quality improvement in electron microscopy, as well as in other types of optical devices, such as telescopes, ophthalmoscopes and endoscopes [28]. It is widely used for so-called blind deconvolution procedures [8, 85], for automated defocus and astigmatism correction in electron microscopy [4, 22, 33, 52, 77, 87, 95, 96]. To this end, one needs on one hand a careful analysis of the Fourier transform, on the other hand the analysis must be fast. Even if the image f is real-valued its Fourier transform \hat{f} is a complex function in general. The principal method for visualization is computing the *power spectrum* $|\hat{f}|^2$, i.e. the square of the magnitude of the image Fourier transform. The power spectrum is for instance used as a visual support for non-automated corrections performed by an experienced human operator.

We have already shown in the previous chapter that the image power spectrum has a direct relation with the derivative-based sharpness function. Though the image Fourier transform was considered only for analytical computations. In this chapter we introduce a semi-automated method for defocus and astigmatism correction based on the Fourier transform, where the numerical computations of the *discrete Fourier transform* are involved. Three operators $\mathcal{H}_q, q = 0, 1, 2$ corresponding to the three control variables d, d_x, d_y (*sharpness operator, x-stigmatic operator, y-stigmatic operator*) are chosen to simplify the defocus and astigmatism correction for a non-experienced human operator. The power spectrum's *mathematical moments* are expressed in terms of these operators [65]. The three real-valued functions $S_q^{\text{ft}}, q = 0, 1, 2$ (*sharpness function, x-stigmatic function, y-stigmatic function*) derived from sharpness/stigmatic operators are used as a basis for an automated application. In the largest part of this chapter we use the notation S_q instead of S_q^{ft} for the Fourier transform-based sharpness, x-stigmatic and y-stigmatic functions.

In Section 4.1 the general properties of the discrete Fourier transform are introduced. Section 4.2 discusses the relation of a power spectrum orientation with defocus and astigmatism based on the practical example with the Gaussian point spread function. Section 4.3 defines the typical power spectrum benchmarks that we are going to use further. Section 4.4 explains the method of orientation identification, which involves computing sharpness/stigmatic operators as well as computing sharpness/stigmatic functions. Section 4.6 connects

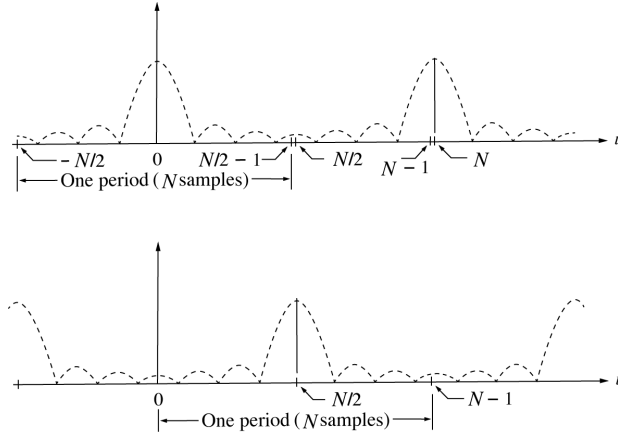


FIGURE 4.1: Power spectrum in one-dimension before and after the shift.

the Fourier transform-based sharpness/stigmatic functions with the derivative-based sharpness function studied in the previous chapter. In Section 4.7 we explain the way of computing the sharpness/stigmatic operators and functions for the discrete power spectrum. Section 4.8 illustrates results of numerical experiments with Gaussian benchmarks and SEM experimental images.

4.1 Discrete Fourier transform

The components of the *discrete Fourier transform* of the discrete image (2.28) are defined as, cf. [19, 42, 43]

$$\hat{f}_{k,l} := \sum_{i,j=0}^{N-1} f_{i,j} e^{-2\pi i \frac{ik+jl}{N}}, \quad (4.1)$$

with the inverse discrete Fourier transform components

$$f_{i,j} := \frac{1}{N^2} \sum_{k,l=0}^{N-1} \hat{f}_{k,l} e^{2\pi i \frac{ik+jl}{N}}. \quad (4.2)$$

By letting $i, j, k, l = -\infty, \dots, \infty$, by direct substitution in (4.1) and (4.2) it can be shown that the discrete Fourier transform and its inverse components are periodic in both directions, i.e.

$$\hat{f}_{k,l} = \hat{f}_{k+N,l} = \hat{f}_{k,l+N} = \hat{f}_{k+N,l+N}, \quad f_{i,j} = f_{i+N,j} = f_{i,j+N} = f_{i+N,j+N}.$$

This periodicity is just a mathematical property of the discrete Fourier transforms. In general the image f is not periodic. Usually only the values of $i, j, k, l = 0, \dots, N-1$ are considered. The origin of the Fourier transform

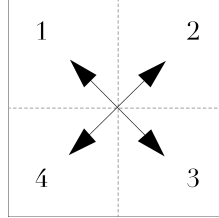


FIGURE 4.2: The shift of the Fourier transform two-dimensions.

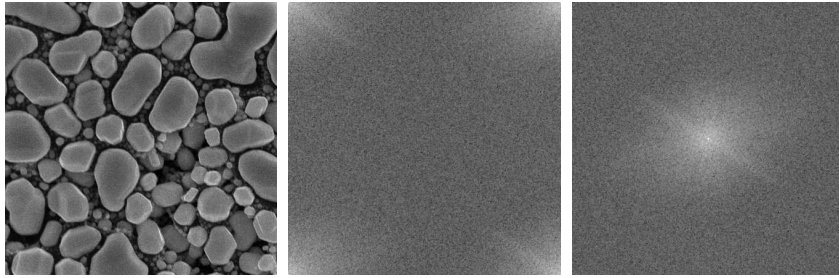


FIGURE 4.3: SEM experimental image, its power spectrum before and after the shift.

(the zero-frequency component) is equal to the sum of the entries in the image $\hat{f}_{0,0} = \sum_{i,j} f_{i,j}$.

Figure 4.1 shows a power spectrum in one-dimension. The *shift operator* moves the zero-frequency component to the center of the computed array. It is used for visualizing the Fourier transformed functions with the zero-frequency component in the middle of the spectrum [44]. Figure 4.2 shows schematically the shift operator for the two-dimensional case. Figure 4.3 shows a SEM experimental image, and its power spectra before and after applying the shift operator. If not mentioned otherwise we visualize power spectra after the shift.

We can premultiply the images by a *window function*, for instance by

$$W(\mathbf{u}) := \begin{cases} (1 + \cos(\pi \frac{\|\mathbf{u}\|}{U}))/2, & \text{if } \|\mathbf{u}\| \leq U, \\ 0, & \text{otherwise,} \end{cases} \quad (4.3)$$

where U is the support size of the Fourier transform. Figure 4.4 shows the difference between the power spectra of the non-windowed image and the windowed image. The non-windowed power spectrum (Figure 4.4(c)) has two orthogonal lines crossing the origin. They are missing in the windowed power spectrum (Figure 4.4(d)). The lines in the non-windowed power spectrum (Figure 4.4(c)) are results of the image discontinuity on the boundary. They would reduce the efficiency of the method described below. We will apply the window function to all images described below, before computing their Fourier transforms.

The power spectrum's entries vary quite nonlinearly: low frequencies (pixels

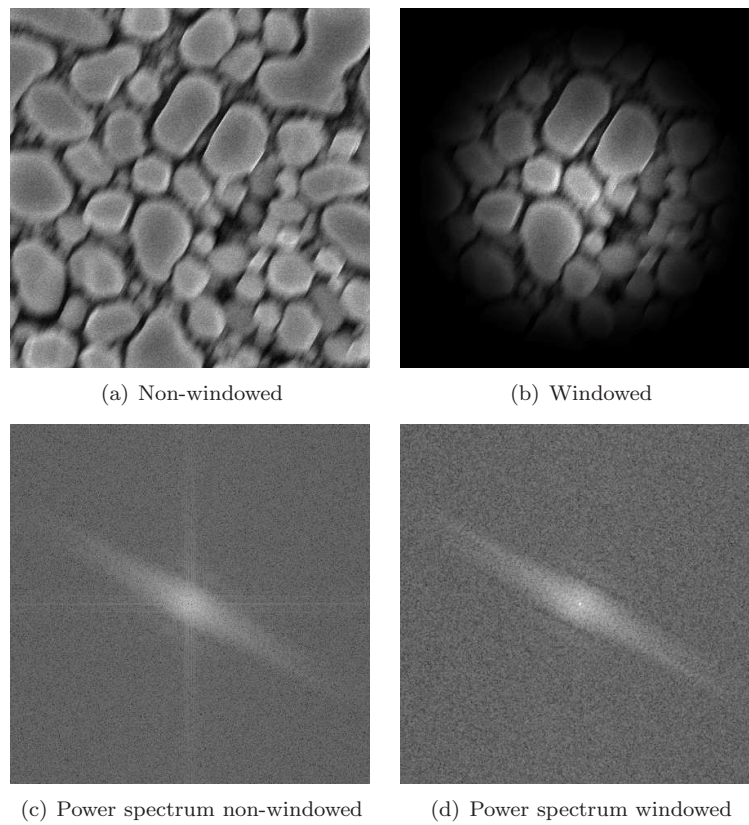


FIGURE 4.4: 4.4(a) A SEM experimental image of gold-on-carbon (a non-windowed image); 4.4(b) the experimental image multiplied by the window function (a windowed image); 4.4(c) power spectrum of the experimental non-windowed image; 4.4(d) power spectrum of the experimental windowed image.

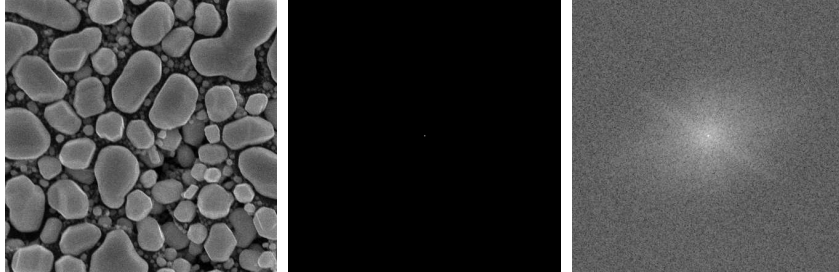


FIGURE 4.5: SEM experimental image and its power spectrum before and after logarithmic scale.

with indices close to $(i, j) = (0, 0)$ have much larger values than high frequencies. Therefore, following a suggestion from [30], we use a logarithmic scale for its visualization

$$\check{p} := \log(C + |\hat{f}|^2). \quad (4.4)$$

To ensure that the values of the logarithmic scale are positive the value of $C = 1$ is usually applied [19]. Figure 4.5 shows the SEM experimental image and its power spectrum before and after logarithmic scaling. All the figures in this chapter related to the experiments with the real-world data show the logarithmic scale (4.4) of power spectra.

4.2 Power spectrum orientation

We illustrate the relation between the power spectrum orientation and defocus and astigmatism with the help of three numerical (synthetic) examples. The left column of figures 4.6-4.8 shows identical experimental SEM image (we consider it as an object function ψ) of tin balls and its power spectrum, which is nearly rotationally symmetric. The object ψ is convolved with a Gaussian point spread functions with parameters (2.15) $\sigma = 1, \varsigma = 0$ (Figure 4.6), $\sigma = 5, \varsigma = 0$ (Figure 4.7), $\sigma = 4, \varsigma = 1, \theta = 0$ (Figure 4.8). The top row of each figure contains the spatial representation, and the bottom row shows the frequency representation, where *convolution* becomes multiplication. The influence of the point spread function's parameters is visible in the result of the convolution in the spatial representation (image f - the top right of each of the figures), as well as in the frequency representation.

For a rotationally symmetric point spread function with a relatively small width (Figure 4.6) the final image f resembles its original ψ . Its rotationally symmetric power spectrum has a relatively large width. For a rotationally symmetric point spread function with a larger width (Figure 4.7), its power spectrum has a smaller width, as well as the power spectrum of f (power spectrum's intensity decreases), and the image f itself looks more blurred than in Figure

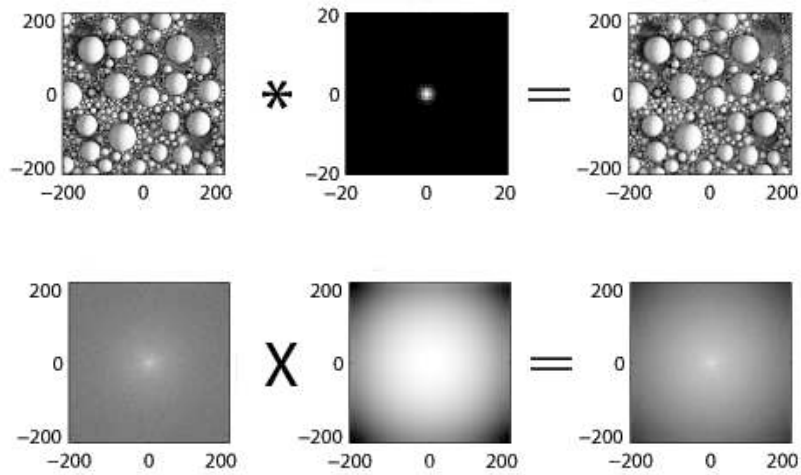


FIGURE 4.6: The top row represents real space; from left to right: experimental SEM image, which is considered to be the object function ψ , Gaussian point spread function g ($\sigma = 1, \zeta = 0$), numerical result of their convolution f . The bottom row shows them in the Fourier space. PS denotes the Power Spectrum. PSF denotes the point spread function.

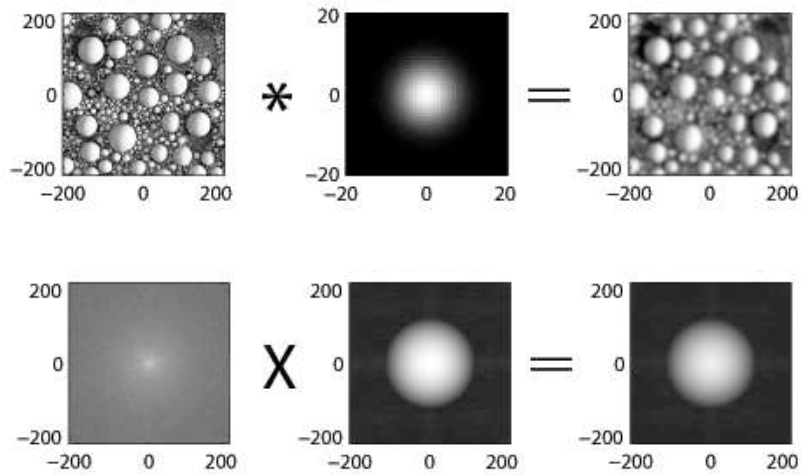


FIGURE 4.7: Similar to Figure 4.6 numerical computation, but with the Gaussian point spread function parameters $\sigma = 5, \zeta = 0$. PS denotes the Power Spectrum. PSF denotes the point spread function.

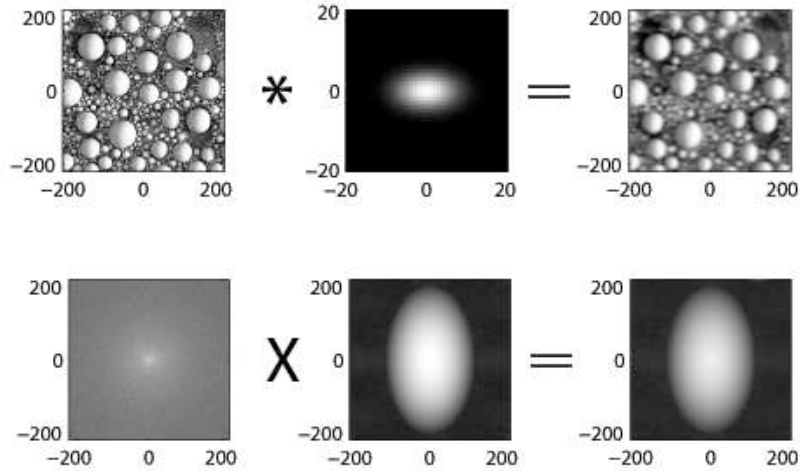


FIGURE 4.8: Similar to Figure 4.6 numerical computation, but with the Gaussian point spread function parameters $\sigma = 4, \zeta = 1, \theta = 0$. PS denotes the Power Spectrum. PSF denotes the point spread function.

4.6. Finally for a non-symmetric point spread function, the power spectrum of f becomes non-symmetric as well (Figure 4.8).

The above presented characteristics of the power spectrum are used by a human operator for defocus and astigmatism corrections. For instance for an *amorphous* object (which has a rotationally symmetric power spectrum) the operator tries to obtain an image f with a rotationally symmetric power spectrum as intense as possible. If the power spectrum of an object ψ is not rotationally symmetric (the object has a strong preferential direction) the situation is more complex. In this case it is important to compare the difference between power spectra obtained for different values of the controls \mathbf{d} and to find a center of symmetry [52].

Looking at the power spectrum's shape, it is often difficult for a non-experienced human operator to understand which of the controls \mathbf{d} has to be adjusted in which direction. In the next section we provide a method to assist a human operator to correct the image. In fact this method can be used as a basis for automated defocus and astigmatism correction.

4.3 Power spectrum benchmarks

Power spectra of images recorded by a number of optical devices including electron microscopes can accurately be approximated by a Lévi stable density

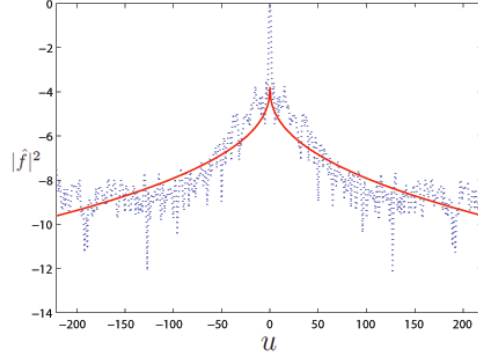


FIGURE 4.9: One-dimensional cut of the power spectrum of the SEM experimental image of tin balls (dotted line), its approximation with the Lévi stable density function (solid line). The data is plotted with the logarithmic scale.

function [7, 8]

$$|\hat{f}|^2 = A \cdot \left(|\hat{g}_1|^2 \circ \mathcal{J}_{\mathbf{a}} \circ \mathcal{R}_{\zeta} \right), \quad (4.5)$$

where the operators $\mathcal{J}_{\mathbf{a}}, \mathcal{R}_{\zeta}$ are defined in (2.1), (2.3), $\mathbf{a} := (a, b)^T \in \mathbb{R}^2$ and A is a constant. The function \hat{g}_1 is defined in (2.14), where it depends on the Lévi stable density parameter β . Figure 4.9 shows a one-dimensional cut of the power spectrum of the SEM experimental image of tin balls, used in the previous section as the object function (dotted line). Its approximation with the Lévi stable density function is plotted as a solid line. The data are on logarithmic scale. If $\beta = 1$ in (2.14), the power spectrum is a Gaussian function.

Below we will use polar coordinates in frequency representations

$$(r, \varphi) \in \{\mathbb{R}^+ \times [0, 2\pi)\}, \quad u = r \cos \varphi, \quad v = r \sin \varphi.$$

Property 4.1. *For the linear image formation model, a Gaussian point spread function and a benchmark object (2.11), the image power spectrum is a Gaussian function.*

Proof. According to the definitions

$$\begin{aligned} |f|>^2 &= |\hat{\psi}|^2 |\hat{\varrho}_{\sigma}|^2 = C e^{-\gamma^2 \|\mathbf{u}\|^2} \mathcal{J}_{\sigma} \mathcal{R}_{\theta} e^{-\|\mathbf{u}\|^2} = \\ &C e^{-r^2 \left(((\sigma + \varsigma)^2 + \gamma^2) \cos^2(\varphi + \zeta) + ((\sigma - \varsigma)^2 + \gamma^2) \sin^2(\varphi + \zeta) \right)}. \end{aligned}$$

□

Corollary 4.1. *It follows that in the case of the Gaussian benchmark the power spectrum parameters \mathbf{a} and ζ can be expressed as*

$$a^2 = (\sigma + \varsigma)^2 + \gamma^2, \quad b^2 = (\sigma - \varsigma)^2 + \gamma^2, \quad \zeta = \theta.$$

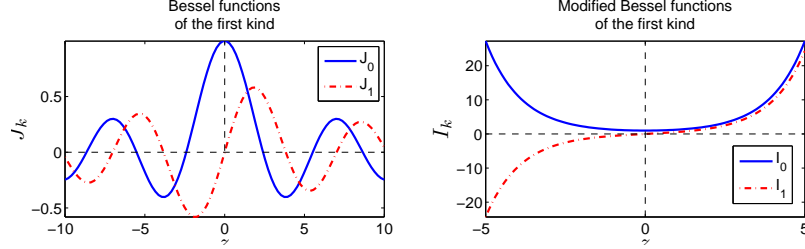


FIGURE 4.10: Bessel functions of the first kind and modified Bessel functions of the first kind for $k = 0, 1$.

4.4 Orientation identification

Based on the polar coordinates in the frequency representation $u = r \cos \varphi$, $v = r \sin \varphi$, we define three operators $\mathcal{H}_q : \mathbb{R}^2 \rightarrow \mathbb{R}^+$, $q = 0, 1, 2$ as

$$\mathcal{H}_0 f := \mathcal{H}_0[f(\mathbf{x})](r) := \oint_0^{2\pi} |\hat{f}|^2 d\varphi, \quad (4.6)$$

$$\mathcal{H}_1 f := \mathcal{H}_1[f(\mathbf{x})](r) := \oint_0^{2\pi} |\hat{f}|^2 \cos 2\varphi d\varphi, \quad (4.7)$$

$$\mathcal{H}_2 f := \mathcal{H}_2[f(\mathbf{x})](r) := \oint_0^{2\pi} |\hat{f}|^2 \sin 2\varphi d\varphi. \quad (4.8)$$

In (4.7) and (4.8) $\cos 2\varphi$ and $\sin 2\varphi$ play the roles of the weight functions. Below it will be illustrated how they help to obtain information about the directions of the *stigmators* d_x and d_y .

For the Gaussian benchmark power spectrum one can rewrite the expressions for the *sharpness operator* and *stigmatic operators* (4.6) - (4.8) as follows

$$\mathcal{H}_0 f = 2\pi e^{-\frac{1}{2}(a^2+b^2)r^2} I_0\left(\frac{1}{2}(a^2-b^2)r^2\right), \quad (4.9)$$

$$\mathcal{H}_1 f = 2\pi e^{-\frac{1}{2}(a^2+b^2)r^2} I_1\left(\frac{1}{2}(a^2-b^2)r^2\right) \cos 2\zeta, \quad (4.10)$$

$$\mathcal{H}_2 f = 2\pi e^{-\frac{1}{2}(a^2+b^2)r^2} I_1\left(\frac{1}{2}(a^2-b^2)r^2\right) \sin 2\zeta. \quad (4.11)$$

Here I_k is the *modified Bessel function of the first kind*, which can be expressed as, cf. [1]

$$I_k(z) := \frac{1}{\pi} \int_0^\pi e^{z \cos \varphi} \cos k\varphi d\varphi,$$

which is related to the *Bessel function of the first kind* as in [1]

$$I_k(z) := i^{-k} J_k(iz).$$

Figure 4.10 shows the Bessel functions of the first kind and modified Bessel functions of the first kind for $k = 0, 1$. From the representation (4.9)-(4.11) it

is clear that for the Gaussian benchmark $\mathcal{H}_0 f$ is always positive and the sign of $\mathcal{H}_1 f$, $\mathcal{H}_2 f$ is constant for $r \in \mathbb{R}^+$ and depends on the point spread function parameters in the following way

$$\begin{aligned}\operatorname{sgn}(\mathcal{H}_1 f) &= \operatorname{sgn}(a^2 - b^2) \operatorname{sgn}(\cos 2\zeta) = \operatorname{sgn}(\zeta) \operatorname{sgn}(\cos 2\theta), \\ \operatorname{sgn}(\mathcal{H}_2 f) &= \operatorname{sgn}(a^2 - b^2) \operatorname{sgn}(\sin 2\zeta) = \operatorname{sgn}(\zeta) \operatorname{sgn}(\sin 2\theta).\end{aligned}$$

As a consequence the directions of the stigmator control variables can easily be identified from the signs of stigmatic operators

$$\operatorname{sgn}(d_x - d_{x_0}) = \operatorname{sgn}\mathcal{H}_1 f, \quad \operatorname{sgn}(d_y - d_{y_0}) = \operatorname{sgn}\mathcal{H}_2 f.$$

It is important to note that for $a = b$ both stigmatic operators are uniformly zero.

Figure 4.13 shows the sharpness and stigmatic operators computed for the Gaussian benchmark. The signs of the operators suggest the stigmator controls needed for correction and their direction. A more detailed description of the figure is provided below in Subsection 4.8.1. It is important to note that one image is not enough for astigmatism correction. An elliptic power spectrum does not necessarily imply a *stigmatic image*. If we consider a Gaussian elliptic benchmark object we will observe that its power spectrum is elliptic as well, even without any astigmatism present. In order to estimate the presence of astigmatism in the system in general one needs to record a few images with different control values. This leads to the *focus series method*, explained next.

4.5 The focus series method

The operators (4.6)-(4.8) are related to the mathematical moments of the power spectrum $|\hat{f}|^2$

$$\mathcal{M}_{k,l} := \iint_{-\infty}^{+\infty} u^k v^l |\hat{f}|^2 d\mathbf{u} = \int_0^\infty r^{k+l+1} \oint_0^{2\pi} |\hat{f}|^2 \cos^k \varphi \sin^l \varphi d\varphi dr,$$

which are widely used in various applications for orientation identification and other purposes [80, 97, 99]. The 0-th moment is given by

$$\mathcal{M}_{0,0} = \int_0^\infty r(\mathcal{H}_0 f) dr. \quad (4.12)$$

The image f in this thesis is real-valued, which implies that its Fourier transform is conjugate symmetric about the origin $\hat{f}(\mathbf{u}) = \hat{f}^*(-\mathbf{u})$. Thus the power spectrum is also symmetric about the origin

$$|\hat{f}(\mathbf{u})|^2 = |\hat{f}(-\mathbf{u})|^2. \quad (4.13)$$

The symmetry of the power spectrum (4.13) implies that the 1-st moments are equal to zero, i.e. $\mathcal{M}_{1,0} = \mathcal{M}_{0,1} = 0$, or

$$\oint_0^{2\pi} |\hat{f}|^2 \cos \varphi d\varphi dr = 0, \quad \oint_0^{2\pi} |\hat{f}|^2 \sin \varphi d\varphi dr = 0.$$

The 2-nd moments are:

$$\mathcal{M}_{2,0} = \int_0^\infty r^3 \oint_0^{2\pi} |\hat{f}|^2 \cos^2 \varphi d\varphi dr = \frac{1}{2} \int_0^\infty r^3 (\mathcal{H}_0 f + \mathcal{H}_1 f) dr, \quad (4.14)$$

$$\mathcal{M}_{0,2} = \int_0^\infty r^3 \oint_0^{2\pi} |\hat{f}|^2 \sin^2 \varphi d\varphi dr = \frac{1}{2} \int_0^\infty r^3 (\mathcal{H}_0 f - \mathcal{H}_1 f) dr, \quad (4.15)$$

$$\mathcal{M}_{1,1} = \int_0^\infty r^3 \oint_0^{2\pi} |\hat{f}|^2 \cos \varphi \sin \varphi d\varphi dr = \frac{1}{2} \int_0^\infty r^3 (\mathcal{H}_2 f) d\varphi dr. \quad (4.16)$$

Let us further introduce three sharpness/stigmatic functions applied to the image f

$$S_q[f] := \int_0^\infty r^3 (\mathcal{H}_q f) dr, \quad q = 0, 1, 2. \quad (4.17)$$

Note that we have the relation to the second mathematical moments as follows

$$\mathcal{M}_{2,0} = \frac{1}{2}(S_0 + S_1), \quad \mathcal{M}_{0,2} = \frac{1}{2}(S_0 - S_1), \quad \mathcal{M}_{1,1} = \frac{1}{2}S_2.$$

In a more general case one could consider

$$S_q^{(\vartheta)}[f] := \int_0^\infty \vartheta(r) (\mathcal{H}_q f) dr, \quad q = 0, 1, 2, \quad (4.18)$$

where ϑ plays the role of a weight function. For instance the sharpness function $S_0^{(r)}$ based on the 0-th mathematical moment (4.12) is widely used for autofocus in different applications. In a generalized case it is defined (cf. [52, 87])

$$S_{0,0} := \iint_{\Omega} |\hat{f}|^2 dx = \int_{r_{min}}^{r_{max}} r \int_0^{2\pi} |\hat{f}|^2 d\varphi dr, \quad (4.19)$$

where $\Omega := \{(x, y) \in \mathbb{R}^2, r_{min}^2 < x^2 + y^2 < r_{max}^2\}$, and r_{min} and r_{max} are the low frequency band and high frequency band. In the next subsection next to the functions (4.17), which can be expressed through (4.18) as $S_q[f] = S_q^{(r^3)}[f]$, $q = 0, \dots, 2$, we will consider $S_q^{(r)}$, $q = 0, 1, 2$. In the general case ϑ is not necessarily related to the mathematical moments of the power spectrum.

4.5.1 A Lévi stable density benchmark

For the Lévi stable density power spectrum it follows that

$$S_0 = \frac{A}{4} \pi \Gamma(1 + \frac{2}{\beta}) \frac{1}{ab} \frac{a^2 + b^2}{a^2 b^2}, \quad (4.20)$$

$$S_1 = \frac{A}{4} \pi \Gamma(1 + \frac{2}{\beta}) \frac{1}{ab} \frac{b^2 - a^2}{a^2 b^2} \cos 2\zeta, \quad (4.21)$$

$$S_2 = \frac{A}{4} \pi \Gamma(1 + \frac{2}{\beta}) \frac{1}{ab} \frac{a^2 - b^2}{a^2 b^2} \sin 2\zeta, \quad (4.22)$$

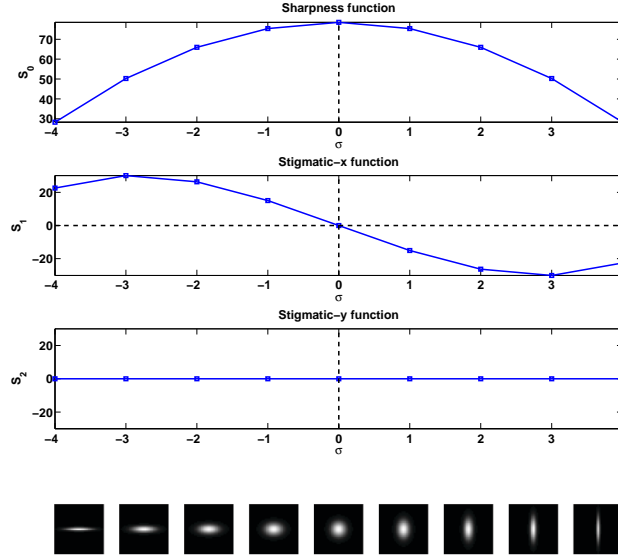


FIGURE 4.11: Sharpness and stigmatic functions for the Gaussian benchmark power spectrum.

and

$$S_0^{(r)} = A\pi\Gamma\left(1 + \frac{1}{\beta}\right)\frac{1}{ab}, \quad (4.23)$$

$$S_1^{(r)} = A\pi\Gamma\left(1 + \frac{1}{\beta}\right)\frac{1}{ab}\frac{b-a}{a+b}\cos 2\zeta, \quad (4.24)$$

$$S_2^{(r)} = A\pi\Gamma\left(1 + \frac{1}{\beta}\right)\frac{1}{ab}\frac{a-b}{a+b}\sin 2\zeta, \quad (4.25)$$

where Γ is the Gamma function

$$\Gamma(z) := \int_0^{+\infty} t^{z-1}e^{-t}dt, \quad z > 0.$$

These observations lead to the *focus series method*. One can record a *through-focus series* in order to find the optimum of the sharpness function and to identify the defocus value d_0 . From the same experimental data *stigmatic functions* can be computed, which would provide the information about the astigmatism. It is important to note that they do not produce directly the ideal stigmator control values d_{x_0} and d_{y_0} . However, one can clearly identify whether stigmators are supposed to be adjusted, and in which direction.

Figure 4.11 shows the Fourier transform-based sharpness and stigmatic functions computed for the Gaussian power spectrum with $\varsigma > 0, \theta = 0$, plotted as a function of σ . For the particular case of a Gaussian power spectrum $\beta = 1$, and

it follows that $\Gamma(1 + \frac{2}{\beta}) = \Gamma(3) = 2$ in (4.20)-(4.22). The function S_0 shows the general behaviour of the ideal sharpness function: It reaches its single optimum at $\sigma = 0$. The functions S_1, S_2 have a different behaviour, which is related to the stigmatic control variables. For instance, the stigmatic function S_1 is equal to zero at $\sigma = 0$ and has two local optima: a maximum and a minimum. It means that the x-stigmatic control d_x has to be adjusted, the distribution of the optima signs around zero identifies direction of this adjustment. The function S_2 is uniformly zero, which means that $d_y = d_{y_0}$.

The parameter ζ can be calculated directly from the values of S_1 and S_2

$$\zeta = \begin{cases} \frac{1}{2} \arctan(\frac{S_2}{S_1}), & \text{if } S_1 \neq 0, \\ \frac{\pi}{4}, & \text{if } S_1 = 0, \quad S_2 \neq 0, \\ \mathbb{R}, & \text{if } S_1 = 0, \quad S_2 = 0. \end{cases} \quad (4.26)$$

4.6 A link to the derivative-based approach

In this section we denote the functions (4.17) as S_q^{ft} . We show the connection between the Fourier transform-based sharpness/stigmatic functions (4.17) and the derivative-based sharpness function studied in the previous chapter.

$$S_{k,l}^{\text{der}} := \left\| \frac{\partial^k}{\partial x^k} \frac{\partial^l}{\partial y^l} f \right\|_{L^2}^2.$$

We find using Parseval's identity

$$S_{k,l}^{\text{der}} = \frac{1}{2\pi} \|u^k v^l \hat{f}\|_{L^2}^2 = \frac{1}{2\pi} \iint_{-\infty}^{\infty} u^{2k} v^{2l} |\hat{f}|^2 d\mathbf{u}.$$

From this it follows that the derivative-based sharpness function is related to the moments of the power spectrum as

$$\mathcal{M}_{2k,2l} = 2\pi S_{k,l}^{\text{der}}[f].$$

The Fourier transform-based sharpness and x-stigmatic functions can be expressed via the derivative-based sharpness function as

$$S_0^{\text{ft}} = 2\pi(S_{1,0}^{\text{der}} + S_{0,1}^{\text{der}}), \quad (4.27)$$

$$S_1^{\text{ft}} = 2\pi(S_{1,0}^{\text{der}} - S_{0,1}^{\text{der}}). \quad (4.28)$$

From the definition of stigmatic operators we then have

$$\mathcal{H}_1 \mathcal{R}_{\pi/4} f = 2 \oint_0^\pi |\mathcal{R}_{\pi/4} \hat{f}|^2 \cos 2\varphi d\varphi = 2 \oint_0^\pi |\hat{f}|^2 \cos 2(\varphi + \frac{\pi}{4}) d\varphi = -\mathcal{H}_2 f,$$

and as a consequence

$$S_2^{\text{ft}}[f] = -S_1^{\text{ft}}[\mathcal{R}_{\pi/4} f] = 2\pi(S_{0,1}^{\text{der}}[\mathcal{R}_{\pi/4} f] - S_{1,0}^{\text{der}}[\mathcal{R}_{\pi/4} f]). \quad (4.29)$$

Thus the Fourier transform-based sharpness/stigmatic functions can be easily computed through the derivative-based sharpness function (4.27)-(4.29). However, for the computations of sharpness/stigmatic operators discrete Fourier transform calculations are required. It is clear that though analytically the expressions for the Fourier transform-based functions and the derivative-based functions are equal, numerically we will obtain different results due to a variety of numerical errors. The experimental comparison of the quality of these results will be provided in Chapter 6. In this section all numerical computations are based on the discrete Fourier transform using the scheme explained in the next section.

It is important to note that for the numerical computation of the right hand side of (4.29) we do not necessarily need to rotate the discrete image, but can compute a directional derivative instead. Computations of directional derivatives for the discrete image is discussed in Section 2.7.

Similarly for the intensity-based sharpness function

$$S^{\text{int}} := \|f\|_{L^2}^2,$$

we have

$$\mathcal{M}_{0,0} = S_0^{\text{ft},(r)}[f] = 2\pi S_{0,0}^{\text{der}}[f].$$

4.7 Discretization

In this section we explain how the Fourier transform-based sharpness/stigmatism operators and functions can be computed numerically. Due to the power spectrum's symmetry about the origin (4.13) we have

$$\mathcal{H}_0(r) := 2 \oint_0^\pi |\hat{f}|^2 d\varphi, \quad (4.30)$$

$$\mathcal{H}_1(r) := 2 \oint_0^\pi |\hat{f}|^2 \cos 2\varphi d\varphi, \quad (4.31)$$

$$\mathcal{H}_2(r) := 2 \oint_0^\pi |\hat{f}|^2 \sin 2\varphi d\varphi. \quad (4.32)$$

The representation (4.30)-(4.32) enables us to do the numerical evaluation faster.

To execute the method we first need to discretize the functions. Let us assume that the continuous power spectrum $|\hat{f}|^2$ has a period $2U$. Define the frequency mesh points as $u_i := i\Delta u$, $i = -n, \dots, n$, where

$$\Delta u := \frac{2U}{(2n+1)-1} = \frac{U}{n}, \quad (4.33)$$

and define

$$p_{i,j} := |\hat{f}(u_i, u_j)|^2. \quad (4.34)$$

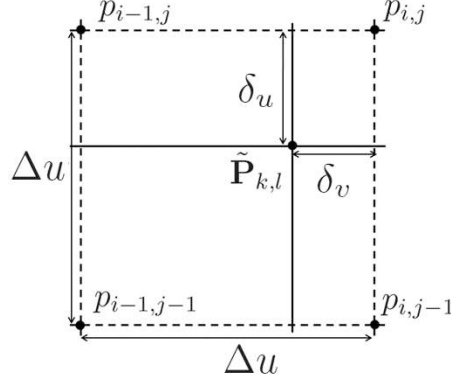


FIGURE 4.12: Linear interpolation of the power spectrum values.

We define the radial mesh points $r_k := k\Delta r + \frac{\Delta r}{2}$, $k = 0, \dots, N-1$, where $\Delta r := \frac{U}{N}$, and the angular mesh points $\varphi_l := l\Delta\varphi + \frac{\Delta\varphi}{2}$, $l = 0, \dots, M-1$, where $\Delta\varphi := \frac{\pi}{M}$. Further we define radial mesh points

$$u^{(k,l)} := r_k \cos \varphi_l, \quad v^{(k,l)} := r_k \sin \varphi_l. \quad (4.35)$$

Each point $(u^{(k,l)}, v^{(k,l)})$ is between four mesh points

$$u_{i-1} \leq u^{(k,l)} \leq u_i, \quad v_{j-1} \leq v^{(k,l)} \leq v_j, \quad i, j \in \{-n+1, \dots, n\}.$$

We compute entries of $\tilde{\mathbf{P}}_{k,l} \in \mathbb{R}^{N \times M}$ using bilinear interpolation (see Figure 4.12)

$$\tilde{\mathbf{P}}_{k,l} = \left(\left(1 - \frac{\delta_u}{\Delta u}\right) p_{i-1,j-1} + \frac{\delta_u}{\Delta u} p_{i,j-1} \right) \left(1 - \frac{\delta_v}{\Delta v}\right) + \left(\left(1 - \frac{\delta_u}{\Delta u}\right) p_{i-1,j} + \frac{\delta_u}{\Delta u} p_{i,j} \right) \frac{\delta_v}{\Delta v},$$

where

$$\delta_u := u^{(k,l)} - u_{i-1}, \quad \delta_v := v^{(k,l)} - v_{j-1}.$$

For $k \in \{1, \dots, N\}$ we approximate the values of $\mathcal{H}_0 f(r_k)$, $\mathcal{H}_1 f(r_k)$, $\mathcal{H}_2 f(r_k)$ with the repeated midpoint numerical integration rule

$$\mathcal{H}_0 f(r_k) \doteq 2\Delta\varphi \sum_{l=1}^M \tilde{\mathbf{P}}_{k,l}, \quad (4.36)$$

$$\mathcal{H}_1 f(r_k) \doteq 2\Delta\varphi \sum_{l=1}^M \tilde{\mathbf{P}}_{k,l} \cos 2\varphi_l, \quad (4.37)$$

$$\mathcal{H}_2 f(r_k) \doteq 2\Delta\varphi \sum_{l=1}^M \tilde{\mathbf{P}}_{k,l} \sin 2\varphi_l. \quad (4.38)$$

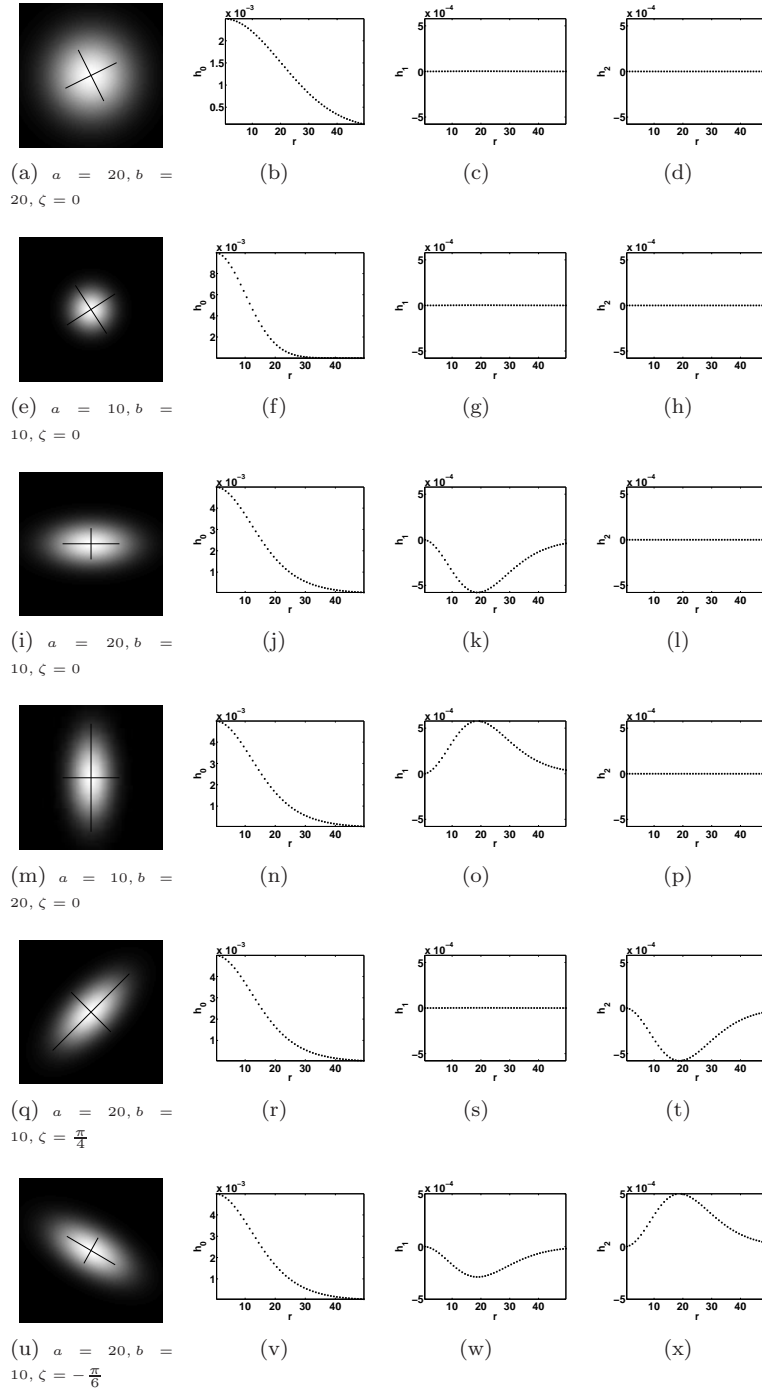


FIGURE 4.13: Numerical computations for a Gaussian benchmark power spectrum.

Then the sharpness/stigmatic function values are approximated with the mid-point rule as well

$$S_q \doteq \Delta r \sum_{k=1}^N r_k^3 h_q(r_k), \quad q = 1, 2, 3. \quad (4.39)$$

4.8 Numerical experiments

In this section we consider two situations to illustrate the performance of the method: a Gaussian benchmark and SEM experimental data.

4.8.1 A Gaussian benchmark

First we consider a Gaussian function sampled at $(2n+1) \times (2n+1) = 101 \times 101$ frequency mesh points. To compute the sharpness and stigmatic operators $\mathcal{H}_0 f, \mathcal{H}_1 f, \mathcal{H}_2 f$, for the discretization of the polar radius r we choose for $N = n = 50$ data points and for the discretization of the polar angle φ we choose for $M = 2(2n+1) = 202$. We take $\Delta u = 1$ in (4.33), thus $U = n$. Each row in Figure 4.13 shows a Gaussian function with particular parameter values a, b, ζ and corresponding $\mathcal{H}_0 f, \mathcal{H}_1 f, \mathcal{H}_2 f$ functions. The rotation angle ζ is numerically estimated according to (4.26) and indicated by two orthogonal lines plotted above each of the Gaussian functions.

From these pictures we observe the following:

1. Figures 4.13(a)-4.13(d): the Gaussian is rotationally symmetric ($a = b = 20$). For this reason the computed values of $\mathcal{H}_1 f$ and $\mathcal{H}_2 f$ are equal to zero.
2. Figures 4.13(e)-4.13(h): the Gaussian is rotationally symmetric ($a = b = 10$). The values of $\mathcal{H}_1 f$ and $\mathcal{H}_2 f$ are equal to zero. The Gaussian width is smaller than in Figure 4.13(a). The integral of the function $\mathcal{H}_0 f$ in Figure 4.13(f) decreases in comparison with the previous experiment (Figure 4.13(b)). The goal of a human operator in this case is to find the control variables with the smallest point spread function width, and as a consequence with the intensity of $\mathcal{H}_0 f$ as large as possible.
3. Figures 4.13(i)-4.13(l): an elliptic Gaussian. The widths $a > b$ and as a consequence all values of $\mathcal{H}_1 f$ are smaller than zero. The values of $\mathcal{H}_2 f$ are equal to zero, because $\zeta = 0$.
4. Figures 4.13(m)-4.13(p): an elliptic Gaussian. The widths $b > a$ and as a consequence all values of $\mathcal{H}_1 f$ are larger than zero. The values of $\mathcal{H}_2 f$ are equal to zero, because $\zeta = 0$.
5. Figures 4.13(q)-4.13(t): an elliptic Gaussian with $a > b, \zeta = \frac{\pi}{4}$. As a consequence $\mathcal{H}_1 f = 0, \mathcal{H}_2 f < 0$.

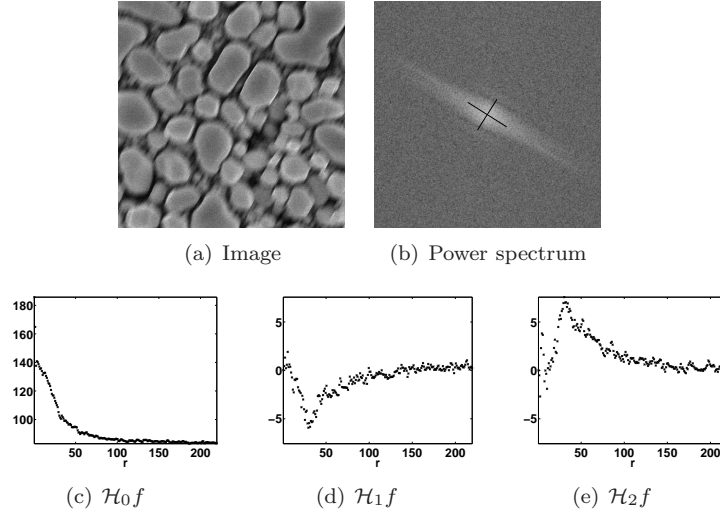


FIGURE 4.14: From left to right, from top to bottom: in-focus astigmatism-free tin balls image, logarithmic scale of its power spectrum, numerically computed operators $\mathcal{H}_0f(r)$, $\mathcal{H}_1f(r)$, $\mathcal{H}_2f(r)$.

6. Figures 4.13(u)-4.13(x): an elliptic Gaussian with $a > b$, $\zeta = -\frac{\pi}{6}$. As a consequence $\mathcal{H}_1f < 0$, $\mathcal{H}_2f > 0$.

The numerical results agree with the analytical observations and can guide an unexperienced human operator, giving visual suggestion about a proper choice of stigmator controls and their directions.

4.8.2 SEM images

In the next example the operators \mathcal{H}_0f , \mathcal{H}_1f , \mathcal{H}_2f are computed for power spectra of real-world SEM images. The rotation angle ζ is numerically estimated using (4.26). Each of the figures 4.14-4.18 show

- SEM experimental image;
- Logarithmic scale of its power spectrum;
- Operators \mathcal{H}_0f , \mathcal{H}_1f , \mathcal{H}_2f computed with (4.36)-(4.38).

The size of each experimental image is $(2n + 1) \times (2n + 1) = 441 \times 441$ pixels. The following values are chosen for discretization $\Delta u = 1$, $N = n = 220$, $M = 2(2n + 1) = 882$. For each operator \mathcal{H}_0f , \mathcal{H}_1f , \mathcal{H}_2f the values of S_0 , S_1 , S_2 are computed with (4.39). Further, the angle ζ is identified using (4.26). Two orthogonal lines above each power spectrum visualize the observed value of ζ . The operators \mathcal{H}_0f , \mathcal{H}_1f , \mathcal{H}_2f shown in the figures are computed for the logarithmic scale of the power spectrum for a convenient visualization. The

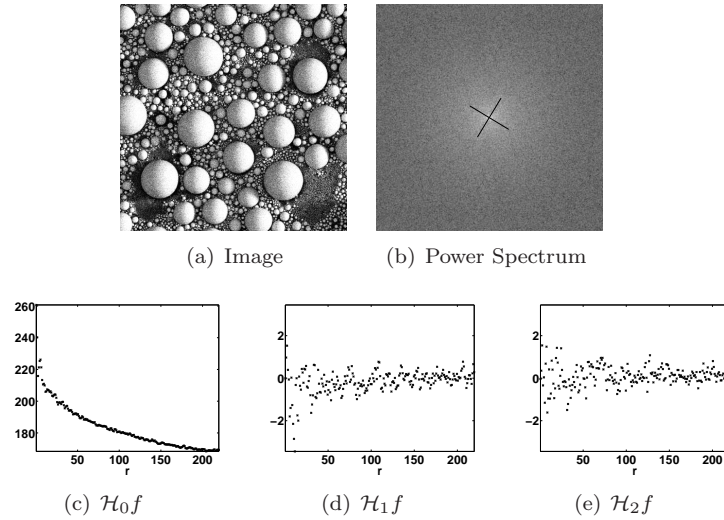


FIGURE 4.15: From left to right, from top to bottom: tin balls image with astigmatism, logarithmic scale of its power spectrum, numerically computed operators $\mathcal{H}_0 f(r)$, $\mathcal{H}_1 f(r)$, $\mathcal{H}_2 f(r)$.

angle ζ shown by the two orthogonal lines is estimated from sharpness/stigmatic function values before the logarithmic scale.

Figure 4.14 shows the experiment for *gold-on-carbon stigmatic image*. Adjustment of both d_x and d_y is necessary to improve the image quality. The signs of the curves $\mathcal{H}_1 f$ and $\mathcal{H}_2 f$ indicate the direction of change. Figure 4.15 shows in-focus astigmatism-free image of tin balls. Both $\mathcal{H}_1 f$ and $\mathcal{H}_2 f$ is just numerical noise around zero: hence astigmatism correction is not needed in this case. Figure 4.16 shows an out-of-focus stigmatic image of the same sample. We can see how the image quality decreases and power spectrum shape changes. The y-stigmatic operator suggests the change of d_y , while adjustment of d_x is not necessary. Figure 4.17 shows a magnified image of tin balls. The image does not show a lot of details and as a consequence its power spectrum has only a few values of low frequencies different from noise. Because of the lack of information in the Fourier space of the image, it is difficult to analyse this power spectrum and to draw the conclusions about the presence of astigmatism in the image. For this type of samples the method of orientation identification as well as other Fourier transform-based methods might fail. In such a case the method of simultaneous defocus and astigmatism correction described in Chapter 7 will be more useful than the Fourier transform-based techniques. Figure 4.18 shows one more out-of-focus and stigmatic image. We can clearly see that correction of both d_x and d_y is needed.

Figure 4.19 shows sharpness and stigmatic functions (S_q on the left and $S_q^{(r)}$ on the right) computed for experimental gold-on-carbon through-focus series.

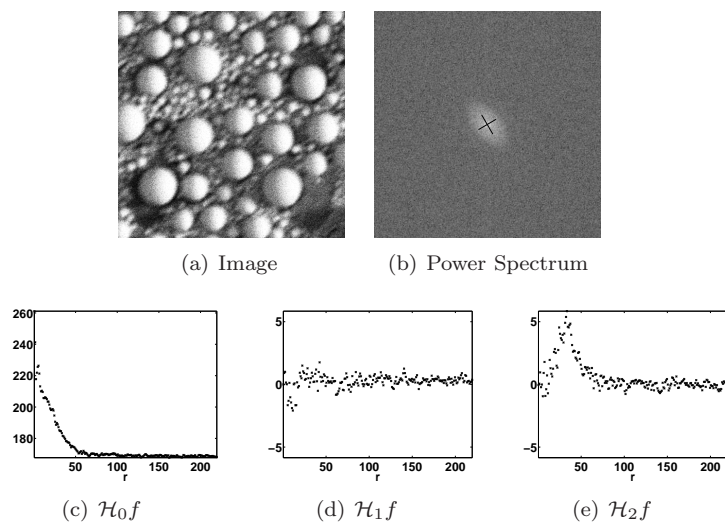


FIGURE 4.16: From left to right, from top to bottom: SEM experimental image, logarithmic scale of its power spectrum, numerically computed operators $\mathcal{H}_0 f(r)$, $\mathcal{H}_1 f(r)$, $\mathcal{H}_2 f(r)$.

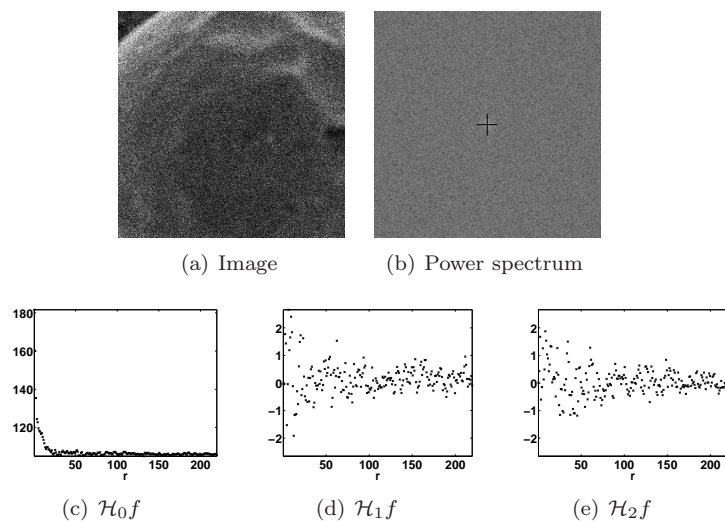


FIGURE 4.17: From left to right, from top to bottom: gold-on-carbon image with astigmatism, logarithmic scale of its power spectrum, numerically computed operators $\mathcal{H}_0 f(r)$, $\mathcal{H}_1 f(r)$, $\mathcal{H}_2 f(r)$.

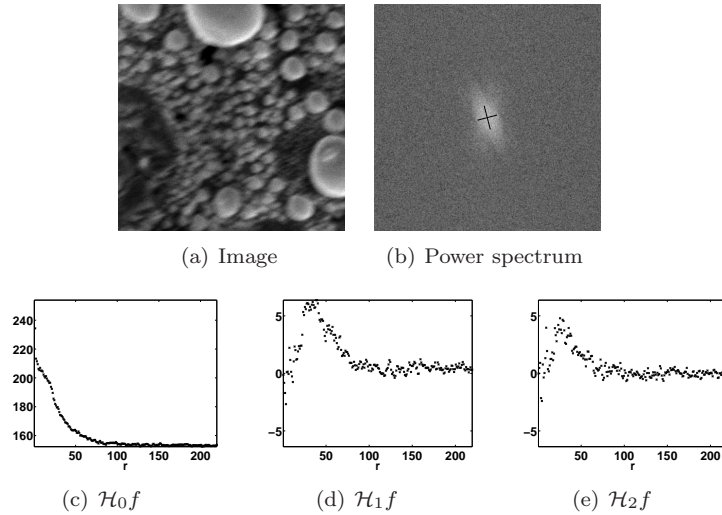


FIGURE 4.18: From left to right, from top to bottom: SEM experimental image, logarithmic scale of its power spectrum, numerically computed operators $\mathcal{H}_0 f(r)$, $\mathcal{H}_1 f(r)$, $\mathcal{H}_2 f(r)$.

The shapes of the functions are similar to the shapes obtained for the benchmarks. They indicate the astigmatism presence in the through-focus series.

The alternative approach to extracting power spectrum parameters could be simply fitting it with a continuous model [8] by minimizing least square difference of continuous and experimental data with the help of an iterative method (for instance the Newton method). In [8] this is done for one-dimensional case under the assumption that the power spectrum is rotationally symmetric. For the one-dimensional case this would take much longer computational time than computing of sharpness/stigmatic operators, which is non-iterative.

Another approach for extracting sharpness/stigmatic information from power spectrum could be based on a non-iterative fitting of discrete power spectrum with a set of basis functions, for instance via the projection method [66]. In this case the sharpness/stigmatic operator could be pre-computed analytically for the given set of basis functions. However, for the set of basis function explored so far, the approach is still slower than direct computation of sharpness/stigmatic operator described in this chapter. More details on this approach can be found in Appendix B.

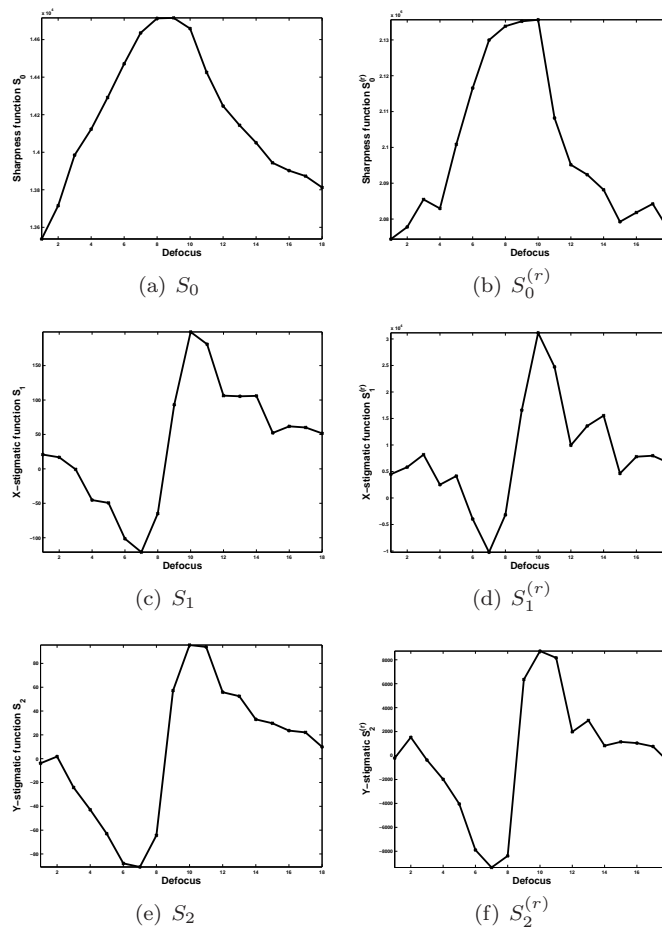


FIGURE 4.19: Sharpness, x-stigmatic and y-stigmatic functions computed for SEM experimental gold-on-carbon through-focus series. Horizontal axes show defocus in arbitrary units.

Chapter 5

Statistics-based approaches

In this chapter we discuss alternative sharpness functions: autocorrelation-based [63], variance-based, intensity-based and histogram-based sharpness functions denoted as S^{ac} , S^{var} , S^{int} and S^{his} respectively. In the next chapter these functions will be applied to a set of SEM experimental data.

5.1 Autocorrelation-based sharpness function

In Subsection 5.1.1 we define the autocorrelation operator and its properties. Then two sharpness functions based on the first and second derivatives of the image's autocorrelation are introduced and discussed in Subsection 5.1.2. The second derivative autocorrelation-based sharpness function turns out to be equivalent to the derivative-based sharpness function (3.19). In Subsection 5.1.3 a formulation of the discrete sharpness functions is given.

5.1.1 Image autocorrelation

The *autocorrelation* of a continuous image $f \in \mathbb{L}_2(\mathbb{R}^2)$ is defined as

$$\check{f}(\mathbf{x}) = \iint_{-\infty}^{\infty} f(\mathbf{x}')f^*(\mathbf{x}' - \mathbf{x})d\mathbf{x}'.$$

Property 5.1. *The autocorrelation function reaches its maximum at the origin*

$$\max_{\mathbf{x}} \check{f}(\mathbf{x}) = \check{f}(0, 0).$$

Proof. See for instance [38]. □

Corollary 5.1. *The autocorrelation's maximum is equal to the squared L_2 -norm of the image*

$$\check{f}(0) = \|f\|_{L_2}^2.$$

Corollary 5.2. *The Fourier transform of the autocorrelation is equal to the image power spectrum*

$$\mathfrak{F}\check{f} = |\hat{f}|^2.$$

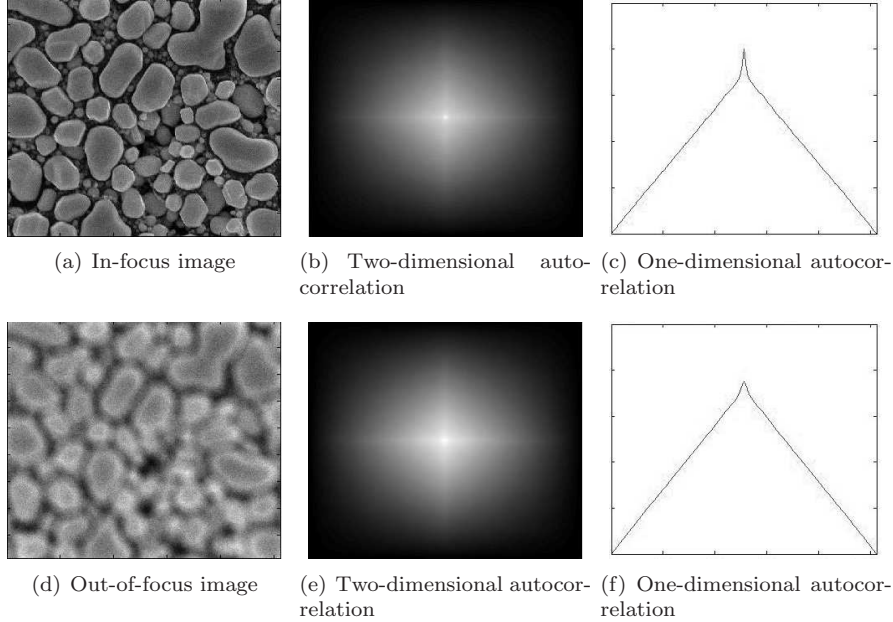


FIGURE 5.1: In-focus and out-of-focus SEM experimental images and their autocorrelations. Most important in the autocorrelation function is the shape of its peak, which indicates the sharpness of the image.

Corollary 5.3. *For the object benchmark (2.11) and a Gaussian point spread function, image autocorrelation is a Gaussian function*

$$\check{f}(x) = \frac{C\sqrt{\pi}}{\sqrt{\gamma^2 + \sigma^2 + \alpha^2}} e^{-\frac{\|x\|^2}{4(\gamma^2 + \sigma^2 + \alpha^2)}}. \quad (5.1)$$

5.1.2 The sharpness function

A number of autocorrelation-based sharpness functions can be found in the literature [24, 53, 88]. These functions are based on heuristic knowledge of the autocorrelation behaviour for *in-focus* and *out-of-focus* images. One example of such a behaviour is shown in Figure 5.1. In general the sharp edges in the image result in a small amount of highly correlated regions. Out-of-focus images contain large smooth areas, which are highly correlated. As a consequence the autocorrelation peak of the in-focus image is narrow and high, while the autocorrelation peak of the out-of-focus image is broad and low [28]. These observations lead to the following sharpness function based on the first derivative of the image autocorrelation (cf. [88, 89])

$$S^{\text{ac}_1} := \frac{\partial}{\partial x} \check{f} \Big|_{\mathbf{x}=(\epsilon, 0)}, \quad \epsilon > 0, \quad (5.2)$$

where ϵ is small.

Property 5.2. *For the benchmark object (2.11) and a Gaussian point spread function, the autocorrelation-based sharpness function (5.2) is given by*

$$S^{\text{ac}_1}(\sigma) = \frac{C\sqrt{\pi}\epsilon}{2(\gamma^2 + \sigma^2 + \alpha^2)^{3/2}} e^{-\frac{\epsilon^2}{4(\gamma^2 + \sigma^2 + \alpha^2)}}. \quad (5.3)$$

Proof. The proof follows directly from the definition of the sharpness function (5.2) and Corollary 5.3. \square

The derivative of the sharpness function (5.3)

$$\frac{\partial}{\partial \sigma} S^{\text{ac}_1} = \frac{C\sqrt{\pi}\sigma(\epsilon^2 - 6(\gamma^2 + \sigma^2 + \alpha^2))e^{-\frac{\epsilon^2}{4(\gamma^2 + \sigma^2 + \alpha^2)}}}{4(\gamma^2 + \sigma^2 + \alpha^2)^{7/2}}$$

has roots at

$$\sigma_0 = 0, \quad \sigma_{1,2} = \pm \sqrt{-\gamma^2 - \alpha^2 + \frac{\epsilon}{6}}.$$

In order for the sharpness function to have a single optimum at $\sigma_0 = 0$, we can choose $\epsilon < 6(\gamma^2 + \alpha^2)$ for the benchmark. In general it is difficult to guess a proper value of ϵ , since the value depends on the object function ψ . For a very small ϵ the sharpness function will be very sensitive to noise in the image formation. In practice for most applications ϵ is determined experimentally [88, 89].

A more natural way to define the parameter-independent autocorrelation-based sharpness function is the evaluation of the second derivative of the autocorrelation at the origin

$$S^{\text{ac}_2}(\sigma) := -\frac{\partial^2}{\partial x^2} \check{f} \Big|_{\mathbf{x}=(0,0)}.$$

For instance for the benchmark (5.1), the autocorrelation second derivative changes monotonically with the width of the point spread function. Below we link the autocorrelation-based sharpness function and the derivative-based sharpness function

$$S^{\text{der}} := \left\| \frac{\partial}{\partial x} f \right\|_{L_2}^2, \quad (5.4)$$

that was studied in Chapter 3.

Property 5.3. *The autocorrelation-based and derivative-based sharpness functions are identical*

$$S^{\text{ac}_2} = S^{\text{der}}.$$

Proof. It follows from Corollary 5.2 that

$$-\frac{\partial^2}{\partial x^2} \check{f} = -\frac{\partial^2}{\partial x^2} \mathfrak{F}^{-1} |\hat{f}|^2 =$$

$$-\frac{1}{2\pi} \frac{\partial^2}{\partial x^2} \iint_{-\infty}^{\infty} e^{i\mathbf{u}\cdot\mathbf{x}} |\hat{f}|^2 d\mathbf{u} = \frac{1}{2\pi} \iint_{-\infty}^{\infty} u^2 e^{i\mathbf{u}\cdot\mathbf{x}} |\hat{f}|^2 d\mathbf{u}.$$

Because of Parseval's identity we have

$$\begin{aligned} S^{\text{ac}_2} &= -\frac{\partial^2}{\partial x^2} \check{f} \Big|_{\mathbf{x}=(0,0)} = \frac{1}{2\pi} \iint_{-\infty}^{\infty} u^2 |\hat{f}|^2 d\mathbf{u} = \\ &= \frac{1}{2\pi} \|\omega \hat{f}\|_{L_2}^2 = \left\| \frac{\partial}{\partial x} f \right\|_{L_2}^2 =: S^{\text{der}}. \end{aligned}$$

□

5.1.3 Discretization

The autocorrelation coefficients for the discrete image are defined as in [88]

$$\check{f}_{k,l} := \sum_{i,j} f_{i,j} f_{i+k,j+l}. \quad (5.5)$$

An approximation of the autocorrelation-based sharpness function s^{ac_1} is found by taking *finite difference* (cf. [88, 89])

$$S^{\text{ac}_1} \doteq s^{\text{ac}_1} := \check{f}_{0,0} - \check{f}_{1,0}. \quad (5.6)$$

The discrete derivative-based sharpness function s^{der} is defined as in (3.28)

$$S^{\text{der}} \doteq s^{\text{der}} := \sum_{i,j} (f_{i,j} - f_{i+k,j})^2. \quad (5.7)$$

This follows directly from a two-point *finite difference* approximation for the first derivative and the mid-point numerical integration rule. It is straightforward to check that the discrete derivative-based sharpness function and the discrete second-derivative autocorrelation-based sharpness function are identical

$$\begin{aligned} s^{\text{der}} &= \sum_{i,j} f_{i,j}^2 - 2 \sum_{i,j} f_{i,j} f_{i+k,j} + \sum_i f_{i+k,j}^2 = \\ &= 2 \left(\sum_{i,j} f_{i,j}^2 - \sum_{i,j} f_{i,j} f_{i+k,j} \right) = 2(\check{f}_{0,0} - \check{f}_{k,0}) = \\ &= -\check{f}_{-k,0} + 2\check{f}_{0,0} - \check{f}_{k,0} = s^{\text{ac}_2} \doteq S^{\text{ac}_2}. \end{aligned}$$

5.2 Variance- and intensity-based sharpness functions

For an image f with compact support \mathbb{X} , its mean value $E[f]$ is defined as

$$E[f] := \bar{f} := \left(\iint_{\mathbb{X}} f d\mathbf{x} \right) / \left(\iint_{\mathbb{X}} d\mathbf{x} \right) \quad (5.8)$$

The variance-based sharpness function is defined as (cf. [16, 84])

$$S^{\text{var}}[f] := \|f - \bar{f}\|_{L_2}^2. \quad (5.9)$$

A discussion on this variance-based sharpness function's robustness with respect to the image noise can be found in Appendix C. Consider the amplitude image function

$$f^{(A)} := f - \bar{f}.$$

It follows from the definition that the mean value of the image amplitude function is equal to zero

$$E[f^{(A)}] = \left(\iint_{\mathbb{X}} f \, d\mathbf{x} - \bar{f} \iint_{\mathbb{X}} d\mathbf{x} \right) / \left(\iint_{\mathbb{X}} d\mathbf{x} \right) = 0.$$

In some applications the amplitude image function $f^{(A)}$ is used instead of f for the sharpness analysis (cf. [16]). In this case we obtain $S^{\text{var}}[f^{(A)}] = S^{\text{var}}[f - \bar{f}] = \|f\|_{L_2}^2$, which is known as the intensity-based sharpness function

$$S^{\text{int}} := \|f\|_{L_2}^2. \quad (5.10)$$

The intensity-based sharpness function is equal to the height of the autocorrelation peak, and for this reason is intuitively used for autofocus [53].

In the property and two corollaries below we consider a rotationally-symmetric point spread function, i.e. $\varsigma = 0$ in (2.15).

Property 5.4. *The sharpness function (5.10) can be expressed as follows*

$$S^{\text{int}}(\sigma) = \frac{1}{2\pi} \iint_{-\infty}^{\infty} |\hat{\psi}(\mathbf{u})|^2 e^{-\sigma^2 \beta \|\mathbf{u}\|^2} e^{-\alpha^2 \|\mathbf{u}\|^2} d\mathbf{u}. \quad (5.11)$$

Corollary 5.4. *The sharpness function (5.10) is smooth, and is strictly increasing for $\sigma < 0$ and strictly decreasing for $\sigma > 0$.*

Corollary 5.5. *For $\alpha > 0$ the sharpness function (5.10) has a finite maximum at $\sigma = 0$*

$$\max_{\sigma} S^{\text{int}}(\sigma) = S^{\text{int}}(0), \quad \forall \alpha > 0.$$

For the benchmark object (2.11) and the Gaussian point spread function the intensity-based sharpness function can be computed directly

$$\begin{aligned} S^{\text{int}}(\sigma) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-((\varsigma-\sigma)^2 + \alpha^2 + \gamma^2)u^2} du \int_{-\infty}^{\infty} e^{-((\varsigma+\sigma)^2 + \alpha^2 + \gamma^2)v^2} dv = \\ &= \frac{1}{2\pi} \left((\varsigma^2 + \sigma^2 + \alpha^2 + \gamma^2)^2 - 4\varsigma^2\sigma^2 \right)^{1/2}. \end{aligned}$$

It trivially follows that in this case $\frac{\partial}{\partial \sigma} S^{\text{int}}$ has three roots σ for $\varsigma > \sqrt{\alpha^2 + \gamma^2}$ and one root at $\sigma = 0$ otherwise. Thus the intensity-based sharpness function

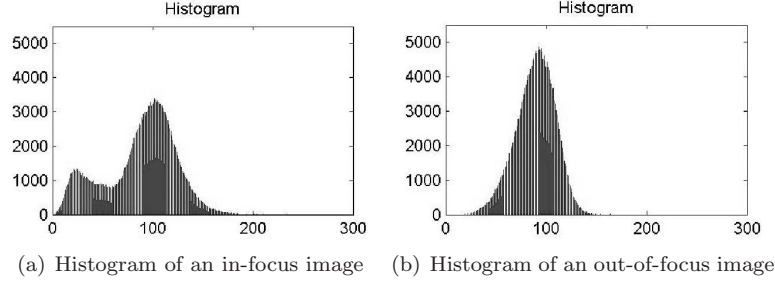


FIGURE 5.2: Histograms computed for in-focus and out-of-focus SEM experimental images of the same object.

has a single optimum even for larger amounts of astigmatism than the derivative-based sharpness function, which has local optima if $\zeta > \frac{1}{\sqrt{2}}\sqrt{\alpha^2 + \gamma^2}$ (Chapter 3).

For the discrete image the mean value is approximated by

$$\bar{f} \doteq \bar{F} := \frac{\sum_{i,j} \Delta x^2 f_{i,j}}{\sum_{i,j} \Delta x^2} = \frac{1}{N^2} \sum_{i,j} f_{i,j} \quad (5.12)$$

and the discrete variance-based sharpness function is

$$S^{\text{var}} \doteq s^{\text{var}} := \frac{1}{N^2} \sum_{i,j} (f_{i,j} - \bar{F})^2. \quad (5.13)$$

Similarly applying the *mid-point rule* and considering the threshold parameter as in [69], we obtain

$$S^{\text{int}} \doteq s^{\text{int}} = \sum_{i,j} |f_{i,j}|^2, \quad f_{i,j} > \Theta. \quad (5.14)$$

5.3 Histogram-based sharpness function

In most applications the unscaled image \mathbf{F} is a matrix of natural intensity values. Let

$$\tilde{\mathbf{f}} = (\tilde{f}_k)_{k=1}^L, \quad \tilde{f}_{k-1} < \tilde{f}_k,$$

be a set of all *pixel* values in the image \mathbf{F} , i.e. $f_{i,j} \in \mathbf{F} \Leftrightarrow \exists k$ such that $f_{i,j} = \tilde{f}_k \in \tilde{\mathbf{f}}$. The vector $\mathbf{h} = (h_k)_{k=1}^L$, where h_k is the number of pixels with the value \tilde{f}_k in the image \mathbf{F} , is called the *histogram* of the image \mathbf{F} . Then the probability of a pixel value equal to \tilde{f}_k is $\frac{h_k}{N^2}$.

Figure 5.2 shows the histograms of in-focus and out-of-focus SEM experimental images. The horizontal axis on each diagram represents the pixel gray values, and the vertical axis the number of counts \mathbf{h} . The in-focus image has

the whole range of the pixel values, including pixels equal to 0 and to 255. The out-of-focus image has less contrast, and its values in this case are spread between 12 and 130. These observations lead to the histogram-based sharpness function, known as histogram range [69]

$$S^{\text{his}_r} := \max_{k, h_k \neq 0} h_k - \min_{k, h_k \neq 0} h_k. \quad (5.15)$$

It is clear from the above example that the larger the range the more contrast the image has, and the more information it contains. Other histogram-based sharpness functions are the *entropy* (cf. [51])

$$S^{\text{his}_e} := - \sum_{k, h_k \neq 0} \frac{h_k}{N^2} \log_2 \frac{h_k}{N^2}. \quad (5.16)$$

and the threshold image count (cf. [69])

$$S^{\text{his}_t} := \sum_{k=1}^n h_k, \quad \tilde{f}_n \leq \Theta \quad \tilde{f}_{n+1} > \Theta. \quad (5.17)$$

Chapter 6

Assessing sharpness functions

For SEM *static autofocus* applications (see Section 2.6) this chapter is devoted to determine a *sharpness function* that will properly perform for a large variety of samples. To this end we consider the discrete sharpness functions introduced above, namely the *derivative-based* sharpness function (Section 3.5), the *autocorrelation-based* sharpness function (Subsection 5.1.3), the *variance-based* and the *intensity-based* sharpness functions (Section 5.2), the *histogram-based* sharpness function (Section 5.3). This collection of sharpness functions includes the ones from the classical paper on autofocus [69] that assesses them for fluorescence *light microscopy*. In addition our derivative-based and autocorrelation-based sharpness functions are extended with parameters, which lead to improvements. We consider another group of sharpness functions based on the image *Fourier transform*, cf. Section 4.5.

The *Fourier transform-based* sharpness functions were not included in the earlier assessments [69, 73, 40] of the sharpness functions because of the fact that they require more CPU time. With the desktop CPU's available nowadays the calculation of the *discrete Fourier transform* only takes milliseconds. In particular in electron microscopy the Fourier transform computation of an image in a *through-focus series* takes less time than the acquisition of another image. Thus the image Fourier transform can easily be computed without slowing down the autofocus procedure.

We discuss how to rank sharpness functions and apply them to 14 experimental SEM through-focus series of the samples with different geometries, some of which contain *astigmatism*. Our ranking procedure is a modified version of [40, 69]. For the ranking we use the electron microscopy instead of the light microscopy images, as in [40, 69]. Based on our ranking procedure the derivative-based sharpness function scores best, the Fourier transform-based function scores almost equal.

Section 6.1 lists the sharpness functions to be ranked. Sections 6.2 and 6.3 describe the experimental SEM data sets and the ranking procedure respectively. Section 6.4 provides the ranking results. Finally in Section 6.5 we discuss the numerically computed sharpness functions, in particular which aspects of the sharpness functions mainly influence their positions in the ranking.

6.1 The sharpness functions

For the discrete image $\mathbf{F} := (f_{i,j})_{i,j=1}^N$ we list a number of the sharpness functions to be assessed. For $k \in \mathbb{N}$, $p = \{1, 2\}$ and $\Theta \in \mathbb{R}^+$ the discrete *derivative-based sharpness function* is defined as in (3.27)

$$s_x^{\text{der}} := \sum_{i,j} |f_{i,j} - f_{i,j+k}|^p, \quad |f_{i,j} - f_{i,j+k}|^p > \Theta. \quad (6.1)$$

We remark that k (*pixel difference*) and Θ (*threshold*) influence the *noise sensitivity* of the sharpness function. The papers [6], [69] only use $k = 1, 2$. Below we experimentally show that larger values of k often lead to higher scores in the assessment. Let s_y^{der} be the sharpness function that computes the norm of the pixel difference in vertical direction. Then the form that generalizes derivative-based sharpness functions is

$$s^{\text{der}} := \frac{1}{2}(s_x^{\text{der}} + v s_y^{\text{der}}), \quad v \in \{0, 1\}. \quad (6.2)$$

We also consider the discrete derivative-based sharpness (cf. (3.30))

$$s^{\text{der,t}} := \sum_{i,j} ((\mathbf{F} * \mathbf{G}_1)_{i,j}^2 + (\mathbf{F} * \mathbf{G}_2)_{i,j}^2), \quad ((\mathbf{F} * \mathbf{G}_1)_{i,j}^2 + (\mathbf{F} * \mathbf{G}_2)_{i,j}^2) > \Theta, \quad (6.3)$$

where $\mathbf{G}_1, \mathbf{G}_2$ are the so-called *Sobel operators*, defined in (3.31), and $*$ denotes discrete convolution.

The discrete *autocorrelation difference sharpness function* is defined as in (5.6), Subsection 5.1.3

$$s_x^{\text{acr}} := \check{f}_{0,k} - \check{f}_{0,k+l}. \quad (6.4)$$

Similarly to the derivative-based sharpness function both vertical and horizontal directions can be considered

$$s^{\text{acr}} := \frac{1}{2}(s^{\text{acr,x}} + v s^{\text{acr,y}}), \quad v = \{0, 1\}. \quad (6.5)$$

For $p = \{1, 2\}$ the discrete *variance-based sharpness function* and the *intensity-based sharpness functions* are defined as in (5.13), (5.14) Section 5.2

$$s^{\text{var}} := \sum_{i,j} (f_{i,j} - \bar{F})^2, \quad (6.6)$$

$$s^{\text{int}} := \sum_{i,j} |f_{i,j}|^p, \quad f_{i,j} > \Theta. \quad (6.7)$$

The *histogram-based sharpness functions* are defined, as in (5.15)-(5.17)

$$s^{\text{his}_r} := \max_{k, h_k \neq 0} h_k - \min_{k, h_k \neq 0} h_k. \quad (6.8)$$

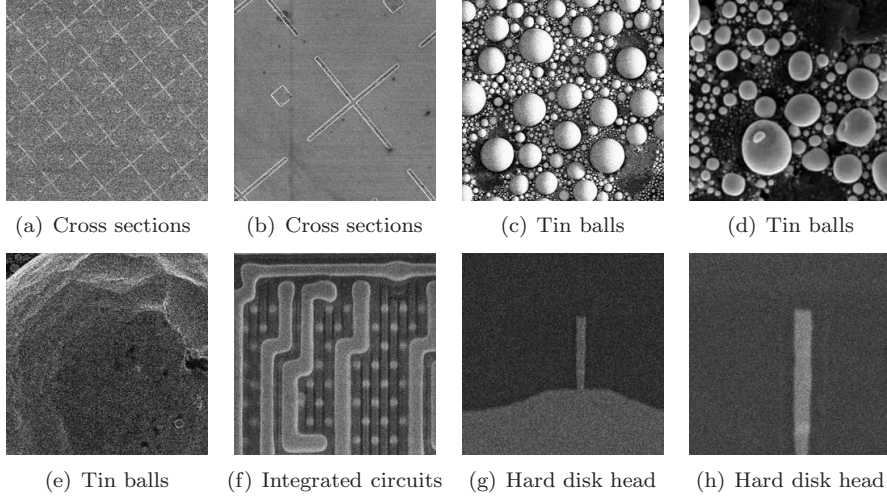


FIGURE 6.1: In-focus images from experimental SEM through-focus series without astigmatism.

$$s^{\text{his}_e} := - \sum_{k, h_k \neq 0} \frac{h_k}{N^2} \log_2 \frac{h_k}{N^2}. \quad (6.9)$$

$$s^{\text{his}_t} := \sum_{k=1}^n h_k, \quad \tilde{f}_n \leq \Theta, \quad \tilde{f}_{n+1} > \Theta. \quad (6.10)$$

The discrete *Fourier transform-based sharpness function* is defined (cf. [52, 87])

$$s^{\text{ft}} := \sum_{i,j} \left| \hat{f}_{i,j} \right|^2, \quad i, j \in \{h, \dots, n_0 - l\} \cup \{n_0 + l, \dots, N - h\}, \quad (6.11)$$

where l and h are the low and the high band frequencies, $n_0 = \text{mod}(N/2) + 1$. In (6.11) $\hat{f}_{i,j}$ are the entries of the *discrete Fourier transform*. Before computing the discrete Fourier transform the images are premultiplied by the *window function* (4.3).

6.2 The images

Fourteen experimental *through-focus series* are recorded with a FEI Strata SEM at *magnifications* from $15.000\times$ to $25.000\times$. The number of images in a through-focus series varies from 11 to 53. Figures 6.1 and 6.2 show the in-focus images from each of the series. The size of each image is 442×442 pixels. The images shown in Figure 6.1 are from astigmatism-free through-focus series. Those in Figure 6.2 are from the through-focus series with *astigmatism*.

Figures 6.1(c), 6.1(d), 6.1(e), 6.2(c), 6.2(d) show the images of a *tin balls* sample; Figures 6.1(a), 6.1(b), 6.2(a), 6.2(b) show images of a *cross sections*

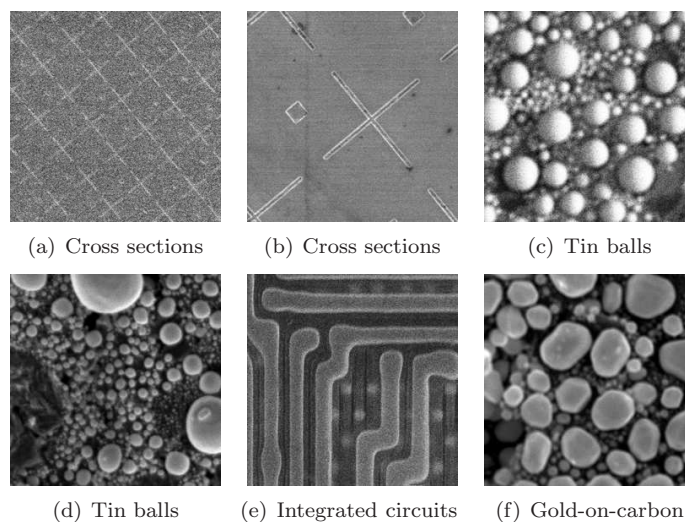


FIGURE 6.2: In-focus images from experimental SEM through-focus series with astigmatism.

sample; Figures 6.1(f), 6.2(e) show images of a *integrated circuits* sample; Figures 6.1(g), 6.1(h) show images of a *hard disk head* sample; Figure 6.2(f) show images of a *gold-on-carbon* sample.

The images contain various details. Series of images, such as in Figure 6.1(c), contain fine details, i.e. the distance between neighbouring objects in an image is several *pixels*. Other series, like in Figure 6.1(b) and Figure 6.1(g), contain only coarse details (the distance between two objects is more than 100 pixels). There are also periodic images (such as in Figure 6.1(a)).

6.3 Assessment

In the recent paper with the assessment of sharpness functions [40] several ranking scores, based on the previous papers [69, 73] are explained. We consider the four ranking scores described in [40].

- **Accuracy score** z_{acc} measures the distance between the best focus position, determined by a professional human operator and the maximum of a sharpness function.
- **Local maxima score** z_{lm} counts the number of local (false) maxima in a sharpness function.
- **Range score** z_{ran} estimates the interval around a sharpness function global maximum, where the function is concave.

- **Noise score** z_{noise} estimates the noise amplitude in a sharpness function.

We do not consider the dynamic ranking scores defined in [40], because in this chapter we discuss only *static autofocus* (see Section 2.6). We omit the ranking score that estimates the *half width at half maximum* of a sharpness function peak [40, 73]. The narrow sharpness function peak is not always beneficial for static autofocus in electron microscopy: if the defocus step in a through-focus series is large (*course focusing*), the sharpness function peak can be unobserved and autofocus procedure will fail. Earlier the computational time required by a sharpness function was also considered as a ranking score [69]. Nowadays the computational time of all sharpness functions considered in this chapter is lower than the image acquisition time in electron microscope. In general the sharpness function value for a current image in a through-focus series can be computed in parallel to the acquisition of the next image in the through-focus series at no extra cost time.

Consider a vector of the sharpness function values computed for a through-focus series of K images with the k_0 -th image being in-focus

$$\mathbf{s} := (s_k)_{k=1}^K. \quad (6.12)$$

The sharpness function values are scaled between 0 and 1, i.e. $s_k \in [0, 1]$. Figure 6.3 schematically shows possible examples of the sharpness function behaviour for $K = 7$, $k_0 = 3$. The notation SF in the figure means the *Sharpness Function*. We take

$$\vartheta := \begin{cases} 1 & \text{if } \max_k s_k \in \{s_1, s_K\}, \\ 0 & \text{otherwise.} \end{cases}$$

In Figure 6.3 ϑ is equal to 1 for the sharpness function on the right and is equal to 0 for other functions.

Let δ_{acc} be the number of defocus steps between the in-focus image and the sharpness function maximum. Then the maximum number of defocus steps between the in-focus image and the sharpness function maximum is

$$\delta_{\text{acc}} \leq \delta_{\text{acc}}^{\text{max}} := \max(K - k_0, k_0 - 1).$$

Let δ_{lm} be the number of the local maxima in the sharpness function, then

$$\delta_{\text{lm}} \leq \delta_{\text{lm}}^{\text{max}} := \lfloor \frac{K-1}{2} \rfloor - 1 + \vartheta.$$

Let δ_{ran} be the monotonicity range in defocus steps around the maximum, then

$$\delta_{\text{ran}} \leq \delta_{\text{ran}}^{\text{max}} := (K - 1) - 2 + \vartheta.$$

Let δ_{noise} be the noise amplitude in the sharpness function, then

$$\delta_{\text{noise}} < \delta_{\text{noise}}^{\text{max}} := K - 3 + \vartheta.$$

The ranking results in scores which are real numbers

$$z_{\text{acc}}, z_{\text{ran}}, z_{\text{lm}}, z_{\text{noise}} \in [0, 1].$$

They are equal to 1 in the ideal situation and equal to 0 in the worst situation. We define

$$z_{\text{score}} := 1 - \frac{\delta_{\text{score}}}{\delta_{\text{score}}^{\max}},$$

where *score* denotes either *accuracy*, *local maxima*, *range* or *noise*.

For the total overall score of sharpness function values we use the weighted sum of scores described above

$$z_{\text{tot}} = \frac{1}{2}(z_{\text{acc}} + \frac{1}{3}(z_{\text{ran}} + z_{\text{lm}} + z_{\text{noise}})) \in [0, 1].$$

Equal weight is given to z_{acc} and the sum of $z_{\text{ran}}, z_{\text{lm}}, z_{\text{noise}}$. The reason for this choice is the importance of accuracy for *static autofocus* and the fact that the scores $z_{\text{ran}}, z_{\text{lm}}, z_{\text{noise}}$ are most likely correlated. If a sharpness function does not have local optima then $z_{\text{ran}} = z_{\text{lm}} = z_{\text{noise}} = 1$, and the overall score is high. However, if the maximum of such a sharpness function is far away from the in-focus position (see SF5, Figure 6.3), this sharpness function is meaningless for the autofocus. For an ideal sharpness function, such as SF1, $z_{\text{tot}} = 1$. The ranking score values for sharpness functions from Figure 6.3 are shown in Table 6.7.

The ranking systems described previously are comparative [40, 69, 73]. For instance in [40] the accuracy ranking score is equal to the difference between sharpness function maximum and the in-focus image in defocus steps. The minimum value is equal to zero and is considered to be the best. The maximum value is limited by the number of images in a through-focus series, which depends on a particular experiment. In our assessment procedure the value of each score belongs $[0, 1]$ independently of the number of images in a through-focus series.

Based on the examples presented in Figure 6.3 and in Table 6.7 we illustrate our ranking procedure.

- **Accuracy.** The accuracy score of SF1, SF2, SF4 is equal to 1, because their maxima coincide with the in-focus position. The accuracy of SF5 is equal to 0, because its maximum is as far from the in-focus position as possible. For SF3 the number of defocus steps between sharpness function maximum and the in-focus image is $\delta_{\text{acc}} = 1$, thus $z_{\text{acc}} = 1 - \frac{\delta_{\text{acc}}}{\delta_{\text{acc}}^{\max}} = \frac{2}{3}$.
- **Local maxima.** For seven images $\delta_{\text{lm}}^{\max} = 3$. Thus for SF1 and SF5 $z_{\text{lm}} = 1$, for SF2 and SF4 $z_{\text{lm}} = 0$, and for SF3 with the number of local maxima $\delta_{\text{lm}} = 1$, $z_{\text{lm}} = 1 - \frac{\delta_{\text{lm}}}{\delta_{\text{lm}}^{\max}} = \frac{1}{2}$.
- **Range.** The functions SF1 and SF5 have $z_{\text{ran}} = 1$, because they are monotone on both sides of the global maximum. The functions SF2 and SF4 have $z_{\text{ran}} = 0$, because because their monotonicity interval around

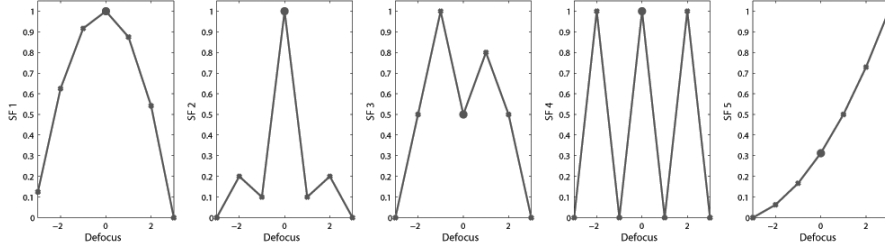


FIGURE 6.3: Benchmark examples of a general sharpness function behaviour. The notation SF denotes the *Sharpness Function*. The benchmark functions are ordered according to the ranking, i.e. SF1 scores best and SF5 - worst (see Table 6.7 for the details).

TABLE 6.1: Ranking of the sharpness function benchmarks, shown in Figure 6.3. The notation SF denotes the *Sharpness Function*.

SF	Accuracy z_{acc}	Local maxima z_{lm}	Range z_{ran}	Noise z_{noise}	Total score z_{tot}
SF1	1.00	1.00	1.00	1.00	1.00
SF2	1.00	0.00	0.00	0.45	0.58
SF3	0.67	0.506	0.25	0.68	0.57
SF4	1.00	0.00	0.00	0.00	0.50
SF5	0.00	1.00	1.00	1.00	0.50

the global maximum is minimal (two defocus steps). For SF3 this area is $\delta_{ran} = 3$ defocus steps, then for SF3 $z_{ran} = 1 - \frac{\delta_{ran}}{\delta_{ran}^{max}} = \frac{1}{4}$.

- **Noise level.** For SF1 and SF5, $z_{noise} = 1$, because they do not have local maxima and minima. For SF4, $z_{noise} = 0$, because it has as much maxima and minima as possible, with the highest sum amplitude. For SF3 the total amplitude of global minima and maxima is $\delta_{noise} = 0.8 + (1 - 0.5) = 1.3$, $z_{noise} = 1 - \frac{\delta_{noise}}{\delta_{noise}^{max}} = 1 - \frac{1.3}{4} = 0.675$.

More details about our ranking procedure can be found in [62].

6.4 Ranking results

The sharpness functions are applied to each of the images in the SEM through-focus series described in Section 6.2 and then ranked according to the procedure described in the previous section. The results are averaged for all given through-focus series. Parameterized sharpness functions are evaluated for various parameter values. For example, the derivative-based sharpness function

TABLE 6.2: The derivative-based sharpness functions (6.2)-(6.3) ranking. The squared gradient without threshold that takes into account the difference between pixels in vertical and horizontal directions scores best. Rows 1-8 represent the function (6.2). Row 9 represents (6.3). The function (6.3) has the lowest overall score.

\mathcal{N}	Sharpness function parameters				Ranking scores				
	Vert.d. v	Power p	Pix.diff. k	Thresh. Θ [%]	Accur. z_{acc}	False max. z_{lm}	Range z_{ran}	Noise z_{noise}	Overall z_{tot}
1	0	2	1...441	0	0.9767	0.9627	0.9033	0.9591	0.9592
2	1	1	1...441	0	0.9940	0.9449	0.8766	0.9485	0.9587
3	1	2	1...441	0	0.9583	0.9565	0.9033	0.9569	0.9486
4	0	1	1...441	0	0.9767	0.9140	0.8376	0.9291	0.9351
5	1	1	1	0...100	0.8773	0.8081	0.5574	0.8297	0.8045
6	0	2	1	0...100	0.8267	0.8449	0.5589	0.8471	0.7885
7	0	1	1	0...100	0.8273	0.8068	0.5856	0.8331	0.7846
8	1	2	1	0...100	0.8178	0.8506	0.5186	0.8427	0.7776
9	-	-	-	0...100	0.7809	0.8277	0.5744	0.8554	0.7667

TABLE 6.3: The Fourier transform-based sharpness functions (6.11) ranking. The Fourier transform-based sharpness function with the zero low frequency band scores best.

\mathcal{N}	Sharpness function parameters		Ranking scores				
	Low freq. band l	High freq. band h	Accur. z_{acc}	False max. z_{lm}	Range z_{ran}	Noise z_{noise}	Overall z_{tot}
1	0	1...219	0.9708	0.9472	0.9070	0.9604	0.9545
2	2	1...219	0.9734	0.9416	0.8850	0.9518	0.9498

TABLE 6.4: The statistics-based sharpness functions (6.5)-(6.6) ranking. Rows 1-4 represent the function (6.5). Row 5 represents (6.6). The autocorrelation-based sharpness scores best.

\mathcal{N}	Sharpness function parameters			Ranking scores				
	Vert.d. v	Autocor. coeff. k	Autocor. coeff. l	Accur. z_{acc}	False max. z_{lm}	Range z_{ran}	Noise z_{noise}	Overall z_{tot}
1	0	1	1...50	0.9311	0.9208	0.8587	0.9450	0.9197
2	1	1	1...50	0.9726	0.8839	0.7801	0.8868	0.9115
3	0	1...50	1	0.9264	0.8840	0.8226	0.9097	0.8992
4	1	1...50	1	0.9482	0.8579	0.7131	0.8839	0.8832
5	-	-	-	0.8713	0.9038	0.7383	0.9191	0.8625

TABLE 6.5: The histogram-based sharpness functions (6.8)-(6.10) ranking.

\mathcal{N}	Sharpness function parameters		Ranking scores				
	Sharpness function	Thresh. Θ [%]	Accur. z_{acc}	False max. z_{lm}	Range z_{ran}	Noise z_{noise}	Overall z_{tot}
1	s_{his_t}	0...100	0.8787	0.8855	0.5671	0.8811	0.8283
2	s_{his_e}	-	0.7359	0.8273	0.4769	0.8662	0.7297
3	s_{his_r}	-	0.2856	0.6944	0.1954	0.7367	0.4139

TABLE 6.6: The intensity-based sharpness functions (6.7) ranking. Squared intensity has the highest overall score. Absolute intensity has the lowest overall score.

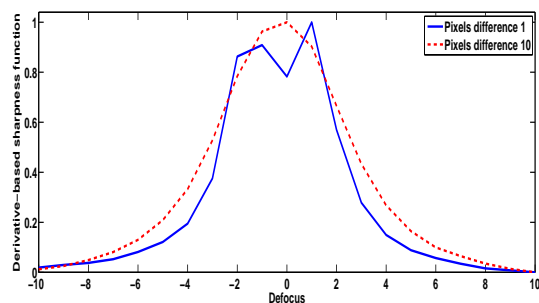
\mathcal{N}	Sharpness function parameters		Ranking scores				
	Power p	Thresh. Θ [%]	Accur. z_{acc}	False max. z_{lm}	Range z_{ran}	Noise z_{noise}	Overall z_{tot}
1	2	0...100	0.8207	0.8376	0.6294	0.8679	0.7995
2	1	0...100	0.1304	0.8110	0.3574	0.8045	0.3940

TABLE 6.7: The sharpness function (6.1)-(6.8) ranking.

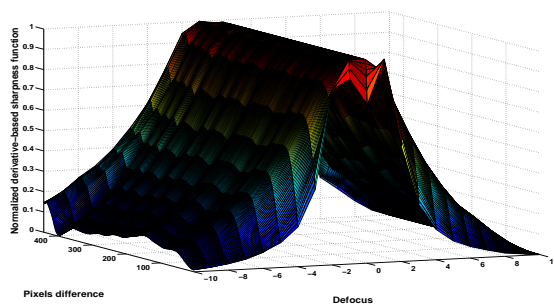
\mathcal{N}	Sharpness function family	Accuracy z_{acc}	Local maxima z_{lm}	Range z_{ran}	Noise z_{noise}	Overall score z_{tot}
1	Derivative	0.9767	0.9627	0.9033	0.9591	0.9592
2	Fourier transform	0.9708	0.9472	0.9070	0.9604	0.9545
3	Statistics	0.9311	0.9208	0.8587	0.9450	0.9197
4	Histogram	0.8787	0.8855	0.5671	0.8811	0.8283
5	Intensity	0.8207	0.8376	0.6294	0.8679	0.7995

(6.2) is applied to the gold-on-carbon through-focus series (Figure 6.2(f)) for $v = 0$, $p = 2$, $\Theta = 0$ and the pixel difference parameter $k = 1, \dots, 441$. For each k we obtain a different one-dimensional sharpness function with different z_{score} values. The results with the highest z_{tot} is chosen. The same procedure is repeated for each of the through-focus series. Ranking scores obtained for each through-focus series are averaged, and the final result can be seen in in first row ($\mathcal{N}=1$) of Table 6.2. Table 6.2 shows the performance of derivative-based sharpness functions when varying of different parameters separately (for the details see columns *sharpness function parameters*). The results for each of the 5 sharpness function families are shown in 5 different tables Tables 6.2-6.6. Table 6.7 summarizes the results according to the highest overall scores z_{tot} from Tables 6.2-6.6. The derivative-based and Fourier transform-based sharpness functions have shown the best overall performance.

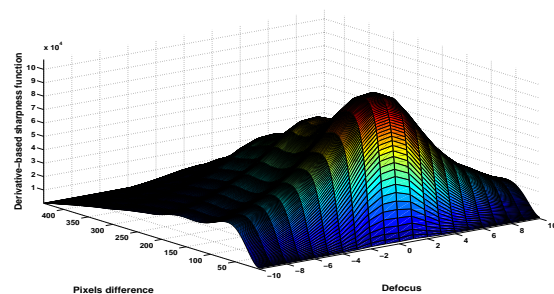
Table 6.4 shows that the variance-based sharpness function ($\mathcal{N} = 5$) scores considerably lower than the autocorrelation-based sharpness functions ($\mathcal{N} = 1, \dots, 4$) probably because we score the best (variation over several parameters) to a single variance-based score. Through Table 6.5 we can see the histogram-based sharpness function with threshold count (6.10) has much higher ranking score than *entropy* (6.9) and range (6.8). We note that entropy and range functions as well as the variance are "single-parameter" scores. Table 6.6 shows that the intensity-based sharpness function (6.7) shows a better performance for $p = 2$.



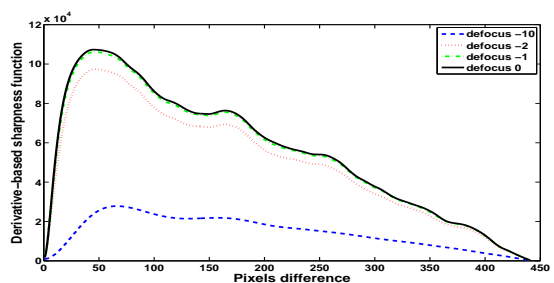
(a) Sharpness function values are scaled between 0 and 1.



(b) Sharpness function values are scaled between 0 and 1.



(c) Sharpness function values are not scaled.



(d) Sharpness function values are not scaled.

FIGURE 6.4: Derivative-based sharpness functions (6.2) computed for a gold-on-carbon through-focus series with astigmatism for $p = 2$, $\Theta = 0$.

6.5 Numerical results

The ranking of some of the sharpness functions in the previous section is obtained by maximization over a range of parameters. This section will address results for specific parameter choices.

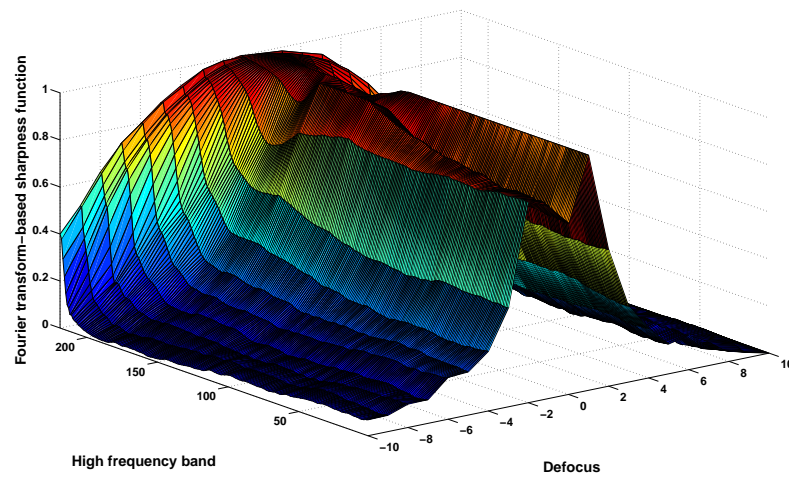
For derivative-based sharpness functions it is clear that varying the pixel difference parameter gives much better results than varying the threshold. Two sharpness functions with pixels difference parameters $k = 1, 10$, corresponding to $\mathcal{N} = 1$ in Table 6.2 are plotted for the gold-on-carbon through-focus series with astigmatism (Figure 6.2(f)). For $k = 1$ the sharpness function has a double-peak. This behaviour was discussed and demonstrated analytically for the benchmark in Section 3.4. In Figure 6.4(a) for $k = 1$, one obtains $z_{\text{tot}} = 0.7431$ due to the error in accuracy, one local maximum and a lower monotonicity range. For $k = 10$ the double-peak disappears and $z_{\text{tot}} = 1$. Figures 6.4(c) and 6.4(b) show the sharpness function surface for $k \in \{1, \dots, 441\}$ for the gold-on-carbon series with astigmatism (Figure 6.2(f)). In Figure 6.4(c) the sharpness functions are not scaled. In figures 6.4(b) and 6.4(a) the sharpness function values are scaled between 0 and 1 for each value of k . Figure 6.4(d) shows sharpness function for several defocus values plotted versus k .

The size of each image in the through-focus series is 442×442 pixels. Ranking shows that for $k \in \{6, \dots, 421\}$ $z_{\text{tot}} = 1$. The same effect is observed for most of the other experimental through-focus series: the derivative-based sharpness functions do not fail even if the pixel difference k is large. For the threshold parameter $\Theta = 0$, the power parameter $p = 2$ and $v = 0$ in (6.1), we get a sharpness function that takes into account the difference between pixels only in horizontal direction with the highest overall score from table Table 6.2 ($\mathcal{N} = 2$)

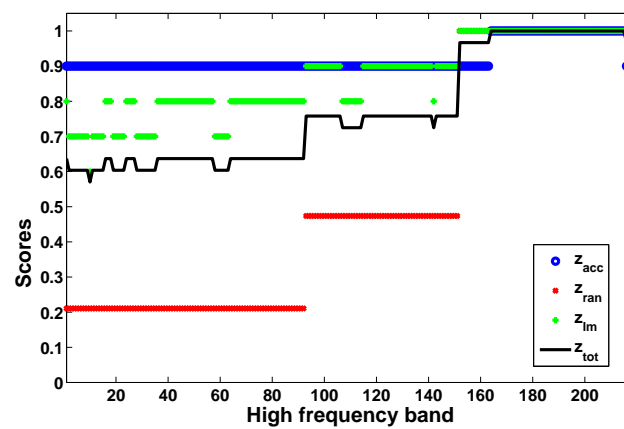
$$\begin{aligned}
 s_x^{\text{der}} &= \sum_{i=1}^N \sum_{j=1}^{N-k} (f_{i,j} - f_{i,j+k})^2 = \\
 &= \sum_{i=1}^N \sum_{j=1}^{N-k} f_{i,j}^2 + \sum_{i=1}^N \sum_{j=k+1}^N f_{i,j}^2 - 2 \sum_{i=1}^N \sum_{j=1}^{N-k} f_{i,j} f_{i,j+k}. \quad (6.13)
 \end{aligned}$$

The first two terms of (6.13) are intensity-based sharpness functions (6.7), applied to the parts of the image. The third term is a discrete autocorrelation coefficient $\check{f}_{0,k}$, defined in (5.5). When k is increasing the autocorrelation coefficient tends to zero. Thus for $k \rightarrow \infty$ the derivative sharpness function (6.13) converges to the sum of the intensity-based sharpness function values (6.7) applied to small parts of discrete image. This is clearly observed in Figure 6.4(c) and Figure 6.4(d): with increasing k the sharpness function peak-width first increases and then decreases.

It was shown in Section 3.4 that for the benchmark the derivative-based sharpness function has a local minimum in the in-focus position in the case of astigmatism. Due to the non-*amorphous* nature of most of the samples this minimum could be shifted to left or right. Because the first and the second terms



(a) Sharpness function.



(b) Ranking scores.

FIGURE 6.5: Scaled Fourier transform-based sharpness function for gold-on-carbon through-focus series with astigmatism and its ranking scores .

of (6.13) are the intensity-based sharpness function applied to different parts of an image, there is a high chance that the local minimum of both functions will be obtained in different positions. Thus the composition of the two will average the local minimum effect or even will help to get rid of it.

Figure 6.5 shows the Fourier transform-based sharpness functions for a low frequency band equal to 2 with different high frequency bands for gold-on-carbon series with astigmatism. The high frequency band is changing from 1 to 219. For certain high frequencies the similar double peak effect, as for the derivative-based sharpness function is observed. It shows the change of the ranking scores $z_{\text{acc}}, z_{\text{ran}}, z_{\text{lm}}, z_{\text{tot}}$ described in Section 6.3 due to the change of the high frequency band parameter. It clearly shows, how the sharpness function parameter varying can influence the sharpness function quality.

Figures 6.6-6.9 show the autocorrelation-based, derivative-based and intensity-based sharpness functions applied to four of the through-focus series. For instance for the autocorrelation-based sharpness function computed for the gold-on-carbon: Figure 6.7(a) shows the function itself plotted versus defocus and different values of parameter l . Similarly for the derivative-based and intensity-based sharpness functions, the parameters k and Θ are varied. Figure 6.7(d) shows the plots of the sharpness function quality ranking. The horizontal axis corresponds to the changing parameter (in this case l) and the vertical axis shows the ranking scores. Figure 6.7(d) shows the legend with the score parameter identification: blue - accuracy, red - local maxima, green - monotonicity range, cyan - robustness to noise. Most important is the black solid line, which corresponds to the total score $z \in [0, 1]$. If $z = 1$ the sharpness function has the ideal shape.

For the tin balls (see Figure 6.6) all three functions have comparably high score: the functions have a single optimum at the in-focus image. For the gold-on-carbon (see Figure 6.5) the autocorrelation-based and derivative-based functions have a double peak for the small values of parameters l and k respectively, as the result of the astigmatism present in the through-focus series. However, the intensity-based sharpness function does not show such behaviour for the same set of experimental data. This is also visible at the quality ranking plots and correspond to the theoretical observations about the intensity-based function (Section 5.2).

For the integrated circuits (see Figure 6.8) we again observe a double peak in the autocorrelation-based and derivative-based functions. The double peak of the intensity-based function is not that visible with the eyes. However, the ranking plot shows that the monotonicity range of the intensity-based sharpness function is poor as well. The possible reason is the stronger amount of astigmatism present in the through-focus series.

The images of the hard disk head are very noisy. This results in oscillations in all three sharpness functions (see Figure 6.9). We can see that the derivative-based sharpness function is most sensitive to noise. Increasing the values of the corresponding parameters l, k, Θ helps to decrease the amount of noise in the sharpness function.

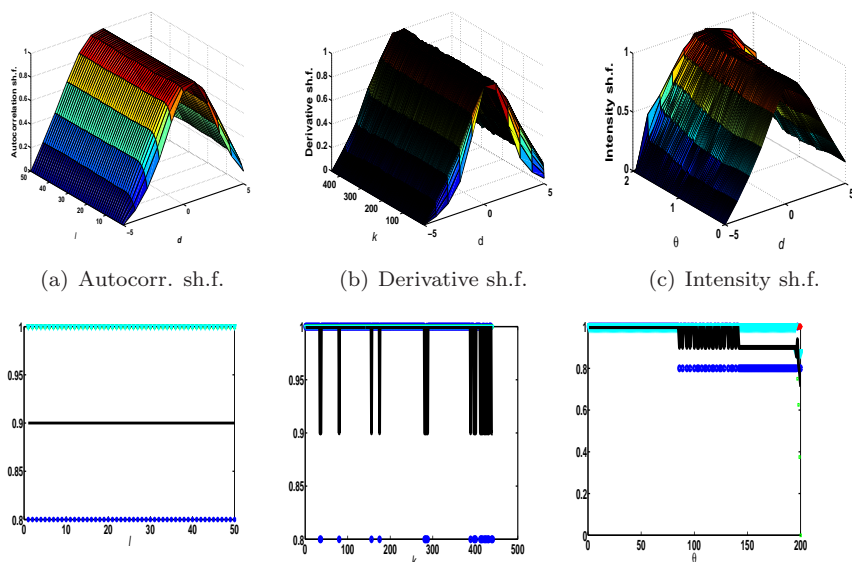


FIGURE 6.6: Sharpness functions and their scores for a tin balls through-focus series in Figure 6.1(c).

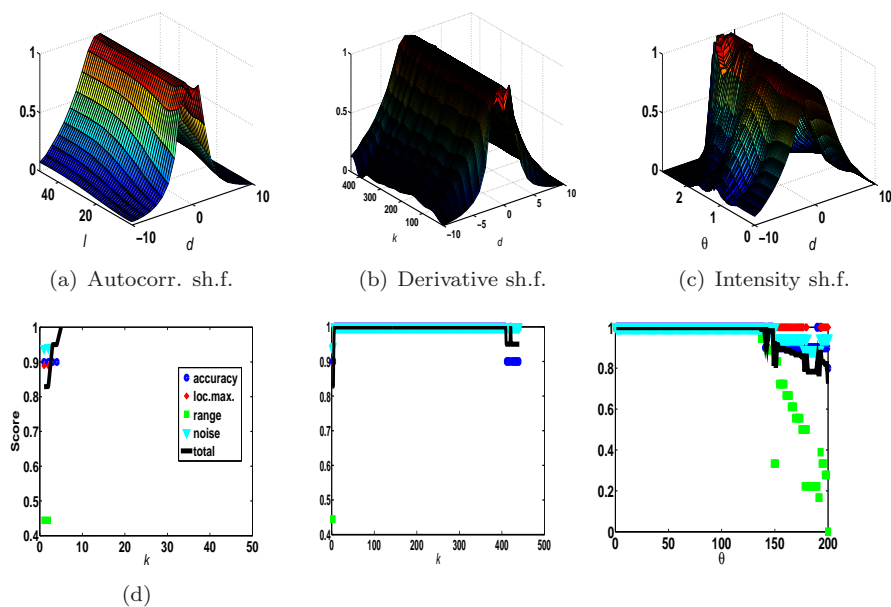


FIGURE 6.7: Sharpness functions and their quality ranking scores computed for the gold-on-carbon through-focus series with astigmatism (Figure 6.2(f)).

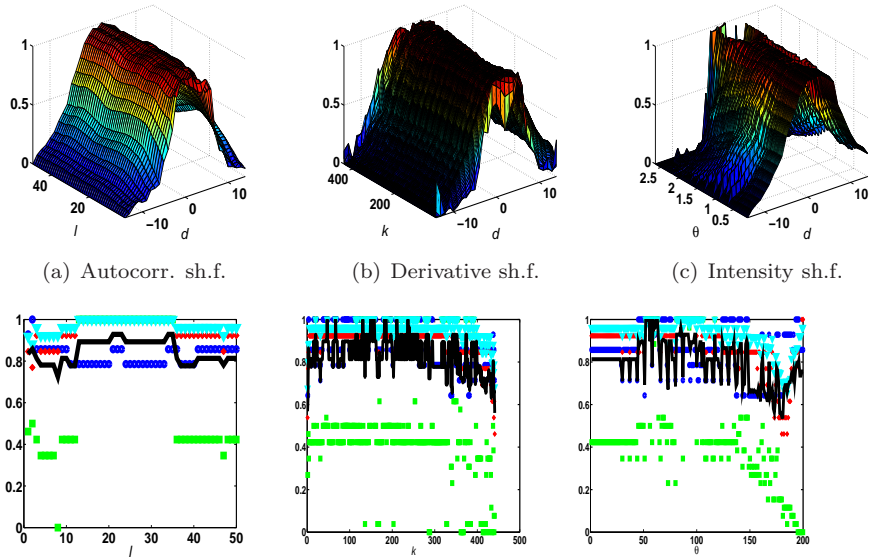


FIGURE 6.8: Sharpness functions and their quality ranking scores computed for the integrated circuits through-focus series with astigmatism (Figure 6.2(e)).

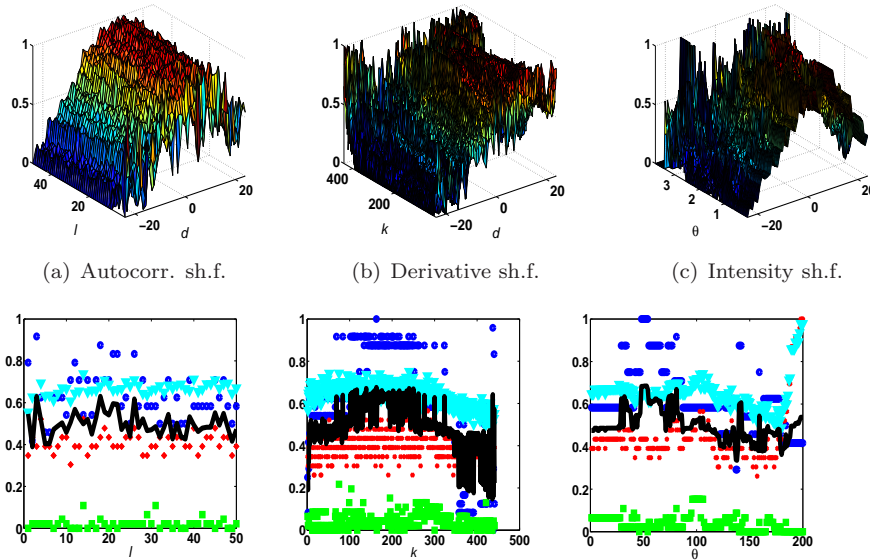


FIGURE 6.9: Sharpness functions and their quality ranking scores computed for the hard disk head through-focus series (Figure 6.1(g)).

Chapter 7

Simultaneous defocus and astigmatism correction

In this chapter we introduce a method for simultaneous autofocus and twofold astigmatism correction for application to an arbitrary sample in a low-to-medium *magnification* (i.e. resolution 1 nm or larger). This magnification is typical for many STEM applications that require automation (electron tomography [31, 35, 84, 98], Critical Dimension STEM [91]). Our method corrects defocus and astigmatism determining the optimum of a sharpness function in the space of the three microscope control variables \mathbf{d} (defocus, x-stigmator, y-stigmator). In particular we apply the variance-based sharpness function (5.9). The variance-based sharpness function has been applied earlier to the images from the various types of microscopy [84, 69, 16]. For fluorescent light microscopy it is shown experimentally that out of a large amount of sharpness functions variance is the better one [40], for non-fluorescence microscopy normalized variance is the better one [73].

This chapter consists of two sections. In Section 7.1 we describe numerical simulations based on a linear image formation model and the wave aberration point spread function explained in Chapter 2 and realistic physical values. Numerical simulations show that the variance-based sharpness function reaches its maximum at the *Scherzer defocus* point with zero astigmatism. This is demonstrated for the synthetic *amorphous* images and the ellipsoid particles image with and without noise. From observations about the behaviour of the variance-based sharpness function we draw the conclusion that it is important to optimize the function in the space of the three control variables simultaneously.

Section 7.2 discusses derivative-free optimization general properties required by simultaneous defocus and astigmatism correction. A brief summary of the Nelder-Mead simplex method and the interpolation-based trust-region method for the three-parameter optimization is provided. The results of the method's on-line performance in the STEM microscope will be illustrated in Chapter 8.

TABLE 7.1: Realistically chosen physical values of the STEM parameters C_s and E used for the numerical computations of STEM point spread function. The wavelength λ is computed through the electron voltage E . Scherzer values $\eta_{A_{Sh}}$, $R_{A_{Sh}}$, d_{Sh} are computed from (2.21), (2.23) and (2.24).

Parameter name	Notation	Value	Units
Spherical aberration	C_s	1.07	mm
Electron energy	E	300	keV
Electron wavelength	λ	1.9×10^{-3}	nm
Aperture radius	$R_{A_{Sh}}$	5.3	nm
Scherzer defocus	d_{Sh}	-55.2	nm
Convergent semi angle	$\eta_{A_{Sh}}$	10.2	mrad

7.1 Simulations

In this section we simulate the variance-based sharpness function in order to study its behaviour. In particular we show that for a given object function the image variance reaches its maximum in a three parameter space at the Scherzer defocus point with zero astigmatism (2.26).

7.1.1 Description

Figure 7.1 shows the amorphous object function (Figure 7.1(a)) and the ellipsoid particles object function (Figure 7.1(b)) simulated by procedure described in [68]. A choice for the object function of an amorphous sample of uniform thickness is a random draw from a *Poisson distribution* with mean value equal to the expected number of atoms [14]. For the second object the ellipsoid particles are dispersed on an amorphous background. All particles have parallel main axes with different widths in orthogonal directions. This results in a strong preferential direction in the image, as is evidenced by the Fourier transform (Figure 7.1(d)), differently from the Fourier transform of amorphous object function (Figure 7.1(c)), that has a circular shape. The preferential direction could cause difficulties for the Fourier transform-based correction methods, because in general we expect elliptic shape of Fourier transform in the case of astigmatism.

The size of both objects is 100×100 pixels. We assume the *pixel width* to be 0.5 nm for both objects. We use the physical model described in Section 2.4.2 in order to obtain a synthetic experimental image, i.e. a simulated object function convolved with a point spread function. The noise ε is simulated as a Poisson distribution applied to the synthetic image

Table 7.1 shows the parameter values used in the computations. We took the realistic values for the spherical aberration C_s and the electron voltage E . The wavelength λ is computed through the electron voltage E [12]. The Scherzer values $\eta_{A_{Sh}}$, $R_{A_{Sh}}$, d_{Sh} are computed from (2.21), (2.24) and (2.23). The largest defocus deviation that does not influence the image quality is known

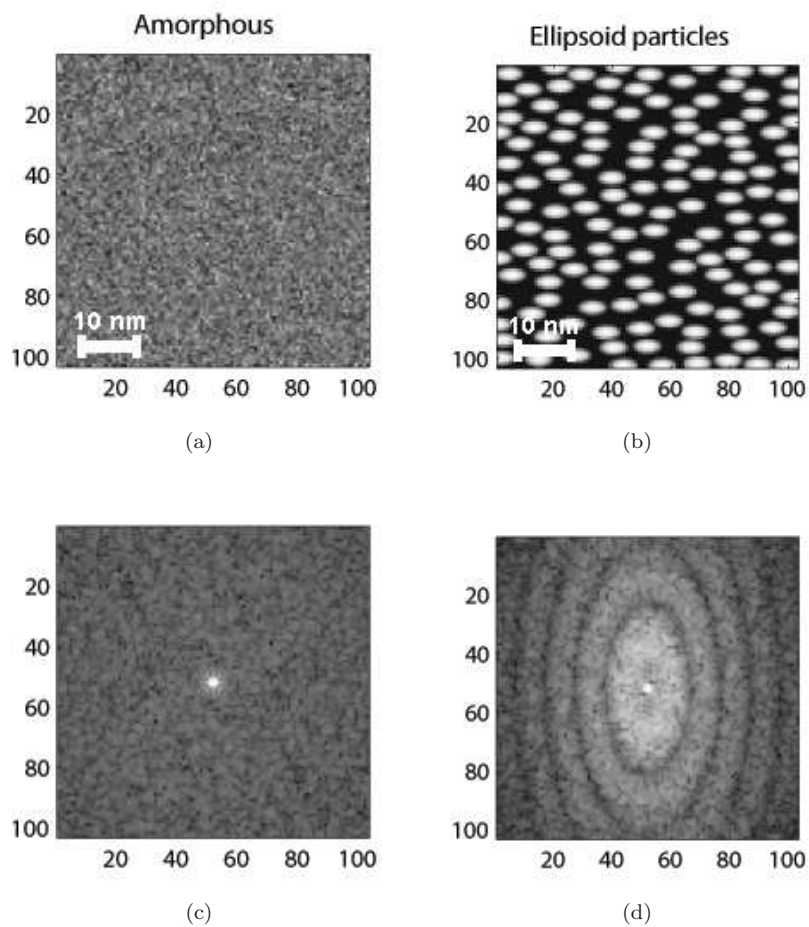


FIGURE 7.1: 7.1(a) Amorphous object function; 7.1(b) ellipsoid particles object function; 7.1(c) the Fourier transform of the amorphous object function; 7.1(d) the Fourier transform of the ellipsoid particles object function. The Fourier transform of the ellipsoid particles object function is elliptic due to the ellipticity of the particles.

TABLE 7.2: Defocus and astigmatism ranges and steps used for numerical calculations of the variance-based sharpness function domain.

Parameter name	Notation	Value	Units
Defocus step size	Δd	9.2	nm
Astigmatism step size	ΔC_b	9.2	nm
	ΔC_c	9.2	nm
Amount of defocus steps	N_d	40	-
Amount of astigmatism steps	N_{C_b}	20	-
	N_{C_c}	20	-
Defocus boundary values	$[d_{min}, d_{max}]$	[-239.2, 128.8]	nm
Astigmatism boundary values	$[C_{b_{min}}, C_{b_{max}}]$	[-92,92]	nm
	$[C_{c_{min}}, C_{c_{max}}]$	[-92,92]	nm

as a *tolerable defocus error* and is defined (cf. [68]), as

$$d_{et} := \frac{\lambda}{2\eta_{ASh}^2}. \quad (7.1)$$

For our simulation we take the defocus step equal to the tolerable defocus error

$$\Delta d = d_{et}. \quad (7.2)$$

Parameters C_b and C_c in (2.19) indicate the distance between the two focal points in the case of astigmatism. We can thus set the astigmatism step sizes equal to the defocus step sizes

$$\Delta C_b = \Delta C_c = \Delta d. \quad (7.3)$$

The numerical computations are performed for N_d defocus d steps around the Scherzer defocus d_{Sh} , N_{C_b} steps of C_b around C_{b_0} and N_{C_c} steps of C_c around C_{c_0} , i.e. the parameter domain Ω is

$$\Omega = [d_{min}, d_{max}] \times [C_{b_{min}}, C_{b_{max}}] \times [C_{c_{min}}, C_{c_{max}}], \quad (7.4)$$

where

$$d_{min} = d_{Sh} - \frac{N_d}{2}\Delta d, \quad d_{max} = d_{Sh} + \frac{N_d}{2}\Delta d, \quad (7.5)$$

$$C_{b_{min}} = C_{b_0} - \frac{N_{C_b}}{2}\Delta C_b, \quad C_{b_{max}} = C_{b_0} + \frac{N_{C_b}}{2}\Delta C_b, \quad (7.6)$$

$$C_{c_{min}} = C_{c_0} - \frac{N_{C_c}}{2}\Delta C_c, \quad C_{c_{max}} = C_{c_0} + \frac{N_{C_c}}{2}\Delta C_c. \quad (7.7)$$

The computed values are shown in Table 7.2.

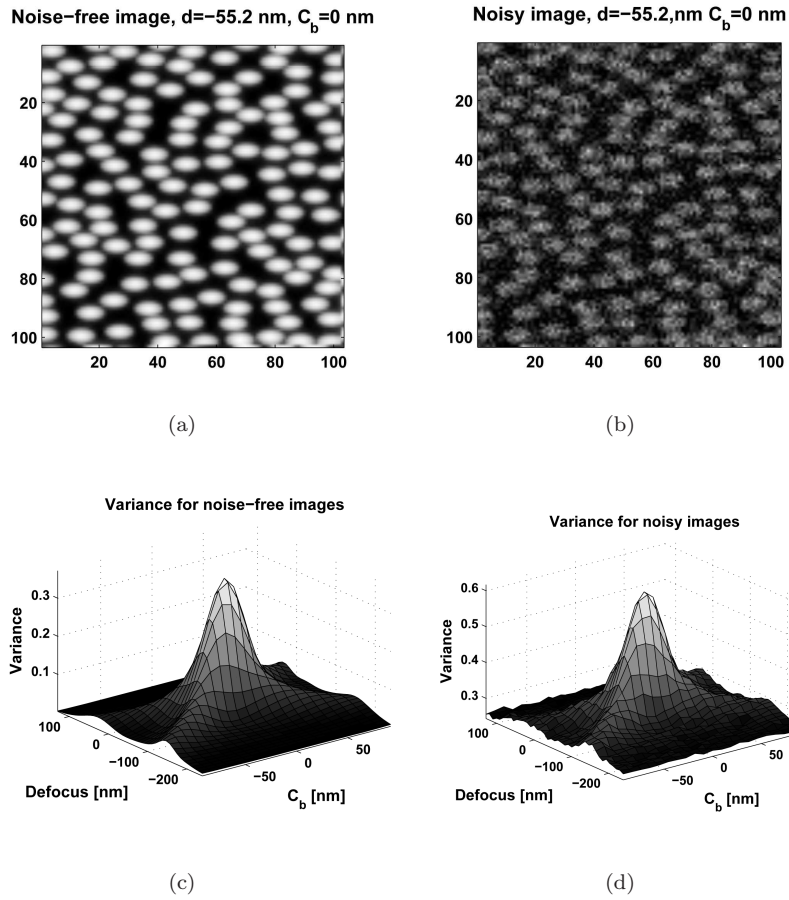


FIGURE 7.2: The synthetic ellipsoid particles image at the Scherzer defocus point with zero astigmatism 7.2(a): without noise; 7.2(b): with noise. The variance-based sharpness function computed for synthetic ellipsoid particles images 7.2(c): without noise and 7.2(d): with noise.

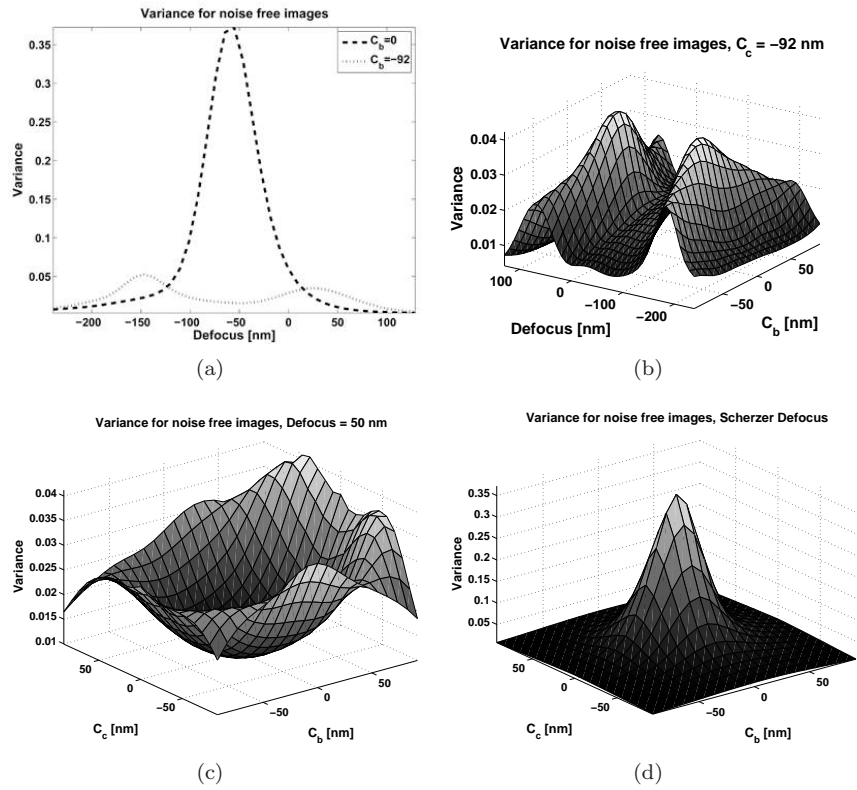


FIGURE 7.3: Graphs of the variance-based sharpness function for the ellipsoid particles images without noise. 7.3(a): For $C_b = 0$ nm, the variance-based sharpness function reaches its maximum at the Scherzer defocus point, while for $C_b = -92$ nm variance has a local minimum at the Scherzer defocus point and two maxima on the left and the right. The variance-based sharpness functions plotted in two-parameter spaces for the fixed values of the third parameter 7.3(b): $C_c = -92$ nm; 7.3(c): $d = 50$ nm; 7.3(d): $d = d_{Sh}$ (Scherzer defocus).

7.1.2 Numerical results

Four numerical experiments are carried out: 1) amorphous without noise; 2) amorphous with noise; 3) ellipsoid particles without noise; 4) ellipsoid particles with noise. In all four cases the simulated variance-based sharpness function reaches its maximum at the Scherzer defocus point with zero astigmatism within the defocus and astigmatism errors (7.2)-(7.3).

Figure 7.2 shows the results of numerical computations for the ellipsoid particles object without and with noise. The figure shows the variance functions in the space of parameters d and C_b for a fixed $C_c = 0$. In our simulation in the noise-free case the variance function has no other maxima. We can clearly see that when noise is added to the images, the variance function becomes noisy as well, i.e. it may have a lot of local optima, especially further away from the global optimum. A similar shape of the variance function is obtained for the amorphous object.

It is important to note that the variance has local optima in one-parameter and two-parameter spaces, which are no longer the optima in the encompassing three-parameter space. For instance the variance has a local *minimum* at the Scherzer defocus point with the astigmatism $C_b \gg 0$ (figures 7.3(a)-7.3(b)). This local minimum is not a minimum anymore in the two-parameter space (Figure 7.2(c)). This corresponds to the earlier observations: in [16] it is shown analytically for SEM that the variance function has local optima in one-parameter spaces. In [16] the symmetry between the two local optima in case of an amorphous object is used for the calibration of defocus and astigmatism. In the general case of a non-amorphous object the positions of the two local maxima are not exactly symmetric (see for instance Figure 7.2(a)). This was also shown in experiments for SEM [62, 61]. Similar phenomena are observed in a two-parameter space: variance as a function of C_b and C_c reaches a local minimum at the zero astigmatism if the defocus value is far away from the Scherzer point (figures 7.3(c)-7.3(d)). These minima are not minima anymore in a three-parameter space. From the above observations that the variance reaches its maximum at the Scherzer defocus point and zero astigmatism, and from the existence of the local maxima with unpredictable behaviour in lower dimensions we draw the conclusion that it is important to optimize the sharpness function in the three-parameter space at the same time.

7.2 Derivative-free optimization

A few things are worth mentioning for the optimization of the sharpness function S in electron microscopy

- The sharpness function S might have local optima due to noise in the image formation;
- The sharpness function S might have local optima due to the objects' geometry ψ ;

- At different time points t_1, t_2 we may have

$$S_{t_1}(\mathbf{d}) \neq S_{t_2}(\mathbf{d}) \quad (7.8)$$

not only due to noise in the image formation, but also due to instabilities, such as sample drift, sample contamination and hysteresis of the magnetic lens;

- Determining the sharpness function value is much faster than an the *recording of an image*. Therefore, and also because repeated recordings can damage or destruct the sample, one has to optimize the defocus and astigmatism with just few recordings (few function evaluations);
- Analytical derivative information is not available. Calculating (approximate) derivatives with *finite differences* would dramatically increase the amount of necessary image recordings, and is therefore not a serious option. Besides it would not be very accurate due to noise.

The *Nelder-Mead simplex method* is designed to find a local optimum of a function (in our case it is the variance-based sharpness function). It makes no assumptions about the shape of the function and does not use derivative information. The Nelder-Mead simplex method has successfully been applied before to industrial problems, where derivative-free optimization is necessary (see for instance [79]). A detailed description of the Nelder-Mead simplex method for n dimensions can be found in [10, 58].

A *simplex* is a geometrical object in n dimensions that consists of $n + 1$ points. In the case of defocus and twofold astigmatism correction $n = 3$, and the simplex is a tetrahedron. The Nelder-Mead simplex method requires $n + 1$ points, defining an initial simplex to start. Let the initial value of the defocus and astigmatism parameters be \mathbf{d}_{init} . Then the column entries of the initial simplex $\mathbf{D} \in \mathbb{R}^{n \times n+1}$ are constructed as

$$\mathbf{D}^i = \mathbf{d}_{init} + \Delta \mathbf{d}_i \mathbf{e}_{i-1}, \quad i = 1, \dots, 4, \quad (7.9)$$

where $\mathbf{e}_0 = 0$, $\mathbf{e}_i, i = 1, \dots, 3$ is a unit vector, and

$$\Delta \mathbf{d} \in \mathbb{R}^3 \quad (7.10)$$

is the input parameter for the optimization.

During every *iteration* the Nelder-Mead simplex method first *evaluates* the variance function for a finite number of points (between 1 and $n + 2 = 5$). In our case one *function evaluation* corresponds to one *image recording* and the computation of the sharpness function values. It means that between 1 and 5 images may be recorded during one iteration of the Nelder-Mead simplex method. Then the vertex corresponding to the lowest function value in a simplex is replaced by a new point with a higher function value. The algorithm stops if for a *stopping tolerance* criterion d_{tol}

$$\|\mathbf{d}^{(1)} - \mathbf{d}^{(2)}\| < d_{tol}, \quad (7.11)$$

where $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$ are the two points of the simplex, corresponding to the highest variance values. We have to keep in mind that the *defocus* and *stigmator* control variable values in a real-world machine have different scalings. Thus we actually have to separate the tolerances for different dimensions. The second method we consider is the *interpolation-based trust-region method* (or the *Powell method*), following Powell's UOBYQA (Unconstrained Optimization BY Quadratic Approximation) [57]. The interpolation-based trust-region method employs a series of local quadratic models that interpolate m sample points of the sharpness function. The quadratic model is constructed using standard Lagrangian basis functions and is minimized within the trust-region to obtain a trial point. The repeated quadratic model minimization converges to the nearest local minimum [60]. Each computed trial point is accepted or rejected based on an iteration. Each trust-region iteration replaces at most one interpolation point in the set, keeping m fixed. In other words, at most only one function evaluation per iteration is required. When an interpolation point in the set is replaced by a new one, the Lagrangian functions are adapted as well, leading to a new quadratic model of the sharpness function. The iterative process is stopped if the length of the trial step is below the *stopping tolerance* or the maximum allowable number of function evaluations is reached. The detailed description of Powell's UOBYQA can be found in [56, 57]. In the next chapter we show for a real-world application that for an image of a general sample the algorithms converge to the optimal value of the control variables in typically 20-50 image recordings.

Chapter 8

On-line STEM application

The automated defocus and astigmatism correction methods described in the previous chapters are implemented and tested *on-line*. In the previous chapters all outputs were computed *off-line*, i.e. for the data recorded in advance. The *on-line* testing described in this chapter implies that the algorithms *run* (i.e. the algorithms are executed) in a control loop with the electron microscope.

8.1 The experimental set-up

The automated defocus and astigmatism correction methods described in the previous chapters are implemented and tested *on-line*. Figure 8.1 shows the three main parts of the experimental set-up:

- The prototype of a FEI Tecnai F20 STEM acquires the images.
- A Java-based experimental platform (called *EXPLA*), which consists of a core that connects to the *TEMScripting* interface for a FEI microscope control, and an application control framework. *EXPLA* is used to extract the images from the microscope for the specific control settings of the microscope. It can be controlled using Matlab [68].
- The *particle analysis algorithm* includes as a substep the automated defocus and astigmatism correction. All the methods are implemented in Matlab V7.5 (R2007b).

The goal of the particle analysis algorithm is to analyse the nanoparticle distribution, in particular to count the amount of specific particles that only occur sparsely in another carrier, and to estimate the statistical distribution of their locations and sizes. The algorithm records images at different stage positions and magnifications. The ideal control variables (*defocus* and *stigmators*) value \mathbf{d}_0 depends on the machine control changes (stage position and magnification) as well as on the sample and environment instabilities. If the image is *stigmatic* or *out-of-focus*, the particle analysis algorithm might give errors. For this reason it is important to run the algorithms of automated defocus and astigmatism correction with a certain time periodicity.

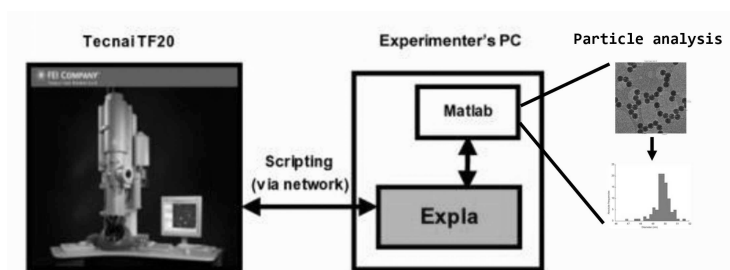


FIGURE 8.1: Experimental set-up. The diagrams are taken from [15].

Within the *Condor project* the experimental set-up is referred to as a *Concept Car* [15]. This chapter describes and discusses the outputs obtained within the Concept Car framework, related to the automated defocus and astigmatism correction. It involves three different algorithms, related as shown in Figure 8.2. The three algorithms are numbered according to the list below.

1. The *focus series algorithm* (introduced in Section 4.5). It is applied to improve the starting guess \mathbf{d}_{init} , which is used to determine the initial simplex (7.9). The output of one of the runs of the algorithm is described in Section 8.2.
2. The simultaneous defocus and astigmatism correction algorithm. This algorithm is based on the three-parameter derivative-free optimization, described in Chapter 7. It turns out that the algorithm successfully finds proper control variable values at a speed and with an accuracy comparable to a human operator (Section 8.3).
3. The *fast autofocus* algorithm. This algorithm implements the derivative-based approach described in Chapter 3. The ideal defocus control variable value d_0 in electron microscopy is drifting much faster than the ideal stigmator control variables values d_{x_0}, d_{y_0} . For this reason the fast autofocus algorithm must be executed more often than the simultaneous defocus and astigmatism correction. It also requires much less *image recordings* than the simultaneous defocus and astigmatism correction. The outputs of the algorithm's runs are described in Section 8.4.

In the runs described below the *electron voltage* of the microscope is taken 200 keV, *camera length* - 200 nm, *spot size* - 7, *dwell time* - 5 μs . If not mentioned otherwise, the image sizes in each run are 256×256 pixels, the pixel depth is 16 bits. The defocus values are given either in μm or nm . Stigmator values are given in non-physical machine units (*arbitrary units*). The runs take place at low-to-medium magnifications. We premultiply the discrete image by the *window function* (4.3) before the *discrete Fourier transform* computation. All power spectra shown are in logarithmic scale.

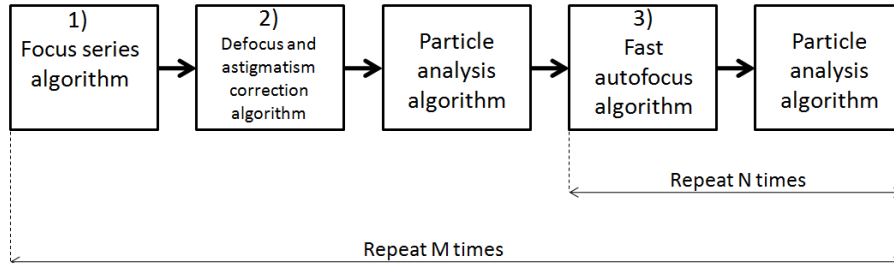


FIGURE 8.2: The simplified workflow diagram of the particle analysis algorithm and defocus/astigmatism correction.

8.2 The focus series algorithm

Figure 8.3 contains a STEM image used by the particle analysis algorithm. The image power spectrum and the Fourier transform-based *sharpness* and *stigmatic operators* (4.6)-(4.8) are displayed. The outputs shown in Figure 8.3 are similar to those described in Section 4.8.2 for SEM. The only difference is that the computations are running on-line for a real-world application, instead of off-line data recorded in advance. From the behaviour of the sharpness/stigmatic operators we can see that the image is strongly stigmatic and the microscope requires adjustments of both stigmator control variables (as explained in Section 4.4).

Figure 8.4 shows Fourier transform-based sharpness and stigmatic functions computed for the through-focus series of the images with particles (the image shown in Figure 8.3 is one of the images from this through-focus series). The functions give a suggestion for the initial step for the simultaneous defocus and astigmatism optimization: the optimum of the sharpness function points to the defocus values which is close to ideal. The shapes of the stigmatic functions indicate the directions of change of the stigmator control variables, as was explained in Section 4.5. For illustration purpose, the number of the images in the through-focus series is 41 (too high for a real-world application). Most on-line runs use around 9 images.

8.3 Simultaneous defocus and astigmatism correction

In this section we describe the outputs of the simultaneous defocus and astigmatism correction algorithm, which is based on the sharpness function optimization introduced in Chapter 7. In order to obtain in-focus astigmatism-free images the variance-based sharpness function (5.13) is optimized with the *Nelder-Mead simplex method* in a three-parameter space of defocus and stigmators control variables. The *Nelder-Mead simplex method* implementation is the modified

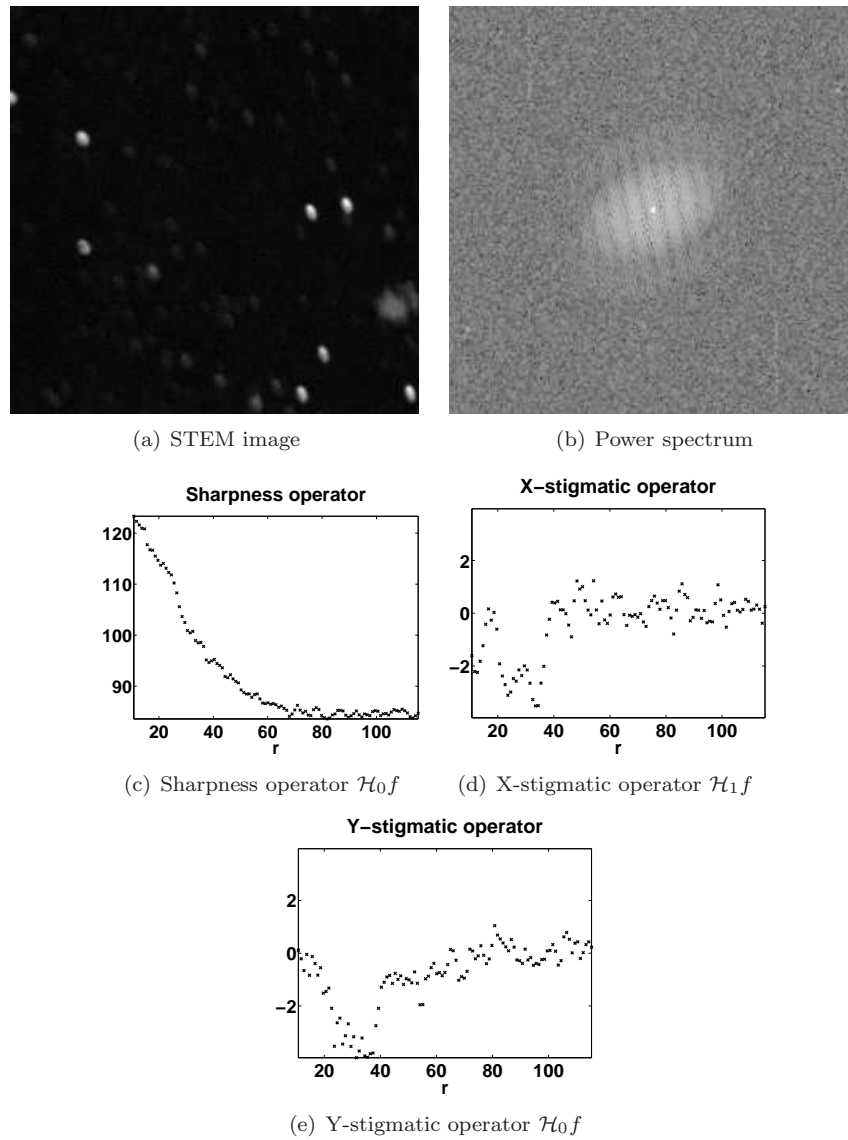


FIGURE 8.3: The output of the focus series algorithm. The Fourier transform-based sharpness and stigmatic operators (4.6)-(4.8) are computed for the STEM image.

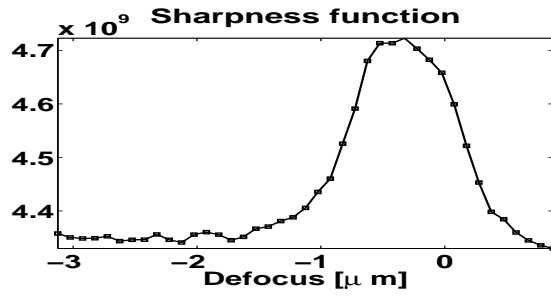
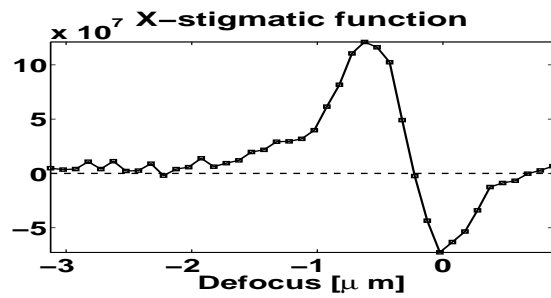
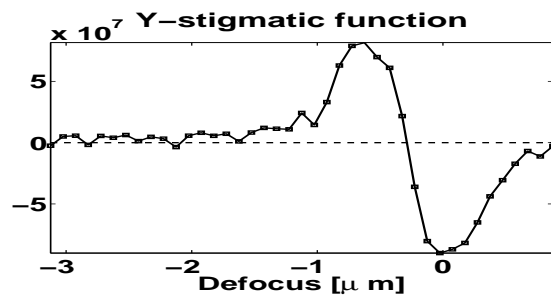
(a) Sharpness function S_0^{ft} (b) X-stigmatic function S_1^{ft} (c) Y-stigmatic function S_2^{ft}

FIGURE 8.4: The output of the focus series algorithm. The Fourier transform-based sharpness and stigmatic functions (4.17) are computed for a STEM through-focus series.

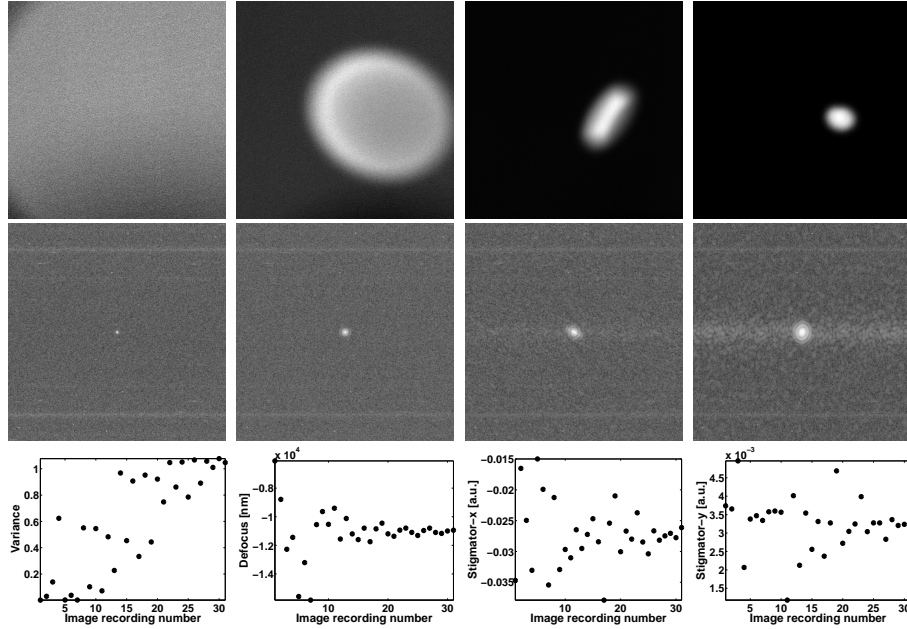


FIGURE 8.5: Simultaneous defocus and astigmatism correction for a *gold particle* with a radius of about 20 nm. The first row, from left to right shows 1st function evaluation, 6th function evaluation, 17th function evaluation, 31st function evaluation; the second row contains the Fourier transforms of the four images; the third row displays the image variance, defocus, x-stigmator and y-stigmator values plotted versus the numbers of image recordings.

version of the code available in the Matlab Optimization Toolbox V3.1.2 (*fminsearch* function). We provide the outputs of the algorithm for gold particle samples (Subsection 8.3.1) and the carbon cross grating sample (Subsection 8.3.2), which are designed for microscope calibration. Subsection 8.3.3 describes the experimental comparison of the Nelder-Mead simplex method and the Powell interpolation-based trust-region method in the content of the simultaneous defocus and astigmatism correction.

8.3.1 Gold particle samples

In this subsection we discuss three outputs of the simultaneous defocus and astigmatism correction algorithm for three gold particles samples. The image size in each run is 512×512 pixels.

First the simultaneous defocus and astigmatism correction is performed for the out-of-focus stigmatic image of a gold particle with a radius of 20 nm shown in Figure 8.5. As mentioned in Section 7.2 during defocus and astigmatism optimization one *function evaluation* corresponds to one *image recording* and the

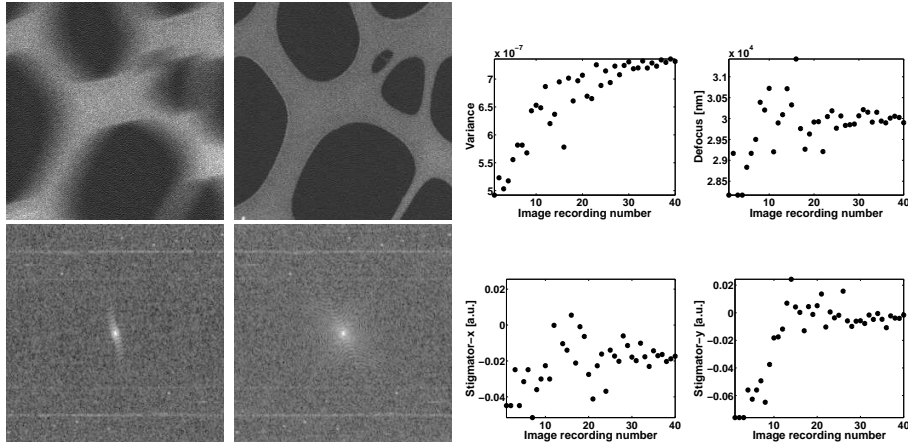


FIGURE 8.6: Simultaneous defocus and astigmatism correction for a *gold particle* sample. The figure shows images before and after optimization, their Fourier transforms, and the values of image variance, defocus, x-stigmator and y-stigmator plotted versus the numbers of image recordings.

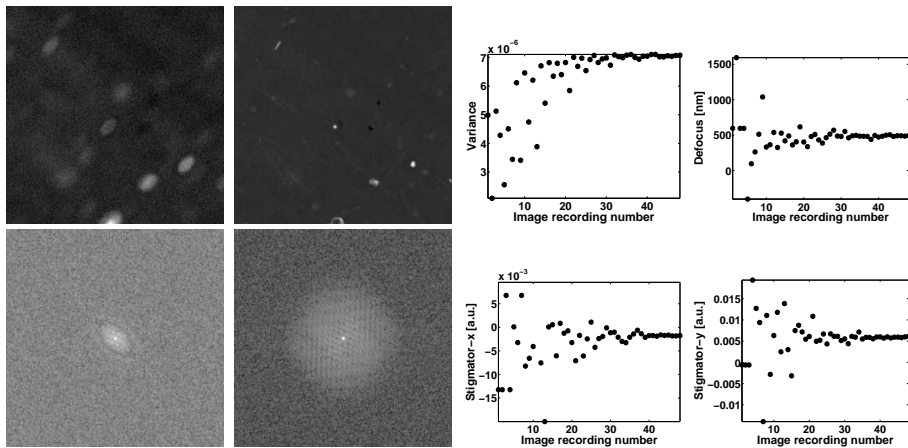


FIGURE 8.7: Simultaneous defocus and astigmatism correction for a *gold particle* sample. The figure shows images before and after optimization, their Fourier transforms, and the values of image variance, defocus, x-stigmator and y-stigmator plotted versus the numbers of image recordings.

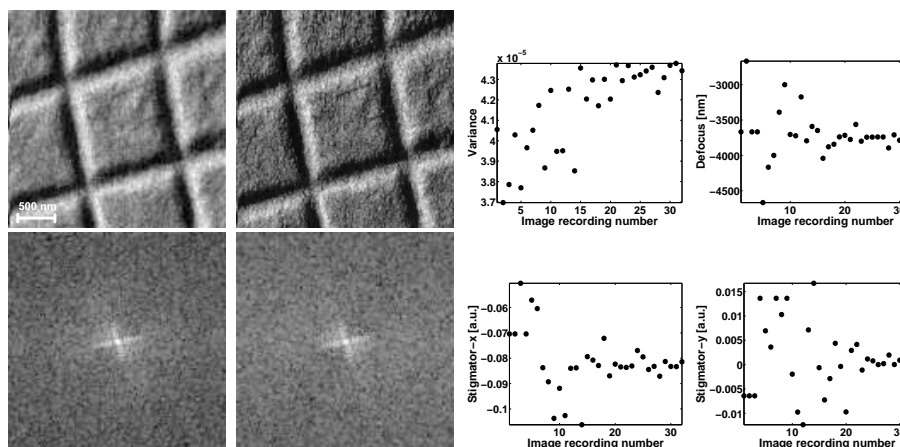


FIGURE 8.8: Simultaneous defocus and astigmatism correction for a *carbon cross grating* sample at the magnification $40\,000\times$. The figure shows images before and after optimization, their Fourier transforms, and the values of image variance, defocus, x-stigmator and y-stigmator plotted versus the numbers of image recordings.

computation of the sharpness function values. The figure shows the images of four image recordings (1st, 6th, 17th, 31st), image variance, defocus and stigmator values plotted versus image recording numbers. The Fourier transforms of the images only have a few Fourier components. For this reason the sample would be difficult to analyze with Fourier transform-based techniques. The Nelder-Mead simplex method converges in 31 function evaluations. The image quality at the last function evaluation is equivalent to the one obtained by a human operator. The accuracy is estimated visually as well as by comparing the image variance values. The original gold particle is a bit elliptic, which causes the Fourier transform of the final image to be a bit elliptic as well.

Figures 8.6 and 8.7 show outputs for the different gold particles samples. Each figure displays images before and after optimization, their Fourier transforms, and the values of image variance, defocus, x-stigmator and y-stigmator plotted versus the numbers of image recordings. The images in Figure 8.6 show mostly a carrier and a small gold particle on the left bottom. The run shown in Figure 8.7 requires a larger number of function evaluations due to a smaller *stopping tolerance* (7.11) value.

8.3.2 Carbon cross grating sample

This subsection describes five outputs of simultaneous defocus and astigmatism correction runs for a carbon cross grating sample at five different *magnifications*. The *pixel width* in the microscope changes linearly with the inverse magnification. The magnifications and pixel widths corresponding to each of the runs

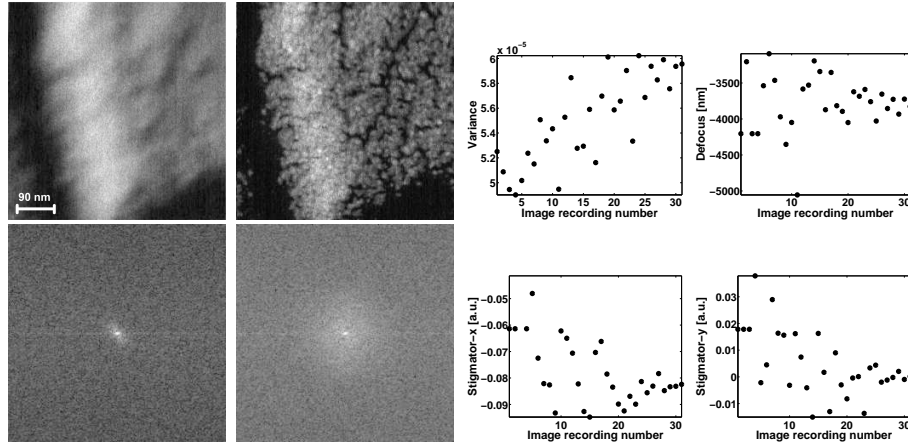


FIGURE 8.9: Simultaneous defocus and astigmatism correction for a *carbon cross grating* sample at the magnification $225\,000\times$. The figure shows images before and after optimization, their Fourier transforms, and the values of image variance, defocus, x-stigmator and y-stigmator plotted versus the numbers of image recordings.

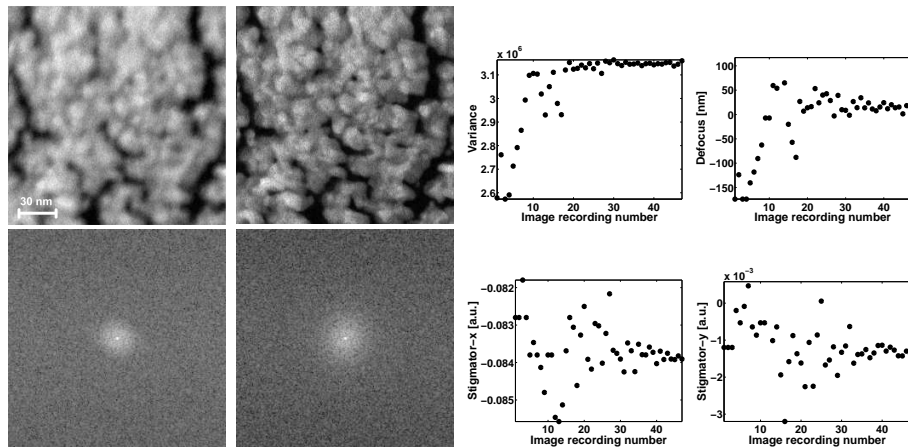


FIGURE 8.10: Simultaneous defocus and astigmatism correction for a *carbon cross grating* sample at the magnification $630\,000\times$. The figure shows images before and after optimization, their Fourier transforms, and the values of image variance, defocus, x-stigmator and y-stigmator plotted versus the numbers of image recordings.

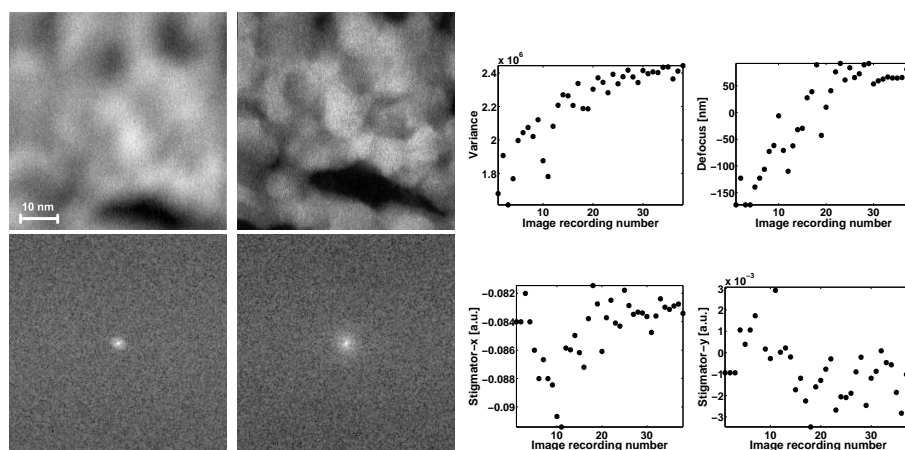


FIGURE 8.11: Simultaneous defocus and astigmatism correction for a *carbon cross grating* sample at the magnification $1\,800\,000\times$. The figure shows images before and after optimization, their Fourier transforms, and the values of image variance, defocus, x-stigmator and y-stigmator plotted versus the numbers of image recordings.

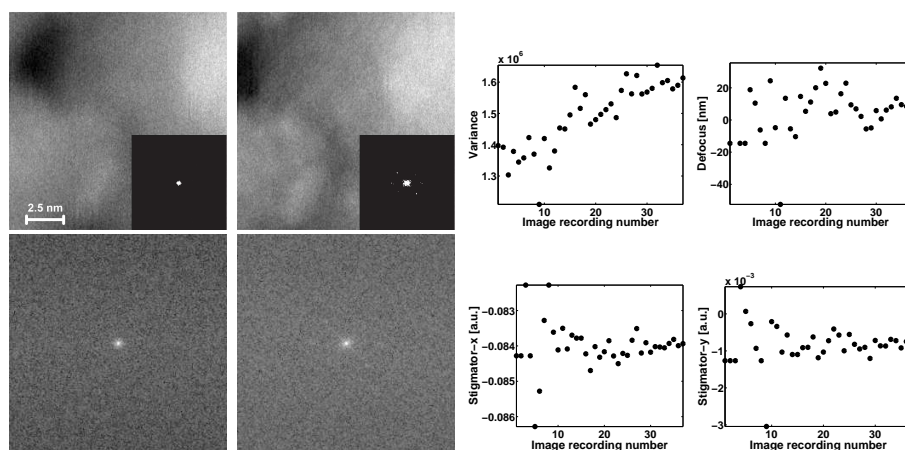


FIGURE 8.12: Simultaneous defocus and astigmatism correction for a *carbon cross grating* sample at the magnification $7\,200\,000\times$ (*high-resolution* imaging mode). The figure shows images before and after optimization, their Fourier transforms, and the values of image variance, defocus, x-stigmator and y-stigmator plotted versus the numbers of image recordings. The same Fourier transforms but with a threshold are shown above the images.

TABLE 8.1: A summary of the *magnifications* and the *pixel widths* corresponding to the five runs of the simultaneous defocus and astigmatism correction algorithm for the carbon cross grating sample (figures 8.8-8.12). The pixel width in the electron microscope changes linearly with its magnification.

\mathcal{N}	Magnification	Pixel width [nm]
1	40 000 \times	10.5
2	225 000 \times	1.9
3	630 000 \times	0.6596
4	1 800 000 \times	0.2309
5	7 200 000 \times	0.0577

are summarized in Table 8.1. Figures 8.8-8.12 show the output of the five runs, which correspond to the rows $\mathcal{N}=1-5$ in Table 8.1. Magnification increases from figure to figure as a consequence.

In the first run (Figure 8.8) the Fourier transform of the images is strongly influenced by crosses from the sample's geometry. It is difficult to detect differences between the Fourier transforms before and after improvement, though differences are visible in the real space: the corrected image shows more details than the image before the correction.

In the next run (Figure 8.9) the original image is strongly stigmatic. The presence of astigmatism shows as stretching in the image itself and as elliptic shape of its Fourier transform. It is important to note that the Fourier transform of the original stigmatic image has only a few frequency components, which might be difficult to analyze with Fourier transform-based methods.

In run three (Figure 8.10) the *stopping tolerance* of the Nelder-Mead simplex method (7.11) is lowered. Here Nelder-Mead requires 47 function evaluations, which is more than in the previous runs. However, during the last 15 recordings the variance value is almost constant, it does not increase anymore. A proper choice of the stopping tolerance is important in order to achieve a high-quality image while simultaneously avoiding unnecessary function evaluations.

In run four (Figure 8.11) the stopping tolerance (7.11) value is increased. The algorithm requires 35 function evaluations. The maximum value of the variance function is obtained in the last recording. Probably if the stopping tolerance would be smaller, the image quality could be improved even further. As in the previous runs we do not observe a large difference between the Fourier transforms of the images before and after the improvement.

The last run (Figure 8.12) is executed in the *high-resolution* mode of the electron microscope (the mode that allows imaging at atomic scale). Fourier transforms with a threshold equal to 164 [pixel gray value] are displayed above the images to illustrate the image quality improvement. In general we remark that running the algorithm in the high-resolution mode is tricky due to instabilities caused by the environment and/or the microscope (for instance, image drift and contamination). In this mode the object can be strongly displaced and con-

TABLE 8.2: Nelder-Mead vs. Powell: the experimental comparison for the simultaneous defocus and astigmatism correction.

\mathcal{N}	Magnification	Powell		Nelder-Mead		Ratio R
		Improvement v	Evaluations E	Improvement v_{NM}	Evaluations E_{NM}	
1	10000×	3.16	22	2.5	49	2.81
2	10000×	7.45	25	4.15	66	4.74
3	10000×	10.38	24	10.8	35	1.4
4	40000×	2.03	20	2.24	31	1.4
5	40000×	8.09	27	8.16	28	1.03
6	160000×	20.49	18	21.71	23	1.21

taminated between two image acquisitions, which influences the sharpness function to be optimized. The application of the algorithm to the high-resolution images should be the topic of the future research.

8.3.3 Nelder-Mead vs. Powell

In this subsection we describe six experiments, summarized in Table 8.2. Each experiment consists of two runs of the simultaneous defocus and astigmatism correction algorithm. The first run uses the *Nelder-Mead simplex method* (or the Nelder-Mead method), the second run uses the *interpolation-based trust-region method* (or the Powell method). The methods are executed for the same starting point $\mathbf{d}_0 \in \mathbb{R}^3$ with identical initial parameters. The initial *simplex* for the Nelder-Mead method, consisting of $n + 1 = 4$ points, is a subset of initial trust-region for the Powell method, consisting of $\frac{1}{2}(n + 1)(n + 2) = 10$ points.

In all cases summarized in Table 8.2 both methods turn out to converge, i.e. they both correct defocus and astigmatism successfully. However, the number of required *function evaluations* is different. The first column of Table 8.2 indicates the experiment number, the second column shows the related magnification of the microscope. The columns *Improvement* correspond to the relative variance improvement

$$v := \frac{S_{final} - S_{initial}}{S_{initial}} \quad (8.1)$$

for Nelder-Mead and Powell methods respectively. In (8.1) $S_{initial} := S(\mathbf{d}_0)$. We take into account the relative value of variance change (8.1) instead of the absolute value, because the absolute value changes due to noise and instabilities (7.8). The columns *Evaluations* indicate the number of function evaluations for each of the methods. The final ratio is computed as a ratio of the average sharpness function growth per *iteration* for both methods

$$R := \left(\frac{v_P}{E_P}\right) / \left(\frac{v_{NM}}{E_{NM}}\right). \quad (8.2)$$

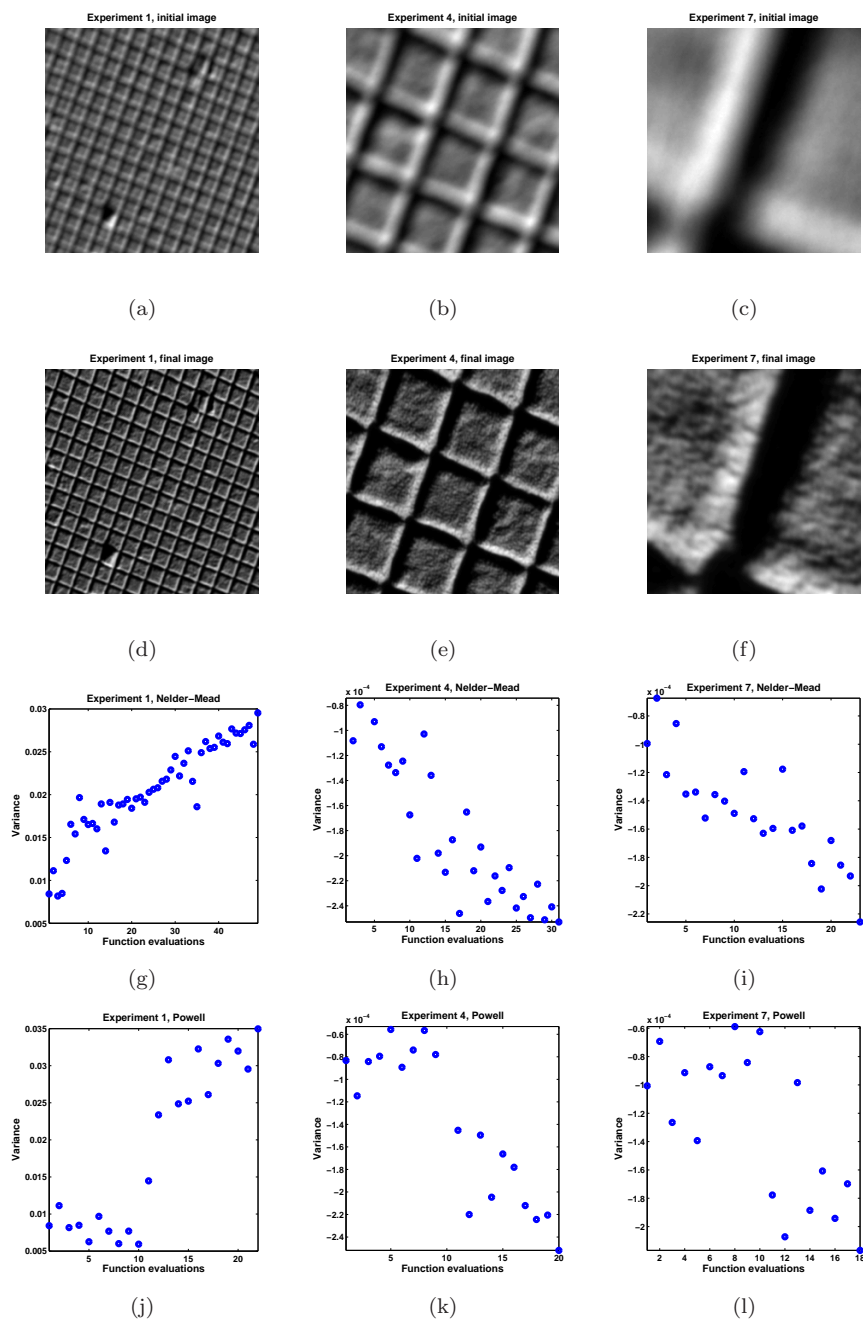


FIGURE 8.13: Simultaneous defocus and astigmatism correction. Illustration for experiments 1,4,7 from Table 8.2.

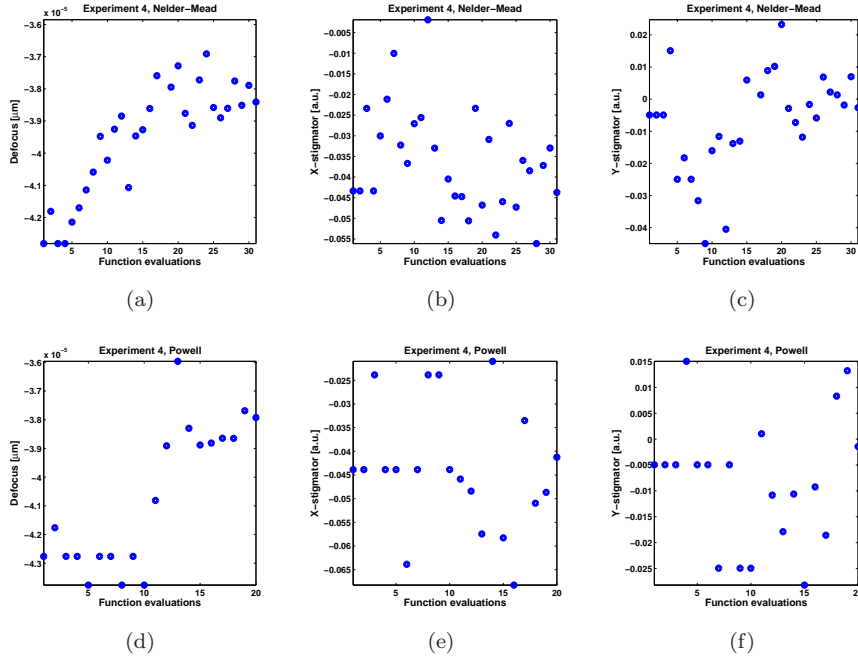


FIGURE 8.14: Simultaneous defocus and astigmatism correction for the *carbon cross grating* sample. Parameters changes in Experiment 4 from Table 8.2.

We can see that in all six cases $R > 1$, which indicates that the Powell method shows a better performance than the Nelder-Mead method, i.e. it requires fewer function evaluations in order to achieve the same quality image as the Nelder-Mead method. In some cases this difference is large (Experiment 2) and in other the performances of the algorithms is almost equal (Experiment 5).

Figure 8.13 illustrates the performance of the algorithms for the experiments 1,4,7 from Table 8.2. Figures 8.13(a), 8.13(b), 8.13(c) show out-of-focus, stigmatic images of a carbon cross grating sample before the optimization. For every experiment both the Powell method and the Nelder-Mead method are executed and lead to the similar final outputs. Figures 8.13(d), 8.13(e), 8.13(f) show images after optimization. The change of variance versus function evaluations for Nelder-Mead is shown in figures 8.13(g), 8.13(h), 8.13(i) and for Powell in figures 8.13(j), 8.13(k), 8.13(l). These plots clearly show that though Powell requires a larger number of function evaluations to start up it reaches the optimum faster. We can observe that the variance magnitude (along the vertical axes of the plots) changes ten times, when magnifications changes from 10000 \times to 40000 \times . It deals with the fact that the observed sample's geometry changes. It is one of the reasons why it is important to observe relative change in variance, while the absolute value of it does not play any role.

Figure 8.14 shows the change of the three components of \mathbf{d} in Experiment 4

from Table 8.2. Both the Nelder-Mead method and the Powell method converge to similar vectors \mathbf{d} , though Nelder-Mead requires more function evaluations. Note that stigmator values in this experiment are initially close to ideal, while defocus is far off.

As mentioned in Section 7.2 one *function evaluation* corresponds to one *image recording* and the computation of the sharpness function. One image recording in STEM costs about 1-30 seconds depending on microscopic parameters such as the *dwell time* (time required for recording of one pixel) and the number of pixels in the image. The total optimization time for instance for Experiment 1 is about 1 minute 25 seconds for the Nelder-Mead method and 35 seconds for the Powell method.

The experiments described in this subsection show that the Powell method uses fewer function evaluations than Nelder-Mead. The possible explanation for it could be that the Powell method uses $\frac{1}{2}(n+1)(n+2) = 10$ points (and hence 10 function evaluations) to initialize the method. These points are used to approximate the derivative of the sharpness function [60], which helps the method to make a rapid initial step towards the optimum, as can be clearly seen in the fourth row of Figure 8.13. The use of the quadratic approximation improves the overall performance of the method because of the explicit approximation of the gradient and the Hessian of the objective function. On the other hand, Nelder-Mead uses only $n+1 = 4$ points (and hence 4 function evaluations) to initialize the method. Thus there is less information available to determine the search direction.

8.4 Fast autofocus method

In this section we describe five on-line runs of the *fast autofocus* algorithm, which was introduced in Section 3.6. The outputs of one run for a *carbon cross grating* and four runs for a *gold particles* sample are illustrated.

Figure 8.15 shows the output for carbon cross grating. The initial position of the machine defocus is $d_1 = -3 \mu m$. The defocus step $\Delta d = 5 \mu m$ is chosen. The two intermediate images with $d_2 = -8 \mu m$ and $d_3 = 2 \mu m$ are obtained. For each of the images the derivative-based sharpness function (3.19) is computed. The position of the in-focus image is calculated from the given three images and corresponds to $d_4 = 0.3 \mu m$, which is within the *tolerable defocus error* for the given machine settings. The improvements of the image quality are visible in Figure 8.15.

The output of four fast autofocus algorithm runs is shown in Figure 8.16. The first two columns show autofocus based on three images, and the third and the fourth columns show autofocus based on four images. For the four-image autofocus we used the least squares fitting described in Appendix A. The recording of the fourth image might improve the final image quality (the sharpness function has a lower value). The improvement takes place, but it is not that strong. The difference between the fourth and the fifth output images in these runs is not distinguishable by a standard human eye.

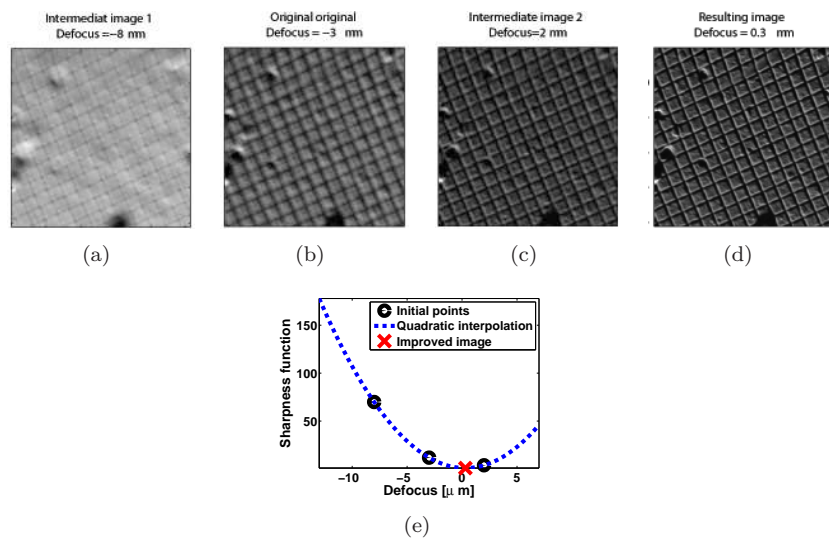


FIGURE 8.15: Derivative-based fast autofocus algorithm, image improvement for the *carbon cross grating* sample. The plot shows the fitting of the three data points with a quadratic polynomial and thus obtaining the in-focus image.

For all runs of the algorithm within our experimental setup the value of the artificial blur parameter α defined in Section 2.2 is equal to 3. As it is described in Section 4.8.2, the choice of the artificial blur is made experimentally by computing the sharpness function for experimental data *off-line* and fitting it with a quadratic polynomial. The value of α corresponding to the smallest approximation error is chosen.

The values of parameter Δd are chosen experimentally as well. It is clear that Δd would be larger than the *tolerable defocus error*, but not too large. In the real-world application Δd should change proportionally to the magnification.

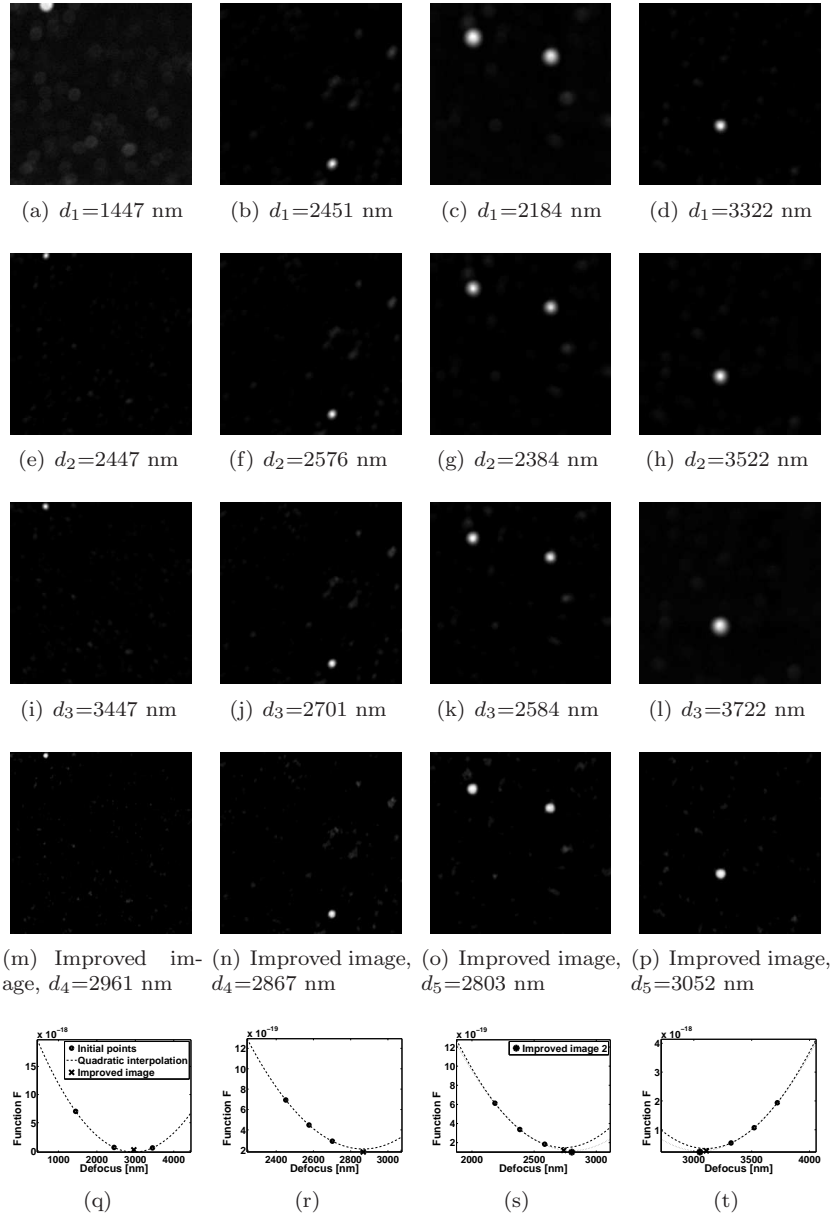


FIGURE 8.16: Fast autofocus. Defocus improvement via interpolation of the derivative-based sharpness function by a quadratic polynomial.

Chapter 9

Future recommendations

Several methods introduced for automated focusing and astigmatism correction have been examined and tested on-line for SEM and STEM with a variety of samples. The methods do not require any special hardware and most of them are generally applicable. They could be used in a number of applications, such as particle analysis or electron tomography, as well as for other types of microscopes and different optical devices. Below we provide a list of suggestions for possible topics of the future study.

- **Automatic choice of input parameters.** Though the methods are automated, they do require input parameters. For instance the simultaneous defocus and astigmatism correction method requires the size of the *simplex* and the *stopping tolerance* as input. The fast autofocus method requires the size of initial step and the artificial blur value. These parameters can be used for entire families of sample geometries and for similar microscope settings. An investigation of the automatic choice of these parameters might be a valuable contribution.
- **Noise quantification.** Noise usually leads to local peaks in the sharpness functions (see for instance the simulation results in Subsection 7.1.2). The quantification of the influence of noise on the methods could be a topic of a study.
- **Alternative sharpness functions for optimization.** The method of simultaneous defocus and astigmatism correction (Chapter 7) makes use of the variance-based sharpness function. However, other sharpness functions (for example, derivative-based) could be used instead for optimization in a three-parameter space. A different choice of the sharpness function might improve the performance of the method.
- **Performance improvement.** The capacity of modern processors allows computations of the sharpness function values within a negligible amount of time. However, image recording might require a noticeable amount of time. The simultaneous defocus and astigmatism correction method described in Chapter 7 is iterative, and requires about 30-50 image recordings. The run-time of the method can be lowered by decreasing the input image size, subsampling the input image, decreasing the *dwell time* (time

required for recording of one pixel) or choosing a faster derivative-free optimization alternative to the *Nelder-Mead simplex method* or the *Powell method*.

- **Application to high-resolution images.** We have applied the methods on-line to images with low-to-medium magnification. In general we observed that applying the method in *high-resolution* mode is not always easy due to instabilities of the environment or the microscope (for instance, image drift and contamination). The application of the methods to high-resolution images should be the topic of future research.

Appendix A

Fast autofocus iterative algorithm

Algorithm A.1. *Fast autofocus (iterative algorithm)*

1. Let d_2 be the current defocus control value of the optical device. Choose a Δd , then $d_1 := d_2 - \Delta d, d_3 := d_2 + \Delta d$.
2. Record three images at d_1, d_2, d_3 and compute $S_1^{-1/2}, S_2^{-1/2}, S_3^{-1/2}$. We set $N=3$.
3. Fit N given points with a quadratic polynomial for instance with the *least squares* and estimate d_{N+1} , the optimum of the quadratic polynomial.

For $N > 3$ we obtain the overdetermined system

$$\underbrace{\begin{pmatrix} 1 & d_1 & d_1^2 \\ \dots & \dots & \dots \\ 1 & d_N & d_N^2 \end{pmatrix}}_{\mathbf{D}} \underbrace{\begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix}}_{\mathbf{c}} = \underbrace{\begin{pmatrix} S_1 \\ \dots \\ S_N \end{pmatrix}}_{\mathbf{s}}.$$

This can e.g. be solved by least squares, giving

$$\mathbf{D}^T \mathbf{D} \mathbf{c} = \mathbf{D}^T \mathbf{s},$$

compute \mathbf{c} ; estimate the sharpness function optimum

$$d_{N+1} = -\frac{c_1}{2c_2}.$$

as the optimum of the polynomial.

4. If for the given tolerance $d_{tol} \in \mathbb{R}$, $|d_N - d_{N+1}| < d_{tol}$, stop. Else, compute $S_{N+1} = S(d_{N+1})$ and go to the previous step.

The parameter d_{tol} can be determined from the knowledge of the property of the optical device. For instance in electron microscopy the *tolerable defocus error* is defined in (7.1). More steps ($N > 3$) of the iterative method are required only if very accurate focusing is needed. In the third step of Algorithm A.1 different numerical methods could be used. The choice of the method is not very important for $N=4$, however.

Appendix B

Power spectrum analytical approximation

The goal of this appendix is to explore ways of fast and accurate approximation of the logarithm of the power spectrum \check{p} . We approximate it with a finite linear combination s of analytical shapes a_i , i.e.,

$$s(\mathbf{u}) = \sum_i c_i a_i(\mathbf{u}). \quad (\text{B.1})$$

To this end a *projection method* is employed. From (B.1) it follows that

$$\iint_{-\infty}^{\infty} (\check{p} - s) a_i d\mathbf{u} = 0 \quad \forall i. \quad (\text{B.2})$$

As a consequence

$$\sum_j c_j \underbrace{\iint_{-\infty}^{\infty} a_i a_j d\mathbf{u}}_{\mathbf{A}_{i,j}} = \underbrace{\iint_{-\infty}^{\infty} \check{p} a_i d\mathbf{u}}_{\mathbf{b}_i}. \quad (\text{B.3})$$

The vector $\mathbf{c} := (c_j)_{j=1}^M$ is a solution of the system $\mathbf{A}\mathbf{c} = \mathbf{b}$, which can be solved by a direct solver. The method does not have a restriction with respect to the power spectrum rotational symmetry. As typical shape functions a_i , the monomials could be used in order to obtain the mathematical moments. For numerical robustness Legendre polynomials or Chebyshev polynomials could be applied.

Let the matrix \mathbf{S} be the approximation of the discrete power spectrum logarithm $\check{\mathbf{S}}$. The *least squares* difference between them is defined as

$$\|\check{\mathbf{P}} - \mathbf{S}\|_{LS} := \left(\frac{1}{N^2} \sum_{i,j} (\check{\mathbf{P}}_{i,j} - \mathbf{S}_{i,j})^2 \right)^{1/2}. \quad (\text{B.4})$$

If the least squares difference (B.4) is of the same order as the noise level in the image, the approximation has a reasonable quality.

Computed data sets	Computational time (sec)
A	600
b	2
c	0

TABLE B.1: Projection method, computational time.



FIGURE B.1: A tin balls sample: power spectrum, its approximation and their profiles before the shift.

Experimental images of two samples are recorded with a FEI Strata SEM. The size of every image is 442×442 pixels. We apply the projection method with Chebyshev polynomials as basis functions to $M = 20^2$ basis functions. The method is applied before the shift operator (see Figure B.1). The appropriate power spectra \tilde{p} and the related projection approximation s are shown in figures B.2-B.3. Figure B.1 corresponds to the second column of Figure B.2 except for the shift. In Figure B.1 results are shown before the shift operator. The least square difference reaches about 5%, which is not large if we take into account the high amount of noise in the power spectrum.

Table B.1 shows an overview of the computational time required on a modern laptop computer. The computation of the matrix **A** takes most of the time. However, it can be computed in advance because it does not depend on the power spectrum itself. The calculation of the vector **c** takes only around 2 seconds.

The method can be used for the fast fitting of the recorded discrete data to the analytical model. It can be useful for the blind deconvolution techniques and automated defocus and astigmatism correction algorithms. The method can be used as well as a power spectrum noise reduction technique.

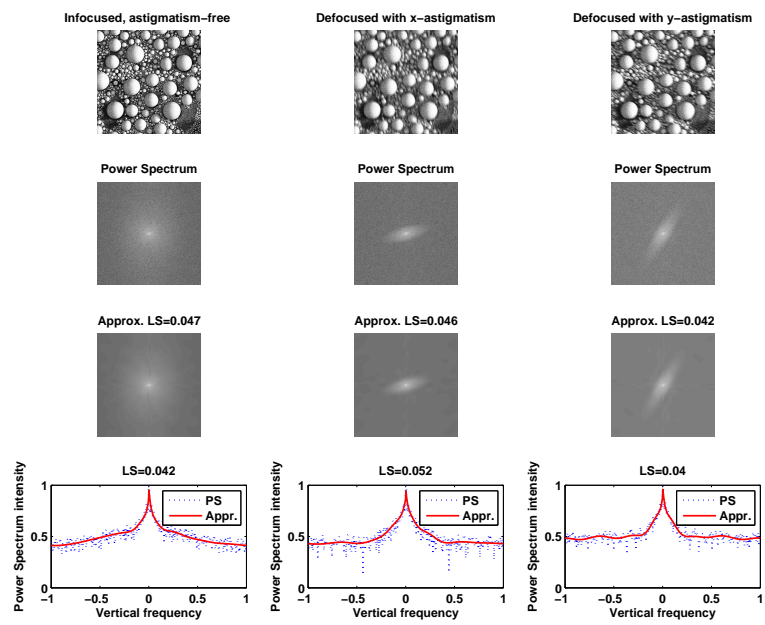


FIGURE B.2: A tin balls sample: power spectrum, its approximation and their profiles after the shift.

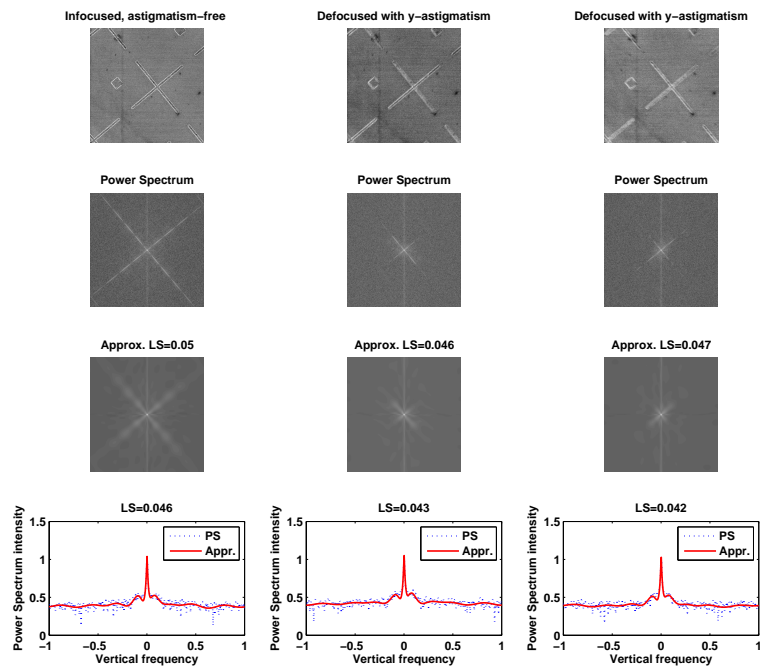


FIGURE B.3: A cross sections sample: power spectrum, its approximation and their profiles after the shift.

Appendix C

The image variance, robustness to noise

In order to avoid the influence of noise on the sharpness function it is suggested to compute the *covariance* of two images recorded at the same sample area

$$\text{cov}[f_1, f_2] := E[(f_1 - E[f_1])(f_2 - E[f_2])],$$

instead of the variance of a single image [16]. Consider two images corresponding to the same object and the same point spread function corrupted by different noise functions

$$f_1 := f^0 + \varepsilon_1, \quad f_2 := f^0 + \varepsilon_2, \quad (\text{C.1})$$

where the natural property of the noise function holds

$$E[\varepsilon_1] = 0, \quad E[\varepsilon_2] = 0.$$

The noise functions in (C.1) are independent on each other and on the image f^0 , thus

$$\text{cov}[\varepsilon_1, \varepsilon_2] = 0, \quad \text{cov}[f^0, \varepsilon_1] = 0, \quad \text{cov}[f^0, \varepsilon_2] = 0.$$

Consequently

$$S^{\text{var}}[f_1] = S^{\text{var}}[f^0] + S^{\text{var}}[\varepsilon_1], \quad S^{\text{var}}[f_2] = S^{\text{var}}[f^0] + S^{\text{var}}[\varepsilon_2],$$
$$\text{cov}[f_1, f_2] = S^{\text{var}}[f^0].$$

Thus the covariance of two images appears to be noise-independent. However, in real-world situations this would not always hold, due to the environment and machine instabilities, such as sample drift and contamination. Also recording two images might be time consuming. The technique described above might be better applicable to SEM than to STEM, due to a shorter *image recording* time and more stable characteristics of the microscope.

Frequently used symbols

$a_{k,l}$	Object pixel values	u	Frequency coordinate
C_s	Spherical aberration	v	Frequency coordinate
d	Defocus control variable	\mathbf{u}	Vector of frequency coordinates
d_x	X-stigmator control variable	x	Spatial coordinate
d_y	Y-stigmator control variable	y	Spatial coordinate
\mathbf{d}	Vector of control variables	\mathbf{x}	Vector of spatial coordinates
\mathbf{d}_0	Ideal value of control variables	\mathbb{X}	Image domain
d_{Sh}	Scherzer defocus value	α	Artificial blur control variable
\mathfrak{F}	Fourier transformation	β	Lévi stable density function parameter
f	Continuous image function	γ	Benchmark object parameter
\hat{f}	Image Fourier Transform	ε	Noise function
\bar{f}	Image mean value	ζ	Power spectrum rotation angle
\check{f}	Image autocorrelation	θ	Point spread function rotation angle
$f_{i,j}$	Image pixel values	λ	Wavelength of electrons
\mathbf{F}	Discrete image matrix	ψ	Object function
g	Gaussian function	ϱ	Lévi stable density function
\mathcal{H}_q	Defocus/stigmator operators	σ	Width of the point spread function
i	Complex number $\sqrt{-1}$	ς	Astigmatism parameter of the point spread function
\mathcal{J}_w	Stretching operator	τ	Pixel width
\mathcal{J}_w	Stretching matrix	ω	Frequency variable in one-dimensional setting
$\mathcal{M}_{k,l}$	Moments of the power spectrum	φ	Polar angle in Fourier space
\mathcal{R}_θ	Rotation operator	χ	Wave aberration function
\mathcal{R}_θ	Rotation matrix	Γ	Gamma function
r	Polar radius in Fourier space	Θ	Pixel threshold parameter
S	Sharpness function		
S^{der}	Derivative-based sharpness functions		
S^{ft}	Fourier transform-based sharpness functions		
S^{ac}	Autocorrelation-based sharpness functions		
S^{int}	Intensity-based sharpness functions		
S^{var}	Variance-based sharpness functions		
S^{his}	Histogram-based sharpness functions		

Bibliography

- [1] M. Abramowitz and I.A. Stegun. *Handbook on mathematical functions with formulas, graphs, and mathematical tables*, chapter Modified Bessel functions, pages 374–377. Dover, New York, 9th edition, 1972.
- [2] S. Amelinckx, D. Van Dyck, J. Van Landuyt, and G. Van Tendeloo. *Handbook of Microscopy: Applications in Materials Science, Solid-State Physics and Chemistry Applications*, volume 2. Weinheim : VCH, 1997.
- [3] M. Azaria, I. Vitsnedel, and Y.Y. Zeevi. The design of two-dimensional gradient estimators based on one-dimensional operators. *IEEE. Trans. Im. Proc.*, 5(1):155–159, 1996.
- [4] N. Baba, K. Terayama, T. Yoshimizu, N. Ichise, and N. Tanaka. An auto-tuning method for focusing and astigmatism correction in HAADF-STEM, based on the image contrast transfer function. *Ultramicroscopy*, 50:163–176, 2001.
- [5] K.N. Bozhilov. Comparison of experimental and calculated spatial resolution in X-ray microanalysis of oxide and silicate minerals. *Microsc. Microanal.*, 15(Suppl 2):1124–1125, 2009.
- [6] J.F. Brenner, B.S. Dew, J. Brian Horton, T. King, P.W. Neurath, and W.D. Selles. An automated microscope for cytologic research a preliminary evaluation. *J. Histochem. Cytochem.*, 24(1):100–111, 1976.
- [7] A.S. Carasso. The APEX method in image sharpening and the use of low exponent Lévy stable laws. *SIAM J. Appl. Math.*, 63(2):593–618, 2002.
- [8] A.S. Carasso, D.S. Bright, and A.E. Vladar. APEX method and real-time blind deconvolution of scanning electron microscopy imagery. *Opt. Eng.*, 41(10):2499–2514, 2002.
- [9] FEI Company. *All you wanted to know about electron microscopy...* FEI Electron Optics, 2006.
- [10] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to derivative-free optimization*. MPS-SIAM series on optimization, 2009.
- [11] J.M. Cowley. Adjustment of a STEM instrument by use of shadow images. *Ultramicroscopy*, 4:413–418, 1979.
- [12] M. de Graef. *Introduction to Conventional Transmission Electron Microscopy*. Cambridge University Press, 2003.
- [13] N. Dellby, O.J. Krivanek, P.D. Nellist, P.E. Batson, and A.R. Lupini. Progress in aberration-corrected scanning transmission electron microscopy. *J. Electron Microsc.*, 50(3):177–185, 2001.
- [14] A.J. Den Dekker, S. Van Aert, A. Van den Bos, and D. Van Dyck. Maximum likelihood estimation of structure parameters from high resolution electron microscopy images. Part I: A theoretical framework. *Ultramicroscopy*, 104:83–106, 2005.
- [15] W. Doornbos, R. Van den Broek, M.E. Rudnaya, D. Langswierdt, and N. Venema. Automation of a FEI electron microscope. In *Proc. of Embedded Systems Institute symposium*, pages 60–61, Eindhoven, The Netherlands, 2009.
- [16] S.J. Erasmus and K.C.A. Smith. An automatic focusing and astigmatism correction system for the SEM and CTEM. *J. Microsc.*, 127(2):185–199, 1982.

- [17] A. Erteza. Sharpness index and its application to focus control. *Appl. Opt.*, 15(4):877–881, 1976.
- [18] L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever. Scale and differential structure of images. *Im. Vis. Comp.*, 10(6):376–388, 1992.
- [19] R.C. Gonzalez, R.E. Woods, and S.L. Eddings. *Digital image processing using MATLAB*. Upper Saddle River: Pearson Prentice Hall, 2004.
- [20] P.J. Goodhew, J. Humphreys, and R. Beanland. *Electron microscopy and analysis*. London : Taylor & Francis, 3rd edition, 2001.
- [21] J.W. Goodman. *Fourier Optics*. Roberts & Company, 3rd edition, 2006.
- [22] M. Haider, H. Muller, and S Uhlemann. *Advances in imaging and electron physics*, volume 153, chapter Present and Future Hexapole Aberration Correctors for High-Resolution Electron Microscopy, pages 43–120. Academic press, Amsterdam, The Netherlands, 2008.
- [23] M. Haider, S. Uhlemann, and J. Zach. Upper limits for the residual aberrations of a high-resolution aberration-corrected STEM. *Ultramicroscopy*, 81:163–175, 2000.
- [24] V. Hilsenstein. Robust autofocusing for automated microscopy imaging of fluorescently labelled bacteria. In *Proc. International Conference on Digital Image Computing: Techniques and Applications*, 2005.
- [25] Embedded Systems Institute. <http://www.esi.nl>.
- [26] R.A. Jarvis. Focus optimization criteria for computer image processing. *Microscope*, 24:163–180, 1976.
- [27] C.B. Johnson. A method for characterizing electro-optical device modulation transfer function. *Photograph. Sci. Eng.*, 14(413–415), 1970.
- [28] S. Jutamulia, T. Asakura, R.D. Bahuguna, and C. De Guzman. Autofocusing based on power-spectra analysis. *Applied optics*, 33(26):6210–6212, 1994.
- [29] J. Kautsky, J. Flusser, B. Zitova, and S. Simberova. A new wavelet-based measure of image focus. *Pattern recognition letters*, 23:1785–1794, 2002.
- [30] E.J. Kirkland. *Advanced Computing in Electron Microscopy*. Plenum Press, 1998.
- [31] M. Koguchi, H. Kakibayashi, R. Tsuneta, M. Yamaoka, T. Niino, N. Tanaka, K Kase, and M. Iwaki. Three-dimensional STEM for observing nanostructures. *J. Electron Microsc.*, 50(3):235–241, 2001.
- [32] O.L. Krivanek, N. Dellby, and A.R. Lupini. Towards sub-Å electron beams. *Ultramicroscopy*, 78:1–11, 1999.
- [33] O.L. Krivanek and P.E. Mooney. Applications of slow-scan CCD cameras in transmission electron microscopy. *Ultramicroscopy*, 49:95–108, 1993.
- [34] E. Krotkov. Focusing. *Int. J. Comput. Vis.*, 1:223–237, 1987.
- [35] R. Krsmanovic, S. Bals, G. Berton, and G. Van Tendeloo. Structural characterization of Er-doped $\text{Li}_2\text{O}-\text{Al}_2\text{O}_3-\text{SiO}_2$ glass ceramics. *Optical Materials*, 30:1183–1188, 2008.
- [36] K. Kumar, M. Pisarenco, M. Rudnaya, V. Savcenko, and S. Srivastava. Shape reconstruction techniques for optical sectioning of arbitrary objects. *Mathematics-in-Industry Case Studies J.*, 3:19–36, 2011.
- [37] T. Le, R. Chartrand, and T.J. Asaki. A variational approach to reconstructing images corrupted by poisson noise. *J. Math. Imaging Vis.*, 27(3):257–263, 2007.
- [38] Y.W. Lee. *Statistical theory of communication*. John Wiley & Sons, New York, 1967.
- [39] G. Ligthart and F.C.A. Groen. A comparison of different autofocus algorithms. In *Proc. 6th Int. Joint Conf. on Pattern Recognition*, pages 597–600, Munich, Germany, 1982.
- [40] X.Y. Liu, W.H. Wang, and Y. Sun. Dynamic evaluation of autofocusing for automated microscopic analysis of blood smear and pap smear. *J. Microsc.*, 227:15–23, 2007.

- [41] A.R. Lupini and S.J. Pennycook. Rapid autotuning for crystalline specimens from an inline hologram. *J. Electron Microsc.*, 57(6):195–201, 2008.
- [42] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition, 1999.
- [43] The MathWorks Inc. Function reference: Volume 2 (F-O). MATLAB 7, 1984–2007.
- [44] The MathWorks Inc. Image processing toolbox 6 user guide. MATLAB 7, 1993–2008.
- [45] P. Meer and I. Weiss. Smoothed differentiation filters for images. *J. Visual Communication and Image Representation*, 3(1):58–72, 1992.
- [46] S. Morigi, L. Reichel, F. Sgallari, and A. Shyshkov. Cascadic multiresolution methods for image deblurring. *SIAM J. Imaging Sci.*, 1(1):51–74, 2007.
- [47] R.A. Muller and A. Buffington. Real-time correction of atmospherically degraded telescope images through image sharpening. *J. Opt. Soc. Am.*, 64(9):1200–1210, 1974.
- [48] K. Nakamae, M. Chikahisa, and Fujioka H. Estimation of electron probe profile from SEM image through wavelet multiresolution analysis for inline SEM inspection. *Image. Vis. Comp.*, 25:1117–1123, 2007.
- [49] S. K. Nayar and Y. Nakagawa. Shape from focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(8):824–831, 1994.
- [50] P.D. Nellist and S.J. Pennycook. Incoherent imaging using dynamically scattered coherent electrons. *Ultramicroscopy*, 78:111–124, 1999.
- [51] B. Nys, I. Geuens, J. Naudts, R. Gijbels, W. Jacob, and P. Van Espen. A convenient method for autofocusing based on image information content. *J. of Computer-Assisted Microscopy*, 2(2):115–123, 1990.
- [52] K.H. Ong, J.C.H. Phang, and J.T.L. Thong. A robust focusing and astigmatism correction method for the scanning electron microscope. *Scanning*, 19:553–563, 1997.
- [53] K.H. Ong, J.C.H. Phang, and J.T.L. Thong. A robust focusing and astigmatism correction method for the scanning electron microscope—part II: Autocorrelation-based coarse focusing method. *Scanning*, 20:324–334, 1997.
- [54] A. Papoulis. *Signal analysis*. McGraw-Hill book company, 1977.
- [55] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(7):629–639, 1990.
- [56] M.J.D. Powell. On the Lagrange functions of quadratic models that are defined by interpolation. Technical report no. damptp 2000/na10, Department of applied mathematics and theoretical physics, University of Cambridge, 2000.
- [57] M.J.D. Powell. UOBYQA: Unconstrained optimization by quadratic approximation. *Math. Programming*, 92:555–582, 2002.
- [58] W.H. Press, S.A. Teukolsky, and W.T. Vetterling. *Numerical recipes : The art of scientific computing*. Cambridge University Press, 3rd edition, 2007.
- [59] W.D. Riecke. *Topics in Current Physics: Magnetic Electron Lenses*, chapter Practical lens design, pages 164–351. Springer-Verlag, 1982.
- [60] M.E. Rudnaya, S.C. Kho, R.M.M. Mattheij, and J.M.L. Maubach. Derivative-free optimization for autofocus and astigmatism correction in electron microscopy. In *Proc. 2nd International Conference on Engineering Optimization*, Lisbon, Portugal, 2010.
- [61] M.E. Rudnaya, R.M.M. Mattheij, and J.M.L. Maubach. Iterative autofocus algorithms for scanning electron microscopy. *Microscopy and Microanalysis*, 15(Suppl 2):1108–1109, 2009.
- [62] M.E. Rudnaya, R.M.M. Mattheij, and J.M.L. Maubach. Evaluating sharpness functions for automated scanning electron microscopy. *Journal of Microscopy*, 240:38–49, 2010.
- [63] M.E. Rudnaya, R.M.M. Mattheij, J.M.L. Maubach, and H. ter Morsche. Autocorrelation-based sharpness functions. In *Proc. 3rd IEEE International Conference on Signal Processing Systems*, Yantai, China, 2011.

- [64] M.E. Rudnaya, R.M.M. Mattheij, J.M.L. Maubach, and H. ter Morsche. Gradient-based sharpness function. In *Proc. International Conference of Applied and Engineering Mathematics*, volume 1, pages 301–306, London, UK, 2011.
- [65] M.E. Rudnaya, R.M.M. Mattheij, J.M.L. Maubach, and H.G. ter Morsche. Orientation identification of the power spectrum. *Optical Engineering*, accepted, 2011.
- [66] M.E. Rudnaya, J.M.L. Maubach, and R.M.M. Mattheij. Scanning electron microscopy: Power spectrum analysis. In *Proc. of 9th Multinational Conference on Microscopy*, pages 185–186, Graz, Austria, 2009.
- [67] M.E. Rudnaya, H.G. ter Morsche, J.M.L. Maubach, and R.M.M. Mattheij. A derivative-based fast autofocus method in electron microscopy. *Journal of Mathematical Imaging and Vision*, accepted, 2011.
- [68] M.E. Rudnaya, W. Van den Broek, R.M.P. Doornbos, R.M.M. Mattheij, and J.M.L. Maubach. Autofocus and twofold astigmatism correction in HAADF-STEM. *Ultramicroscopy*, 111:1043–1054, 2011.
- [69] A. Santos, C.O. De Solórzano, J.J. Vaquero, J.M. Peña, N. Malpica, and F. Del Pozo. Evaluation of autofocus functions in molecular cytogenetic analysis. *J. Microsc.*, 188:264–272, 1997.
- [70] H. Sawada, T. F. Sannomiya, F. Hosokawa, T. Nakamichi, T. Kaneyama, T. Tomita, Y. Kondo, T. Tanaka, Y. Oshima, Tanishiro Y., and Takayanagi K. Measurement method of aberration from ronchigram by autocorrelation function. *Ultramicroscopy*, 108:1467–1475, 2008.
- [71] O. Scherzer. The theoretical resolution limit of the electron microscope. *J. Appl. Phys.*, 20:20–29, 1949.
- [72] J. C. H. Spence. *Experimental high-resolution electron microscopy*. New York: Oxford U. Press., 1988.
- [73] Y. Sun, S. Duthaler, and B.J. Nelson. Autofocusing in computer microscopy: Selecting the optimal focus algorithm. *Microsc. Res. Tech.*, 65:139–149, 2004.
- [74] N. Tanaka, J.J. Hu, and N. Baba. An on-line correction method of defocus and astigmatism in HAADF-STEM. *Ultramicroscopy*, 78:103–110, 1999.
- [75] The Central Microscopy Research Facilities techniques. <http://www.uiowa.edu/cemrf/methodology/index.htm>.
- [76] A. Tejada, A.J. den Dekker, and W. Van den Broek. Introducing measure-by-wire, the systematic use of control theory in transmission electron microscopy. *Ultramicroscopy*, accepted, 2011.
- [77] A. Tejada, W. Van den Broek, S. Van der Hoeven, and A.J. den Dekker. Towards STEM control: Modeling framework and development of a sensor for defocus control. In *Proc. 48th IEEE Conference on Decision and Control*, pages 8310–8315, 2009.
- [78] A. Tejada, S. Van der Hoeven, A.J. den Dekker, and P.M.J. Van den Hof. Towards automatic control of scanning transmission electron microscopes. In *Proc. 18th IEEE International Conference on Control Applications*, pages 788–793, 2009.
- [79] A. Thornton, T. Wienhart, O. Bokhove, B. Zhang, K. Kumar, M. Pisarenco, M.E. Rudnaya, V. Savcenko, J. Rademacher, J. Zijlstra, A. Szabelska, M. Van der Schans, V. Timperio, and F. Veerman. Modeling and optimization of algae growth. In *Proc. 72nd European study group Mathematics with industry*, pages 54–85, Amsterdam, The Netherlands, 2010.
- [80] G. Tzimiropoulos, N. Mitianoudis, and T. Stathaki. A unifying approach to moment-based shape orientation and symmetry classification. *IEEE Trans. Image Proc.*, 18(1):125–139, 2009.
- [81] S. Van Aert, A. J. den Dekker, A. Van den Bos, and D. Van Dyck. High-resolution electron microscopy: From imaging toward measuring. *IEEE Trans. Instrum. Meas.*, 51(4):611–615, 2002.

-
- [82] P.J. Van Bree, C.M.M. Van Lierop, and P.P.J. Van den Bosch. Electron microscopy experiments concerning hysteresis in the magnetic lens system. In *Proc. IEEE Conference on Control Applications (CCA), Multi-Conference on Systems and Control*, Yokohama, Japan, 2010.
- [83] P.J. Van Bree, C.M.M. Van Lierop, and P.P.J. Van den Bosch. On hysteresis in magnetic lenses of electron microscopes. In *Proc. ISIE*, Bari, Italy, 2010.
- [84] W. Van den Broek. *Advanced Focus Methods in Electron Microscopy: Tomographic Reconstruction of the EELS Data Cube Autofocus of HAADF-STEM Images*. PhD thesis, University of Antwerp, 2007.
- [85] W.E. Vanderlinde and J.N. Caron. Blind deconvolution of SEM images. In *Proc. 33rd International Symposium for Testing and Failure Analysis*, 2007.
- [86] S. Vitha, V.M. Bryant, A. Zwa, and A. Holzenburg. 3D confocal imaging of pollen. *Microsc. Microanal.*, 15(Suppl 2):622–623, 2009.
- [87] A.E. Vladar, M.T. Postek, and M.P. Davidson. Image sharpness measurement in scanning electron microscopy—Part II. *Scanning*, 20:24–34, june 1998.
- [88] D. Vollath. Automatic focusing by correlative methods. *J. Microsc.*, 147:279–288, Sep 1987.
- [89] D. Vollath. The influence of the scene parameters and of noise on the behaviour of automatic focusing algorithms. *J. Microsc.*, 151:133–146, 1997.
- [90] de V. Vriendt. Fast computation of unbiased intensity derivatives in images using separable filters. *Int. J. Comp. Vis.*, 13(3):259–269, 1994.
- [91] H. Wang, J. Fang, and M. Lederman. Critical dimension monitoring with STEM in thin film magnetic head wafer production. *Microsc. Microanal.*, 14(Suppl 2):1026–1027, 2008.
- [92] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. John Wiley & Sons, USA, 1949.
- [93] G. Yang and B.J. Nelson. Wavelet-based autofocusing and unsupervised segmentation of microscopic images. In *Proc. of the Intl. Conference on Intelligent Robots and Systems*, 2003.
- [94] T.T.E. Yeo, S.H. Ong, Jayasooriah, and R. Sinniah. Autofocusing for tissue microscopy. *Image Vis. Comput.*, 11(10):629–639, 1993.
- [95] J. Zach and M. Haider. Correction of spherical and chromatic aberration in a low-voltage sem. *Optik*, 98:112–118, 1995.
- [96] F. Zemlin, K. Weiss, W. Schiske, W. Kunath, and K.H. Herrmann. Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms. *Ultramicroscopy*, 3:49–60, 1978.
- [97] Y. Zhang, Y. Zhang, and W. Changyun. A new focus measure method using moments. *Image Vis. Comp.*, 18:959–965, 2000.
- [98] U. Ziese, C. Kubel, A.J. Verkleij, and A.J. Koster. Three-dimensional localization of ultrasmall immuno-gold labels by HAADF-STEM tomography. *J. Struct. Biol.*, 138:58–62, 2002.
- [99] J. Zunic, P.L. Rosin, and L. Kopanja. On the orientability of shapes. *IEEE Trans. Image Proc.*, 15(11):3478–3487, 2006.

Index

- aberration, 3, 5, 7, 15
 - chromatic, 3
 - spherical, 3, 15, 16, 18, 92
- amorphous, 7, 14, 51, 85, 91–93, 97
- aperture function, 16
- arbitrary units, 102
- astigmatism, 3, 5–7, 15, 18, 19, 43, 75, 77, 78, 94
- autocorrelation, 13, 14, 28, 67, 68
- autofocus, 5, 20, 21, 25, 26, 33, 36, 39, 40, 55, 71, 119
 - dynamic, 22
 - static, 21, 75, 79, 80
- automated focusing, *see* autofocus

- Bessel function, 53

- camera length, 102
- Concept Car, 102
- Condor project, 3–5, 102
- convergence semi-angle, 16
- convolution, 12, 38, 49–51
- covariance, 127
- crystalline, 7

- defocus, 5, 7, 8, 16, 18, 20, 21, 39, 40, 43, 99, 101, 102
- differential operator, 23
- dwel time, 102, 115, 119

- electron tomography, 5, 91, 119
- electron voltage, 92, 102
- entropy, 73, 83
- EXPLA, 101

- fast autofocus, 8, 38, 40, 102, 115–117, 121
- finite difference, 23, 70, 98
- focus series method, 8, 54, 56, 102, 103
- focusing, 5
 - course, 22, 79
 - fine, 22, 30, 32
- Fourier transform, 7, 11, 15, 75
 - discrete, 45–47, 49, 58, 75, 77, 102
- frequency coordinate, 11
- function evaluation, 98, 106, 112, 115

- Gaussian function, 12, 14, 23
 - derivative, 38

- high-resolution, 2, 3, 110–112, 120
- histogram, 72
- hysteresis, 3, 21

- image formation model, 12
 - linear, 12, 26
- image recording, 3, 8, 98, 99, 102, 106, 115, 119, 127
- in-focus, 5, 6, 15, 21, 25, 68, 72, 77, 78
- interpolation-based trust-region, *see* Powell method
- iteration, 98, 99, 112

- Lévi stable density, 14, 15, 26, 52, 55
- least squares, 121, 123
- light microscopy, 1, 3, 7, 75

- magnification, 1, 12, 40, 43, 77, 91, 101, 108, 111
- mid-point rule, 22, 72

- Nelder-Mead simplex method, 8, 22, 98, 99, 103, 111–115, 120
- noise, 12, 22, 26, 34, 38, 43, 76, 79, 80, 87, 97, 98, 119

- object function, 12–14
- off-line, 101, 103, 116
- on-line, 101, 103, 119
- out-of-focus, 5, 6, 21, 25, 29, 63, 68, 72, 73, 101, 106, 114

- Parseval’s identity, 26, 33
- particle analysis algorithm, 40, 101, 103
- pixel, 2, 25, 78, 85, 115
 - difference, 23, 38, 76, 83, 85

- image, 22, 38, 72, 73
- object, 13, 29, 31, 34
- width, 13, 29, 40, 43, 92, 108, 111
- point spread function, 12, 14, 15
 - Gaussian, 14, 26, 27, 34–37, 50–52, 68, 69, 71
 - width, 14, 15, 21
- Poisson distribution, 92
- Powell method, 22, 99, 112–115, 120
- power spectrum, 45, 124–126
 - Gaussian, 14, 52–54, 56, 60–62
- projection method, 123
- ranking, 9, 75, 79–81, 86, 87
- resolution, 1, 2, 12
- resolving power, *see* resolution
- Ronchigram, 7
- rotation matrix, 11, 15
- rotation operator, 11, 15
- run, 101, 102
- sample, 2, 16
 - carbon cross grating, 40, 41, 108–110, 114–116
 - cross sections, 77, 78, 126
 - gold particles, 40, 42, 106, 107, 115
 - gold-on-carbon, 63, 78
 - hard disk head, 77, 78
 - integrated circuits, 77, 78
 - tin balls, 52, 77, 78, 124, 125
- Scanning Electron Microscope (SEM), 2, 5, 6, 25, 38, 47–50, 52, 62, 64–66, 68, 72, 75, 77, 78, 81, 103, 119, 124, 127
- Scanning Transmission Electron Microscope (STEM), 2, 8, 38, 40, 104
- Scherzer, 18
 - aperture, 18, 92
 - convergence semi-angle, 18, 92
 - defocus, 9, 18, 91, 92, 94–97
- shadow image, *see* Ronchigram
- sharpness function, 20, 21, 75, 78, 81
 - autocorrelation-based, 75, 76, 82
 - derivative-based, 40, 75, 76, 82
 - Fourier transform, 56
 - Fourier transform-based, 45, 75, 77, 82, 83, 103, 105
 - histogram-based, 75, 76, 82, 83
 - intensity-based, 25, 75, 76, 83
 - variance-based, 75, 76, 82
- sharpness operator, 45, 53, 61, 103, 104
- shift operator, 47, 124
- signal-to-noise ratio, 3
- simplex, 98, 112, 119
- Sobel operator, 38, 76
- spatial coordinate, 11
- spot size, 102
- standard deviation, 14
- stigmatic function, 45, 56, 57, 103, 105
- stigmatic image, 5, 15, 54, 63, 101, 114
- stigmatic operator, 45, 53, 54, 57, 61, 103, 104
- stigmator, 5, 18–20, 35, 53, 54, 56, 57, 99, 101–103, 115
- stopping tolerance, 98, 99, 108, 111, 119
- stretching matrix, 12, 15
- stretching operator, 11
- TEMScripting, 101
- through-focus series, 8, 20, 39–44, 56, 65, 75, 77, 79, 80, 83, 103
- tolerable defocus error, 94, 115, 116, 121
- Transmission Electron Microscope (TEM), 1, 2
- wave aberration function, 16
- wavelength, 1, 3, 16, 92
- window function, 47, 77, 102

Summary

Automated focusing and astigmatism correction in electron microscopy

Nowadays electron microscopy still requires an expert operator in order to manually obtain in-focus and astigmatism-free images. Both the defocus and the twofold astigmatism have to be adjusted regularly during the image recording process. Possible reasons are for instance the instabilities of the environment and the magnetic nature of some samples. For some applications the high level of repetition severely strains the required concentration. Therefore, a robust and reliable autofocus and twofold astigmatism correction algorithm is a necessary tool for electron microscopy automation.

Most of the automatic focusing methods are based on a sharpness function, which delivers a real-valued estimate of an image quality. In this thesis we study sharpness functions based on image derivative, image Fourier transform, image variance, autocorrelation and histogram. A new method for rapid automated focusing is developed, based on a quadratic interpolation of the derivative-based sharpness function. This function has been already used before on heuristic grounds. In this thesis we give a more solid mathematical foundation for this function and get a better insight into its analytical properties.

Further we consider a focus series method, which could act as an extension for an autofocus technique. The method is meant to obtain the astigmatism information from the focus series of images. The method is based on the moments of the image Fourier transforms. After all the method of simultaneous defocus and astigmatism correction is developed. The method is based on a three-parameter optimization (Nelder-Mead simplex method or interpolation-based trust region method) of a sharpness function.

All the three methods are employed for the scanning transmission electron microscopy. To be more specific, we have implemented them in the FEI scanning transmission electron microscope and successfully tested their performance as a part of a real-world application.

Acknowledgements

This thesis records the research that I conducted between 2007 and 2011, as a member of Centre for Analysis, Scientific computing and Applications, Department of Mathematics and Computer Science, Eindhoven University of Technology within Condor project. I would like to express my gratitude to all those who have contributed to the thesis. First of all, I would like to thank my daily supervisor Dr. Jos Maubach and my promotor Prof. Bob Mattheij for giving me the opportunity to work on the project. I am grateful for all the things I learned from them.

Next, I would like to thank the other members of the defense core committee for the time they have spent on reading my thesis and for their valuable feedback: Prof. Luc Florack, Prof. Lothar Reichel, Prof. Dirk van Dyck and Prof. Sandra van Aert. I am thankful to other committee members for joining the defense ceremony: Dr. Seyno Sluiterman and Dr. Hennie ter Morsche.

I am thankful to my FEI advisor Dr. Seyno Sluiterman for sharing great ideas and for the valuable guidance. Next, I would like to thank the colleagues from FEI company for many useful things I learned from them, for providing us with experimental data and technical support: Auke van Balen, Rob van Vucht, Max Otten, Michael Janus, Pybe Faber, Yuri Rikers, Sander Henstra, Maarten Kuijper, Erik Franken.

I am thankful to Richard Doornbos for help and support with practical experiments. It was a real pleasure to make the things running on the microscope. I am grateful to the other colleagues from Embedded Systems Institute for providing the management support of the project: Jan Schuddemat, Sijr van Loo, Jacques Verriet, Teade Punter.

I would like to thank the colleagues from Condor project for interesting discussions, cooperation and the time they have spent to get to know my work and give a valuable feedback: Wouter van den Broek, Sandra van Aert, Dirk van Dyck (EMAT, University of Antwerpen), Arturo Tejada Ruiz, Arjan den Dekker, Sarah van der Hoeven (Technical University of Delft), Patrick van Bree (TU/e). I am thankful to Dries Langsweirdt (Technical University of Leuven) and Nol Venema (Technolution) for the test software development.

I am very thankful to Dr. Hennie ter Morsche for supporting me in the important part of this research. Without Hennie I would not cope with all the proofs

provided in this thesis. I would like to thank Prof. Lothar Reichel and Dr. Michiel Hochstenbach for sharing ideas and knowledge. I am thankful to Prof. Luc Florack, Dr. Remco Duits and Dr. Evgeniya Balmashnova for sharing the image analysis ideas. I am grateful to Sinatra Kho for providing the codes and sharing knowledge about optimization.

I want to thank my office mate Patricio Rosen Esquivel for keeping up the good and positive atmosphere in our room HG 8.87 for all these years, and especially for assistance with the organizing of cooking events. I am grateful to other colleagues without who cooperated for organizational part of cooking events would not be possible: Nico Banagaaya, Kundan Kumar, Valeriu Savcenco, Michiel Renger. I am thankful to my colleagues from CASA Kundan Kumar, Valeriu Savcenco, Maxim Pisarenco and Sudhir Srivastava for encouraging me to get involved into other interesting scientific projects. I am grateful to Maria Ugrumova for providing the introduction to the life in the Netherlands when I just arrived, and keeping up the good atmosphere by singing and smiling. I would like to thank Volha Shchetnikava for being always optimistic and cheerful. I am grateful to Mark van Kraaij for all the time he spent on explaining to me all the details of Latex and Matlab. I would like to thank Mirela Darau for warm conversations and tea breaks.

I would like to thank all other CASA and LIME colleagues for creating a nice and friendly working environment: Oleg Matveichuk, Hans Groot, Neda Sepasian, Ali Etaati, Tasnim Fatima, Yves van Gennip, Shruti Gumaste, Andriy Hlod, Bart Janssen, Peter int Panhuis, Miguel Patricio Dias, Erwin Vondenhoff, Fan Yabin, Rostyslav Polyuga, Qingzhi (Darcy) Hou, Roxana Ionutiu, Godwin Kakuba, Agnieszka Lutowska, Zoran Ilievski, Georg Prokert, Sorin Pop, Prof. Peletier, Prof. Wil Schilders, Bas van der Linden, Ronald Rook, Marese Wolfs-v.d. Hurk. I am thankful to Dr. Rene van Hassel, Dr. Jan ten Thije Boonkamp, Prof. Jan de Graaf, Dr. Adrean Muntean for cooperation in preparing the CASA seminars and guiding me with teaching the courses. I want to thank Enna van Dijk for her kindness and support.

At the end I would like to thank my farther Evgenii Rudnyi for supporting me for all this years. Also I would like to thank Tamara Bechtold for helping me to find this PhD position. I am thankful to Olia Ganzha for continuous support for all these years, help with the vector graphics for some figures of this thesis and sharing the ideas for the cover design.

Maria Rudnaya, Eindhoven 2011

Curriculum Vitae

Maria Rudnaya was born in Moscow, Russia, on the 20th of November in 1984. She graduated at the faculty of Mathematics and Computer Science in Lomonosov Moscow State University in 2007.

In October 2007 she started her PhD research in The Centre for Analysis, Scientific computing and Applications (CASA) at the Technische Universiteit Eindhoven under supervision of prof.dr. Robert Mattheij and dr. Joseph Maubach, the results of which are presented in this thesis.