Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Sensitivity Analysis for Indirect Measurement in Scatterometry and the Reconstruction of Periodic Grating Structures

Hermann Gross[1] and Andreas Rathsfeld[2]

[1] *Department 8.4 of Mathematical Modelling and Data Analysis, Physikalisch-Technische Bundesanstalt Abbestr. 2–12, 10587 Berlin, Germany E-mail: Hermann.Gross@ptb.de*

[2] *Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany E-mail: rathsfeld@wias-berlin.de*

submitted: September 18, 2006

No. 1164
Berlin 2006

1991 *Mathematics Subject Classification.* 78A46, 65N30, 65K05.

Key words and phrases. diffraction gratings, inverse problems, sensitivity analysis.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

In this work, we discuss some aspects of numerical algorithms for the determination of periodic surface structures (gratings) from light diffraction patterns. With decreasing structure details of lithography masks, increasing demands on suitable metrology techniques arise. Methods like scatterometry as a non-imaging indirect optical method are applied to simple periodic line structures in order to evaluate the quality of the manufacturing process. Using scatterometry, geometrical parameters of periodic structures including period (pitch), side-wall angles, heights, top and bottom widths of trapezoid shaped bridges can be determined. The mathematical model for the scattering is based on the time-harmonic Maxwell's equations and reduces in case of grating structures to the Helmholtz equation. For the numerical simulation, e.g. finite element methods can be applied to solve the corresponding boundary value problems. More challenging is the inverse problem, where the grating geometry is to be reconstructed from the measured diffraction patterns. Restricting the class of gratings and the set of measurements, the inverse problem can be reformulated as a non-linear operator equation in Euclidean spaces. The operator maps the parameters describing the grating to special efficiencies of plane wave modes diffracted by the grating. We employ a Newton type iterative method to solve this operator equation. The reconstruction properties and the convergence of the numerical algorithm, however, is controlled by the local conditioning of the non-linear mapping, i.e. by the condition numbers of its Jacobian matrix. To improve the convergence of the iteration and the accuracy of the reconstruction, we determine optimal sets of efficiencies for the measurements by optimizing the condition numbers of the corresponding Jacobians. Numerical examples for a chrome-glass mask and for an inspecting light of wave length 632.8 nm confirm that an optimization of the measurement data results in better solutions.

1 Introduction

The investigation of micro- or nano-structured surfaces regarding their structure geometries and dimensions can be performed in a rapid and non-destructive way by the measurement and analysis of light diffraction by the structured surfaces. Non-imaging metrology methods like scatterometry are in contrast to optical microscopy non diffraction limited and they grant access to the geometrical parameters of periodic structures (cf. Figure 1) like structure width (critical dimension CD), period (pitch), side-wall angle or height of trapezoidal bridges (lines) [23, 15]. An important application of scatterometric metrology is the evaluation of structure dimensions on photo-masks and wafers in lithography [30, 29]. In particular in the semiconductor industry both the feature sizes and the required limit of measurement uncertainty decrease continuously. Besides conventional microscopical metrology techniques like atomic force, electron and optical microscopy, scatterometry is an important tool for the characterization of such structures (cf. e.g. [35, 13]). However, scatterometric methods require a-priori information. Typically, the surface structure is sought in a certain class of gratings described by a finite number of parameters, and these parameters are confined to certain intervals.

The conversion of measurement data into the desired geometrical parameters depends crucially on a high precision rigorous modelling of the light-structure interaction, which includes the vectorial and the 3D character of light and structure, respectively. More precisely, the mathematical modelling of scatterometry requires the computation of the relation between the input (the incoming wave) and output (diffraction efficiencies, phase shifts). These quantities are described by Maxwell's equations [27, 4, 7] which reduce to the two-dimensional Helmholtz equation since geometry and material properties are invariant in one direction. The typical transmission conditions of electro-magnetic fields turn into continuity and jump conditions for the transverse field components, and the radiation conditions at infinity are well established.

For the numerical solution of the Helmholtz equation there exists a whole variety of different methods. We mention here the rigorous coupled wave analysis (cf. [24, 32, 25, 20, 21]), the so-called C method (cf. [6]), and the boundary element method (cf. the references in [18]). On the other hand, the most popular method to solve boundary value problems for elliptic partial differential equations is the finite element method (FEM). The truncation of the infinite domain to a finite domain of computation can be accomplished by coupling with boundary elements. FEM has been applied to gratings e.g. by Urbach [34], Bao [3], and Elschner et al [10]. We refer also to the alternative FEM approach proposed for the Maxwell system by Schaedle, Zschiedrich, Burger, Klose, and Schmidt [31] which includes absorbing boundary conditions (perfectly matched layers) and domain decomposition techniques. To improve the computation of highly oscillatory fields, generalized finite element methods are available (cf. e.g. [16, 22, 5, 9]).

Apart from the forward computations of the Helmholtz equation, the solution of the inverse problem, i.e. the reconstruction of the grating profiles and interfaces from measured or simulated diffraction data, is the essential task of the indirect measurements in scatterometry. This problem is strongly related to optimization problems for the design of diffractive optics (cf. [33]). Our approach here employs:

- FEM computation of efficiencies corresponding to a given grating (cf. [12])
- FEM computation for the derivatives of the efficiencies with respect to geometry parameters of the grating (cf. [11])
- Iterative algorithm of Newton type proposed by Al-Assaad and Byrne [1] modified for box constraints

It is well known that the solution of the inverse problem might fail if it is based on insufficient or improper input data. Studies with simulated data for a typical grating representing a photolithograhic mask [13] show a strong dependence of the reconstruction result on the subset of efficiencies chosen from the set of all available efficiencies. Based on a sensitivity analysis, we propose an algorithm for finding sets of efficiency data suitable for the inverse problem. Indeed, high local sensibility of the geometry parameters with respect to the measured efficiency values is expressed by the well conditioning of the Jacobi matrix of the mapping efficiency values to geometry parameters. Equivalently, high sensibility means small condition numbers for the Jacobian of the mapping geometry parameters to efficiency values. Hence, we choose our optimal measurement set of efficiency values by minimizing the condition numbers of the corresponding Jacobians.

This paper is organized as follows: In Section 2 the mathematical formulation of the scattering problem and the numerical algorithms for solving the inverse problem are discussed. An FEM-based Newton type method to solve the reconstruction problem is described and the convergence properties are indicated. Section 3 presents a sensitivity analysis for the indirect measurements in scatterometry which enables an optimal choice for the set of measurement values, i.e. an algorithm for optimizing the reconstruction setting is proposed. In Section 4 we employ the FEM package DIPOG (cf. [9]) and recover lithographic masks from scatterometric data. In particular, we consider periodic chrome-glass gratings with coated trapezoidal bridges and a period of 1 120 nm. We simulate diffracted efficiency measurements under inspecting light of a wavelength of 632.8 nm. From a huge set of measurement data we choose a subset optimal for the reconstruction, i.e. with minimal condition numbers of the Jacobians. Using this we are able to reconstruct a given grating with high accuracy. In particular we investigate the bias of different initial solutions and perturbed efficiencies on the accuracy of the reconstruction. We end up in a discussion of further research and open questions.

2 Mathematical Formulation of the Scattering Problem

2.1 Electro-Magnetic Fields Scattered by a periodic grating structure

We consider the diffraction by a periodic surface structure like in Figure 1 which is called grating. In order to describe the mathematical model (cf. e.g. the monographs [27, 33]), we choose a rectangular coordinate system such that the z-axis shows in the direction of the grooves and that the y-axis is orthogonal to

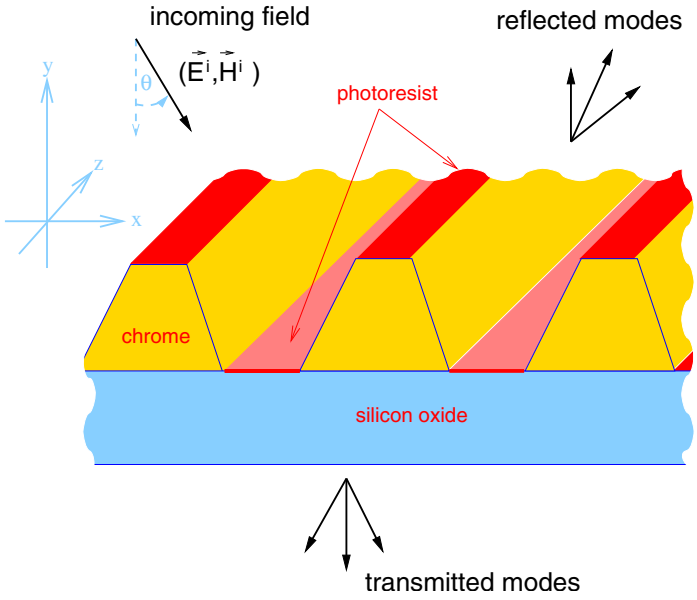


Figure 1: Periodic grating structure.

the surface plane containing the grooves. Hence, the geometric and material properties are constant in the z direction and periodic in the direction of x. The width of groove and bridge in x direction is the periodicity d (pitch) of the grating. The refractive index of the cover material is $\mathbf{n}^+$, that of the substrate under the grating surface structure $\mathbf{n}^-$. The grating part consists of several grating materials with different refractive indices (e.g. the two materials photoresist and chrome in Figure 1). Below the grating structure there may exist some coated layers, again, with different refractive indices (no layers in Figure 1). For simplicity sake, we consider classical diffraction, i.e. we suppose that a plane wave is incident from above with a direction located in the $x-y$ plane, i.e. in the plane perpendicular to the grooves. The angle of incidence θ is the angle between y-axis and direction of incidence. The wave length of the light in air is λ and we start with the case of TE polarization where the electric field vector is parallel to the grooves, i.e. it shows in the z direction. Hence, if μ_0 is the magnetic permeability of vacuum and c the speed of light, then the transverse z-coordinate of the electric field is given as

$$\begin{aligned} \mathcal{E}_z^{incident}(x,y,z,t) &:= E_z^{incident}(x,y,z)\ \exp(-\mathbf{i}\omega t)\,, \qquad \omega := \frac{2\pi c}{\lambda}\,, \\ E_z^{incident}(x,y,z) &:= \exp\left(\mathbf{i}(k^+\sin\theta\, x - k^+\cos\theta\, y)\right), \qquad k^+ := \omega\sqrt{\mu_0\varepsilon_0}\,\mathbf{n}^+. \end{aligned} \tag{2.1}$$

The light is diffracted by the grating structure. Besides some evanescent part the diffracted light splits into a finite number of reflected and transmitted TE polarized plane wave modes, the propagation directions of which are independent of the grating geometry and the grating materials. Simulating the diffraction means to determine the amplitude and the phase of the reflected and transmitted modes.

In the case of fields invariant in the z-direction the full system of Maxwell's equations reduces to a boundary value problem for the Helmholtz equation. More precisely, the transverse component E_z satisfies the scalar two-dimensional Helmholtz equation $\{\triangle + k^2\}E_z = 0$ in any domain of the cross section plane (x-y plane) with constant material as well as some transmission conditions on the interfaces between materials of different refractive indices. The wave number k is equal to ω/c times the refractive index of the material. Thus we can determine E_z numerically by approximate methods for elliptic partial differential equations, e.g. by the finite element method (cf. e.g. [34, 3, 10]).

Above the grating ($y \geq y_{max}$) resp. beneath the grating structure ($y \leq y_{min}$) the component E_z admits an expansion into the Rayleigh series of the form

$$E_z(x,y) = \sum_{n=-\infty}^{\infty} A_n^+ \exp\left(\mathbf{i}(\alpha_n x + \beta_n^+ y)\right) + A_0^{inc} \exp\left(\mathbf{i}(\alpha x - \beta_0^+ y)\right) , \quad \text{if } y \geq y_{max} \tag{2.2}$$

$$E_z(x,y) = \sum_{n=-\infty}^{\infty} A_n^- \exp\left(\mathbf{i}(\alpha_n x - \beta_n^- y)\right) , \quad \text{if } y \leq y_{min} \tag{2.3}$$

$$\beta_n^{\pm} := \sqrt{[k^{\pm}]^2 - [\alpha_n]^2} , \quad k^{\pm} := \frac{\omega \mathbf{n}^{\pm}}{c} , \qquad A_0^{inc} := 1$$

$$\alpha := k^+ \sin\theta, \qquad \alpha_n := k^+ \sin\theta + \frac{2\pi}{d} n .$$

Here d is the period of the grating and the complex constants $A_n^{\pm}$ are the so-called Rayleigh coefficients. The exclusions of incoming plane waves resp. of exponentially growing waves $\exp(\mathbf{i}(\alpha_n x - \beta_n^+ y))$ from (2.2) means that the Rayleigh expansion satisfies the outgoing wave (radiation) condition above the grating structure. Similarly, the exclusion of $\exp(\mathbf{i}(\alpha_n x + \beta_n^- y))$ from (2.3) implies the radiation condition below the structure. Note that the Rayleigh series expansions (2.2) and (2.3) are used also to define non-local boundary conditions at the upper and lower boundaries of the rectangular domain for the finite element computation when the infinite domain is restricted to a finite section. The interesting Rayleigh coefficients are those with $n \in \mathcal{U}^{\pm}$,

$$\mathcal{U}^{\pm} := \begin{cases} \{n \in \mathbb{Z} : |\alpha_n| < k^{\pm}\} & \text{if } \Im\mathrm{m}\, k^{\pm} = 0 \\ \emptyset & \text{if } \Im\mathrm{m}\, k^{\pm} > 0 \end{cases} .$$

Indeed, these coefficients $A_n^{\pm}$ describe magnitude and phase shift of the propagating plane waves. More precisely, the modulus $|A_n^{\pm}|$ is the amplitude of the nth reflected resp. transmitted wave mode and $\arg[A_n^{\pm}/|A_n^{\pm}|]$ the phase shift in comparison to the phase of the incoming wave. The terms with $n \notin \mathcal{U}^{\pm}$ lead to evanescent waves, only. The optical efficiencies of the grating are defined by

$$e_n^{\pm} := \frac{\beta_n^{\pm}}{\beta_0^+} \frac{|A_n^{\pm}|^2}{|A_0^{inc}|^2}, \quad (n,\pm) \in \left\{(n,+) : \ n \in \mathcal{U}^+\right\} \cup \left\{(n,-) : \ n \in \mathcal{U}^-\right\}, \tag{2.4}$$

which is the ratio of energy of the incident wave entailed to the nth propagating mode. Note that these efficiencies of propagating modes exist for non-absorbing materials, i.e. for $\Im\mathrm{m}\, k^{\pm} = 0$.

The case of TM polarization is quite similar to TE. Indeed, this time the vector of the magnetic field $\mathcal{H}$ points into the direction of the grooves, i.e. in the direction of the z axis. Analogously to formula (2.1) given above for the incident electric field, we get

$$\begin{aligned} \mathcal{H}_z^{incident}(x,y,z,t) &:= H_z^{incident}(x,y,z)\ \exp(-\mathbf{i}\omega t)\,, \\ H_z^{incident}(x,y,z) &:= \exp\left(\mathbf{i}(k^+ \sin\theta\, x - k^+ \cos\theta\, y)\right), \quad k^+ := \omega\sqrt{\mu_0 \varepsilon_0}\, \mathbf{n}^+, \end{aligned} \tag{2.5}$$

for the z component of the incident magnetic field $\mathcal{H}_z^{incident}$. This z component of the complete field H_z satisfies the Helmholtz equation $\{\triangle + k^2\} H_z = 0$ in any domain of the cross section plane with constant materials. In comparison to the TE case, the transmission conditions on the interfaces are different. We can solve the transmission problem of the Helmholtz equation by the finite element method (cf. e.g. [34, 3, 10]). Again we have a finite number of transmitted and reflected modes and the Rayleigh expansions hold for E_z replaced by H_z. More precisely, the Rayleigh coefficients are the $B_n^{\pm}$ of the expansions

$$H_z(x,y) = \sum_{n=-\infty}^{\infty} B_n^+ \exp\left(\mathbf{i}(\alpha_n x + \beta_n^+ y)\right) + B_0^{inc} \exp\left(\mathbf{i}(\alpha x - \beta_0^+ y)\right) , \quad \text{if } y \geq y_{max} , \quad B_0^{inc} := 1 , \tag{2.6}$$

$$H_z(x,y) = \sum_{n=-\infty}^{\infty} B_n^- \exp\left(\mathbf{i}(\alpha_n x - \beta_n^- y)\right) , \quad \text{if } y \leq y_{min} \tag{2.7}$$

The objective is to compute the Rayleigh coefficients. They result from the finite element solution of the new transmission problems and from the Fourier series expansion (2.6) and (2.7). The efficiencies (compare (2.4)) are computed by

$$e_n^{\pm} = \frac{\beta_n^{\pm}}{\beta_0^{+}} \frac{[k^{+}]^2}{[k^{\pm}]^2} \frac{|B_n^{\pm}|^2}{|B_0^{inc}|^2}, \quad (n,\pm) \in \Big\{(n,+):\ n \in \mathcal{U}^{+}\Big\} \cup \Big\{(n,-):\ n \in \mathcal{U}^{-}\Big\}. \tag{2.8}$$

The surface geometry of the periodic grating structure is determined by one period of the cross section in the $x-y$ plane. If the domains of different materials are separated by polyhedral interfaces, then the corresponding interface profile curves of the cross section are polygonal. In other words, the geometry is determined by the x- and y-coordinates of the polygonal corners which can be considered as parameters in a polygonal representation of a class of geometries. Additional parameters are the refractive indices of the grating materials. Now we fix the state of polarization to TE or TM, the wave length, and the angle of incidence. Then the efficiencies $e_n^{\pm}$, the phase shifts $\arg[A_n^{\pm}/|A_n^{\pm}|]$ and $\arg[B_n^{\pm}/|B_n^{\pm}|]$ are functions of the geometric and material parameters. The gradients of these functions with respect to the parameter set can be computed by approximate methods, e.g. by the finite element method (cf. [11] for the derivatives w.r.t. the points and derive similar formulas for the derivatives w.r.t. the refractive indices).

2.2 Inverse Problem of Scatterometry

In scatterometry, we need to solve the inverse problem corresponding to the computation described in the last section. In other words, we suppose the efficiencies $e_n^{\pm}$ of the propagating reflected and transmitted plane wave modes (cf. (2.4) and (2.8)) and the phase shift differences $\arg[A_n^{\pm}/|A_n^{\pm}|] - \arg[B_n^{\pm}/|B_n^{\pm}|]$ are measured for a fixed finite set of wave lengths and a fixed finite set of incidence angles and polarization states. From all these data or from a certain part of it, we try to determine the periodic grating structure corresponding to the data. Unfortunately, this inverse problem is, from the mathematical point of view, a severely ill-posed problem. Consequently, the determination of the geometry and the refractive indices is extremely difficult. Tiny errors in the measurement data usually result in huge errors of the reconstructed grating structure. Applying regularization techniques the solution of inverse problems improve. Nevertheless the accuracy is much less than that of well posed problems. A better reconstruction is possible only if more a-priori knowledge is used. This can be realized by restricting the search to a smaller class of grating structures determined by a small number of parameters. Thus the ill-posed inverse problem is reduced to a well-posed problem of parameter reconstruction.

We consider a fixed class of gratings which can be described by the parameter vector $h = (h_n)_{n\in\mathcal{N}}$ of real parameters h_n depending on a complex index $n \in \mathcal{N}$ of a finite index set $\mathcal{N}$ of cardinality N. These parameters can be e.g.:

- geometry parameters like heights, widths or corner coordinates of triangular resp. trapezoidal domains occupied by certain material components
- the real or imaginary part of the refractive indices of certain material components

We suppose that the grating depends on the geometry parameters in such a way that the efficiencies and phase shifts depend smoothly on these parameters. The dependence on the refractive indices is smooth if the substrate and cover materials are fixed since the corresponding partial differential operators are analytic functions with respect to the constant coefficients of the Helmholtz equation and the transmission conditions. If the geometry parameters describe a polygonal structure and if each change in the parameters results in a homeomorphic change of the geometry only, then the dependence on the geometry parameters is smooth, too (cf. [11], where the derivative w.r.t. the geometry parameters is represented as a derivative of a fixed differential operator composed by an analytic family of isomorphisms of the underlying domain). Besides the smooth dependence, we suppose that the set of admissible (feasible) parameters is defined by lower and upper bounds h_n^{lo} and h_n^{up}, i.e., that the parameters satisfy the constraints

$$h_n^{lo} \le h_n \le h_n^{up},\ n \in \mathcal{N}. \tag{2.9}$$

Clearly, for each fixed class of geometries a lot of parametric descriptions are possible. To simplify the numerical optimization algorithm for the reconstruction problem, we recommend to choose the parameter set such that the constraints (feasibility conditions) for a meaningful parameter set are simple. Indeed we shall assume that the only restrictions are the upper and lower bounds given in (2.9) for all parameters. On the other hand, each parameter can be multiplied by a normalization factor. We suppose that these normalization factors are defined such that the desired accuracy tolerances for all parameters coincide.

We denote the efficiency and the phase shift difference values forming the data of measurement, from which the grating structure is to be reconstructed, by $(E_m)_{m\in\mathcal{M}_0}$. More precisely, for each parameter set $h = (h_n)_{n\in\mathcal{N}}$, there corresponds a set of efficiency and phase shift difference values $(E_m^0(h))_{m\in\mathcal{M}_0}$. The measured values corresponding to the exact solution which is to be reconstructed are denoted by $(E_m^{ex})_{m\in\mathcal{M}_0}$. Hence, the inverse problem is to find the parameter set $h^{op} = (h_n^{op})_{n\in\mathcal{N}}$ which satisfies (2.9) and

$$\left(E_m^0(h_n^{op})\right)_{m\in\mathcal{M}_0} = \left(E_m^{ex}\right)_{m\in\mathcal{M}_0} \tag{2.10}$$

2.3 Numerical Algorithms for the Inverse Problem of Scatterometry

If the measurement values are not accurate, then the existence of a solution for (2.10) cannot be guaranteed. In this case the difference of the left and right-hand side in (2.10) should be kept minimal. In other words, the inverse problem can be reformulated as the following optimization problem: Find a parameter vector $h^{op} = (h_n^{op})_{n\in\mathcal{N}}$ such that

$$\begin{aligned} f\left(h^{op}\right) &= \min_{\substack{h=(h_n)_{n\in\mathcal{N}}:\\ h_n^{lo}\le h_n\le h_n^{up}}} f\left(h\right), \qquad (2.11)\\ f\left(h\right) &:= \left\| \left(E_m^0(h)\right)_{m\in\mathcal{M}_0} - \left(E_m^{ex}\right)_{m\in\mathcal{M}_0} \right\|^2 := \sum_{m\in\mathcal{M}_0} \left|E_m^0(h) - E_m^{ex}\right|^2. \end{aligned}$$

To (2.11) we can apply any of the known optimization methods. A first algorithm is to simulate the diffraction for all gratings with parameters h from a fine grid of the box defined by the constraints (2.9) and to determine that set of parameters h from this grid for which the functional value $f(h)$ is minimal. This algorithm of best fitting (cf. e.g. [28]) requires a huge amount of work for the precomputation of the data $E_m^0(h)$ for h from the grid of parameter sets. The subsequent determination of the minimum over the grid can be fast. Alternatively, stochastic global algorithms like e.g. simulated annealing (cf. [19]) or genetic algorithms (cf. [2]) render the solution of (2.11) with high probability. Gradient based local methods like e.g. conjugate gradient method (cf. e.g. [26]) or interior point method (cf. e.g. [17]) render at least so-called local solutions. Recall that a local solution of (2.11) is a solution $h^{locop} = (h_n^{locop})_{n\in\mathcal{N}}$ for which the constraints $h_n^{lo} \le h_n^{locop} \le h_n^{up}$, $n \in \mathcal{N}$ are satisfied and for which there exists a small $\varepsilon > 0$ such that $f(h^{locop}) \le f(h)$ holds at least for all $h = (h_n)_{n\in\mathcal{N}}$ with $\|h^{locop} - h\| \le \varepsilon$ and (2.9). In contrast to local solutions, the solution of (2.11) is called global. If a local method is applied, then a clever guess of the initial solution or a certain number of restarts from different initial solutions can help to find the global solution.

Unfortunately, the efficiency of the optimization algorithm depends strongly on the right scaling of the objective function. Locally, the objective function is like

$$\begin{aligned} f(h) &\approx \left\| \left(E_m^0(h^{ex})\right)_{m\in\mathcal{M}_0} + A^{ex}_{\mathcal{M}_0\mathcal{N}}(h - h^{ex}) - \left(E_m^{ex}\right)_{m\in\mathcal{M}_0} \right\|^2 \qquad (2.12)\\ A^{ex}_{\mathcal{M}_0\mathcal{N}} &:= \left(\frac{\partial E_m^0}{\partial h_n}\left(h^{ex}\right) \right)_{\substack{m\in\mathcal{M}_0\\ n\in\mathcal{N}}} \in \mathbb{R}^{M_0\times N}. \end{aligned}$$

The gradient of this quadratic approximate functional is $2[A^{ex}_{\mathcal{M}_0\mathcal{N}}]^T A^{ex}_{\mathcal{M}_0\mathcal{N}} h$ plus some constant vector. Consequently, neglecting the constraints, the minimization of the quadratic functional is equivalent to

finding the zero of its gradient, i.e., to solving a matrix equation with the matrix $[A^{ex}_{\mathcal{M}_0\mathcal{N}}]^T A^{ex}_{\mathcal{M}_0\mathcal{N}}$. Hence, the condition number $\mathrm{cond}([A^{ex}_{\mathcal{M}_0\mathcal{N}}]^T A^{ex}_{\mathcal{M}_0\mathcal{N}})$ controls the conditioning of f.

Recall that the condition number of a selfadjoint matrix C is defined as $\mathrm{cond}(C) = \|C\| \cdot \|C^{-1}\|$, where $\|C\|$ is the operator norm of C acting as a linear mapping in the Euclidean space $\mathbb{R}^N$. In other words, $\mathrm{cond}(C)$ is the ratio of the largest eigenvalue of C divided by the smallest eigenvalue. It is well known that the condition number of a matrix $\mathrm{cond}(C)$ is a measure for the solvability properties of the matrix equation $Cx = y$ with $x, y \in \mathbb{R}^N$. A minimal value $\mathrm{cond}(C) = 1$ implies that C is the identity matrix with perfect solvability properties, whereas equations $Cx = y$ with large $\mathrm{cond}(C)$ are difficult to solve. Matrices with small condition number are called well conditioned.

An alternative to the optimization methods for the objective functional f from (2.11) is to apply a Newton type method directly to the operator equation (2.10). Similarly to (2.12), we conclude

$$\begin{aligned} (E^{ex}_m)_{m\in\mathcal{M}_0} &= \left(E^0_m(h^{ex})\right)_{m\in\mathcal{M}_0} \approx \left(E^0_m(h)\right)_{m\in\mathcal{M}_0} + A_{\mathcal{M}_0\mathcal{N}}(h^{ex}-h) \qquad (2.13)\\ & \qquad\qquad A_{\mathcal{M}_0\mathcal{N}} := \left(\frac{\partial E^0_m}{\partial h_n}(h)\right)_{\substack{m\in\mathcal{M}_0\\ n\in\mathcal{N}}}\\ A_{\mathcal{M}_0\mathcal{N}}(h^{ex}-h) &\approx (E^{ex}_m)_{m\in\mathcal{M}_0} - \left(E^0_m(h)\right)_{m\in\mathcal{M}_0} \end{aligned}$$

The least square solution of the last non-quadratic linear system is given by

$$h^{ex} \approx h + \left[[A_{\mathcal{M}_0\mathcal{N}}]^T A_{\mathcal{M}_0\mathcal{N}}\right]^{-1} [A_{\mathcal{M}_0\mathcal{N}}]^T \left[(E^{ex}_m)_{m\in\mathcal{M}_0} - \left(E^0_m(h)\right)_{m\in\mathcal{M}_0}\right] .$$

In view of this formula Drège, Al-Assaad, and Byrne [8] have suggested the following iterative scheme. Choose an initial solution $h^0 = (h^0_n)_{n\in\mathcal{N}}$ and, for any iterative solution $h^l = (h^l_n)_{n\in\mathcal{N}}$, $l \geq 0$, define a new iterate $h^{l+1} = (h^{l+1}_n)_{n\in\mathcal{N}}$ by

$$\begin{aligned} h^{l+1} &= h^l + \left[[A^l_{\mathcal{M}_0\mathcal{N}}]^T A^l_{\mathcal{M}_0\mathcal{N}}\right]^{-1} [A^l_{\mathcal{M}_0\mathcal{N}}]^T \left[(E^{ex}_m)_{m\in\mathcal{M}_0} - \left(E^0_m(h^l)\right)_{m\in\mathcal{M}_0}\right] , \qquad (2.14)\\ & \qquad A^l_{\mathcal{M}_0\mathcal{N}} := \left(\frac{\partial E^0_m}{\partial h_n}\left(h^l\right)\right)_{\substack{m\in\mathcal{M}_0\\ n\in\mathcal{N}}} . \end{aligned}$$

If the iterative solution of (2.14) does not satisfy the constraints $h^{lo}_n \leq h^{l+1}_n \leq h^{up}_n$, then we define $h^{l+1} = (h^{l+1}_n)_{n\in\mathcal{N}}$ to be the solution of the following optimization problem of a convex quadratic functional defined over a box domain.

$$f_l\left(h^{l+1}\right) = \min_{\substack{h=(h_n)_{n\in\mathcal{N}}:\\ h^{lo}_n\leq h_n\leq h^{up}_n}} f_l(h) , \quad f_l(h) := \left\| \left(E^0_m(h^l)\right)_{m\in\mathcal{M}_0} + A^l_{\mathcal{M}_0\mathcal{N}}(h-h^l) - (E^{ex}_m)_{m\in\mathcal{M}_0} \right\|^2 . \qquad (2.15)$$

Clearly, any solution of (2.15) satisfying the constraint conditions (2.9) strictly, i.e. $h^{lo}_n < h_n < h^{up}_n$, $n \in \mathcal{N}$, solves the equation (2.14).

In some cases this Newton type method converges quite fast although only first order derivatives are used. The number of iterations is less than those for the conjugate gradient and interior point methods applied to (2.11). The iteration steps are even faster for the Newton type method (2.15) since no line search with time consuming function evaluations is needed. However, similarly as for the gradient based methods applied to (2.11), the conditioning of the matrices $[A^l_{\mathcal{M}_0\mathcal{N}}]^T A^l_{\mathcal{M}_0\mathcal{N}}$ plays a crucial role (cf. (2.14) or look at the gradient of f_l).

The theoretical convergence properties of the Newton type method are, roughly speaking, as follows. If the equation (2.10) has a solution, then the iteration converges very fast (quadratically). If (2.10) has no exact solution but a generalized solution is defined by (2.11) and if the efficiency values for the generalized solution are close to the measured data (deviation $\|(E^0_m(h^{op}))_{m\in\mathcal{M}_0} - (E^{ex}_m)_{m\in\mathcal{M}_0}\|$ less than a certain problem dependent small ε), then the iteration still converges fast (linearly) to the generalized

approximate solution. Finally, if the efficiency values for the generalized solution are not close to the measured data, then the iteration need not to be convergent. After sufficiently large iteration steps, however, the deviation of the iterative solutions from the generalized solution is less than a certain constant times the deviation $\|(E_m^0(h^{op}))_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0}\|$ of the efficiency values for the generalized solution from the measured data.

To prepare the precise formulation of convergence properties for the method (2.15), we recall the following condition and definition. Any local minimum h^∞ satisfies the necessary KKT conditions (cf. e.g. [26]) for a minimum of the functional f if, for any $1 \le n \le N$, one of the following alternatives is satisfied:

$$\begin{aligned}
&\text{either } h_n^{lo} < h_n^\infty < h_n^{up} \text{ and } \partial f/\partial h_n(h^\infty) = 0\\
&\text{or } h_n^{lo} = h_n^\infty \text{ and } \partial f/\partial h_n(h^\infty) \ge 0\\
&\text{or } h_n^\infty = h_n^{up} \text{ and } \partial f/\partial h_n(h^\infty) \le 0
\end{aligned}$$

If the inequalities for the cases $h_n^{lo} = h_n^\infty$ and $h_n^\infty = h_n^{up}$ are strict, then the KKT conditions are said to be satisfied with strict complementarity. We have the following local convergence results.

Lemma 2.1. *i) If the sequence of iterative solutions of (2.15) converges to a limit h^∞, then this limit satisfies the constraints (2.9) and at least the necessary optimality conditions, i.e. the KKT conditions.*

ii) Suppose h^{op} is the optimal solution of (2.11). Hence the KKT conditions hold. However, we suppose a little bit more. Namely the KKT conditions should hold with strict complementarity. Moreover, we assume that the prescribed values are attained, i.e. $(E_m^0(h^{op}))_{m\in\mathcal{M}_0} = (E_m^{ex})_{m\in\mathcal{M}_0}$. Finally, we assume that the chosen set of measurements is large enough such that the matrix $[A^{op}_{\mathcal{M}_0\mathcal{N}_{na}}]^T A^{op}_{\mathcal{M}_0\mathcal{N}_{na}}$ is invertible and that the kernel of the matrix $A^{op}_{\mathcal{M}_0\mathcal{N}}$ is trivial, where

$$A^{op}_{\mathcal{M}_0\mathcal{N}_{na}} := \left(\frac{\partial E_m^0}{\partial h_n}(h^{op})\right)_{\substack{m\in\mathcal{M}_0\\ n\in\mathcal{N}_{na}}}, \quad A^{op}_{\mathcal{M}_0\mathcal{N}} := \left(\frac{\partial E_m^0}{\partial h_n}(h^{op})\right)_{\substack{m\in\mathcal{M}_0\\ n\in\mathcal{N}}},$$

and $\mathcal{N}_{na} := \{n \in \mathcal{N} : h_n^{lo} < h_n^{op} < h_n^{up}\}$ is the set of non-active indices. Then, for any initial solution h^0 sufficiently close to h^{op}, the iteration (2.15) converges quadratically, i.e. $\|h^l - h^{op}\| \longrightarrow 0$ for $l \to \infty$ and $\|h^{l+1} - h^{op}\| \le \text{const.}\, \|h^l - h^{op}\|^2$.

iii) Suppose the assumptions of ii) are satisfied with the exception that the prescribed values are not attained. Instead suppose that the difference norm $\|(E_m^0(h^{op}))_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0}\|$ is sufficiently small. Then, for any initial solution h^0 sufficiently close to h^{op}, the iteration (2.15) converges linearly, i.e. we get the estimate $\|h^{l+1} - h^{op}\| \le q\|h^l - h^{op}\|$ with a constant $0 < q < 1$.

iv) Suppose the assumptions of iii) are satisfied with the exception that $\|(E_m^0(h^{op}))_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0}\|$ is not sufficiently small. Additionally, suppose $[A_{\mathcal{M}_0\mathcal{N}_{na}}]^T A_{\mathcal{M}_0\mathcal{N}_{na}}$ is invertible and the inverse is uniformly bounded for all h satisfying (2.9), where

$$A_{\mathcal{M}_0\mathcal{N}_{na}} := \left(\frac{\partial E_m^0}{\partial h_n}(h)\right)_{\substack{m\in\mathcal{M}_0\\ n\in\mathcal{N}_{na}}}.$$

Then there is a constant $c > 0$ and a small constant $\delta > 0$ such that: For all sequences of iterative solutions defined by (2.15) with

$$\begin{aligned}
&\|h^0 - h^{op}\| \le \delta\,,\\
&\forall n \in \mathcal{N}:\ h_n^l = h_n^{up} \Leftrightarrow h_n^{op} = h_n^{up}\,, &(2.16)\\
&\forall n \in \mathcal{N}:\ h_n^l = h_n^{lo} \Leftrightarrow h_n^{op} = h_n^{lo}\,, &(2.17)
\end{aligned}$$

the iterative solutions of (2.15) satisfy the estimate $\|h^l - h^{op}\| \le c\|(E_m^0(h^{op}))_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0}\|$ provided l is sufficiently large.

Proof. i) If the sequence of iterative solutions converges, we can consider the limit $l \to \infty$ in the optimality relations $f_l(h) \geq f_l(h^{l+1})$. For any $h \in \mathbb{R}^N$, the limit point h^∞ satisfies

$$\begin{aligned}
\left\| \left(E_m^0(h^\infty)\right)_{m\in\mathcal{M}_0} + A^\infty_{\mathcal{M}_0\mathcal{N}}(h-h^\infty) - (E_m^{ex})_{m\in\mathcal{M}_0} \right\|^2 &\geq \left\| \left(E_m^0(h^\infty)\right)_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0} \right\|^2 , \\
A^\infty_{\mathcal{M}_0\mathcal{N}} &:= \left(\frac{\partial E_m^0}{\partial h_n}(h^\infty) \right)_{\substack{m\in\mathcal{M}_0 \\ n\in\mathcal{N}}} , \\
\left\| \left(E_m^0(h)\right)_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0} \right\|^2 + \mathcal{O}\left(\|h-h^\infty\|^2\right) &\geq \left\| \left(E_m^0(h^\infty)\right)_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0} \right\|^2 .
\end{aligned}$$

Obviously, the last equation implies that h^∞ satisfies the necessary KKT conditions.

ii) The gradient of f at h^{op} is equal to $2[A^{op}_{\mathcal{M}_0\mathcal{N}}]^T((E_m^0(h^{op}))_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0}))$. Due to the KKT conditions we have

$$[A^{op}_{\mathcal{M}_0\mathcal{N}_{na}}]^T \left(\left(E_m^0(h^{op})\right)_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0}) \right) = 0 . \tag{2.18}$$

We denote the orthogonal projection of $\mathbb{R}^N$ onto the vectors with non-zero components only for index in $\mathcal{N}_{na}$ by P and set $Q = I - P$. Using the Taylor series expansion of E_m^0 at h^l, we conclude from (2.18) that

$$\begin{aligned}
\left(E_m^0(h^l)\right)_{m\in\mathcal{M}_0} + A^l_{\mathcal{M}_0\mathcal{N}}(h^{op}-h^l) + \mathcal{O}\left(\left\|h^{op}-h^l\right\|^2\right) &= E_m^0(h^{op}) , \\
A^l_{\mathcal{M}_0\mathcal{N}} &:= \left(\frac{\partial E_m^0}{\partial h_n}(h^l) \right)_{\substack{m\in\mathcal{M}_0 \\ n\in\mathcal{N}}} , \\
\left[\left[A^{op}_{\mathcal{M}_0\mathcal{N}_{na}}\right]^T A^l_{\mathcal{M}_0\mathcal{N}_{na}} \right] \left[P(h^{op}-h^l)\right] &= \left[A^{op}_{\mathcal{M}_0\mathcal{N}_{na}}\right]^T \Big[(E_m^{ex})_{m\in\mathcal{M}_0} - \left(E_m^0(h^l)\right)_{m\in\mathcal{M}_0} \\
&\quad - A^l_{\mathcal{M}_0\mathcal{N}} Q(h^{op}-h^l) \Big] + \mathcal{O}\left(\left\|h^{op}-h^l\right\|^2\right) , \\
\left[\left[A^{l}_{\mathcal{M}_0\mathcal{N}_{na}}\right]^T A^l_{\mathcal{M}_0\mathcal{N}_{na}} \right] \left[P(h^{op}-h^l)\right] &= \left[A^{l}_{\mathcal{M}_0\mathcal{N}_{na}}\right]^T \Big[(E_m^{ex})_{m\in\mathcal{M}_0} - \left(E_m^0(h^l)\right)_{m\in\mathcal{M}_0} \\
&\quad - A^l_{\mathcal{M}_0\mathcal{N}} Q(h^{op}-h^l) \Big] \\
&\quad + \left[\left[A^{op}_{\mathcal{M}_0\mathcal{N}_{na}}\right]^T - \left[A^{l}_{\mathcal{M}_0\mathcal{N}_{na}}\right]^T \right] \left((E_m^{ex})_{m\in\mathcal{M}_0} - \left(E_m^0(h^{op})\right)_{m\in\mathcal{M}_0} \right) \\
&\quad + \mathcal{O}\left(\left\|h^{op}-h^l\right\|^2\right) .
\end{aligned} \tag{2.19}$$

On the other hand, if we try to solve the optimization problem (2.15) under the only restrictions $Qh^{l+1} = Qh^{op}$, then the iteration vector h^{l+1} is the solution of

$$\begin{aligned}
\left[\left[A^{l}_{\mathcal{M}_0\mathcal{N}_{na}}\right]^T A^l_{\mathcal{M}_0\mathcal{N}_{na}} \right] \left[P(h^{l+1}-h^l)\right] &= \left[A^{l}_{\mathcal{M}_0\mathcal{N}_{na}}\right]^T \Big[(E_m^{ex})_{m\in\mathcal{M}_0} - \left(E_m^0(h^l)\right)_{m\in\mathcal{M}_0} \\
&\quad - A^l_{\mathcal{M}_0\mathcal{N}} Q(h^{op}-h^l) \Big]
\end{aligned} \tag{2.20}$$

If we compare the right-hand sides of (2.20) and (2.19) and if we take into account that by assumption the second term on the right-hand side of (2.19) vanishes, then we conclude $\|h^{l+1}-h^{op}\| \leq \text{const.}\, \|h^l-h^{op}\|^2$. In other words, if h^{l+1} from (2.20) is really the minimal solution of (2.15), then the quadratic convergence follows. However, if h^l and therewith h^{l+1} is close to h^{op}, then (2.20), the continuity of the mappings, and the complementarity of the KKT condition for h^{op} imply that the solution h^{l+1} of (2.20) satisfies the KKT conditions for f_l, too. This condition, however, is sufficient for h^{l+1} to be the minimal solution since the functional f_l is quadratic and, due to the triviality of the kernel $A^l_{\mathcal{M}_0\mathcal{N}}$, strictly convex.

iii) If the prescribed value is not attained, but the difference norm $\|(E_m^0(h^{op}))_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0}\|$ is smaller than a small ε, then the second term on the right-hand side of (2.19) is less than a constant times $\varepsilon \left\|h^l-h^{op}\right\|$. Consequently, comparing (2.19) and (2.20), we arrive at the estimate

$$\left\|h^{l+1}-h^{op}\right\| \leq \text{const.}\, \varepsilon \left\|h^l-h^{op}\right\| + \text{const.} \left\|h^l-h^{op}\right\|^2 \tag{2.21}$$

which proves the linear rate of convergence if const. $\varepsilon < q < 1$.

iv) The last assertion follows from (2.21) which holds with ε equal to $\|(E_m^0(h^{op}))_{m\in\mathcal{M}_0} - (E_m^{ex})_{m\in\mathcal{M}_0}\|$. However, since a convergence of the iterative solutions cannot be proved anymore, the derivation of (2.21) requires the uniform bound of the $\{[A_{\mathcal{M}_0\mathcal{N}_{na}}]^T A_{\mathcal{M}_0\mathcal{N}_{na}}\}^{-1}$ for all h. Moreover, the limit argument showing that a vector h^{l+1} with $Qh^{l+1} = Qh^{op}$ is really the next iterate does not work either. Consequently, we have to assume the a-priori restriction (2.16) and (2.17) for the iterative solutions. □

3 Sensitivity Analysis of Geometric Parameters with respect to Measurement Data

3.1 Measured Data

We introduce a "complete" set of all possible measurements. From this set, we have to extract a smaller set of measurements which is optimal for the reconstruction of the geometry and material composition of the grating. Here, optimality will be a compromise between high accuracy of the reconstructed data on the one hand and the fast access to measurement data as well its fast processing in a numerical algorithm on the other hand.

The complete data of measurement is a vector $(E_m)_{m\in\mathcal{M}}$ of real values $E_m \in \mathbb{R}$ with m running over a general finite index set $\mathcal{M}$ of M indices. For a given optical grating, E_m can be either the efficiency of a given reflected resp. transmitted mode under TE resp. TM polarization or the phase difference between the TE and the TM polarized modes for a given order of reflection resp. transmission. The efficiencies and phase differences can even be multiplied by a normalization factor. All together, the complex index m comprises the following information:

- type of measurement, i.e. efficiency or phase difference
- angle of incidence of the plane wave inciting the mode
- wave length of the incident light
- order of the reflected resp. transmitted mode
- polarization type TE or TM of the inciting plane wave if E_m is an efficiency[1]
- normalization factor of the measurement value

Of course, we must be careful joining completely different entities E_m into one vector. Should the efficiency numbers be given in per cent or should they be normalized to one? Should the angles of phase shift differences be given in degrees or in radians? All these normalization factors surely have a great influence on the sensitivity and on the inverse problems. We suggest that the measurement entities should be normalized such that the expected average measurement errors coincide for all $E_m,\ m \in \mathcal{M}$.

A simple example for such an $(E_m)_{n\in\mathcal{M}}$ can be the vector of all reflected and transmitted efficiencies corresponding to the propagating modes obtained for the incidence angles $\theta = -85°, -80°, \ldots, 85°$ and for the polarization states TE and TM. From $\mathcal{M}$, we shall extract a subset $\mathcal{M}_0 \subseteq \mathcal{M}$ of cardinality M_0, and the reconstruction of the grating will be based on the data $(E_m)_{m\in\mathcal{M}_0}$. The task is to find an optimal subset $\mathcal{M}_0$ to guarantee a fast and accurate reconstruction.

3.2 Completion of the Set of Parameters

So far, for each parameter vector $h = (h_n)_{n\in\mathcal{N}}$, there exists a unique grating in the grating class, and thus a unique set of data $(E_m)_{m\in\mathcal{M}}$ corresponding to this grating. This data can be obtained by measurement. We suppose that the noise of the measured data is negligible. In other words, we have a vector valued

[1] In case of phase shifts, the phase difference between TE and TM polarization is meant and no polarization type is to be fixed.

function

$$\mathbb{R}^N \ni h = (h_n)_{n\in\mathcal{N}} \mapsto (E_m)_{m\in\mathcal{M}} \in \mathbb{R}^M, \quad E_m = E_m^0(\dots, h_n, \dots) \tag{3.1}$$

defining a smooth N-dimensional submanifold of the Euclidean space $\mathbb{R}^M$.

For the next section, we need a one-to-one correspondence between parameters and complete data of measurement. Therefore, we introduce a supplement vector of parameters $(\tilde{h}_n)_{n\in\widetilde{\mathcal{N}}}$ of $\widetilde{N} := M - N$ real numbers such that the mapping

$$\mathbb{R}^M \ni \Big((h_n)_{n\in\mathcal{N}}, (\tilde{h}_n)_{n\in\widetilde{\mathcal{N}}}\Big) \mapsto (E_m)_{m\in\mathcal{M}} \in \mathbb{R}^M, \quad E_m = E_m(\dots, h_n, \dots, \tilde{h}_n, \dots) \tag{3.2}$$

is one-to-one at least between a neighborhood of the point $((h_n^{ex})_{n\in\mathcal{N}}, (0)_{n\in\widetilde{\mathcal{N}}})$ and a neighborhood of the image point. The functions of (3.2) are extension of those in (3.1) satisfying

$$E_m^0(\dots, h_n, \dots) = E_m(\dots, h_n, \dots, 0, \dots, 0).$$

Moreover, the extensions can be chosen such that the new tangent vectors at $((h_n^{ex})_{n\in\mathcal{N}}, (0)_{n\in\widetilde{\mathcal{N}}})$ are orthogonal to the old ones, i.e.

$$\left\langle \left(\frac{\partial E_m}{\partial h_n}\right)_{m\in\mathcal{M}}, \left(\frac{\partial E_m}{\partial \tilde{h}_{n'}}\right)_{m\in\mathcal{M}} \right\rangle \;:=\; \sum_{m\in\mathcal{M}} \frac{\partial E_m}{\partial h_n}\frac{\partial E_m}{\partial \tilde{h}_{n'}} \;=\; 0, \quad n\in\mathcal{N},\ n'\in\widetilde{\mathcal{N}} \tag{3.3}$$

where the arguments of the functions in (3.3) are the parameters $h_n = h_n^{ex},\ n\in\mathcal{N}$ and $\tilde{h}_n = 0,\ n\in\widetilde{\mathcal{N}}$.

Due to the one-to-one correspondence of the mapping in (3.2), we may consider the parameters as a function of the measured data, i.e.

$$\begin{aligned} \mathbb{R}^M \ni (E_m)_{m\in\mathcal{M}} \mapsto \Big((h_n)_{n\in\mathcal{N}}, (\tilde{h}_n)_{n\in\widetilde{\mathcal{N}}}\Big) \in \mathbb{R}^M, \qquad & h_n = h_n(\dots, E_m, \dots), \\ & \tilde{h}_n = \tilde{h}_n(\dots, E_m, \dots). \end{aligned} \tag{3.4}$$

Clearly, these functions are the inverse functions of (3.2), i.e.

$$\begin{aligned} E_m &= E_m\Big(\dots, h_n(\dots, E_{m'}, \dots), \dots, \tilde{h}_n(\dots, E_{m'}, \dots), \dots\Big),\ m\in\mathcal{M}, \\ h_n &= h_n\Big(\dots, E_m(\dots, h_{n'}, \dots, \tilde{h}_{n'}, \dots), \dots\Big),\ n\in\mathcal{N}, \\ \tilde{h}_n &= \tilde{h}_n\Big(\dots, E_m(\dots, h_{n'}, \dots, \tilde{h}_{n'}, \dots), \dots\Big),\ n\in\widetilde{\mathcal{N}}. \end{aligned} \tag{3.5}$$

3.3 Derivatives of Parameters with respect to Measurement Data

The numerical expression of the sensitivity of the reconstructed parameter h_n with respect to the measurement data value E_m is the partial derivative $\partial h_n/\partial E_m$. Unfortunately, this derivative cannot be computed directly. However, we can easily determine the derivatives $\partial E_m/\partial h_n$ of the inverse functions (cf. [11] and derive similar variational formulas for the derivatives w.r.t. the refractive indices) and get $\partial h_n/\partial E_m$ from these. To this end we derive the subsequent Equation (3.7).

Applying the chain rule to (3.5), we arrive at

$$\delta_{m,m'} \;=\; \frac{\partial E_m}{\partial E_{m'}} \;=\; \sum_{n\in\mathcal{N}} \frac{\partial E_m}{\partial h_n}\frac{\partial h_n}{\partial E_{m'}} + \sum_{n\in\widetilde{\mathcal{N}}} \frac{\partial E_m}{\partial \tilde{h}_n}\frac{\partial \tilde{h}_n}{\partial E_{m'}}.$$

This is equivalent to the system of matrix equations

$$e_{m'} = B\xi_{m'} + \widetilde{B}\tilde{\xi}_{m'}, \quad m' \in \mathcal{M} \tag{3.6}$$

$$B := \left(\frac{\partial E_m}{\partial h_n}\right)_{\substack{m\in\mathcal{M}\\ n\in\mathcal{N}}} \in \mathbb{R}^{M\times N}, \quad \widetilde{B} := \left(\frac{\partial E_m}{\partial \tilde{h}_n}\right)_{\substack{m\in\mathcal{M}\\ n\in\tilde{\mathcal{N}}}} \in \mathbb{R}^{M\times \tilde{N}},$$

$$\xi_{m'} := \left(\frac{\partial h_n}{\partial E_{m'}}\right)_{n\in\mathcal{N}} \in \mathbb{R}^N, \quad \tilde{\xi}_{m'} := \left(\frac{\partial \tilde{h}_n}{\partial E_{m'}}\right)_{n\in\tilde{\mathcal{N}}} \in \mathbb{R}^{\tilde{N}}, \quad e_{m'} := (\delta_{m,m'})_{m\in\mathcal{M}} \in \mathbb{R}^M.$$

Now we observe that the columns of the matrices B and $\widetilde{B}$ are the tangent vectors satisfying the orthogonality relations (3.3). Consequently, the image space of B is orthogonal to the image space of $\widetilde{B}$, and (3.6) implies that the vectors $\xi_{m'}$ are the solutions of the optimization problems

$$\min_{\xi_{m'}\in\mathbb{R}^N} \left\| B\xi_{m'} - e_{m'}\right\|^2, \quad m' \in \mathcal{M},$$

including the Euclidean norm $\|\cdot\|$. In other words, the vectors $\xi_{m'}$ of partial derivatives $\partial h_n/\partial E_{m'}$ are the zeros of the gradient of the quadratic functional, i.e., the solutions of the linear systems

$$[B^T B]\,\xi_{m'} = B^T e_{m'}, \quad m' \in \mathcal{M}. \tag{3.7}$$

3.4 Optimizing the Reconstruction by a Good Choice of Measurements

Suppose a grating determined by the "exact" parameter values $(h_n^{ex})_{n\in\mathcal{N}}$ is given. For this grating, we measure the data $(E_m)_{m\in\mathcal{M}_0}$ and denote the corresponding values by $(E_m^{ex})_{m\in\mathcal{M}_0}$. Using these, we try to reconstruct the exact parameter values $(h_n^{ex})_{n\in\mathcal{N}}$. Clearly, the efficiency and accuracy of the reconstruction will depend on the choice of $\mathcal{M}_0$. Moreover, reconstruction is meaningful only if the number N of values to be reconstructed is less or equal to the number M_0 of known data. A reconstruction from less data is possible only in rare cases of degenerate mappings from parameter data to measurements. Hence, we always suppose $N \le M_0$.

We introduce the Jacobi matrix $A_{\mathcal{M}\mathcal{N}}$ and its submatrix $A_{\mathcal{M}_0\mathcal{N}}$ by the formulae

$$A_{\mathcal{M}\mathcal{N}} := \left(\frac{\partial E_m^0}{\partial h_n}\right)_{\substack{m\in\mathcal{M}\\ n\in\mathcal{N}}} \in \mathbb{R}^{M\times N}, \quad A_{\mathcal{M}_0\mathcal{N}} := \left(\frac{\partial E_m^0}{\partial h_n}\right)_{\substack{m\in\mathcal{M}_0\\ n\in\mathcal{N}}} \in \mathbb{R}^{M_0\times N}.$$

Here the derivatives in the matrix entries are taken at the parameter vector h^{ex}. Reconstructing the parameter values h_n^{ex} from the measured values E_m, $m \in \mathcal{M}_0$ means to solve the non-linear equations (2.10). Locally, this mapping behaves like the linear mapping represented by its Jacobian matrix $A_{\mathcal{M}_0\mathcal{N}}$ (cf. (2.13)). In other words, (2.10) is a system which is easy to solve locally for values h close to the solution h^{ex} if the condition number $\mathrm{cond}([A_{\mathcal{M}_0\mathcal{N}}]^T A_{\mathcal{M}_0\mathcal{N}})$ of the product of $A_{\mathcal{M}_0\mathcal{N}}$ and its transposed $[A_{\mathcal{M}_0\mathcal{N}}]^T$ is small.

Now a good choice of a set $\mathcal{M}_0 \subseteq \mathcal{M}$ is that index set for which $\mathrm{cond}([A_{\mathcal{M}_0\mathcal{N}}]^T A_{\mathcal{M}_0\mathcal{N}})$ is small. Suppose we have a huge finite sequence $\mathcal{M}_{0,j}$, $j = 1,\dots,J$ of subsets of $\mathcal{M}$ such that the cardinality $M_{0,j}$ of $\mathcal{M}_{0,j}$ is greater or equal to N the number of parameters. These are the index sets of admissible measurement data $(E_m)_{m\in\mathcal{M}_{0,j}}$ from which we may choose $\mathcal{M}_0 = \mathcal{M}_{0,j}$ for the reconstruction task. Admissible means acceptable from the view point of technical or algorithmic restrictions fixed by the user. For the simplest case, the sequence of $\mathcal{M}_{0,j}$ could be the sequence of all subsets $\mathcal{M}_0 \subseteq \mathcal{M}$ of a fixed cardinality $M_{0,j} = M_0 > N$. In this sequence of index sets we choose the best $\mathcal{M}_{0,k}$, $1 \le k \le J$ for the reconstruction (2.10) by solving

$$\mathrm{cond}\left([A_{\mathcal{M}_{0,k}\mathcal{N}}]^T A_{\mathcal{M}_{0,k}\mathcal{N}}\right) = \min_{j=1,\dots,J} \mathrm{cond}\left([A_{\mathcal{M}_{0,j}\mathcal{N}}]^T A_{\mathcal{M}_{0,j}\mathcal{N}}\right). \tag{3.8}$$

Unfortunately, this is a discrete optimization problem which requires the test of the condition number for $\binom{M}{M_0}$ submatrices. To avoid this huge amount of computation we propose the following algorithm to determine a suboptimal solution of the problem (3.8). The main idea is not to choose the best submatrix at once, but to determine this submatrix by choosing its rows step by step.

The optimality of the solution of (3.8) is of local nature since the Jacobi matrix $A_{\mathcal{MN}}$ is defined for the fixed parameter set h^{ex}. In order to make the notion of optimality more stable, we recommend to compute $A_{\mathcal{MN}}$ at h^{ex} and at a few small perturbations of h^{ex}. For the optimization the single condition numbers in (3.8) can be replaced by the maximum of the condition numbers taken over h^{ex} and its perturbations.

3.5 Fast Choice of a Suboptimal Set of Measurement Data

We denote the jth row vector of $A_{\mathcal{MN}}$ by $a_j := (\partial E_j^0/\partial h_n)_{n\in\mathcal{N}}$. Here j is an index from the index set $\mathcal{M}$ which we identify with the set $\mathcal{M} = \{j : 1 \le j \le M\}$. Our task is to choose indices $j_1,\ j_2,\ \dots,\ j_{M_0}$ with $M_0 \ge N$ such that the condition number $\mathrm{cond}([A_{\mathcal{M}_0\mathcal{N}}]^T A_{\mathcal{M}_0\mathcal{N}})$ with $A_{\mathcal{M}_0\mathcal{N}}$, consisting of the M_0 rows $a_j,\ j = j_1,\ , j_2, \dots,\ j_{M_0}$, is small.

We choose an integer parameter M_C which is larger than M_0 the desired number of measurements but less than M such that the computer capacity allows to determine the optimal submatrix of M_0 rows a_j with a_j chosen from a set of M_C rows a_j. Now we take the first M_C rows a_j with $j \in \mathcal{M}_{00} := \{1, \dots, M_C\}$ and determine the index set $\mathcal{M}_1 \subseteq \mathcal{M}_{00}$ of optimal rows $a_j,\ j \in \mathcal{M}_1$ such that

$$\mathrm{cond}\left([A_{\mathcal{M}_1\mathcal{N}}]^T A_{\mathcal{M}_1\mathcal{N}}\right) = \min_{\mathcal{M}'=\{j_1,\dots,j_{M_0}\}\subseteq\mathcal{M}_{00}} \mathrm{cond}\left([A_{\mathcal{M}'\mathcal{N}}]^T A_{\mathcal{M}'\mathcal{N}}\right).$$

Next we form the new subset $\mathcal{M}_{01}$ as the union of $\mathcal{M}_1$ with $\{M_C + 1, \dots, M_C + (M_C - M_0)\}$ and determine the index set $\mathcal{M}_2 \subseteq \mathcal{M}_{01}$ of optimal rows $a_j,\ j \in \mathcal{M}_2$ such that

$$\mathrm{cond}\left([A_{\mathcal{M}_2\mathcal{N}}]^T A_{\mathcal{M}_2\mathcal{N}}\right) = \min_{\mathcal{M}'=\{j_1,\dots,j_{M_0}\}\subseteq\mathcal{M}_{01}} \mathrm{cond}\left([A_{\mathcal{M}'\mathcal{N}}]^T A_{\mathcal{M}'\mathcal{N}}\right).$$

Then we form the subset $\mathcal{M}_{02}$ as the union of $\mathcal{M}_2$ with $\{M_C + (M_C - M_0) + 1, \dots, M_C + 2(M_C - M_0)\}$ and determine the index set $\mathcal{M}_3 \subseteq \mathcal{M}_{02}$ of optimal rows $a_j,\ j \in \mathcal{M}_3$ such that

$$\mathrm{cond}\left([A_{\mathcal{M}_3\mathcal{N}}]^T A_{\mathcal{M}_3\mathcal{N}}\right) = \min_{\mathcal{M}'=\{j_1,\dots,j_{M_0}\}\subseteq\mathcal{M}_{02}} \mathrm{cond}\left([A_{\mathcal{M}'\mathcal{N}}]^T A_{\mathcal{M}'\mathcal{N}}\right).$$

We continue this way for further $k-3$ steps until $M_C + (k-1)\cdot(M_C - M_0) + 1 > M$ and the set of all indices $\{1, 2, \dots, M\}$ is exhausted. With the last optimal set of indices $\mathcal{M}_k$ we form the almost optimal set of measurements $\{E_j,\ j \in \mathcal{M}_k\}$.

3.6 Scaling of the Reconstruction Algorithm for a Fixed Set of Measurements

To improve the function f, we can scale the parameters replacing h_n by $h_n^{\#}/r_n$ with real scaling factors $r_n > 0$. Moreover, we can introduce positive weights w_m for the terms in the objective functional of (2.11). Together, we end up with the optimization problem

$$f^{\#}\Big(\dots, h_n^{\#op}, \dots\Big) = \min_{(h_n^{\#})_{n\in\mathcal{N}}} f^{\#}\Big(\dots, h_n^{\#}, \dots\Big), \tag{3.9}$$

$$f^{\#}\Big(\dots, h_n^{\#}, \dots\Big) := \sum_{m\in\mathcal{M}_0} w_m \big|E_m(\dots, h_n^{\#}/r_n, \dots) - E_m^{ex}\big|^2.$$

From this, the solution of the original problem is obtained by the backward substitution $h_n^{op} = h_n^{\#op}/r_n$. The local conditioning of the scaled optimization problem (3.9) is characterized by the condition number

$\mathrm{cond}([A^{\#}_{\mathcal{M}_0\mathcal{N}}]^T A^{\#}_{\mathcal{M}_0\mathcal{N}})$ of the corresponding Jacobi matrix $A^{\#}_{\mathcal{M}_0\mathcal{N}}$ corresponding to functional $f^{\#}$ given as

$$A^{\#}_{\mathcal{M}_0\mathcal{N}} \;=\; \left(\sqrt{w_m}\,\delta_{m,m'}\right)_{m,m'\in M_0} A_{\mathcal{M}_0\mathcal{N}} \left(\frac{\delta_{n,n'}}{r_n}\right)_{n,n'\in N}.$$

Clearly, the conditioning can be improved if we find scaling factors $r_n^{op} > 0$ and weights w_m^{op} such that the corresponding matrix $A^{\#op}_{\mathcal{M}_0\mathcal{N}}$ satisfies

$$\mathrm{cond}\left([A^{\#op}_{\mathcal{M}_0\mathcal{N}}]^T A^{\#op}_{\mathcal{M}_0\mathcal{N}}\right) \;=\; \min_{\substack{r_n:\,0<r_n,\;n\in\mathcal{N}\\ w_m:\,0<w_m,\;m\in\mathcal{M}_0}} \mathrm{cond}\left([A^{\#}_{\mathcal{M}_0\mathcal{N}}]^T A^{\#}_{\mathcal{M}_0\mathcal{N}}\right). \tag{3.10}$$

Since a solution of the last optimization problem is not so easy, we recommend to apply heuristic rules. For instance, the scaling factors r_n should be chosen such that the Euclidean norm of the rows in the matrix $A^{\#}_{\mathcal{M}_0\mathcal{N}}$ are almost of modulus equal to one. This corresponds to a preconditioning of $A^T_{\mathcal{M}_0\mathcal{N}} A_{\mathcal{M}_0\mathcal{N}}$ such that the preconditioned matrix $[A^{\#}_{\mathcal{M}_0\mathcal{N}}]^T A^{\#}_{\mathcal{M}_0\mathcal{N}}$ has diagonal entries of size one. Alternatively, the `LAPACK` routine `DGEEQU` can be applied to $A_{\mathcal{M}_0\mathcal{N}}$, which determines scaling factors $R(m),\ m = 1,\ldots,M_0$ and $C(n),\ n = 1,\ldots,N$ such that the condition number of $A_{\mathcal{M}_0\mathcal{N}}$ is reduced by scaling the matrix from the right by the $R(m)$ and from the left by the $C(n)$. Hence, we can choose $r_n = 1/C(n),\ n = 1,\ldots,N$ and $w_m := R(m)^2, m = 1,\ldots,M_0$.

Note that the last scaling is done to improve the convergence of numerical schemes after the normalization factors for the parameters (cf. Subsection 2.2) and the optimal set of measurements with their normalization factors (cf. Sect. 3.4 and 3.5) have been chosen. In other words, a first scaling by normalization factors is done to adapt the setting to the intrinsic requirements of the application and to enhance the accuracy of the reconstruction. The scaling of this section is not meant to improve accuracy, but only to improve the convergence of the numerical optimization algorithms. Indeed, in some of our numerical experiments the scaling has been important. Namely, if one or two components of the gradients of the objective functional (2.11) taken at the iterative solutions are ten or hundred times larger in magnitude than the others, then the iterative approximate solution is changed w.r.t. the components of large magnitude until the large components of the gradients are reduced to almost zero. Practically, however, they are reduced to the numerical error of the gradient computation. If these errors are still larger in magnitude than the remaining small gradient components, then the local optimization relies on false gradients and convergence breaks down.

4 Numerical Reconstruction of Periodic Chrome on Glass Masks

4.1 General Class of Gratings in the Numerical Experiments

As an example (cf. the mask CoG1 in [35]) we consider gratings with a cross-section shown in Figure 2 illuminated by plane waves with the fixed wave length $\lambda = 632.8$ nm. The period can simply be determined from the measured directions of the diffracted modes and the formulae (2.2) and (2.3). Therefore, we may suppose that the period d is known. The material properties are fixed by the refractive index which is 1.4571 for the silicon oxide in the substrate, $3.7329 + i\,3.8113$ for the chrome in the trapezoidal bridge, and $3.1185 + i\,0.3802$ for the oxide layer covering the bridge. The geometry of the grating structure is determined by the eight parameters p_i defined in Figure 2. Since we wish to recover the geometry data from the diffraction efficiencies and since these efficiencies are independent of a shift of the grating structure into the direction of the x-axis, we always fix at least one of the parameters p_7 and p_8. In order to avoid extra constraints like $0 \le p_3 < p_2 \le d$ in the optimization algorithm, we switch to relative x-coordinates and choose the internal parameters of optimization by

$$h_i := p_i,\ i = 1,4, \qquad h_i := \frac{p_i}{d},\ i = 2,5,7, \qquad h_i := \frac{p_i}{p_{i-1}},\ i = 3,6,8.$$

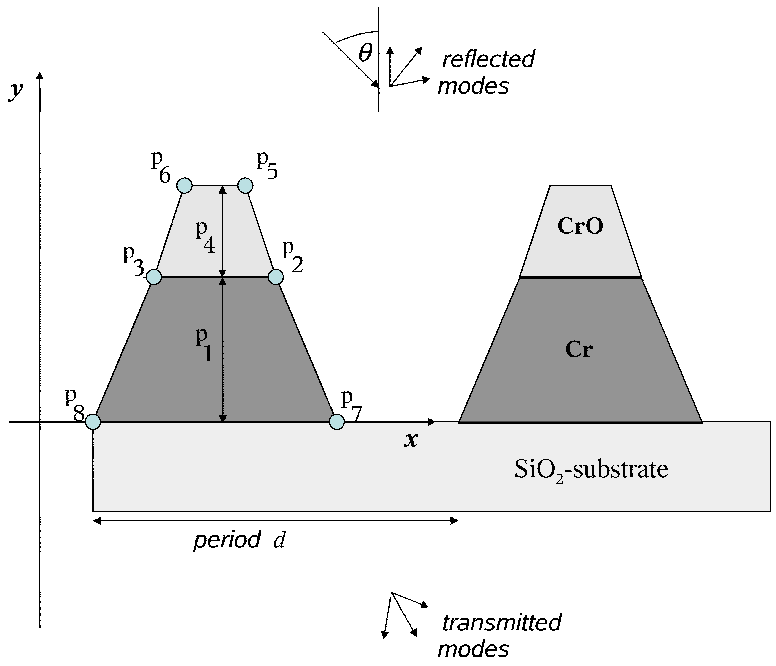


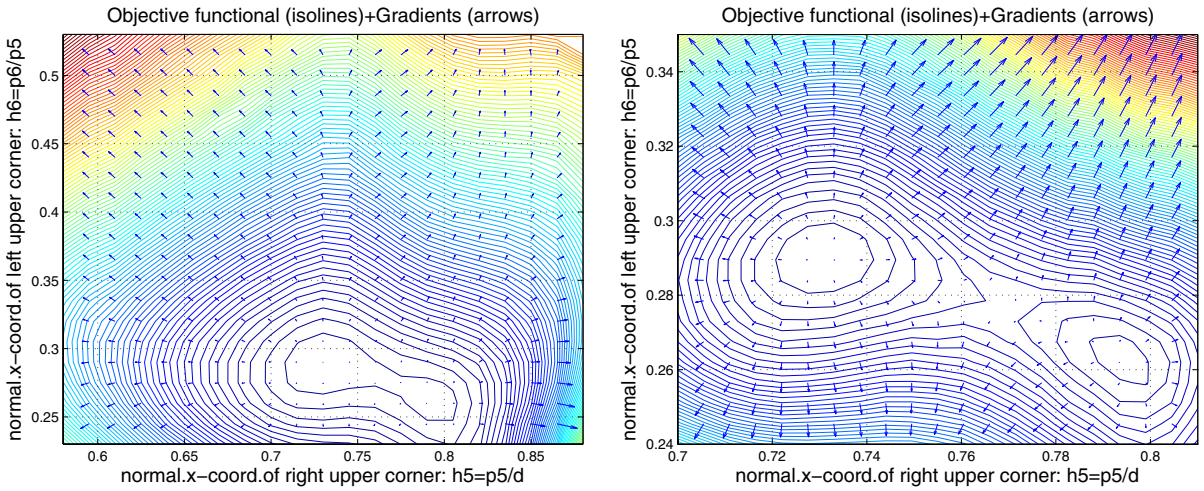
Figure 2: Scheme of the periodic grating structure chrome on glass (CoG) and its geometrical parameters: p_1 - height of Cr layer; p_2, p_3 - x coordinates of the upper right and left corners of the Cr layer, p_4 - height of CrO layer; p_5, p_6 - x coordinates of the upper right and left corners of the CrO layer; p_7, p_8 - x coordinates of the the lower right and left corners of the bridge.

The aim of all calculations is to recover the grating of Figure 2 determined by a bridge height of $p_1 = 50$ nm, a layer width of $p_4 = 23$ nm, a top CD of $p_5 - p_6 = 580$ nm, and side-wall angles of 73° on all lateral sides of the chrome and the oxide layer trapezoid. All measurement data in the present paper are simulated using the FEM package DIPOG (cf. [9]). However, using the same numerical method for simulation and reconstruction often yields too optimistic reconstruction results. Consequently, we performed the simulation over different types of FEM-grids and on much higher levels of discretization.

4.2 Local Minimum for the Two Parameter Case

In our first test we fix all parameters except p_5 and p_6. The measurement data are the three phase shift differences $\arg[A_1^-/|A_1^-|] - \arg[B_1^-/|B_1^-|]$, $\arg[A_0^+/|A_0^+|] - \arg[B_0^+/|B_0^+|]$, and $\arg[A_0^-/|A_0^-|] - \arg[B_0^-/|B_0^-|]$, corresponding to plane wave illumination under an angle θ of $-20°$, $40°$, and $40°$, respectively. Note that this set of measurement data is the optimal subset (cf. Subsection 3.4) of all phase shift differences for all angles $\theta = -80°, -70°, \ldots, 80°$. Plots of the objective function f from (2.11) are presented in Figure 3 for the arguments in $[0.575, 0.875] \times [0.225, 0.525]$ and in $[0.70, 0.81] \times [0.24, 0.35]$, respectively. Besides the global minimum for $h_5 = h_5^{ex} = 0.73007$ and $h_6 = h_6^{ex} = 0.29068$, there appears an additional local minimum for $h_5 = 0.793$ and $h_6 = 0.265$ with a value of the objective functional quite close to that of the global optimum. If the initial values of the optimization algorithm are chosen on the side of the additional local minimum, then the local algorithm converges to this. The global minimum h^{ex} is not reconstructed. In general, we have observed that an increase of M_0, i.e. of data in the measurement set, is helpful to overcome this problem. Though, especially for large number of parameters N, the existence of additional minimal solutions cannot be excluded, their distance to the global solution is larger and their attained functional value is larger than for the case of small M_0.

In our further tests we restrict the measurement data to sets of efficiency values. Indeed, our numerical experience has revealed that excluding phase shift differences from the optimal measurement data can



(a) function f over $[0.575,0.875]\times[0.225,0.525]$ (b) zoom in over $[0.70,0.81]\times[0.24,0.35]$

Figure 3: Objective functional depending on two parameters indicated by isolines. The functional is defined by an optimal choice of measurement set consisting of three phase shift differences. The arrows indicate approximate gradients computed by the FEM. Two local minima appear.

double the optimal condition numbers in (3.8). However, performance and reconstruction properties of the optimization algorithms for optimal measurement data with and without phase shift differences are almost comparable.

4.3 Reconstruction of Three Parameters

For the second test, we reconstruct three geometry parameters. The other parameters are fixed. A good reconstruction is possible for the parameters p_1, p_4, and p_7. However, the class of corresponding gratings consists of convex bridges, only. Since the results for the reconstruction of p_1, p_4, and p_7 are similar to those for the non-convex case of p_3, p_6, and p_8, we present the results for the last case, only. We consider the complete data set $(E_m)_{m\in\mathcal{M}}$ of all reflected and transmitted efficiencies (i.e. all indices $n \in \mathcal{U}^{\pm} = \mathcal{U}^{\pm}(\theta)$ with plus and minus sign in upper index of (2.4)) corresponding to the incident angles $\theta = -80°, -70°, \ldots, 80°$ and to both classical states of polarization (TE and TM). From these $M = 300$ measurement values the optimal set of three, four, and five values can be determined in accordance with (3.8) by computing the condition numbers for all possible subsets of three, four, and five values. The best choices are given in Table 1. We do not know of any different way to construct measurement data, and the results of Table 1 seem not to suggest any simple algorithm.

For the reconstruction, we have introduced the upper and lower bounds $0 \le h_k \le 0.4$, $k = 3, 6, 8$ to prescribe the box constraints (2.9). At first we consider the optimal three measured efficiencies. The corresponding global minimum of f from (2.11) is attained for $h_3^{op} = 0.2797$, $h_6^{op} = 0.2907$, and $h_8^{op} = 0.2564$, i.e. the optimal solution h^{op} coincides with the exact solution h^{ex} to be reconstructed. Two further local minima can be observed for $h_3 = 0.2344$, $h_6 = 0.3474$, and $h_8 = 0.1540$ as well as for $h_3 = 0.2623$, $h_6 = 0.2670$, and $h_8 = 0.2037$. The local optimization algorithm (2.15) starting from different initial solutions converges to either of these three minima. If eight different initial solutions with values for $h_k^0 \in \{0.1, 0.3\}$, $k = 3, 6, 8$ are chosen, then algorithm (2.15) converges twice to the optimal solution and six times to a local minimum. To improve the reconstruction more measurement data should be used. Thus, at second we consider the optimal five measured efficiencies. In this case, besides the optimal h^{op} there appears an additional local minimum for f from (2.11) at the boundary of the admissibility domain, namely $h_3 = 0.2381$, $h_6 = 0.4$, and $h_8 = 0.3003$. Starting the optimization algorithm (2.15) from the eight initial solutions with $h_k^0 \in \{0.1, 0.3\}$, the optimal solution is recovered

number M_0 of measurement values	polarization state	angle of incidence θ	reflected or transmitted mode	order n of diffracted mode
3	TE	20°	transmitted	-1
	TM	20°	reflected	-1
	TM	40°	reflected	-1
4	TE	−60°	transmitted	-1
	TM	−10°	reflected	-2
	TE	20°	transmitted	-1
	TE	30°	transmitted	-1
5	TE	−60°	transmitted	-1
	TM	−10°	reflected	-2
	TE	20°	transmitted	-1
	TE	30°	transmitted	-1
	TE	80°	reflected	-1

Table 1: Optimal measurement data to reconstruct the parameters p_3, p_6, and p_8, i.e. corresponding local mapping has smallest condition numbers (cf. (3.8)). Optimality is defined via the Jacobian taken at parameter set h^{ex}.

level of discretization	h_3	h_6	h_8
3	0.28663	0.29082	0.24794
4	0.28276	0.29047	0.25248
5	0.28092	0.29075	0.25491
6	0.28010	0.29064	0.25580
7	0.27983	0.29068	0.25618
exact values	0.27967	0.29068	0.25638

Table 2: Reconstruction of the parameters h_3, h_6, and h_8 from almost optimal ten measurement values by the optimization algorithm (2.15) computed over several discretization levels. Initial solution is $h_3^0 = 0.2$, $h_6^0 = 0.4$, and $h_8^0 = 0.2$.

already six times. Finally, we consider an almost optimal measurement set of ten measured efficiencies produced by the algorithm of Subsection 3.5 setting $M_0 = 10$ and $M_C = 40$. Using this and taking initial solutions with $h_k^0 \in \{0.1, 0.3\}$ or even $h_k^0 \in \{0., 0.4\}$, the algorithm (2.15) converges always to the global optimum, i.e. the solution is recovered. The convergence of the recovered solution depending on the discretization level is recorded in Table 2. Note that a discretization level increased by one means that the meshsize of the underlying FEM grid is halved. Seems as if the solution can be computed upto any prescribed accuracy. Of course, in applications measurement uncertainties and a restricted validity of the mathematical model (i.e. the periodicity assumption and the idealized trapezoidal structure) limit the accuracy of reconstruction.

In our last numerical test with three parameters, we have taken the same angles of incidence $\theta \in \{-40, -30, 0, 20, 30, 40, 60, 70\}$ as for the almost optimal choice of ten measurement values. This time, however, we have included the efficiencies of all 136 propagating TE and TM modes into the measurement set. Taking the eight initial solutions with $h_k^0 \in \{0, 0.4\}$, $k = 3, 6, 8$, the Newton method converges always to a good approximation of the exact parameter set. However, in comparison to the almost optimal set of ten measurement values, the reconstruction errors are slightly larger. For instance, the approximate values at level seven are $h_3 = 0.27984$, $h_6 = 0.29037$, and $h_8 = 0.25617$ (cf. Table 2).

As mentioned at the end of Subsection 3.3 the optimality of the measurement set is of local nature. To underline this fact we list the condition numbers $\mathrm{cond}([A_{\mathcal{M}_0\mathcal{N}}]^T A_{\mathcal{M}_0\mathcal{N}})$ computed at h^{ex} and at small perturbations of this parameter set in Table 3. A small perturbation of the parameter set can spoil the conditioning essentially. Fortunately, the deterioration is not so drastic for larger numbers M_0 of

number M_0 of measurement values	$h_3 = h_3^{ex}$ $h_6 = h_6^{ex}$ $h_8 = h_8^{ex}$	$h_3 = 1.05 \cdot h_3^{ex}$ $h_6 = h_6^{ex}$ $h_8 = h_8^{ex}$	$h_3 = h_3^{ex}$ $h_6 = 0.95 \cdot h_6^{ex}$ $h_8 = h_8^{ex}$	$h_3 = h_3^{ex}$ $h_6 = h_6^{ex}$ $h_8 = 1.05 \cdot h_8^{ex}$	$h_3 = 1.05 \cdot h_3^{ex}$ $h_6 = 0.95 \cdot h_6^{ex}$ $h_8 = 1.05 \cdot h_8^{ex}$
3	1.50	495.75	11.39	6.31	8.97
4	1.11	6.72	3.16	3.60	3.44
5	1.06	6.29	3.11	3.45	3.49

Table 3: Condition numbers for Jacobians of the mapping geometry parameters to measurement data. Varying parameters are h_3, h_6, and h_8. Measurement sets (cf. Table 1) are optimal at parameter set h^{ex}. Condition numbers are computed at h^{ex} and at slightly perturbed parameter sets.

measurement values. To obtain measurement sets which are almost optimal in a larger neighborhood of the exact parameter set, we recommend to optimize the condition numbers of the Jacobian simultaneously at h^{ex} and at small perturbations of this parameter set (cf. the end of Subsection 3.4). On the other hand, the optimized condition numbers are sufficiently small such that a further reduction in the sense of Subsection 3.6 is not needed.

4.4 Reconstruction of Seven Parameters

Now we reconstruct the seven parameters $h_k,\ k = 1, \ldots, 6, 8$. Clearly, the more parameters are to be reconstructed the closer is the parameter reconstruction problem to a discretization of the ill-posed inverse problem of reconstructing the grating geometry without any a-priori knowledge. Thus the parameter determination becomes more difficult. Larger condition numbers and more local minima will complicate our reconstruction. Similarly, to the three parameter case we can compute almost optimal measurement data using the setting of Subsection 3.4 and the algorithm of Subsection 3.5 this time for M_0 measurement values with $M_0 = 7, 10, 18$. In accordance with the comment at the end of Subsection 3.4, we replace $\mathrm{cond}([A_{\mathcal{M}_0\mathcal{N}}]^T A_{\mathcal{M}_0\mathcal{N}})$ for the Jacobian at h^{ex} with the maximum of $\mathrm{cond}([A_{\mathcal{M}_0\mathcal{N}}]^T A_{\mathcal{M}_0\mathcal{N}})$ taken over the Jacobians computed at ten parameter sets

$$
\begin{aligned}
&h_k = h_k^{ex},\ k = 1,2,3,4,5,6,8,\\
&h_k = h_k^{ex},\ k = 2,3,4,5,6,8, \quad h_1 = 1.05 \cdot h_1^{ex},\\
&h_k = h_k^{ex},\ k = 1,3,4,5,6,8, \quad h_2 = 1.05 \cdot h_2^{ex},\\
&h_k = h_k^{ex},\ k = 1,2,4,5,6,8, \quad h_3 = 1.05 \cdot h_3^{ex},\\
&h_k = h_k^{ex},\ k = 1,2,3,5,6,8, \quad h_4 = 0.95 \cdot h_4^{ex},\\
&h_k = h_k^{ex},\ k = 1,2,3,4,6,8, \quad h_5 = 1.05 \cdot h_5^{ex},\\
&h_k = h_k^{ex},\ k = 1,2,3,4,5,8, \quad h_6 = 1.05 \cdot h_6^{ex},\\
&h_k = h_k^{ex},\ k = 1,2,3,4,5,6, \quad h_8 = 1.05 \cdot h_8^{ex},\\
&h_k = h_k^{ex},\ k = 2,5,6, \quad h_k = 1.05 \cdot h_k^{ex},\ k = 1,3,4,8,\\
&h_k = 1.05 \cdot h_k^{ex},\ k = 1,2,3,4,5,6,8.
\end{aligned}
$$

For instance, applying the algorithm of Subsection 3.5 with $M_0 = 18$ and $M_C = 27$ and with the above mentioned ten Jacobians, we get the optimal choice of eighteen measurement values given in Table 4. The condition numbers of the corresponding Jacobians at h^{ex} and some of the perturbed measurement sets are listed in Table 5 for different numbers M_0 of measurement values. As expected the condition numbers are larger than for the three parameter case in Table 3 but they decrease with increasing M_0. Moreover, similarly to Subsection 4.3, seven measurement values are not sufficient for a reconstruction of the parameters if their initial values are not close enough to the expected values.

Even for larger values of M_0 the advantage of the optimal choice compared to a naive or guideless choice

polarization state	angle of incidence θ	reflected or transmitted mode	order n of diffracted mode	efficiencies in per cent at exact set of parameters
TE	$-80°$	reflected	1	0.769504
TM	$-80°$	reflected	2	0.550241
TE	$-30°$	transmitted	-1	8.438866
TE	$-20°$	reflected	2	0.286397
TE	$30°$	reflected	-2	0.396901
TE	$30°$	transmitted	1	8.438866
TE	$40°$	transmitted	-3	0.978679
TE	$50°$	reflected	-2	0.303465
TE	$50°$	transmitted	-3	1.133749
TE	$60°$	reflected	-3	0.331527
TE	$60°$	transmitted	-3	1.038697
TE	$60°$	transmitted	-2	0.709608
TM	$60°$	reflected	-2	0.502725
TM	$60°$	transmitted	-2	0.509878
TE	$70°$	reflected	-3	0.287705
TE	$70°$	reflected	-2	0.224387
TE	$70°$	transmitted	-3	0.777185
TE	$80°$	transmitted	-3	0.428565

Table 4: Optimal measurement data consisting of eighteen efficiency values to reconstruct the parameters p_1, p_2, ..., p_6, and p_8. Optimality is defined via the Jacobians taken at parameter set h^{ex} and nine perturbations. Additional we list the corresponding efficiencies.

of measurements can be demonstrated. For example, we can take the eighteen reflected efficiency values of polarization state TE at the orders $0, 1, 2$ resp. -2, -1, 0 and at the incident angles $\theta = -80°$, $-50°$, $-20°$ resp. $10°$, $40°$, $70°$ for a naive choice. Starting from the same initial values $h_k^0 = 0.8 \cdot h_k^{ex}$, $k \neq 4$ and $h_4^0 = 1.95 \cdot h_4^{ex}$, quite different reconstruction results emerge for the optimal and the naive subset, respectively. This is illustrated in Figure 4. Obviously the naive choice of a subset does not yield an acceptable reconstruction (see lower profile in Figure 4). The corresponding optimal value of the objective functional f (cf. Equation (2.11)) is 2.65 which is relatively large compared to the $1.6 \cdot 10^{-5}$ in the case of the optimal subset. In particular, the heights and the side wall angles of the Cr0 and the Cr layers are extremely far from the expected values for the naive choice. In order to find an explanation for the different behaviour, we compare the shapes of the corresponding objective functionals f. Figure 5 shows the shapes of f if the two parameters h_1 and h_4, i.e. the heights of the Cr and the CrO layers, vary and if the remaining five parameters are fixed to the corresponding expected values. Surely, mountain areas like that in the middle of the graph on the right-hand side will prevent the iterative solution to move from the side with only local minima to the side with the global one.

Furthermore, for the optimal set of eighteen parameters (cf. Table 4) the optimization algorithm (2.15) converges to the exact solution for all initial vectors h^0 chosen such that $h_k^0 \in \{0.8 \cdot h_k^{ex}, 1.2 \cdot h_k^{ex}\}$. The convergence for the initial value $h_k^0 = 0.8 \cdot h_k^{ex}$ is presented in Table 6. Figure 6 shows the details of the convergence for all $2^7 = 128$ initial vectors. The deviations of the optimized parameters from their expected values are presented in this figure. The optimizations were excuted at discretization level 5 and were started with the different initial solutions. The parameter specific differences in the deviations are clearly displayed. Only the deviations of the parameters $h_1 = p_1$ and $h_4 = p_6$ (cf. Figure 2) are almost 0 at this discretization level. However, the lower and upper quartiles[2] are very close to the medians[3] (shown as boxes around the medians in Figure 6(b)) for all seven parameters confirming a good convergence

[2]Recall that the lower quartile of a finite sequence of numbers is the threshold such that one fourth of the numbers is less and three fourth greater than the treshold. Similarly, the upper quartile is the threshold such that one fourth of the numbers is greater and three fourth less than the treshold.

[3]Recall that the median of a finite sequence of numbers is the number in the middle if the sequence is sorted with respect to magnitude.

number M_0 of measurement values	$h_1 = h_1^{ex}$ $h_2 = h_2^{ex}$ $h_3 = h_3^{ex}$ $h_4 = h_4^{ex}$ $h_5 = h_5^{ex}$ $h_6 = h_6^{ex}$ $h_8 = h_8^{ex}$	$h_1 = 1.05 \cdot h_1^{ex}$ $h_2 = h_2^{ex}$ $h_3 = h_3^{ex}$ $h_4 = h_4^{ex}$ $h_5 = h_5^{ex}$ $h_6 = h_6^{ex}$ $h_8 = h_8^{ex}$	$h_1 = h_1^{ex}$ $h_2 = h_2^{ex}$ $h_3 = h_3^{ex}$ $h_4 = 0.95 \cdot h_4^{ex}$ $h_5 = h_5^{ex}$ $h_6 = h_6^{ex}$ $h_8 = h_8^{ex}$	$h_1 = h_1^{ex}$ $h_2 = h_2^{ex}$ $h_3 = 1.05 \cdot h_3^{ex}$ $h_4 = h_4^{ex}$ $h_5 = h_5^{ex}$ $h_6 = h_6^{ex}$ $h_8 = h_8^{ex}$	$h_1 = 1.05 \cdot h_1^{ex}$ $h_2 = h_2^{ex}$ $h_3 = 1.05 \cdot h_3^{ex}$ $h_4 = 1.05 \cdot h_4^{ex}$ $h_5 = h_5^{ex}$ $h_6 = h_6^{ex}$ $h_8 = 1.05 \cdot h_8^{ex}$
7	239.52	255.52	239.23	264.36	666.03
10	142.31	194.80	137.86	236.47	169.41
18	89.12	124.02	102.74	153.00	158.60

Table 5: Condition numbers for Jacobians of the mapping geometry parameters to measurement data. Varying parameters are $h_k,\ k = 1, 2, 3, 4, 5, 6, 8$. Measurement sets (cf. Table 4 for $M_0 = 18$) are almost optimal. Condition numbers are computed at h^{ex} and at slightly perturbed parameter sets.

level of discretization	h_1	h_2	h_3	h_4	h_5	h_6	h_8
3	0.04719	0.74705	0.26449	0.02184	0.73551	0.32085	0.27829
4	0.04939	0.73908	0.27645	0.02276	0.73134	0.29526	0.26069
5	0.04988	0.73720	0.27897	0.02295	0.73035	0.29177	0.25755
6	0.04997	0.73665	0.27954	0.02299	0.73016	0.29099	0.25674
exact values	0.05000	0.73635	0.27967	0.02300	0.73007	0.29068	0.25638

Table 6: Reconstruction of the parameters $h_k,\ k = 1, 2, 3, 4, 5, 6, 8$ from almost optimal eighteen measurement values by the optimization algorithm (2.15) computed over several discretization levels. Initial solution is $h_k^0 = 0.8 \cdot h_k^{ex},\ k = 1, 2, 3, 4, 5, 6, 8$.

behaviour for all initial vectors with $h_k^0 \in \{0.8 \cdot h_k^{ex}, 1.2 \cdot h_k^{ex}\}$.

Typically, the accuracy of the reconstructed parameters is not uniform if the parameters are of different nature. In our example the heights $h_1 = p_1$ and $h_4 = p_4$ seem to have a much stronger impact on the efficiencies than the lateral dimensions. Consequently, the accuracy of the reconstruction should be higher for the heights than for the remaining parameters. Note, however, that the scaling of the parameters ($r_n = 1$ with r_n defined in Subsection 3.6) is chosen such that a change in each of the parameters by say 0.01 corresponds, roughly speaking, to a change of a geometric dimension like height, diameter or shift in position by 0.01 μm. In other words, a uniform accuracy for all parameters would have been desirable.

4.5 Noisy Data for the Measurement

In order to study the influence of measurement uncertainties on the accuracy of the reconstruction we repeat the test calculations of the previous Subsection with perturbed sets of efficiencies. The optimal set of eighteen efficiency values from the last column of Table 4 is superimposed by a normally distributed noise signal. The signal to noise ratio of the scatterometric setups available at the PTB laboratories [35] is about 10^4 and, in first approximation, independant of the intensity of the incident light. Therefore, in a first test, the efficiencies of the optimal set of Table 4 is superimposed repeatedly with a relative noise level of $1.1 \cdot 10^{-4}$ resulting in 128 perturbed sets of eighteen efficiencies. Then these perturbed sets of simulated efficiencies are used as measurement data in convergence tests for the 128 initial vectors with $h_k^0 \in \{0.8 \cdot h_k^{ex}, 1.2 \cdot h_k^{ex}\}$. For a second test we halve the noise level. The results for the reconstruction of the seven unknown parameters are presented in Figure 7 as parameter specific plots of the medians and their lower and upper quartiles. The maximal deviation over all 128 results of the optimization algorithm is 0.01049 for the higher noise level and 0.0043 for the lower. All together these results meet our expectations and yield a first estimate for the uncertainties of the parameters to be reconstructed.

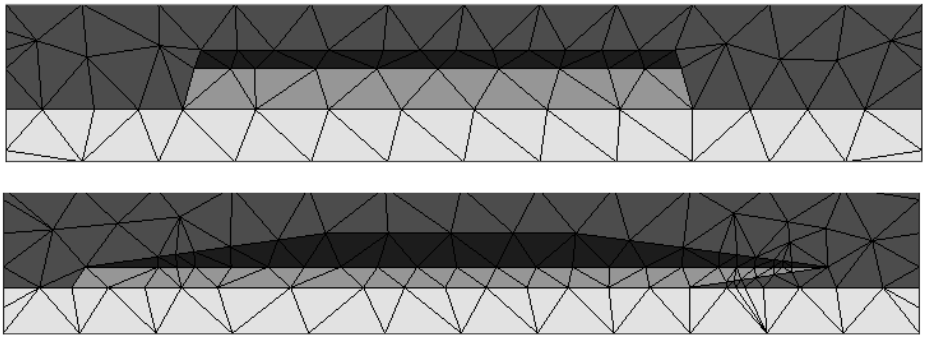


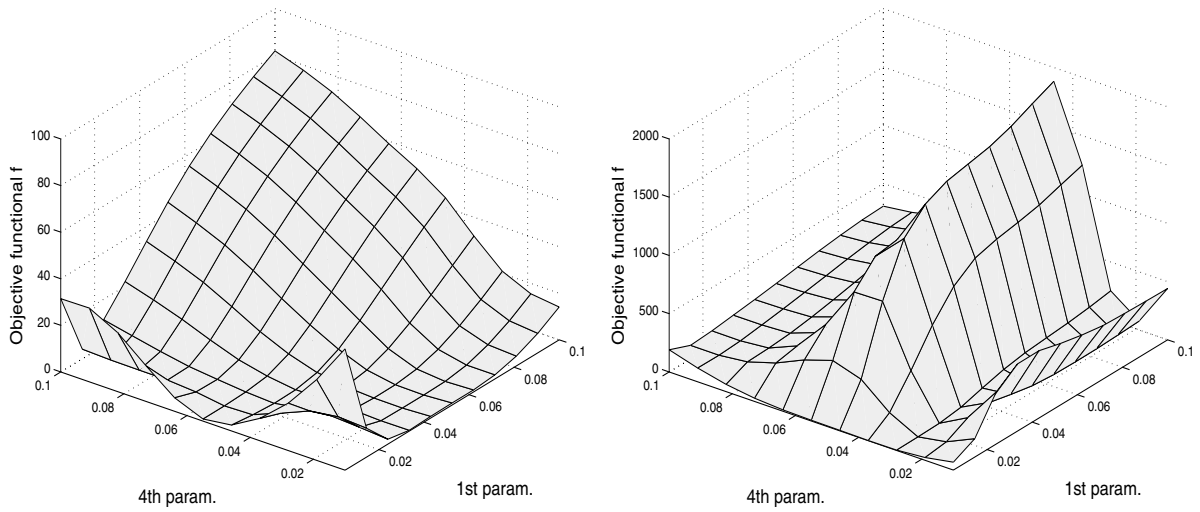
Figure 4: Optimization results: Upper profile - good reconstruction for the optimal subset of Table 4; lower profile - unacceptable reconstruction for an intuitively chosen subset (see text for details).

5 Conclusions and Further Research

In scatterometry the variability of the efficiency patterns depends on the parameters of the inspecting light and on the geometry of the scattering probe. To solve the reconstruction problem, i.e. to determine the geometrical parameters of the inspected grating from the measured efficiency pattern, we have formulated the inverse problem as a non-linear operator equation in Euclidean spaces. The operator maps the vector of sought parameters to a vector of certain efficiencies of the plane wave modes diffracted by the grating. A FEM based Newton type method has been described to solve the corresponding operator equation. The convergence of the numerical algorithm and the quality of the reconstruction results is strongly affected by the local conditioning of the non-linear mapping, i.e. by the condition number of the Jacobian matrix for parameter sets close to the solution vector. As a consequence, for a fixed positive integer M_0, we have presented an algorithm to determine almost optimal sets of M_0 efficiencies for the measurements by optimizing the condition numbers of the corresponding Jacobians. These Jacobians are taken at the expected set of parameters h^{ex} and, in order to improve the stability of the almost optimal set of measurements, at some small perturbations.

As numerical examples we have considered a periodic chrome-glass grating of periodically allocated Cr bridges with a height of 50 nm coated by a CrO layer with a width of 23 nm. The period has been fixed to 1 120 nm. Reflected and transmitted efficiency measurements have been simulated under inspecting light of a wavelength of 632.8 nm and for the polarization types TE and TM. The angle of the incident wave θ has been varied in the set $\{-80°, -70°, \ldots, 80°\}$. From this set of all simulated efficiency measurements, we have determined almost optimal subsets of measurements with different sizes M_0. We have presented the convergence properties and the reconstruction quality of the proposed algorithm applying them to the almost optimal maesurement sets. Additionally, we have investigated the bias of different initial solutions h_k^0 and perturbed efficiencies values on the reconstruction results. We have observed the following:

- Good reconstruction results are possible using mesurement data consisting only of efficiency measurements but not of phase shift differences.

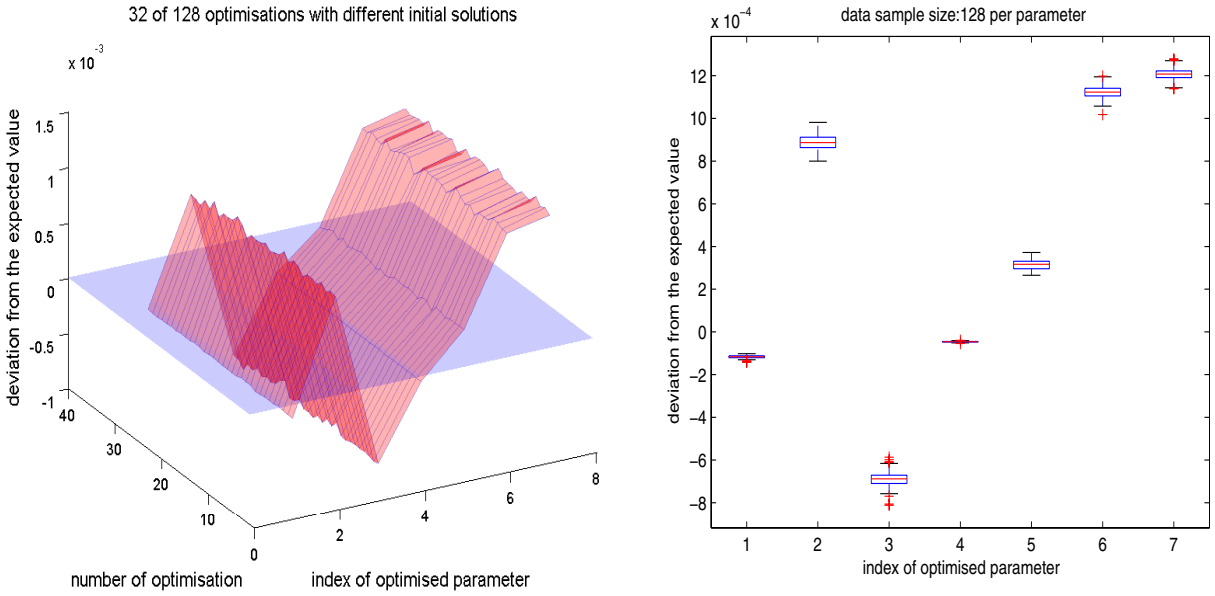


(a) (b)

Figure 5: Objective functional depending on h_1 and h_4 (a) for the optimal subset of Table 4; (b) for an intuitively chosen subset (see text for details).

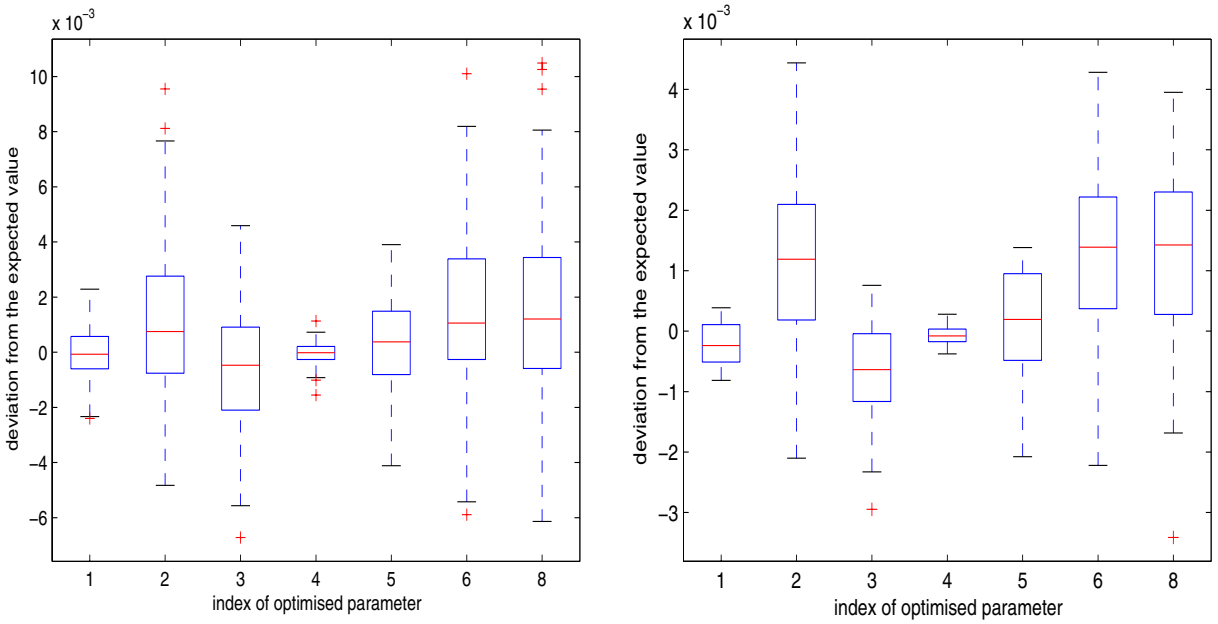
- The measurement data for the reconstruction must be chosen carefully. For some sets of measurement reconstruction is not possible. We recommend to determine the measurement data optimizing the local conditioning of the mapping parameters to efficiencies.
- The number of measurement values M_0 should be chosen sufficiently large in order to avoid false solutions of the local optimization algorithms.
- A fast convergence of the algorithm and an accurate reconstruction is possible even for seven parameters. According to our numerical tests, the reconstruction is independent of the initial solution at least if the last deviates from the exact solution by no more than 20%.
- The reconstruction algorithms yield parameter specific differences in the deviations from the exact values of the parameters of reconstruction.
- The algorithm converges even if the optimal subset of measurement data is superimposed with noise of different levels.

Clearly, reconstructing a large number of parameters is like solving the complete ill-posed inverse problem. So it remains an interesting question, how many parameters can be reconstructed and whether it is possible to adopt regularization procedures for ill-posed problems. We have restricted our considerations so far to chrome on glass masks inspected under the wave length of 632.8 nm. Our next task will be to apply the presented techniques to EUV masks inspected under 13.6 nm (cf. [35]). Finally, our tests with simulated data have demonstrated that the grating structures can be determined from the mathematical point of view. To see what happens in real life, we still have to apply our techniques to measured data. We are planning to do this in the next future both for the 632.8 nm case treated above and for the EUV case.



(a) (b)

Figure 6: Deviations from the expected parameter values for different initial vectors (a) as a mesh plot of the first 32 optimizations and (b) as a plot of the medians including the lower and upper quartiles as boxes around the medians. The whiskers extending from each end of the boxes indicate the extent of the rest of the deviation data.



(a) (b)

Figure 7: Deviations from the expected parameter values for perturbed efficiencies and for different initial vectors (a) for a noise level of $1.1 \cdot 10^{-4}$ and (b) for a noise level of $0.55 \cdot 10^{-4}$.

References

[1] R.M. Alassaad and D.M. Byrne, Error Analysis in Inverse Scatterometry I: Modeling, *Optical Engineering*, to appear.

[2] T. Baeck, *Evolutionary algorithms in theory and practice*, Oxford University Press, 1996.

[3] G. Bao, Finite element approximation of time harmonic waves in periodic structures, *SIAM J. Numer. Anal.*, **32** (1995), pp. 1155–1169.

[4] G. Bao and D.C. Dobson, Modeling and optimal design of diffractive optical structures, *Surveys on Mathematics for Industry*, **8** (1998), pp. 37–62.

[5] O. Cessenat and B. Depres. Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem, *SIAM J. Numer. Anal.*, **35** (1998), pp. 255-299.

[6] J. Chandezon, D. Maystre, and G. Raoult, A new theoretical method for diffraction gratings and its applications, *J. Opt. (Paris)*, **11** (1980), pp. 235–241.

[7] X. Chen and A. Friedmann, Maxwell's equation in a periodic structure, *Trans. Amer. Math Soc.*, **323** (1991), pp. 811–818.

[8] E.M. Drège, R.M. Al-Assaad, and D.M. Byrne, Mathematical analysis of inverse scatterometry, *Proc. of SPIE*, **4689** (2002), pp. 151–162.

[9] J. Elschner, R. Hinder, A. Rathsfeld, and G. Schmidt, *DIPOG Homepage*, http://www.wias-berlin.de/software/DIPOG

[10] J. Elschner, R. Hinder, and G. Schmidt, Finite element solution of conical diffraction problems, *Advances in Comp. Math.*, **16** (2002), pp. 139–156.

[11] J. Elschner and G. Schmidt, Conical diffraction by periodic structures: Variation of interfaces and gradient formulas, *Math. Nachr.*, **252** (2003), pp. 24–42.

[12] J. Elschner and G. Schmidt, Numerical solution of optimal design problems for binary gratings, *J. Comput. Physics*, **146** (1998), pp. 603–626.

[13] H. Gross, R. Model, M. Bär, M. Wurm, and B. Bodermann, Mathematical modelling of indirect measurements in scatterometry, *Measurememt*, to appear.

[14] Ch. Großmann and J. Terno, *Numerik der Optimierung*, Teubner Studienbücher der Mathematik, Teubner Stuttgart, 1997.

[15] H.T. Huang and F.L. Terry Jr., Erratum to "Spectroscopic ellipsometry and reflectometry from gratings (Scatterometry) for critical dimension measurement and in situ, real-time process monitoring" [Thin Solid Films 455-456 (2004) 828-836], *Thin Solid Films*, **468** (2004), pp. 1155–1169.

[16] F. Ihlenburg, *Finite element analysis of acoustic scattering*, Springer Verlag New-York, Berlin Heidelberg, Applied Mathematical Sciences **132**, 1998.

[17] F. Jarre and J. Stoer, *Optimierung*, Springer Verlag New York Berlin Heidelberg, 2004.

[18] B.H. Kleemann, *Elektromagnetische Analyse von Oberflächengittern von IR bis XUV mittels einer parametrisierten Randintegralmethode: Theorie, Vergleich und Anwendungen.* Dissertation, TU Ilmenau, 2002, Mensch und Buch Verlag Berlin, 2003.

[19] P.J.M. van Laarhoven and E.H.L. Aarts, *Simulated annealing: Theory and Applications*, D. Reidel Publishing Company, Mathematics and Applications, Member of the Kluwer Academic Publishing Group, Dodrecht Boston Lancaster Tokyo, 1988.

[20] P. Lalanne and G.M. Morris, Highly improved convergence of the coupled-wave method for TM-polarization, *JOSA A*, **13** (1996), pp. 779–784.

[21] L. Li, New formulation of the Fourier modal method for crossed surface-relief gratings, **JOSA A**, **14** (1997), pp. 2758–2767.

[22] J.M. Melenk and I. Babuška, The partition of unity method: Basic theory and applications, *Comp. Methods Appl. Mech. Eng.*, **139** (1998), pp. 289–314.

[23] B.K. Minhas, S.A. Coulombe, S. Sohail, H. Naqvi, and J.R. McNeil, Ellipsometric scatterometry for metrology of sub-0.1μm-linewidth structures, *Appl. Optics*, **37**, No. 22 (1998), pp. 5112–5115.

[24] M.G. Moharam and T.K. Gaylord, Rigorous coupled wave analysis of planar grating diffraction, *J. Opt. Soc. Amer.*, **71** (1981), pp. 811–818.

[25] M.G. Moharam, E.B. Grann, D.A. Pommet, and T.K. Gaylord, Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: enhanced transmittance matrix approach, *JOSA A*, **12** (1995), pp. 1077–1086.

[26] J. Nocedal and S.J. Wright, *Numerical optimization*, Springer Verlag New-York Berlin Heidelberg, Springer Series in Operation Research, 2000.

[27] R. Petit, (ed.), *Electromagnetic theory of gratings*, Springer-Verlag Berlin 1980.

[28] J. Pomplun, *Rigorous FEM-simulation of Maxwell's equations for EUV-Lithography*, Konrad Zuse Institut, Berlin, Diplomarbeit, 2006.

[29] C.J. Raymond, M.R. Murnane, S.L. Prins, S. Sohail, H. Naqvi, and J.R. McNeil, Multiparameter grating metrology using optical scatterometry, *J. Vac. Sci. Technol.*, **15** (2) (1997), pp. 361–368.

[30] C.J. Raymond, M.R. Murnane, S. Sohail, H. Naqvi, and J.R. McNeil, Metrology of subwavelength photoresist gratings using optical scatterometry, *J. Vac. Sci. Technol.*, **13** (4) (1995), pp. 1484.

[31] A. Schaedle, L. Zschiedrich, S. Burger, R. Klose, and F. Schmidt, *Domain decomposition method for Maxwell's equations: Scattering of periodic structures*, ZIB-Report 06-04, Konrad Zuse Institut, Berlin, 2006.

[32] A. Tavrov, M. Totzeck, N. Kerwien, H.J. Tiziani, Rigorous coupled-wave analysis calculus of submicrometer interference pattern and resolving edge position versus signal-to-noise ratio, *Opt. Eng.*, **41**(8) (2002), pp. 1886–1892.

[33] J. Turunen and F. Wyrowski (eds.), *Diffractive Optics for Industrial and Commercial Applications*, Akademie Verlag, Berlin, 1997.

[34] H.P. Urbach, Convergence of the Galerkin method for two-dimensional electromagnetic problems, *SIAM J. Numer. Anal.*, **28** (1991), pp. 697–710.

[35] M. Wurm, B. Bodermann, and W. Mirandé, Evaluation of scatterometry tools for critical dimension metrology, *DGaO Proceedings*, **106**. Tagung (2005).