# Derivative-free optimization for autofocus and astigmatism correction in electron microscopy

# M.E.Rudnaya, S.C.Kho, R.M.M.Mattheij, J.M.L.Maubach

Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands, tel: +31-(0)-40-247-31-62, e-mail: m.rudnaya@tue.nl

#### Abstract

A simultaneous autofocus and two-fold astigmatism correction method for electron microscopy is described. The method uses derivative-free optimization in order to find a global optimum of an image variance, which is an image quality measure. The Nelder-Mead simplex method and the Powell interpolation-based trust-region method are discussed and compared for an application running on a scanning transmission electron microscope.

**Keywords:** Derivative-free optimization, electron microscopy, autofocus, astigmatism correction, image quality measure.

## 1. Introduction

Electron microscopy is a powerful tool in semiconductor industry, life and material sciences. Both the defocus and the two-fold astigmatism have to be adjusted regularly during the image recording process in an electron microscope. It has to do with regular operations such as inserting a new sample, changing the stage position or magnification. Other possible reasons are for instance instabilities of the electron microscope (instabilities of electron beam, stage drift, etc.) and magnetic nature of some samples. Nowadays it still requires an expert operator in order to manually and repeatedly obtain in-focus and astigmatism-free images. For the next instrument generations the manual operation has to be automated. One of the reasons is that for some applications the high level of repetition severely strains the required concentration. Therefore, a robust and reliable simultaneous autofocus and two-fold astigmatism correction algorithm is a necessary tool for electron microscopy applications.

For a specific geometry of input sample we consider a microscopic image as a function  $f(\mathbf{x}, \mathbf{p})$  of spacial coordinates  $\mathbf{x}$  and machine parameters  $\mathbf{p}$ . For an image  $f(\mathbf{x}, \mathbf{p})$  its quality (sharpness) can be estimated by a real-valued quality measure  $\sigma(f(\mathbf{x}, \mathbf{p}))$ . Image quality measure  $F(\mathbf{p}) := \sigma(f(\mathbf{x}, \mathbf{p}))$ reaches its minimum, when an image has highest possible quality, i.e. the microscopic parameters  $\mathbf{p}$  are at optimal focus and astigmatism. The optimization goal is then to minimize the image quality measure of the parameters defocus, x-stigmator and y-stigmator, to be denoted by  $p_1$ ,  $p_2$  and  $p_3$  respectively, i.e.

$$\min_{\mathbf{p}\in\mathcal{R}^3} F(\mathbf{p}), \quad \mathbf{p} := (p_1, p_2, p_3)^T.$$
(1)

Thus, the problem of automated defocus and astigmatism correction in electron microscopy can be considered as at least a three-parameter optimization problem (one focus and two astigmatism parameters)  $\mathbf{p} \in \mathcal{R}^3$ . The evaluation function is an image  $f(\mathbf{x}, \mathbf{p})$  recording with computing of an image quality measure  $\sigma(f(\mathbf{x}, \mathbf{p}))$ . To this end the derivative-free Nelder-Mead simplex method is used in [7]. The approach proves to work both for simulated images and in a real-world application of Scanning Transmission Electron Microscopy. The accuracy of the method is comparable to a human operator and the speed is faster. In this paper we replace Nelder-Mead simplex method by the Powell interpolation-based trust-region method, and show that the second one requires less function evaluations in general and thus has a better performance.

In the next section a short introduction into electron microscopy, defocus and astigmatism correction is given. The linear image formation model is presented. As image quality measure  $\sigma(f(\mathbf{x}, \mathbf{p}))$  we introduce and define the variance. In Section 3 an overview of the two derivative-free optimization methods (Nelder-Mead and Powell methods) is given. Section 4 describes experimental results from a real-world application. Section 5 summarizes the paper with discussion and conclusions.

## 2. Autofocus and astigmatism correction in electron microscopy

The resolution of a typical light microscope is 200 nm, being limited only by the wavelength of visible light.

The electron microscope uses electrons instead of photons used in optical microscopy. The wavelength of electrons is much smaller than the wavelength of photons (for example, for 100-keV electrons,  $\lambda = 3.7$  pm), which makes it possible to achieve much higher magnifications than in light microscopy.

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 400x magnification. The simplest Transmission Electron Microscope (TEM) is an exact analogue of the light microscope (see Figure 1(a)). The illumination coming from the electron gun is concentrated on the specimen with the condenser lens. The electrons transmitted through a specimen 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 is larger and more complex than the diagram suggests: High vacuum, long electron path, highly stabilized electronic supplies for electron lenses are required [9]. The accelerating voltage of the TEM electron source is 100-1000 kV.

In a Scanning Electron Microscope (SEM) a fine probe of electrons with energies from a few hundred eV to tens of keV is focused at the surface of the specimen and scanned across it (see Figure 1(b)). A current of emitted electrons is collected, amplified and used to modulate the brightness of a cathoderay tube. Scanning Transmission Electron Microscope (STEM) is a combination of SEM and TEM. A fine probe of electrons is scanned over a specimen and transmitted electrons are being collected to form an image signal [3]. Resolution of STEM achieves 0.05 nm nowadays. Figure 1(c) shows STEM (FEI company). The resolution in electron microscopy is limited by abberations of the magnetic lens, but not by the wavelength, as in light microscopy.



Figure 1: 1(a) Schematic drawings of light microscope and transmission electron microscope (taken from [9]); 1(b) schematic drawings of scanning electron microscope (taken from [9]); 1(c) FEI's scanning transmission electron microscope.

Astigmatism is a lens abberation that deals with the fact that the lens is not perfectly symmetric. Figure 2(a) shows a ray diagram for an astigmatism-free situation. The only adjustable parameter in order to get a sharp image is the current through magnetic lens; it changes the lens focal length and focuses the magnetic beam on the image plane. This is the parameter  $p_1$  in Eq.(1). Astigmatism implies that the rays traveling through a horizontal plane will be focussed at a different focal point with the rays traveling through vertical plane (Figure 2(b)). Due to the presence of astigmatism the electron beam becomes elliptic. Two stigmators (X-stigmator and Y-stigmator) that correspond to parameters  $p_2$  and  $p_3$  in Eq.(1) adjust the final beam shape, by applying a correcting field (Figures 2(c), 2(d)).

Let  $f(\mathbf{x}, \mathbf{p})$  be an image, where  $\mathbf{x} = (x, y)$  are the spatial coordinates and  $\mathbf{p} \in \mathcal{R}^3$  is a vector of defocus and two-fold astigmatism parameters. A linear image formation model is given as in [4]

$$f(\mathbf{x}, \mathbf{p}) = \psi_0(\mathbf{x}) * h(\mathbf{x}, \mathbf{p}) + \epsilon(\mathbf{x}) := \iint_{\mathbf{X}} \psi_0(\mathbf{x}') h(\mathbf{x} - \mathbf{x}', \mathbf{p}) d\mathbf{x}' + \epsilon(\mathbf{x}),$$
(2)



Figure 2: 2(a) Ray diagram for astigmatism-free lens, the lens has one focal point; 2(b) ray diagram for lens with astigmatism, the lens has two focal points; 2(c) and 2(d) show schemes of X- and Y-stigmators correspondingly that are used for correction of astigmatic electron beam.

where \* denotes convolution,  $\psi_0(\mathbf{x})$  is a function that describes microscopic object geometry,  $h(\mathbf{x}, \mathbf{p})$  is an intensity of a probe function (or a point spread function) and  $\epsilon(\mathbf{x})$  is noise. The mean value of the image  $f(\mathbf{x}, \mathbf{p})$  is

$$\bar{f}(\mathbf{p}) := \frac{\iint_{\mathbf{X}} f(\mathbf{x}, \mathbf{p}) d\mathbf{x}}{\iint_{\mathbf{X}} d\mathbf{x}}.$$
(3)

The image quality is estimated by measure  $\sigma$  which is a real-valued function of the image. An overview of different image quality measures can be found in [8]. In this paper, as well as in [7], the image variance is used as an image quality measure

$$F(\mathbf{p}) := \sigma(f(\mathbf{x}, \mathbf{p})) = -\frac{1}{\bar{f}^2(\mathbf{p})} \iint_{\mathbf{X}} \left| f(\mathbf{x}, \mathbf{p}) - \bar{f}(\mathbf{p}) \right|^2 d\mathbf{x}.$$
(4)

The use of variance is based on the assumption that a defocused (out-of-focus) image has a lower contrast than the in-focus one. Image variance values  $F(\mathbf{p})$  as well as variance shape F are not known in advance, because they depend on the object geometry  $\psi_0(\mathbf{x})$ , which is generally unknown. It has been proven in [2] that the variance of an amorphous image reaches its global optimum (in our case minimum) for ideal defocus and astigmatism in SEM under assumptions of a Gaussian point spread function, amorphous sample and noise-free situation ( $\epsilon(\mathbf{x}) = 0$  in Eq.(2)). It has been shown numerically in [7] that variance reaches its global optimum for ideal defocus and astigmatism in STEM for abberationbased point spread function and typical types of samples.

The image quality measure optimization in electron microscopy has the following characteristics:

- The objective function  $F(\mathbf{p})$  might be noisy due to the noise  $\epsilon(\mathbf{x})$  in image formation;
- The objective function  $F(\mathbf{p})$  might have local optima due to the sample's geometry  $\psi_0(\mathbf{x})$ ;
- At different moments in time

$$t_1 \neq t_2 : F_{t_1}(\mathbf{p}) \neq F_{t_2}(\mathbf{p}) \tag{5}$$

not only due to the noise  $\epsilon(\mathbf{x})$ , but also due to instabilities, like sample drift, sample contamination and hysteresis of magnetic lens.

- Computations determining an image quality measure cost much less time than an image recording. Therefore, and also because repeated recordings can damage or destruct the sample, it is needed to optimize the focus and astigmatism with just few recordings (few function evaluations);
- Analytical derivative information is not available. Calculating (approximately) derivatives with finite differences would dramatically increase the amount of necessary image recordings, and is therefore not discussable.

#### 3. Derivative-free optimization

In this section we give a short overview of Nelder-Mead simplex method[1] and the Powell interpolationbased trust-region method [5], which both serve to find a local optimum of a function. Both methods do not use derivative information. Nelder-Mead makes no assumption about the objective function shape, while Powell method assumes quadratic shapes. In the next section the results of comparison of both methods for autofocus and astigmatism correction application in STEM are presented. Every iteration of the Nelder-Mead method is based on a simplex. A simplex in n dimensions is the convex hull of n+1 points. For our application Eq.(1) with one defocus and two astigmatism parameters to replace n = 3, and the simplex is a tetrahedron. The Nelder-Mead simplex method initially needs n+1 points for the definition of a starting simplex. In electron microscopy the starting point is initial position of microscope defocus and astigmatism parameters  $\mathbf{p}_0 \in \mathcal{R}^{n \times 1}$ . Thus, for the given input parameter  $\rho_{beg,NM} > 0$  and the vertices of initial simplex  $\mathbf{P} \in \mathcal{R}^{n \times n+1}$  are constructed as

$$\mathbf{P}_i = \mathbf{p}_0 + \rho_{beg,NM} \mathbf{e}_i, \quad i = 0 \dots 3$$

where  $\mathbf{e}_i$ , i = 1...3 is a unit vector and  $\mathbf{e}_0 = \mathbf{0}$ . For every iteration the Nelder-Mead simplex method evaluates the function for a finite number of points (between 1 and n+2). Further, Nelder-Mead replaces the point with the largest variance in the simplex, with one with a lower variance. The algorithm stops when the difference in variance for a small simplex according to the input tolerance parameter  $\rho_{end,N} > 0$ becomes small

$$\|\mathbf{p}_{k,1}-\mathbf{p}_{k,2}\|_{\infty} < \rho_{end,NM},$$

where  $\mathbf{p}_{k,1}, \mathbf{p}_{k,2}$  are the two points of the simplex on kth iteration with the lowest variance. Thus, the method requires two input parameters:

$$\rho_{beg,N}, \rho_{end,NM} > 0.$$

The Nelder-Mead implementation used for the application described in the following section is based on the code available in the Matlab Optimization Toolbox V3.1.2.



Figure 3: The flowchart of the interpolation-based trust-region method.

The second method we consider to solve problem Eq.(1) is an interpolation-based trust region, following Powell's UOBYQA (Unconstrained Optimization BY Quadratic Approximation) [5]. The interpolation-based trust-region employs a series of local quadratic models that interpolate m sample points of the objective function. The quadratic model is constructed using standard Lagrangian basis functions and is minimized within the trusted region of the current iterate(s) to obtain a trial point. The repeated quadratic model minimization converges to the nearest local minimum. Each computed trial point is tested by a certain mechanism whether it is accepted or rejected. 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 objective function. The iterative process is stopped if the length of the trial step is below tolerance or the maximum allowable number of function evaluations is reached.

Starting the method requires an initial point  $\mathbf{p}_0 \in \mathcal{R}^n$ , and an initial trust-region radius  $\rho_{beg} \in \mathcal{R}$ . The stopping criterion parameters need to be supplied as well, i.e., the maximum allowable function evaluations  $c_{max} \in \mathcal{N}$ , and the final trust-region radius  $\rho_{end} \in \mathcal{R}$ . The objective function is approximated by a quadratic model

$$Q(\mathbf{p}_k + \Delta \mathbf{p}) = c_Q + \mathbf{g}_Q^T \Delta \mathbf{p} + \frac{1}{2} \Delta \mathbf{p}^T \mathbf{G}_Q \Delta \mathbf{p}$$
(6)

that is constructed by interpolating a uni-solvent set of points  $\{\mathbf{p}_i\}_{i=1}^m$  which interpolates F

$$Q(\mathbf{p}_i) := F(\mathbf{p}_i), \qquad i = 1, 2, \dots, m,\tag{7}$$

where the initial interpolation points have been chosen automatically according to the initial point and the initial trust-region radius. In the quadratic model Eq.(6), the coefficients are the scalar  $c_Q$ , the vector  $\mathbf{g}_Q$ , and the symmetric matrix  $\mathbf{G}_Q$ . The point  $\mathbf{p}_k \in {\{\mathbf{p}_i\}_{i=1}^m}$  is the center of the trust-region and is the latest best point in the interpolation set, i.e.,  $Q(\mathbf{p}_k) = F(\mathbf{p}_k) = \min{\{F(\mathbf{p}_j), j = 1, 2, ..., m\}}$ . It also acts as the center of Taylor expansion of F. To describe the coefficients of Eq.(6) uniquely, we require m, the number of interpolation points, to be  $\frac{1}{2}(n+1)(n+2)$ . For application Eq.(1), since we have n = 3, we need 10 initial interpolation points and hence, 10 function evaluations to initialize the method.

The quadratic model is constructed using Lagrangian basis functions. Each Lagrangian function,  $\ell_j, j = 1, \ldots, m$ , is a quadratic polynomial  $\ell_j$  from  $\mathcal{R}^n$  to  $\mathcal{R}$  that has property  $\ell_j(\mathbf{p}_i) := \delta_{ij}, i = 1, \ldots, m$ , and takes the form in an analogous way to that of Eq.(6). The use of Lagrange functions has three-fold advantages: they are useful to maintain the non-singularity of the interpolation system, they provide an error bound of the quadratic model, they can be updated efficiently if there is a change of point in the interpolation set [6].

We work explicitly with the coefficients of the quadratic model Q, and the Lagrange function  $\ell_j, j = 1, \ldots, m$ . They are carried out and updated throughout the computation. The initial coefficients of the Lagrangian functions are obtained by solving systems of linear equations. As soon as the coefficients of all Lagrangian functions are available, obtaining the coefficients of the quadratic model are trivial. This approach is simple and easy to implement but might not be the most efficient one. In [5, 6], Powell describes a more efficient approach in finding the analytical expression of the Lagrange functions and their corresponding derivatives by making use of the structure of the initial interpolation points, i.e., they are constructed on the lattice points.

The quadratic model Eq.(6) is used in a trust-region calculation to find a trial step  $\Delta \mathbf{p}$  such that the quadratic model is minimum within the trust-region

$$\min_{||\Delta \mathbf{p}|| \le \tau} Q(\mathbf{p}_k + \Delta \mathbf{p}) \tag{8}$$

where  $|| \cdot ||$  denotes Euclidean 2-norm and  $\tau$  is the outer trust-region radius. In the method, we make use of two radii to define the trust-region. They are the inner and the outer trust-region radii, denoted by  $\rho$  and  $\tau$ , respectively, where  $\tau \ge \rho$  and  $\rho_{beg} \ge \rho \ge \rho_{end}$ . The inner radius is never increased since this will require decrement later which can reduce the efficiency of the algorithm. It is decreased if no further improvement in the optimization can be obtained with the current  $\rho$ . On the other hand, the outer radius is allowed to increase and decrease depending on the success in the optimization, expressed in the ratio of the reduction in the objective function and the reduction in the quadratic model. The advantage of using the two radii is to allow the trial step length to exceed the inner radius.

The trial point  $(\mathbf{p}_k + \Delta \mathbf{p})$  is accepted if it maintains the non-singularity of the interpolation system. Otherwise, we either improve the geometry of the interpolation system or decrease the outer trust-region radius. To improve the geometry of the interpolation set, we typically move a point that is far away from the current best point. However, in general the process is assisted by the Lagrange functions that exploit the error bound estimate of the quadratic model derived by Powell [6].

The modification of the trust-region radii is described in the following paragraph. After solving the trust-region subproblem, we calculate the ratio of the reduction in the objective function and the reduction in the quadratic model:

$$r := \frac{F(\mathbf{p}_k) - F(\mathbf{p}_k + \Delta \mathbf{p})}{Q(\mathbf{p}_k) - Q(\mathbf{p}_k + \Delta \mathbf{p})}$$
(9)

The ratio Eq.(9) is typical in trust-region framework. It controls the the increase or decrease of the outer trust-region radius,  $\tau$ . We typically increase the radius if the ratio is relatively high and decrease it otherwise. The precise updating is given as follow:

$$\tau = \begin{cases} \max\left[\tau, \frac{5}{4} ||\Delta \mathbf{p}||, \rho + ||\Delta \mathbf{p}||\right], & r \ge 0.7; \\ \max\left[\frac{1}{2}\tau, ||\Delta \mathbf{p}||\right], & 0.1 < r < 0.7; \\ \frac{1}{2} ||\Delta \mathbf{p}||, & r \le 0.1. \end{cases}$$
(10)

If  $\tau \leq \frac{3}{2}\rho$  we set  $\tau = \rho$  since  $\rho \leq \tau$  is mandatory. Moreover, as mentioned previously that the inner trust-region radius is reduced if no further reduction can be obtained from the current value. If  $\rho > \rho_{end}$ , the updating is done as follow

$$\rho = \begin{cases}
\rho_{end}, & \rho_{end} < \rho \le 16\rho_{end}; \\
\sqrt{\rho \ \rho_{end}}, & 16\rho_{end} < \rho \le 250\rho_{end}; \\
0.1\rho, & \rho > 250\rho_{end}.
\end{cases}$$
(11)

The numbers in Eq.(10) and Eq.(11) are recommended by Powell in [5], and we follow this without any changes.

Once an interpolation point in the set is replaced by a new point, while keeping all the remaining points, all the coefficients of the Lagrangian functions and the quadratic model need to be updated. This is where the use of the Lagrangian functions is handy that they provide an efficient updating procedure as explained in [6]. The steps from solving the trust-region subproblem and updating the interpolation points, trust-region radii, the coefficients of Lagrangian functions and quadratic model are repeated until one of the two stopping criterion are satisfied. We stop if the maximum allowable number of function evaluations is reached ( $c > c_{max}$ ) or if the inner trust-region radius is smaller than the final trust-region radius ( $\rho < \rho_{end}$ ). The first is related to the expensive computation of the objective value while the second is related to the accuracy of the method. We summarize the method in an flowchart shown in Figure 3.

Table 1: Microscope settings during experiments.

Electron voltage	Camera length	Spot size	Dwell time	Image size	Pixel depth
200 keV	200  mm	8	$1.25 \ \mu s$	$256 \times 256$ pixels	16-bit

# 4. Experimental results

The experiments with the method application are performed on a FEI Tecnai F20 STEM electron microscope with the settings summarized in Table 1. We use the carbon cross grating sample, which is designed for microscope calibration. In every experiment the correction of defocus and astigmatism is performed with the help of Powell and Nelder-Mead methods sequentially. The methods are applied to the same starting point  $\mathbf{p}_0$  with the equal sets of initial parameters

$$\rho_{beg,NM} = \rho_{beg},\tag{12}$$

$$\rho_{end,NM} = \rho_{end}.\tag{13}$$

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

Table 2: Experimental results



Figure 4: Illustration of application work for experiments N=1,4,7 from Table 2.



Figure 5: Parameter changes in Experiment 4.

Due to Eq.(12) the initial simplex for Nelder-Mead, consisting of n + 1 = 4 points, is a subset of initial trust-region for Powell, consisting of  $\frac{1}{2}(n+1)(n+2) = 10$  points.

In all the test cases summarized in Table 2 both methods are able to find the global optimum successfully. However, the amount of required function evaluations appears to be different. Though equality Eq.(13) takes place, the two stopping criteria have different meanings for Powell and Nelder-Mead methods. As a result Nelder-Mead method aims to achieve lower objective function values than Powell method and as a consequence performs larger amount of function evaluations. In order to make a truthful comparison with Powell we take the last function evaluations of Nelder-Mead away till the variance values similar to the last values of Powell is reached. The amount of function evaluations left for Nelder-Mead are shown in Table 2. Still the amount of required function evaluations is different. The first column of Table 2 indicates the experiment number. The second column shows magnification of the microscope, for which the experiment is performed. The columns *Improvement* correspond to the relative variance improvement

$$v := \frac{F_{final} - F_{initial}}{F_{initial}} \tag{14}$$

for Nelder-Mead and Powell methods correspondingly. In Eq.(14)  $F_{initial} := F(\mathbf{p}_0)$ . We take into account the relative value of variance change Eq.(14) instead of the absolute value, because the absolute value changes due to noise and instabilities Eq.(5). The columns *Evaluations* indicate amount of function evaluations for each of the methods. The final ratio is computed as a ratio of average function improvement per iteration for both methods

$$R := \left(\frac{v_P}{E_P}\right) / \left(\frac{v_{NM}}{E_{NM}}\right). \tag{15}$$

We can see that in all the six cases R > 0, which means that Powell method shows a higher performance than Nelder-Mead method, i.e. it required less function evaluations in order to achieve the same quality image as Nelder-Mead method. However, in some cases this difference is large (Experiment 2) and sometimes the performances of the algorithms is almost equal (Experiment 5).

Figure 4 gives an illustration of application work for experiments N=1,4,7 from Table 2. Figures 4(a), 4(b), 4(c) show defocused, stigmatic images of carbon cross grating sample before the optimization. For every experiment both Powell and Nelder-Mead optimization are applied and leaded to the similar final results. Figures 4(d), 4(e), 4(f) show images after optimization. The change of variance versus function

evaluations for Nelder-Mead is shown in figures 4(g), 4(h), 4(i) and for Powell in figures 4(j), 4(k), 4(l). These plots clearly show that though Powell requires a larger amount of initial function evaluations it reaches the optimum faster. We can observe that the variance magnitude (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 5 shows the change of parameters  $\mathbf{p}$  during optimizations in Experiment 4. The values of defocus are given in  $\mu m$  and stigmator values are given in arbitrary machine units. Both Nelder-Mead and Powell methods converge to the same values, though Nelder-Mead requires more function evaluations. We can observe that stigmator values in this experiment are initially close to ideal, while defocus is far off.

As it is mentioned before, one function evaluation corresponds to one image recording. One image recording in STEM costs about 1-30 seconds depending on microscopic parameters like dwell time and amount of pixels in the image. In the case of described experiment for the settings summarized in Table 1 the total optimization time for instance for Experiment 1 is about 1 minute 25 seconds for Nelder-Mead method and 35 seconds for Powell method.

## 5. Discussion and conclusions

The experiments with the real microscopy application have shown that the Powell method uses fewer function evaluations than Nelder-Mead and is therefore faster than the application in [7]. 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 objective function which helps the method for this application to make an rapid initial step towards the optimum, which can be clearly seen in the 4th row of Figure 4. The use of the quadratic model in the method 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, it is less flexible to decide to which direction to go. Moreover, the progress towards the optimum is only based on function values without even implicit information on the objective function derivative.

#### Acknowledgements

We kindly acknowledge R.Doornbos (ESI, The Netherlands) for assistance with obtaining experimental data and W. Van den Broek (EMAT, Belgium) for thoughtful discussions.

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.

## References

- A.R. Conn, K. Scheinberg, L. N. Vicente, Introduction to derivative-free optimization, MPS-SIAM series on optimization, Philadelphia, 2009.
- [2] S.J. Erasmus, K.C.A. Smith, An automatic focussing and astigmatism correction system for SEM and CTEM, *Journal of Microscopy*, 127 (1982) 185–199.
- [3] S.J. Goodhew, J. Humphreys, R. Beanland, *Electron microscopy and analysis*, 3rd ed., Taylor & Francis, London, 2001.
- [4] E.J. Kirkland, Advanced Computing in Electron Microscopy, Plenum Press, New York, 1998.
- [5] M. J. D. Powell, UOBYQA: unconstrained optimization by quadratic programming, Mathematical Programming, 92, 555-582, 2002.
- [6] 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, England, 2000.
- [7] M.E. Rudnaya, W. Van den Broek, R.M.P. Doornbos, R.M.M. Mattheij, J.M.L. Maubach, Autofocus and two-fold astigmatism correction in HAADF-STEM, CASA-Report 10-09, Eindhoven University of Technology, The Netherlands (http://www.win.tue.nl/analysis/reports/rana10-09.pdf).

- [8] M.E. Rudnaya, R.M.M. Mattheij, J.M.L. Maubach, Evaluating sharpness functions for automated Scanning Electron Microscopy, *Journal of Microscopy*, 2010, in press.
- [9] The Central Microscopy Research Facilities techniques (http://www.uiowa.edu/ cemrf/methodology/index.htm).