MASTER'S THESIS

Josef Navrátil

Rigorous Electromagnetic Theory of the Optical Response of Periodic Nanostructures

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague 11.5.2018   ....

Title: Rigorous Electromagnetic Theory of the Optical Response of Periodic Nanostructures

Author: Josef Navrátil

Department: Institute of Physics of Charles University

Supervisor: RNDr. Roman Antoš, Ph.D., Institute of Physics of Charles University

Abstract: The diffraction of light is an important phenomenon with wide physical and engineering applications and diffraction gratings are optical components with a periodic structure which are used to diffract light into several beams propagating in various directions. Direct methods like AFM or SEM proved to be insufficient to study the shape of planar diffraction gratings and therefore they must be supplemented with results obtained from optical spectroscopy. Computer simulations are the integral part of this method. This Thesis is focused on two particular simulation methods - the RCWA and the C-Method. It gives a rich theoretical introduction, discusses the weaknesses of these methods and also describes improvements of the RCWA using the Airy-like series and proper Fourier factorization. Both methods are implemented, tested on simple examples and afterwards the convergence for particular cases is investigated. The C-Method and the modified RCWA algorithm exhibit excellent convergence. At the end, the numeric results are compared with experiments, giving a very good agreement in the ellipsometric parameters.

Keywords: diffraction, relief gratings, RCWA, C-Method, ellisometry, spectroscopy, Fourier Factorization

Contents

Chapter 1

Introduction

1.1 A short excursion to history

The scattering of light is a physical phenomenon well-known from everyday life as the rainbow, glittering of the water surface, polished gemstones or prisms. The diffraction gratings, which are periodic structures diffracting the light are also well-known from everyday life, as DVD discs, holograms or a back of colorful spider in Fig. 1.1[1]. A bird feather is the first known studied diffraction grating; in 1673 observed J. Gregory the diffraction pattern produced by a sunlight going through a feather. The first man-made grating was done by D. Rittenhouse in 1785 from hairs strung between two screws. The appearance of diffraction orders was studied by T. Young in 1803. Another contribution was done by Fresnel in 1818; he studied the bending of light and invented the so-called Fresnel zones. The mathematical formalism of the wave theory was for the first time done by G.R. Kirchhoff.
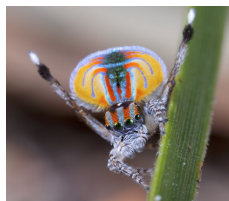

Figure 1.1: Peacock spider as a walking diffraction grating

The foundations of optical spectroscopy were laid by J. von Frauenhof in 1821. He designed the first spectroscope for the purpose of observing the star Sirius. In 1887 A. Michelson developed the interferometer, which he used to measure the speed of light and it still nowadays has its role in optical instruments. In the late 19th century, probably in 1888, P. Drude invented the experimental technique called ellipsometry. A special kind of this technique called spectroscopic ellipsometry will be described later here in detail.

The significant contribution in a theoretical study of optics and waves has been done by Lord Rayleigh between the years 1871–1899. For example, he was able to explain why the sky is blue, and the Rayleigh-Jeans law of radiation of the black body was the first step towards the quantum mechanics.

Extensive study of advanced optical properties of gratings, which cannot be explained by Fraunhofer's grating equation, was initiated by the discovery of diffraction anomalies by R. W. Wood in 1902 [1].

In 1934 C.V. Raman wrote a series of papers [2]–[4]. The first paper was concerned with the color of the Indian roller bird. He proposed, without further investigation, that the colourful feathers of this bird could be caused not by the thin-film interference, but by diffraction by small cavities. The subsequent papers are concerned with the coupled wave theory.

---

[1] Wikipedia, The Free Encyclopedia, s.v. "Maratus," (accessed March 1st, 2018)

The further development in this field was determined by the discovery of X-ray generators, lasers and CCD sensors. These devices are now standard equipment in optical laboratories.

The theoretical models of diffraction gratings were affected by the development of computers, which offered much faster data processing and short computation time.

The main application of the diffraction grating is in spectroscopy. There are three main advantages of gratings against prisms. The first one is their planar structure and it means compact size. The second one is that gratings can work in spectral regions, where no transparent glass exists with dispersion sufficient enough. The third one is a possibility of their variability in shapes, unlike to prisms whose properties are determined by the groove angle and material choice. The disadvantages are lack of diffraction efficiency and resolution, and their efficiency depends on the polarization, as will be seen in the forthcoming chapters.

However, although spectroscopy is the main area of application of diffraction gratings, it is not the only one. There are application in astronomy, lasers, optical fibers, surface-plasmons grating detectors and many others. And as mentioned in the introduction, nature uses gratings to do beautiful patterns on animals, coronas or glittering stones.

1.2 Planar gratings

The aim of this work is to study the diffraction of light on relief planar gratings with a certain shape (sinusoidal, triangular, etc.). These gratings are isotropic in one given direction and if the incident light is in the plane perpendicular to this direction, the whole problem can be considered as planar. There are various methods for modeling the optical response of planar gratings, e.g. integral and differential methods, finite elements methods, etc.

The aim of this Thesis is to study two very popular methods — the RCWA (Rigorous Coupled-Wave Analysis) and the C-Method. Both of them transform the system of partial differential equations in the Fourier space in order to get the infinite-dimensional algebraic problem which is afterwards truncated, and it gives a finite-dimensional system of linear equations.

The RCWA was developed in 1969, but it has underwent significant improvements over the past 60 years. Although it was primarily designed to model gratings with lamellar profile, later it was extended to more general profiles using the so-called staircase approximation. The former T-matrix algorithm used in this method turn out to be unstable for particular problems and therefore was replaced with the S-matrix algorithm [5]. This extension is described in Section 3.5. But still the convergence of this method was acceptable only for dielectric and shallow gratings and for gratings illuminated by X-ray. Until 1998 it was not known whether this problem has its origin in physics or whether there are some problems arising from the discretization.

The revolutionary work of G. Granet, M. Neviére, L. Li and E. Popov [6], [7], [8] revealed that the reason is the incorrect formulation of the gratings equations and developed new formulation which uses correct factorization and the derived algorithm converges much faster. A new formulation of the RCWA for curved gratings, which uses the correct factorization rules, was introduced by I. Gushchin and A. Tischenko in 2012, [9]. The detailed discussion of these problems and about the correct formulation of the equations can be found in Sections 3.3.2, 3.5. The physical explanation for a poor convergence for metallic gratings illuminated by p-polarized light is in the presence of surface plasmons. These particles cause a high intensity of light on the grating surface, which in a combination with Gibbs effect and an incorrect formulation of the equations does a significantly large error [10]. For the case of dielectic gratings or metallic gratings illuminated by s-polarized light the intensity on the grating surface is low and therefore the error is decreased, cf. Fig. 1.2, and see also [10].

The second method is the C-Method. It was developed by J. Chandezon et. al. in 1980, [11]. The main idea is in a coordinate transformation, where the curved profile of the grating is straightened leading to more complex equations, but with simple boundary conditions. Due to the risk of invoking Rayleigh hypothesis the real eigenvalues of the approximation are replaced with real propagation orders. The final solution of the problem is in matching the electric and magnetic fields on the (line) interface between the substrate and superstrate of the grating. This methods is primarily focused on shallow gratings, which
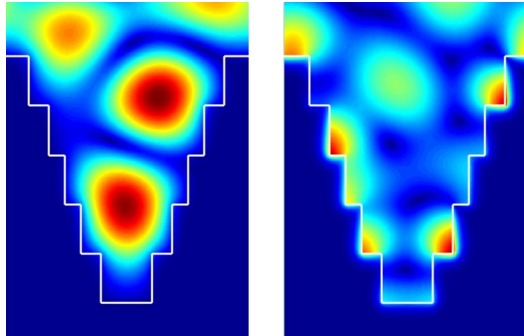
Figure 1.2: Diffraction on a staircase grating, s-polarization left, p-polarization right

can be highly conductive and coated (e.g. with a layer of some oxide). The description of this algorithm can be found in Chapter 4. It has to be noted that the C-Method does not suffer from an incorrect factorization or Gibbs phenomenon, unlike to the RCWA. On the other hands, it has a poor convergence for deep gratings and is more complicated to implement.

After careful consideration we decided to add one section devoted to the formulation and existence of weak solutions of grating equations. The weak formulation is a necessary step towards Finite Element Method, which plays an essential role in a numerical solution of Maxwell equations. Unlike to Modal Methods, for the FEM there are some rigorous results concerning the existence a uniqueness of solution and error of approximate solution. But on the other hand, the FEM are difficult to implement. FEM solvers are common part of commercial software (e.g. JCMsuite). More details can be found in Subsection 2.6.

At the end of this paragraph, let us shortly summarize the content of this Thesis. The Chapters 2 – 4 are reviewing known results about planar diffraction gratings. In particular, Chapter 2 is focused on the analytical properties of diffraction gratings, with a special emphasis to rigorous mathematical formulation of the whole problem. Chapter 3 is about the RCWA and its improvements from [9]. The process of matching the interface conditions is different to the one from [9] and uses a modification the Airy-like series algorithm introduced in [12], [13]. The subsequent Chapter is explaining the basics of the C-Method. The methods described in Chapters 3, 4 were implemented in MATLAB, the correctness of implementation was tested on trivial examples and tabular results from the respective papers. Afterwards these implementations were used to model particular real systems and results were compared with the experimental ones from [12]. All of this is summarized in Chapter 5. The last Chapter is devoted to the discussion about obtained results and to the summary remarks.

1.3 Experimental techniques, Four-zone null ellipsometry

Having a periodic surface (grating) the main question at this point is what to measure and how to measure it. There are two basic sets of parameters — geometric and material. Among geometric parameters there are thin film properties (layer thickness, interface roughness), critical dimensions (period, linewidth, element profile) and others like line-width roughness, line-edge roughness, ... Among the material properties there are optical properties like refractive index, extinction coefficient, material anisotropy, and magnetic properties.

The surface can be analyzed directly by methods like AFM, SEM, ... They measure the profile by "touching" or "feeling" the surface with a mechanical probe. The probe is moved around the surface and gathers the information about the sample. The main advantage of this technique is in providing the direct image of the surface. On the other hand, the method is expensive, it can require further modification of the sample, which will disable it for further usage, it is cumbersome — requires vacuum or a mechanical contact with the sample and finally there are some fundamental physics barriers which cannot be overcome, e.g. misalignment for the probe method.

Another possibility is a use of optical techniques, which give the information about the sample by solving the inverse problem — optical parameters are measured for the sample and its properties are found by fitting then to a computer simulation. The time consuming computer simulation and fitting process are one of the disadvantages of this method. Another disadvantage is that it cannot determine the structure without approximate knowledge of it, and also a sensitivity to too many fitting parameters. In contrast, there are no systematic errors caused by a mechanical contact, it has a higher precision (down to 1 nm), it is sensitive to ultrathin features (native oxides, interlayer effects), it has a possibility of monitor features beneath the surface, it is capable of determining line and edge roughness and there is a possibility of a unique analysis of various material phenomena like e.g. nanocomposition.

Here we are going to briefly describe the experimental techniques for obtaining the so-called el-lipsometric parameters, the algorithms for computer simulations will be extensively discussed in the forthcoming Sections.

General setting of an experimental apparatus is depicted in Fig. 1.3. The lamp generates the light
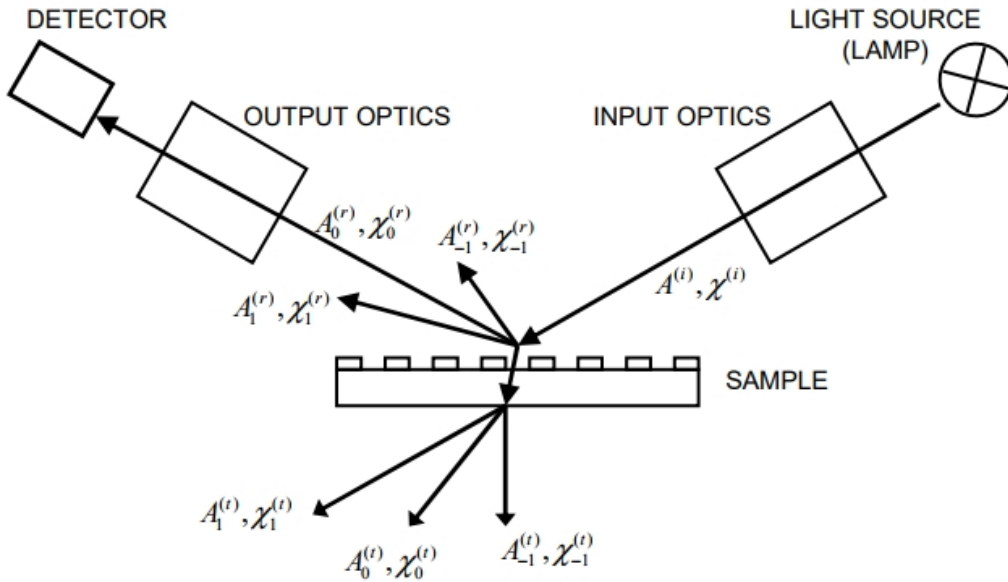

Figure 1.3: General setting of the optical experiment [12]

flux, which goes through the input optics to became a plane wave with the amplitude A_i and polarization ellipse $\chi^{(i)}$ incident onto the sample. After impinging onto the planar 1D-periodic grating the light flux is scattered into several modes — diffraction orders, which have both reflection and transmission modes. The mode of interest (usually 0th mode of reflection of transmission goes through the output optics and its intensity I_r or I_t respectively is measured by the detector. Sometimes the reflected and transmitted zeroth-order mode are measured at the same time to get additional information. The ellipsometric techniques measure the change of the polarization of the incident light, and therefore provide the ratio $\chi_n^{(r,t)}/\chi^{(i)}$ of the complex χ-numbers.

Polarization of light The harmonic wave is considered to propagate in the $z-$direction and its electric field intensity can be written as

$$\mathbf{E}(\mathbf{r},t) = \begin{pmatrix} A_x \exp(i\phi_x) \\ A_y \exp(i\phi_y) \end{pmatrix} \exp(i(\omega t - k_z z)) = \mathbf{J} \exp(i(\omega t - k_z z)).$$

The vector

$$J = \begin{pmatrix} j_x \\ j_y \end{pmatrix} = \begin{pmatrix} A_x \exp(i\phi_x) \\ A_y \exp(i\phi_y) \end{pmatrix} = \begin{pmatrix} \tilde{A}_x \\ \tilde{A}_y \exp(i\delta) \end{pmatrix},$$

where $\delta = \varphi_y - \varphi_x$, is called Jones vector and it determines the polarization state of the light. The transformation between polarization states is provided by the so-called Jones matrix. In two dimensions, it is

$$P = \begin{pmatrix} p_{xx} & p_{xy} \\ p_{yx} & p_{yy} \end{pmatrix}.$$

The polarization state of light which was initially in the state J_0 and went through several optical devices with polarization matrices $P_1, ... P_n$ (in this ordering) can be described by

$$J_n = P_n P_{n-1} \cdots P_2 P_1 J_0.$$

Spectroscopic ellipsometer The arrangement of spectroscopic ellipsometer is in Fig. 1.4.
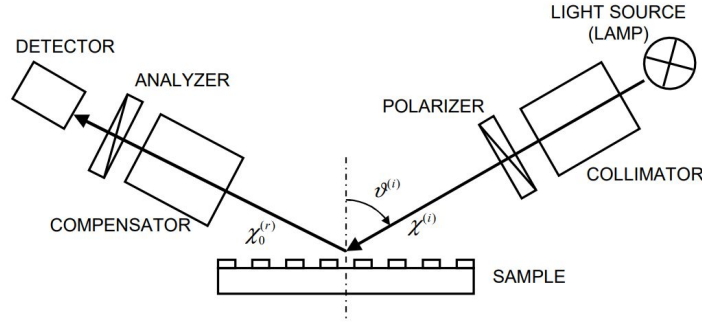

Figure 1.4: Spectroscopic ellipsometer

The light from the source (lamp) goes through the collimator, followed by the polarizer. The polarized light impinges onto the sample, the diffracted wave goes to the compensator, through the analyzer and ends in the detector. The change of polarization in each component is described by the Jones matrix. The (Jones) matrix of the linear polarizer is

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

the matrix of the compensator is

$$P = \begin{pmatrix} \exp\left(i\frac{\delta}{2}\right) & 0 \\ 0 & \exp\left(-i\frac{\delta}{2}\right) \end{pmatrix},$$

and finally the matrix of the sample in the basis of its spectral reflection is

$$P = \begin{pmatrix} r_s & 0 \\ 0 & r_p \end{pmatrix},$$

where r_s, r_p are the reflection coefficients, see also (2.27), below. Let $\alpha, \beta, \gamma, \delta$ be the rotation angles of the polarizer, compensator and analyzer respectively. The Jones vector of the beam outgoing from the analyzer and incoming into the detector has in the basis composed of analyzer eigenvectors the form

$$J_{out} = PR(\beta - \gamma)C(\delta)R(\gamma)SR(-\alpha)J_{in},$$

where

$$R(\omega) = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix},$$

denotes the matrix of rotation by an angle ω, where $J^{(i)} = [1, 0]^T$ is the Jones vector of polarized incident beam in the basis of polarizer eigenmodes. The output intensity being detected is

$$I^{(0)} = I^{(i)}(J^{(0)})^\dagger J^{(0)} = I^{(i)} L,$$

where

$$\begin{aligned} L =& r_s \cos(\overline{\alpha})[\exp(i\delta)\cos(\gamma)\cos(\overline{\beta} - \gamma) - \sin(\gamma)\sin(\overline{\beta} - \gamma)] + \\ &+ r_p \cos(\overline{\alpha})[\exp(i\delta)\cos(\gamma)\cos(\overline{\beta} - \gamma) - \sin(\gamma)\sin(\overline{\beta} - \gamma)], \end{aligned} \tag{1.1}$$

with $\overline{\alpha} = \alpha - \pi/2$, $\overline{\beta} = \beta - \pi/2$. The data analyzed in this Thesis comes from the method called Four-zone null ellipsometry, the apparatus has compensator with $\delta = \pi/2$ and $\gamma^{(\pm)} = \pm\frac{\pi}{4}$.

Four-zone null ellipsometry The null-ellipsometry is an experimental technique which gives the ellipsometric ratio $\rho = r_s/r_p$ by searching null-intensity positions of the polarizer and analyzer according to fixed positions of the compensator, i.e. angles for which $L = 0$, then

$$\rho = -\tan(\overline{\alpha})\frac{\pm 1 + i\tan(\overline{\beta} \mp \frac{\pi}{4})}{1 \mp i\tan(\overline{\beta} \mp \frac{\pi}{4})},$$

see (1.1). Rewriting it as $\rho = \tan\Psi \exp(i\Delta)$ there can be found by a use of the identity $\exp(2i\omega) = (1 + i\tan(\omega))/(1 - i\tan(\omega))$ that

$$\tan\Psi \exp(i\Delta) = \tan(\overline{\alpha})\exp(2i(\overline{\beta} - \frac{\pi}{4})),$$

$$\tan\Psi \exp(i\Delta) = -\tan(\overline{\alpha})\exp(2i(\overline{\beta} - \frac{\pi}{4})),$$

which corresponds to $\gamma^{(+)}$, $\gamma^{(-)}$ respectively. Each of these equations have two solutions — here comes the name "Four-zone null ellipsometry". To eliminate the measurement errors, the four angles are measured and the ellipsometer parameters are calculated as the averages

$$\Psi = \frac{1}{4}\left(\overline{\alpha}_1 - \overline{\alpha}_2 + \overline{\alpha}_3 - \overline{\alpha}_4\right),$$

$$\Delta = \frac{1}{2}\left(\overline{\beta}_1 - \overline{\beta}_2 + \overline{\beta}_3 - \overline{\beta}_4\right).$$

Data processing Once the simulated and measured diffraction efficiencies are known, the shape of grating profile can be determined by solving the inverse problem. Let

$$\cos(\mathcal{E}_j) = \mathcal{S}_{e,j} \cdot \mathcal{S}_{m,j},$$

where

$$S_{e,j} = \begin{pmatrix} \sin(2\psi_{e,j})\cos(\Delta_{e,j}) \\ \sin(2\psi_{e,j})\sin(\Delta_{e,j}) \\ \cos(2\psi_{e,j}) \end{pmatrix},$$

and

$$S_{m,j} = \begin{pmatrix} \sin(2\psi_{m,j})\cos(\Delta_{m,j}) \\ \sin(2\psi_{m,j})\sin(\Delta_{m,j}) \\ \cos(2\psi_{m,j}) \end{pmatrix},$$

denote a vector of the $j-$th experimental and modelled ellipsometric values. The sum of squares of differences

$$\mathcal{E}_{LS}^2 = \sum_{j=1}^{M} \mathcal{E}_j^2,$$

will be minimized, giving the desired fitted parameters. The averaged error is then

$$\mathcal{E} = \frac{1}{M} \sum_{j=1}^{M} \mathcal{E}_j.$$

Chapter 2

EM fields in a periodic medium

2.1   The Maxwell equations

In the forthcoming sections we will mostly follow the derivations in [12], [14] and [15].

The propagation of light in a medium is described by the Maxwell equations

$$\text{curl } \mathbf{E}(\mathbf{r}, t) = -\mu_0 \frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t} \qquad \text{div}\, (\varepsilon_0 \overleftrightarrow{\varepsilon_r}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}, t)) = 0,$$

$$\text{curl } \mathbf{H}(\mathbf{r}, t) = \varepsilon_0 \overleftrightarrow{\varepsilon_r}(\mathbf{r}) \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} \qquad \text{div } \mathbf{H}(\mathbf{r}, t) = 0,$$

(2.1)

where $\mathbf{E}$ is a vector of electric field, $\mathbf{H}$ is a vector of magnetic field, $\overleftrightarrow{\varepsilon_r}(\mathbf{r})$ is a tensor of relative permittivity, ε_0 is permittivity of vacuum, μ_0 is magnetic permeability of vacuum, $\mathbf{r} \in \mathbb{R}^3$ is a vector of spatial coordinates and $t \in \mathbb{R}$ is time. During this Thesis we will suppose that

$$\overleftrightarrow{\varepsilon_r}(\mathbf{r}) = \varepsilon_r(\mathbf{r}) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

i.e. the medium is isotropic. The electric and magnetic fields are considered to be in a form of harmonic monochromatic waves and can be written as

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) \mathrm{e}^{-i\omega t}, \quad \mathbf{H}(\mathbf{r}, t) = \mathbf{H}_0(\mathbf{r}) \mathrm{e}^{-i\omega t}.$$

Here $\mathbf{E}_0(\mathbf{r})$ and $\mathbf{H}_0(\mathbf{r})$ are initial states of the system, which can also depend on ω. Now clearly

$$\frac{\partial \mathbf{H}(\mathbf{r}, t)}{\partial t} = -i\omega \mathbf{H}(\mathbf{r}, t), \quad \frac{\partial \mathbf{E}(\mathbf{r}, t)}{\partial t} = -i\omega \mathbf{E}(\mathbf{r}, t).$$

and the system (2.1) can be rewritten as

$$\text{curl } \mathbf{E}_0(\mathbf{r}) = i\mu_0 \omega \mathbf{H}_0(\mathbf{r}) \qquad \text{div}\, (\varepsilon_0 \varepsilon_r(\mathbf{r}, \omega) \cdot \mathbf{E}_0(\mathbf{r})) = 0,$$

$$\text{curl } \mathbf{H}_0(\mathbf{r}) = -i\varepsilon_0 \varepsilon_r(\mathbf{r}) \omega \mathbf{E}_0(\mathbf{r}) \qquad \text{div } \mathbf{H}_0(\mathbf{r}) = 0.$$

(2.2)

The permittivity function $\varepsilon_r(\mathbf{r}, \omega)$ is in general a complex valued function which can depend on ω. The Ampère circuital law can be rewritten as

$$\text{curl } \mathbf{H}_0(\mathbf{r}) = -i\omega \left(\varepsilon(\mathbf{r}, \omega) - \frac{i\sigma(\mathbf{r}, \omega)}{\omega} \right) \mathbf{E}_0(\mathbf{r}),$$

where $\varepsilon(\mathbf{r})$ is a real valued permittivity and σ is the conductivity of the medium [16]. It is possible to redefine ε and σ so that ε_r is preserved:

$$\varepsilon \to \varepsilon + \varepsilon', \quad \sigma \to \sigma - \frac{\omega}{4\pi i} \varepsilon'.$$

The processes behind ε, σ are distinguishable only in the DC case, where ε describes the "bound charges", which are bound to the equilibrium positions and are stretched to new equilibrium positions by DC current, and σ describes the "free charges", which can move freely over arbitrary distance in response to DC field. In the case of AC field this distinctions blurs. At low frequencies $\omega \ll 1/\tau$, where τ is a relaxation time, the distinction can still be preserved - the free charges velocities will respond in phase with the field, while the bound charge velocities will respond out of phase with the field. At higher frequencies the distinction between free and bound electrons disappears - the convention is to denote by σ the response of electrons in partially filled bans and by ε the response of the electrons in completely filled bands. The response of all electrons is summed up into the single dielectric constant

$$\varepsilon_r(\mathbf{r}, \omega) = \varepsilon(\mathbf{r}, \omega) + \frac{i\sigma(\mathbf{r}, \omega)}{\omega},$$

which allows to use the same notation for metals and insulators. More details can be found in most of classical books about electrodynamics, e.g. [16], Appendix K or in [17], Section 2. In one example in Section 5 the real part of the square root of ε will be the refractive index, the complex part will be the extinction coefficient, see Fig 5.15.

For clarity we will drop the notation for the space variable $\mathbf{r}$, denote the fields simply $\mathbf{E}$ and $\mathbf{H}$ and also make a rescaling

$$\mathbf{r} \to k_0 \mathbf{r}, \ \mathbf{H} \to c\mu_0 \mathbf{H}_0, \ \nabla \to \frac{1}{k_0}\nabla, \ \mathbf{E} \to \frac{1}{\varepsilon_0}\mathbf{E}, \tag{2.3}$$

where c is the speed of light and $k_0 = 2\pi/\lambda$ is a wavenumber, with λ being the wavelength. Since $k_0 = \omega/c$ and $\varepsilon_0\mu_0 = c^{-2}$, the system (2.2) has after the rescaling the form

$$\begin{aligned} \operatorname{curl} \mathbf{E} &= i\mathbf{H} & \operatorname{div}(\varepsilon_r \cdot \mathbf{E}) &= 0, \\ \operatorname{curl} \mathbf{H} &= -i\varepsilon_r \mathbf{E} & \operatorname{div} \mathbf{H} &= 0. \end{aligned} \tag{2.4}$$

The rigorous formulation of boundary conditions and of the space of solutions will be given in the Section 2.6. As stated in the introduction our interest will be structures periodic in a given direction.
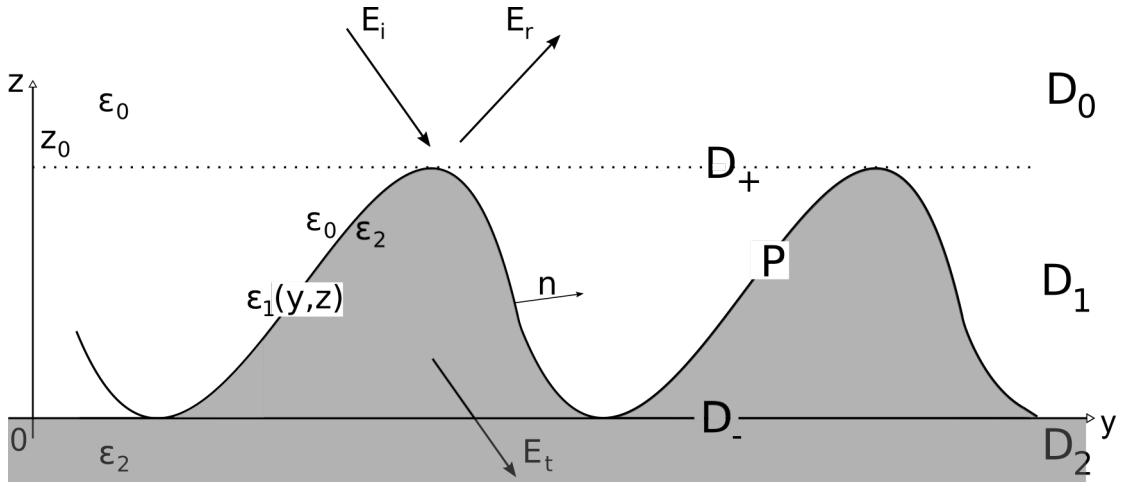
Figure 2.1: 1D-grating

We will consider a planar grating made of a material with the relative permittivity ε_r. We let $(x, y, z) \in \mathbb{R}^3$ denote the Cartesian coordinate system. The function ε_r is then supposed to be periodic in certain two directions, and we let z denote the axis perpendicular to these directions and z_0 the top of grating profile, see Fig. 2.1. The non-scaled height of the grating is $d = z_0/k_0$.

One of the most important type of a grating in this Thesis will be a relief grating, where the permittivity is defined by

$$\varepsilon_r(\mathbf{r}) = \begin{cases} \varepsilon_{r,0} & \text{if } (x,y,z) \in D_+, \\ \varepsilon_{r,2} & \text{if } (x,y,z) \in D_-, \end{cases}$$

with D_+ and D_- being the regions above and below grating and $\mathbf{r}$ denotes a vector of spatial coordinates. Since the system is scaled to $\varepsilon_0 = 1$, we will relabel $\varepsilon_{r,0}, \varepsilon_{r,2}$ to $\varepsilon_0, \varepsilon_2$, respectively. The grating profile P is assumed to be a bounded function. The grating is illuminated from D_+ by an incident wave

$$\mathbf{E}_i = \mathbf{E}_{0,i} \exp(i\mathbf{q_i r}) = \mathbf{E}_{0,i} \exp\left(in_0(y \sin\vartheta_i - z \cos\vartheta_i)\right),$$

where $\mathbf{q_i} = [0, n_0 \sin(\vartheta_i), -n_0 \cos(\vartheta_i)]$ is the wave vector, ϑ_i is the angle of incidence, $n_0 = \sqrt[c]{\varepsilon_0}$ is a refractive index of ambient media in D_+ and $\mathbf{E}_{0,i}$ is a vector of polarization of the incident wave. If the medium in D_+ is the air, then $n_0 = 1$. The coordinate system is chosen so that $\vartheta_i \in (0, \pi/2)$. This incident wave induces diffracted waves $\mathbf{E}_r$ (reflected) in D_+ and $\mathbf{E}_t$ (transmitted) wave in D_-. The diffracted fields must satisfy so-called radiation conditions:

Diffracted fields must remain bounded and propagate upwards in D_+ as $z \to \infty$,

Diffracted fields must remain bounded and propagate downwards in D_- as $z \to -\infty$, $\quad$ (2.5)

and these two conditions are supplemented with so-called interface conditions:

$$\begin{aligned} (\mathbf{E}_i|_{z\to P+} + \mathbf{E}_r|_{z\to P+}) \times \mathbf{n} &= \mathbf{E}_t|_{z\to P-} \times \mathbf{n}, \quad \text{for all } x, y \in \mathbb{R}, \\ (\mathbf{H}_i|_{z\to P+} + \mathbf{H}_r|_{z\to P+}) \times \mathbf{n} &= \mathbf{H}_t|_{z\to P-} \times \mathbf{n} \quad \text{for all } x, y \in \mathbb{R}, \end{aligned} \quad (2.6)$$

where $\mathbf{n}$ is a unit normal to the profile oriented towards region D_+, and the limits $z \to P\pm$ means that for fixed x the coordinate z converges to the grating profile in the respective domains. The conditions (2.6) express the fact that on the grating profile the tangential component of the electric field and the magnetic field strength must be continuous. This follows simply from the Faraday's law. Let us note, that it is true only under the assumption that the permittivities of the media are bounded. However, it will be always valid in all studied systems herein.

In practical computations it can be in some cases, see Section 2.3 and the following, advantageous to divide the problem into three regions — the region D_0 above the grating top, the grating region D_1 and the substrate region D_2, see Fig. 2.1. More precisely, the permittivity is defined by

$$\varepsilon_r(\mathbf{r}) = \begin{cases} \varepsilon_0 & \text{if } (x,y,z) \in D_0, \\ \varepsilon_{r,1} & \text{if } (x,y,z) \in D_1 \\ \varepsilon_2 & \text{if } (x,y,z) \in D_2, \end{cases}$$

The radiation conditions (2.5) are now supposed for the sets D_0, D_2, the interface conditions (2.6) must be fulfilled on every interface. In some cases it is convenient to divide the area D_1 even in more layers. Such procedure will be discussed in Sections 2.5, 3.5. For simplicity we relabel $\varepsilon_1 := \varepsilon_{r,1}$.

2.2 One-dimensional isotropic gratings

This Thesis is focused on 1D gratings. The permittivity function is invariant in one given direction, we let denote it as x. The function ε_r satisfies

$$\varepsilon_r(x, y, z) = \varepsilon_r(y, z) \text{ for all } (x, y, z) \in \mathbb{R}^3 \qquad (\varepsilon_r \text{ is indepdent of } x) \qquad (2.7)$$

$$\varepsilon_r(x, y + \Lambda_0, z) = \varepsilon_r(x, y, z) \text{ for all } (x, y, z) \in \mathbb{R}^3 \quad (\varepsilon_r \text{ is periodic in } y)$$

here Λ_0 denotes the periodicity of the grating. More precisely, if Λ is the period of the grating in the non-scaled coordinates, in the scaled coordinates the length of the period is $\Lambda_0 = \Lambda k_0 = 2\pi\Lambda/\lambda$, see (2.3).

Due to (2.7) the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ are independent of x. Keeping in mind that $\mathbf{E} = \mathbf{E}(y, z)$, $\mathbf{H} = \mathbf{H}(y, z)$ the equations (2.4) have the form

$$\text{curl } \mathbf{E} = \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}, \frac{\partial E_x}{\partial z}, -\frac{\partial E_x}{\partial y} \right) = i(H_x, H_y, H_z),$$

$$\text{curl } \mathbf{H} = \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}, \frac{\partial H_x}{\partial z}, -\frac{\partial H_x}{\partial y} \right) = i\varepsilon_r(E_x, E_y, E_z). \tag{2.8}$$

The vector normal to the profile is now $\mathbf{n} = (0, n_y, n_z)$ and the interface conditions (2.6) reduces to

$$E_{i,x}|_{z\to P+} + E_{r,x}|_{z\to P+} = E_{t,x}|_{z\to P-}, \tag{2.9}$$

$$E_{i,y}|_{z\to P+}n_z - E_{i,z}|_{z\to P+}n_y + E_{r,y}|_{z\to P+}n_z - E_{r,z}|_{z\to P+}n_y = E_{t,y}|_{z\to P-}n_z - E_{t,z}|_{z\to P-}n_y, \tag{2.10}$$

$$H_{i,x}|_{z\to P+} + H_{r,x}|_{z\to P+} = H_{t,x}|_{z\to P-}, \tag{2.11}$$

$$H_{i,y}|_{z\to P+}n_z - H_{i,z}|_{z\to P+}n_y + H_{r,y}|_{z\to P+}n_z - H_{r,z}|_{z\to P+}n_y = H_{t,y}|_{z\to P-}n_z - H_{t,z}|_{z\to P-}n_y, \tag{2.12}$$

The equations (2.8) together with (2.9)–(2.12) can be separated into two independent sets. We split the vector $\mathbf{E}$ as

$$\mathbf{E} = (E_x, 0, 0) + (0, E_y, E_z) = \mathbf{E}_{\hat{s}} + \mathbf{E}_{\hat{p}},$$

and the vector $\mathbf{H}$ then must be splitted as

$$\mathbf{H} = (0, H_y, H_z) + (H_x, 0, 0) = \mathbf{H}_{\hat{s}} + \mathbf{H}_{\hat{p}}.$$

The field having $\mathbf{E} = (E_x, 0, 0)$ is called TE-polarized (also s-polarized), the field with $\mathbf{H} = (H_x, 0, 0)$ is called TM-polarized (also p-polarized). The shortcut TE means "Transverse Electric", the vector of polarization of electric wave is perpendicular to the direction of wave propagation. The shortcut TM means "Transverse Magnetic", expressing that the vector of polarization of magnetic wave is perpendicular to direction of wave propagation.

TE polarization   The equations for TE polarization have the form

$$\frac{\partial E_x}{\partial z} = iH_y, \tag{2.13}$$

$$\frac{\partial E_x}{\partial y} = -iH_z, \tag{2.14}$$

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = -i\varepsilon_r E_x, \tag{2.15}$$

and with the interface conditions (2.9), (2.12). Formal putting of (2.13) and (2.14) into (2.15) gives one second-order equation

$$-\Delta E_x = \varepsilon_r E_x, \tag{2.16}$$

with interface conditions

$$E_{r,x}|_{z\to P+} - E_{t,x}|_{z\to P-} = -E_{i,x}|_{z\to P+} = -E_{0,ix}\exp(in_0(y\sin(\vartheta_i) - z\cos(\vartheta_i)))|_{z\to P+},$$

$$\mathbf{n} \cdot \nabla E_{r,x}\Big|_{z\to P+} - \mathbf{n} \cdot \nabla E_{t,x}\Big|_{z\to P-} = -\mathbf{n} \cdot \nabla E_{i,x}\Big|_{z\to P+} =$$

$$= -iE_{0,ix}\mathbf{n} \cdot \mathbf{q}_i \exp(in_0(y\sin(\vartheta_i) - z\cos(\vartheta_i)))\Big|_{z\to P+} \tag{2.17}$$

and the radiation conditions (2.5).

TM polarization   The equations for TM polarization are

$$\frac{\partial H_x}{\partial z} = -i\varepsilon_r E_y, \tag{2.18}$$

$$\frac{\partial H_x}{\partial y} = i\varepsilon_r E_z, \tag{2.19}$$

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = iH_x, \tag{2.20}$$

with the interface conditions (2.11), (2.10). The similar procedure for TM-polarized system formally yields

$$-\mathrm{div}\left(\frac{1}{\varepsilon_r}\nabla H_x\right) = H_x \tag{2.21}$$

with the radiation condition (2.5) and with the interface conditions

$$H_{r,x}|_{z\to P+} - H_{t,x}|_{z\to P-} = -H_{i,x}|_{z\to P+} = -H_{0,x}\exp(in_0(y\sin(\vartheta_i) - z\cos(\vartheta_i)))\Big|_{z\to P+}$$

$$\frac{1}{\varepsilon_0}\mathbf{n}\cdot\nabla H_{r,x}\Big|_{z\to P+} - \frac{1}{\varepsilon_2}\mathbf{n}\cdot\nabla H_{t,x}\Big|_{z\to P-} = -\frac{1}{\varepsilon_0}\mathbf{n}\cdot\nabla H_{i,x}\Big|_{z\to P-} =$$

$$= -iH_{0,x}\mathbf{n}\cdot\mathbf{q}_i\exp(in_0(y\sin(\vartheta_i) - z\cos(\vartheta_i)))\Big|_{z\to P+}, \tag{2.22}$$

where $\mathbf{H}_0$ can be found from the field of the incident electric wave as

$$\mathbf{H}_0 = \mathbf{q}_i \times \mathbf{E}_{0,i}.$$

If there are more interfaces, the condition (2.22) must be considered for every interface separately.

The general system is now described as a superposition of TE and TM polarized states (s- and p-polarized states respectively), see Remark on general polarization on pg. 17.

2.3   The Rayleigh expansion and pseudoperiodicity

It can be useful to divide the whole problem into three regions in z as $(-\infty, 0)$, $(0, z_0)$, (z_0, ∞). In these three regions the relative permittivity is

$$\varepsilon_r(y,z) = \begin{cases} \varepsilon_0 & \text{if } z > z_0, \\ \varepsilon_1(y,z) & \text{if } z \in (0, z_0) \\ \varepsilon_2 & \text{if } z < 0 \end{cases}$$

In such a case, it is necessary to consider four interface conditions (one pair for each interface):

$$\begin{aligned}
(\mathbf{E}_i|_{z\to z_0+} + \mathbf{E}_r|_{z\to z_0+}) \times \mathbf{n} &= (\mathbf{E}_+|_{z\to z_0-} + \mathbf{E}_-|_{z\to z_0-}) \times \mathbf{n} \\
(\mathbf{H}_i|_{z\to z_0+} + \mathbf{H}_r|_{z\to z_0-}) \times \mathbf{n} &= (\mathbf{H}_+|_{z\to z_0-} + \mathbf{H}_-|_{z\to z_0-}) \times \mathbf{n} \\
(\mathbf{E}_+|_{z\to 0+} + \mathbf{E}_-|_{z\to 0+}) \times \mathbf{n} &= \mathbf{E}_t|_{z\to 0-} \times \mathbf{n} \\
(\mathbf{H}_+|_{z\to 0+} + \mathbf{H}_-|_{z\to 0+}) \times \mathbf{n} &= \mathbf{H}_t|_{z\to 0-} \times \mathbf{n},
\end{aligned} \tag{2.23}$$

for all $y \in \mathbb{R}$, cf. (2.6).

TE polarization   The periodicity of grating induces a translation symmetry of the electric and magnetic field in the substrate and superstrate media, which can be described as

$$\mathbf{E}_r(x, y + \Lambda_0, z) = \mathbf{E}_r(x, y, z)\exp[iq_0\Lambda_0] \quad \text{for all } x, y \in \mathbb{R}, z > z_0$$

$$\mathbf{E}_t(x, y + \Lambda_0, z) = \mathbf{E}_t(x, y, z)\exp[iq_0\Lambda_0] \quad \text{for all } x, y \in \mathbb{R}, z < 0,$$

where $q_0 = n_0 \sin(\vartheta_i)$, z_0 is a top of the grating and 0 is its bottom. The nonzero component of $\mathbf{E}_r, \mathbf{E}_t$

$$E_{rx}(x, y, z) \exp(-iq_0 y), \quad E_{tx}(x, y, z) \exp(-iq_0 y),$$

are periodic and can be written in a form of Fourier series

$$E_{rx}(y, z) \exp(-iq_0 y) = \sum_{m=-\infty}^{\infty} E_{rx}^m(z) \exp(imqy),$$

$$E_{tx}(y, z) \exp(-iq_0 y) = \sum_{m=-\infty}^{\infty} E_{tx}^m(z) \exp(imqy),$$

with $q = \lambda/\Lambda = 2\pi/\Lambda_0$ (λ denotes the wavelength of the incident wave). Now

$$E_{rx}(x, y, z) = \sum_{m=-\infty}^{\infty} E_{rx}^m(z) \exp(iq_m y),$$

$$E_{tx}(x, y, z) = \sum_{m=-\infty}^{\infty} E_{tx}^m(z) \exp(iq_m y),$$

where $q_m := n_0 \sin(\vartheta_i) + mq$. The function ε_r is constant in half-planes $z > z_0$ and $z < 0$. This formulae, used in (2.16) yields two ODE's:

$$\sum_{m=-\infty}^{\infty} \left(\frac{\mathrm{d}^2 E_{rx}^m(z)}{\mathrm{d}z^2} + (\varepsilon_0 - q_m^2) E_{rx}(z) \right) \exp(iq_m y) = 0, \quad \text{for } z > z_0, \ y \in \mathbb{R},$$

$$\sum_{m=-\infty}^{\infty} \left(\frac{\mathrm{d}^2 E_{tx}^m(z)}{\mathrm{d}z^2} + (\varepsilon_2 - q_m^2) E_{tx}(z) \right) \exp(iq_m y) = 0 \quad \text{for } z < 0, \ y \in \mathbb{R},$$

which splits into an infinite system of ODE's

$$\frac{\mathrm{d}^2 E_{rx}^m(z)}{\mathrm{d}z^2} + s_{r,m}^2 E_{rx}^m(z) = 0 \ \text{ for } z > z_0, \qquad \frac{\mathrm{d}^2 E_{tx}^m(z)}{\mathrm{d}z^2} + s_{t,m}^2 E_{tx}^m(z) = 0 \ \text{ for } z < 0,$$

where $s_{r,m} = \pm\sqrt[c]{\varepsilon_0 - q_m^2}$, $s_{t,m} = \pm\sqrt[c]{\varepsilon_2 - q_m^2}$. The coefficients $s_{t,m}$ are chosen in a way that $\mathfrak{Re}(s_{t,m}) + \mathfrak{Im}(s_{t,m}) > 0$, similarly for $s_{r,m}$. The fields are then

$$E_{rx}(x, y, z) = \sum_{m=-\infty}^{\infty} A_{rx}^m \exp(iq_m y + is_{r,m} z) + B_{rx}^m \exp(iq_m y - is_{r,m} z) = 0, \quad z > z_0$$

$$E_{tx}(x, y, z) = \sum_{m=-\infty}^{\infty} A_{tx}^m \exp(iq_m y - is_{t,m}) + B_{tx}^m \exp(iq_m y + is_{t,m}) = 0, \quad z < 0,$$

where $A_{rx}^m, A_{tx}^m, B_{rx}^m, B_{tx}^m$ are constants. Due to the radiation condition the downward propagation of the reflected waves and upward propagation of the transmitted waves must be excluded, i.e. $B_{rx}^m = 0$, $B_{tx}^m = 0$ for all $m \in \mathbb{Z}$, and hence, the final form of the fields outside the grating region is

$$\mathbf{E}_i(y, z) = \mathbf{A}_i \exp(iq_0 y - is_{i,0} z) = \mathbf{A}_i \exp(in_0(\sin(\vartheta_i)y - \cos(\vartheta_i)z))$$

$$\mathbf{E}_r(y, z) = \sum_{m=-\infty}^{\infty} \mathbf{A}_r^m \exp(iq_m y + is_{r,m} z), \quad \text{for all } y \in \mathbb{R}, z > z_0$$

$$\mathbf{E}_t(y, z) = \sum_{m=-\infty}^{\infty} \mathbf{A}_t^m \exp(iq_m y - is_{t,m}) \quad \text{for all } y \in \mathbb{R}, \ z < 0,$$

where only x-components of the vectors are nonzero. It remains to describe a wave propagation in the grating region D_1. The Rayleigh expansion is not possible here, but it is possible to treat it via Floquet theorem — which treats the fields via pseudo-Fourier series in a similar way as the Rayleigh series in previous derivations. Let $z_1 \in (0, z_0)$ be arbitrary fixed. Due to the periodicity of the grating, the electric field is again periodic here, which mathematically stated means that

$$\mathbf{E}(y + \Lambda_0, z_1) = \mathbf{E}(y, z_1) \exp[iq_0\Lambda_0].$$

Using Floquet Theorem the electric field can be written as

$$\mathbf{E}(y, z_1) = \mathbf{e}(y, z_1) \exp(iq_0 y),$$

with $e(y, z_1)$ being periodic in y with the periodicity Λ_0. Therefore $e(y, z_1)$ can be expanded into the Fourier series

$$\mathbf{e}(y, z_1) = \sum_{n=-\infty}^{\infty} \mathbf{e}_n(z_1) \exp(inqy),$$

and the entire field is then

$$\mathbf{E}(y, z) = \sum_{n=-\infty}^{\infty} \mathbf{e}_n(z) \exp(iq_n y),$$

with $q_n = q_0 + nq$. In the same way we can expand the magnetic fields. The permittivity function ε_1 is periodic in y for any fixed $z_1 \in (0, z_0)$ and can be expanded into a Fourier series

$$\varepsilon_1(y, z) = \sum_{m=-\infty}^{\infty} \hat{\varepsilon}_m(z) e^{imqy}, \quad z \in (0, z_0), \quad y \in \mathbb{R}.$$

For TE polarization the vector $\mathbf{e} = (e_x, 0, 0)$ and

$$E_x(y, z) = \sum_{n=-\infty}^{\infty} e_{x,n}(z) \exp(iq_n y).$$

For simplicity we drop the index x. Inserting this expansion of E_x and ε_1 into (2.16) gives

$$\left[\frac{d^2}{dy^2} + \frac{d^2}{dz^2}\right] \sum_{n=-\infty}^{\infty} e_n(z) \exp(iq_n y) = -\left(\sum_{m=-\infty}^{\infty} \hat{\varepsilon}_m \exp(imqy)\right)\left(\sum_{n=-\infty}^{\infty} e_n(z) \exp(iq_n y)\right),$$

and the use of Laurent rule for a multiplication of two series

$$\left(\sum_{m=-\infty}^{\infty} f_m\right)\left(\sum_{n=-\infty}^{\infty} g_n\right) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f_{n-m} g_m,$$

gives a system

$$\left[\frac{d^2}{dy^2} + \frac{d^2}{dz^2}\right] \sum_{n=-\infty}^{\infty} e_n(z) \exp(iq_n y) = -\sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} e_n(z)\hat{\varepsilon}_{n-m}(z) \exp(iq_m y).$$

The exponential functions are linearly independent which leads to a system of coupled ODE's

$$\frac{d^2 e_n(z)}{dz^2} = q_n^2 e_n(z) - \sum_{m=-\infty}^{\infty} \hat{\varepsilon}_{n-m}(z) e_m(z) = \left(\text{diag}(\mathbf{q})^2 \mathbf{e}(z) - [\![\varepsilon_1]\!] \, \mathbf{e}(z)\right)_n, \quad (2.24)$$

where $(\llbracket \varepsilon_1 \rrbracket)_{mn} := \varepsilon_{m-n}(z)$ is a Toeplitz matrix of Fourier coefficients of ε_1. If ε_1 is independent of z, then $\llbracket \varepsilon_1 \rrbracket$ is a constant matrix and has a form

$$\llbracket \varepsilon_1 \rrbracket = \begin{pmatrix} \hat{\varepsilon}_0 & \hat{\varepsilon}_{-1} & \hat{\varepsilon}_{-2} & \hat{\varepsilon}_{-3} & \cdots \\ \hat{\varepsilon}_1 & \hat{\varepsilon}_0 & \hat{\varepsilon}_{-1} & \hat{\varepsilon}_{-2} & \ddots \\ \hat{\varepsilon}_2 & \hat{\varepsilon}_1 & \hat{\varepsilon}_0 & \hat{\varepsilon}_{-1} & \ddots \\ \hat{\varepsilon}_3 & \hat{\varepsilon}_2 & \hat{\varepsilon}_1 & \hat{\varepsilon}_0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

As there are four sets of constants $A_{rx}^m, A_{tx}^m, A_+^m, A_-^m$ to determine, this second order ODE must be complemented with four initial (interface) conditions derived from (2.23), see also (2.17).

TM polarization Similarly the TM polarization can be treated using Rayleigh expansion and Floquet Theorem. Above and below the grating region the similar procedure as for TE polarization leads to series

$$\mathbf{H}_i(y,z) = \mathbf{A}_i \exp(in_0(\sin(\vartheta_i)y - \cos(\vartheta_i)z)) = \mathbf{A}_i \exp(iq_0 y - is_{0,i}z)$$

$$\mathbf{H}_r(y,z) = \sum_{m=-\infty}^{\infty} \mathbf{A}_r^m \exp\left(i(q_m y - s_{r,m}z)\right), \quad \text{for all } y \in \mathbb{R}, z > z_0$$

$$\mathbf{H}_t(y,z) = \sum_{m=-\infty}^{\infty} \mathbf{A}_t^m \exp\left(i(q_m y - s_{t,m}z)\right), \quad \text{for all } y \in \mathbb{R}, \ z < 0.$$

and thence, for the TM polarized field $\mathbf{H} = (H_x, 0, 0)$ the nonzero components are

$$H_{ix}(x,y,z) = A_{ix,0} \exp(i(q_0 y - s_{i,0}z)), \quad z > z_0$$

$$H_{rx}(x,y,z) = \sum_{m=-\infty}^{\infty} A_{rx}^m(z) \exp\left(iq_m y\right), \quad z > z_0$$

$$H_{tx}(x,y,z) = \sum_{m=-\infty}^{\infty} A_{tx}^m(z) \exp\left(iq_m y\right), \quad z < 0.$$

The expansion in the grating region is more complex. Let ε_1 be independent of z. Similar procedure as for TE case gives that the field H_x in the grating region can be expressed as

$$H_x(y,z) = \sum_{m=-\infty}^{\infty} h_m(z) \exp(iq_m y).$$

The equation (2.21) can be rewritten as

$$\partial_z^2 H_x - \varepsilon_1 \partial_y \left(\frac{1}{\varepsilon_1} \partial_y H_x\right) = \varepsilon_1 H_x.$$

and putting the Fourier series of ε_1 and H_x and using the Laurent rule leads to

$$\frac{dh_k(z)}{dz^2} = \sum_{n=-\infty}^{\infty} \llbracket \varepsilon_1 \rrbracket_{kn} \left(\sum_{m=-\infty}^{\infty} \left(\left\llbracket \frac{1}{\varepsilon_1} \right\rrbracket_{nm} \right) q_m q_n h_m(z) - \delta_{mn} h_m(z) \right). \tag{2.25}$$

This equation must be supplemented with the four respective initial (interface) conditions. It will be seen later that although this formulation is correct, the simple truncation is without further modifications inconvenient as a numerical method.

General polarization A general polarization state can be expressed as a superposition of TM and TE polarized states:

$$\mathbf{E}_i = \mathbf{E}_{i,\hat{s}} + \mathbf{E}_{i,\hat{p}} = E_{i,s}\mathbf{x} + E_{i,p}(\mathbf{y}\cos(\vartheta_i) + \mathbf{z}\sin(\vartheta_i)),$$
$$\mathbf{A}_r^m = \mathbf{A}_{r,\hat{s}}^m + \mathbf{A}_{r,\hat{p}}^m = A_{r,s}^m\mathbf{x} + A_{r,p}^m(-\mathbf{y}\cos(\vartheta_i) - \mathbf{z}\sin(\vartheta_i)),$$
$$\mathbf{A}_t^m = \mathbf{A}_{t,\hat{s}}^m + \mathbf{A}_{t,\hat{p}}^m = A_{t,s}^m\mathbf{x} + A_{t,p}^m(\mathbf{y}\cos(\vartheta_i) + \mathbf{z}\sin(\vartheta_i)),$$

for all $m \in \mathbb{Z}$ and similarly for the magnetic fields

$$\mathbf{H}_i = \mathbf{H}_{i,\hat{s}} + \mathbf{H}_{i,\hat{p}} = H_{i,s}\mathbf{x} + H_{i,p}(\mathbf{y}\cos(\vartheta_i) + \mathbf{z}\sin(\vartheta_i)),$$
$$\mathbf{A}_r^m = \mathbf{A}_{r,\hat{s}}^m + \mathbf{A}_{r,\hat{p}}^m = A_{r,s}^m\mathbf{x} + A_{r,p}^m(-\mathbf{y}\cos(\vartheta_i) - \mathbf{z}\sin(\vartheta_i)),$$
$$\mathbf{A}_t^m = \mathbf{A}_{t,\hat{s}}^m + \mathbf{A}_{t,\hat{p}}^m = A_{t,s}^m\mathbf{x} + A_{t,p}^m(\mathbf{y}\cos(\vartheta_i) + \mathbf{z}\sin(\vartheta_i)),$$

for all $m \in \mathbb{Z}$.

2.4   Energy balance

The energy conservation is used to get quantities of the grating problem which can be measured experimentally. The energy of the incident wave must be the same as the energy of the reflected and transmitted waves. The intensity is a size of Poynting vector and the energy balance can be expressed as an equality between the z-propagating orders

$$\sum_{m \in U_0} s_{r,m}|A_{rx}^m|^2 + \sum_{m \in U_2} \frac{\varepsilon_0}{\varepsilon_2} s_{t,m}|A_{tx}^m|^2 = |s_{i,0}|^2,$$

where U_0, U_2 contains the indices of z-propagating orders in the respective domains. Dividing by $|s_{i,0}|^2$ leads to an energy balance

$$\sum_{m \in U_0} s_{r,m}\frac{|A_{rx}^m|^2}{|s_{i,0}|^2} + \sum_{m \in U_2} s_{t,m}\frac{\varepsilon_0}{\varepsilon_2}\frac{|A_{tx}^m|^2}{|s_{i,0}|^2} = 1.$$

The diffraction efficiencies are then defined by

$$\begin{aligned} r_{0,m} &= \frac{|A_{rx}^m|^2}{|s_{i,0}|^2}s_{0,m}, \quad m \in U_0 \\ t_{2,m} &= \frac{\varepsilon_0}{\varepsilon_2}\frac{|A_{tx}^m|^2}{|s_{i,0}|^2}s_{2,m}, \quad m \in U_2. \end{aligned} \tag{2.26}$$

Of a special interest will be the zeroth diffraction order. The generalized amplitude reflection and transmission coefficients

$$\begin{aligned} \mathfrak{r}_m &:= \frac{A_{rx}^m}{A_{ix}}, \\ \mathfrak{t}_m &:= \frac{A_{tx}^m}{A_{ix}} \end{aligned} \tag{2.27}$$

are found for both polarizations and the ellipsometry parameters can be calculated as

$$\begin{aligned} \rho_r^{(m)} &= \frac{\mathfrak{r}_m^p}{\mathfrak{r}_m^s} = \tan\psi^{(r)}\exp(i\Delta^{(r)}), \\ \rho_t^{(m)} &= \frac{\mathfrak{t}_m^p}{\mathfrak{t}_m^s} = \tan\psi^{(t)}\exp(i\Delta^{(t)}), \end{aligned}$$

where s, p denotes the polarization of the incident wave (TE and TM respectively). Of a special interest are the zeroth orders from which the ellipsometry parameters can be obtained via the formula

$$\psi = \arctan(|\rho_r^{(0)}|),$$
$$\delta = \text{Arg}(\rho_r^{(0)}).$$

These ellipsometry parameters can be measured experimentally, the experimental techniques are explained in Section 1.3.

2.5   Multilayered gratings and staircase approximation

The grating region or substrate can be composed of more layers. The technique to solve these systems is to solve equations of light propagation in each layer separately, and then join them by a use of the interface conditions. More precisely, the system which consists of N layers with the non-intersecting profiles P_k splits into

$$-\Delta E_x = \varepsilon_{r,j} E_x, \ j = 0, ..., N+1 \tag{2.28}$$

where

$$\varepsilon_r = \begin{cases} \varepsilon_{r,0} = \varepsilon_0 & \text{if } (y, z) \in \text{superstrate}, \\ \varepsilon_{r,1} & \text{if } (y, z) \in \ 1^{\text{st}} \text{ layer}, \\ \dots, \\ \varepsilon_{r,N} & \text{if } (y, z) \in \ N^{\text{th}} \text{ layer}, \\ \varepsilon_{r,N+1} = \varepsilon_2 & \text{if } (y, z) \in \ \text{substrate}. \end{cases} \tag{2.29}$$

To avoid the presence of too much indices in the expression it will be made a relabeling $E \equiv E_x$. The interface conditions are

$$E_i|_{z \to P_{0+}} + E_r|_{z \to P_{0+}} = E_{+,1}|_{z \to P_{0-}} + E_{-,1}|_{z \to P_{0-}},$$
$$E_{+,k}|_{z \to P_{k+}} + E_{-,k}|_{z \to P_{k+}} = E_{+,k+1}|_{z \to P_{k-}} + E_{-,k+1}|_{z \to P_{k-}} \quad \text{for } k = 1, ..., N-1$$
$$E_{+,N}|_{z \to P_{N+}} + E_{-,N}|_{z \to P_{N+}} = E_t|_{z \to P_{N-}},$$
$$\mathbf{n} \cdot (\nabla E_i + \nabla E_r)\Big|_{z \to P_{0+}} = \mathbf{n} \cdot (\nabla E_{+,1} + \nabla E_{-,1})\Big|_{z \to P_{0-}} \tag{2.30}$$
$$\mathbf{n} \cdot (\nabla E_{+,k} + \nabla E_{-,k})\Big|_{z \to P_{k+}} = \mathbf{n} \cdot (\nabla E_{+,k+1} + \nabla E_{-,k+1})\Big|_{z \to P_{k-}} \quad \text{for } k = 1, ..., N-1$$
$$\mathbf{n} \cdot (\nabla E_{+,N} + \nabla E_{-,N})\Big|_{z \to P_{N+}} = \mathbf{n} \cdot \nabla E_t\Big|_{z \to P_{N-}},$$

where $+$ and $-$ index denotes the incoming and outgoing fields w.r.t. to the boundary and a number in the index determines the permittivity (layer). For a system consisting of N parallel layers having the interfaces in the planes $z_0 > z_1 > ... > z_{N+1} = 0$, which will be the case of the FMM, the permittivity and the interface conditions can be formulated as follows:

$$\varepsilon_r = \begin{cases} \varepsilon_0 = 1 & \text{if } z > z_0, \\ \varepsilon_{r,1} & \text{if } z_1 < z < z_0, \\ \dots, \\ \varepsilon_{r,N} & \text{if } 0 < z < z_N, \\ \varepsilon_{r,N+1} = \varepsilon_2 & \text{if } z < 0. \end{cases} \tag{2.31}$$

The interface conditions for such problem are derived from (2.30) by putting $P_i := z_i, \ i = 0, ..., N$.

The equations in the case of TM polarization are

$$\partial_z^2 H_x - \varepsilon_r \partial_y \left(\frac{1}{\varepsilon_r} \partial_y H_x \right) = \varepsilon_r H_x.$$

where ε_r is given by (2.29). The interface conditions are

$$H_r|_{z \to P_{0+}} + H_i|_{z \to P_{0+}} = H_{+,1}|_{z \to P_{0-}} + H_{-,1}|_{z \to P_{0-}},$$

$$H_{+,k}|_{z \to P_{k+}} + H_{-,k}|_{z \to P_{k+}} = H_{+,k+1}|_{z \to P_{k-}} + H_{-,k+1}|_{z \to P_{k-}} \quad \text{for } k = 1, ... N-1$$

$$H_{+,N}|_{z \to P_{N+}} + H_{-,N}|_{z \to P_{N+}} = H_t|_{z \to P_{N-}},$$

$$\mathbf{n} \cdot \frac{1}{\varepsilon_0} \left(\nabla H_i + \nabla H_r \right) \Big|_{z \to P_{0+}} = \mathbf{n} \cdot \frac{1}{\varepsilon_{r,1}} \left(\nabla H_{1,+} + \nabla H_{1,-} \right) \Big|_{z \to P_{0-}} \tag{2.32}$$

$$\mathbf{n} \cdot \frac{1}{\varepsilon_{r,k}} \left(\nabla H_{+,k} + \nabla H_{-,k} \right) \Big|_{z \to P_{k+}} = \mathbf{n} \cdot \frac{1}{\varepsilon_{r,k+1}} \left(\nabla H_{+,k+1} + \nabla H_{-,k+1} \right) \Big|_{z \to P_{k-}}$$

$$\text{for } k = 1, ... N-1$$

$$\mathbf{n} \cdot \frac{1}{\varepsilon_{r,N}} \left(\nabla H_{+,N} + \nabla H_{-,N} \right) \Big|_{z \to P_{N+}} = \frac{1}{\varepsilon_2} \nabla H_t \Big|_{P_{N-}}.$$

on every interface. Especially, the permittivity for the problem with N parallel interfaces in the planes $z_0 > z_1 > ... > z_{N+1} = 0$ is (2.31) and the interface conditions are derived from (2.32) by a simple substitution $P_k := z_k$, $k = 0, ..., N$.

This approach can also be used to treat curved gratings, e.g. sinusoidal or triangular via Fourier Modal Method using a so-called staircase approximation. The grating region is divided into several layers, each respective to the shape of grating, see Fig. 2.2 and the technique for multilayered grating is applied on this system.
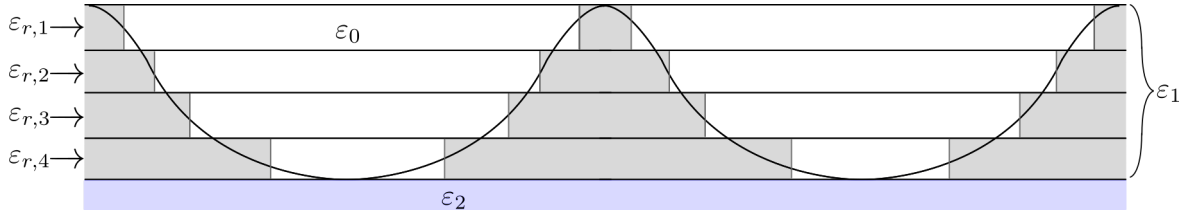

Figure 2.2: Staircase approximation

The equations for grating region in TE and TM polarization can be derived in the same way as (2.24), (2.25). This derived numerical scheme will be called the Lalanne–Morris technique (LMT). However, without further modifications it cannot be truncated, because then it does not converge to a sinusoidal grating as $N \to \infty$, see [18], Chapter VI 5.3, [19]. The formulation which truncation converges to the exact result was developed in [9] and will be described in Section 3.5. Another way how to treat sinusoidal or in general gratings with $C^{0,1}$ profile is in use of Coordinate transformation method, which will be discussed in Chapter 4.

2.6 Variational formulation, well-posedness and the FEM

The RCWA and the C-Method are not the only approximation technique which can be used to solve the grating problem. Often used method is also the Finite Element Method (FEM), which will be shortly described here, because much more rigorous results are known for this problem. The two most important problems are related with the well-posedness. The first one is the direct problem:

Given the grating geometry and incident field, solution of the Maxwell equation predicts the behavior of the outgoing fields.

The second problem is related to the inverse problem:

Given the incident field and outgoing fields, the grating profile is uniquely determined.

19

These two problems were an interest of several papers and books, e.g. [20], [21], [22]. Here we will shortly review the results given in [20] and [22], Chapter 6.

Direct problem The case of TE polarization will be discussed first. Since the existence and uniqueness will be dependent on the angular frequency ω of the light, it is necessary to go back to the non-scaled problem, which is a Helmholtz equation

$$\triangle E_x + k^2 E_x = 0 \quad \text{in } \mathbb{R}^2, \tag{2.33}$$

where $k^2 = \mu\varepsilon_r\omega^2$. This equation is supplemented with the radiation conditions and the assumption of periodicity of ε_r. The last assumption induces the pseudo-periodicity of the solution. We let denote $k_0 = \mu\varepsilon_0\varepsilon_{r,0}\omega^2$, $k_2 = \mu\varepsilon_0\varepsilon_{r,2}\omega^2$ and suppose that $\text{Im}(k_0) \geq 0$, $\text{Im}(k_2) \geq 0$. If E_x is a pseudo-periodic solution of (2.33), then the function $E_\alpha := \exp(-i\alpha y)E_x$, with $\alpha := 2\pi/\Lambda$, is periodic in y and it is a solution of the problem

$$(\Delta_\alpha + k^2)E_\alpha := (\Delta + 2i\alpha\partial_y - |\alpha|^2)E_x = 0.$$

The function E_α is periodic in y and can be expanded to the Fourier series

$$E_\alpha = \sum_{m=-\infty}^{\infty} e_m(z) \exp(i\alpha_m y).$$

In the domains D_0 and D_2 the function E_α can be represented by the Rayleigh series

$$\begin{aligned}
E_\alpha|_{D_0} &= \sum_{m=-\infty}^{\infty} A_{r,m} \exp[i(\alpha_m y + \beta_{0,m} z)], \\
E_\alpha|_{D_2} &= \sum_{m=-\infty}^{\infty} A_{t,m} \exp[i(\alpha_m y - \beta_{2,m} z)],
\end{aligned} \tag{2.34}$$

where α_m, β_m are non-scaled equivalents of q_m, s_m from previous sections. It can be found that

$$\begin{aligned}
e_m(z) &= e_m(d) \exp[i\beta_{0,n}(z-d)], && \text{for } m \neq 0, \ z \geq d, \\
e_0(z) &= e_m(d) \exp[i\beta_{0,m}(z-d)] + \exp(-i\beta_0 z), && \text{for } z \geq d, \\
e_m(z) &= e_m(0) \exp[-i\beta_{2,n} z] && \text{for } z \leq 0.
\end{aligned} \tag{2.35}$$

The boundary conditions can be then constructed using (2.34) and (2.35):

$$\begin{aligned}
\left.\frac{\partial E_\alpha}{\partial \mathbf{n}}\right|_{z=d} &= T_0(E_\alpha)|_{z=d} - i\beta_0 \exp(-\beta_0 d), \\
\left.\frac{\partial E_\alpha}{\partial \mathbf{n}}\right|_{z=0} &= T_2(E_\alpha)|_{z=0},
\end{aligned} \tag{2.36}$$

where

$$T_j(E_\alpha) = \sum_{m=-\infty}^{\infty} i\beta_{j,m} e_m(z) \exp(-i\alpha_m y).$$

The grating problem can be then formulated as follows. Find $E_\alpha \in H^1(D_1)$ satisfying

$$\Delta_\alpha E_\alpha + k^2 E_\alpha = 0$$

and the boundary conditions (2.36). The weak formulation of the grating problem is

$$-\int_{D_1} \nabla E_\alpha \cdot \nabla\overline{\phi} + 2i\alpha \int_{D_1} \partial_y E_\alpha \overline{\phi} - \int_{D_1} |\alpha|^2 E_\alpha \overline{\phi}+$$

$$+ \int_{\{z=d\}} T_0(E_\alpha)\overline{\phi} + \int_{\{z=0\}} T_2(E_\alpha)\overline{\phi} - 2i\beta_0 \int_{\{z=d\}} \exp(-i\beta_0 d)\overline{\phi} = 0, \qquad (2.37)$$

$$\text{for all } \phi \in \mathrm{H}^1(D_1).$$

The problem in TM polarization can be treated similarly. The elliptic equation is

$$\nabla_\alpha \cdot \left[\left(\frac{1}{k^2}\right)\nabla_\alpha H_\alpha\right] + H_\alpha = 0,$$

with $\nabla_\alpha := \nabla + i(\alpha, 0)$. The boundary conditions can be derived similarly as for TE case. The weak formulation in the case of TM polarization is

$$-\frac{1}{k^2}\int_{D_1} \nabla H_\alpha \cdot \nabla\overline{\phi} + \int_{D_1} \left(\omega^2 - \frac{\alpha^2}{k^2}\right) H_\alpha \overline{\phi} + i\alpha \int_{D_1} \frac{1}{k^2}\left(\partial_y H_\alpha \overline{\phi}\right) - i\alpha \int_{D_1} \frac{1}{k^2} H_\alpha \overline{\partial_y \phi}+$$

$$+ \int_{\{z=d\}} \frac{1}{k_0^2} T_0(H_\alpha)\overline{\phi} + \int_{\{z=0\}} \frac{1}{k_2^2} T_2(H_\alpha)\overline{\phi} - 2i\beta_0 \int_{\{z=d\}} \frac{1}{k_2^2}\exp(-i\beta_0 d)\overline{\phi} = 0,$$

$$\text{for all } \phi \in \mathrm{H}^1(D_1).$$

$$(2.38)$$

As now the weak formulation of both problems is done, it is possible to discuss the results related to it. The existence and uniqueness of the solution of both problems was discussed in [23] and [24]. The results can be summarized as follows:

The diffraction problem has a unique solution up to a set of countable frequencies ω_j,

$$|\omega_j|^2 \to \infty.$$

It is necessary to emphasize the role of the frequency ω here. As the form of this results suggests, the authors have rewritten the weak problem as an operator equation and then used the Fredholm theory. The second result concerns a special case:

If $Im\ \varepsilon_0 > 0$ or $Im\ \varepsilon_2 > 0$, then the solution of the grating problem is unique.

This condition is clearly not satisfied for dielectric gratings. The smooth dependence on data was proved in [23]. However, the continuous dependence on the grating profile was proved only for the case of C^1 smoothness. More precisely, there exists a constant $C > 0$ such that if f, g are two gratings profile and E_{xf} and E_{xg} are respective solutions of the grating problem, then

$$\|E_{xf} - E_{xg}\| \le C\|f - g\|_{C^1}.$$

Analogously for TM polarization.

Since the existence results are based on a variational formulation of the problem, they are not applicable for a formulation used in Fourier Modal Methods. However, the approximations obtained by the FMM were compared to the FEM ones (see e.g. [25], [26]) and a good coincidence in most standard problems suggests that the FMM give a fair approximation of the solution of grating problem.

Finite element methods A solution of the grating problem can be approximated also by the use of the Finite Element Method. It is based on the formulation (2.37) for TE polarization and (2.38) for TM polarization. The general formulation is

$$a_{TE}(E_\alpha, \phi) = (f_{TE}, \phi), \quad a_{TM}(H_\alpha, \phi) = (f_{TM}, \phi)$$

where

$$a_{TE}(E_\alpha, \phi) := -\int_{D_1} \nabla E_\alpha \cdot \nabla\overline{\phi} + 2i\alpha \int_{D_1} \partial_y E_\alpha \overline{\phi} - \int_{D_1} |\alpha|^2 E_\alpha \overline{\phi} + \int_{\{z=d\}} T_0(E_\alpha)\overline{\phi} + \int_{\{z=0\}} T_2(E_\alpha)\overline{\phi},$$

$$(f_{TE}, \phi) := 2i\beta_0 \int_{\{z=z_0\}} \exp(-i\beta_0 d)\overline{\phi},$$

$$a_{TM}(H_\alpha, \phi) := -\frac{1}{k^2}\int_{D_1} \nabla H_\alpha \cdot \nabla\overline{\phi} + \int_{D_1} \left(\omega^2 - \frac{\alpha^2}{k^2}\right) H_\alpha \overline{\phi} + i\alpha \int_{D_1} \frac{1}{k^2}\left(\partial_y H_\alpha \overline{\phi}\right) - i\alpha \int_{D_1} \frac{1}{k^2} H_\alpha \overline{\partial_y\phi} +$$

$$+ \int_{\{z=d\}} \frac{1}{k_0^2} T_0(H_\alpha)\overline{\phi} + \int_{\{z=0\}} \frac{1}{k_2^2} T_2(H_\alpha)\overline{\phi},$$

$$(f_{TM}, \phi) = 2i\beta_0 \int_{\{z=d\}} \exp(-i\beta_0 d)\overline{\phi}$$

A finite-dimensional subspace of the space of solutions will be denoted as $\{S^h\}_{h\in[0,1]}$, typically piecewise polynomials. The finite-dimensional approximation is found by solving

$$u^h: \quad a_{TE/TM}(u^h, \phi^h) = (f_{TE/TM}, \phi^h) \quad \text{for all } \phi^h \in S^h. \tag{2.39}$$

By choosing a basis $\{\phi_1, ...\phi_k\}$ of S^h the system can be written as a finite-dimensional algebraic problem. Solving this algebraic problem gives the desired approximation u^h. A vast number of convergence properties are known for this method, we will mention only the most important ones. The first one is for TE polarization:

There exists a constant $h_0 \in (0, 1]$ such that for any h, $0 < h < h_0$, the discretized problem (2.39) attains a unique solution u^h and

$$\|E_\alpha - u^h\|_{L^2(\Omega)} \leq Ch^2.$$

The second one is for TM polarization:

Assume that the equation (2.39) has unique solution. Let $f \in L^2(\Omega)$. For any given $\delta > 0$ there exists $h_1 = h_1(\delta)$ such such that for all $0 < h < h_1$,

$$\|H_\alpha - u^h\|_{H_1} \leq \delta\|f\|_{L^2(\Omega)}.$$

Uniqueness of inverse problem The inverse problem can be formulated as follows:

Given the incident wave $E_i = \exp(i\alpha y - i\beta z)$, with the angle of incidence $-\pi/2 < \vartheta_i < \pi/2$ and the diffracted fields $\mathbf{E}_r, \mathbf{E}_t$, *determine the grating profile.*

Much less is known here about the existence, uniqueness and stability, especially for TM case. In the TE case the system is described by the equation

$$\Delta E_x + k^2 E_x = 0,$$

with the b.c. (2.36). In general, the inverse problem is underdetermined. Despite of this underdetermination, there are some uniqueness results. Let us assume that two profiles a_1, a_2 are sufficiently smooth, and let us denote $T := \max\{a_1(y), a_2(y)\} - \min\{a_1(y), a_2(y)\}$.

Assume that $E_1(y, z_0) = E_2(y, z_0)$. Suppose that one of the following conditions is true:

- *k has a nonzero imaginary part*

- *k is real and T satisfies $k^2 < 2(T^{-2} + \Lambda^{-2})$.*

Then $a_1(y) = a_2(y)$.

If the first condition is true, than one have a global uniqueness results. However, if the second condition is true, then the result is only local — i.e. two grating profiles are identical if and only if they generate the same diffraction patterns and the area between them is "sufficiently small".

Much more important problem is the stability result.

The Hausdorff distance between two domains D_1, D_2 will be defined by

$$d(D_1, D_2) = \max\{\rho(D_1, D_2), \rho(D_2, D_1)\},$$

where

$$\rho(D_1, D_2) = \sup_{x \in D_1} \inf_{y \in D_2} |x - y|.$$

Let

$$D = \{(y, z)|a(y) < z < z_0\}, \quad \text{and} \quad D_h = \{(y, z)|a(y) + h\sigma_h(y)\mu(y) < z < d\},$$

for all $h < h_0$, where h_0 is a certain threshold and $\mu(y)$ is a normal to the profile $P \equiv \{z = a(y)\}$. Assume that the profile $a_h(y) := a(y) + h\sigma_h(y)\mu(y)$ is periodic with the period Λ and there exists $C > 0$ so that $|\sigma_h(y)| \leq C$ for all $y \in \mathbb{R}$. Suppose

$$C_1 h \leq d(D, D_h) < C_2 h \quad \text{for } h \text{ sufficiently small.}$$

Suppose E_x, E_{xh} being solutions of the scattering problems with the profiles S, S_h respectively. Then

$$d(D_h, D) \leq C\|u_h|_{z=d} - u|_{z=d}\|_{H^{1/2}},$$

where $H^{1/2}$ is a fractional Sobolev space.

Chapter 3

The Rigorous Coupled Wave Analysis

The Rigorous Coupled Wave Analysis (RCWA) was constructed for rectangular gratings. For such gratings, the system can be divided into three regions. The first is the superstrate medium, typically air or vacuum, the third is a substrate medium and the one between them is a grating region, see Fig. 3.1. In the grating region the relative permittivity is considered to be periodic and constant in z direction. The problem is solved separately in each region and these solutions are joined together using the interface conditions.
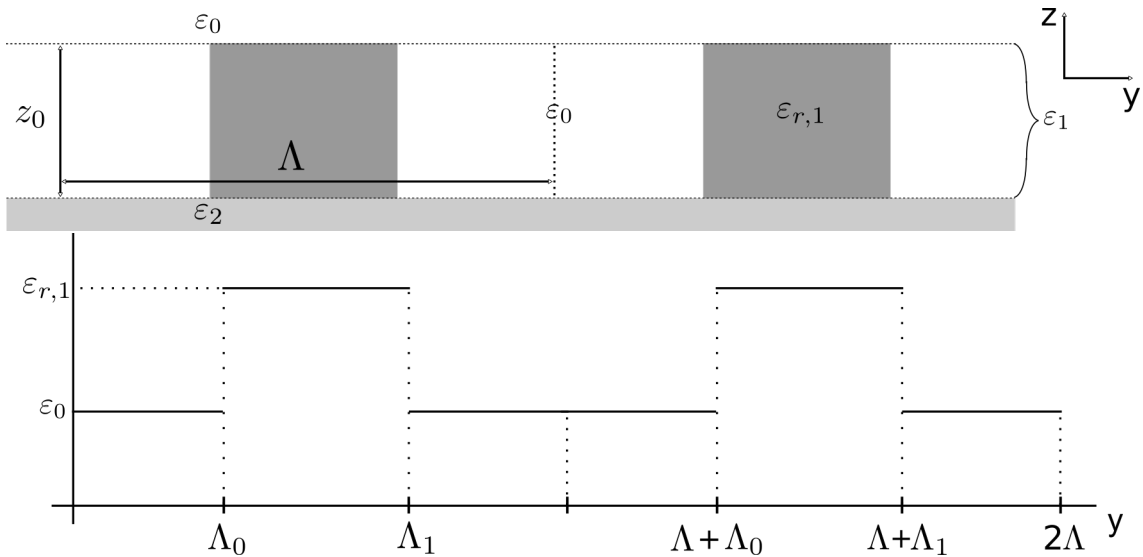

Figure 3.1: Rectangular grating — the checkerboard is a grating region with the permittivity ε_r

In the grating region the permittivity $\varepsilon_1 = \varepsilon_r$ is depending only on y and therefore the matrix $[\![\varepsilon_1]\!]$ is constant. The idea is to truncate the fields and approximate the solution of the infinite-dimensional ODE systems (2.24), (2.25) by a solution of a linear algebraic system. The interface conditions will be truncated as well in order to get a system of linear equations. Such systems can be solved easily by a suitable software package (in this Thesis it will be MATLAB 2014b). The procedure for TE polarization will be described in the first subsection, for TM polarization in the second subsection the following subsection will be devoted to explaining the problems originating from the truncation and the last section will contain a description of the algorithm modeling the optical response of curved gratings.

Remark 1. *Using the definition of Fourier series it can be found that the function ε_1 from the Fig. 3.1 can be written as*

$$\varepsilon_1 = \sum_{n=-\infty}^{\infty} \frac{i}{2\pi n} \left(\exp(-i\pi wn) - \exp(i\pi wn) \right) \left(\varepsilon_{r,1} - \varepsilon_0 \right) \exp(inqy), \tag{3.1}$$

where w is a fill factor, that means the ratio

$$w = \frac{\Lambda_1 - \Lambda_0}{\Lambda}.$$

3.1 Matrix formulation of problem in TE polarization

According to Rayleigh Theorem the total fields in the upper and lower semi-infinite regions can be written as

$$E_x(y,z) = E_{ix}(y,z) + E_{rx}(y,z) = A_i \exp(iq_0 y - is_{i,0}z) + \sum_{m=-\infty}^{\infty} A_{rx}^m \exp(iq_m y + is_{r,m}z), \quad z > z_0$$

$$E_x(y,z) = E_{tx}(y,z) = \sum_{m=-\infty}^{\infty} A_{tx}^m \exp(iq_m y - is_{t,m}z), \quad z < 0,$$

and in the grating region the field is

$$E_{1x}(y,z) = \sum_{n=-\infty}^{\infty} e_n(z) \exp(iq_n y),$$

where $e_n(z)$ is a solution of (2.24).

The system (2.24) is a linear system of ODE's and therefore it is convenient to make an Ansatz $e_n(z) := e_n^{(s)} \exp[isz]$. It leads to an infinite-dimensional algebraic eigenvalue problem

$$\left([\![\varepsilon_1]\!] - \mathbf{q}^2 \right) \mathbf{e}^{(s)} = s^2 \mathbf{e}^{(s)}, \tag{3.2}$$

where $\mathbf{q}^2 = \operatorname{diag}(q_m^2)$. The key step here is the truncation of the fields. By cutting the Rayleigh expansion of the fields from m_1 in the lower index and m_2 in the upper index, i.e.

$$E_x(y,z) = A_i \exp(iq_0 y - is_{i,0}z) + \sum_{m=m_1}^{m_2} A_{rx}^m \exp(iq_m y + is_{r,m}z), \quad z > z_0$$

$$E_{tx}(y,z) = \sum_{m=m_1}^{m_2} A_{tx}^m \exp(iq_m y - is_{t,m}z), \quad z < 0,$$

$$E_{1x}(y,z) = \sum_{m=m_1}^{m_2} e_m(z) \exp(iq_m y), \quad z \in (0, z_0),$$

$$\varepsilon_r(y) = \sum_{m=m_1}^{m_2} \varepsilon_m e^{imqy},$$

the equation (3.2) is reduced to a finite-dimensional eigenvalue problem

$$\mathbb{C}\boldsymbol{\nu} = \mu\boldsymbol{\nu},$$

with $\mathbb{C} = \left([\![\varepsilon]\!] - \mathbf{q}^2 \right)_{[m_1, m_2]}$ being a truncation of $[\![\varepsilon]\!] - \mathbf{q}^2$, $\mu \in \mathbb{R}$ and $\boldsymbol{\nu} = (e_{m_1}, ... e_{m_2}) \in \mathbb{R}^M$, where $M := m_2 - m_1 + 1$ is a total number of indices. Such eigenvalue equation has M *complex* solutions μ_j. The propagation numbers are the complex square roots of μ_j:

$$s_j^{\pm} = \pm \sqrt[c]{\mu_j},$$

where the plus sign corresponds to upward modes and minus sign corresponds to downward modes. Since $s_j^- = -s_j^+$ it is convenient to drop the superscript and relabel $s_j := s_j^+$, then $s_j^- = -s_j$. By introducing a matrix diagonalizer $\mathbb{G} = [\boldsymbol{\nu}_{m_1}, ..., \boldsymbol{\nu}_{m_2}]$ and a diagonal matrix $\boldsymbol{\mu} = \mathrm{diag}(\mu_{m_1}, ... \mu_{m_2})$ the matrix $\mathbb{C}$ can be written as

$$\mathbb{C} = \mathbb{G}\boldsymbol{\mu}\mathbb{G}^{-1}.$$

The so-called propagation matrices are defined by

$$\begin{aligned}
\mathbb{P}_z^+ &= \mathbb{P}_z = \mathbb{G}\exp(i\mathbf{s}z)\mathbb{G}^{-1}, \\
\mathbb{P}_z^- &= \mathbb{P}_z^{-1} = \mathbb{G}\exp(-i\mathbf{s}z)\mathbb{G}^{-1},
\end{aligned} \tag{3.3}$$

where $\mathbf{s} = \mathrm{diag}(s_{m_1}, ... s_{m_2})$. These matrices determine a transformation of the electric and magnetic field between planes with different z coordinate. The general solution of the equations in the grating region can be written as

$$\mathbf{e}_{1x}(z) = \left(\exp(i\sqrt{\mathbb{C}}z)\mathbf{A}_+ + \exp(-i\sqrt{\mathbb{C}}z)\mathbf{A}_-\right),$$

with the vectors $\mathbf{A}_+, \mathbf{A}_-$ which should be determined from the interface conditions. As can be seen from the form of the solution the field in the grating region is composed of two fields — upward and downward. The magnetic field intensity is found from the equation (2.13) as

$$H_{1y,m}(y, z) = -i\frac{\partial E_x}{\partial z} = \left(\sqrt{\mathbb{C}}\exp(i\sqrt{\mathbb{C}}z)\mathbf{A}_+ - \sqrt{\mathbb{C}}\exp(-i\sqrt{\mathbb{C}}z)\mathbf{A}_-\right)\exp(iq_m y).$$

The matrix $\sqrt{\mathbb{C}}$ is called "admittance matrix", is denoted as $\mathbb{Y}_1$ and can be computed from

$$\mathbb{Y}_1 := \sqrt{\mathbb{C}} = \mathbb{G}\mathbf{s}\mathbb{G}^{-1}. \tag{3.4}$$

Theoretically it could happen that the matrix $\mathbb{C}$ is not diagonalizable, in such a case it is necessary to replace the diagonal matrix $\boldsymbol{\mu}$ with the Jordan matrix. However, we did not hit this situation in our computations so far.

The final solution of the grating problem is made by matching the boundaries.

Matching the boundaries — infinitely deep grating Let the interface be in the line $z = 0$, and the grating region be in half-plane $z \le 0$. Such system can describe a grating with the depth significantly larger than the period. The vector of Fourier coefficients of the electric field of the incident wave is

$$\mathbf{A_i} = [\cdots, 0, A_i, 0, \cdots],$$

where the coefficient A_i is precisely at the $(m_2 - m_1)/2$ position of the M-sized column vector. We introduce vectors

$$\begin{aligned}
\mathbf{I} &= \mathbf{A_i}, \\
\mathbf{r} &= \mathbf{A}_r = [A_{rx}^{m_1}, ..., A_{rx}^{-1}, A_{rx}^0, A_{rx}^1, ..., A_{rx}^{m_2}], \\
\mathbf{t} &= \mathbf{A}_t = [A_{tx}^{m_1} ..., A_{tx}^{-1}, A_{tx}^0, A_{tx}^1, ..., A_{tx}^{m_2}].
\end{aligned}$$

There is only one interface and with the conditions

$$\begin{aligned}
E_{ix}|_{z\to0+} + E_{rx}|_{z\to0+} &= E_{tx}|_{z\to0-}, \\
H_{iy}|_{z\to0+} + H_{ry}|_{z\to0+} &= H_{tx}|_{z\to0-},
\end{aligned}$$

which in the matrix form is

$$\begin{aligned}
\mathbf{I} + \mathbf{r} &= \mathbf{t}, \\
(-\mathbb{Y}_0)\mathbf{I} + \mathbb{Y}_0\mathbf{r} &= (-\mathbb{Y}_1)\mathbf{t},
\end{aligned}$$

where $\mathbb{Y}_0 = \mathrm{diag}(\mathbf{s}_r)$. The reflection matrices and transmission matrices defined by a relation

$$\mathbf{r} = \widehat{\mathbb{R}}_{01}\mathbf{I}, \quad \mathbf{t} = \mathbb{T}_{01}\mathbf{I}$$

can be directly computed as

$$\widehat{\mathbb{R}}_{01} = (\mathbb{Y}_0 + \mathbb{Y}_1)^{-1}(\mathbb{Y}_0 - \mathbb{Y}_1).$$

and

$$\mathbb{T}_{01} = \mathbb{1} + \widehat{\mathbb{R}}_{01} = 2(\mathbb{Y}_0 + \mathbb{Y}_1)^{-1}\mathbb{Y}_0.$$

Once the vectors $\mathbf{r}, \mathbf{t}$ are known, the diffraction efficiencies can be found from (2.26).

Matching the boundaries — grating with the finite depth There are two interfaces which will be used to determine the unknown constants A_{n+}, A_{n-}, A_{tx}, A_{rx}. We introduce vectors

$$\mathbf{I} = \exp(-i\mathbf{s}_{i,0}z_0)\mathbf{A_i},$$
$$\mathbf{r} = \exp(i\mathbf{s}_r z_0)\mathbf{A}_r = \exp(i\mathbf{s}_r z_0)[A_{rx}^{m_1}, ..., A_{rx}^{-1}, A_{rx}^0, A_{rx}^1, ..., A_{rx}^{m_2}],$$
$$\mathbf{t} = \mathbf{A}_t = [A_{tx}^{m_1}..., A_{tx}^{-1}, A_{tx}^0, A_{tx}^1, ..., A_{tx}^{m_2}].$$

The shift in z is here for a better numerical behavior, the field above grating region is then

$$E_x(y, z) = I_0 \exp(iq_0 y - is_0(z - z_0)) + \sum_{m=m_1}^{m_2} r_m \exp(iq_m y + is_{r,m}(z - z_0)) \quad z > z_0.$$

At the interface $z = z_0$ the argument of the exponential function is independent of $s_{r,m}$, which ensures good numerical behavior even when there are some $s_{r,m}$ with a large negative complex part. The upward and downward fields are

$$E_{1x}^+(y,z) = \sum_{n=m_1}^{m_2}\sum_{m=m_1}^{m_2}\left(\exp(i\sqrt{\mathbb{C}}z)\right)_{nm} A_{m+}\exp(iq_n y) = \sum_{n=m_1}^{m_2}\sum_{m=m_1}^{m_2}\left(\mathbb{P}_z^+\right)_{nm} A_{m+}\exp(iq_n y),$$

$$E_{1x}^-(y,z) = \sum_{n=m_1}^{m_2}\sum_{m=m_1}^{m_2}\left(\exp(-i\sqrt{\mathbb{C}}(z - z_0))\right)_{nm} A_{m-}\exp(iq_n y) = \sum_{n=m_1}^{m_2}\sum_{m=m_1}^{m_2}\left(\mathbb{P}_{(z-z_0)}^-\right)_{nm} A_{m-}\exp(iq_n y).$$

Again, the shift at E_{1x}^- is due to a better numerical behavior. Let us also define admittance matrices $\mathbb{Y}_0 = \mathrm{diag}(\mathbf{s}_{r,m})$ and $\mathbb{Y}_1 = \mathbb{G}\mathbf{s}_+\mathbb{G}^{-1} = -\mathbb{G}\mathbf{s}_-\mathbb{G}^{-1}$, $\mathbb{Y}_2 = \mathrm{diag}(\mathbf{s}_2)$. Now the interface condition on the first interface is

$$E_{ix}|_{z\to z_0+} + E_{rx}|_{z\to z_0+} = E_{1x}^+|_{z\to z_0-} + E_{1x}^-|_{z\to z_0-},$$

which after the truncation and writing into the matrix form is

$$\mathbf{I} + \mathbf{r} = \mathbb{P}_{z_0}^+\mathbf{A}_+ + \mathbf{A}_-.$$

Since $\partial_\mathbf{n}E_x = \partial_z E_x = iH_y$, the second equation is for the H_y field

$$H_{iy}|_{z\to z_0+} + H_{ry}|_{z\to z_0+} = H_{1y}^+|_{z\to z_0-} + H_{1y}^-|_{z\to z_0-},$$

which in the matrix form is

$$\mathbb{Y}_0(\mathbf{I} - \mathbf{r}) = \mathbb{Y}_1(\mathbb{P}_{z_0}^+\mathbf{A}_+ - \mathbf{A}_-).$$

The interface conditions on the second interface are

$$E_{1x}^+|_{z\to0+} + E_{1x}^-|_{z\to0+} = E_{tx}|_{z\to0},$$
$$H_{1x}^+|_{z\to0+} + H_{1x}^-|_{z\to0+} = H_{tx}|_{z\to0},$$

and corresponding matrix form is

$$\mathbf{A}_+ + \mathbb{P}_{z_0}^+ \mathbf{A}_- = \mathbf{t}$$
$$\mathbb{Y}_1(\mathbf{A}_+ - \mathbb{P}_{z_0}^+ \mathbf{A}_-) = \mathbb{Y}_2\mathbf{t},$$

where $\mathbb{Y}_2 = \text{diag}(\mathbf{s}_t)$. We used the identity $\mathbb{P}_{z_0}^+ = \mathbb{P}_{-z_0}^-$. Let us recall the all fields are supposed to be truncated. In sum the matrix form of the interface conditions is

$$\mathbf{I} + \mathbf{r} = \mathbb{P}_{z_0}^+ \mathbf{A}_+ + \mathbf{A}_-.$$
$$\mathbb{Y}_0(\mathbf{I} - \mathbf{r}) = \mathbb{Y}_1(\mathbb{P}_{z_0}^+ \mathbf{A}_+ - \mathbf{A}_-),$$
$$\mathbf{A}_+ + \mathbb{P}_{z_0}^+ \mathbf{A}_- = \mathbf{t},$$
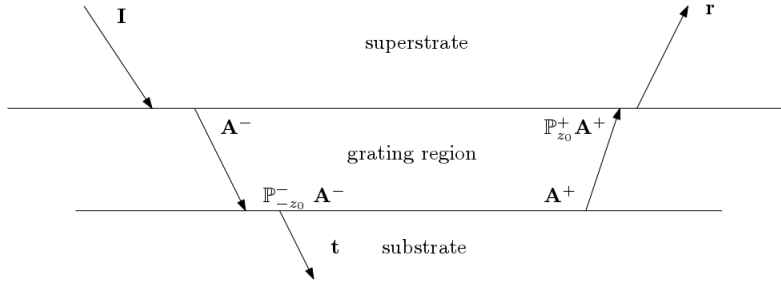$$\mathbb{Y}_1(\mathbf{A}_+ - \mathbb{P}_{z_0}^+ \mathbf{A}_-) = \mathbb{Y}_2\mathbf{t}.$$

Figure 3.2: Diffraction between two interfaces

These equations can be solved to obtain

$$\mathbf{r} = \left(\widehat{\mathbb{R}}_{01} + \mathbb{T}_{10}\mathbb{P}_{z_0}^+ \widehat{\mathbb{R}}_{12}\mathbb{P}_{z_0}^+(\mathbb{1} - \mathbb{Q})^{-1}\mathbb{T}_{01}\right)\mathbf{I},$$
$$\mathbf{t} = \mathbb{T}_{12}\mathbb{P}_{z_0}^+(\mathbb{1} - \mathbb{Q})^{-1}\mathbb{T}_{01}\mathbf{I}, \tag{3.5}$$

where

$$\widehat{\mathbb{R}}_{ij} = (\mathbb{Y}_i + \mathbb{Y}_j)^{-1}(\mathbb{Y}_i - \mathbb{Y}_j)$$
$$\mathbb{T}_{ij} = \mathbb{1} + \widehat{\mathbb{R}}_{ij}, \tag{3.6}$$
$$\mathbb{Q} = \widehat{\mathbb{R}}_{10}\mathbb{P}_{z_0}^+ \widehat{\mathbb{R}}_{12}\mathbb{P}_{z_0}^+.$$

3.2 Matrix formulation of problem in TM polarization

Assuming the dependence $H_x^{(s)}(y,z) = h_x^{(s)}(y)\exp[isz]$ and making the procedure similar to TE polarization leads to an algebraic equation

$$\sum_m [\![\varepsilon_1]\!]_{nm} \sum_p \left(\delta_{mp} - q_m q_p \left[\!\left[\frac{1}{\varepsilon_1}\right]\!\right]_{mp}\right) h_p = s^2 h_n, \tag{3.7}$$

where h_n are components of $h_x^{(s)}$ in the Fourier basis. Again, the fields will be truncated from m_1 to m_2, which produces the finite-dimensional eigenvalue problem. To find the remaining constants, the interface condition will be used. Unfortunately, the equation (3.7) exhibits a poor convergence as $m \to \infty$. Up to 1996 it was not known, whether this is a consequence of physical properties of the system or the problem is in mathematics. G. Granet & B. Guizal proposed in [8] to make a different factorization, namely to substitute $[\![\varepsilon_1]\!]$ by $[\![1/\varepsilon_1]\!]^{-1}$ and $[\![1/\varepsilon_1]\!]$ by $[\![\varepsilon_1]\!]^{-1}$. This substantially improved the convergence. The rigorous explanation appeared in [6] and will be described in the forthcoming subsection. After the correct factorization the system is in the form

$$\sum_{m=m_1}^{m_2} \left[\!\!\left[\frac{1}{\varepsilon_1}\right]\!\!\right]_{nm}^{-1} \sum_{p=m_1}^{m_2} \left(\delta_{mp} - q_m\, [\![\varepsilon_1]\!]_{mp}^{-1}\, q_p\right) h_p = s^2 h_n. \tag{3.8}$$

By defining the matrix $\mathbb{C} := [\![1/\varepsilon_1]\!]^{-1}\left(\mathbb{1} - \mathbf{q}\,[\![\varepsilon_1]\!]^{-1}\,\mathbf{q}\right)$ the equation (3.8) is rewritten simply as

$$\mathbb{C}\boldsymbol{\nu} = \mu\boldsymbol{\nu}.$$

This eigenvalue problem has M complex solutions μ_j, $j = \{m_1, \cdots, m_2\}$ and corresponding M eigenvectors. The propagation modes are $s_j^{\pm} := \pm\sqrt[c]{\mu_j}$.

Matching the boundaries — infinitely deep grating The continuity condition on the first interface is

$$H_{ix}|_{z \to z_0+} + H_{rx}|_{z \to z_0+} = H_{tx}|_{z \to 0}.$$

which in the Fourier space is

$$\mathbf{I} + \mathbf{r} = \mathbf{t},$$

and for E field it is

$$E_{iy}|_{z \to z_0+} + E_{ry}|_{z \to z_0+} = E_{ty}|_{z \to 0},$$

which in the Fourier space is

$$\widehat{\mathbb{Z}}_0(\mathbf{I} - \mathbf{r}) = \widehat{\mathbb{Z}}_1\mathbf{t}.$$

Procedure similar to TE case gives

$$\widehat{\mathbb{R}}_{01} = \left(\widehat{\mathbb{Z}}_0 + \widehat{\mathbb{Z}}_1\right)^{-1}\left(\widehat{\mathbb{Z}}_0 - \widehat{\mathbb{Z}}_1\right), \quad \mathbb{T}_{01} = 2\left(\widehat{\mathbb{Z}}_0 + \widehat{\mathbb{Z}}_1\right)^{-1}\widehat{\mathbb{Z}}_0,$$

where

$$\widehat{\mathbb{Z}}_0 = \left[\!\!\left[\frac{1}{\varepsilon_0}\right]\!\!\right]\mathbb{G}\,\mathrm{diag}(\mathbf{s}_r)\mathbb{G}^{-1}, \quad \widehat{\mathbb{Z}}_1 = \left[\!\!\left[\frac{1}{\varepsilon_1}\right]\!\!\right]\mathbb{G}\,\mathrm{diag}(\mathbf{s})\mathbb{G}^{-1}. \tag{3.9}$$

Matching the boundaries — grating with the finite depth The interface conditions for the first interface are

$$H_{ix}|_{z \to z_0+} + H_{rx}|_{z \to z_0+} = H_{1x}^+|_{z \to z_0-} + H_{1x}^-|_{z \to z_0-},$$
$$E_{iy}|_{z \to z_0+} + E_{ry}|_{z \to z_0+} = E_{1y}^+|_{z \to z_0-} + E_{1y}^-|_{z \to z_0-},$$

which in the Fourier space is

$$\mathbf{I} + \mathbf{r} = \mathbb{P}_{z_0}^+\mathbf{A}_+ + \mathbf{A}_-,$$
$$\widehat{\mathbb{Z}}_0(\mathbf{I} - \mathbf{r}) = \widehat{\mathbb{Z}}_1(\mathbb{P}_{z_0}^+\mathbf{A}_+ - \mathbf{A}_-),$$

where $\mathbb{P}_{z_0}^+ = \exp(i\sqrt{\mathbb{C}}z_0) = \mathbb{G}\exp(i\mathbf{s}z_0)\mathbb{G}^{-1}$. The interface conditions on the second interface are

$$H_{1x}^+|_{z\to 0+} + H_{1x}^-|_{z\to 0+} = H_{tx}|_{z\to 0},$$
$$E_{1y}^+|_{z\to 0+} + E_{1y}^-|_{z\to 0+} = E_{ty}|_{z\to 0},$$

and the matrix form is

$$\mathbf{A}_+ + \mathbb{P}_{z_0}^+\mathbf{A}_- = \mathbf{t},$$
$$\widehat{\mathbb{Z}}_1(\mathbf{A}_+ - \mathbb{P}_{z_0}^+\mathbf{A}_-) = \widehat{\mathbb{Z}}_2\mathbf{t},$$

where $\widehat{\mathbb{Z}}_2 = [\![1/\varepsilon_2]\!]\mathbb{G}\,\mathrm{diag}(\mathbf{s}_t)\mathbb{G}^{-1}$. The vectors $\mathbf{r}, \mathbf{t}$ can obtained from (3.5), where now

$$\widehat{\mathbb{R}}_{ij} = \left(\widehat{\mathbb{Z}}_i + \widehat{\mathbb{Z}}_j\right)^{-1}\left(\widehat{\mathbb{Z}}_i - \widehat{\mathbb{Z}}_j\right),$$
$$\mathbb{T}_{ij} = 2\left(\widehat{\mathbb{Z}}_i + \widehat{\mathbb{Z}}_j\right)^{-1}\widehat{\mathbb{Z}}_i, \quad i,j \in \{1,2\}. \tag{3.10}$$

3.2.1 Staircase approximation

The main idea of the staircase approximation used to model curved gratings was already described in Section 2.5. The grating region is sliced into N layers, each with the width z_k/N, $k = 1,...N$. In every slice the permittivity function is constant in z and staircase-like in y, see Fig. 2.2 and 3.1. The matrix $[\![\varepsilon_1]\!]$ is constructed from the expansion (3.1). The fill factor in a given point y_k is found from the relation $y_k = w(z_k)$, where w is the inverse function of the profile and z_k is a given nodal point, for equidistant points it is $z_k = (2k-1)/(2N)z_0$, with $k = 1,...N$.

The equation describing propagation in each slice is (3.2) for TE polarization and (3.8) for TM polarization. The propagation matrix $\mathbb{P}_j$ and the impedance and admittance matrices $\mathbb{Y}_i$ or $\widehat{\mathbb{Z}}_i$ for each slice can be found from (3.3) and (3.4) for TE polarization, (3.9) for TM polarization. The reflection and transmission matrices on each interface are found from (3.6) for TE and (3.10) for TM polarization. Finally, the reflection matrix is found by the iteration using the Airy-like reflection series [12]

$$\widehat{\mathbb{R}}_{0,j+1} = \widehat{\mathbb{R}}_{0,j} + \mathbb{T}_{j,0}\mathbb{P}_j\widehat{\mathbb{R}}_{j,j+1}\mathbb{P}_j(1 - \mathbb{Q}_j)^{-1}\mathbb{T}_{0,j},$$
$$\widehat{\mathbb{R}}_{j+1,0} = \widehat{\mathbb{R}}_{j+1,j} + \mathbb{T}_{j,j+1}\mathbb{P}_j\widehat{\mathbb{R}}_{j,0}\mathbb{P}_j(1 - \tilde{\mathbb{Q}}_j)^{-1}\mathbb{T}_{j+1,j},$$
$$\mathbb{T}_{0,j+1} = \mathbb{T}_{j,j+1}\mathbb{P}_j(1 - \mathbb{Q}_j)T_{0,j},$$
$$\mathbb{T}_{j+1,0} = \mathbb{T}_{j,0}\mathbb{P}_j(1 - \tilde{\mathbb{Q}}_j)T_{j+1,j}.$$

where

$$\mathbb{Q}_j = \widehat{\mathbb{R}}_{j,0}\mathbb{P}_j\widehat{\mathbb{R}}_{j,j+1}\mathbb{P}_j, \quad \tilde{\mathbb{Q}}_j = \widehat{\mathbb{R}}_{j,j+1}\mathbb{P}_j\widehat{\mathbb{R}}_{j,0}\mathbb{P}_j$$

Using the identities $\mathbf{r} = \widehat{\mathbb{R}}_{0,N+1}\mathbf{I}$, $\mathbf{t} = \mathbb{T}_{0,N+1}\mathbf{I}$ it is possible to calculate the diffraction efficiencies from (2.26).

3.3 Discretization and convergence

This subsection is devoted to a discussion about issues connected with the truncation. The main questions which arise from the truncation are the following — whether the truncation of the matrices is correct, whether the eigenvalues of the truncated matrix converge to the eigenvalues of the original problem and whether the truncated interface conditions converge to the infinite-dimensional ones. And finally, whether there arise other problems resulting from the discretization. The first two questions will be partially answered, but the others are open so far. A lack of rigorous results regarding the rate of convergence of discretized equations is still a crucial problem of the RCWA and the C-Method.

Let us remind that $\mathbb{C} := [\![\varepsilon_1]\!] - \mathrm{diag}(\mathbf{q})^2$. The first question is if the eigenvalues of the truncated equation converge as the truncation numbers go to infinity.

3.3.1 Convergence of eigenvalues

Theorem 1 (Poincare–von Koch)**.** *For the determinant of the matrix $A = (a_{ij})$ and all of its minors to be absolutely convergent, it is sufficient that*

$$\sum_{i,k} |a_{ik}| < \infty.$$

This theorem is applicable for volume gratings, where $\hat{\varepsilon}_n \leq O(1/n^2)$. Unfortunately, it is insufficient for surface-relief grating, because $\varepsilon_n \leq O(1/n)$. Here it is necessary to find a better alternative.

Let us suppose that the matrix elements a_{ik} are functions of a parameter τ in a domain T in the complex plane.

Theorem 2 (von Koch (1892))**.** *For the determinant of $A(\tau)$ and all of its minors to be absolutely and uniformly convergent in a domain T, it is sufficient that there exists a sequence of nonzero numbers, $\{x_i\}$, such that*

$$\sum_{i,k} \left| a_{i,k}(\tau) \frac{x_i}{x_k} \right|$$

is uniformly convergent in T. If the above condition is satisfied, the determinant remains absolutely and uniformly convergent when a row or a column of $A(\tau)$ is replaced by a sequence of bounded numbers.

But there is much more in the von Koch work. He showed that many results of classical linear algebra remain unchanged. For instance, Laplace expansion, Cramer's rule and necessary and sufficient condition for a homogeneous system to have a nonzero solution is the zero determinant. This theorem can be directly applied to the grating problem in TE polarization. The eigenvalue equation is (3.2), which can be in the Fourier space rewritten as

$$\sum_{m} \left(\delta_{mn} + \frac{\tilde{\varepsilon}_{n-m}}{\hat{\varepsilon}_0 - q_n^2 - s} \right) E_{zm} = 0, \tag{3.11}$$

where $\tilde{\varepsilon}_{n-m} = \hat{\varepsilon}_{m-n}$ if $n \neq m$, $\tilde{\varepsilon}_0 = 0$ and $\hat{\varepsilon}_0$ is a zero coefficient in the Fourier series expansion of ε_1 and there is also assumed $s \neq \hat{\varepsilon}_0 - q_n^2$. According to [6], the eigenvalues and eigenvectors of truncated equation (3.11) converge to eigenvalues and eigenvectors of the full problem (3.2). The convergence is not uniform, and this non-uniformity probably plays an important role in the overall convergence of the RCWA and the C-Method.

Unfortunately, there is not a rigorous proof of this statement for the case of TM polarization, i.e. for the equation (3.7), but in particular problems a good coincidence with physical measurement and the approximations obtained by a use of FEM suggest the convergence of this method.

The intuitive possibility how to truncate the system is to take $N \in \mathbb{N}$ and truncate $m = -M, ..., M$. The approximation obtained from such truncated system are not accurate enough when the incident angle is near grazing. Therefore it is necessary to do it in a different way. The correct truncation is $m = -m_1, ..., m_2$, where $m_1 = -[q_0 \Lambda_0] - [M/2]$, $m_2 = -[q_0 \Lambda_0] + [M/2]$ and M is an order of truncation. In most cases, this reduces to a standard bounds $m_1 = -m_2 = M/2$, but if the incidence angle is large, then m_1 is different from $-m_2$.

3.3.2 Li factorization rules

In 1996 published Li the paper [6], where he derived the rules for a convergence of a product of two Fourier series. This paper has significantly influenced the development of the RCWA. In this chapter we will post his three factorization theorems and explain theirs meaning in more details.

Definition 1. *We let $\mathcal{P}$ denote the set of 2π-periodic real valued functions which are square integrable and piecewise C^2-continuous. The set of abscissae of discontinuities of f is defined by*

$$U_f := \{x_i | f(x_i+) \neq f(x_i-), \ x_i \in [0, 2\pi]\}.$$

Let $f, g \in \mathcal{P}$, and $U_{f,g} = U_f \cap U_g$. The functions f, g have complementary discontinuities if and only if $h(x) := f(x)g(x)$ is continuous in all points $x \in U_{f,g}$. Otherwise we say that f, g have concurrent discontinuities.

Definition 2. *The Fourier series of $h = fg$ with $2M + 1$ harmonics is*

$$h^{(M)}(x) := \sum_{n=-M}^{M} h^{(n)} \exp(inx),$$

with the Fourier coefficients

$$h^{(n)} := \sum_{m=-\infty}^{\infty} f_{n-m} g_m.$$

Let us remind the product rule, which says that the value of function h at a point $x \in [0, 2\pi]$ can be written as

$$h(x) = \sum_{n=-\infty}^{\infty} h_n \exp(inx) = f(x)g(x) = \sum_{n=-\infty}^{\infty} h^{(n)} \exp(inx) = \sum_{n=-\infty}^{\infty} f_{n-m} g_m \exp(inx),$$

where h_n are Fourier coefficients of h.

Theorem 3. *Let $f, g \in \mathcal{P}$, $h = fg$ and f and g have no concurrent discontinuities. Then the truncated Laurent Fourier series $h^{(M)}(x)$ converges, i.e.*

$$\lim_{M \to \infty} h^{(M)}(x) = h(x).$$

Theorem 4. *Let $f, g \in \mathcal{P}$, $h = fg$. Let discontinuities of f and g be complementary. Additionally let $f(x) \neq 0$ for all $x \in [0, 2\pi)$. If f satisfies either one of the following conditions*

- *$Re(1/f)$ does not change sign in $[0, 2\pi)$ and $\inf_{x \in [0,2\pi)} |Re(1/f(x))| > 0$,*

- *$Im(1/f)$ does not change sign in $[0, 2\pi)$ and $\inf_{x \in [0,2\pi)} |Im(1/f(x))| > 0$,*

then the truncated Inverse Fourier series $\tilde{h}^{(M)}$ converges, i.e.

$$\lim_{M \to \infty} \tilde{h}^{(M)}(x) = \lim_{M \to \infty} \sum_{m=-M}^{M} \sum_{n=-M}^{M} \frac{1}{\tilde{f}_{n-m}} g_m \exp(inx) = h(x),$$

where $\tilde{f}_m$ are the Fourier coefficients of the function f^{-1}.

Let $x_i \in U_{f,g}$. We denote $\hat{f}_i := f(x_i + 0) - f(x_i - 0)$.

Theorem 5. *Let $f, g \in \mathcal{P}$ have concurrent discontinuities, $h = fg$. Then the truncated Laurent Fourier series $h^{(M)}(x)$ has the following error behaviour*

$$h^{(M)}(x) = h_M(x) - \sum_{x_i \in U_{f,g}} \frac{\hat{f}_i \hat{g}_i}{2\pi^2} \Phi_M(x - x_i) - o(1),$$

where $o(1)$ uniformly tends to zero for $M \to \infty$ and

$$\Phi_M(x) := \sum_{i=1}^{M} \frac{\cos(nz)}{n} \sum_{|m| > M} \frac{1}{m - n}.$$

The function Φ_M satisfies

$$\lim_{M \to \infty} = \begin{cases} 0 & \text{if } x \neq 0 \\ \frac{\pi^2}{4} & \text{if } x = 0. \end{cases}$$

3.4 Application of Li factorization rules

The three above stated theorems allows us to divide the product of two periodic function into two groups:

1. A product of two functions with no concurrent discontinuities

2. A product of two functions with complementary discontinuities

3. A product of two functions with concurrent but non-complementary discontinuities

The first type is represented e.g. by the functions E_x and ε in TE polarization. E_x component of the electric field is continuous, and ε is discontinuous, and the application of Theorem 3 yields that it is possible to truncate the product of ε and E_x.

The second type is represented by E_y and ε. The product of these two functions is $\varepsilon E_y := D_y$, and D_y is always continuous by Faraday's Law.

The third type can be represented by $\varepsilon(E_z + E_y)$. However, it seems that every product which is relevant in grating problems can be decomposed into a sum of products which can be factorized by first or second rule.

For illustrative purposes, let $[\,\cdot\,]$ denote a vector of the Fourier components of a function, and $[\![\,\cdot\,]\!]$ the Toeplitz matrix generated by this function. Then the first and second product can be written

$$[\varepsilon E_x] = [\![\varepsilon]\!][E_x],$$

$$[\varepsilon E_y] = \left[\!\!\left[\frac{1}{\varepsilon}\right]\!\!\right]^{-1} [E_y].$$

3.5 First order methods for TM polarization (the Normal Vector Method)

In the case of TM polarization and highly conducting grating it can be better or even necessary to avoid the rewriting the Maxwell equations as a second-order equation (as it was done in Section 2.3) and solve the problem (2.18), (2.19) as a first order equation. Suppose the form of the solution in the grating region as

$$H_x = \sum_{m=-\infty}^{\infty} h_m(z) \exp(iq_m y),$$

$$E_y = \sum_{m=-\infty}^{\infty} e_m(z) \exp(iq_m y).$$

Assuming $e_m(z), h_m(z) \sim \exp(isz)$ and substitution into (2.18)–(2.20) leads to

$$is\mathbf{h}_x = -i\mathbf{d}_y, \tag{3.12}$$

$$i\mathbf{q}\mathbf{h}_x = i\mathbf{d}_z, \tag{3.13}$$

$$s\mathbf{e}_y = i\mathbf{q}\mathbf{e}_z - i\mathbf{h}_x, \tag{3.14}$$

where $\mathbf{q} = \operatorname{diag}(q_m)$ and $\mathbf{d}_{y/z} = [\![\varepsilon_r]\!]\mathbf{e}_{y/z}$. The classical staircase approximation now leads to an infinite-dimensional algebraic problem

$$s \begin{pmatrix} \mathbf{e}_y \\ \mathbf{h}_x \end{pmatrix} = \begin{pmatrix} 0 & \mathbb{1} - \mathbf{q}[\![\varepsilon_r^{-1}]\!]\mathbf{q} \\ [\![\varepsilon_r]\!] & 0 \end{pmatrix} \begin{pmatrix} \mathbf{e}_y \\ \mathbf{h}_x \end{pmatrix}.$$

This formulation is for gratings with non-staircase profile correct as long as the system is infinite-dimensional. According to previous Section it is not possible to truncate this system, because we used

Laurent rule for the components $\mathbf{D}_y, \mathbf{D}_z$, which in general are discontinuous. Hence, the equations must be derived in a different way. We will follow the derivation of the eigenvalue problem from [9], [18], and to match the interface condition we will use our own algorithm using the Airy-like series.

The basic idea is to rewrite the equations (3.12)–(3.14) using the tangent E_t and normal D_n components of the fields which are continuous according to Faraday's law instead of the E_y, E_z or D_y, D_Z components, which in general can be discontinuous. Let $\phi(y)$ be the normal to the lamellar grating boundary in the $y - z$ plane. Then

$$E_n = E_y \cos(\phi(y)) - E_z \sin(\phi(y))$$
$$E_t = E_y \sin(\phi(y)) + E_z \cos(\phi(y)),$$

the symbol E_t denotes the component of electric field tangent to the boundary, E_n the normal component. Assume that the periodic function $a(y)$ determining the profile is continuously differentiable. Then

$$\cos(\phi(y)) = \frac{1}{\sqrt{1 + [a'(y)]^2}} = \sum_{n=-\infty}^{\infty} c_m \exp(iq_m y),$$
$$\sin(\phi(y)) = \frac{a'(y)}{\sqrt{1 + [a'(y)]^2}} = \sum_{n=-\infty}^{\infty} s_m \exp(iq_m y).$$

The assumption of continuous differentiability can be relaxed to a continuity of $\sin(y), \cos(y)$ at the points y for which $\mathbf{E}_y, \mathbf{E}_z$ are discontinuous, i.e. for the points on the grating surface. For more details see Appendices in [7].

The factorization of Fourier components e_{nm} and e_{tm} of $\mathbf{E}_n, \mathbf{E}_t$ gives

$$(\mathbf{e}_n)_m = \sum_j \left([\![c]\!]_{mj}(\mathbf{e}_y)_j - [\![s]\!]_{mj}(\mathbf{e}_z)_j \right),$$
$$(\mathbf{e}_t)_m = \sum_j \left([\![s]\!]_{mj}(\mathbf{e}_y)_j + [\![c]\!]_{mj}(\mathbf{e}_z)_j \right),$$

where $[\![s]\!]_{mj}$, $[\![c]\!]_{mj}$ are Toeplitz matrices of Fourier components of $\sin(\phi(y))$, $\cos(\phi(y))$. The components $\mathbf{d}_y, \mathbf{d}_z$ of the electric displacement are transformed similarly

$$(\mathbf{d}_y)_m = \sum_j \left([\![c]\!]_{mj}(\mathbf{d}_n)_j + [\![s]\!]_{mj}(\mathbf{d}_t)_j \right),$$
$$(\mathbf{d}_z)_m = \sum_j \left(-[\![s]\!]_{mj}(\mathbf{d}_n)_j + [\![c]\!]_{mj}(\mathbf{d}_t)_j \right).$$

The fields $\mathbf{E}_t$ and $\mathbf{D}_n$ are continuous, hence, it is possible to use the first and the second Li rule:

$$\mathbf{d}_t = [\![\varepsilon_r]\!]\mathbf{e}_t,$$
$$\mathbf{d}_n = \left[\!\left[\frac{1}{\varepsilon_r}\right]\!\right]^{-1} \mathbf{e}_n.$$

Using these expansions (3.12) can be rewritten as

$$i s \mathbf{h}_x = -i([\![c]\!]\mathbf{d}_n + [\![s]\!]\mathbf{d}_t) = -i \left([\![c]\!] \left[\!\left[\frac{1}{\varepsilon_r}\right]\!\right]^{-1} \mathbf{e}_n + [\![s]\!][\![\varepsilon_r]\!]\mathbf{e}_t \right) =$$

$$= -i \left([\![c]\!] \left[\!\left[\frac{1}{\varepsilon_r}\right]\!\right]^{-1} ([\![c]\!]\mathbf{e}_y - [\![s]\!]\mathbf{e}_z) + [\![s]\!][\![\varepsilon_r]\!]([\![s]\!]\mathbf{e}_y + [\![c]\!]\mathbf{e}_z) \right),$$

and similarly for (3.13). For clarity we introduce the matrices

$$[\![A]\!] = [\![c]\!] \left[\!\!\left[\frac{1}{\varepsilon_r}\right]\!\!\right]^{-1} [\![c]\!] + [\![s]\!][\![\varepsilon_r]\!][\![s]\!]$$

$$[\![B]\!] = [\![s]\!][\![\varepsilon_r]\!][\![s]\!] - [\![c]\!] \left[\!\!\left[\frac{1}{\varepsilon_r}\right]\!\!\right]^{-1} [\![c]\!]$$

$$[\![C]\!] = [\![c]\!][\![\varepsilon_r]\!][\![s]\!] + [\![s]\!] \left[\!\!\left[\frac{1}{\varepsilon_r}\right]\!\!\right]^{-1} [\![c]\!]$$

$$[\![F]\!] = [\![s]\!] \left[\!\!\left[\frac{1}{\varepsilon_r}\right]\!\!\right]^{-1} [\![s]\!] + [\![c]\!][\![\varepsilon_r]\!][\![c]\!]$$

and write (3.12), (3.13) as

$$s(\mathbf{h}_x)_m = \sum_j ([\![A]\!]_{mj}(\mathbf{e}_y)_j + [\![B]\!]_{mj}(\mathbf{e}_z)_j)$$

$$q_m(\mathbf{h}_x)_m = \sum_j ([\![C]\!]_{mj}(\mathbf{e}_y)_j + [\![F]\!]_{mj}(\mathbf{e}_z)_j).$$

Insertion of (3.13) into (3.14) finally yields an eigenvalue problem which produces a system of eigenvalue equations

$$\mathbb{M} \begin{pmatrix} \mathbf{e}_y \\ \mathbf{h}_x \end{pmatrix} = s \begin{pmatrix} \mathbf{e}_y \\ \mathbf{h}_x \end{pmatrix},$$

with the matrix

$$\mathbb{M} = \begin{pmatrix} -\mathbf{q}^2 [\![F]\!]^{-1} [\![C]\!] & \mathbb{1} - \mathbf{q}^2 [\![F]\!]^{-1} \mathbf{q}^2 \\ [\![A]\!] - [\![B]\!][\![F]\!]^{-1}[\![C]\!] & -[\![B]\!][\![F]\!]^{-1}\mathbf{q}^2 \end{pmatrix}. \tag{3.15}$$

This matrix can be now truncated to get a finite-dimensional eigenvalue problem. In the plane $z = 0$ the column vector $\mathbf{f}_0 = [\mathbf{e}_{0y}, \mathbf{h}_{0x}]$ can be written as

$$\mathbf{f}_0 = \sum_{j=0}^{2n_{\mathrm{Dim}}} g_{0j}\boldsymbol{\nu}_j = \mathbb{G}\mathbf{g}_0,$$

where $\mathbb{G}$ is a matrix composed of the eigenvectors of $\mathbb{M}$. The propagation the z-direction is described by

$$\mathbf{f}(z) = \sum_{n=0}^{2n_{\mathrm{Dim}}} g_{0j}\boldsymbol{\nu}_j \exp(is_j z) = \mathbb{G}\exp(i\mathbf{s}z)\mathbb{G}^{-1}, \tag{3.16}$$

where $\mathbf{s} = (s_1, \cdots, s_{2n_{\mathrm{Dim}}})$ is a vector of eigenvalues. The unknown coefficients have to be determined from the interface conditions. There are several ways how to treat this problem, the most popular are T- or S-matrix algorithms. We use a generalization of the Airy-like series, which is based on the S-matrix algorithm. The basic idea can be found e.g. in [13].

The equations in TE polarization remains unchanged and it is possible to solve the whole problem using the second-order method described before.

3.5.1 Matching the boundaries and the Airy-like series

The first step is in computing the propagation matrices for every layer and reflection and transmission matrices for every interface. The eigenvalues s_n will be divided into two sets — s^+, s^-. The set s^+

contains the positive real eigenvalues and the complex eigenvalues with the positive imaginary part. The set s^- contains the negative real eigenvalues and the complex eigenvalues with the negative imaginary part. Both sets have the same number of elements. The eigenvectors are divided into two sets ν^+, ν^- as well. The positive propagation modes can be then expressed as

$$\begin{pmatrix} \mathbf{e}_y^+ \\ \mathbf{h}_x^+ \end{pmatrix} = \begin{pmatrix} \mathbb{G}_e^+ & \mathbb{G}_e^- \\ \mathbb{G}_h^+ & \mathbb{G}_h^- \end{pmatrix} \begin{pmatrix} \mathbf{g}^+ \\ 0 \end{pmatrix},$$

where $\mathbb{G}_e^{\pm}$ and $\mathbb{G}_h^{\pm}$ are matrices generated from the eigenvectors respective to eigenvalues from s^+, s^-. It can be proceed similarly for the negative modes $\mathbf{g}^-$. The propagation matrices in positive and negative directions can be found using (3.16) as

$$\mathbb{P}_z^+ = \mathbb{G}_h^+ \exp(i\mathbf{s}_+ z)(\mathbb{G}_h^+)^{-1}, \quad \mathbb{P}_z^- = \mathbb{G}_h^- \exp(i\mathbf{s}_- z)(\mathbb{G}_h^-)^{-1}.$$

Note the plus sign in the definition of $\mathbb{P}_z^-$. The the impedance matrices are computed from

$$\mathbf{e}_y^+ = \mathbb{G}_e^+ \mathbf{g}^+ = \mathbb{G}_e^+ (\mathbb{G}_h^+)^{-1} \mathbf{h}_x^+ = \widehat{\mathbb{Z}}^+ \mathbf{h}_x^+, \quad \mathbf{e}_y^- = \mathbb{G}_e^- (\mathbb{G}_h^-)^{-1} \mathbf{h}_x^- = \widehat{\mathbb{Z}}^- \mathbf{h}_x^-,$$

i.e.

$$\widehat{\mathbb{Z}}^+ = \mathbb{G}_e^+ (\mathbb{G}_h^+)^{-1}, \quad \widehat{\mathbb{Z}}^- = \mathbb{G}_e^- (\mathbb{G}_h^-)^{-1}.$$

The reflection and transmission matrices on the interface between the environment i and $i+1$ can be found similarly as above. The continuity on the interface means

$$H_{ix} + H_{rx} = H_{tx},$$
$$E_{iy} + E_{ry} = E_{ty},$$

and in the Fourier space it is

$$\mathbf{I} + \mathbf{r} = \mathbf{t},$$
$$\widehat{\mathbb{Z}}_i^- \mathbf{I} + \widehat{\mathbb{Z}}_i^+ \mathbf{r} = \widehat{\mathbb{Z}}_{i+1}^- \mathbf{t},$$

and for the interface between $i+1$ and i it is

$$\mathbf{I} + \mathbf{r} = \mathbf{t},$$
$$\widehat{\mathbb{Z}}_{i+1}^+ \mathbf{I} + \widehat{\mathbb{Z}}_{i+1}^- \mathbf{r} = \widehat{\mathbb{Z}}_i^+ \mathbf{t},$$

yielding the result

$$\widehat{\mathbb{R}}_{j,j+1} = (\widehat{\mathbb{Z}}_{j+1}^- - \widehat{\mathbb{Z}}_j^+)^{-1}(\widehat{\mathbb{Z}}_j^- - \widehat{\mathbb{Z}}_{j+1}^-), \quad \mathbb{T}_{j,j+1} = (\widehat{\mathbb{Z}}_{j+1}^- - \widehat{\mathbb{Z}}_j^+)^{-1}(\widehat{\mathbb{Z}}_j^- - \widehat{\mathbb{Z}}_j^+),$$
$$\widehat{\mathbb{R}}_{j+1,j} = (\widehat{\mathbb{Z}}_j^+ - \widehat{\mathbb{Z}}_{j+1}^-)^{-1}(\widehat{\mathbb{Z}}_{j+1}^+ - \widehat{\mathbb{Z}}_j^+), \quad \mathbb{T}_{j+1,j} = (\widehat{\mathbb{Z}}_j^+ - \widehat{\mathbb{Z}}_{j+1}^-)^{-1}(\widehat{\mathbb{Z}}_{j+1}^+ - \widehat{\mathbb{Z}}_{j+1}^-).$$

It is easy to verify that for isotropic environment with $\widehat{\mathbb{Z}}^+ = -\widehat{\mathbb{Z}}^-$ this reduces to (3.10). The reflection matrix between the first and second interface will be found from the interface conditions

$$\mathbf{I} + \mathbf{r} = \mathbb{P}^+ \mathbf{A}_+ + \mathbf{A}_-,$$
$$\widehat{\mathbb{Z}}_0^- \mathbf{I} + \widehat{\mathbb{Z}}_0^+ \mathbf{r} = \widehat{\mathbb{Z}}_1^+ \mathbb{P}^+ \mathbf{A}_+ + \widehat{\mathbb{Z}}_1^- \mathbf{A}_-,$$
$$\mathbf{A}_+ + \mathbb{P}^- \mathbf{A}_- = \mathbf{t},$$
$$\widehat{\mathbb{Z}}_1^+ \mathbf{A}_+ + \widehat{\mathbb{Z}}_1^- \mathbb{P}^- \mathbf{A}_- = \widehat{\mathbb{Z}}_2^- \mathbf{t},$$

with $\mathbb{P}^+ := \mathbb{P}_{\pm z_0}^{\pm}$ to make the formulae more clear. Using the third and the fourth equation it is possible to find

$$\mathbf{A}_+ = \widehat{\mathbb{R}}_{12} \mathbb{P}^- \mathbf{A}_-.$$

Inserting this into the first and the second equation, eliminating $\mathbf{A}_-$ and expressing it in the form $\mathbf{I} = \widehat{\mathbb{R}}_{02}\mathbf{r}$ gives the result

$$\widehat{\mathbb{R}}_{02} = \widehat{\mathbb{R}}_{01} + \mathbb{T}_{10}\mathbb{P}^+\widehat{\mathbb{R}}_{12}\mathbb{P}^-(1 - \widehat{\mathbb{R}}_{10}\mathbb{P}^+\widehat{\mathbb{R}}_{12}\mathbb{P}^-)^{-1}\mathbb{T}_{01}.$$

The transmission matrix can be then found as

$$\mathbb{T}_{02} = \mathbb{T}_{12}\mathbb{P}^-(1 - \widehat{\mathbb{R}}_{10}\mathbb{P}^+\widehat{\mathbb{R}}_{12}\mathbb{P}^-)^{-1}\mathbb{T}_{01}.$$

The upward matrices $\widehat{\mathbb{R}}_{20}$ and T_{20} are computed analogously. The reflection and transmission matrices between layers 0 and $J + 1$, $J \geq 1$ are found interatively using the Airy-like series:

$$\widehat{\mathbb{R}}_{0,j+1} = \widehat{\mathbb{R}}_{0,j} + \mathbb{T}_{j,0}\mathbb{P}_j^+\widehat{\mathbb{R}}_{j,j+1}\mathbb{P}_j^-(1 - \mathbb{Q}_j^+)^{-1}\mathbb{T}_{0,j},$$
$$\widehat{\mathbb{R}}_{j+1,0} = \widehat{\mathbb{R}}_{j+1,j} + \mathbb{T}_{j,j+1}\mathbb{P}_j^-\widehat{\mathbb{R}}_{j,0}\mathbb{P}_j^+(1 - \mathbb{Q}_j^-)^{-1}\mathbb{T}_{j+1,j},$$
$$\mathbb{T}_{0,j+1} = \mathbb{T}_{j,j+1}\mathbb{P}_j^-(1 - \mathbb{Q}_j^+)^{-1}\mathbb{T}_{0,j},$$
$$\mathbb{T}_{j+1,0} = \mathbb{T}_{j,0}\mathbb{P}_j^+(1 - \mathbb{Q}_j^-)^{-1}\mathbb{T}_{j+1,j},$$

where

$$\mathbb{Q}_j^+ = \widehat{\mathbb{R}}_{j,0}\mathbb{P}_j^+\widehat{\mathbb{R}}_{j,j+1}\mathbb{P}_j^-,$$
$$\mathbb{Q}_j^- = \widehat{\mathbb{R}}_{j,j+1}\mathbb{P}_j^-\widehat{\mathbb{R}}_{j,0}\mathbb{P}_j^+.$$

Once the reflection and transmission matrices are known, diffraction efficiencies and ellipsometric parameters can be calculated using the definitions in Section 2.4. The described algorithm was implemented in MATLAB.

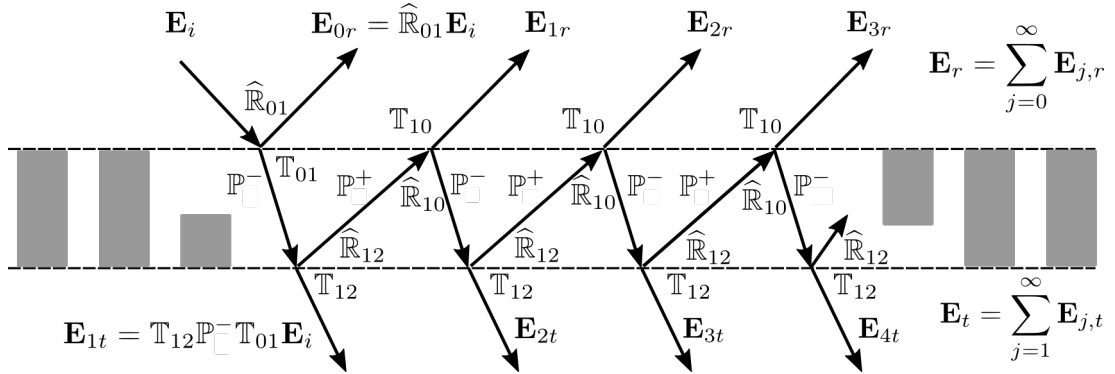Let us mention at the end of this Section that we will refer to this method as to the Normal Vector Method (NVM).

Figure 3.3: Airy-like series

Chapter 4

The C-Method

The coordinate transformation method, also the Chandezon method or abr. the C-Method is known since 1980. The basic idea is to make a coordinate transformation w.r.t. the shape of grating, and instead of solving simple equations on a complex domain to solve two more complex equations on half-planes. The procedure described in the paper [27] will be the primary source of this section, but we will also give some comments, extensions, remarks and observations from [15]. During the description of this method we will give some remarks related to a rescaling of the system at the beginning in order to show numerical problems which arise in a non-scaled system.

Let $\mathbf{x}, \mathbf{y}, \mathbf{z}$ be an orthonormal base in $\mathbb{R}^3$ and (x, y, z) corresponding Cartesian coordinates, and let the 2D surface be placed in $\mathbf{yz}$ plane, with grooves parallel to $\mathbf{y}$ direction. It represents 1D grating. This surface separates two semi-infinite homogeneous and isotropic media with constant permittivities denoted as ε_0 and ε_2. The grating profile P in Cartesian coordinates is determined by a function $z = a(y)$, with a being a differentiable function with the periodicity Λ. The translation coordinate system (x, y, z) is related with the Cartesian coordinate system by

$$x = w, \quad y = v, \quad z = u + a(v).$$

The inverse transformation is

$$w = x, \quad v = y, \quad u = z - a(y).$$

In the new coordinates the grating profile coincides with the $u = 0$ plane. Let us start from the system with TM-polarized incident wave, which is described by the Helmholtz equation

$$\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon_r}\frac{\partial H_x}{\partial y}\right) + \frac{\partial}{\partial z}\left(\frac{1}{\varepsilon_r}\frac{\partial H_x}{\partial z}\right) + H_x = 0, \quad \text{for all } (y, z) \in \mathbb{R}^2,$$

where $\varepsilon_r = \varepsilon_0$ in $\{u > 0\}$, $\varepsilon_r = \varepsilon_2$ in $\{u < 0\}$ are relative permitivitties above and below the grating respectively. Since ε_r is constant in each of the domains $D_\pm$, it is possible to split this equation to these sub-domains and write it as two separate equations

$$\begin{aligned}
\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \varepsilon_0\right) H_x &= 0, \quad \text{in } D_+ \\
\left(\frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} + \varepsilon_2\right) H_x &= 0, \quad \text{in } D_-,
\end{aligned} \tag{4.1}$$

joined with the interface conditions.

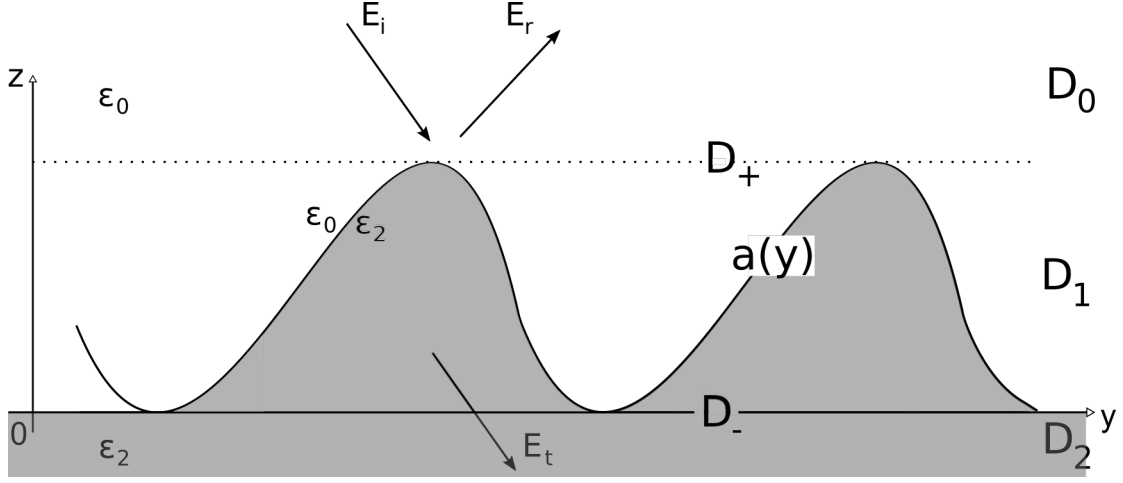Let us place here the Fig. 2.1 once more, see Fig. 4.1. In the domains D_0 and D_2 the field can be written

Figure 4.1: Grating

using Rayleigh expansions

$$H_x^{(1)} = A_i \exp[iq_0 y - is_{i,0}z] + \sum_{m=-\infty}^{\infty} A_r^m \exp(iq_m y + is_{r,m}z) + B_r^m \exp(iq_m y - is_{r,m}z)$$

$$H_x^{(2)} = \sum_{m=-\infty}^{\infty} A_t^m \exp(iq_m y - is_{t,m}z) + B_t^m \exp(iq_m y + is_{t,m}z),$$

where

$$q_m = n_0 \sin\vartheta + mq, \quad s_{r,m} = (\varepsilon_0 - q_m^2)^{\frac{1}{2}}, \quad s_{t,m} = (\varepsilon_2 - q_m^2)^{\frac{1}{2}},$$
$$\mathrm{Re}[s_{r,m}] + \mathrm{Im}[s_{r,m}] > 0, \quad \mathrm{Re}[s_{t,m}] + \mathrm{Im}[s_{t,m}] > 0,$$

and $q = \lambda/\Lambda$ and $A_r^m, B_r^m, A_t^m, B_t^m, A_i$ are constant amplitudes. These amplitudes has to satisfy $B_r^m = B_t^m = 0$ for all $m \in \mathbb{Z}$ in order to the field H_y be finite in $\pm\infty$ and propagating upwards in D_0 and downwards in D_2, as follows from the radiation conditions. The coefficients A_r^m, A_t^m has to be determined from the interface condition in the plane $v = 0$. In the domain D_1 the Rayleigh expansion is not possible. It could happen that one of the $s_{r/t,m}$ is equal to zero, it means the corresponding diffracted wave propagates parallel to the grating plane. The most simple solution is to change slightly the angle of incidence.

After the change of variables $v = y$, $u = z - a(y)$ the equations (4.1) changes to

$$\frac{\partial^2 H_x}{\partial v^2} - 2\dot{a}\frac{\partial^2 H_x}{\partial v \partial u} - \ddot{a}\frac{\partial H_x}{\partial u} + (1 + \dot{a}^2)\frac{\partial^2 H_x}{\partial u^2} + \varepsilon_0 H_x = 0,$$
$$\frac{\partial^2 H_x}{\partial v^2} - 2\dot{a}\frac{\partial^2 H_x}{\partial v \partial u} - \ddot{a}\frac{\partial H_x}{\partial u} + (1 + \dot{a}^2)\frac{\partial^2 H_x}{\partial u^2} + \varepsilon_2 H_x = 0,$$

with $\dot{a} = \partial a/\partial v$. The interface conditions are

$$\lim_{u \to 0+} H_x(v, u) = \lim_{u \to 0-} H_x(v, u), \quad \lim_{v \to 0-} (E_y + \dot{a}E_z)(v, u) = \lim_{v \to 0+} (E_y + \dot{a}E_z)(v, u), \text{ for all } u, v \in \mathbb{R}. \quad (4.2)$$

By defining $H_x' := -i\partial H_x/\partial u$ and using the identity

$$2\dot{a}\frac{\partial^2}{\partial v \partial u} + \ddot{a}\frac{\partial}{\partial u} = \left(\frac{\partial}{\partial v}\dot{a} + \dot{a}\frac{\partial}{\partial v}\right)\frac{\partial}{\partial u}$$

these two second-order PDE's can be rewritten as a system of two first-order PDE's:

$$\begin{bmatrix} \varepsilon_0 + \dfrac{\partial^2}{\partial v^2} & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} H_x \\ H'_x \end{pmatrix} = \frac{1}{i} \begin{bmatrix} i\left(\dfrac{\partial}{\partial v}\dot{a} + \dot{a}\dfrac{\partial}{\partial v}\right) & 1 + \dot{a}^2 \\ 1 & 0 \end{bmatrix} \frac{\partial}{\partial u} \begin{pmatrix} H_x \\ H'_x \end{pmatrix}. \tag{4.3}$$

$$\begin{bmatrix} \varepsilon_2 + \dfrac{\partial^2}{\partial v^2} & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} H_x \\ H'_x \end{pmatrix} = \frac{1}{i} \begin{bmatrix} i\left(\dfrac{\partial}{\partial v}\dot{a} + \dot{a}\dfrac{\partial}{\partial v}\right) & 1 + \dot{a}^2 \\ 1 & 0 \end{bmatrix} \frac{\partial}{\partial u} \begin{pmatrix} H_x \\ H'_x \end{pmatrix}. \tag{4.4}$$

This systems (4.3), (4.4) will be solved by a use of Fourier Methods. Since a is periodic, it can be expanded to a Fourier series

$$a(y) = \sum_{m=-\infty}^{\infty} a_m e^{iqmy}.$$

The equation (4.1) is not directly dependent on u, hence, we can assume an exponential dependence $\exp(i\rho u)$ of H_y in u. As now we have a problem with a straight boundary, the modes of the propagation in the v direction are dependent on $\exp(iq_m y)$ and the derivatives are

$$\frac{\partial}{\partial v} \to i\mathbf{q}, \quad \frac{\partial}{\partial u} \to i\rho.$$

By introducing a vector $\mathbf{F} = (\mathbf{H}_x, \mathbf{H}'_x)$ of the Fourier components of H_x, H'_x and the matrix

$$(\dot{\mathbf{a}})_{mn} = \dot{a}_{m-n} = \frac{q}{2\pi} \int_0^{\frac{2\pi}{q}} \dot{a}(t) \exp\left(-i(m-n)qt\right) \mathrm{d}t, \tag{4.5}$$

the systems (4.3), (4.4) can be rewritten in the Fourier space as two equations with infinite-dimensional matrix equations

$$\begin{bmatrix} (\varepsilon_0 + \mathbf{q}^2) & 0 \\ 0 & \mathbb{1} \end{bmatrix} \mathbf{F}^+ = \rho^+ \begin{bmatrix} -(\mathbf{q}\cdot\dot{\mathbf{a}} + \dot{\mathbf{a}}\cdot\mathbf{q}) & (\mathbb{1} + \dot{\mathbf{a}}\cdot\dot{\mathbf{a}}) \\ \mathbb{1} & 0 \end{bmatrix} \mathbf{F}^+,$$

$$\begin{bmatrix} (\varepsilon_2 + \mathbf{q}^2) & 0 \\ 0 & \mathbb{1} \end{bmatrix} \mathbf{F}^- = \rho^- \begin{bmatrix} -(\mathbf{q}\cdot\dot{\mathbf{a}} + \dot{\mathbf{a}}\cdot\mathbf{q}) & (\mathbb{1} + \dot{\mathbf{a}}\cdot\dot{\mathbf{a}}) \\ \mathbb{1} & 0 \end{bmatrix} \mathbf{F}^-,$$

which after simple modifications leads to two eigenvalue problems

$$\frac{1}{\rho^+}\mathbf{F}^+ = \begin{bmatrix} -\left(\mathbf{s}_r^2\right)^{-1}(\mathbf{q}\cdot\dot{\mathbf{a}} + \dot{\mathbf{a}}\cdot\mathbf{q}) & -\left(\mathbf{s}_r^2\right)^{-1}(\mathbb{1} + \dot{\mathbf{a}}\cdot\dot{\mathbf{a}}) \\ \mathbb{1} & 0 \end{bmatrix} \mathbf{F}^+,$$

$$\frac{1}{\rho^-}\mathbf{F}^- = \begin{bmatrix} -\left(\mathbf{s}_t^2\right)^{-1}(\mathbf{q}\cdot\dot{\mathbf{a}} + \dot{\mathbf{a}}\cdot\mathbf{q}) & -\left(\mathbf{s}_t^2\right)^{-1}(\mathbb{1} + \dot{\mathbf{a}}\cdot\dot{\mathbf{a}}) \\ \mathbb{1} & 0 \end{bmatrix} \mathbf{F}^-, \tag{4.6}$$

where $\mathbf{q} = \mathrm{diag}(q_m)$, $\mathbf{s}_t = \mathrm{diag}(s_{t,m})$, $\mathbf{s}_r = \mathrm{diag}(s_{r,m})$. These eigenvalue equations can be solved numerically after a proper truncation, giving the eigenvalues ρ^+, ρ^-.

Remark 2. *The non-scaled formula for $\dot{a}$ is*

$$\dot{a} = \frac{1}{\Lambda} \int_0^{\Lambda} \dot{a}(t) \exp\left(i(m-n)\frac{2\pi}{\Lambda}t\right) \mathrm{d}t.$$

The argument in the exponential is of an order 10^7, which is improper for numerical computations. The scaling can be done by making a substitution $\xi \rightarrow 2\pi t/\lambda$ in the integral in (4.5). The matrix $\dot{\mathbf{a}}$ has then the components

$$\dot{\mathbf{a}}_{mn} = \frac{qk_0}{2\pi} \int_0^{\frac{2\pi}{q}} \frac{da\,(\xi)}{d\xi} \exp(iq(m-n)\xi)d\xi,$$

which, with the abuse of notation $a(\xi) \equiv k_0 a(\xi)$, is the formula (4.5).

Due to the radiation conditions, all real eigenvalues with negative real part and all eigenvalues with negative imaginary part has to be discarded in the domain D^+ (more precisely, the amplitude corresponding to this mode will be put zero). All real eigenvalues with positive real part and all eigenvalues with positive imaginary part has to be discarded in the domain D^-. Let us also remind that the eigenvalues of the infinite-dimensional eigenvalue equation are $s_{r,m}, s_{t,m}$, because the transformation of coordinates cannot change the eigenvalues. However, it is not possible to replace the eigenvalues of truncated problem with its Rayleigh counterparts $s_{r/t,m}$, it could happen that truncated solutions will not converge to the solutions of the original problem [27]. In practice, the real eigenvalues are replaced by their Rayleigh counterparts $s_{r/t,m}$, where m runs over real positive eigenvalues ρ^+. Hence, the function H_x can be written in the domain D^+ as a series

$$H_x^+ = A_i \exp[iq_0 y - is_{i,0}z] + \sum_{n \in U^+} A_r^n \exp[iq_n y + is_{r,n}z] + \sum_m \exp(iq_m y) \sum_{l \in V^+} F_{ml}^+ \exp(i\rho_l^+ u)C_l^+,$$

where the first term represents the incident wave, the second contains the Rayleigh waves and the third one represents the diffracted waves. In the domain D^- the series is

$$H_x^- = \sum_{n \in U^-} A_t^n \exp[iq_n y + is_{t,n}z] + \sum_m \exp(iq_m y) \sum_{l \in V^-} F_{ml}^- \exp(i\rho_l^- u)C_l^-.$$

Here $A_r^n, A_t^n, C_l^\pm$ are the unknown diffraction amplitudes, which will be determined from the interface conditions (4.2), $F_{ml}^\pm$ are the elements of the H_x part of the l-th eigenvectors respective to $\rho_l^\pm$ and $U^\pm, V^\pm$ denote the sets of indices for the propagating and the evanescent orders in domains $D^\pm$. More precisely, the sets U^+, U^- contains the positive and negative real modes from $s_{r/t}$ respectively. The sets $V^\pm$ are complements to $U^\pm$. The problems regarding the convergence of eigenvalues are discussed in [28].

To use the boundary conditions, it is convenient to reformulate this field in terms of u and y, since then the boundary lies in the plane $u = 0$. The procedure will be shown on the incident wave, the remaining terms are analogous. Using the coordinate $z = v + a(u)$ and expanding $\exp(a(y))$ into the Fourier series give

$$A_i \exp[iq_0 y - is_{i,0}z] = A_i \exp\left[iq_0 y - is_{i,0}u\right] \sum_{m=-\infty}^{\infty} (L_m[-s_{i,0}] \exp(iqmy)) =$$

$$= A_i \sum_{m=-\infty}^{\infty} L_m\left[-s_{i,0}\right] \exp[iq_m] \exp[-is_{i,0}u],$$

where $L_m[-s_{i,0}]$ are coefficients of the Fourier series of $\exp(a(y))$, i.e.

$$L_m[-s_{i,0}] = \frac{q}{2\pi} \int_0^{\frac{2\pi}{q}} \exp\left[-is_{i,0}a(t) - iqmt\right] dt.$$

By defining the function

$$L_m[-\gamma] = \frac{q}{2\pi} \int_0^{\frac{2\pi}{q}} \exp\left[-i\gamma a(t) - iqmt\right] dt.$$

and making the analogous steps for other terms in $H_x^{\pm}$ it is possible to rewrite $H_x^{\pm}$ as

$$H_x^+ = \sum_m \exp(iq_m y) \times$$

$$\times \left(A_i L_m[-s_{i,0}] \exp(-is_{i,0}u) + \sum_{n \in U^+} L_{m-n}[s_{r,n}] \exp(is_{r,n}u) A_r^n + \sum_{l \in V^+} F_{ml}^+ \exp(i\rho_l^+ u) C_l^+ \right),$$

$$H_x^- = \sum_m \exp(iq_m y) \left(\sum_{k \in U^-} L_{m-n}[s_{t,k}] \exp(-is_{t,k}u) A_t^k + \sum_{p \in V^-} F_{mp}^- \exp(i\rho_p^- u) C_p^- \right).$$

Remark 3. *Let us make again a remark related to a scaling properties. The non-scaled version of L_m is*

$$L_m(\gamma) = \frac{1}{\Lambda} \int_0^{\Lambda} \exp[i\gamma a(t) - \frac{2\pi}{\Lambda} imt] dt.$$

This definition of L_m is useless for numerical computations. The function $a(y) - imy$ has for low m typically size about 10^{-7} and from numerical point of view this $L_m(\gamma)$ balances between zero and $\pm$Inf. This problem does not appear in the rescaled system.

The fields in D^+ and D^- must be equal at the boundary $u = 0$ therefore the coefficients of the Fourier series of $H_x^{\pm}$ must be equal when $u = 0$, which leads to the equations

$$A_i L_m[-s_{i,0}] + \sum_{n \in U^+} L_{m-n}[s_{r,n}] A_r^n + \sum_{l \in V^+} F_{ml}^+ C_l^+ = \sum_{n \in U^-} A_t^n L_{m-n}[-s_{t,n}] + \sum_{l \in V^-} F_{ml}^- C_l^-.$$

The matrix formulation of this system is

$$\left(F_{mn}^{R+}, F_{ml}^+, -F_{mk}^{R-}, -F_{mp}^- \right) \begin{pmatrix} A_r^n \\ C_l^+ \\ A_t^k \\ C_p^- \end{pmatrix} = -F_{m0}^{R,in},$$

where

$$F_{mn}^{R+} = L_{m-n}[s_{r,n}], \quad -F_{mk}^{R-} = L_{m-k}[-s_{t,k}], \quad F_{m0}^{R,in} = A_i L_m[-s_{i,0}].$$

After truncation to N orders this system has N equations for $2N$ variables. The second set of N equations in constructed from matching the tangential components of the electric fields. Here $\mathbf{t} = \mathbf{z} + \dot{a}\mathbf{y}$ and consequently

$$E_t = E_y + \dot{a} E_z.$$

It follows from the Maxwell equations that

$$E_z = \frac{1}{i\varepsilon_k} \frac{\partial H_x}{\partial y}, \quad E_y = -\frac{1}{i\varepsilon_k} \frac{\partial H_x}{\partial z}, \quad k = 0, 2,$$

which is expressed in the u, v coordinate system as

$$E_t \equiv G(u, v) = \frac{1}{i\varepsilon_k} \left[\dot{a} \frac{\partial H_x}{\partial v} - (1 + \dot{a}^2) \frac{\partial H_x}{\partial u} \right].$$

The continuity of the tangential components of the electric fields then means

$$\lim_{u \to 0+} \frac{1}{\varepsilon_0} \left[\dot{a} \frac{\partial H_x}{\partial v} - (1 + \dot{a}^2) \frac{\partial H_x}{\partial u} \right] = \lim_{u \to 0-} \frac{1}{\varepsilon_2} \left[\dot{a} \frac{\partial H_x}{\partial v} - (1 + \dot{a}^2) \frac{\partial H_x}{\partial u} \right],$$

Matching the coefficients of the tangential components of the electric fields yields the system

$$\left(G^{R+}_{mn}, G^{+}_{ml}, -G^{R-}_{mk}, -G^{-}_{mp}\right) \begin{pmatrix} A^n_r \\ C^+_l \\ A^k_t \\ C^-_p \end{pmatrix} = -G^{R,in}_{m0},$$

where

$$G^{R+}_{mn} = \frac{1}{\varepsilon_0} \sum_l \left[(\dot{a})_{m-l} q_l - (1 + \mathbf{a} \cdot \mathbf{a})_{ml} s_{r,n}\right] L_{l-n}[+s_{r,n}],$$

$$G^{R-}_{mk} = \frac{1}{\varepsilon_2} \sum_l \left[(\dot{a})_{m-l} q_l + (1 + \mathbf{a} \cdot \mathbf{a})_{ml} s_{t,k}\right] L_{l-k}[-s_{t,k}],$$

$$G^{R,in}_{m0} = \frac{1}{\varepsilon_0} \sum_l \left[(\dot{a})_{m-l} q_l + (1 + \mathbf{a} \cdot \mathbf{a})_{ml} s_{i,0}\right] L_l[-s_{i,0}],$$

$$G^{+}_{mn} = \frac{1}{\varepsilon_0} \sum_l \left[(\dot{a})_{m-l} q_l - (1 + \mathbf{a} \cdot \mathbf{a})_{ml} \rho^+_n\right] F^+_{ln},$$

$$G^{-}_{mp} = \frac{1}{\varepsilon_2} \sum_l \left[(\dot{a})_{m-l} q_l - (1 + \mathbf{a} \cdot \mathbf{a})_{ml} \rho^-_p\right] F^-_{lp}.$$

This altoghether will produce after the truncation a system of $2N$ equations for $2N$ unknowns

$$\begin{pmatrix} F^{R+}_{mn}, F^{+}_{ml}, -F^{R-}_{mk}, -F^{-}_{mp} \\ G^{R+}_{mn}, G^{+}_{ml}, -G^{R-}_{mk}, -G^{-}_{mp} \end{pmatrix} \begin{pmatrix} A^n_r \\ C^+_l \\ A^n_t \\ C^-_p \end{pmatrix} = - \begin{pmatrix} F^{R,in}_{m0} \\ G^{R,in}_{m0} \end{pmatrix}, \tag{4.7}$$

which has to be solved. The diffraction efficiencies can be calculated as

$$\eta^r_n = \frac{s_{r,n}}{s_{i,0}} |A^r_n|^2,$$

for the reflected orders and

$$\eta^t_n = \frac{\varepsilon_2 s_{t,n}}{\varepsilon_0 s_{i,0}} |A^t_n|^2,$$

for the transmitted orders, where $n \in U^{\pm}$ for the respective directions.

Remark to truncation   As was advertised, the key step in the C-method is a truncation of the infinite-dimensional system to a certain order. In numerical computations the truncation interval $[m_1, m_2]$ is chosen so that

$$m_1 = -\left[\frac{\alpha_0}{K}\right] - \frac{N-1}{2}, \quad m_2 = -\left[\frac{\alpha_0}{K}\right] + \frac{N-1}{2},$$

with N being an order of truncation. This gives a better convergence in general. The eigenvalue problems (4.6) has then the size $2N \times 2N$ and produces two sets of $2N$ eigenvalues (one for each medium). The N eigenvalues from each set will be discarded. The remaining N eigenvalues in each set is then divided into two sets — one contains the real eigenvalues, the second one the eigenvalues with nontrivial imaginary part. The real eigenvalues are replaced by their Rayleigh counterparts. It is not possible to say generally how many eigenvalues will contains the set $U^{\pm}$, $V^{\pm}$, we know only that $U^+ \cup V^+, U^- \cup V^-$ contains

both N eigenvalues. Therefore the sum

$$H_x^+ = \sum_m \exp(iq_m y) \times$$

$$\times \left(A_i L_m[-s_{i,0}] \exp(-is_{i,0}u) + \sum_{n \in U^+} L_{m-n}[s_{r,n}] \exp(is_{r,n}u) A_r^n + \sum_{l \in V^+} F_{ml}^+ \exp(i\rho_l^+ u) C_l^+ \right),$$

will contain $2N$ terms, similarly for H_x^- and electric components. The number of reflected and transmitted propagation orders corresponds to a number of elements of the sets U^+ and U^- respectively.

Chapter 5

Implementation and comparison of the described methods

5.1 Testing the implementation of the C-Method

The C-method was implemented on a basis of the paper [27]. The implementation was tested at first on a simple reference problem of a diffraction on a line interface. The exact solution is known as the Fresnel equations. The results will be demonstrated on two particular examples. The first one has $n_1 = n_2$, i.e. it is a simple propagation of light, the second one has parameters $n_1 = 1$, $n_2 = 2.65$, the incident angle is $15°$.

The diffraction efficiencies computed from the Fresnel equations for the case $n_2 = 1$ were $\eta_0^t = 1$ and zero others. The reflection amplitude was $\eta_0^r = 0.1937$ for the case $n_2 = 2.65$, the transmission amplitude was $\eta_0^t = 0.8063$, others were equal to zero (TM polarization). The results obtained from our implementation perfectly corresponds to the analytical results computed by using the Fresnel equations. The transmitted wave propagated under the angle $5°36'$. It is possible to make a conclusion that our implementation gives correct results for simple cases. The amplitudes of waves for two particular cases can be seen on the following two figures.
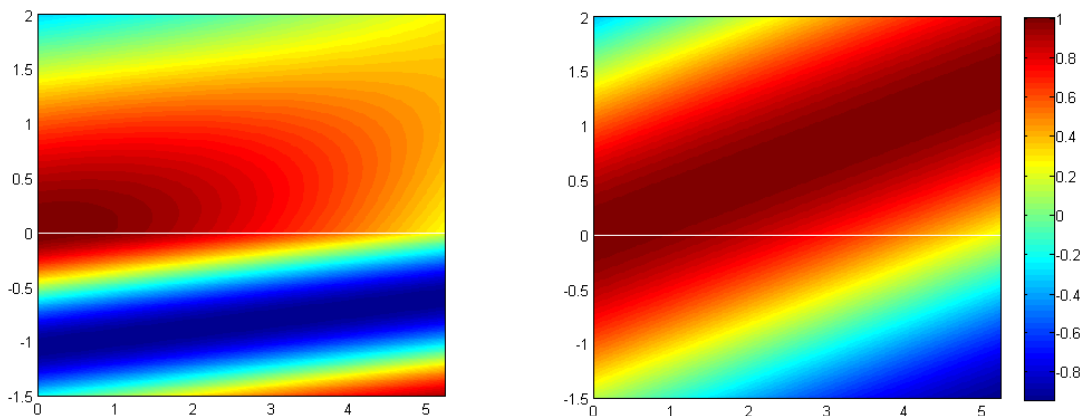
Figure 5.1: Propagation of light (right) and bending of light (left)

The second set of tests was made on the reference examples from the paper [27]. The first one was the diffraction of light with the wavelength λ incident under the angle $15°$ on an asymmetrical shallow

Order	TE	TM	TE from [27]	TM from [27]
η_{-2}^r	0.3025e-2	0.4823e-2	0.3025e-2	0.4823e-2
η_{-1}^r	0.2770	0.3466	0.2770	0.3466
η_0^r	0.4363	0.3062	0.4363	0.3062
η_1^r	0.1508	0.1902	0.1508	0.1902

Table 5.1: Results for a grating with asymmetrical profile.

Order	TE	TM	TE from [27]	TM from [27]
η_{-2}^r	0.2994e-2	0.5962e-3	0.2982e-2	0.5882e-3
η_{-1}^r	0.6392e-3	0.1005e-2	0.6300e-3	0.9762e-3
η_0^r	0.1969e-2	0.1887e-3	0.1963e-2	0.1847e-3
η_1^r	0.1254e-2	0.9483e-3	0.1252e-2	0.9344-e3
η_{-3}^t	0.5272e-1	0.1237e-2	0.5274e-1	0.1219e-2
η_{-2}^t	0.1348	0.1324	0.1347	0.1320
η_{-1}^t	0.1281	0.1712	0.1280	0.1710
η_0^t	0.1585	0.1134	0.1586	0.1138
η_1^t	0.4457	0.5314	0.4457	0.5317
η_2^t	0.7337e-1	0.4723e-1	0.7337e-1	0.4726e-1

Table 5.2: Results for a grating with asymmetrical profile.

grating with the profile

$$z = a(y) := 0.1d \sin\left(\frac{\pi}{d}y\right) + 0.2d \cos\left(\frac{2\pi}{d}y - \frac{5\pi}{9}\right),$$

and with $n_1 = 1$, $n_2 = 1 + 5i$, $d = \lambda$, the truncation numbers were $m_1 = -6$, $m_2 = 4$. The results are written in the Tab. 5.1. They perfectly corresponds to those from [27].

The second grating had a sinusoidal profile

$$a(y) := d \cos\left(\frac{\pi}{d}y\right),$$

the refractive indices were $n_1 = 1$, $n_2 = 1.5$ and the truncation numbers were $m_1 = -28$, $m_2 = 26$ and the depth was $d = \lambda$. The results are summarized in the Tab. 5.2. It is possible to see a slight difference in the diffraction efficiencies to those from [27], which is probably due to differences in numerical implementation (the paper was published in 1999). The diffraction orders with a small amplitude differ most, but the relative error is small. The intensity plot can be found in Fig. 5.2

These results will be later compared with the results obtained by a use of the LMT and the NVM.

Convergence of the C-Method The convergence of the C-Method was tested on two examples introduced above. The computation was started with a certain order and quit when MATLAB reported the singularity of the system (4.7). The first tested case was a shallow grating with the asymmetrical profile and s-/p- polarized incident light having the wavelength $\lambda = 300$ nm. The solver worked up to the order 80, the system (4.7) became numerically singular for larger truncation orders. The diffraction efficiencies stabilized rapidly around certain values for the truncation numbers $N \geq 6$. To qualitatively measure the oscillations, we defined an error

$$\mathcal{E}_r := \sum_{N=-2}^{1} \left(\frac{\eta_N^r - \eta_{ref,N}^r}{\eta_{ref,N}^r}\right)^2,$$
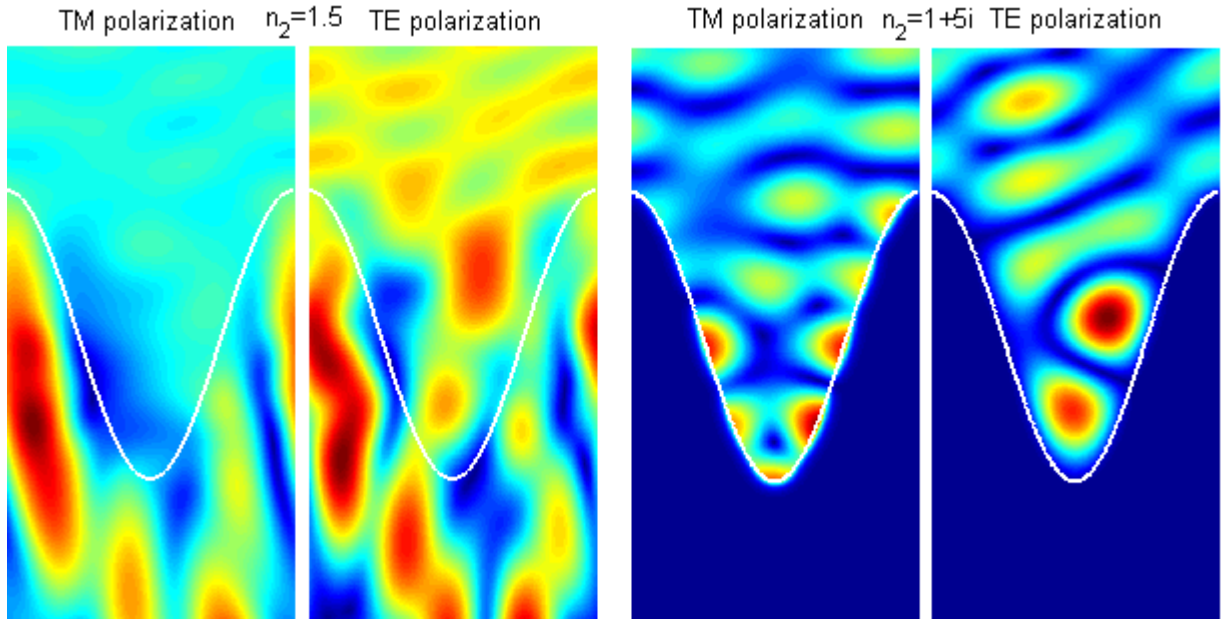
Figure 5.2: Intensity plots for sinusoidal grating with a various refractive index and incident light polarization

where $\eta^r_{ref,N}$ is a reference value found as the mean value of diffraction efficiencies for truncation number between 20 and 60. The graph of error can be found in Fig. 5.3. The results are very stable w.r.t. truncation order, and the the maximal difference between the largest and lowest value for truncation numbers between 15 and 80 is of an order 10^{-25}. The graph of computation time is in Fig. 5.4.

The second tested case was the sinusoidal grating, with $d = \Lambda = 600$ nm, $\lambda = 300$ nm. The error was defined here as

$$\mathcal{E}_r := \sum_{N=-2}^{1} \left(\frac{\eta^r_N - \eta^r_{ref,N}}{\eta^r_{ref,N}} \right)^2 + \sum_{N=-3}^{2} \left(\frac{\eta^t_N - \eta^t_{ref,N}}{\eta^t_{ref,N}} \right)^2 .$$

The reference solution was found again as a mean value of results for truncation numbers from $N = 15$ to $N = 25$. The graph in Fig. 5.5 shows the dependence of error on the truncation number, it is possible to see much larger oscillations than in previous case. The convergence here is good as well, but the numerical errors are present already for the truncation number 30. Significant increases are observed for the highest truncation number in s- and p-polarization, despite the fact that MATLAB did not report any problem. The main problem here is very fast decay of coefficients of L_m fields. In general, the C-method is considered to be ineffective for deep gratings, the main reason is the slow convergence of eigenvalues of truncated equation to Rayleigh eigenvalues [29]. The eigenvalues in our implementation were compared with the ones in [29], giving a perfect agreement.

5.2 Testing the implementation of the LMT and the NVM

The implementation of the LMT was done by Dr. Antoš (supervisor of this Thesis) and is already debugged. The NVM implementation was made by the author of this Thesis. The both codes were at first tested on simple examples of line interface, giving a perfect correspondence.

After that we used more complex situations. The first example was grating with the refractive index $n_2 = 1.5$, second grating had the refractive index $n_2 = 1+5i$. Since the LMT implementation is not using the correct factorization rules, there can be expected a poor convergence for dense optical environment
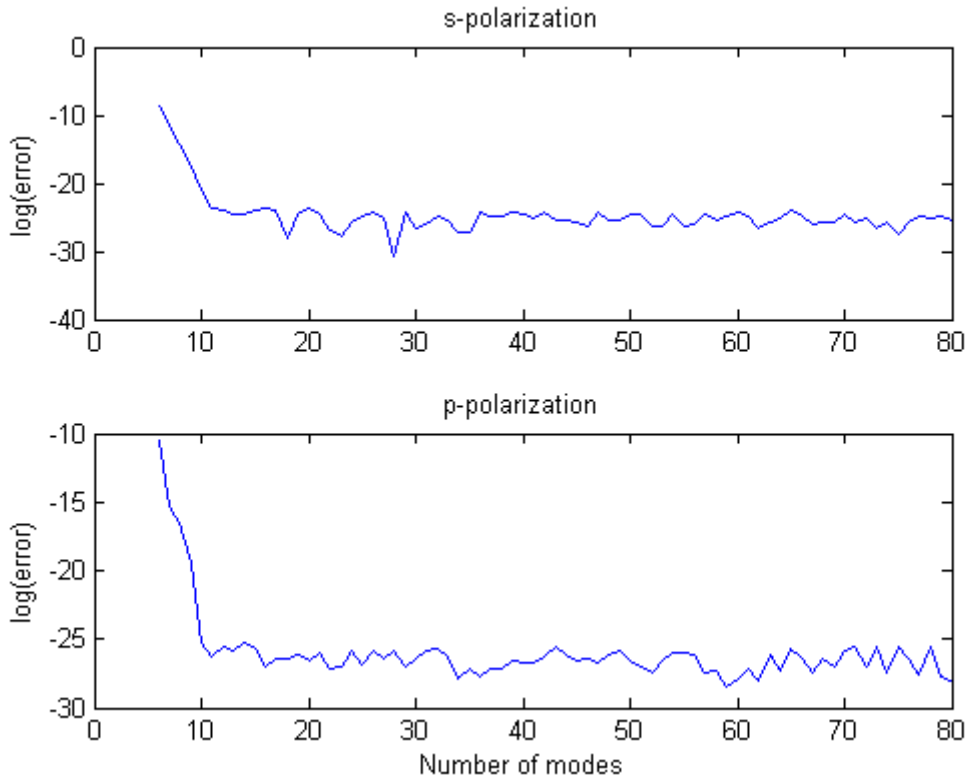
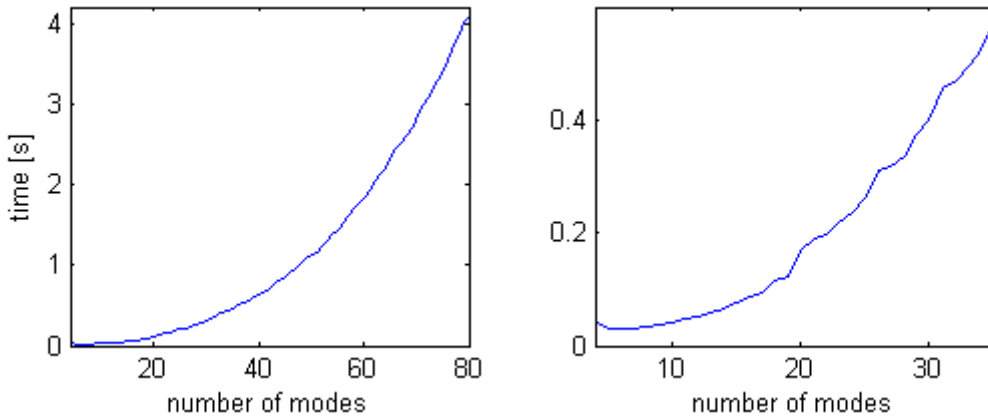Figure 5.3: Error in s- and p- polarization for the C-Method (shallow grating)

Figure 5.4: Computation time w.r.t. number of modes for the C-Method. The shallow grating is on the left, the sinusoidal on the right.

in TM polarization, and even poorer convergence for strongly conductive gratings. On the other hands, the convergence for shallow gratings will be not much affected by the incorrect factorization, because $E_n \approx E_x$, $E_t \approx E_z$. For this reason we decided to skip the shallow grating as a test example and use the deep sinusoidal grating with $n_2 = 1 + 5i$ as an example to demonstrate the benefits of the NVM. The question of convergence of these systems is in fact a two-dimensional problem, because it depends
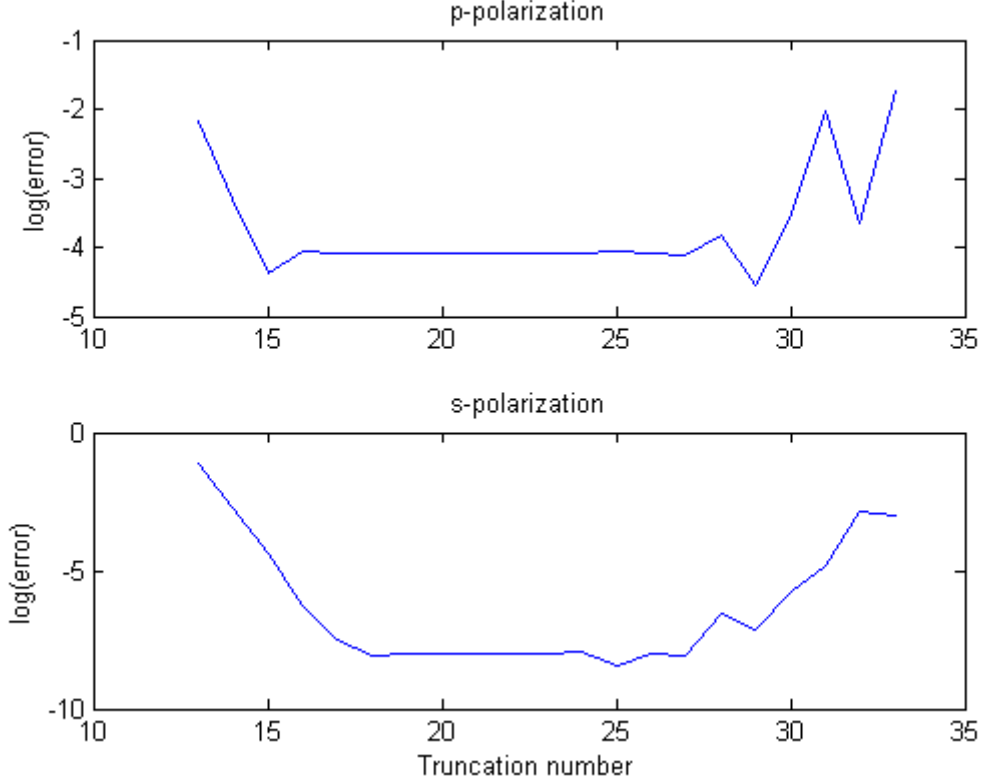
Figure 5.5: Error in s- and p- polarization (sinusoidal grating)

on the number of modes and on the number of slices. Therefore the results are plot in two-dimensional graph. The essential step is a suitable definition of discretization error. The definition of error is again adopted from the paper [7]:

$$\mathcal{E}_r := \sum_{N=-2}^{1} \left(\frac{\eta_N^r - \eta_{ref,N}^r}{\eta_{ref,N}^r} \right)^2,$$

where η_{ref}^r is a reference diffraction efficiency vector for reflected orders and the summation numbers corresponds to the propagating diffraction orders. Similarly the error for transmission orders is defined by

$$\mathcal{E}_t := \sum_{N=-3}^{2} \left(\frac{\eta_N^t - \eta_{ref,N}^t}{\eta_{ref,N}^t} \right)^2.$$

Due to a lack of rigorous results and explicit examples the error was related to the results computed from the C-Method and from the NVM with a twice number of modes and slices (double-step method).

The logarithm of error w.r.t. number of slices and the number of modes for sinusoidal grating with $n_2 = 1.5$ is plotted in Fig. 5.7. There are six plots in this figure. The label N_s on axis x means the number of slices. The label N_m means the truncation number. The shortcut DS means that the reference solution was computed with the double-step method. More precisely, the error here is related to the reference solution obtained by a use of the NVM with the number of modes $N_m = 60$ and the number of slices $N_s = 150$. The shortcut C-M means C-Method and the error in this plot is related to the reference solution obtained by a use of the C-Method. It can be seen that the NVM is a slightly

more precise, but the difference against the LMT is not large. The Fig 5.8 contains a plot with the error $\mathcal{E}_t$. Again, values obtained by the NVM are slightly more accurate than the LMT, but the difference is not large. The Fig. 5.9 contains plots for highly conductive sinusoidal grating with $n_2 = 1 + 5i$. The error levels of s-polarization and NVM are good and comparable to previous values. However, the results obtained by the LMT are very inaccurate and the convergence is very poor (see the positive sign of the logarithm of the error). The LMT is not a good choice for this case. The reference solution for the DS-error was obtained by a choice of parameters $N_m = 100$, $N_s = 300$.

The plot of the intensity of fields looks similar for the LMT method and the C-method, cf. Fig. 5.2, 5.6, but the LMT one was obtained for $N_m = 200$, $N_s = 100$. The subroutine for plotting the intensity obtained from NVM was unfortunately not yet finished.
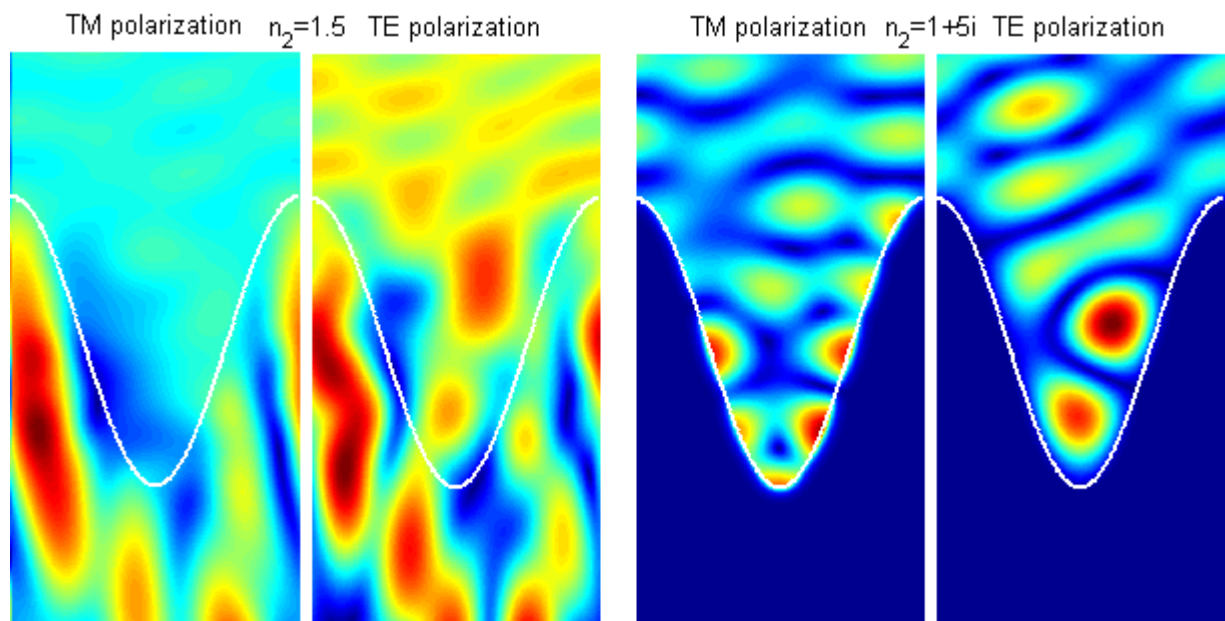
Figure 5.6: Intensity plots for sinusoidal grating with a various refractive index and incident light polarization
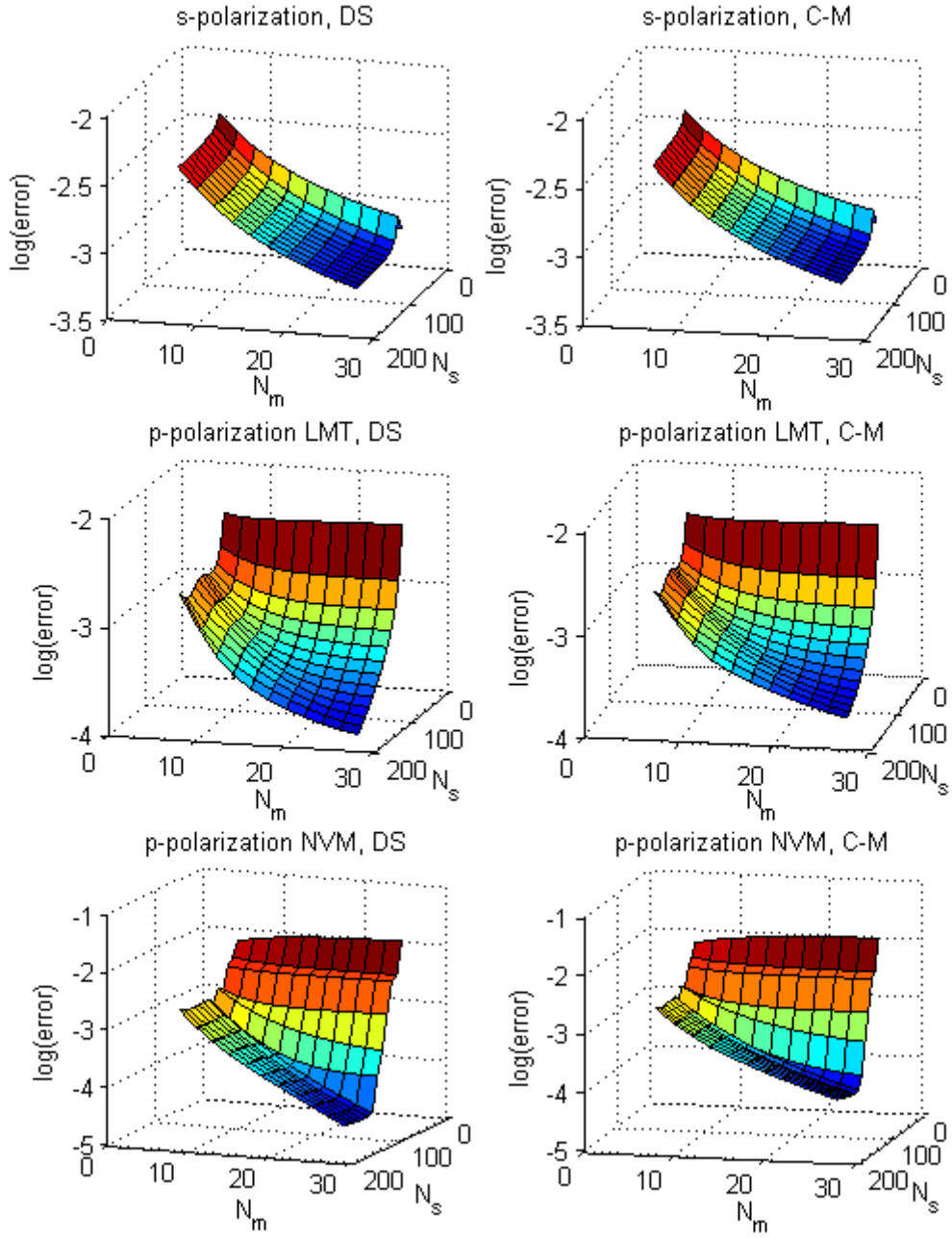
Figure 5.7: Error $\mathcal{E}_r$ (reflected orders) plotted w.r.t. number of slices and number of modes. Sinusoidal grating with $n_2 = 1.5$ and $\Lambda = d$
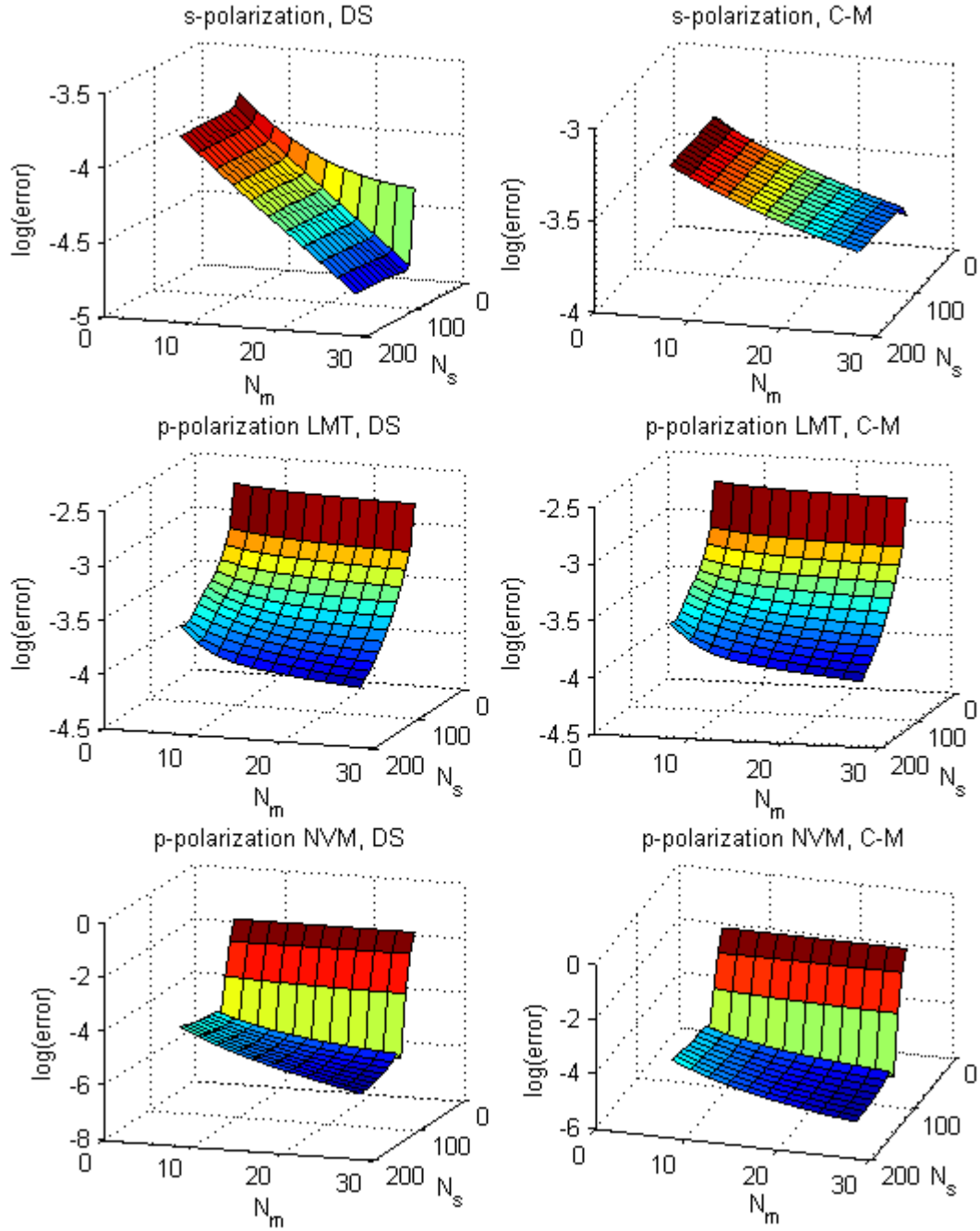
Figure 5.8: Error $\mathcal{E}_t$ (transmitted orders) plotted w.r.t. number of slices and number of modes. Sinusoidal grating with $n_2 = 1.5$ and $\Lambda = d$
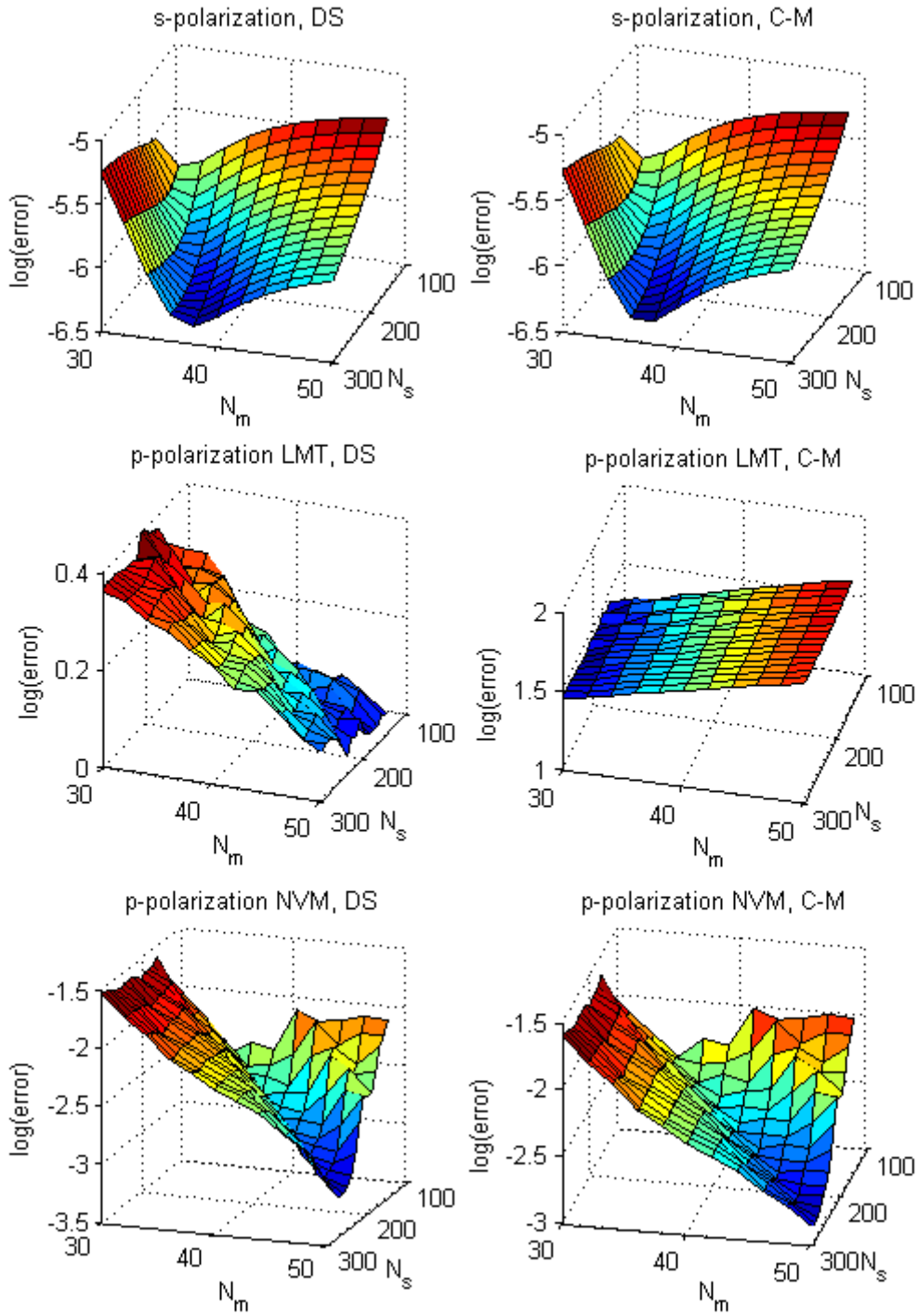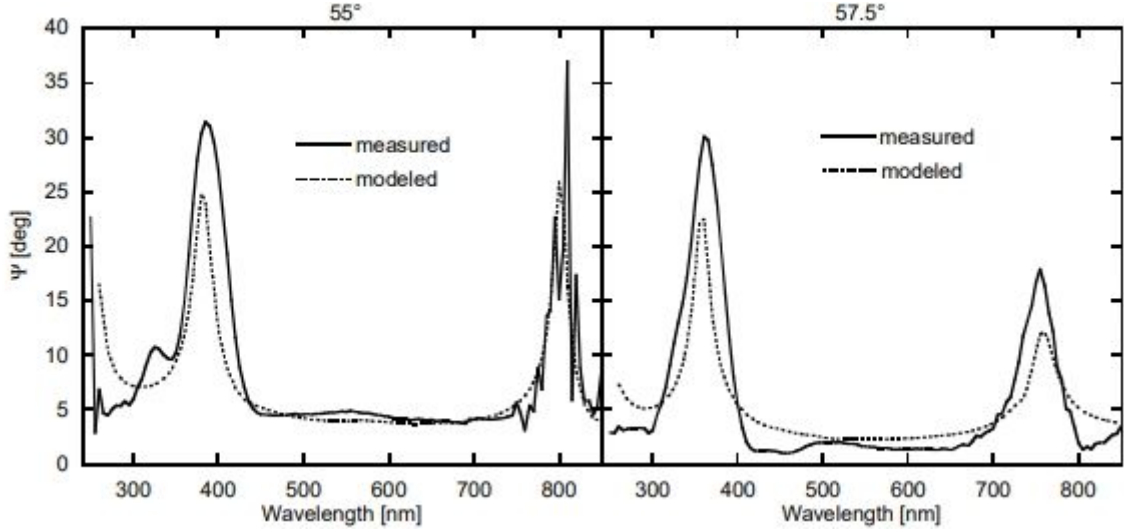
Figure 5.9: Error $\mathcal{E}_r$ (reflected orders) plotted w.r.t. number of slices and number of modes. Sinusoidal grating with $n_2 = 1 + 5i$ and $\Lambda = d$

Figure 5.10: Ellipsometric parameter ψ for the shallow grating with the sinusoidal-triangular profile and $n_2 = 1.52$ (polymer grating) from [30]

5.3 Comparison with the experimental results

Polymer grating At last, the implementation of the C-Method was compared to experimental results from literature. As a model served two gratings — one with a wavy profile which is a combination of sinusoidal and rectangular profile [30] and one which is made from nickel and has a sinusoidal profile. The ratio between the period and the depth of the grating is for the polymer grating more than ten and for nickel grating more than five therefore both gratings are shallow. For this reason the numerical effects described in previous Sections does not play an essential role here.

The first case was a shallow grating with a period $\Lambda = 9365$ nm and depth $d = 620$ nm made of a polymer with the refractive index equal to $n_2 = 1.52$ on the glass substrate with almost the same refractive index. The incident angles were $\vartheta = 55°$ and $\vartheta = 57.5°$.

$$
a(y) = \begin{cases} (1-\sigma)\dfrac{d}{2}\left(1 - \cos\left(\dfrac{2\pi y}{\Lambda}\right)\right) + \dfrac{2\sigma dy}{\Lambda}, & y \in \left[0, \frac{\Lambda}{2}\right] \\[3mm] (1-\sigma)\dfrac{d}{2}\left(1 - \cos\left(\dfrac{2\pi y}{\Lambda}\right)\right) - \dfrac{2\sigma dy}{\Lambda}, & y \in \left[\frac{\Lambda}{2}, \Lambda\right]. \end{cases}
$$

The parameter σ determines the combination between sinusoidal and triangular profile. More precisely, if $\sigma = 1$, the profile is triangular, if $\sigma = 0$, profile is sinusoidal. The fitting of the profile function to the profile scanned by AFM yielded the value $\sigma = 0.6$.

The ellipsometric parameter ψ from [30] (with a use of the LMT) is in Fig. 5.11, the same result computed with a use of our C-Method implementation is in Fig. 5.10. It can be seen that the position of the peaks is almost the same, but the peak for the angle $\vartheta_i = 55°$ is much larger than measured one. Also the peaks are narrower than the measured ones. The main reason for this is in the difference between the refractive index of polymer and glass. Although the difference is small, it causes a back-reflections which were included in the model from [30] but not in our model, see Fig. 5.12.
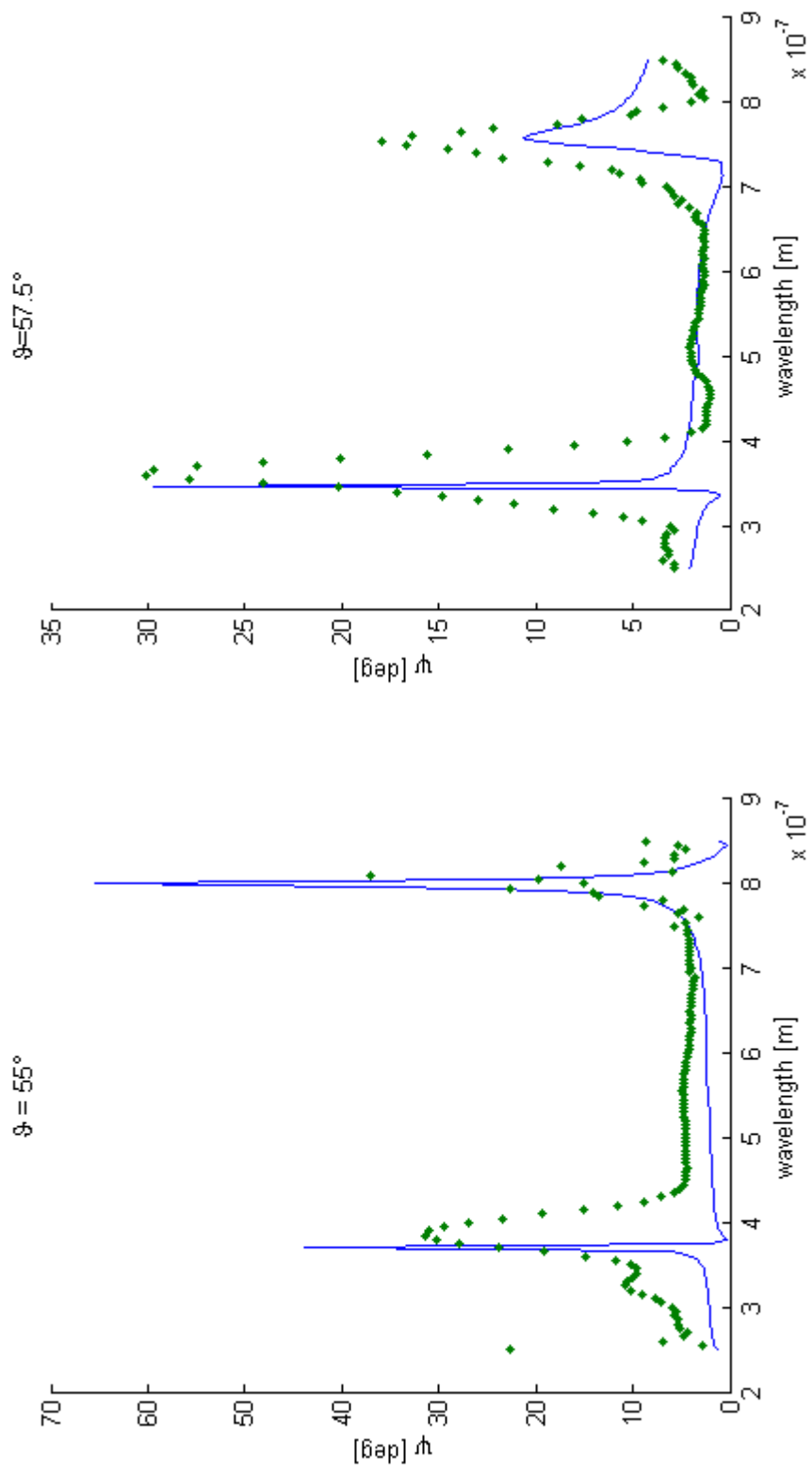
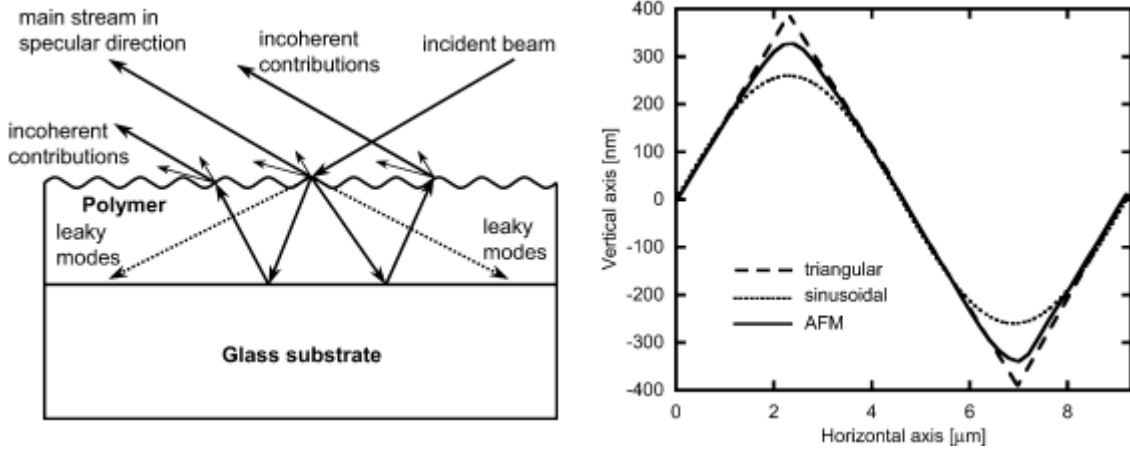Figure 5.11: Ellipsometric parameter ψ for the polymer grating computed from the C-Method

Figure 5.12: Back-reflections on polymer grating (left) and grating profile from AFM (right) [12].
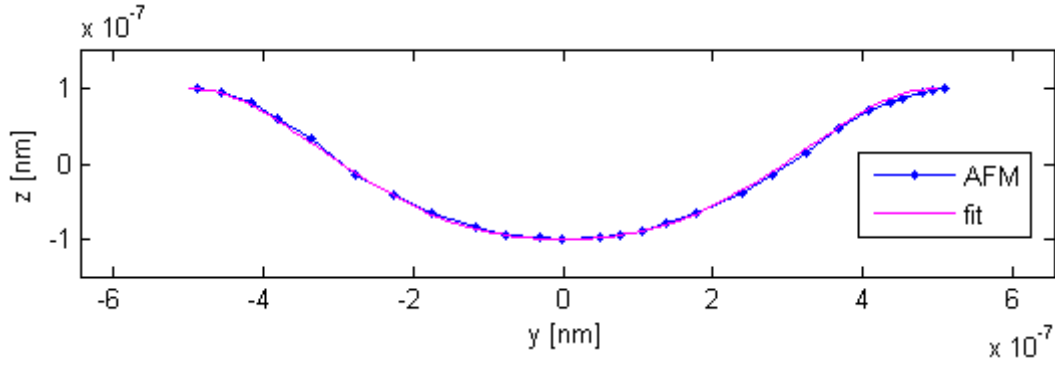

Figure 5.13: Profile of a Nickel grating: AFM (scaled to width 200.4 nm) vs. fit

Nickel grating   The second case was a nickel grating with the refractive index from Fig. 5.15 and a profile of the form

$$z(y) = \frac{d}{2}\left[1 - \cos\left((\pi - h_2)\frac{2y}{\Lambda} + h_2\left(\frac{2y}{\Lambda}\right)^2\right)\right], \quad y \in \left[-\frac{\Lambda}{2}, \frac{\Lambda}{2}\right].$$

The refractive index was measured on a thin layer of nickel put beside the grating. The AFM yielded the period $\Lambda = 917$ nm and the depth $d = 180$ nm. The parameter h_2 determines the shape of the profile. For $h_2 = 0$ the profile is sinusoidal, and for $h_2 = \pi$ the profile is sinusoidal with a quadratic argument. The period is usually determined precisely, but the depth is often inaccurate. Dr. Antoš found by fitting that the depth and h_2 should be $d = 200.4$ nm and $h_2 = 1.0834$, the ellipsometric parameters are in Fig. 5.14. We took this as a reference parameters and computed the ψ, Δ with the C-Method, see Fig. 5.16–5.18 and NVM, see Fig. 5.19–5.21. The modeled curves for ψ, Δ fit well to the experimental ones, despite the fact that our model did not count with the thin NiO_2 layer on the surface of the grating. The reasonable step to further improve the results would be to extend our implementation of the C-Method and the NVM also for the cases of coated gratings, using the procedure described in [15], Chapter 8.4, 12.5. Figure 5.13 gives a comparison between measured (AFM) and fitted profile.
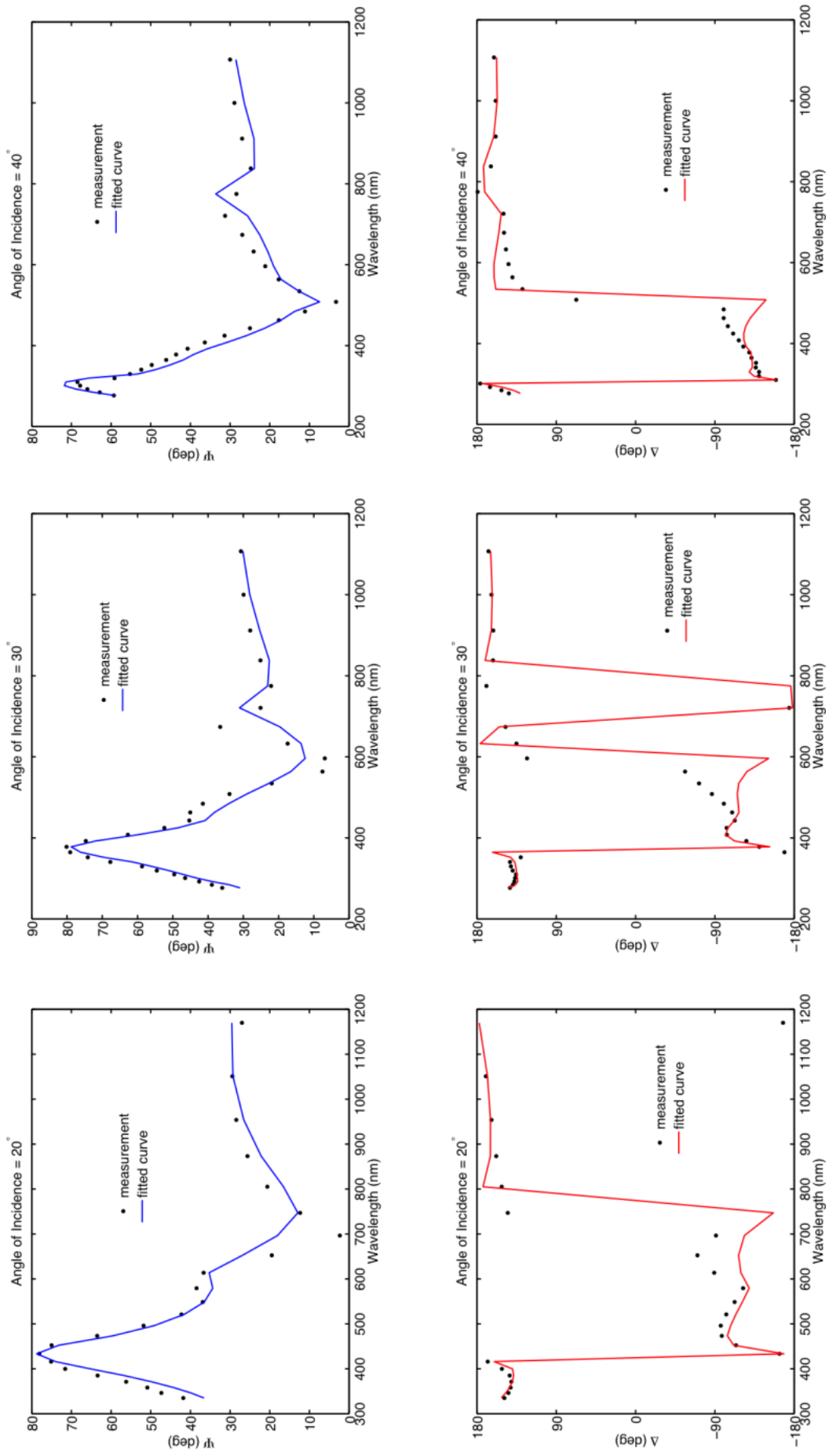
Figure 5.14: Ellipsometric parameter ψ, Δ for the sinusoidal grating made of nickel, the LMT
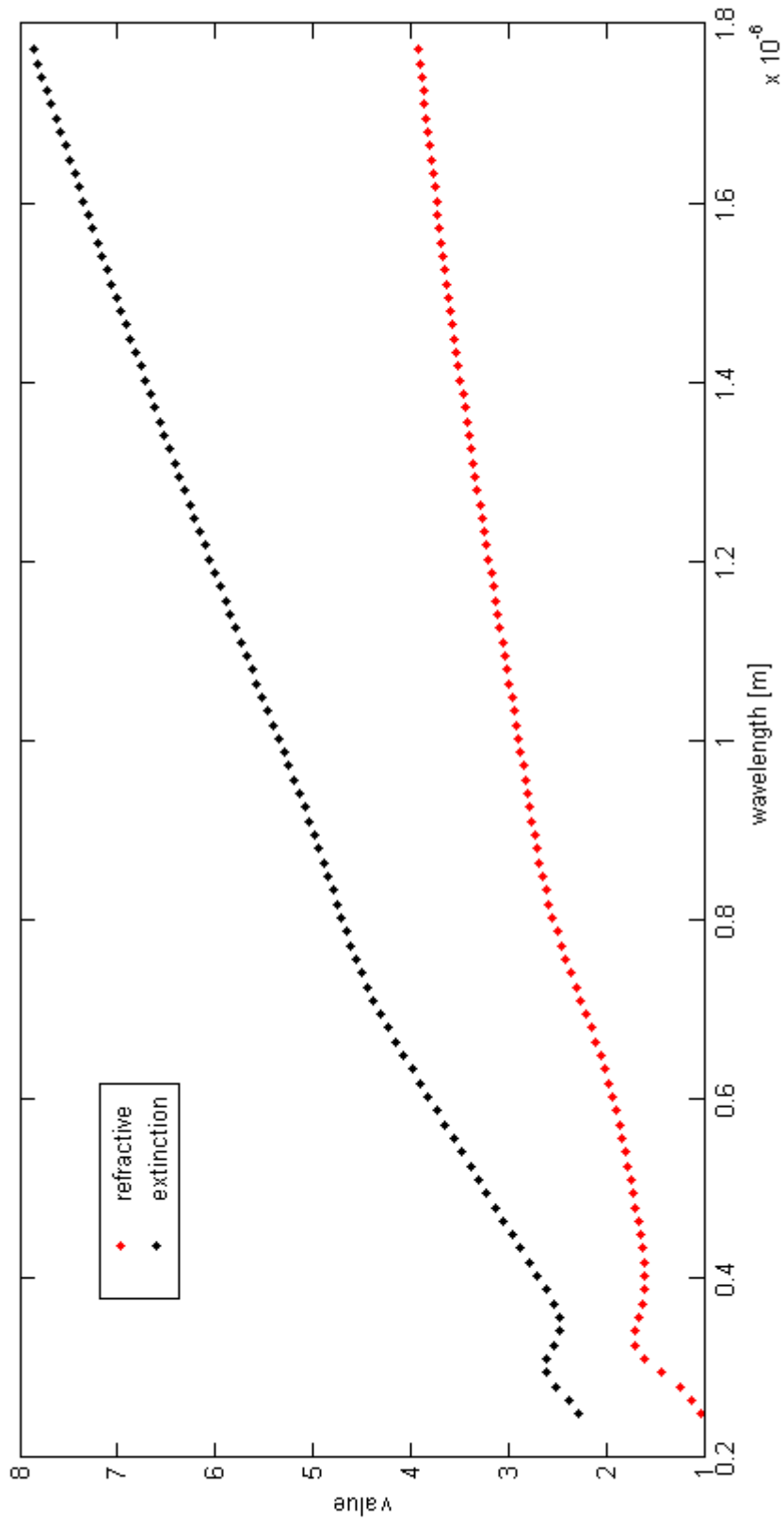
57

Figure 5.15: Nickel refractive index and extinction coefficient of w.r.t. wavelength
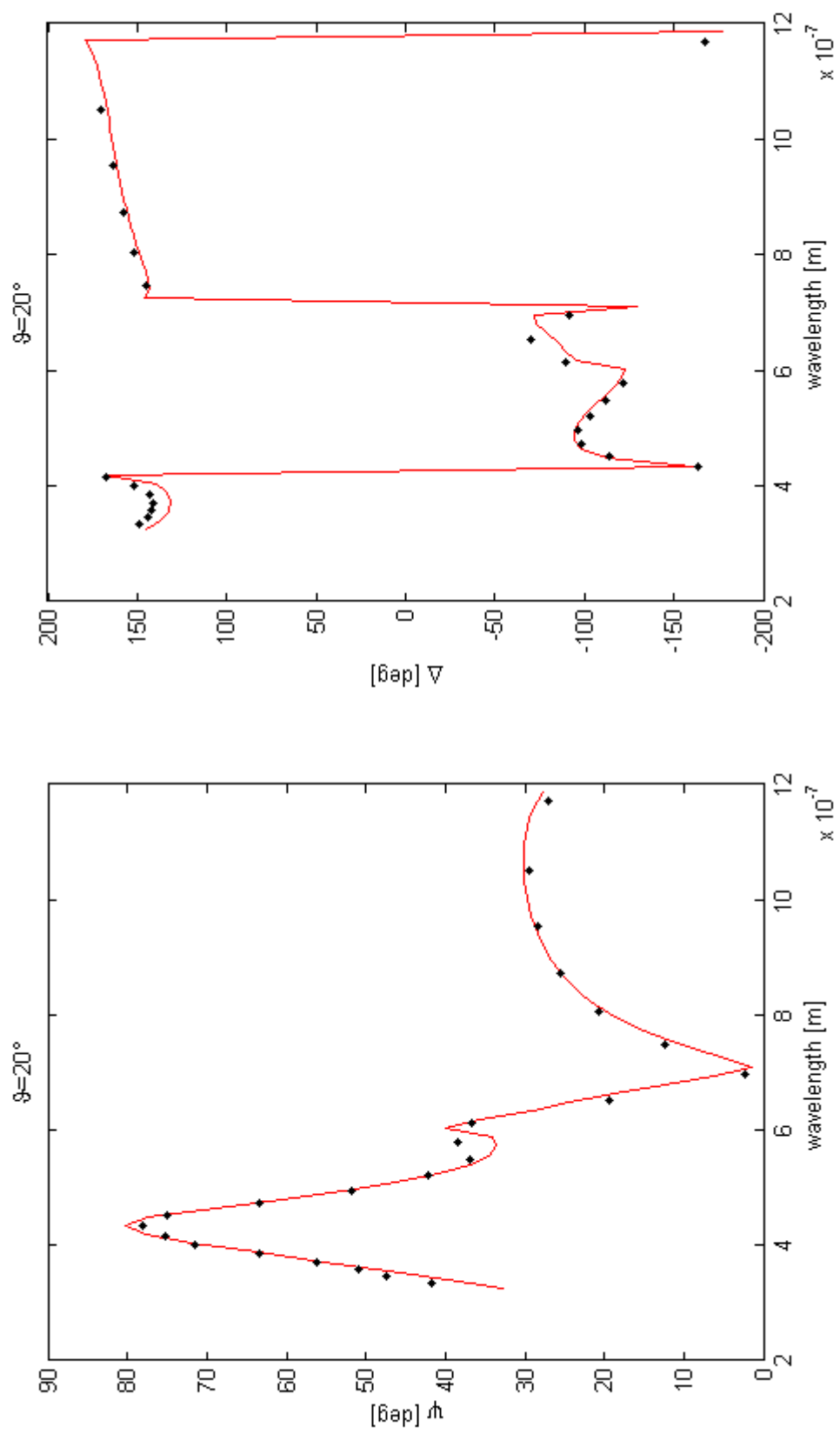
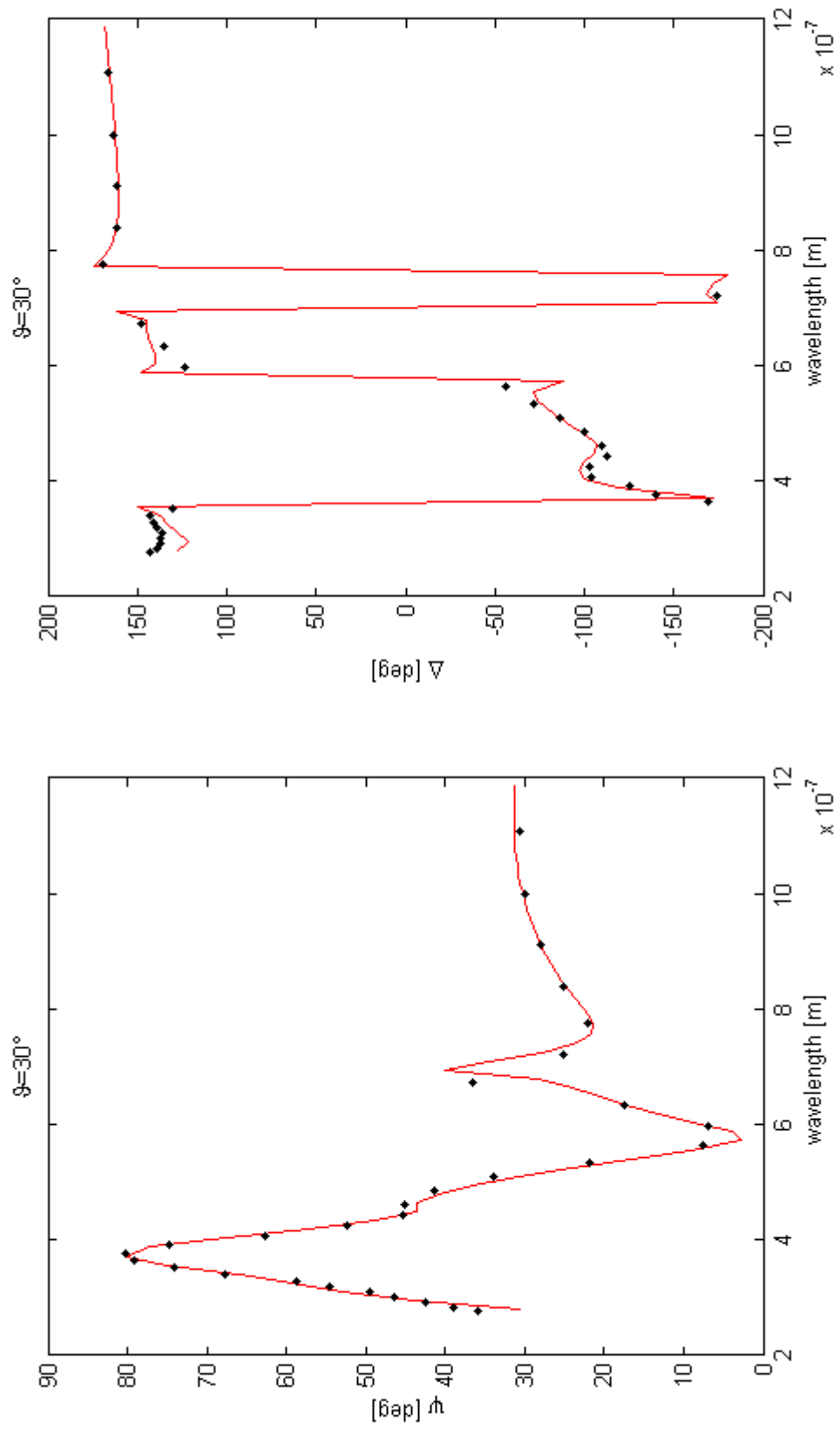Figure 5.16: Ellipsometric parameters for the nickel sinusoidal grating, the C-Method

Figure 5.17: Ellipsometric parameters for the nickel sinusoidal grating, the C-Method
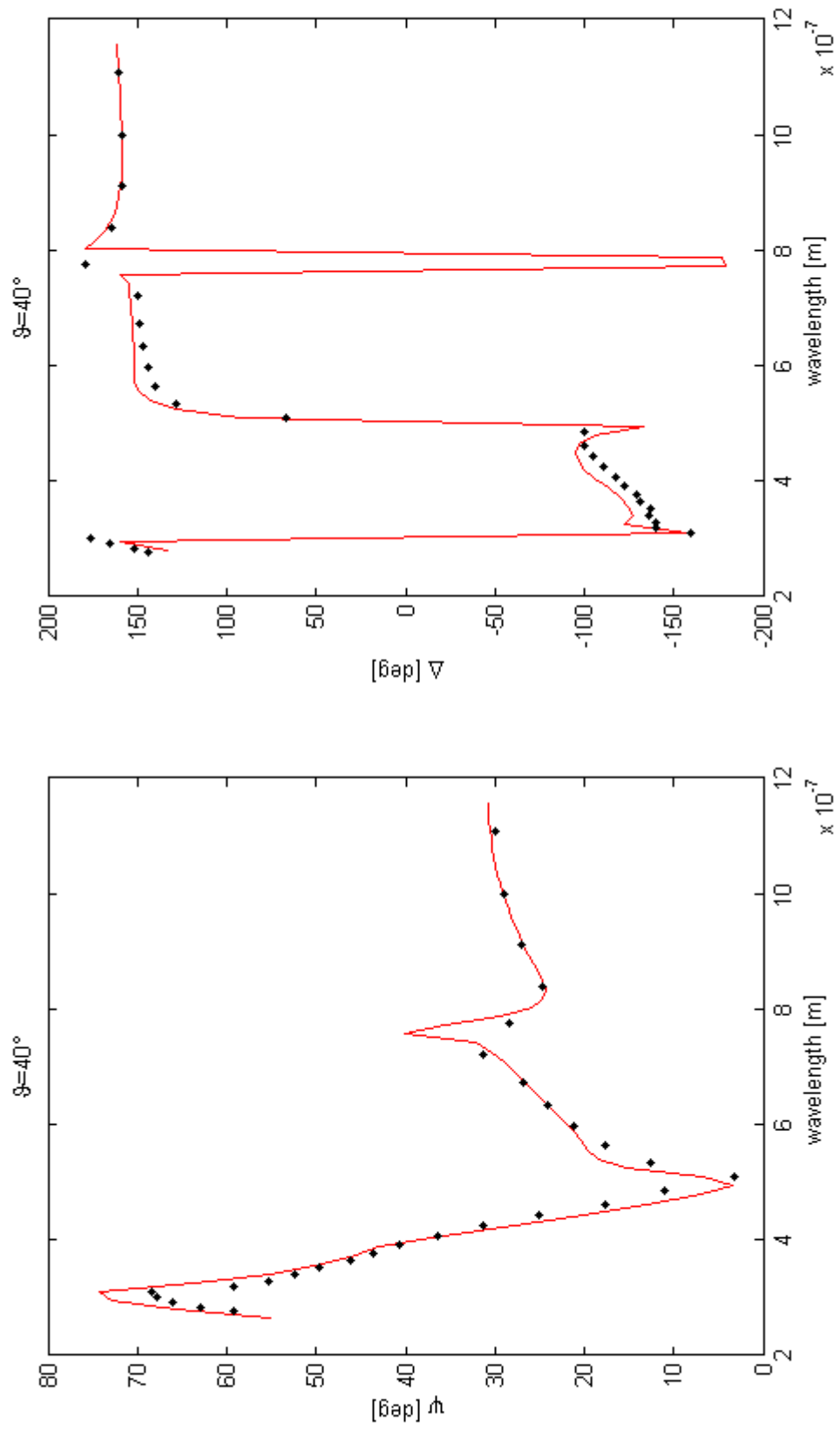
Figure 5.18: Ellipsometric parameters for the nickel sinusoidal grating, the C-Method
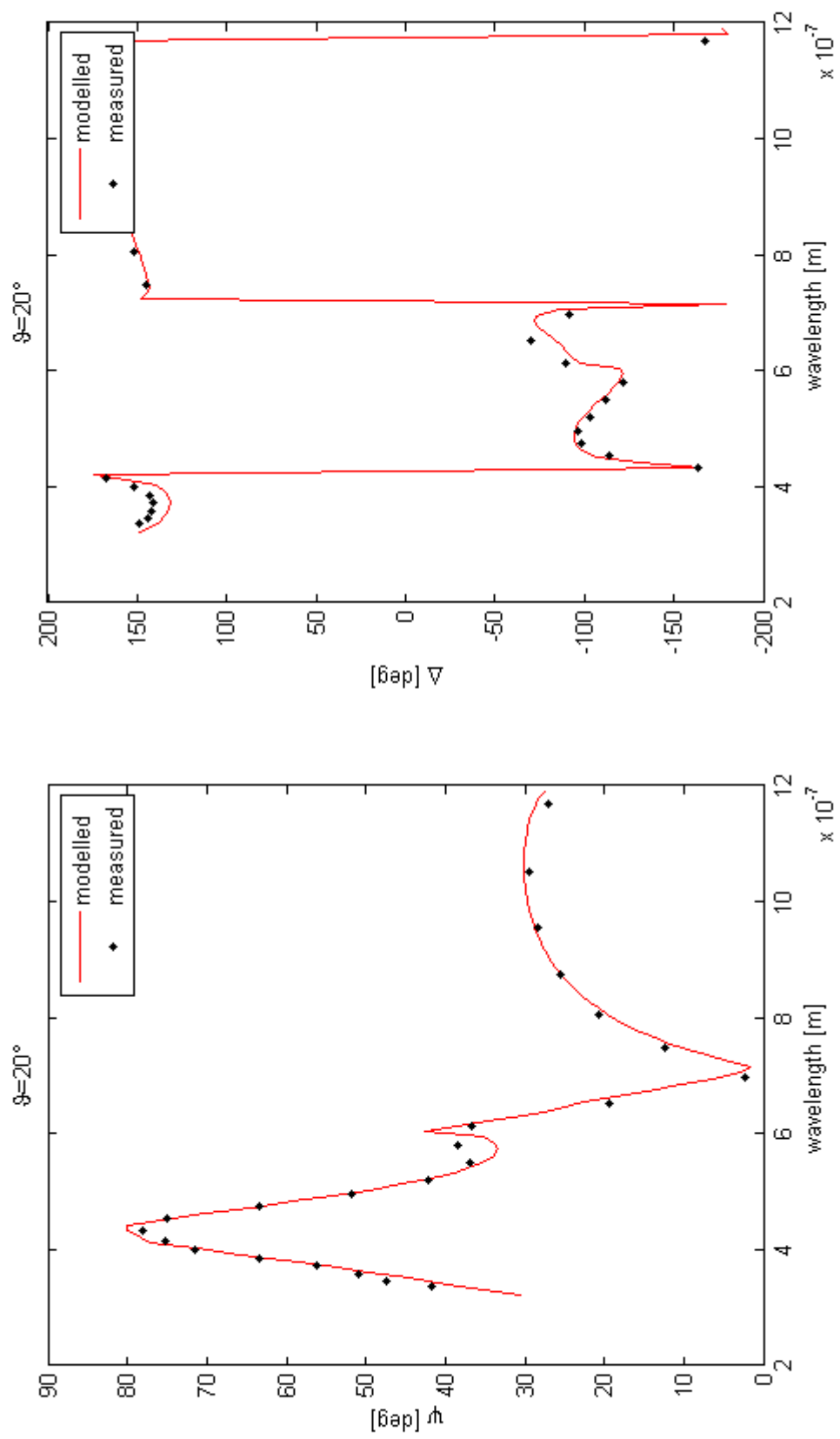
Figure 5.19: Ellipsometric parameters for the nickel sinusoidal grating, NVM
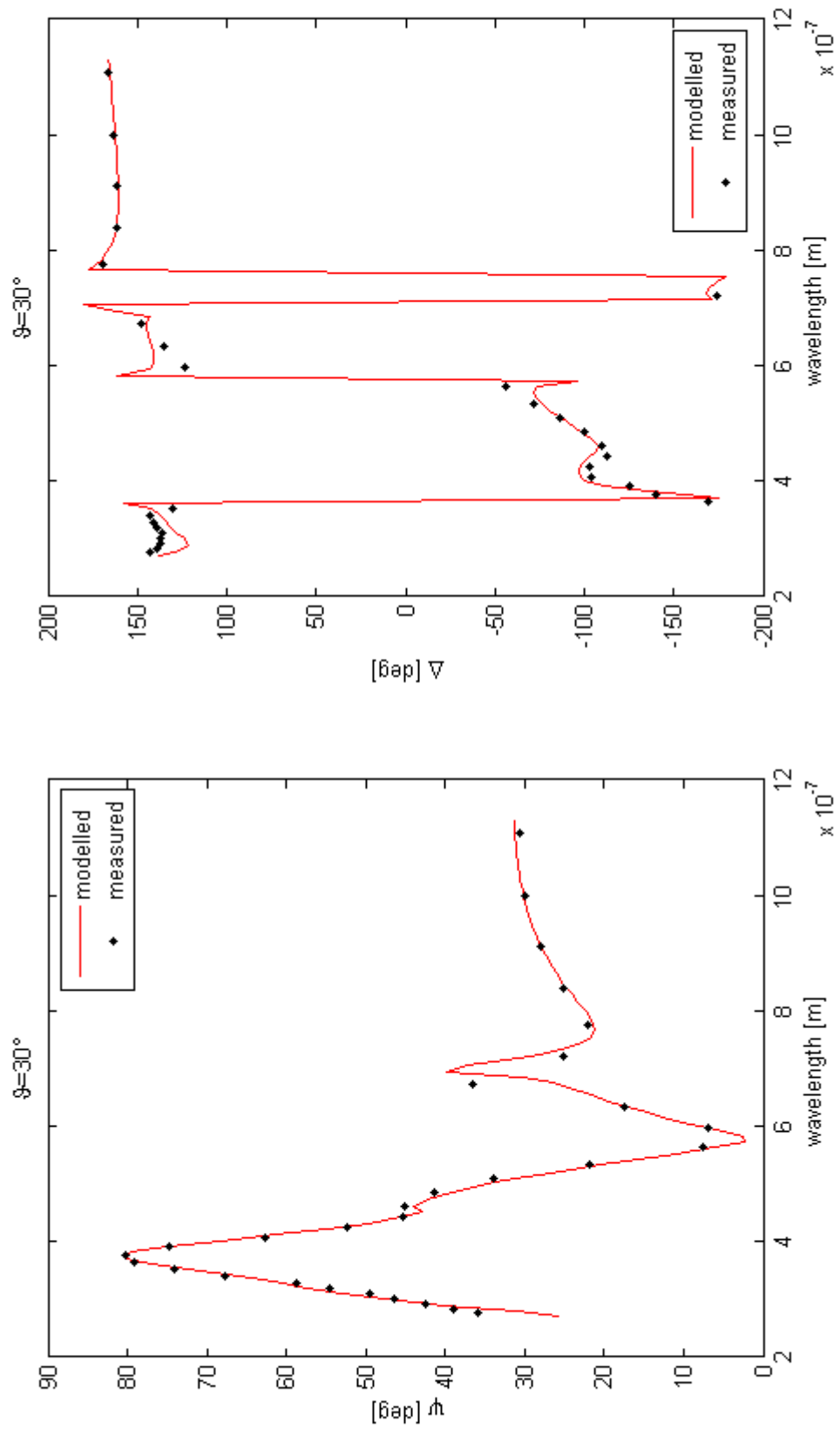
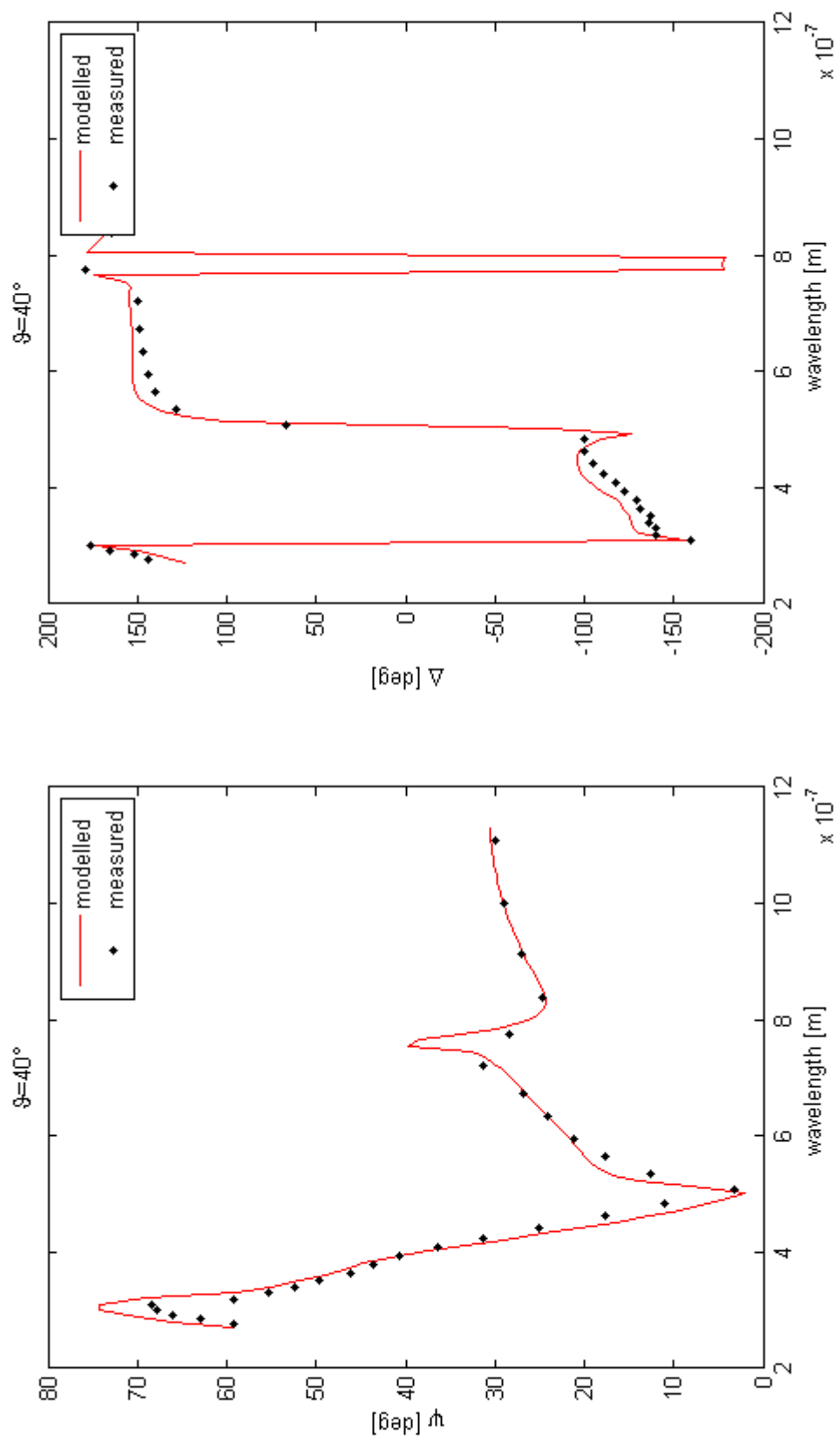Figure 5.20: Ellipsometric parameters for the nickel sinusoidal grating, NVM

Figure 5.21: Ellipsometric parameters for the nickel sinusoidal grating, NVM

Chapter 6

Conclusion

The main goals of this Thesis were in principle satisfied. The theoretical introduction was written in the first part of this Thesis. The short excursion to history, motivation, experimental setting and fitting process were written in Chapter 1. The basic concepts of the theory of periodic nanostructures, namely the Maxwell equations, radiation and interface conditions, the Rayleigh series, the energy conservation and the staircase approximation were described in Sections 2.1–2.4. Also the results concerning the uniqueness and existence of solutions and the FEM were shortly mentioned in Section 2.6. The Chapter 3 was devoted to the numerical method called RCWA. The main concepts of truncating the equations, matching the interface conditions and using the radiation conditions were described in Sections 3.1, 3.2. The problems regarding the truncation of the system to a finite dimension were introduced in Section 3.3, and the revolutionary work of Li [6] was shortly summarized. Afterwards the modified scheme from [9] was shortly described and our own modification of matching the interface conditions on the pseudo-interfaces using the Airy-like series was developed. The Chapter 4 contains the description of C-Method based on [27] together with some remarks related to a scaling of the system.

The hearth of this Thesis is the Chapter 5. The C-Method was implemented in MATLAB using the manual from the paper [27]. The existing implementation of the LMT method was afterwards tested against the MATLAB implementation of the C-method and a modified implementation of the LMT using the factorization rules (NVM) and the Airy-like series. The C-method proved to give more accurate results for diffraction gratings made of a highly conducting material, but the results for deep gratings were not satisfactory. The next step would be to extend this new implementation to coated gratings, but for time reasons it was skipped. The LMT scheme proved to be insufficient in modeling the optical response of deep metallic gratings. On the other hands, the Normal Vector Method supplemented with our modification of Airy-like series was tested on two sinusoidal gratings confirming the fast convergence even for metallic grating, which is in agreement with the results in the paper [9], [7]. All of this was broadly discussed in Sections 5.1,5.2. Also it would be good to supplement the implementation of the NVM with the subroutine for plotting the intensity of the fields and verify that the results corresponds to the Fig. 5.6, 5.2.

Finally, the implementation of the C-Method and NVM were compared to experimental measurements provided by the supervisor giving a good match in the ellipsometric parameters ψ and Δ. The results and graphs together with related comments were written in Section 5.3.

Bibliography

[1] Wood, R.W. (1902). On a remarkable case of uneven distribution of light in a diffraction grat- ingspectrum. Philos. Mag. 4, 396402

[2] Raman, C. V. (1934) The origin of the colours in the plumage of birds, Proc. of the Indian Academy of Sciences - Section A 1, 1–7

[3] Raman, C. V. (1935) The diffraction of light by high frequency sound waves: Part I., II., Proc. of the Indian Academy of Sciences - Section A 4, 406–420

[4] Raman, C. V. (1936) The diffraction of light by high frequency sound waves: Part III.–V., Proc. of the Indian Academy of Sciences - Section A 4, 75–84, 119–125, 459–465

[5] Li, L. (1993). Multilayer modal method for diffraction gratings of arbitrary profile, depth, and permittivity. Journal of the Optical Society of America A, 10(12), 2581. doi:10.1364/josaa.10.002581

[6] Li, Lifeng. (1996) Use of Fourier series in the analysis of discontinuous periodic structures. Journal of the Optical Society of America A 13, no. 9: 1870. doi:10.1364/josaa.13.001870.

[7] Popov, E., & Neviére, M. (2000). Grating theory: new equations in Fourier space leading to fast converging results for TM polarization. Journal of the Optical Society of America A, 17(10), 1773. doi:10.1364/josaa.17.001773

[8] Granet, G., & Guizal, B. (1996). Efficient implementation of the coupled-wave method for metallic lamellar gratings in TM polarization. Journal of the Optical Society of America A, 13(5), 1019. doi:10.1364/josaa.13.001019

[9] Gushchin, I., & Tishchenko, A. V. (2010). Fourier modal method for relief gratings with oblique boundary conditions. Journal of the Optical Society of America A, 27(7), 1575. doi:10.1364/josaa.27.001575

[10] Kim, H., Park, G., & Kim, C. (2012). Investigation of the Convergence Behavior with Fluctuation Features in the Fourier Modal Analysis of a Metallic Grating. Journal of the Optical Society of Korea, 16(3), 196-202. doi:10.3807/josk.2012.16.3.196

[11] Chandezon, J., Raoult, G., & Maystre, D. (1980). A new theoretical method for diffraction gratings and its numerical application. Journal of Optics, 11(4), 235-241. doi:10.1088/0150-536x/11/4/005

[12] R. Antoš (2006) Diffraction on Laterally Structured Aniosotropic Periodic Systems. Ph.D. Thesis. url: https://is.cuni.cz/webapps/zzp/detail/40196/29105862/

[13] Antoš, R., Pištora, J., Mistrík, J., Yamaguchi, T., Yamaguchi, S., Horie, M., . . . Otani, Y. (2006). Convergence properties of critical dimension measurements by spectroscopic ellipsometry on gratings made of various materials. Journal of Applied Physics, 100(5), 054906. doi:10.1063/1.2337256

[14] Hench, J. & Strakoš, Z. The RCWA Method — A Case Study with Open Questions and Perspec- tives of Algebraic Computations. Electronic Transactions on Numerical Analysis, no. 31 (2008). doi:10.1.1.379.8973

[15] Popov, E. (2013). Gratings: theory and numeric applications. Université dAix-Marseille: Institut Fresnel. url: http://www.fresnel.fr/files/gratings/Second-Edition/

[16] Ashcroft, N. W., Mermin, N. D., & Wei, D. (2016). Solid state physics. Singapore: Cengage Learning.

[17] Peatross, Justin, and Michael Ware. Physics of Light and Optics. Provo, UT: Brigham Young University, Department of Physics, 2015.

[18] Neviére, M., & Popov, E. (2003). Light propagation in periodic media: differential theory and design. New York: Marcel Dekker.

[19] Popov, E., Neviére, M., Gralak, B., & Tayeb, G. (2002). Staircase approximation validity for arbitrary-shaped gratings. Journal of the Optical Society of America A, 19(1), 33. doi:10.1364/josaa.19.000033

[20] Bao, G., Dobson, D. C., & Cox, J. A. (1995). Mathematical studies in rigorous grating theory. Journal of the Optical Society of America A, 12(5), 1029. doi:10.1364/josaa.12.001029

[21] Elschner, J., & Hu, G. (2010). Variational approach to scattering of plane elastic waves by diffraction gratings. Mathematical Methods in the Applied Sciences. doi:10.1002/mma.1305

[22] Bao, Gang, Lawrence Cowsar, and Wen Masters. Mathematical Modeling in Optical Science. Philadelphia: SIAM, Society for Industrial and Applied Mathematics, 2001.

[23] Bao, Gang. (1994). Diffractive optics in periodic structures: The TM polarization. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/4436.

[24] Dobson, D. C. (1993). Optimal design of periodic antireflective structures for the Helmholtz equation. European Journal of Applied Mathematics, 4(04). doi:10.1017/s0956792500001169

[25] Solano, M. E., Faryad, M., Lakhtakia, A., & Monk, P. B. (2014). Comparison of rigorous coupled-wave approach and finite element method for photovoltaic devices with periodically corrugated metallic backreflector. Journal of the Optical Society of America A, 31(10), 2275. doi:10.1364/josaa.31.002275

[26] Nesterenko, D. V. (2011). Modeling of diffraction of electromagnetic waves on periodic inhomogeneities by a finite element method coupled with the Rayleigh expansion. Optoelectronics, Instrumentation and Data Processing, 47(1), 68-75. doi:10.3103/s8756699011010109

[27] Li, L., Chandezon, J., Granet, G., & Plumey, J. (1999). Rigorous and efficient grating-analysis method made easy for optical engineers. Applied Optics, 38(2), 304. doi:10.1364/ao.38.000304

[28] Li, L. (1999). Justification of matrix truncation in the modal methods of diffraction gratings. Journal of Optics A: Pure and Applied Optics, 1(4), 531-536. doi:10.1088/1464-4258/1/4/320

[29] Edee, K., Plumey, J., & Chandezon, J. (2013). On the RayleighFourier method and the Chandezon method: Comparative study. Optics Communications, 286, 34-41. doi:10.1016/j.optcom.2012.08.088

[30] Antos, R., Ohlidal, I., Franta, D., Klapetek, P., Mistrik, J., Yamaguchi, T., & Visnovsky, S. (2005). Spectroscopic ellipsometry on sinusoidal surface-relief gratings. Applied Surface Science, 244(1-4), 221-224. doi:10.1016/j.apsusc.2004.09.142