Theory and Implementation of an End-to-End Radio Occultation Simulator

Joel Rasch

Technical Report, Earth and Space Sciences
No. 10, 2014

Chalmers University of Technology
Gothenburg, Sweden

Acknowledgments

This report was produced within the radio occultation subproject of the "Predikterbar prestanda för GNSS-applikationer"-project. Which was a cooperation between Chalmers University of Technology, RUAG Space AB, and OHB Sweden, financed by the Swedish Space Board within the second National Space Research Program (NRFP-2). People that have contributed to this report in different ways are Magnus Bonnedal, Anders Carlström, Patrick Eriksson, Jacob Christensen, and Oscar Isoz.

Contents

Chapter 1

Introduction

Radio occultation (RO) techniques have been developed for many years to study planetary atmospheres. The application to the Earth's atmosphere of these limb-sounding techniques provides an approach for global scale monitoring of tropospheric/ stratospheric temperature, pressure, and humidity profiles with high accuracy and vertical resolution, as well as ionospheric electron density profiles and scintillation properties. When a precision GNSS receiver on board a low earth orbiting (LEO) satellite tracks GNSS signals that are observed to rise or set through the atmosphere (see Fig. 1.1), the arrival time of the received signal is delayed due to the refractive bending and the retardation as it passes the atmosphere. By measuring the accumulated signal carrier phase delay on two different frequencies, the neutral (i.e. not due to the ionosphere) bending angle can be derived. Under the assumption of a spherically symmetric atmosphere it is then possible to determine the neutral atmosphere refractive index profile. The atmospheric parameters can be derived using fundamental physical relations, such as the relation between refractivity and dry air density, the equation of state, and the equation for hydrostatic equilibrium.

EUMETSAT is since 2006 flying an RO instrument named GRAS on board the MetOp satellites [1]. It has since 2007 provided operational data for numerical weather prediction (NWP) and RO measurements has in a few years become one of the most important data for NWP.

In this report we present the results of a project aimed at developing an end-to-end RO simulator. The project was a cooperation between Chalmers University of Technology, RUAG Space AB, and OHB Sweden, and financed by the Swedish Space Board within the second National Space Research Program (NRFP-2). The simulator includes a wave optics propagator (WOP) capable of simulating realistic GNSS signals as they propagate through the Earth's atmosphere using realistic refractivity, a routine for propagating the

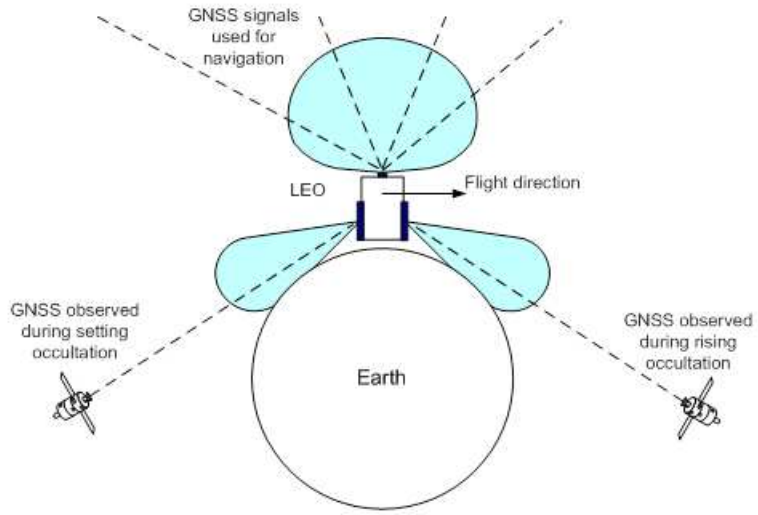

Figure 1.1: Instrument antenna beams.

signal to LEO, a routine for adding noise to the received signal, and a routine for interpreting the signals in the terms of bending angle and impact height (in our case achieved using phase matching). In this report we go through the theory behind each of these techniques, we describe how to implement them in Matlab, and we evaluate the accuracy of each step. In addition to this, and in order to evaluate the results from the code, it was necessary to make a detailed investigation into geometrical optics (GO) and the Abel transform. Mainly to find a fast and accurate way to calculate the results from geometrical optics, but as a spin-off we also present a small investigation into the problem with Abel inversion for super-refractive refractivity profiles.

The report is organized in chapters detailing: the theory behind multiple phase screen (MPS) propagation; the implementation of the MPS method in Matlab; the results on the last phase screen for exponential profiles, and exponential profiles containing local sharp gradients and super-refractive regions; the theory of how to propagate the signal to orbit using a diffractive integral, and how to implement the technique; the unraveling of the signal phase using an estimated generic signal; inversion through phase matching; detailed comparisons between simulations and GO in LEO; and finally a presentation of the numerical techniques used to calculate the Abel transforms, and the problem appearing in the case of super-refractive profiles.

Chapter 2

Theoretical formulation of MPS

2.1 Introduction

In this chapter we shall present the theory behind the MPS, or Fourier split-step technique (FSS). Although the method is quite well known and frequently used for solving Schrödinger type equations in connection with quantum mechanics and wave propagation in non-linear media, there is a marked lack of complete treatments of the problem in the framework of atmospheric wave propagation to be found in the literature. Therefore some time in this chapter is spent on trying to work out the limitations of the technique. The theoretical treatment presented here is far from complete. In particular, when performing the Fourier transform over the vertical coordinate (perpendicular to the wave propagation direction) one assumes that the refractive index has varies slowly with the vertical coordinate. It is unclear under what circumstances this approximation is allowed, and this question has to be deferred due to its mathematical complexity.

2.2 The paraxial approximation

The most general formulation of the wave equation for the electric field in linear media is [2]

$$\nabla^2 \bar{E} - \epsilon\mu \frac{\partial^2 \bar{E}}{\partial t^2} + \nabla(\bar{E} \cdot \nabla(\ln \epsilon)) + \nabla(\ln \mu) \times (\nabla \times \bar{E}) = 0 \tag{2.1}$$

where $\bar{E}$ is the electric field vector, ϵ the permittivity, and μ the permeability. In air we can assume with great accuracy that $\mu \approx \mu_0 = \text{const.}$. Furthermore we can use $n^2/c^2 = \epsilon\mu$, where n is the refractive index, and we get

$$\nabla^2 \bar{E} - \frac{n^2}{c^2}\frac{\partial^2 \bar{E}}{\partial t^2} + 2\nabla(\bar{E} \cdot \nabla(\ln n)) = 0 \tag{2.2}$$

If we restrict our solution to the steady state case we can use a harmonic description for the field, $\bar{E} = \mathcal{R}(\bar{E}_0 \exp(-i\omega t))$, where ω is the field radian frequency. And using $k_0 = \omega/c$ we get

$$\nabla^2 \bar{E}_0 + k_0^2 n^2 \bar{E}_0 + 2\nabla(\bar{E}_0 \cdot \nabla(\ln n)) = 0 \tag{2.3}$$

Now we wish to model the propagation of the field in the atmosphere, where n is actually changing, but Eq. (2.3) is too complicated to use with ease. Fortunately, under the condition that the refractive index changes slowly in comparison to the electric field we can neglect the third term. The precise condition for this approximation to be valid can be found by using l_n as the length scale of variations in n, and λ, the field wavelength, as the length scale for variations in $\bar{E}_0$. We find the order of magnitude of the different terms to be

$$\nabla^2 \bar{E}_0 \sim k_0^2 E_0 \tag{2.4}$$

$$k_0^2 n^2 \bar{E}_0 \sim k_0^2 E_0 \tag{2.5}$$

$$2\nabla(\bar{E}_0 \cdot \nabla(\ln n)) = 2\nabla\left(\bar{E}_0 \cdot \frac{\nabla n}{n}\right) \sim k_0 E_0 \frac{1}{n}\frac{n}{l_n} = \frac{k_0}{l_n} E_0 \tag{2.6}$$

Hence, we can neglect the third term when $k_0^2 E_0 \gg k_0 E_0 / l_n$, or $l_n \gg \lambda$. In the atmospheric conditions we are interested in, super-refractivity occurs when $\partial n/\partial r < -1.57 \times 10^{-7}$ m^{-1} (see chapter 9), so we can put $l_n \sim 10^6$ m as a worst case scale length for n. Furthermore, the GPS signals corresponds to wavelengths around $\lambda \sim 10^{-1}$ m. This means that $\lambda \ll l_n$, and the third term can be safely neglected for any conditions occurring in the Earth's atmosphere. So for our applications we can use

$$\nabla^2 \bar{E}_0 + k_0^2 n^2 \bar{E}_0 = 0 \tag{2.7}$$

But in Cartesian coordinate system this equation must be fulfilled independently by all field components. And the full solution to the wave propagation problem can be found by considering the scalar Helmholtz equation (this includes the solution for the magnetic field)

$$\nabla^2 \psi + k_0^2 n^2 \psi = 0 \tag{2.8}$$

Furthermore we will assume a wave that propagates predominantly in the z-direction. The treatment is simplified significantly by using $\psi(x,y,z) = u(x,y,z)\exp(ik_0 z)$, which changes Eq. (2.8) into

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + 2ik_0\frac{\partial u}{\partial z} + k_0^2(n^2-1)u = 0 \tag{2.9}$$

The paraxial approximation consists of disregarding the second order derivative with respect z. To see when this is allowed we consider a plane wave propagating in free space with an angle θ with respect to the z-axis. The wave is described by

$$\psi(x,y,z) = \psi_0\exp(i\bar{k}_0\cdot\bar{r}) = \tag{2.10}$$

$$\psi_0\exp(ik_0(\sin\theta\cos\phi x + \sin\theta\sin\phi y + \cos\theta z)) \tag{2.11}$$

Now we use the approximations $\sin\theta \approx \theta$, and $\cos\theta \approx 1-\theta^2/2$, which gives

$$\psi(x,y,z) \approx \psi_0\exp(ik_0 z)\exp(ik_0(\theta\cos\phi x + \theta\sin\phi y) - ik_0\frac{\theta^2}{2}z)) \tag{2.12}$$

We can see that the role of $u(x,y,z)$ in Eq. (2.9) is now played by $\phi_0\exp(ik_0(\theta\cos\phi x + \theta\sin\phi y) - ik_0\frac{\theta^2}{2}z))$. We can therefore use Eq. (2.12) to find the order of magnitude of the different derivatives in Eq. (2.9). The derivatives become

$$\frac{\partial u}{\partial z} \approx -ik_0\frac{\theta^2}{2}u \tag{2.13}$$

$$\frac{\partial^2 u}{\partial z^2} \approx -k_0^2\frac{\theta^4}{4}u \tag{2.14}$$

$$\frac{\partial^2 u}{\partial y^2} \approx -k_0^2\theta^2\sin\phi^2 u \tag{2.15}$$

$$\frac{\partial^2 u}{\partial x^2} \approx -k_0^2\theta^2\cos\phi^2 u \tag{2.16}$$

The joint condition for neglecting the second derivative with respect to z in Eq. (2.9) in comparison to the other terms then becomes $\theta \ll 2$, or expressed in degrees $\theta_{\text{deg.}} \ll 115^o$. The paraxial approximation should thus be quite exact around $\theta \approx 10^o$. Provided all the conditions above are satisfied we can finally state the paraxial approximation of the scalar Helmholtz equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + 2ik_0\frac{\partial u}{\partial z} + k_0^2(n^2 - 1)u = 0 \tag{2.17}$$

In the code described in this paper we disregard variations of the wavefront in the horizontal direction, which means that we can neglect the x-part. This finally yields

$$\frac{\partial^2 u}{\partial y^2} + 2ik_0\frac{\partial u}{\partial z} + k_0^2(n^2 - 1)u = 0 \tag{2.18}$$

which will be the basis for all our calculations.

Finally it should be noted that when neglecting the second derivative with respect to z, one also neglects the backwards propagating wave. Fortunately, that wave is of no interest in radio occultation.

2.3 The multiple phase screen technique

In order to solve Eq. (2.18) in a complicated atmosphere one can use different methods, but the most effective seems to be the multiple phase screen, or Fourier split-step technique. There are many different formulations of this technique, but the main trick is always the same [3–7]. The medium is divided into a number of screens extending in the vertical direction. The wave is propagated between the screens using the solution for constant n, and the effect of the inhomogeneous n is included by multiplication when the wave reaches the screen.

To implement this technique one performs spatial Fourier transforms in the y-direction, viz

$$\tilde{u}(q, z) = \mathcal{F}(u(y, z)) = \int_{-\infty}^{\infty} u(y, z)\exp(-iqy)dy \tag{2.19}$$

$$u(y, z) = \mathcal{F}^{-1}(\tilde{u}(q, z)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{u}(q, z)\exp(iqy)dq \tag{2.20}$$

$$\tag{2.21}$$

Applying this Fourier transform to the entire Eq. (2.18), and assuming the n dependence on y can be neglected

$$-q^2\tilde{u} + 2ik_0\frac{\partial \tilde{u}}{\partial z} + k_0^2(n^2 - 1)\tilde{u} = 0 \tag{2.22}$$

This is a differential equation that lacks a general solution due to the fact that n as a function of z is in general quite complicated. If n was constant the solution would be

$$\tilde{u}(q,z) = \tilde{u}(q,0)\exp\left(i\frac{k_0}{2}(n^2-1)z\right)\exp\left(-\frac{i}{2k_0}q^2 z\right) \tag{2.23}$$

And in normal space

$$u(y,z) = \mathcal{F}^{-1}(\tilde{u}(q,z)) = \tag{2.24}$$

$$= \exp\left(i\frac{k_0}{2}(n^2-1)z\right)\mathcal{F}^{-1}\left(\exp\left(-\frac{i}{2k_0}q^2 z\right)\mathcal{F}(u(y,z))\right) \tag{2.25}$$

The FSS or MPS-technique now consists of approximating n as constant and homogeneous over small intervals, Δz, in the z-direction. The solution can then be advanced a step Δz using

$$u(y,z+\Delta z) = \exp\left(i\frac{k_0}{2}(n^2-1)\Delta z\right)\mathcal{F}^{-1}\left(\exp\left(-\frac{i}{2k_0}q^2\Delta z\right)\mathcal{F}(u(y,z))\right) \tag{2.26}$$

To investigate what kind of relative error we introduce over the interval Δz we consider the propagation over a small region where the refractive index is given by

$$n(z) \approx n(z=0) + z\frac{\partial n}{\partial z}|_{z=0} \tag{2.27}$$

Inserting this into Eq. (2.22) and solving yields

$$\tilde{u}(q,z) = \tilde{u}(q,0)\exp\left(i\frac{k_0}{2}(n_0^2-1)z\right)\exp\left(-\frac{i}{2k_0}q^2 z\right)\exp\left(\frac{i}{2}n_0 k_0\frac{\partial n}{\partial z}|_{z=0}z^2\right) \tag{2.28}$$

where $n_0 = n(z=0)$. Comparing with Eq. (2.23) we see that the relative error introduced during a small step Δz is on the order of the factor

$$\exp\left(\frac{i}{2}n_0 k_0\frac{\partial n}{\partial z}|_{z=0}\Delta z^2\right) \approx 1 + \frac{i}{2}n_0 k_0\frac{\partial n}{\partial z}|_{z=0}\Delta z^2 \tag{2.29}$$

Hence, to use the multiple phase screen technique without introducing errors because of the finite number of screens we need to be sure that

$$k_0 \frac{\partial n}{\partial z} \Delta z^2 \ll 1 \tag{2.30}$$

where we have used the fact that $n \approx 1$. The local y-axis makes an angle ξ with respect to the radius vector drawn from the center of the Earth, and the derivative in the z-direction is thus given by

$$\frac{\partial n}{\partial z} \approx \frac{\partial n}{\partial r} \sin \xi \tag{2.31}$$

Using typical numbers for the different terms involved, $k_0 \sim 30$ m^{-1}, $\partial n/\partial r \sim 5 \times 10^{-8}$ m^{-1}, and $\xi \approx 14^o$ (referring to Fig. 3.6 with $H = 100$ km, $L_y = 300$ km, and $R_E = 6371$ km), the requirement becomes

$$\Delta z \ll 2000 \ \text{m} \tag{2.32}$$

But in the presence of super-refraction $\partial n/\partial z \sim 2 \times 10^{-7}$ m^{-1}, and

$$\Delta z \ll 1000 \ \text{m} \tag{2.33}$$

Of course, these approximations are based on a rather small angle of 1 degree. For larger angles the conditions become more restrictive.

It is not immediately obvious what kind of error we introduce by only accounting for the refractivity inhomogeneity in the z-direction. This point is discussed e.g. in Ref. [3], but we will not dwell upon it further, as it appears to be a question of some mathematical intricacy.

One further remark is in order. Using this proceedure for advancing the waves will lead to so-called free space losses, which is nothing stranger than the spreading of the wave, and the decay of the amplitude. If the two dimensional spatial Fourier transform is used, the losses will be those of a spherical wave $\propto 1/r$. If on the other hand, only the y-direction is taken into account (as in this report), there will be no spreading in the x-direction, and the amplitude decays as a cylindrical wave, $\propto 1/\sqrt{r}$. The cylindrical wave is generally described using Bessel functions, but far away from the source ($r \gg \lambda$) one can use

$$u(r) = \frac{\exp(ikr)}{\sqrt{r}} \tag{2.34}$$

2.4 A slightly more accurate MPS technique

The method used in the previous section is quite exact, and for our purposes it is not really necessary to look for more precise methods. It is however

quite common to use a slightly different formulation [3,8], where one does not employ the paraxial approximation. One starts with the Helmholtz equation (2.9) without taking account of the x-direction

$$\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + 2ik_0 \frac{\partial u}{\partial z} + k_0^2(n^2 - 1)u = 0 \tag{2.35}$$

One then performs the Fourier transform to yield

$$-q^2 + \frac{\partial^2 \tilde{u}}{\partial z^2} + 2ik_0 \frac{\partial \tilde{u}}{\partial z} + k_0^2(n^2 - 1)\tilde{u} = 0 \tag{2.36}$$

If one assumes that the z-dependence can be described by

$$u(z) = u_0 \exp(az) \tag{2.37}$$

where a is some constant, we get

$$-q^2 + a^2 + 2ik_0 a + k_0^2(n^2 - 1) = 0 \tag{2.38}$$

With the solution for a

$$a = -ik_0 \pm ik_0 \sqrt{n^2 - \frac{q^2}{k_0^2}} \tag{2.39}$$

Since we have assumed that the main direction of propagation is in the z-direction, we know that $q/k_0 < 1$, and $n > 1$, which means that the term inside the square root is positive. This in turn means that the positive sign before the square root corresponds to a wave moving in the z-direction, whereas the negative sign corresponds to the $-z$-direction. We will limit our interest to the wave moving in the positive z-direction. For our purposes we can assume $n - 1 \ll 1$ and $q/k_0 \ll 1$, which means that we get

$$a = ik_0(n - 1) + ik_0 \left(\sqrt{1 - \frac{q^2}{k_0^2}} - 1 \right) \tag{2.40}$$

The algorithm for advancing the wavefront a small step Δz is then

$$u(y, z + \Delta z) = \exp\left(ik_0(n - 1)\Delta z\right) \mathcal{F}^{-1} \left(\exp\left(ik_0 \left(\sqrt{1 - \frac{q^2}{k_0^2}} - 1 \right) \Delta z \right) \mathcal{F}(u(y, z)) \right) \tag{2.41}$$

We can retrieve Eq. (2.26) by employing the approximations

$$(n^2 - 1)/2 \approx n - 1 \tag{2.42}$$

$$\sqrt{1 - q^2/k_0^2} \approx 1 - \frac{q^2}{2k_0^2} \tag{2.43}$$

Eq. (2.41) is supposed to be slightly more accurate than Eq. (2.26). However, to the knowledge of the author the difference is very small.

Chapter 3

Numerical implementation of MPS

3.1 Introduction

When implementing the multiple phase screen technique numerically three main questions will arise: how to perform the discrete Fourier transform correctly; how to take into account borders (e.g. the Earth's surface); and how to avoid aliasing? We will deal with these things separately.

3.2 The discrete Fourier transform

In our theoretical treatment we used the Fourier transforms

$$\tilde{u}(q,z) = \int_{-\infty}^{\infty} u(y,z)\exp(iqy)dy \tag{3.1}$$

$$u(y,z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{u}(q,z)\exp(-iqy)dq \tag{3.2}$$

But instead of using q we can use $q = 2\pi f_y$, which yields

$$\tilde{u}(f_y,z) = \int_{-\infty}^{\infty} u(y,z)\exp(2\pi i f_y y)dy \tag{3.3}$$

$$u(y,z) = \int_{-\infty}^{\infty} \tilde{u}(f_y,z)\exp(-2\pi i f_y y)df_y \tag{3.4}$$

The numerical implementation has to take place in a finite domain, let us use a vector in the y-plane having length L_y, where the coordinates are given by

$$y[n_y] = n_y \Delta y \quad \text{where} \quad n_y = -M/2, -M/2+1, ..., M/2-1 \tag{3.5}$$

And $\Delta y = L_y/M$. When using the fast Fourier transform, the algorithm works most efficiently when M is a power of 2, therefore we shall always assume that M is even, resulting in integer values for $M/2$. The Fourier transform in the discrete domain is found by forming the Riemann sum

$$\tilde{u}(f_y, z) = \sum_{n_y=-M/2}^{M/2-1} u(y[n_y], z) \exp(2\pi i f_y y[n_y]) \Delta y \tag{3.6}$$

$$\tag{3.7}$$

and inserting the correct value for $y[n_y]$

$$\tilde{u}(f_y, z) = \sum_{n_y=-M/2}^{M/2-1} u(n_y \Delta y, z) \exp\left(2\pi i f_y \frac{n_y L_y}{M}\right) \Delta y \tag{3.8}$$

$$\tag{3.9}$$

Now the discrete spatial frequency must find a suitable definitions. The spatial step is Δy, which implies that the highest frequency we can resolve (due to the Nyqvist criterion) is $1/(2\Delta y)$. It is most practical if the frequency representation has as many vector points as the spatial representation, wherefore the frequency should be defined according to

$$f_y[m_y] = m_y \Delta f_y = \frac{m_y}{M \Delta y} = \frac{m_y}{L_y} \quad \text{where} \quad m_y = -M/2, -M/2+1, ..., M/2-1 \tag{3.10}$$

Using the definitions above the discrete Fourier transform becomes

$$\tilde{u}(f_y[m_y], z) = \sum_{n_y=-M/2}^{M/2-1} u(y[n_y], z) \exp\left(2\pi i \frac{m_y n_y}{M}\right) \Delta y \tag{3.11}$$

$$u(y[n_y], z) = \frac{1}{M} \sum_{m_y=-M/2}^{M/2-1} u(f_y[m_y], z) \exp\left(-2\pi i \frac{m_y n_y}{M}\right) \frac{1}{\Delta y} \tag{3.12}$$

During the progression between phase screens the effect of free space spreading and diffraction effects are incorporated by multiplication with

$$\exp\left(i\frac{k_0}{2}(n^2-1)\Delta z\right) \tag{3.13}$$

and

$$\exp\left(-\frac{i}{2k_0}q^2\Delta z\right) \tag{3.14}$$

In the discrete representation these expressions change into

$$\exp\left(i\frac{k_0}{2}(n(y[n_y],z)^2-1)\Delta z\right) \tag{3.15}$$

and

$$\exp\left(-\frac{i}{2k_0}\left(\frac{2\pi m_y}{L_y}\right)^2\Delta z\right) \tag{3.16}$$

3.3 Discrete Fourier transform in Matlab

Different software have different ways of implementing the discrete Fourier transform. Although Matlab uses a rather straightforward method, it is necessary to be aware of what it actually is doing for the results to be correct. When using the discrete Fourier transform one actually assumes a function that is periodic in the y-direction, with the period L_y. The natural way to represent the function is thus to include one full period on L_y, where the first position in the y vector (i.e. $i_y = 1$, where i_y represents the number of a vector element, going between 1 and M) indicates the start of the period, and the final positions the end of the period ($i_y = M$). Matlab uses a different representation, assuming that the first position ($i_y = 1$) belongs to the middle of the period, the middle to the last, and the last position to the middle. This is more easily explained using illustrations. In Fig. 3.1 we can see a function defined on the y-axis, and how Matlab interprets it.

Due to this peculiar feature of Matlab, if one uses the "natural" way of defining a function, it is necessary to shift the function before performing the discrete Fourier transform in Matlab. To do this one uses the function *fftshift()*, which works for one and two-dimensional Fourier transforms. After performing this shift it is straight forward to take the fast Fourier transform using *fft()*. In Eqs. (3.11) and (3.12) there appears two numerical factors: Δy and $1/(M\Delta y)$. Matlab only takes into account the $1/M$ factor, and if

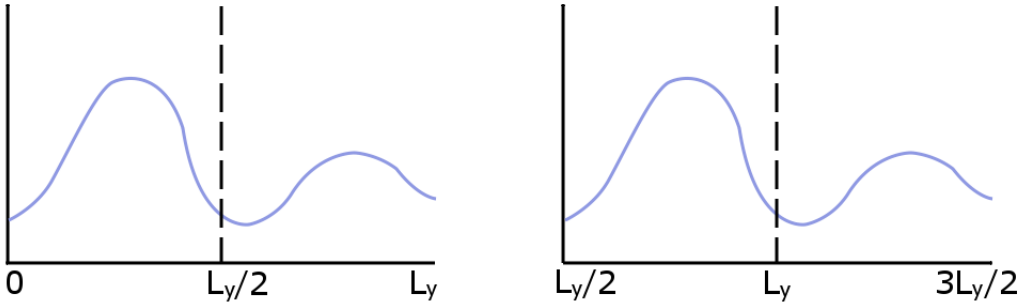

Figure 3.1: To the left, the "natural" way of defining a function on an interval. To the right, how Matlab interprets the function on the left.

needed the other factors will have to be introduced by hand. In our case all operations are linear, and we can forget about the factors Δy and $1/\Delta y$ since they cancel in the end. Before multiplying with Eq. (3.16) it is necessary to undo the previous shift using *ifftshift()*. After multiplication with (3.16) it is necessary to use *fftshift()* again before performing the inverse fast Fourier transform *ifft()*. After the inverse transform one has to use *ifftshift()* again, and after that one can finally multiply by Eq. (3.15).

3.4 Proper windowing

The discrete Fourier transform generates a function which is periodic in the y-direction. This will lead to problems during the progression between screens if not properly handled using a window function. In effect, using the DFT will actually simulate an infinite number of signals being emitted from an infinite number of satellites. Each wavefront starts in its own domain, in the case of our signal it resides mainly in $y \in [0, L_y]$. But this domain will be surrounded by other identical domains. During the progression the signals from these separate domains will start to bleed into each other, eventually creating a terrible mess. The solution for this is to create artificial absorbing boundaries, i.e. a window function. The window function should be smooth (otherwise it will generate unwanted higher frequencies in the wave), approach zero close to the edges, and have no imaginary component. There are of course a great number of ways to construct such a function, but for our purposes it seems that we need nothing more complicated than a function which decreases as a Gaussian close to the edges, viz

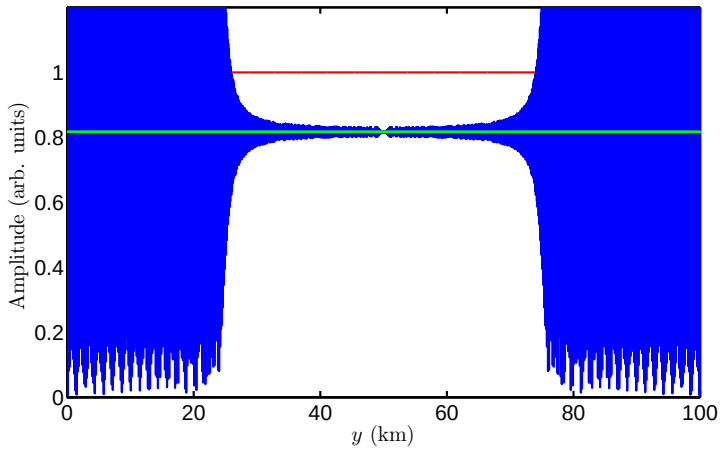

Figure 3.2: The wave amplitude without windowing. The red line is the amplitude on the first screen, the blue line the amplitude on the last screen. The green line is the true amplitude of the signal from the satellite on the last screen.

$$w = \begin{cases} \exp\left(-(\frac{y-a_1L_y}{b_1L_y})^2\right) & \text{if} \quad y < a_1L_y, \\ 1 & \text{if} \quad a_1L_y \leq y \leq L_y(1-a_1), \\ \exp\left(-(\frac{y-L_y(1-a_1)}{b_1L_y})^2\right) & \text{if} \quad L_y(1-a_1) < y. \end{cases} \tag{3.17}$$

As L_y should be around 100-300 km long, suitable values for a_1L_y and b_1L_y seems to be around $20-30$ km and 10 km respectively.

In Figs. 3.2-3.4 we can see a cylindrical wave as it propagates in the z-direction. In Figs. 3.2 and 3.3 we see the effect of no windowing. The adjacent wavefronts bleed into our calculation domain and superposes with the original wave. In Figs. 3.4 and 3.5 we see the same wavefront propagation when the Gaussian window above has been multiplied with the wave amplitude at each phase screen.

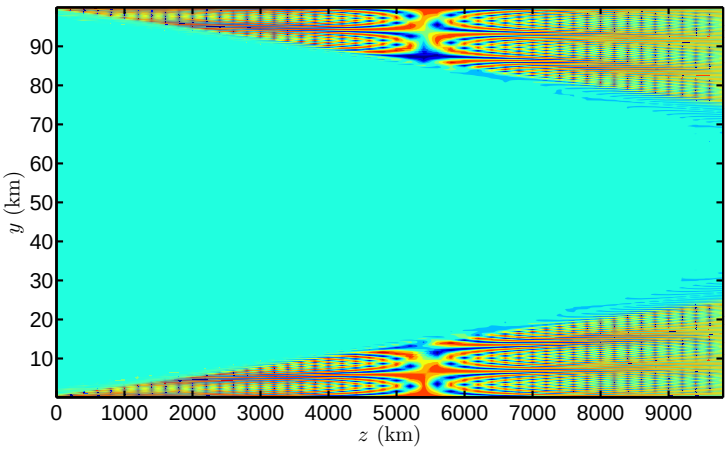

Figure 3.3: The evolution of the wave amplitude without windowing. As the wave progresses from the left side to the right side the adjacent wavefronts start to leak into our calculation box. Blue colors represent low amplitude, close to zero, and red colors represent high amplitude.

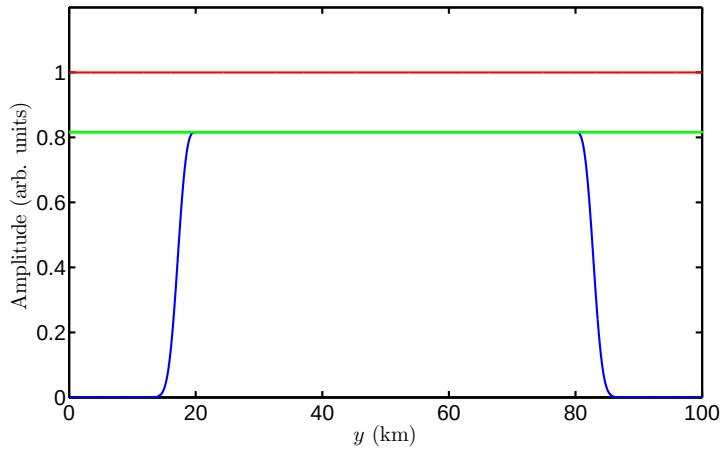

Figure 3.4: The wave amplitude with proper Gaussian windowing.

3.5 Geometry

The most simple geometry for performing the wave propagation is shown in Fig. 3.6. One prescribes a calculation box that has a certain area $L_y \times L_z$, and then one positions the Earth and the GNSS satellite with respect to this box. The wave propagates freely from the satellite to the first screen, which is the left vertical border of the box, wherefrom the multiple phase screen propagation begins. When the calculation reaches the last screen the

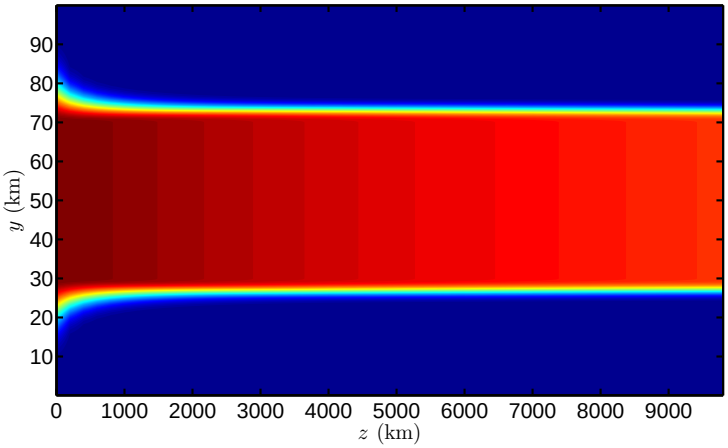

Figure 3.5: The wave amplitude with proper Gaussian windowing. Blue colors represent low amplitude, close to zero, and red colors represent high amplitude, in this case close to unity.

wave leaves the atmosphere, and one can calculate the bending angles and such directly, or propagate the wave to the receiving satellite using different techniques. In this report we have restricted ourselves to interpreting the information gathered on the last screen. The size of the calculation box is of course arbitrary, but if one wants to include the effect of the entire atmosphere and at the same time minimize the size, the proper choice is when the lower corners of the red box in Fig. 3.6 coincides with the outer black circle representing the top of the atmosphere. This gives L_z as a function of R_E, H, L_y, through

$$(R_E + H)^2 = (L_z/2)^2 + (R_E + H - L_y)^2 \tag{3.18}$$

or

$$L_z = 2\sqrt{2L_y(R_E + H) - L_y^2} \tag{3.19}$$

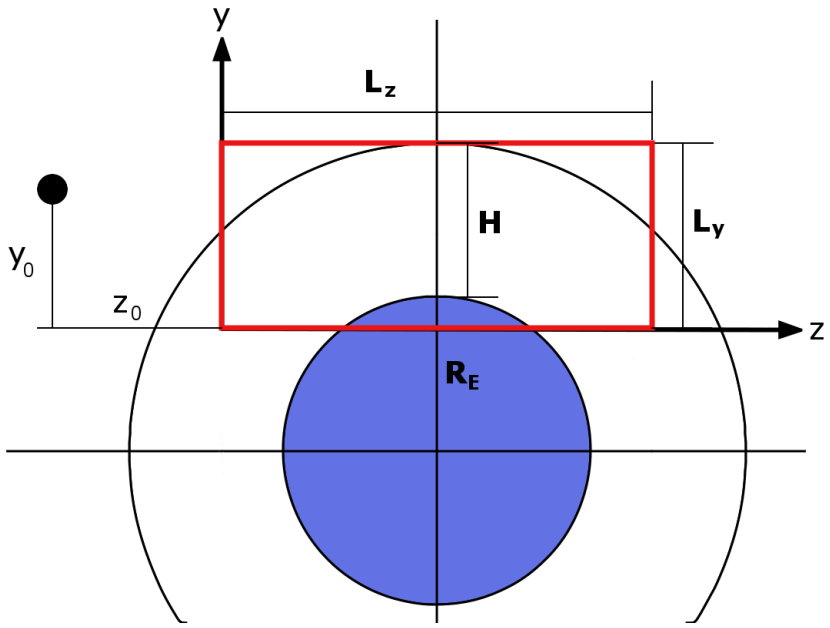

Figure 3.6: The geometry used in the simulations. The GNSS satellite is situated at the black dot. The multiple phase screen routine is performed within the red borders. The blue area represents the Earth, and the large black circle represents the top of the atmosphere.

3.6 Aliasing

Aliasing occurs when the wave contains spatial features which vary more rapidly than the Nyqvist frequency. The effect of this is that repetitive features appear in the signal which have no physical cause. To avoid this the sampling rate has to be sufficiently large. To make sure aliasing does not appear, one has to take into account the position of the transmitting satellite, the distance between screens, and the gradient of the refractive index. We formulate explicit conditions for avoiding aliasing below.

3.6.1 Position of the GNSS satellite

The GNSS satellite is generally assumed to emit cylindrical waves described by

$$u(r) = \frac{\exp(ik_0 r)}{\sqrt{r}} \tag{3.20}$$

where r is the distance from the source ($r = \sqrt{y^2 + z^2}$), and the time factor,

$\exp(-i\omega t)$, is excluded as usual. The satellite is a distance $r = \sqrt{y_0^2 + z_0^2}$ from the lower left vertex of the calculation box where the coordinate origin is located. When the vertical position of the satellite is $y_0 = L_y/2$, the wavefront will hit the first screen perpendicularly in the exact center of the screen. At the corners of the screen the wavefronts make a certain angle with the screen normal amounting to

$$\tan\theta = \frac{L_y}{2z_0} \tag{3.21}$$

when the satellite is not positioned in this way, the angle will differ between different corners, and the largest angle is found from

$$\tan\theta = \frac{d_{y0}}{z_0} \tag{3.22}$$

and

$$d_{y0} = \begin{cases} L_y - y_0 & \text{if} \quad y_0 < L_y/2, \\ y_0 & \text{if} \quad y_0 > L_y/2. \end{cases} \tag{3.23}$$

When these tilted wavefronts hit the screen they will give rise to patterns repeating themselves with a wavelength λ'. This wavelength is given by the angle θ as

$$\lambda' = \frac{\lambda}{\sin\theta} \tag{3.24}$$

For this wave structure to be sampled sufficiently often we need at least two vector points per wavelength λ' [4, 9]. Thus we need a resolution that is fine enough to fulfill

$$\Delta y < \frac{1}{2}\lambda' = \frac{\lambda}{2\sin\theta} \tag{3.25}$$

If computer power is in abundance it might be good to set Δy to the limiting value $\lambda/2$, which will ensure no aliasing even for rather strange satellite positions. The wavelength is around 20 cm, which means that the resolution in that case should be around 10 cm. For a screen height of 100 km this would mean at least 10^6 calculation points, i.e. $M > 10^6 \approx 2^{20}$. On the other hand, the paraxial approximation only works for $\theta < 10^o$, which means that the resolution will never have to be finer than 0.5 m. A height of 100 km then implies $M \approx 2 \times 10^5 \approx 2^{18}$. If one wishes to use a lower resolution, the exact criteria (3.25) that depends on the satellite position will have to

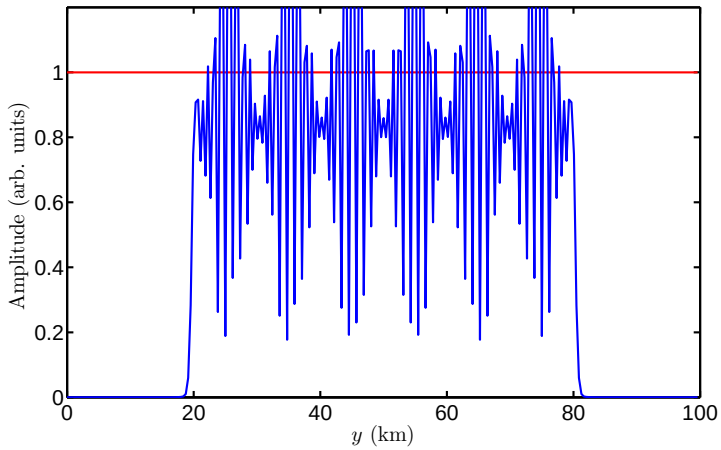

Figure 3.7: The aliasing effect of having too low vertical resolution. The satellite in this case is positioned at $y_0 = L_y/2$, and $z_0 = 20000$ km, transmitting in the L1 band.

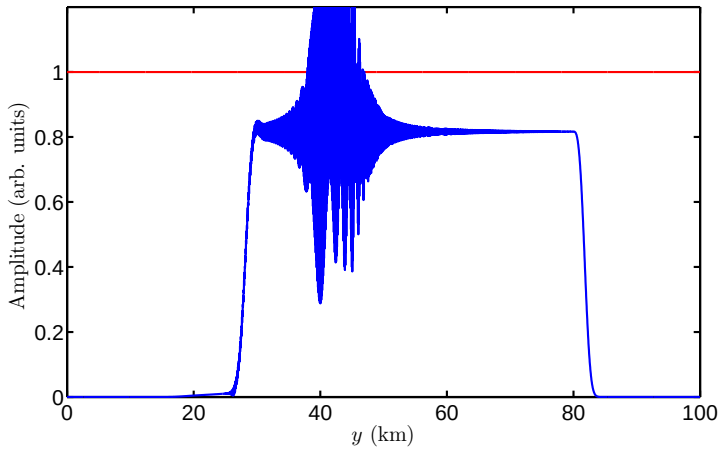

Figure 3.8: The aliasing effect of having too low vertical resolution. The satellite in this case is positioned at $y_0 = 0$, and $z_0 = 20000$ km, transmitting in the L1 band.

be used, and one should not forget the extra bending of the signal in the atmosphere which amounts to a few degrees.

In Fig. 3.7 we can see the effect of having too low resolution in the vertical direction. The signal forms a repetitive pattern of rather low frequency that has no physical cause.

3.6.2 Distance between screens

Knepp [5] noted that multiplication with the exponential factor (3.16)

$$\exp\left(-\frac{i}{2k_0}\left(\frac{2\pi m_y}{L_y}\right)^2 \Delta z\right) \tag{3.26}$$

might give rise to aliasing if the factor differs too much between adjacent grid points. The exponential factor gives rise to a spatial structures where the phase shift between to adjacent grid points is

$$\frac{1}{2k_0}\left(\frac{2\pi(m_y+1)}{L_y}\right)^2 \Delta z - \frac{1}{2k_0}\left(\frac{2\pi m_y}{L_y}\right)^2 \Delta z \tag{3.27}$$

This difference is maximized when $m_y = M/2$ (see Eq. (3.10)), so the maximum phase shift caused by this factor is

$$\frac{2\pi^2 \Delta z(M+1)}{k_0 L_y^2} \tag{3.28}$$

To fulfill the Nyqvist criterion we need this phase difference to be less than π. We can state the "Knepp" criterion as

$$\Delta z < \frac{L_y \Delta y}{\lambda} \tag{3.29}$$

where we have used the fact that $M \gg 1$.
Now we know from the previous section that $\Delta y \sim 0.5$ m for a screen height $L_y = 100$ km, which means that the horizontal step needs to be shorter than 300 km. This simple calculation serves to illustrate that this criterion is of little concern to modern computers and researchers. And furthermore, from condition (2.32) we know that for realistic atmospheric data, the resolution has to be much finer.

Even though this criterion is of little relevance, it is still interesting to demonstrate how this particular form of aliasing appears. In Fig. 3.9 we can see the result of having too low horizontal resolution. In this particular case, the step length is $\Delta z = 8000$ km.

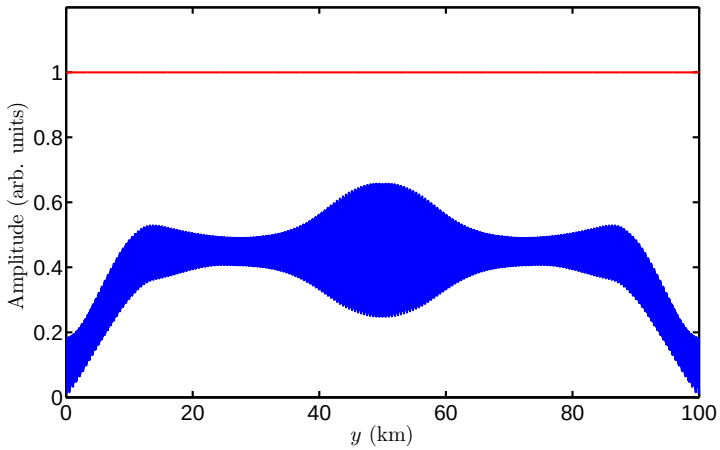

Figure 3.9: The result of having too long distance between the screens, allowing for aliasing to appear.

3.6.3 Refractive index gradient

The exponential factor containing the refractive index may also give rise to aliasing. If the refractive index at different screens is completely uncorrelated, aliasing will arise if the factor (3.15) differs too much between grid points. Similar to the Knepp criterion we find that to avoid aliasing we require

$$\frac{k_0}{2}\left[((n+\delta n)^2-1)-(n^2-1)\right]\Delta z < \pi \tag{3.30}$$

where δn is the maximum difference in the refractive index between grid points. We can approximate $\delta n \approx |\max(\partial n/\partial y)|\Delta y$, giving

$$2\left|\max\left(\frac{\partial n}{\partial y}\right)\right|\Delta y \Delta z < \lambda \tag{3.31}$$

The refractive index on adjacent screens is not uncorrelated in our case, and the requirement has to be made more restrictive. If the wave propagates in a media having a constant gradient in the refractive index over a length L_z, the criteria for avoiding aliasing becomes

$$2\left|\max\left(\frac{\partial n}{\partial y}\right)\right|\Delta y L_z < \lambda \tag{3.32}$$

which depending on L_z can become quite restrictive. Let us use the example of an exponential refractive index profile to see exactly when this problem

appears. It is easiest to use a geometry where the Earth is considered as flat and the y-direction coincides with the r-direction. In this case the refractive index is given by

$$n(y) = 1 + 10^{-6} N_0 \exp(-y/H_{scale}) \tag{3.33}$$

where N_0 is the refractivity at the bottom of the screen, and H_{scale} is the scale length of the atmosphere. The derivative becomes

$$\frac{\partial n}{\partial y} = -\frac{N_0}{H_{scale}} \times 10^{-6} \exp(-y/H_{scale}) \tag{3.34}$$

Since we are using a window function that gradually dampens the signal below $y = a_1 L_y$ and above $y = L_y(1 - a_1)$, the y-position with the highest gradient will not give rise to any problems. Instead we have to take into account the gradient over the entire y-interval and compare it with the window function for the y-interval. The criterion over the y-interval becomes

$$\Lambda \equiv 2 \times 10^{-6} \frac{N_0}{H_{scale}} \frac{\Delta y}{\lambda} L_z \exp(-y/H_{scale}) < 1 \tag{3.35}$$

In Figs. 3.10-3.11 we can see how the violation of this criterion gives rise to aliasing, and how the gradient interplays with the window function. In Fig. 3.10 the point where we violate the criterion (i.e. the blue curve crosses the dashed black line) is situated deep into the part where the window function has damped the wave, and we can see no trace of any numerical problems. In Fig. 3.11 on the other hand, the violation point is situated at a y-value that is above the point where the window function starts to dampen the signal, which results is aliasing.

When taking account of the curvature of the atmosphere as it envelopes the Earth the criteria becomes more complicated. The critical region should combine high gradients in the direction perpendicular to the propagation direction with long propagation path perpendicular to the Earth. This indicates the few km's closest to the Earth's surface as the critical region. Since no exact guidelines can be given for this case it is probably easier to just be careful and investigate how decreasing Δy influences the appearance of the signal amplitude profile.

3.7 Taking the Earth into account

The inclusion of the Earth into the simulation is problematic but unfortunately quite necessary. In a simulation which strives to be as realistic as possible, the finite conductivity, and the polarizability of the Earth should

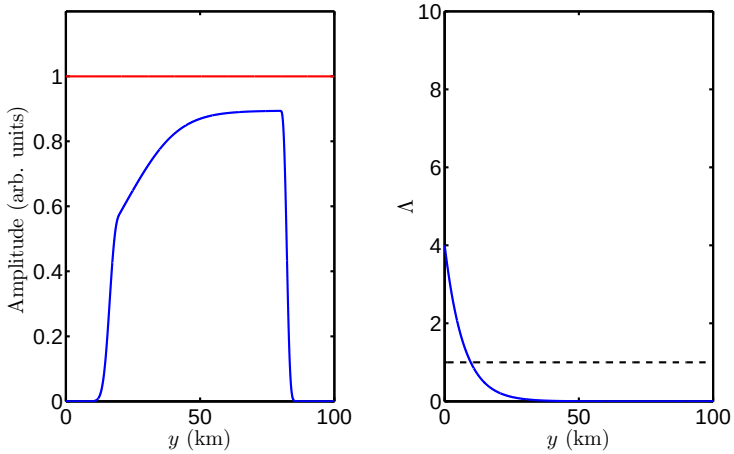

Figure 3.10: Illustration of how high gradients in the refractive index might give rise to aliasing. In this figure the point where the critical factor Λ equals unity is below the point where the window function has dampened the signal significantly, which results in no aliasing.

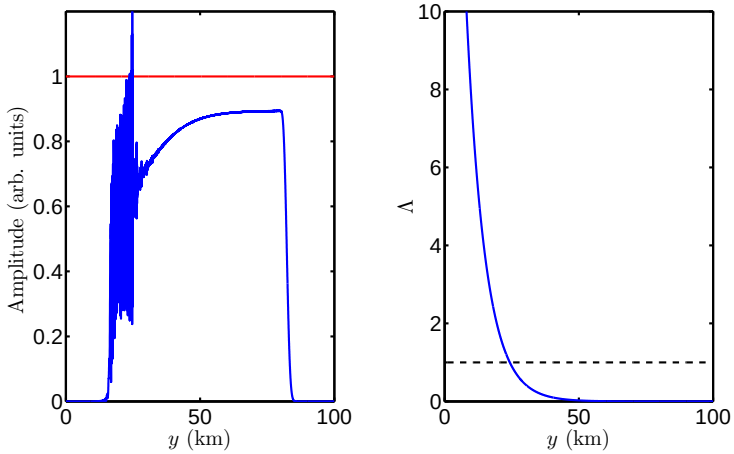

Figure 3.11: Illustration of how high gradients in the refractive index might give rise to aliasing. In this figure the point where the critical factor Λ equals unity is above the point where the window function has dampened the signal significantly, which results serious aliasing.

be taken into account. This will lead to different reflection properties over different types of terrains (compare for example seawater, which is conducting, with desert sand, which is dielectric). Procedures for taking these matters into account can be found in [3, 10]. Initially however, we do not want any

reflections from the Earth. We want to compare the bending angles retrieved using the simulations with those found using the Abel transform (and the equivalent geometrical optics ray tracing routine). The Abel transform takes no regard of the surface properties of the Earth, and only the rays that do not reach the Earth surface result in the occultation. To completely remove reflections from the Earth surface we need to let the wave amplitude decay slowly with increasing penetration into the Earth. A suitable choice for the amplitude damping is an exponential function similar to that used in the window function, viz

$$w_E = \begin{cases} \exp\left(-\frac{(r-R_E)^2}{L_A^2}\right) & \text{if} \quad r \leq R_E, \\ 1 & \text{if} \quad r > R_E. \end{cases} \tag{3.36}$$

where $r = \sqrt{(L_z/2 - z)^2 + (y + R_E + L_y - H)^2}$ is the distance to the center of the Earth, and L_A is the attenuation length. The damping of the amplitude is achieved as previously simply by multiplication with the wave amplitude at each screen. In Fig. 3.12 we can see a comparison of four different attenuation lengths, and in Fig. 3.13 we can see the wavefront amplitude as it is blocked by the Earth.

Naively one would state that the lower right curve looks most promising for good simulations avoiding reflections. That is however before one takes into account the effect of the refractive index. In an ideal atmosphere the refractive index decays exponentially away from the Earth surface. In our simulations we need to specify what the refractive index inside the Earth is. Just putting it equal to some constant will cause a sharp kink in the curve, and the generation of aliasing at the surface. One can let the refractivity continue increasing exponentially inside the Earth. But in that case one needs to make sure that the damping inside the Earth is sufficiently strong to avoid rays passing through the Earth with great bending angles. Another option is letting the refractive index assume the form of a straight line inside the Earth. In any case, there will be a trade-off. Either one allows some reflections, or one allows the wave to travel slightly into the Earth with an unrealistic refractive index.

Below we compare three different functional forms for the refractive index inside the Earth and how they interplay with different attenuation lengths. Outside the Earth the refractive index is approximately given by

$$n(r) = 1 + 10^{-6} N_0 \exp\left(-\frac{1}{H_{scale}}(r - R_E)\right) \tag{3.37}$$

and the forms inside the Earth should match the value of this function and its derivative at $r = R_E$. The three forms are: a constant value inside the Earth

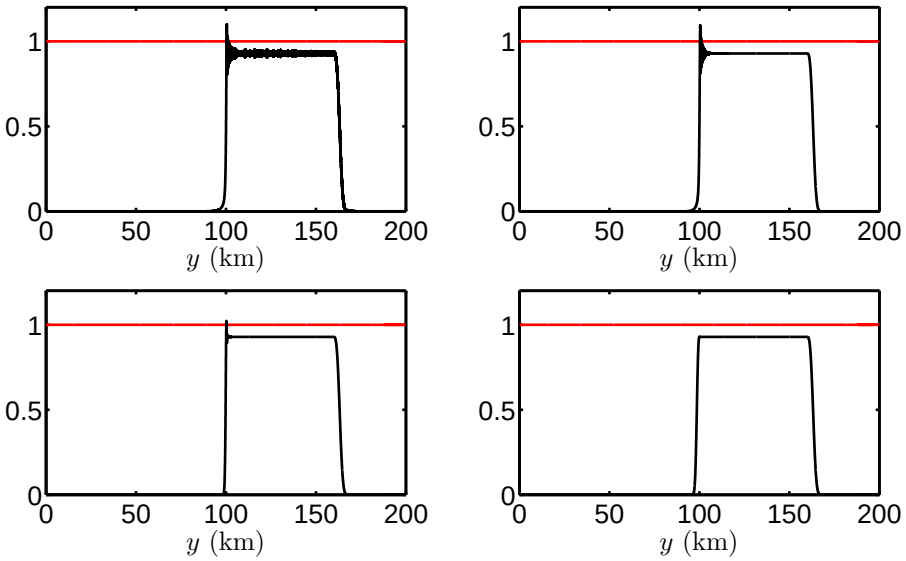

Figure 3.12: The effect of finite conductivity of the Earth's surface. In the upper left panel the wave amplitude is set to zero within the Earth, basically mimicking a perfectly conducting cylinder. This results in massive reflections and a complicated refraction pattern. The upper right panel lets the amplitude fall to zero exponentially (as a Gaussian) with a scale length (i.e. attenuation length) of 50 m. The lower left panel has an attenuation length of 500 m. The lower right panel an attenuation length of 5000 m. In the last case reflections are basically eliminated.

that uses an exponential term to match the derivatives at $r = R_E$; a slanted line with the same slope as the refractive index at $r = R_E$; and an exponential that decays with increasing radius. The mathematical formulation is seen in (3.38)-(3.40).

$$n(r) = 1 + 10^{-6} N_0 \left[2 - \exp\left(\frac{1}{H_{scale}} (r - R_E) \right) \right] \tag{3.38}$$

$$n(r) = 1 + 10^{-6} N_0 \left[1 - \frac{1}{H_{scale}} (r - R_E) \right] \tag{3.39}$$

$$n(r) = 1 + 10^{-6} N_0 \exp\left(-\frac{1}{H_{scale}} (r - R_E) \right) \tag{3.40}$$

In Figs. 3.14-3.16 we see the amplitude of the wave as it has passed through the Earth's atmosphere for the three different functional forms of the refractive index. In Fig. 3.14 we can clearly distinguish the effect of

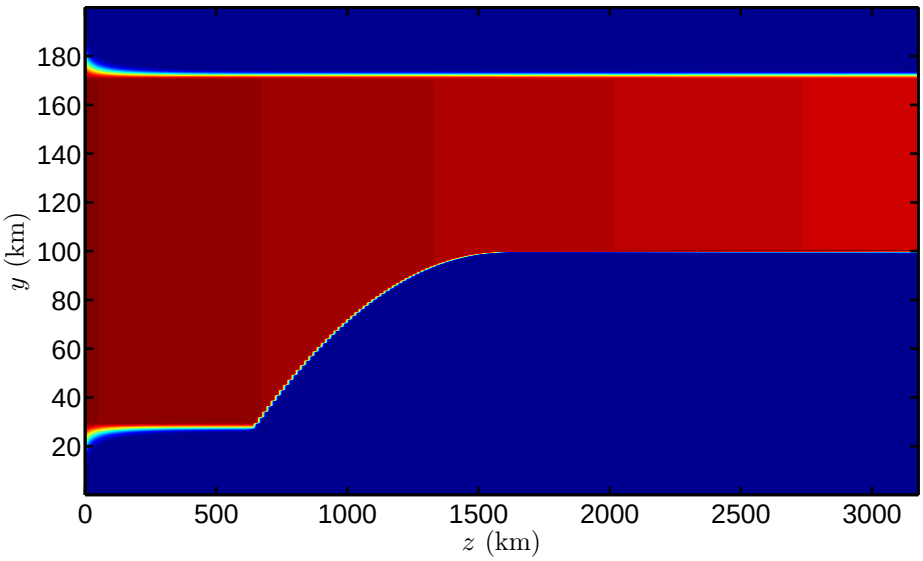

Figure 3.13: The shadowing effect of the Earth without an atmosphere. In this case the damping inside the Earth is slow, with a scale length of 5 km. Blue colors represent low amplitude, whereas red represents amplitude near unity.

the three forms, for in this case the attenuation length is 5 km. The exact features of the curves are too complicated to explain, suffice to say that the results seem to depend too much on the exact form of the refractive index. In Fig. 3.15 the attenuation length is 500 m, and the three different curves are hard to distinguish. In Fig. 3.16 the attenuation length is 50 m, and the three curves are identical and show a distinct refraction pattern. Clearly the intermediate value of 500 m combined with the exponential profile Eq. (3.40) gives the most satisfactory result (if one wishes to eliminate the effects of the Earth's surface), and this combination will be used throughout the remainder of this report. It is interesting to note that the amplitude profile in Fig. 3.15 is very similar in shape to curve A in Fig. 5 in [7]. The curve in [7] continues down to observation points 50 km below that in Fig. 3.15, but this is merely an effect of where the observation plane is positioned.

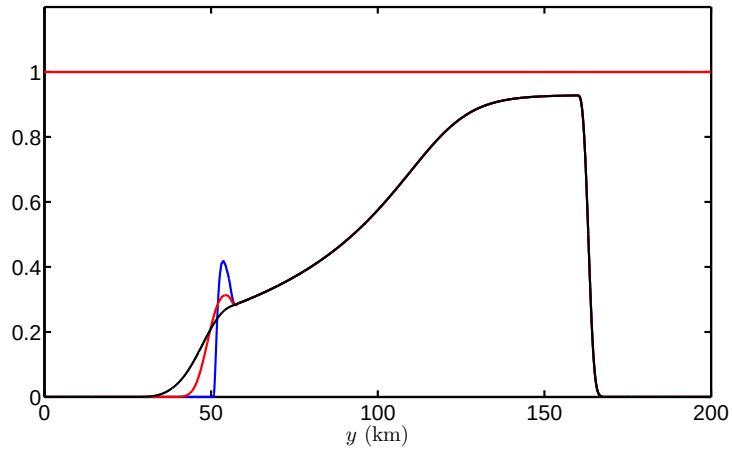

Figure 3.14: The wave amplitude as it has passed through the Earth's atmosphere. The attenuation length is 5 km. The blue curve is the result from using Eq. (3.38) for the refractive index inside the Earth. The red curve is the result from Eq. (3.39), and the black curve from Eq. (3.40).

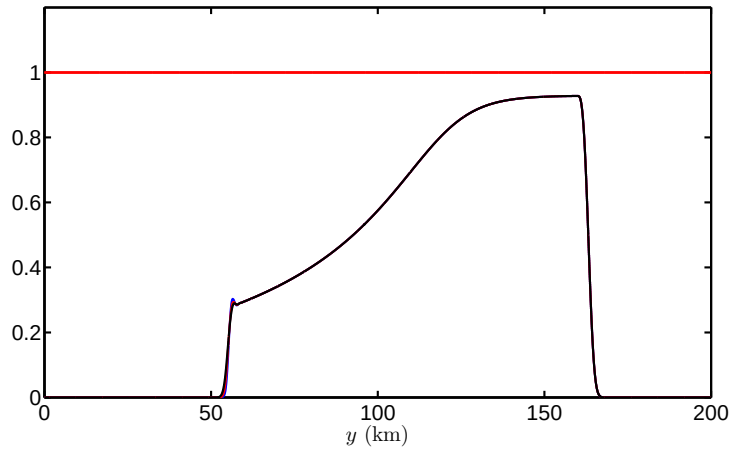

Figure 3.15: Same as in Fig. 3.14 but with an attenuation length of 500 m.

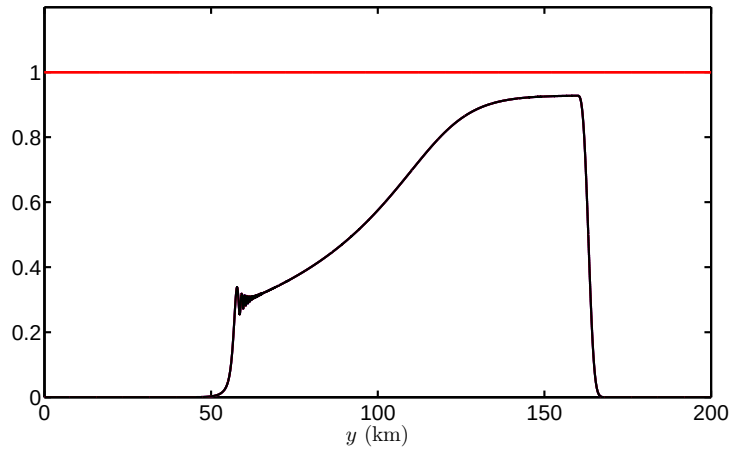

Figure 3.16: Same as in Fig. 3.14 but with an attenuation length of 50 m.

Chapter 4

Results on the last phase screen

4.1 Introduction

Under the assumption that each infinitesimal segment of the received signal on the last phase screen can be linked to one ray it is surprisingly easy to invert the signal and find the bending angle vs. impact height curve using only the received signal phase and geometry. This method is of course completely useless for multipath scenarios, as we will see in section 4.3.3. But logically, in order to evaluate the MPS technique alone we should take some time and analyze the difference between geometrical optics (GO) and simulations on the last phase screen for the case of analytical refractivity profiles. The benefit of using analytical profiles is that no effects of interpolation will appear in the MPS proceedure. We will look at exponential profiles with different gradients, and exponential profiles with a small region of sharper gradient. We will see how the difference between GO and simulations depend on the wavelength, refractivity gradient, number of phase screens, Earth attenuation, and the difference incurred by the truncation error in the numerical evaluation of the Abel integral. Overall the results show that the agreement between GO and simulations is very good.

4.2 Geometrical unwrapping of the phase on the last phase screen

The wavefront strikes the last screen (subscript F) with an angle θ_F with respect to the surface normal. To find this angle we first need to find the relative phase, ψ, of the field on the screen as a function of screen height y, see Fig. 4.1. Every time the real phase goes above π or below $-\pi$, the repre-

sentation of the phase in Matlab jumps to $-\pi$ or π respectively. Therefore, to unravel the relative phase one needs to go through the vector describing the phase in Matlab, and every time the phase performs such a jump, add or subtract 2π. In Matlab one simply uses the function *unwrap()* to perform this proceedure. The result of unwrapping a wavefront that has passed through the Earth's atmosphere, and a reference wave that has suffered no refraction is shown in Fig. 4.2.

Tracking the phase along the screen results in a curve of the phase ψ as a function of position y, as seen in Fig. 4.2. We can take the derivative of this and find the perceived wavelength on the screen

$$\frac{\partial \psi}{\partial y} = \frac{2\pi}{\lambda'} \tag{4.1}$$

The perceived wavelength is related to the real wavelength λ_0 through

$$\lambda' \sin \theta_F = \lambda_0 \tag{4.2}$$

So we get

$$\theta_F = \arcsin \left[\frac{\lambda_0}{2\pi} (\frac{\partial \psi}{\partial y}) \right] \tag{4.3}$$

The angle θ_F calculated this way can be seen in Fig. 4.3.

If one assumes spherical symmetry Bouger's rule will apply [11, 12] (see section 9), and we have the relation

$$rn \sin \phi = a \tag{4.4}$$

Therefore every point on the screen corresponds to a certain distance from the Earth's center r_F, and an angle, ϕ_F, between the ray and the line from the center of the Earth, which when determined will give us the parameters of the GNSS satellite through

$$r_0 \sin \phi_0 = a \tag{4.5}$$

The complete geometry of the problem is seen in Fig. 4.4. We can thus find α through the relations

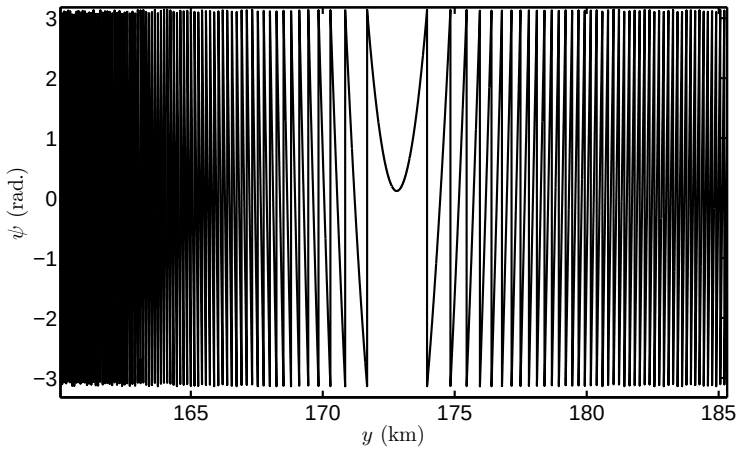

Figure 4.1: An example of the appearance of the phase of the wave as it strikes the final screen.

$$\alpha = \beta_F + \beta_0 + \phi_0 + \phi_F - \pi \tag{4.6}$$

$$\sin\beta_F = \frac{L_z}{2r_F} \tag{4.7}$$

$$\cos(\phi_F - \theta_F) = \frac{L_z}{2r_F} \tag{4.8}$$

$$\cos\beta_0 = \frac{y_0 + D}{r_0} \tag{4.9}$$

$$r_F = \sqrt{(y+D)^2 + (L_z/2)^2} \tag{4.10}$$

$$r_0 = \sqrt{(z_0 + L_z/2)^2 + (y_0 + D)^2} \tag{4.11}$$

$$D = R_E + H - L_y \tag{4.12}$$

$$\sin\phi_0 = \frac{r_F}{r_0}\sin\phi_F \tag{4.13}$$

where y is the coordinate along the vertical direction of the last screen. And we find the corresponding a to the α at a given y through

$$r_F \sin\phi_F = a \tag{4.14}$$

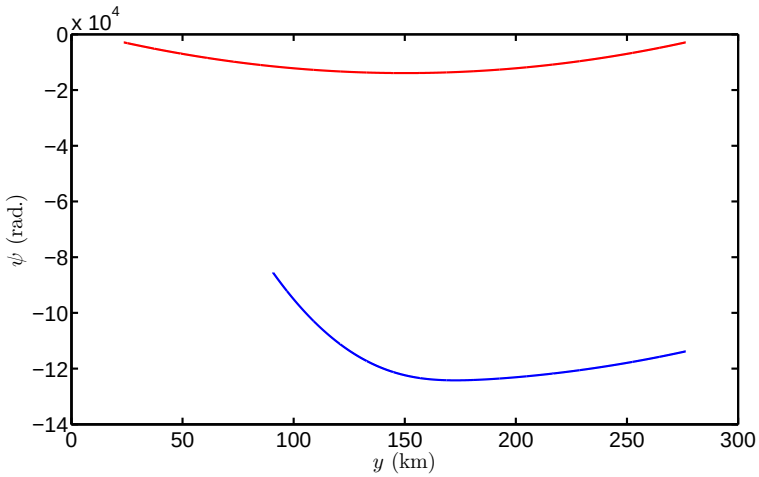

Figure 4.2: The relative phase found by unraveling the phase in Matlab. The blue curve represents the wave that has been refracted by the Earths atmosphere (the exponential profile given in Eq. (4.15)),whereas the red curve is an unrefracted signal. Only the range within which the signal amplitude is higher than 0.01 has been included, which explains the difference in y-range for the two signals. The difference in absolute value for the two signals comes from the fact that it is hard to find the absolute phase of the wavefronts, and no attempt has been made to do so in this case.

4.3 Differences between geometrical optics and simulations on the last phase screen

Geometrical optics is generally thought to provide a very good approximation for radio occultation wave fields. However, there should be minute differences due to diffraction and the finite size of the Fresnel zone. We will quantify these differences and try to identify the mechanisms behind them. We will start the investigation by looking at the results for an exponential profile (having a constant refractivity gradient determined by the atmospheric scale height), and its dependence on the refractivity gradient, the number of phase screens, the vertical resolution of the refractivity data, and the field wavelength. The great benefit of an exponential profile is that it will result in no multipath. We will then look at the results for an exponential profile with a small region having a sharper gradient (not superrefractive). Finally we will look at the results for an exponential profile having a small superrefractive region.

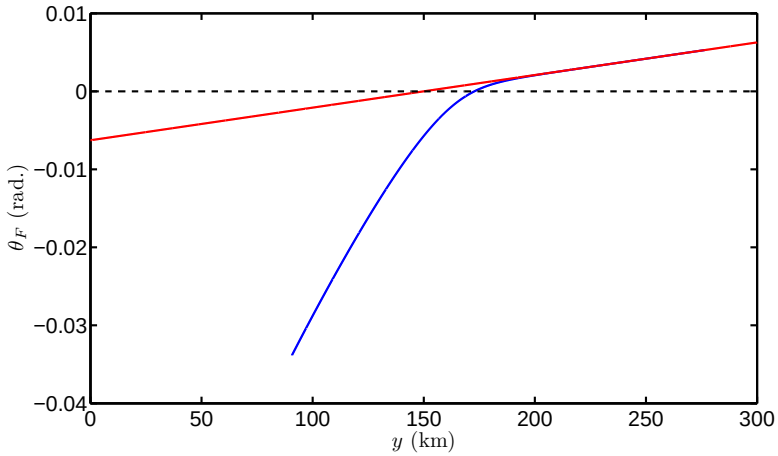

Figure 4.3: The angle with which the wavefront strikes the last screen, found from the derivative of the relative phase. Blue line represents the wave that has been refracted in the Earth's atmosphere, whereas the red curve is an unrefracted wave. The dashed black line is only there to accentuate that the red line has an angle zero precisely at $y = 150$ km, which coincides with the vertical position of the GNSS satellite ($y_0 = 150$ km).

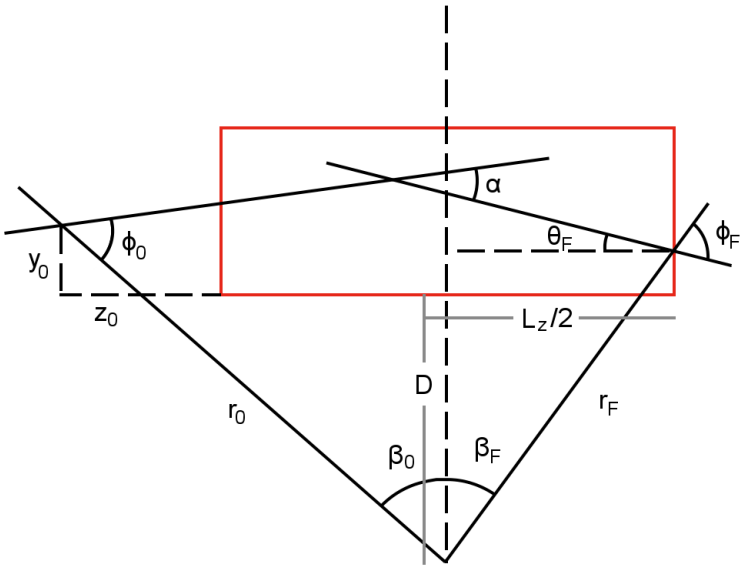

Figure 4.4: The complete geometry needed to find the bending angle α from the parameters used in the simulations.

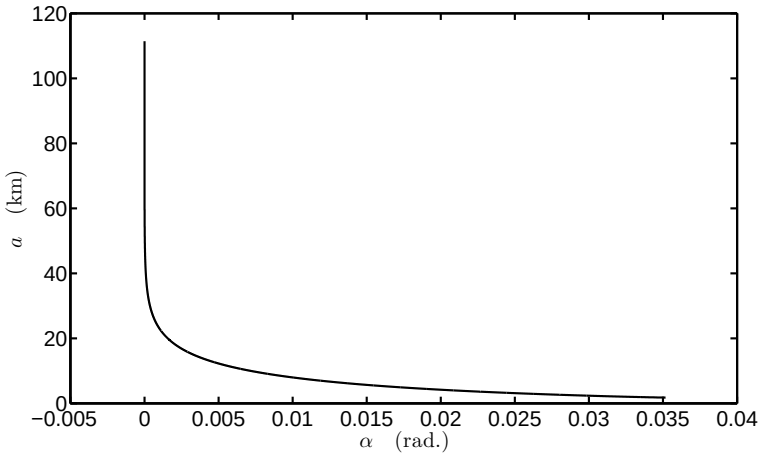

Figure 4.5: Bending angle, α, as a function of impact height, a, on the last phase screen for the exponential profile (4.15). On this scale it is impossible to see any difference between the simulated results and the results from geometrical optics (the Abel transform).

4.3.1 Exponential refractivity profile

In Fig. 4.5 the results for the profile

$$N(h) = N_0 \exp(-h/H_{\text{scale}}) \tag{4.15}$$

where $N_0 = 350$, and the scale height $H_{\text{scale}} = 7$ km, can be seen. On the scale used in Fig. 4.5 it is impossible to detect any difference between geometrical optics and the simulated results. To see how they differ we take the difference between the bending angle from simulations, α, and the bending angle from geometrical optics, α_{Abel}. This difference is shown in Fig. 4.6.

It can be seen in Fig. 4.6 that there is a difference between the results from simulations and GO which is not removed completely by increasing the number of screens. The exact number of screens to use is a complicated issue. There is a lower limit which is set due to the non-aliasing criterion in Sec. 3.6.3. But above this, the only thing that will happen when the number of screens is increased is that the truncation error in every step will be decreased. This can be seen more clearly in Fig. 4.7 where the diffraction pattern becomes more defined as the number of screens is increased. Furthermore, the average value for the difference only changes when going from 100 to 1000 screens, which indicates that no more than 1000 screens are necessary to get the correct average value for α. That this is true for profiles involving regions of sharper gradients will be shown in section 4.3.2.

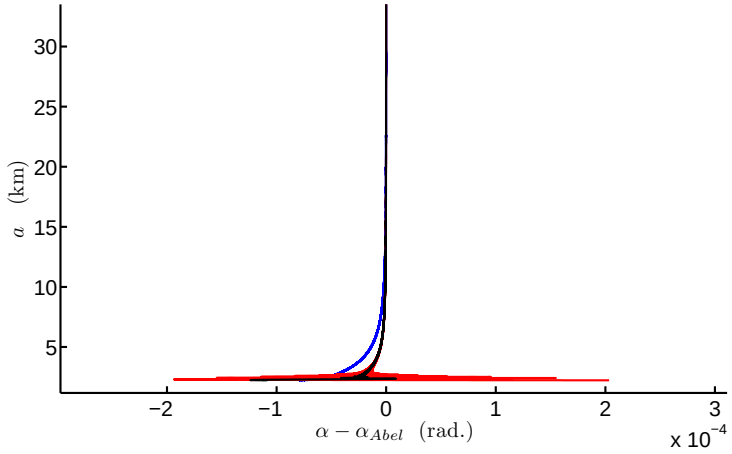

Figure 4.6: The difference between the simulated bending angle and the bending angle from geometrical optics, as a function of impact height on the last phase screen for the exponential profile (4.15). Three different number of screens have been used, 100 screens for the blue curve, 1000 for the black, and 10000 for the red. Clearly, for low a there is a difference between the different curves. This is due to diffraction.

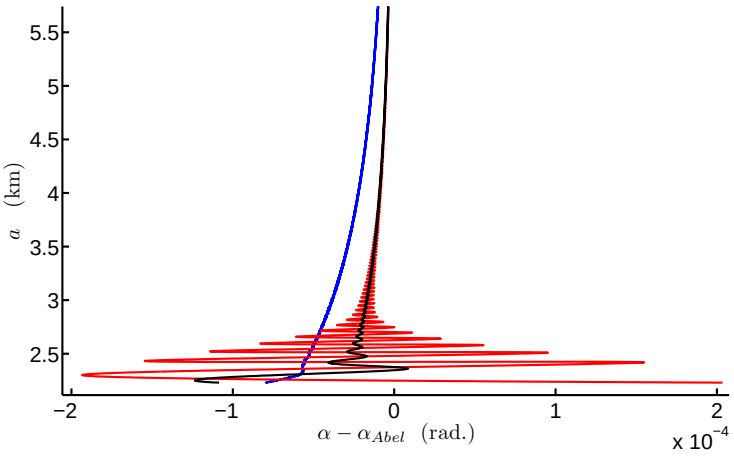

Figure 4.7: Same as in Fig. 4.6 but zoomed in on the lower section to show the diffraction patterns and the average values. Clearly, as the number of screens is increased, the diffraction pattern becomes more pronounced.

To prove that the pattern for low a is due to diffraction we simply need to change the wavelength, for the diffraction effects appear in the shadow zone and the region preceding it. Geometrical optics predict a completely opaque shadow, and although simulations give a field in the shadow zone, no

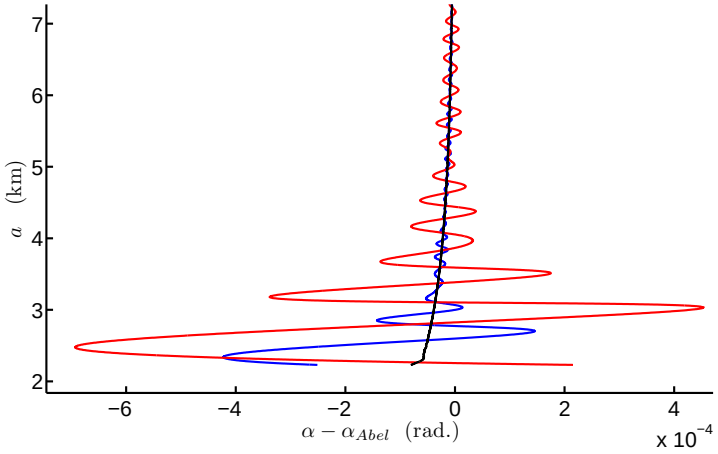

Figure 4.8: The diffraction pattern for the frequency corresponding to L1 (1.57542 GHz), black curve, 10 % of the L1 frequency (blue curve), and 5 % of the L1 frequency (red curve).

comparison can be made with GO there. However, just above the shadow zone there is a region which is also affected by diffraction, and as the wavelength goes to zero the difference between GO and simulations should go to zero in this region. All simulations so far have been performed using the L1 frequency (1.57542 GHz). In Fig. 4.8 we can see the results for L1, L1/10, and L1/20. The results clearly indicate that diffraction becomes stronger with increasing wavelength. And the diffraction pattern seems to have an almost linear dependence on the wavelength. For as the wavelength is doubled between L1/10 and L1/20, the period of the pattern appears to double.

The diffraction pattern should also show some dependence on the impedance of the Earth's surface. To investigate this we compare the patterns created for three different values of the damping length in Eq. 3.36. In Fig. 4.9 the diffraction region is shown for the damping length being 5000 m (black curve), 500 m (blue curve), and 50 m (red curve). Clearly the diffraction pattern becomes more strong for low values of the damping length, and hardly any difference can be discerned between 500 and 5000 m. Most likely this is caused by the complete absorption into the Earth of the part of the wavefront that interacts with the Earth's surface.

Looking further up along the curves (fig. 4.10), where the diffraction pattern has disappeared, a small scale ripple pattern can be seen. This is caused by the truncation error in the numerical integration used for the Abel transform. That this is true is shown in fig. 4.10 where three different step lengths for the height vector is used: 10 m for the black curve, 20 for the red,

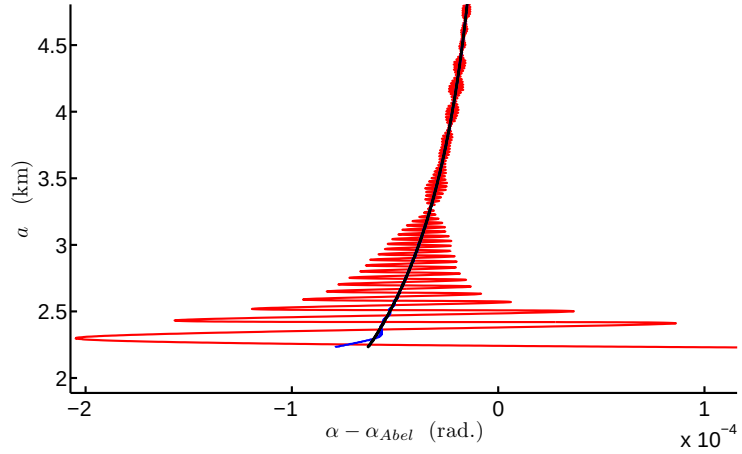

Figure 4.9: The diffraction pattern for three different values of the damping length, black curve is 5000 m, blue curve is 500, and red curve is 50.

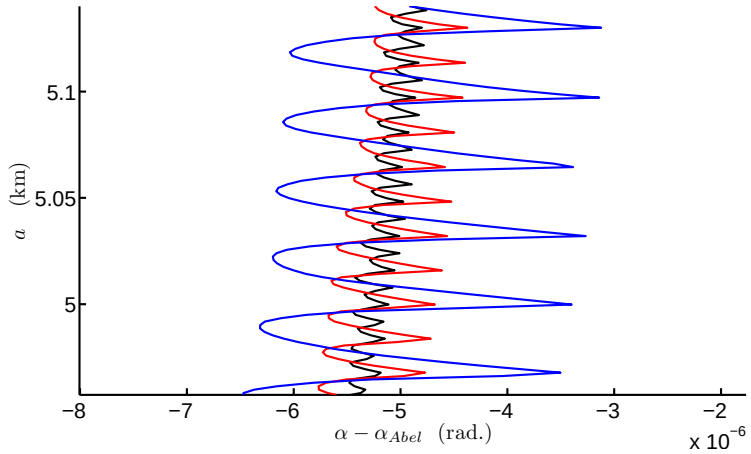

Figure 4.10: The fine scale ripple pattern caused by the truncation error in the numerical Abel transform. The step lengths used in the height vector are 10 m for the black curve, 20 for the red, and 40 for the blue.

and 40 for the blue curve. If the integral could be performed analytically this pattern would disappear. But it can be reduced to any desired level by decreasing the step length.

What remains to be investigated in order to have complete confidence in the simulator for exponential profiles is how the comparison between simulations and GO depends on the refractivity gradient. For this reason we will compare the results for the exponential profile with the scale height set

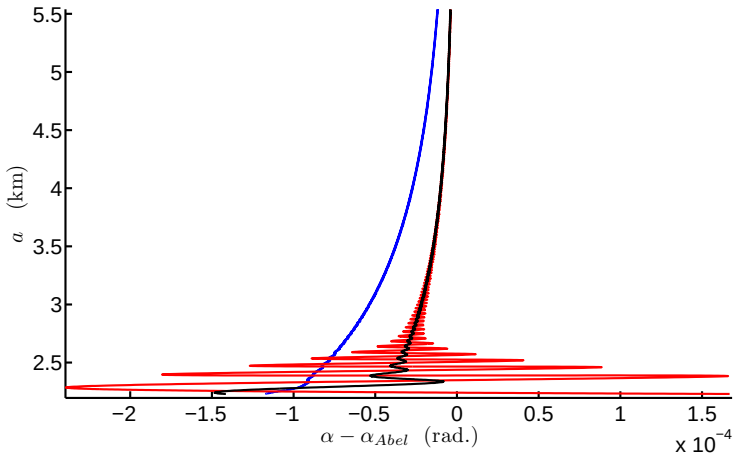

Figure 4.11: The difference between the simulated bending angle and the bending angle from GO, as a function of impact height on the last phase screen for the exponential profile with scale height 6 km. Three different number of screens have been used, 100 screens for the blue curve, 1000 for the black, and 10000 for the red. This figure shows the same thing as Fig. 4.7, namely that the average value for the difference does not change significantly by increasing the number of screens above 1000, whereas the diffraction pattern becomes ever more detailed by increasing the number of screens.

to 6, 7 and 8 km respectively. We know from the investigation above that increasing the number of screens above 1000 does not change the average value for the difference between simulations and GO in the case when the scale height is 7 km. In Fig. 4.11 we can see that this is also true for a scale height of 6 km. We can therefore infer that this is true for all exponential profiles having a less step gradient than the one in Fig. 4.11.

After such preliminary contemplations we can compare the results for the three different scale heights (4.15). This can be seen in Fig. 4.12, where 1000 screens have been used in order to not have to much overlapping diffractive undulations. Clearly the average value for the difference tends to increase with increasing refractivity gradient. This is not at all strange, as any difference between GO and simulations comes from the finite length of the field wavelength in comparison to the length scale for the refractivity gradient both in the atmosphere and in the Earth.

In all the figures above we have used the difference between the bending angles on the x-axis. It is quite hard to judge how large this difference really is in comparison with the bending angle itself. For this reason we shall now quantify the difference for the profile having a 6 km scale height. The

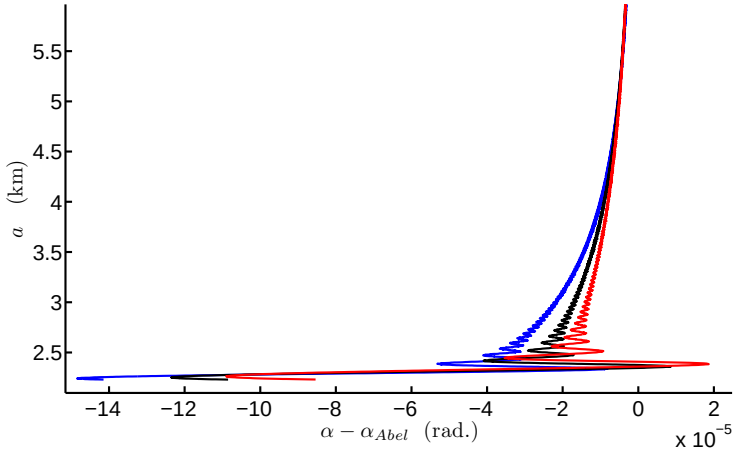

Figure 4.12: A comparison of the difference between simulations and GO for three different scale heights: 6 (blue curve), 7 (black) and 8 km (red). The number of screens used is 1000, and it is clear that the difference between simulations and GO increases with increasing refractivity gradient.

maximum bending angle is found at the lowest impact height, and has a value around 0.03545 radians. At the lowest impact height, the difference in bending angle for 10000 screens oscillate wildly around an average of approximately -0.5×10^{-4} radians, with a maximal negative value of around -2.394×10^{-4}. The maximum deviation between simulations and GO is thus

$$\Delta\alpha_{\max} \approx \frac{2.394 \times 10^{-4}}{3.5 \times 10^{-2}} = 0.00684 \approx 0.7 \ \ \% \tag{4.16}$$

which is quite low.

4.3.2 Sharp gradient profile

A usual situation for the refractivity profiles is that there exist a small region where the gradient is sharper than otherwise (but not superrefractive). Although such regions will perturb the signal amplitude and the resulting bending angle vs. impact height curves significantly they do not appear to give rise to multipath, and can therefore be unraveled directly on the last screen using GO. The refractivity profile can be modeled analytically by

$$N(h) = N_0 \exp(-h/H_{\text{scale}}) + \frac{\Delta N_l}{1 + \exp((h - h_l)/H_l)} \tag{4.17}$$

where H_l and ΔN_l determine the range and depth of the layer, and h_l the height where the layer is situated, whereas N_0 and H_{scale} is just the nor-

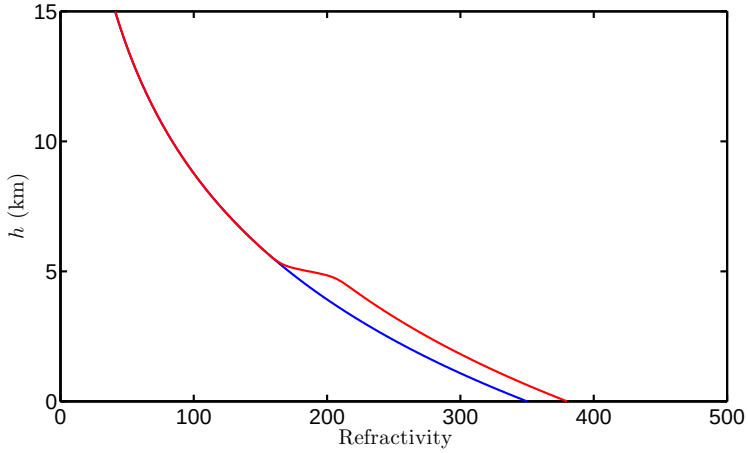

Figure 4.13: The refractivity profile given by Eq. (4.17) (red curve) with $N_0 = 350$, $H_{\text{scale}} = 7$ km, $\Delta N_l = 30$, $H_l = 100$ m, and $h_l = 5$ km. The blue curve is simply the exponential background profile.

mal exponential atmosphere. This type of profile can be seen in Fig. 4.13 alongside a normal exponential profile.

The steepest gradient in this type of profile is found in the middle of the layer, i.e. $h = h_l$, and has the value

$$\frac{\partial N}{\partial h}|_{h=h_l} = -\frac{N_0}{H_{\text{scale}}} \exp(-\frac{h_l}{H_{\text{scale}}}) - \frac{\Delta N_l}{H_l} \tag{4.18}$$

For the profile in Fig. 4.13 the maximal negative gradient is -324 km^{-1}. This should be compared with the criterion for super-refractivity which is

$$\frac{\partial N}{\partial h} < -157 \text{ km}^{-1} \tag{4.19}$$

In this section we will look at a profile which is not superrefractive, and in section 4.3.3 we shall see what happens if the profile is superrefractive. It is easy to decrease the maximum gradient. By simply changing h_l to 500 m we obtain a profile with a maximum gradient of around -84 km^{-1}. We shall investigate this profile for the remainder of this section. First of all it is quite interesting to note how the amplitude (see Fig. 4.14) is modified by adding this region of extra gradient. There is a U-shaped segment in the amplitude profile. Apparently the part of the wavefront that will have tangent height inside the sharp gradient region will be decreased, and more rays will have tangent heights in the region just below it, leading to a peak to the left of the U-shaped region.

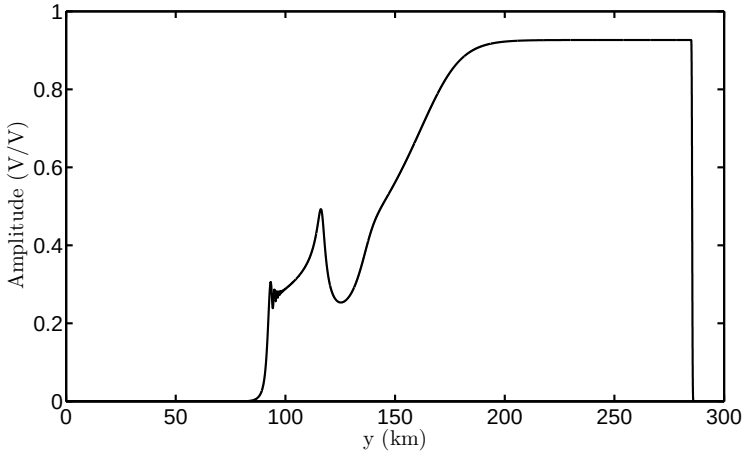

Figure 4.14: The amplitude of the signal as it strikes the last phase screen. Compare with e.g. Fig. 5.11 and note the somewhat depleted region. This is indicative of a region of sharper gradient.

The appearance of the bending angle vs. impact height curve can be seen in Fig. 4.15, where it is easily seen that adding the region of sharper gradient leads to the formation of a hump in the curve. From this figure it is quite hard to see if there is any difference between the simulated curve and the one calculated from GO using the Abel transform.

In Fig. 4.16 we can see a detailed plot of the difference between the bending angle from simulations and GO for 100, 1000 and 10000 screens. Clearly, 100 screens is not sufficient. But increasing the number of screens beyond 1000 seem to alter the average difference very little. The maximum difference in the region influenced by the sharp gradient is around -1.0×10^{-5}. Whereas the bending angle in that region is around 1.7×10^{-2}. This amounts to a deviation between the simulated and GO results of about 0.06 %, implying that the MPS method and GO agrees very well even for slightly more complicated profiles than a simple exponential.

4.3.3 Superrefractive profile

If we use the same profile as in the previous section, but set $H_l = 100$ m we get a maximum negative gradient of -324 km^{-1}. This is roughly twice as steep as the superrefractive limit at -157 km^{-1} and we should expect serious multipath to appear. The amplitude of the signal on the last screen can be seen in Fig. 4.17, and it is clear that the presence of the superrefractive layer has wreaked havoc on all parts of the signal with tangent altitude below h_l.

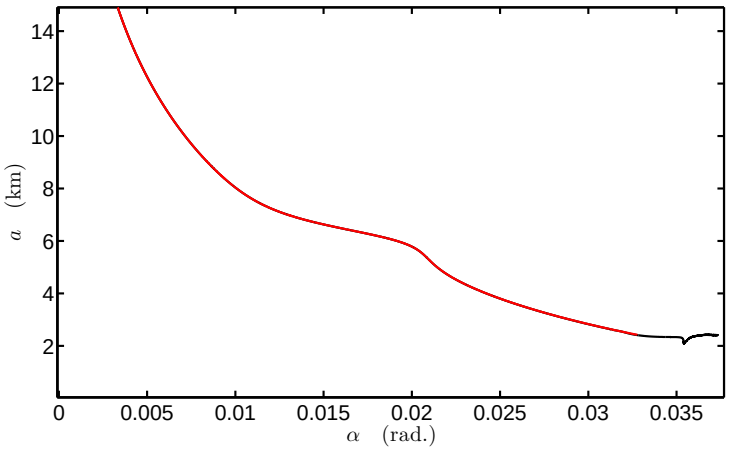

Figure 4.15: The bending angle vs. impact height curves for the sharp gradient profile. Red curve is from GO, whereas black curve is from simulations. The difference between the curves is minimal, except that the curve from simulations extends to lower impact heights; into the shadow region. The effect of adding the region of sharper gradient is obvious in that it adds a hump to the otherwise smooth curve around 6 km.

In Fig. an attempt has been made to unravel and invert the signal on the last phase screen using the GO method. It is quite unsuccessful, but it is interesting to note in which way the method fails. In the region of the superrefractive layer and below it the signal is entirely composed of slanted streaks. This is what happens when one tries to invert a multipath region using GO.

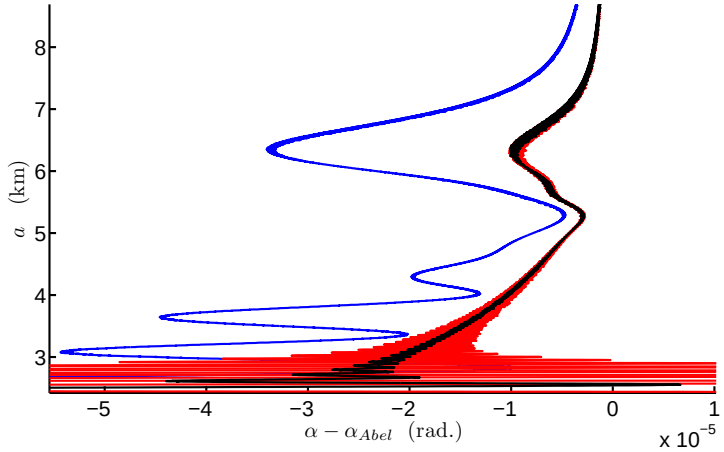

Figure 4.16: The difference between simulations and GO for 100 (blue curve), 1000 (black curve) and 10000 screens (red curve). For low impact heights we see the same features that we have already investigated in the previous section. Around the region of the sharp gradient we see that simulations and GO differ. For 100 screens the difference looks crazy, and clearly the number of screens is to few. For 1000 and 10000 screens the average difference appears to be the same.

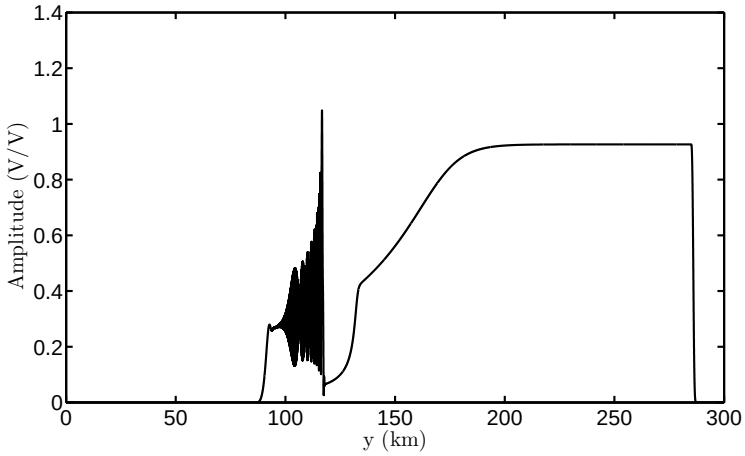

Figure 4.17: The signal amplitude on the last phase screen for $N_0 = 350$, $H_{\text{scale}} = 7$ km, $\Delta N_l = 30$, $h_l = 5$ km, and $H_l = 100$ m. The U-shaped region caused by the superrefractive layer is now seriously depleted, and the region to the left is heavily deformed. This is probably due both to multipath and diffraction.

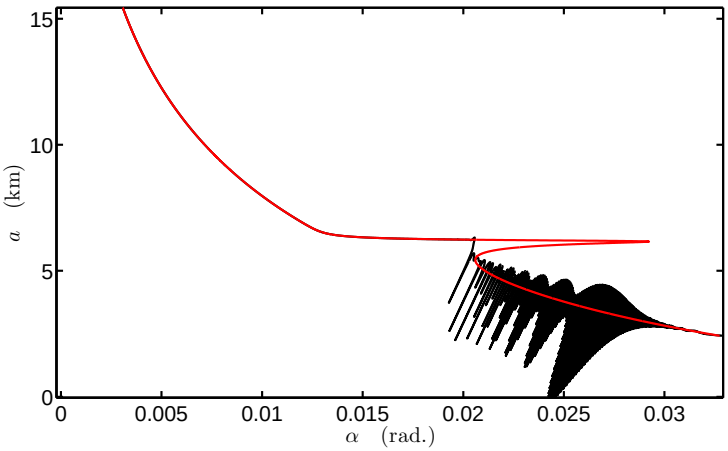

Figure 4.18: A comparison of the calculated bending angle using GO (red curve), and the inverted (using simple GO) simulated signal (black curve), for the exponential profile containing a superrefractive layer. Note the slanted streaking characteristics of the black curve. When using this inversion method they are indicative of multipath. Also note the singularity in the red curve. This is a clear indication of super-refractivity when calculating the bending angle using the Abel transform.

Chapter 5

Propagating to LEO

5.1 Introduction

The propagation of the GNSS signal through the atmosphere is handled with the MPS technique, and from the analysis in the previous sections we know that it works very well. To complete the forward modelling of the RO signal propagation we need to include the step from the last phase screen to the LEO satellite. There are several ways of doing this. The standard way is to use a diffractive integral, which is what we will do, where an integral over the last phase screen is performed for each position in the LEO occultation path. The basic method can be sped up using some more or less sophisticated methods [13, 14]. Alternatively one could simply use the the MPS method with a very large Δz to find the field in each point of the orbit [4]. If the diffractive integral would have to be taken over the entire last phase screen this method would probably be faster, but we shall see that only a fraction of the phase screen around the stationary point needs to be included, which speeds up the process significantly.

5.2 The diffraction integral

In standard textbooks [15, 16] one can find the solution for the Helmholtz equation which describes the radiation received at any point from an opening in an otherwise opaque wall or enclosure, where the field, or the spatial derivate of the field is known inside the opening. This formula is a type of diffraction integral, and in RO publications typically designated with a prefix or combination of prefixes like Kirchoff, Fresnel, or Huygens. It can be directly applied to the problem of propagating the field from the last phase screen to a receiving satellite in LEO if one is working in three dimensions.

So far in our work we have constrained ourselves to two dimensions, and the solution becomes quite different. The solution in this case is not presented in any standard textbooks (at least to the knowledge of this author). A partial derivation of the 2-D diffraction integral can be found in a report by Gorbunov et al. [17], and all simulations using this type of Earth to LEO propagation has used this integral without displaying any relevant derivations. The solution is not overly complicated (although somewhat tedious) wherefore we will give all the details below. The basic technique is to use Green's theorem, assume a certain form for one of the unknown function, assume some specific boundary conditions in a suitable geometry, and find the field at one point by integrating over a finite volume. The presentation below will follow this pattern.

5.2.1 Green's theorem

Green's theorem states that

$$\int_V (u\nabla^2 v - v\nabla^2 u)dV = \oint_S (u\nabla v - v\nabla u)\cdot d\bar{S} \tag{5.1}$$

where u and v are continuous functions having continuous derivatives within the volume V (with volume element dV) which is enclosed by the surface S (with surface element $d\bar{S}$). If both u and v satisfies the Helmholtz equation within V we have

$$\nabla^2 u + k^2 u = 0 \tag{5.2}$$
$$\nabla^2 v + k^2 v = 0 \tag{5.3}$$

meaning that Green's theorem reduces to

$$\oint_S (u\nabla v - v\nabla u)\cdot d\bar{S} = 0 \tag{5.4}$$

To find the diffraction integral from this formula one uses the Green's function which satisfies

$$\nabla^2 G + k^2 G = \delta(r) \tag{5.5}$$

where $\delta(r)$ is the Dirac delta function. By excluding the point where $r \to 0$ the Green's function is a solution of the Helmholtz equation, and actually, by using the Green's function one can find general solutions for the Helmholtz equation having almost any boundary conditions. Now this formula is intended for applications in three dimensions. We want a 2-D solution, and

therefore we state that any derivative in the z-direction yields zero. In this case the volume integral reduces to a surface integral in the xy-plane, and the surface integral reduces to a line integral. The form of the Green's theorem we will use is then

$$\oint_C (u\nabla v - v\nabla u)\cdot \hat{n} dl = \oint_C (u\frac{\partial v}{\partial n} - v\frac{\partial u}{\partial n})dl = 0 \tag{5.6}$$

where C is a curve with the surface normal $\hat{n}$ and length element dl.

5.2.2 The 2-D Green's function

The Green's function for the Helmholtz equation satisfies

$$\nabla^2 G + k^2 G = \delta(r) \tag{5.7}$$

When one knows the Green's function one can construct solutions where the source of the radiation can have complicated shapes, but finding the Green's function is always a bit tricky. In our case we can construct the Green's function by considering the general solution for the Helmholtz equation in cylindrical coordinates, where the only dependence is on the radius r. This is a standard solution which is given by

$$u(r) = AJ_0(kr) + BY_0(kr) \tag{5.8}$$

where J_0 is the zeroth order Bessel function of the first kind, and Y_0 the zeroth order Bessel function of the second kind. This solution describes waves that travel both away and towards the source, and we wish to limit our solution to waves going away from the source at $r = 0$. For this purpose we can use the asymptotic solutions where $r \to \infty$

$$J_0(kr) = \sqrt{\frac{2}{\pi kr}}\cos\left(kr - \frac{\pi}{4}\right) = \frac{1}{2}\sqrt{\frac{2}{\pi kr}}\left(\exp(i(kr-\pi/4)) + \exp(-i(kr-\pi/4))\right) \tag{5.9}$$

$$Y_0(kr) = \sqrt{\frac{2}{\pi kr}}\sin\left(kr - \frac{\pi}{4}\right) = \frac{1}{2i}\sqrt{\frac{2}{\pi kr}}\left(\exp(i(kr-\pi/4)) - \exp(-i(kr-\pi/4))\right) \tag{5.10}$$

Inserting these solutions gives us

$$u(r) = \exp(i(kr-\pi/4))\sqrt{\frac{2}{\pi kr}}\left[\frac{A}{2} + \frac{B}{2i}\right] + \exp(-i(kr-\pi/4))\sqrt{\frac{2}{\pi kr}}\left[\frac{A}{2} - \frac{B}{2i}\right] \tag{5.11}$$

Since we are using a time dependence on the form $\exp(-i\omega t)$ an outward travelling solution requires that we eliminate terms on the form $\exp(-ikr)$, consequently we require

$$\frac{A}{2} - \frac{B}{2i} = 0 \quad \Leftrightarrow \quad B = iA \tag{5.12}$$

and the solution for u becomes

$$u(r) = A\left(J_0(kr) + iY_0(kr)\right) = AH_0(kr) \tag{5.13}$$

where H_0 is the zeroth order Hankel function of the first kind. For $r \to 0$ the Hankel function can be approximated by $2i\ln(kr)/\pi$, wherefore

$$u(kr \to 0) \sim \frac{2iA}{\pi}\ln(kr) \sim \frac{2iA}{\pi}\ln r \tag{5.14}$$

What we have found is actually the Green's function multiplied by some constant, which can be verified by performing the integral over a cylinder with radius R with its central axis at $r \to 0$

$$\int_V (\nabla^2 u + k^2 u)dV = \int_V \nabla^2 u dV + k^2 \int_V u dV = \oint_S \nabla u \cdot d\bar{S} + k^2 \int_V u dV = \tag{5.15}$$

$$= \int_0^{2\pi} \frac{du}{dr}|_{r=R} R d\phi + k^2 \int_0^R \int_0^{2\pi} u(r) r d\phi dr \tag{5.16}$$

Using (5.14) the expression above becomes

$$\int_0^{2\pi} \frac{2iA}{\pi R} R d\phi + k^2 \int_0^R \int_0^{2\pi} \frac{2iA}{\pi}\ln(r) r d\phi dr \tag{5.17}$$

The left term clearly becomes $4iA$ for any R, whereas the right term tends to zero as $R \to 0$. The volume integral therefore becomes

$$\int_V (\nabla^2 u + k^2 u)dV = 4iA \tag{5.18}$$

whereas the Green's function yields an integral

$$\int_V (\nabla^2 G + k^2 G)dV = \int_V \delta(r)dV = 1 \tag{5.19}$$

and hence the Green's function is given by

$$G(r) = \frac{u(r)}{4iA} = \frac{1}{4i}H_0(kr) = -\frac{i}{4}H_0(kr) \tag{5.20}$$

5.2.3 Constructing a solution

Historically there was much controversy over how to perform the integration of Eq. (5.6) using correct boundary conditions. The approach used by Kirchoff was to set $v = \partial v/\partial n = 0$ on s_3 in Fig. 5.1. However, according to a theorem by Riemann, this would imply a v that is zero everywhere. The problem was partially resolved by Sommerfeld [15] by the use of a clever positioning of the "sources" so as to yield $u = 0$ on s_2 and s_3. This method has the advantage of only requiring the knowledge of v, and not its derivative, along s_2 and s_3 in order to find the resulting field at any other point. Due to this positioning of the sources, the Sommerfeld method is limited to plane surfaces. As to which method gives the most accurate result, that seems to depend on the parameter region, and the controversy has not yet been resolved completely. Below we will use the Sommerfeld method, with the small difference that we treat the 2-D problem, whereas Sommerfeld was interested in the 3-D case. The general method is to assume a certain form for u in order to find v at a specific point p_1. The form of u should be some combination of Green's functions that yields appropriate boundary conditions. The integration area is shown in Fig. 5.1. At points p_1 and p_2 two sources described by the 2-D Green's functions having opposite signs are located. The solid (s_3) and dashed line (s_2) that run along the opaque screen lies exactly on the symmetry line between these sources, which means that $u = 0$ on s_2 and s_3. Expression (5.6) then splits into four integrals

$$I_1 + I_2 + I_3 + I_4 = 0 \tag{5.21}$$

where

$$I_j = \int_{s_j} \left[u\frac{\partial v}{\partial n} - v\frac{\partial u}{\partial n} \right] dl \tag{5.22}$$

We can evaluate each contribution separately. The integration over s_4 is performed by assuming that s_4 is semi-circle with a radius that approaches infinity. The integral becomes

$$I_4 = \int_0^{\pi} \left[u\frac{\partial v}{\partial r} - v\frac{\partial u}{\partial r} \right] r d\phi \tag{5.23}$$

The function u is taken to be a combination of the 2-D Green's function, which means that its minimum rate of decay with increasing distance is given by

$$u \sim \frac{\exp(ikr)}{\sqrt{r}} \tag{5.24}$$

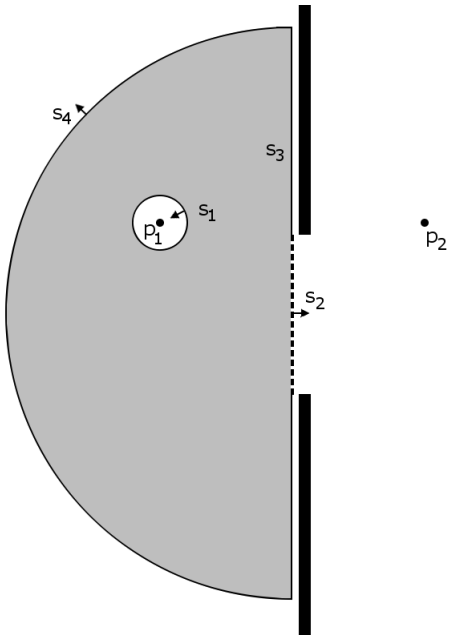

Figure 5.1: The integration area used to find the correct diffraction integral. The dashed line represents the aperture, or in our case the phase screen, and the two black points represent two point sources radiating in a way described by the 2-D Green's function. The two black rectangles represent an opaque screen, and the integration lines are represented by solid and dashed lines.

Inserting this in (5.23) yields

$$I_4 \sim \int_0^\pi \left[\frac{\partial v}{\partial r} + \frac{v}{2r} - ikv\right] \sqrt{r}\exp(ikr)d\phi \tag{5.25}$$

The condition for this integral to vanish is that

$$\lim_{r\to\infty} \sqrt{r}\left[\frac{\partial v}{\partial r} - ikv + \frac{v}{2r}\right] = 0 \tag{5.26}$$

which is true if v decays at least as fast as a cylindrical wave ($\propto \exp(ikr)/\sqrt{r}$), which it always will for a distribution of sources over a finite region. The condition (5.26) is equivalent to the so called Sommerfeld radiation condition [18] which applies to three dimensions and requires that v scales as a spherical wave.

To evaluate I_1 we need to choose the exact form of u. Let us use (it is irrelevant which source is set to be negative, the final result will be the same in both cases)

$$u = \frac{i}{4}H_0(k|\bar{r} - \bar{r}_1|) - \frac{i}{4}H_0(k|\bar{r} - \bar{r}_2|) \tag{5.27}$$

where $\bar{r}_1$ and $\bar{r}_2$ are the position vectors of points p_1 and p_2. Using $H_0(r \to 0) \sim 2i \ln r/\pi$, and using the direction of line element normals indicated by the little arrows in Fig. 5.1, the integral becomes

$$I_1 = -\int_0^{2\pi} \left[u\frac{\partial v}{\partial r} - v\frac{\partial u}{\partial r} \right] Rd\phi \sim \frac{1}{2\pi}\int_0^{2\pi} \left[\ln R \frac{\partial v}{\partial r}|_{r=R} - \frac{v(R)}{R} \right] Rd\phi \tag{5.28}$$

Now as $R \to 0$ the contribution from the left term goes to zero, since v is not supposed to have a singularity at $\bar{r}_1$. The integral becomes

$$I_1 = -v(\bar{r}_1) \tag{5.29}$$

The third integral can be put to zero immediately since $u = 0$, and we assume $v = 0$, on s_3, so

$$I_3 = 0 \tag{5.30}$$

Since the function u has been chosen to yield zero on s_2, the second integral reduces to

$$I_2 = -\int_{s_2} v\frac{\partial u}{\partial n} dl \tag{5.31}$$

In order to evaluate the integral it is necessary to find the form of $\partial u/\partial n$ on s_2. To do this we need to specify a particular coordinate system. Let us say that p_1 is located at $(-x_0, y_0)$, and p_2 at (x_0, y_0), and that the y-axis runs along the line s_2 where $x = 0$. The line normal then points in the same direction as the x-axis. In this case

$$u(x, y) = \frac{i}{4}\left[H_0\left(kr_a\right) - H_0\left(kr_b\right)\right] \tag{5.32}$$

Where $r_a = \sqrt{(x + x_0)^2 + (y - y_0)^2}$ is the distance from p_1 to s_2, and $r_b = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ is the distance from p_2 to s_2.

$$\frac{\partial u}{\partial x} = \frac{ik}{4}\left[\frac{\partial r_a}{\partial x}\frac{\partial H_0}{\partial \zeta_a} - \frac{\partial r_b}{\partial x}\frac{\partial H_0}{\partial \zeta_b}\right] \tag{5.33}$$

where $\zeta_a = kr_a$, and $\zeta_b = kr_b$, which means that $\partial H_0/\partial\zeta_a = \partial H_0/\partial\zeta_b$. The expression becomes

$$\frac{\partial u}{\partial x} = \frac{ik}{4}\left[\frac{x+x_0}{\sqrt{(x+x_0)^2+(y-y_0)^2}} - \frac{x-x_0}{\sqrt{(x-x_0)^2+(y-y_0)^2}}\right]\frac{\partial H_0}{\partial \zeta} \quad (5.34)$$

When $x = 0$ we have

$$\frac{\partial u}{\partial x} = \frac{ik}{2}\frac{x_0}{\sqrt{x_0^2+(y-y_0)^2}}\frac{\partial H_0}{\partial \zeta} \quad (5.35)$$

The second integral then becomes

$$I_2 = -\frac{ik}{2}\int_{s_2} v(y)\frac{x_0}{\sqrt{x_0^2+(y-y_0)^2}}\frac{\partial H_0}{\partial \zeta}dy \quad (5.36)$$

Collating the results for the four integrals (5.29), (5.25), (5.30), and (5.36) with (5.21) we find

$$v(p_1) = -\frac{ik}{2}\int_{s_2} v(y)\frac{x_0}{\sqrt{x_0^2+(y-y_0)^2}}\frac{\partial H_0}{\partial \zeta}dy \quad (5.37)$$

The argument ζ of H_0 is quite large, which means that we can use the expansion

$$H_0(\zeta) = J_0(\zeta)+iY_0(\zeta) \approx \sqrt{\frac{2}{\pi\zeta}}\left[\cos(\zeta-\pi/4)+i\sin(\zeta-\pi/4)\right] = \sqrt{\frac{2}{\pi\zeta}}\exp(i(\zeta-\pi/4)) \quad (5.38)$$

and

$$\frac{\partial H_0}{\partial \zeta} \approx \sqrt{\frac{2}{\pi\zeta}}\exp(i(\zeta-\pi/4))\left(i-\frac{1}{2\zeta}\right) \approx i\sqrt{\frac{2}{\pi\zeta}}\exp(i(\zeta-\pi/4)) \quad (5.39)$$

And finally

$$v(p_1) = \sqrt{\frac{k}{2\pi}}\int_{s_4} v(y)\frac{x_0}{\sqrt{x_0^2+(y-y_0)^2}}\frac{\exp(ik\sqrt{x_0^2+(y-y_0)^2}-i\pi/4)}{(x_0^2+(y-y_0)^2)^{1/4}}dy \quad (5.40)$$

The above form is most well suited for direct implementation in the simulation software, however, one often see it expressed in the following way [8, 9, 12, 14, 17, 19–23].

$$v(p_1) = \sqrt{\frac{k}{2\pi}}\int_{s_4} v(y)\cos\xi\frac{\exp(ik|\bar{r}-\bar{r}_1|-i\pi/4)}{|\bar{r}-\bar{r}_a|^{1/2}}dy \quad (5.41)$$

where ξ is the angle between the vector $\bar{r} - \bar{r}_1$ and the line normal $\hat{n}$, where $\bar{r}_1$ is the vector pointing towards p_1, and $\bar{r}$ points towards a point on line s_2.

5.3 Implementation in Matlab

5.3.1 Introduction

To implement (5.40) in our existing geometry we should swap x_0 for z_r, where z_r is the horizontal distance from the last phase screen to the receiving LEO satellite. Next we replace y_0 with y_r, and note that $dy \to \Delta y$. We get

$$v(p_1) = \sqrt{\frac{k}{2\pi}} \sum_{j=1}^{M-1} v(y[j]) \frac{z_r}{\sqrt{z_r^2 + (y[j] - y_r)^2}} \frac{\exp(ik\sqrt{z_r^2 + (y[j] - y_r)^2} - i\pi/4)}{(z_r^2 + (y[j] - y_r)^2)^{1/4}} \Delta y \tag{5.42}$$

This integral is readily implemented in Matlab. There is one caveat though: a sort of aliasing will appear when the phase screen grid length is too large. The reason for this is the change in the argument of the exponent in (5.40) due to the change in distance to the screen while going from one grid point to the adjacent one. The maximal change occurs when the horizontal distance from the screen is minimal, and the vertical distance from the point being "scanned" is maximal. This occurs at the upper black dot in Fig. 5.2. Before deriving the exact condition for this aliasing to appear, we need to specify the geometry of the problem.

5.3.2 The orbit and the occultation path

The most simple way to treat the orbit of a satellite is to constrain its motion to the same plane as that of the vertical axis (y) and the propagation direction of the field (z), and to assume that it is moving in a perfect circle. The assumption of a perfect circle is in no way mandatory, and one can equally well just use a two vectors r_r and Ω to define occultation path.

If there was in fact no refraction of the wavefront, the complete occultation would occur in the grey band of Fig. 5.2. Since there is some refraction, it is necessary to record the amplitude received by the LEO satellite in a wider band. In the coordinate system attached to the center of the Earth the band goes between $Y = R_E + H - L_y$ and $Y = R_E + H$. For the unrefracted case the occultation would then occur between Ω's given by

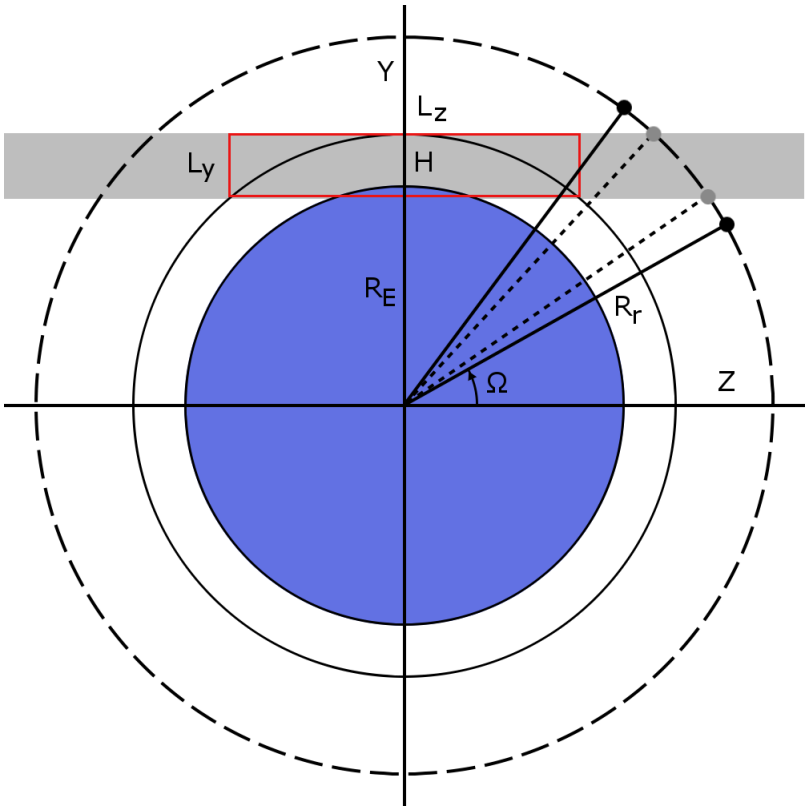

Figure 5.2: The simplest geometry for relating the orbit of the LEO satellite to the simulation geometry. In the case when the LEO orbit is not circular, minor modifications to the path inside the grey area will have to be done.

$$\Omega_{\text{upper}} = \arcsin\left(\frac{R_E + H}{R_r}\right) \tag{5.43}$$

$$\Omega_{\text{lower}} = \arcsin\left(\frac{R_E + H - L_y}{R_r}\right) \tag{5.44}$$

But we wish to extend the occultation band somewhat, so we take the interval to be

$$\Omega_{\text{upper}} = \arcsin\left(\frac{R_E + H + \Delta_+}{R_r}\right) \tag{5.45}$$

$$\Omega_{\text{lower}} = \arcsin\left(\frac{R_E + H - L_y - \Delta_-}{R_r}\right) \tag{5.46}$$

Of course, when simulating the wavefront in a real occultation the wavefront will be artificially limited vertically to the height of the phase screens, and there will be little point in using $\Delta_+ \neq 0$, as it would result in completely unrealistic fields.

The most convenient coordinate system from the point of view of the phase screens is attached to the lower left corner of the red calculation box. And to calculate the diffraction integral we only need to know the position of the satellite relative this system. The transformation to this system is achieved through

$$y = Y - (R_E + H - L_y) \tag{5.47}$$
$$z = Z + L_z/2 \tag{5.48}$$

and the coordinates of the start and end of the occultation are

$$z_{\text{start}} = R_r \cos \Omega_{\text{upper}} \tag{5.49}$$
$$y_{\text{start}} = R_r \sin \Omega_{\text{upper}} \tag{5.50}$$
$$z_{\text{stop}} = R_r \cos \Omega_{\text{lower}} \tag{5.51}$$
$$y_{\text{stop}} = R_r \sin \Omega_{\text{lower}} \tag{5.52}$$

for a setting occultation. For a rising occultation the order is reversed. All the coordinates of the satellite position within the occultation band is then given by $x = R_r \cos \Omega$, and $y = R_r \sin \Omega$. It seems prudent to use a fixed step $\Delta\Omega$ to discretize the trajectory of the satellite.

5.3.3 Aliasing

As stated in the introduction, aliasing appears when the exponent in (5.40) changes too rapidly. The point where it changes most rapidly is where the horizontal distance between the LEO satellite and the screen is minimal, and where the vertical distance to the calculation point is maximal. In Fig. 5.2 this occurs at the upper black dot. The distance from this point to the lower right corner of the calculation box is

$$R = \sqrt{(\Delta_+ + L_y)^2 + (R_r \cos \Omega_{\text{upper}} - L_z/2)^2} \tag{5.53}$$

The distance to the grid point just above it is

$$R' = \sqrt{(\Delta_+ + L_y - \Delta y)^2 + (R_r \cos \Omega_{\text{upper}} - L_z/2)^2} \tag{5.54}$$

So the change in distance between these two points is (provided $R \gg \Delta y$)

$$\Delta R = R - R' \approx \frac{\Delta y(\Delta_+ + L_y)}{\sqrt{(\Delta_+ + L_y)^2 + (R_r \cos\Omega_{\text{upper}} - L_z/2)^2}} \tag{5.55}$$

Since this is the point of maximal difference in distance from the satellite between neighboring grid points it gives rise to the maximal phase difference

$$\Delta\phi_{\max} = 2\pi\frac{\Delta R}{\lambda} \approx \frac{2\pi}{\lambda}\frac{\Delta y(\Delta_+ + L_y)}{\sqrt{(\Delta_+ + L_y)^2 + (R_r \cos\Omega_{\text{upper}} - L_z/2)^2}} \tag{5.56}$$

To avoid aliasing the signal needs to be sampled more than twice per cycle, wherefore the requirement becomes

$$\Delta\phi_{\max} < \pi \quad \Leftrightarrow \quad \Delta y < \frac{\lambda}{2}\frac{\sqrt{(\Delta_+ + L_y)^2 + (R_r \cos\Omega_{\text{upper}} - L_z/2)^2}}{(\Delta_+ + L_y)} \tag{5.57}$$

Let us take a particular example simply to demonstrate how the numerical error appears, and how sensitive the calculation is to violation of requirement (5.57). For the wavelength and screen sizes in our simulations, as the EM field travels on past the last phase screen, it will basically form a beam with roughly the same width as the screen, i.e. L_y. If we choose to project the field at a screen parallel to the last phase screen we should therefore almost replicate the amplitude profile on the last screen. Let us use a phase screen with a height of 300 km, and project it on a screen with a height of 600 km. We want symmetry, which means that $\Delta_+ = \Delta_- = 150$ km. Furthermore, since the screen is parallel to the last phase screen we can just set $(R_r \cos\Omega_{\text{upper}} - L_z/2)$ to some value independent of the other parameters. We set it to 1000 km. Inserting these values into (5.57) gives us

$$\Delta y < 1.22\lambda \ \text{ m} \tag{5.58}$$

And for the L1 band we have $f = 1.57542$ GHz, and $\lambda = 0.1904$ m, which gives us

$$\Delta y < 1.22\lambda \approx 0.23 \ \text{ m} \tag{5.59}$$

In the four figures 5.3-5.6 below we will investigate how sensitive the calculation method is to criteria (5.57).

The four figures 5.3-5.6 illustrate clearly that unwanted features appear when we violate condition (5.57). One can also infer that values of Δy

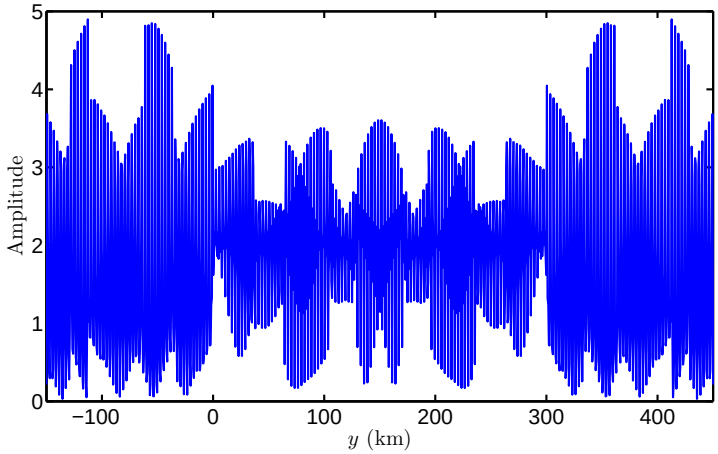

Figure 5.3: The projected field amplitude in the case when $\Delta y = 3$ m, which is far too large according to (5.57), which stipulates that the maximal Δy is around 0.23 m for this particular set of parameters.

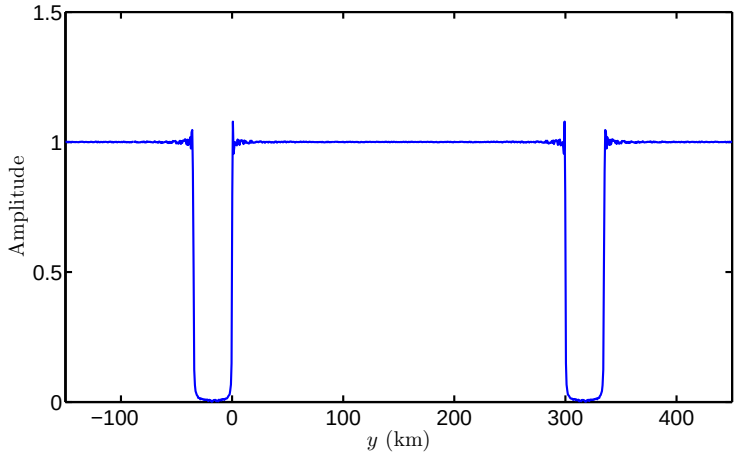

Figure 5.4: The projected field amplitude in the case when $\Delta y = 0.6$ m. Clearly, the magnitude of Δy is still too large.

that are only a fraction larger than the critical value will not give rise to any obvious errors. There will however be differences in the exact shape of the amplitude curve. It seems likely that the curve will continue to change its shape when one changes the grid resolution, and the resolution of the receiving plane (or the number measurement points in the LEO orbit during the occultation). The proper value for Δy will therefore depend on the refractivity data used, and can therefore only be found through trial and

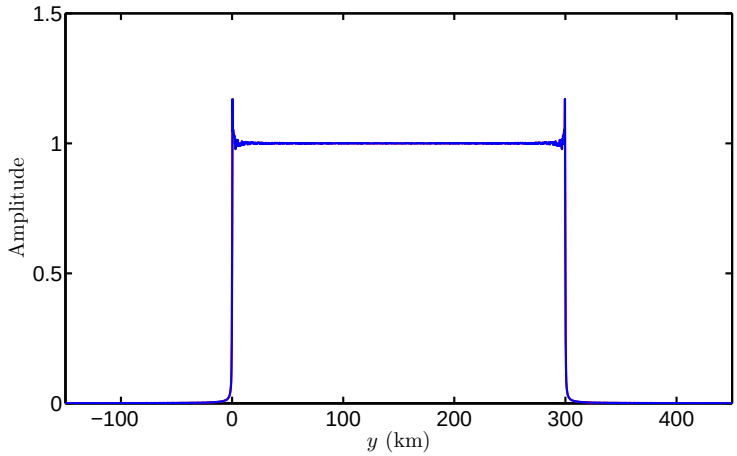

Figure 5.5: The projected field amplitude for the two cases $\Delta y = 0.2$ m (blue curve) and $\Delta y = 0.3$ m (red curve). At this scale there is no way to tell these two curves apart.

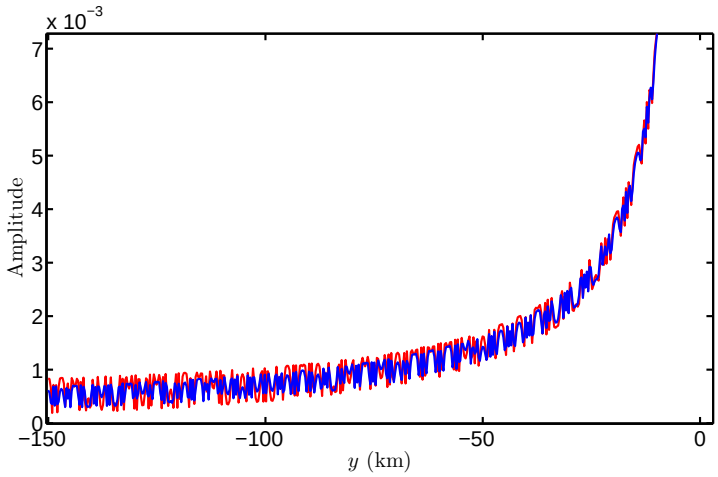

Figure 5.6: A zoom of the lower left corner of Fig. 5.5. Only in this region (and the similar right corner) and at this scale is it possible to see the difference between the curves. It is however quite small.

error in actual realistic simulations.

5.3.4 Stationary phase method

For integrals of the type

$$\int f(y)\exp(ik\phi(y))dy \tag{5.60}$$

the major contribution will come from the region where $\phi'(y) = 0$. In the case when $k \to \infty$ one can regard f as a constant, use a Taylor expansion for ϕ, and expand the integration domain to infinity to get a good approximation for the integral. This is called the stationary phase method. In our case we have $k \approx 30$ and this particular method will not give a very accurate value for the integral. However, the main idea is still sound. Since the main contribution to the integral comes from the region around $\phi'(y) = 0$ we do not have to integrate over the entire phase screen, but only in a limited region around $\phi'(y) = 0$.

In the case when the field on the last phase screen corresponds to a cylindrical wave the diffraction integral becomes

$$v(p_1) = \sqrt{\frac{k}{2\pi}}\int_0^{L_y} A\frac{\exp(ik\sqrt{(z_0+L_z)^2+(y-y_0)^2})}{((z_0+L_z)^2+(y-y_0)^2)^{1/4}}\times \tag{5.61}$$

$$\times\frac{z_r-L_z}{\sqrt{(z_r-L_z)^2+(y-y_r)^2}}\frac{\exp(ik\sqrt{(z_r-L_z)^2+(y-y_r)^2}-i\pi/4)}{((z_r-L_z)^2+(y-y_r)^2)^{1/4}}dy \tag{5.62}$$

Here

$$\phi = \sqrt{(z_0+L_z)^2+(y-y_0)^2}+\sqrt{(z_r-L_z)^2+(y-y_r)^2} \tag{5.63}$$

$$f(y) = A\sqrt{\frac{k}{2\pi}}\frac{1}{((z_0+L_z)^2+(y-y_0)^2)^{1/4}}\frac{z_r-L_z}{((z_r-L_z)^2+(y-y_r)^2)^{3/4}}\exp(-i\pi/4) \tag{5.64}$$

In this case the point of stationary phase can be found from

$$\frac{\partial}{\partial y}\left(\sqrt{(z_0+L_z)^2+(y-y_0)^2}+\sqrt{(z_r-L_z)^2+(y-y_r)^2}\right)=0 \tag{5.65}$$

with the solution

$$y = \frac{y_r(z_0+L_z)+y_0(z_r-L_z)}{z_0+z_r} \tag{5.66}$$

Using this knowledge we can investigate how sensitive the integral is to the integration range, and to the alignment of the integration domain with respect to the stationary phase point. In Fig. 5.7 the real part of the integrand

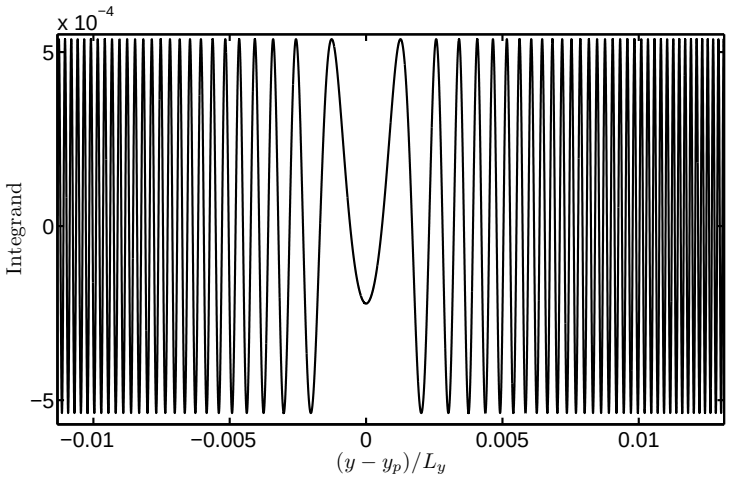

Figure 5.7: The real part of the integrand in Eq. (5.61). The stationary phase point can be seen clearly. The imaginary part of the integrand will display the same behavior.

in (5.61) is shown in the region around the stationary phase point y_p given by Eq. (5.66). In Figs. 5.8 and 5.9 one can see how the value of the integral converges as the integration domain expands. As Δy_p becomes larger than $0.1L_y$ the deviations of the integral from the average value becomes less than 1 %. Lastly, in Fig. 5.10 the effect of having the center of the integration region displaced from the location of the stationary phase is shown. Clearly, as long as the stationary phase point is included in the interval, the result is quite insensitive to the precise positioning of the center of the integration region.

The deviation from the mean value of the integral decreases quite slowly with increasing integration range. But this process can be sped up through the use of a Gaussian weight function in the integrand.

In Figs. 5.11-5.14 we see comparisons between the final amplitude and angle curves in orbit for the three methods: full integration; integration around the stationary phase point; and integration around the stationary phase point with a Gaussian weight. The amplitudes for the three methods do not agree completely, and in the flat region where the wave has undergone little refraction, the unweighted method produces oscillations around the correct value. Although the curves corresponding to the full integration range and the weighted method appear to be close they are not identical. They differ slightly, and it seems to be impossible to find a weight function that makes them agree completely. But in the end, the amplitude is not the most important factor. It is the phase angle that is crucial, in that it determines

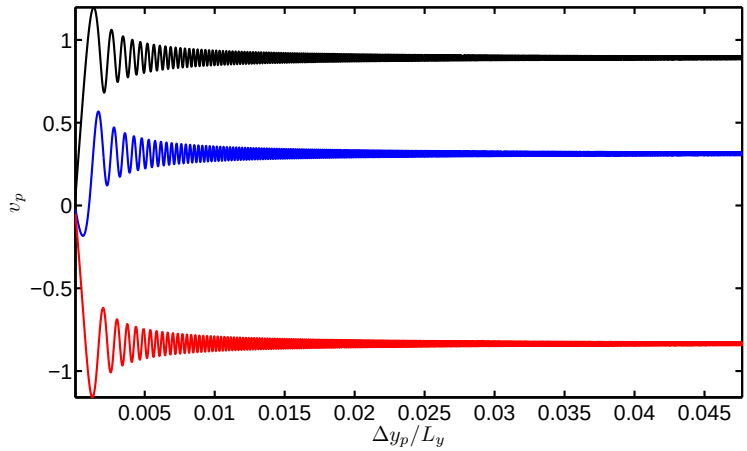

Figure 5.8: The imaginary part (blue curve), real part(red curve), and magnitude of the integral (5.61), and its dependence on the integration interval around the stationary phase point. As the interval grows the integral converges.

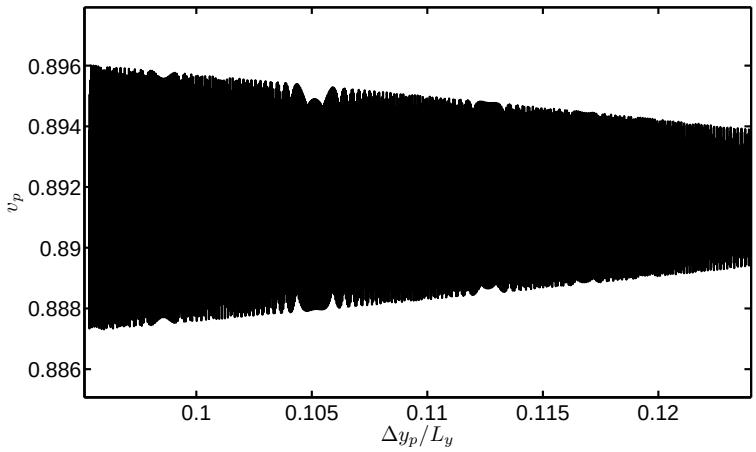

Figure 5.9: A zoom of the magnitude of the integral (5.61) to show that it converges monotonically with increasing integration interval.

the optical path length, and the excess phase. For the three methods in that case it is clear that they produce very similar results. And in fact, when applying the Gaussian weight, the curves agree almost completely. For these particular figures a weight on the form

$$\exp\left(-\frac{(y-y_p)^2}{D^2}\right) \tag{5.67}$$

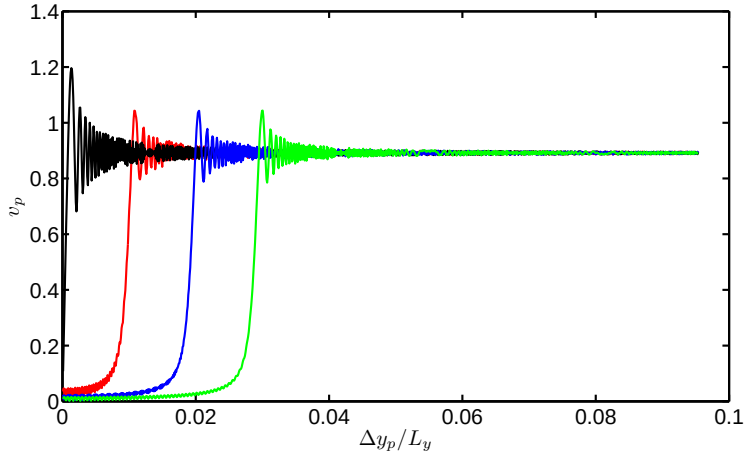

Figure 5.10: The dependence of the magnitude of the integral (5.61) on an offset of the center point of the integration interval from the stationary phase point. The black curve represents a correctly centered interval, the red curve represents an offset by $0.01\Delta y_p/L_y$, the blue curve 0.02, and the green curve 0.03.

has been used, where

$$D = \frac{K}{M}\frac{L_y}{2} \tag{5.68}$$

and

$$K = 0.2M \tag{5.69}$$

The length of the integration interval is then $2KL_y/M$, and the interval is symmetric around the stationary phase point y_p. These particular values were found by trial and error to produce the closest fit between the full integration results and the Gaussian weight method.

5.3.5 Finding the stationary phase point and determining the geometrical ray path

In a simulation where we are using a real refractive index profile there is no way to determine the stationary phase point for each position of the LEO satellite without performing an analysis of the entire signal on the last phase screen, either in order to find some rough formula for the phase as function of y, or to actually locate the stationary phase point for each LEO position. The other alternative is to use a generic atmospheric model to find a rough

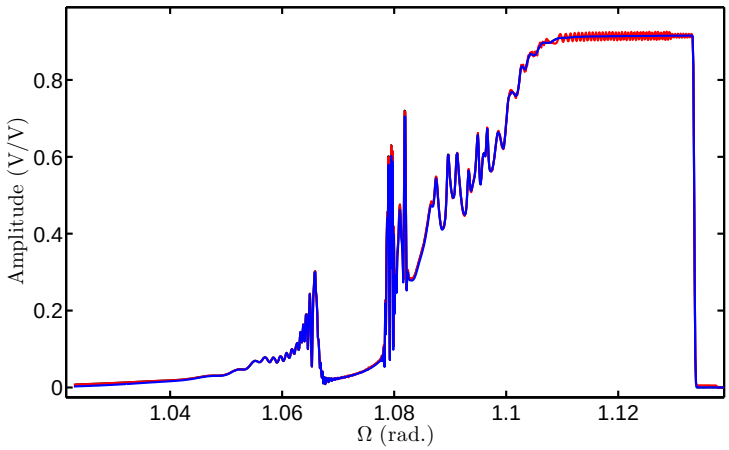

Figure 5.11: The amplitude in orbit for the three methods. Black curve corresponds to integration over the full screen, red to integration over a small interval around the stationary phase point, and blue to integration around the stationary phase point with a Gaussian weight function.

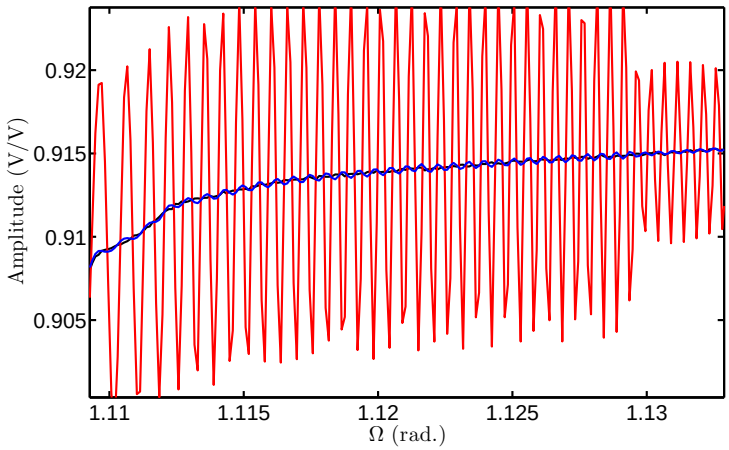

Figure 5.12: A zoom of the amplitude in orbit for the three methods in the flat region corresponding to very low refraction.

value for the stationary phase point. In this case one needs to analyze the geometry in Fig. 5.15. If R_0 and R_r are fixed, then for each value of ϕ_0 there is an associated α and ϕ_r (through $R_0 \sin\phi_0 = R_r \sin\phi_r = \alpha$). Due to the geometry there is one value of Ω_r for each ϕ_r, which means that we can in principle express Ω_r as a function ϕ_r, or the other way around. The last option is favorable since for each point in the orbit we can then find a value

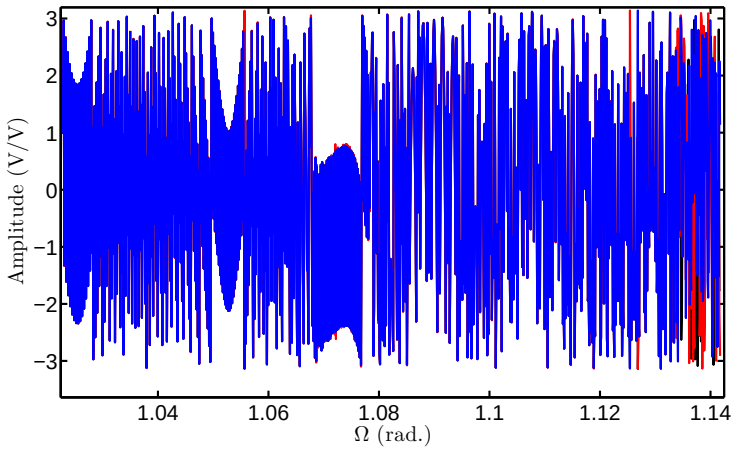

Figure 5.13: The phase angle in orbit for the three methods. Black curve corresponds to integration over the full screen, red to integration over a small interval around the stationary phase point, and blue to integration around the stationary phase point with a Gaussian weight function.

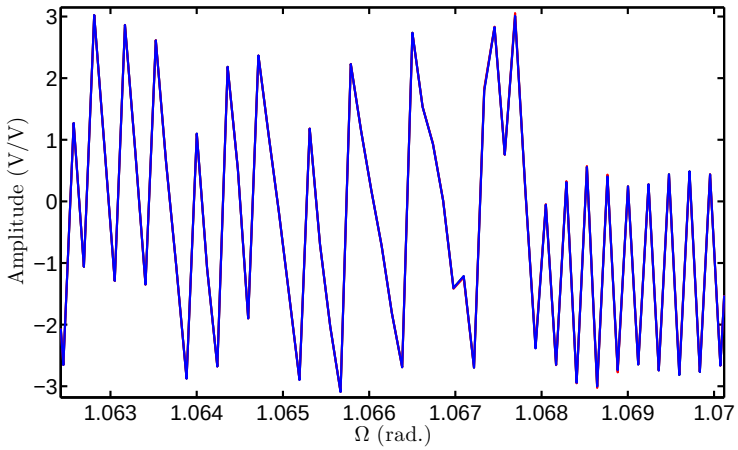

Figure 5.14: A zoom of the phase angle in orbit for an interval with high refraction.

for ϕ_r, and connect this to a position on the last phase screen. This position will in turn roughly correspond to the location of the stationary phase, due to the very nature of light wave propagation as expressed in Fermat's principle.

Given R_0 and the position of the GNSS satellite (Y_0, Z_0) we can calculate Ω_0. From this we can then find Ω_r through

$$\Omega_r = \Omega_0 + \phi_r + \phi_0 - \pi - \alpha \tag{5.70}$$

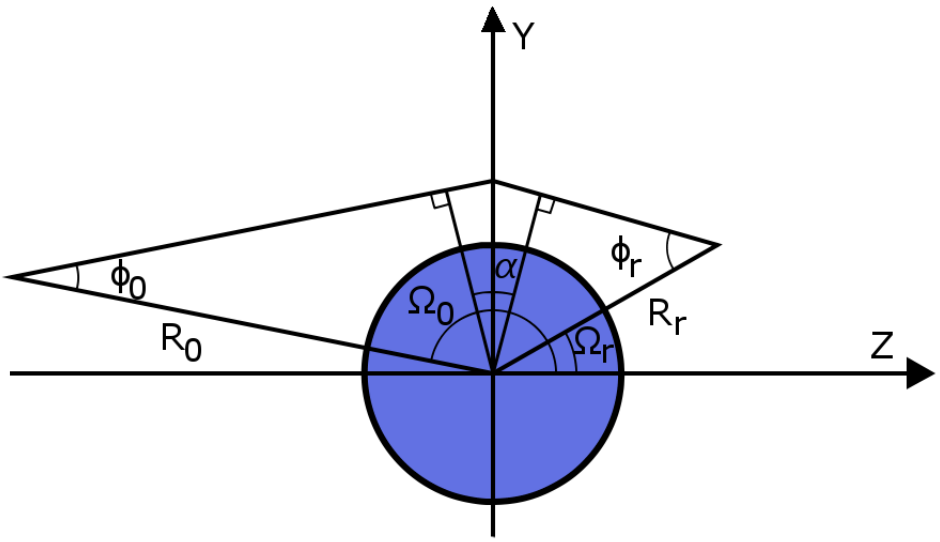

Figure 5.15: The geometry for connecting the geometrical ray parameters with the occultation geometry. The fact that the Y-axis seems to bisect the angle α is merely a coincidence in this particular case.

This can be inverted, and ϕ_r as a function of Ω_r can be found. Once ϕ_r is established, the stationary phase point (Y_p, Z_p) can be found through

$$Y_p^2 + Z_p^2 = (Y_r - Y_p)^2 + (Z_r - Z_p)^2 + R_r^2 - 2\sqrt{(Y_r - Y_p)^2 + (Z_r - Z_p)^2} R_r \cos\phi_r \tag{5.71}$$

Since Z_p is already known, the full solution is thus given by

$$Y_p = -\Lambda - \sqrt{\Lambda^2 + \Gamma} \tag{5.72}$$

where

$$\Lambda = \frac{Y_r}{Y_r^2 - R_r^2 \cos^2\phi_r} \left(Z_p Z_r - R_r^2 \sin^2\phi_r\right) \tag{5.73}$$

$$\Gamma = \frac{1}{Y_r^2 - R_r^2 \cos^2\phi_r} \left(R_r^2 \cos^2\phi_r \left(Z_p^2 - R_r^2 \sin^2\phi_r\right) - \left(R_r^2 \sin^2\phi_r - Z_r Z_p\right)^2\right) \tag{5.74}$$

This formula is quite complicated, but its veracity can be checked by inserting $Y_r = R_r \sin\phi_r$, and $Z_r = R_r \cos\phi_r$, in which case it should reduce to $Y_p = Y_r$.

Chapter 6

Unraveling the phase

6.1 Introduction

In orbit the GNSS signal is sampled at a specific rate, often 50 Hz, and since the occultation only takes some 100 s this means that only around 5000 sample points will be available for interpretation. The interpretation consists of determining the true relative phase of the signal during the entire occultation, and if this is to be done directly, 5000 sample points are far too few. Actually, to perform the unraveling of the phase directly, several hundred thousands of sample points are necessary. The most obvious way to find the phase as a continuous function over the occultation interval is to make a sophisticated guess as to what it should be and then remove this guess from the observed phase. This difference will then be (hopefully) sampled sufficiently often for it to be unraveled directly. In order to execute this scheme we therefore need to produce a good guess for the optical path length of the signal at each point in the occultation region. This is most easily done using a generic exponential refractivity profile, and employing geometrical optics to calculate the optical path length of a whole set of rays.

6.2 Optical path length from geometrical optics

The optical path length, τ, is given by the integral over the physical distance times the refractivity at each point

$$\tau = \int_0^{s_r} n ds \tag{6.1}$$

$$\tau = \int_{r_0}^{r_r} n(r) ds \tag{6.2}$$

Now we already know from section 9 that

$$rn(r) \sin \phi = a = r_t n_t \tag{6.3}$$

At the launch of the ray from the GNSS satellite at r_0 the ray moves towards the Earth, and the angle between the ray and a radial vector drawn from the center of the Earth is ϕ and smaller than π. In this case the increment in the path is related to the increment in the radial distance as

$$ds = -dr/\cos\phi = -\frac{nrdr}{\sqrt{n^2r^2 - a^2}} \tag{6.4}$$

When the ray crosses the tangent point and starts moving away from the Earth we have $\phi > \pi$, which means that

$$\cos\phi < 0 \quad \Leftrightarrow \quad \cos\phi = -\frac{1}{nr}\sqrt{n^2r^2 - a^2} \tag{6.5}$$

and

$$ds = -dr/\cos\phi = \frac{nrdr}{\sqrt{n^2r^2 - a^2}} \tag{6.6}$$

Combining we find

$$\tau = -\int_{r_0}^{r_t} \frac{rn(r)^2}{\sqrt{n(r)^2r^2 - a^2}} dr + \int_{r_t}^{r_r} \frac{rn(r)^2}{\sqrt{n(r)^2r^2 - a^2}} dr \tag{6.7}$$

We use $x = rn(r)$, which means

$$dr = \frac{dx}{n}\left(1 - \frac{x}{n}\frac{\partial n}{\partial x}\right) \tag{6.8}$$

Inserting this yields

$$\tau = -\int_{x_0}^{a} \frac{x}{\sqrt{x^2 - a^2}}\left(1 - x\frac{\partial \ln n}{\partial x}\right) dx + \int_{a}^{x_r} \frac{x}{\sqrt{x^2 - a^2}}\left(1 - x\frac{\partial \ln n}{\partial x}\right) dx \tag{6.9}$$

The first terms of the integrals can be integrated directly

$$\tau = \sqrt{x_0^2 - a^2} + \sqrt{x_r^2 - a^2} + \int_{x_0}^{a} \frac{x^2}{\sqrt{x^2 - a^2}}\frac{\partial \ln n}{\partial x} dx - \int_{a}^{x_r} \frac{x^2}{\sqrt{x^2 - a^2}}\frac{\partial \ln n}{\partial x} dx \tag{6.10}$$

Due to the fact that the refractive index goes to zero at x_0 and x_r we can rewrite the integrals slightly

$$\tau = \sqrt{x_0^2 - a^2} + \sqrt{x_r^2 - a^2} - 2\int_a^\infty \frac{x^2}{\sqrt{x^2 - a^2}} \frac{\partial \ln n}{\partial x} dx \tag{6.11}$$

The remaining integral can be recast in the form

$$-2\int_a^\infty \frac{x^2}{\sqrt{x^2 - a^2}} \frac{\partial \ln n}{\partial x} dx = -2\int_a^\infty \frac{x^2 - a^2}{\sqrt{x^2 - a^2}} \frac{\partial \ln n}{\partial x} dx - 2\int_a^\infty \frac{a^2}{\sqrt{x^2 - a^2}} \frac{\partial \ln n}{\partial x} dx \tag{6.12}$$

And using the fact that

$$\alpha = -2a\int_a^\infty \frac{\partial \ln n}{\partial x} \frac{dx}{\sqrt{x^2 - a^2}} \tag{6.13}$$

Eq. (6.12) becomes

$$-2\int_a^\infty \sqrt{x^2 - a^2} \frac{\partial \ln n}{\partial x} dx + \alpha a \tag{6.14}$$

Finally by using $x_0 = r_0$, $x_r = r_r$, and the results above we can write the optical path length as

$$\tau = \sqrt{r_0^2 - a^2} + \sqrt{r_r^2 - a^2} + \alpha a - 2\int_a^\infty \sqrt{x^2 - a^2} \frac{\partial \ln n}{\partial x} dx \tag{6.15}$$

The first two terms are given directly from the geometry of the system. The $a\alpha$ term can be found using the methods of section 9. The integral term contains no singularities and can be evaluated using standard methods, such as *trapz()* in Matlab. This also means that any interpolation can be used for the refractive index in the integral. This however is of little consequence, as the optical path length is only an aid in finding the true phase function, and the choice of interpolation will have no effect on the final results of the unraveling method.

6.2.1 Reducing the phase

Once the optical path has been calculated, the recorded phase in orbit can be reduced by

$$v' = v \exp(-ik_0\tau) \tag{6.16}$$

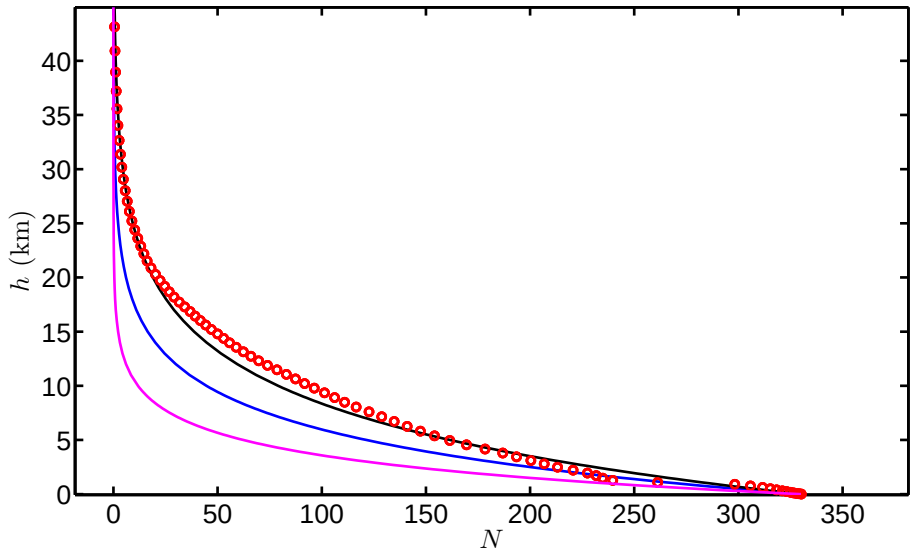

Figure 6.1: Three different generic profiles corresponding to $N_0 = 330$, and $H_{\text{scale}} = 7$ km (black curve), 5 km (blue curve), and 3 km (magenta curve) respectively. The red circles represents a real profile (case 05) that belongs to class 4. 5000 sample points have been used for the entire occultation path.

After this proceedure the phase can be unraveled using a far lower number of sample points than would be required to unravel the original phase. When the correspondence between the true refractivity profile and the generic exponential profile is poor, many sample points are needed to unravel the phase difference. Therefore, too few sample points, and a poorly matched exponential profile will give rise to the same error in the unraveled phase. In fact, the error generated by this proceedure has a very distinct appearance, which is fortunate as it is easy to spot and fix. For this reason this error will be investigated below. The generic profile is on the form

$$N(h) = N_0 \exp(-h/H_{\text{scale}}) \tag{6.17}$$

where $N_0 = 300 - 400$, and $H_{\text{scale}} = 6 - 8$ km, and the refractive index is given by $n = 1 + 10^{-6}N$. In Fig. 6.1 three different generic profiles are shown along with a real class 4 profile.

In Fig. 6.2 one can see the resulting unraveled phase difference using the three generic profiles. The error we are looking for appears in the magenta curve representing the profile with $H_{\text{scale}} = 3$ km. Close to the middle of this curve there is a very distinct zigzag pattern. The phenomena is reminiscent of the type of aliasing one will find when using a video camera to film a

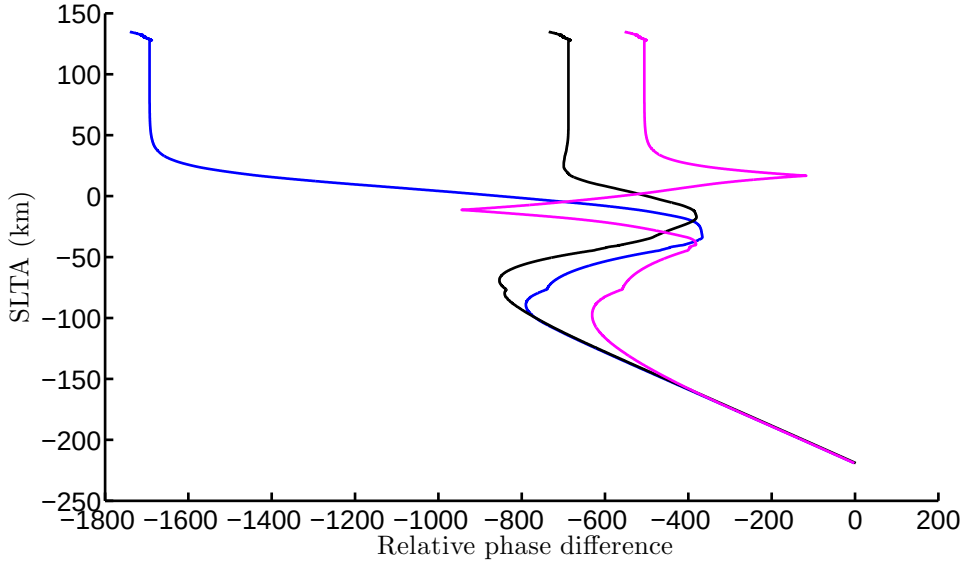

Figure 6.2: The unraveled phase difference using the three generic refractivity profiles from Fig. 6.1, and the case 05 occultation phase. One should note that the black and blue curves are quite smooth, whereas the magenta curve has a very distinct zigzag pattern. This is indicative of a poor match between the generic and true refractivity profile. The straight line tangent altitude (SLTA) on the y-axis is very commonly used, and explained in section B.

spinning wheel, and the wheel appears to go backwards, and sometime reverse its direction of rotation, even though this does not correspond to anything physical. These kinks and slopes disappear when the number of sample points is high enough for the phase difference to be completely unraveled. Therefore there are two ways of removing the error, either the number of sample points is increased (often a very impractical option when trying to simulate the performance of a real satellite on orbit), or the generic refractivity profile must be made to match the true profile better. In the case above 5000 sample points is used, and judging from the results for $H_{\text{scale}} = 5$ km, it appears that this number allows for quite poorly matched generic profiles. In fact, extensive testing with poorly matched profiles have shown that the method is not particularly sensitive.

Chapter 7

Inversion through phase matching

7.1 Introduction

Where it not for multipath, the phase of the signal received in orbit could in principle be directly inverted to yield the impact height vs. bending angle curve using geometrical optics. However, for most atmospheric conditions there will be multipath, and inversion cannot be performed using such crude methods. More sophisticated methods are available, that take into account both the phase and amplitude of the signal. For this reason the different methods are sometimes called radio-holographic. There is no consensus as to which is the best method, as there are several factors to take into account, e.g. speed vs. accuracy. There are basically three methods in use for inversion of RO measurements, these are the canonical transform (CT) [24], the full spectrum inversion (FSI) [25], and phase matching (PM) [26]. As for the benefits of the different methods it is hard to say. For it would require input from somebody that is well versed in all methods. But some things can be stated without detailed knowledge. The CT appears to be quite complicated mathematically, or at least the available publications are somewhat hard to understand. FSI is generally thought to be the fastest method, as it employs the Fourier transform, but it has the disadvantage of only working properly for circular orbits. PM is very similar to FSI, but it has the advantage of being easier to interpret from a physical point of view, and it works for arbitrary orbits. It has the disadvantage of being slower than FSI. We will be using PM in this report.

7.2 Phase matching

The signal that is received in orbit can be represented on the form

$$u(\Omega) = A(\Omega)\exp(ik_0\tau_c(\Omega)) \tag{7.1}$$

where τ_c is the composite optical path of the real signal, which for single ray regions is simply the optical path length of one ray (Eq. (6.15)).

$$\tau = \sqrt{R_0^2 - a^2} + \sqrt{R_r^2 - a^2} + \alpha a - 2\int_a^\infty \sqrt{x^2 - a^2}\frac{d\ln n}{dx}dx \tag{7.2}$$

whereas in the multipath region it is the superposition of several such rays. The phase matching method takes a test ray, τ_t, subtracts it from Eq. (7.1), and integrates over the occultation interval

$$U(\Omega) = \int_{\Omega_{\min}}^{\Omega_{\max}} A(\Omega)\exp(ik_0(\tau_c - \tau_t)) \tag{7.3}$$

The test ray is on the form

$$\tau_t = \sqrt{R_0^2 - a_t^2} + \sqrt{R_r^2 - a_t^2} + \alpha_t a_t \tag{7.4}$$

and α is given by

$$\alpha = \theta(\Omega) + \phi_r(a) + \phi_0(a) - \pi \tag{7.5}$$

$\theta(\Omega)$ is the angle of separation between the transmitter and receiver for a given Ω, and $\phi_{0,r} = \arcsin(a/R_{0,r})$.

Since $k_0\tau_c$ varies so rapidly, and its variation has nothing to do with the equally rapid variation of τ_t, the major contribution to the integral (7.3) will come from the region where $\tau_c \approx \tau_t$, and the actual value of the integral will be on the form

$$U(\Omega) = C(\Omega, a_t)\exp(ik_0(\tau_c(\Omega) - \tau_t(\Omega, a_t))) \tag{7.6}$$

The constant $C(\Omega, a_t)$ has no relevance for our purposes and can be ignored. The phase $\Psi(\Omega, a_t) = \tau_c - \tau_t$ is the relevant bit, and it is easily extracted in Matlab. We know from geometrical optics that in single ray regions it will be

$$\Psi(\Omega, a_t) \approx -2k_0\int_{a_t}^\infty \sqrt{x^2 - a_t^2}\frac{d\ln n}{dx}dx \tag{7.7}$$

If we take the derivative with respect to a_t, we get according to the Leibniz integral rule

$$\frac{\partial \Psi(\Omega, a_t)}{\partial a_t} \approx -2k_0 \left(\int_{a_t}^{\infty} \frac{\partial}{\partial a_t} \left[\sqrt{x^2 - a_t^2} \frac{d \ln n}{dx} \right] - \sqrt{x^2 - a_t^2} \frac{d \ln n}{dx} |_{x=a_t} \right) \tag{7.8}$$

Or

$$\frac{\partial \Psi(\Omega, a_t)}{\partial a_t} \approx 2k_0 \int_{a_t}^{\infty} \frac{a_t}{\sqrt{x^2 - a_t^2}} \frac{d \ln n}{dx} dx = -k_0 \alpha(a_t) \tag{7.9}$$

So, in single ray regions we can find the bending angle through

$$\alpha(a_t) = -\frac{1}{k_0} \frac{\partial \Psi}{\partial a_t} \tag{7.10}$$

The remarkable thing is that the method appears to work in multipath regions as well. To the knowledge of the author nobody has documented the exact reason for this, but a possible explanation is given in section 7.4.

7.3 Implementation in Matlab

There is one problem that appears when the method is implemented in Matlab; the integral does not give behave well if the occultation signal is not sampled sufficiently often. In fact, if one tries to use the method directly on a real signal with some 5000 sample points the output will be horrible. Consequently the signal needs to be "up-sampled" using some suitable method. The most obvious way to do this is to use the unraveled signal that is found using the methods of chapter 6. The reduced phase, which is the difference between the phase of the signal and that of an exponential generic profile, can be unraveled, and the true phase is found by adding the difference to the exponential phase function. Since the true phase function is known one can easily interpolate it to a much finer vector using *interp1* in Matlab. That is, we know $\tau(\Omega)$, and we interpolate it on a Ω_f (which has higher resolution than Ω) to yield $\tau_f(\Omega_f)$. The amplitude of the signal, $A(\Omega)$, varies quite slowly, and can be interpolated to the finer vector directly to yield $A_f(\Omega_f)$. The up-sampled signal is then given by

$$u_f(\Omega_f) = A_f(\Omega_f) \exp(ik_0 \tau_f(\Omega_f)) \tag{7.11}$$

The number of sample points needed can be investigated simply by observing how the integral U changes as the number of sample points M is

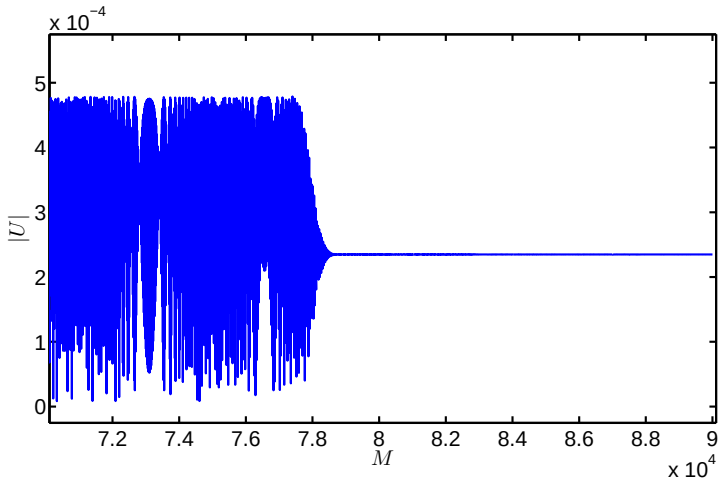

Figure 7.1: An illustration of how the magnitude of the integral Eq. (7.3) depends on the number of sample points, M, for a given test ray τ_t.

increased. The results can be seen in figs. 7.1 and 7.2. Clearly, some 90000 points should be sufficient.

7.4 Possible explanation for the functionality of phase matching in multipath regions

We can formulate a tentative explanation based on Fermat's principle. It states that "rays of light traverse the path of stationary optical length with respect to variations of the path". For multipath regions there are overlapping signals so that the complex signal u is given by

$$u = \sum_{i=1}^{N} u_i = \sum_{i=1}^{N} A_i \exp(ik_0\tau_i) \tag{7.12}$$

Now, according to Fermat's principle we have

$$\delta\tau_i = 0 \tag{7.13}$$

that is, the path is optical path length is stationary for each ray. The variation δ is somewhat arbitrary and depends on the system under scrutiny. In our case we know that the optical path for each ray is given, to a high degree of accuracy, by

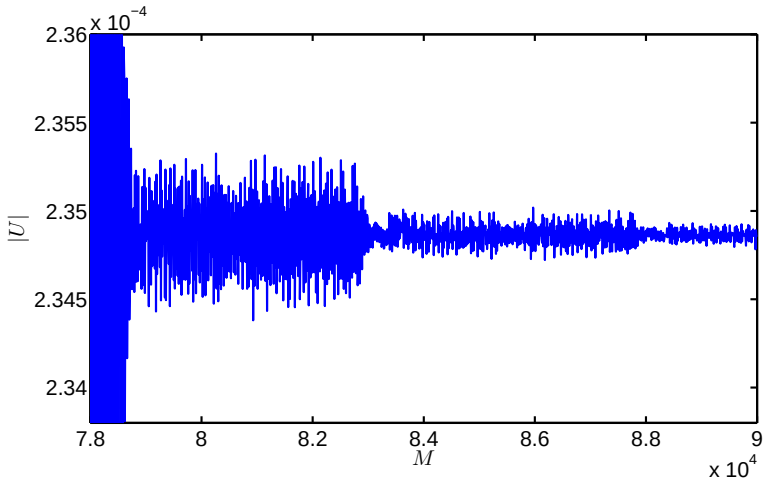

Figure 7.2: Same as fig. 7.1 but zoomed in to reveal the convergence for high M.

$$\tau_i = \sqrt{R_0^2 - a_i^2} + \sqrt{R_r^2 - a_i^2} + \alpha_i a_i - 2\int_{a_i}^{\infty} \sqrt{x^2 - a_i^2}\frac{d\ln n}{dx}dx \qquad (7.14)$$

and the nonmonotonic function n causes several values of a_i and $\alpha_i(\theta, a_i)$ to yield rays that converge at a specific point given by θ. The variation should thus be performed with respect to a while keeping θ fixed. For a particular point in the occultation we thus have

$$\frac{\partial \tau_i}{\partial a}|_{\theta(\Omega)} = 0 \qquad (7.15)$$

where the index indicates that $\theta(\Omega)$ is kept fixed. When we apply the trial ray τ_t and perform the integral we get something like this

$$U(\Omega, a_t) = C(\Omega, a_t)\exp(ik_0(\sum_{i=1}^{N}\tau_i - \tau_t)) \qquad (7.16)$$

Since the point where the exponential varies most slowly should be the point where the trial ray matches one of the rays in the sum, let us say ray j. The phase function then becomes

$$\Psi(\Omega, a_t) \approx \sum_{i=1}^{j-1}\tau_i + \sum_{i=j+1}^{N}\tau_i - 2\int_{a_j}^{\infty}\sqrt{x^2 - a_j^2}\frac{d\ln n}{dx}dx \qquad (7.17)$$

now it appears obvious that when taking the derivative with respect to a_t, the contribution from all rays except j disappears on account of Eq. (7.15), and we are left with

$$\frac{\partial \Psi}{\partial a_t} \approx -k_0 \alpha_j(a_j) \tag{7.18}$$

Chapter 8

Results in LEO

8.1 Introduction

The final stage in the evaluation of the entire WOP package is to make detailed comparisons between simulations with and without noise and GO (using the Abel transform). In the case when we add noise, RUAG has specified how large the total discrepancy (including all sources of noise and error) can be for different heights, and how large filters may be

- 0.5 μrad or 0.2 % of the BA over the altitude range between 35 to 80 km, (whichever value is larger)
- from 0.2 % to 0.5 % of the BA linearly in height between 35 km and 10 km
- from 0.5 % to 5 % of the BA linearly in height between 10 km and the surface
- Any filter that is used should be compatible with the Fresnel zone width

We will try our best to reduce the discrepancy between simulations and GO to under these levels. In order to perform detailed comparisons between simulations and GO we first need to make sure that the same interpolation technique for the refractivity is used in both cases, namely the cubic spline. Then we need to check what the resolution in the refractivity vector needs to be, and how high it must reach, to result in highly accurate Abel transforms. After these initial steps we will compare simulations with and without noise with GO, as well as compare simulations with and without noise with each

other. When performing statistical analysis on the discrepancies we will use the 55 reference atmospheres defined by ECMWF [27]. These atmospheres have been chosen to be representative for all type of climates, covering all seasons, including the most problematic profiles with strong moisture related feautures that lead to ducting and super-refraction.

8.2 A better Abel transform

8.2.1 The linear vs. cubic spline interpolation

The 55 reference cases [27] have a rather course sampling distance in height. For this reason we chose to use a linear interpolation scheme for the numerical Abel transform (see section 9). The WOP on the other hand has employed a cubic spline in order to produce a smooth refractive index profile. This leads to discrepancies between the bending angle curves found from the Abel transform and the WOP and phase matching. This discrepancy will be larger than the tolerances stipulated by RUAG, and it has to be remedied in some way before the accuracy of the simulation package can be evaluated. The easiest way is to stick with the original linear interpolation in the numerical integration used in the Abel transform, but to up-sample the refractive index to a much finer height-vector before applying the Abel transform. The up-sampling should be performed using the cubic spline. Figs. 8.1 to 8.3 demonstrate that the error caused by the interpolation scheme goes from significant for low heights to quite large for medium heights, and back to significant for high heights. Clearly it is necessary to use the same interpolation scheme for simulations and the Abel transform. The remaining question is how high the resolution in the refractivity vector needs to be for the Abel transform to be sufficiently accurate. In Figs. 8.4 to 8.6 the relative difference between the Abel transform using 100 m and 10 m resolution in the up-sampled refractivity is shown. Clearly it is necessary to have a very high resolution to yield sufficiently accurate Abel transforms. In the end 5 m was used for the comparisons in the following sections.

8.2.2 Extending the height of the refractivity

The 55 reference cases [27] generally spans the range 0-80 km above the Earth's surface. If one does not extend the refractivity vector to higher heights the integration in the Abel transform will give inaccurate results. At high impact heights the bending angle will be very small ($\sim 10^{-7}$) and the difference might be hard to see without detailed comparisons. However,

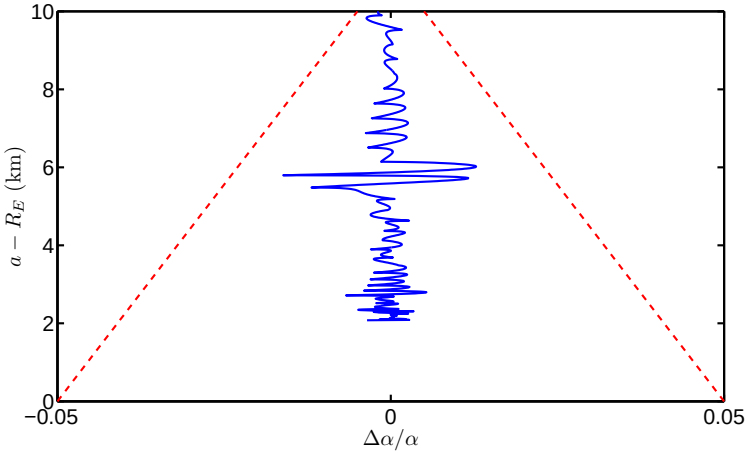

Figure 8.1: The relative difference between the Abel transform using pure linear interpolation and the Abel transform using the cubic spline to generate a smoother refractivity profile on a fine height vector. The curves are for case 01, a class 1 profile. Clearly, within this height segment (0-10 km) the relative difference is within the specified bounds, but not insignificant.

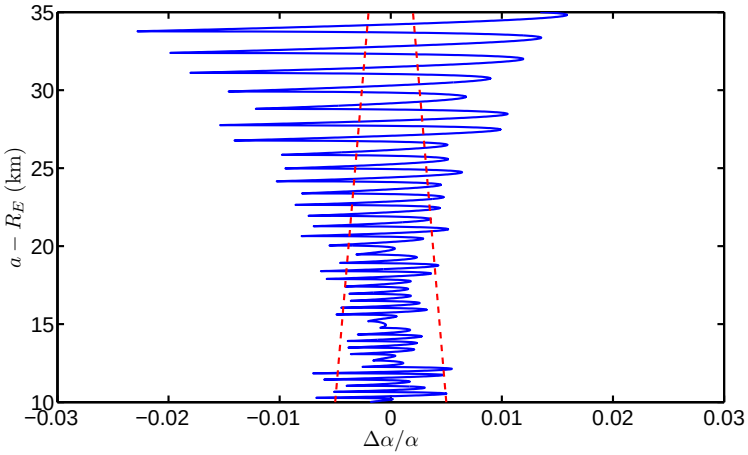

Figure 8.2: Same as Fig. 8.1 but for the height segment 10-35 km. The relative difference is no longer within the specified bounds.

in the WOP it is absolutely necessary to smoothly extend the refractivity to the height of the phase screens, and when comparing between the Abel transform and simulations the difference can become significant indeed. For high heights the refractivity decays exponentially, meaning

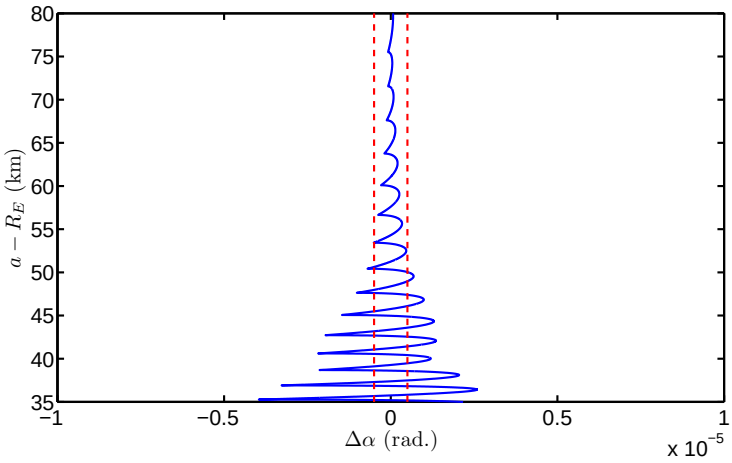

Figure 8.3: Same as Figs. 8.1 and 8.2 but showing the absolute difference for the height segment 35-80 km. The absolute difference is outside the specified bounds for low heights, but decreases with height.

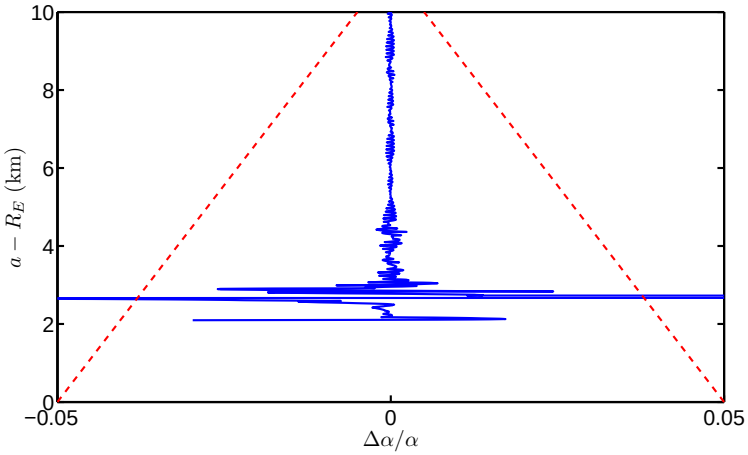

Figure 8.4: The relative difference between cubic spline with 100 and 10 m interpolation length for case 15, a class 4 profile, in the height segment 0-10 km. Only in the most critical regions is the relative difference outside the bounds. But this illustrates the need for high resolution interpolation.

$$N = N_E \exp(-h/H_E) \tag{8.1}$$

where the constants N_E and H_E are valid above the highest point in the reference file height vector. If the index for the final data point is i_f we have

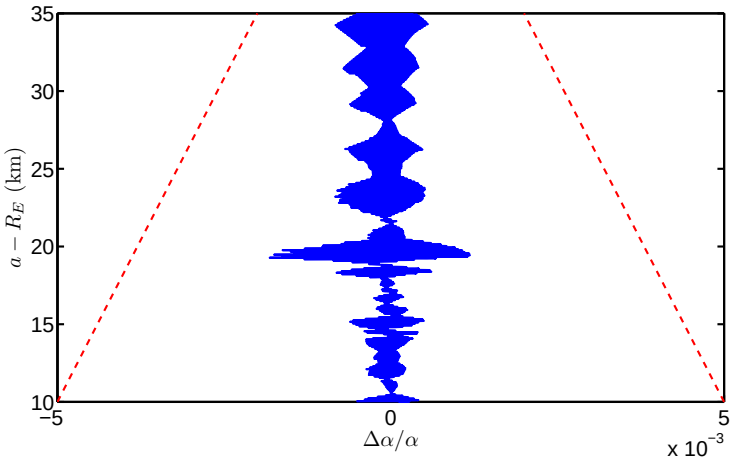

Figure 8.5: Same as Fig. 8.4 but in the height segment 10-35 km. The relative difference is significant.

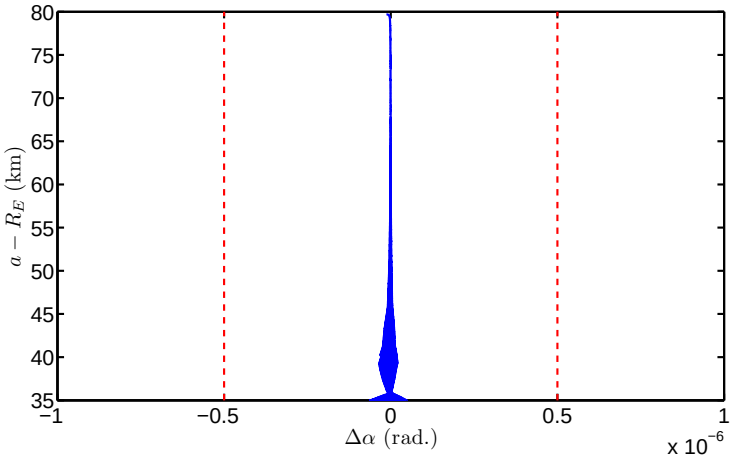

Figure 8.6: Same as Figs. 8.4 and 8.5 but showing the absolute difference in the height segment 35-80 km. The absolute difference is now quite insignificant.

$$H_E = \frac{h(i_f - 1) - h(i_f)}{\ln(N(i_f)/N(i_f - 1))} \tag{8.2}$$

$$N_E = N(i_f) \exp(h(i_f)/H_E) \tag{8.3}$$

In Figs. 8.7 and 8.8 one can see how the bending angle for high heights depends on the final height of the refractivity data. In 8.8 it can be seen that it is not necessary to go beyond 200 km, as the difference between the

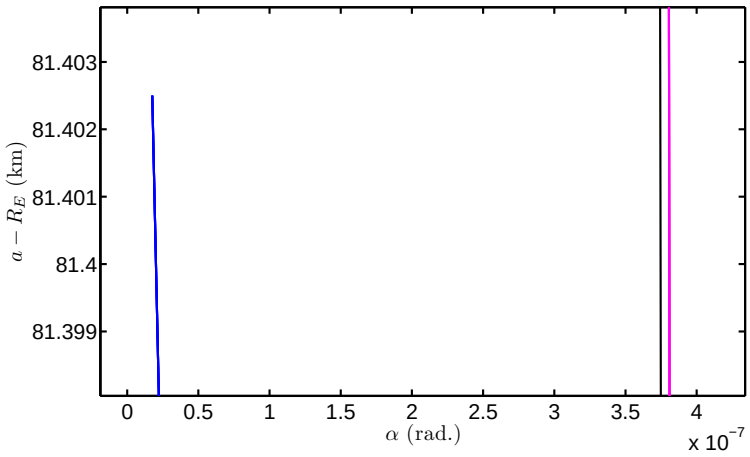

Figure 8.7: The bending angle around 80 km for the maximum heights being 80 km (blue line), 100 km (black line), and 150 km (magenta line).

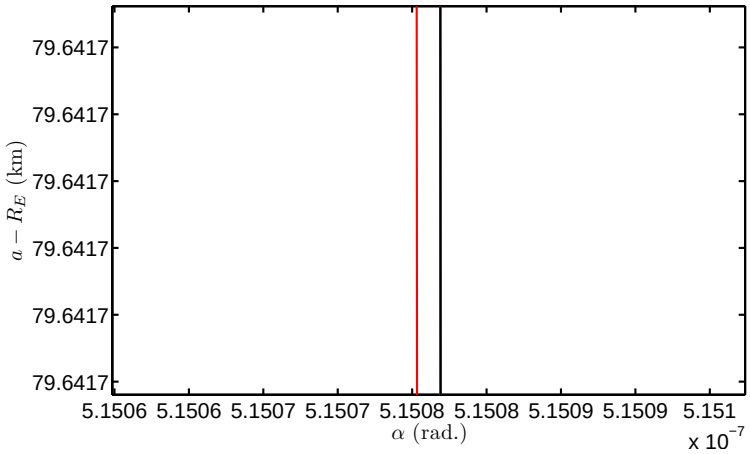

Figure 8.8: The bending angle around 80 km for the maximum heights being 150 km (red line), 200 km (black line), and 250 km (magenta line). The black and magenta lines can hardly be separated, and the difference between the bending angles is very small.

bending angle for 200 and 250 km is no larger than 10^{-12} rad., which is far below the 0.5μrad. RUAG limit.

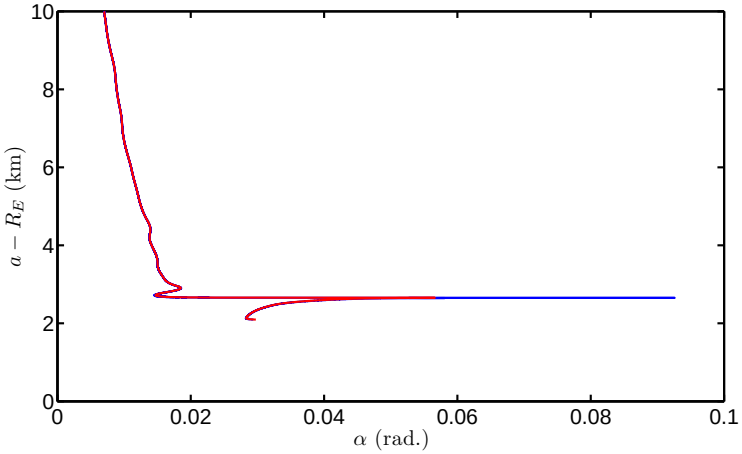

Figure 8.9: The bending angle from simulations (blue curve) and GO (red curve) in the height segment a =0-10 km for case 15 (class 4). On this scale it is very hard to see any difference.

8.3 Comparisons between simulations without noise and GO

When one is pleased with the resolution and accuracy of the Abel transform (GO) one can compare it with the results of the simulations without adding noise. This will provide us with a quick check of the accuracy of the simulation software. Case 15 (class 4) is one of the most problematic profiles, and it contains all the representative features of an interesting profile. Therefore we will display results for this case, and assume that it is a worst case. In Fig. 8.9 we can see a comparison between the simulations (blue curve) and the Abel transform (red curve) in the absence of noise. On this scale it is very hard to see any difference. If we zoom in on the super-refractive region (Fig. 8.10) we see that there indeed are differences. The peaks corresponding to the superrefractive region are slightly misaligned, and the blue curve displays some type of ripples.

It is more useful to look at the difference between the simulations and GO, and see how it complies with the RUAG requirements. In Fig. 8.11 the relative difference between simulations and GO is shown within the height segment $a = 0 - 10$ km. The discrepancy is within the specified bounds everywhere except at the superrefractive region. If one goes to higher impact heights the relative error increases rapidly, and there is no hope of staying within the bounds without applying some kind of filtering.

In Fig. 8.12 a simple Gaussian filter has been applied to the lower segment

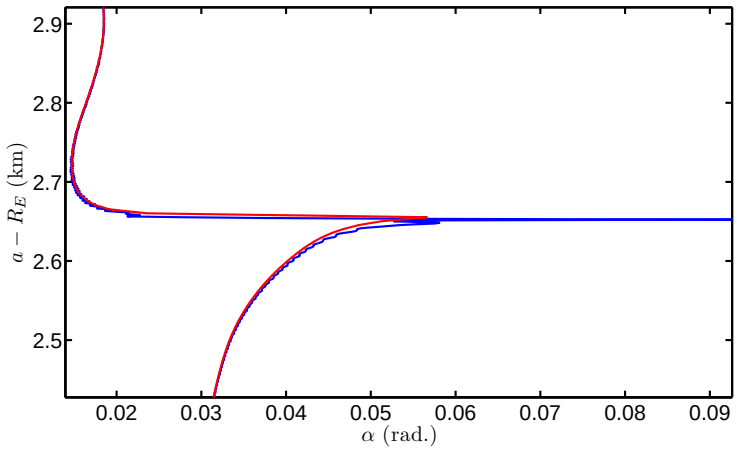

Figure 8.10: Same as Fig. 8.9 but zoomed in around the super-refractive region. The curves are now seen to be slightly misaligned, and some slight rippling in the simulated curve can be seen.

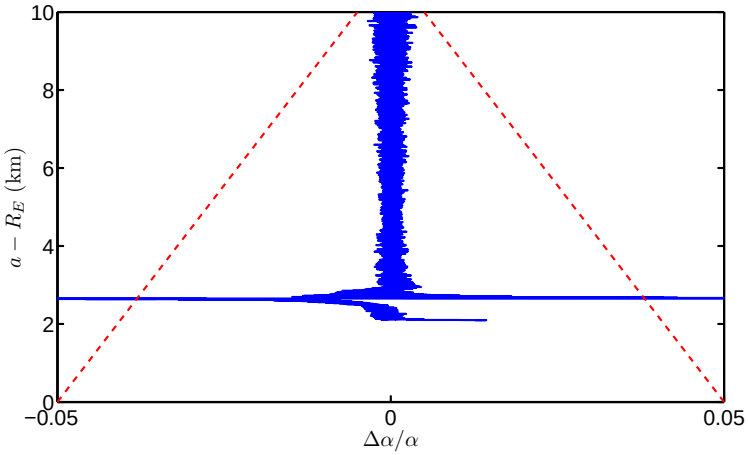

Figure 8.11: The relative difference between simulations and GO between $a = 0$ and 10 km for case 15. Clearly the super-refractive region gives rise to a very large local discrepancy.

of the bending angle. The filter is given by

$$y(x) = A \exp(-x^2/(2\sigma^2)) \tag{8.4}$$

where σ is the standard deviation, x a vector that has a total length given by the window size w, and A is a normalization constant. In Fig. 8.12 the variance was set to $\sigma = 15$ m, and the window size to $w = 150$ m. The noise

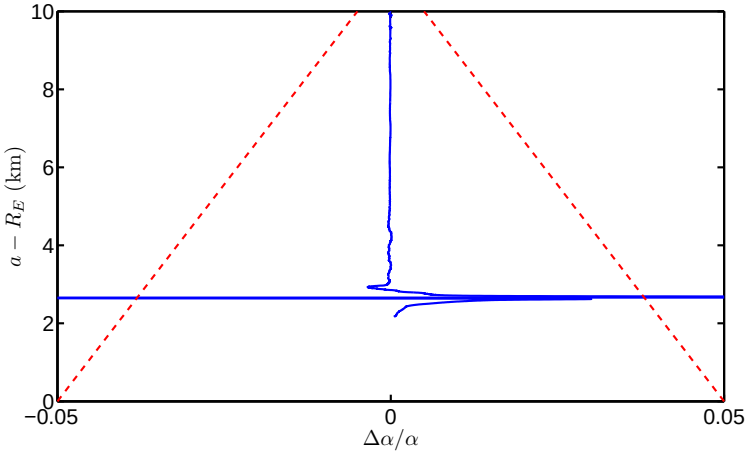

Figure 8.12: Same as Fig. 8.11 but after filtering with $\sigma = 15$ m, and $w = 150$ m. The numerical noise is successfully suppressed everywhere except in the super-refractive region.

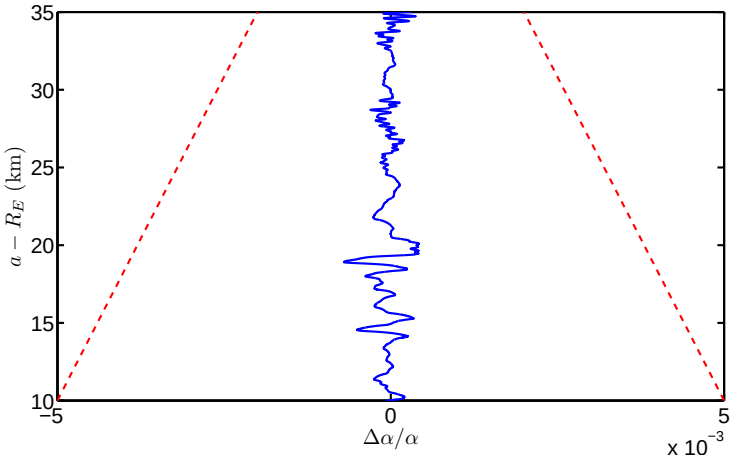

Figure 8.13: Same as in Fig. 8.12 but in the height segment 10-35 km. The numerical noise has been suppressed to a tolerable level using a filter with $\sigma = 60$ m and $w = 600$ m.

has been reduced significantly, but the discrepancy in the superrefractive region remains. In Fig. 8.13 the segment between 10 and 35 km is shown. A Gaussian filter with $\sigma = 60$ m, and $w = 600$ m. The noise has been reduced to a tolerable level. In Fig. 8.14 the segment between 35 and 80 is shown. Using a Gaussian filter with $\sigma = 150$ m and $w = 1500$ m the numerical noise has been reduced to insignificant levels.

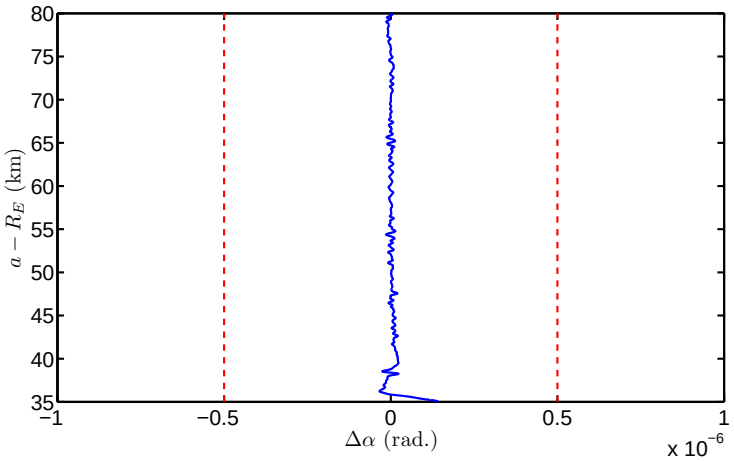

Figure 8.14: Same as in Figs. 8.12 and 8.13 but showing the absolute difference in the height segment 35-80 km. The numerical noise has been suppressed to an insignificant level using $\sigma = 150$ m and $w = 1500$ m.

8.4 Comparisons between simulations with noise and GO

8.4.1 White noise in RO

In real occultations the signal will be full of noise coming from the receiver etc. It is therefore vital that the procedures we are using are robust enough not to be seriously distorted by the noise. To evaluate the capacity of the procedures we will add noise to the signal that is received in LEO, and observe how it affects the end result, i.e. the α vs.a-curves. We will also apply the Gaussian filter in a partially successful attempt to reduce the difference between simulations and GO below the prescribed limits.

The noise we add is white, meaning that it is equally strong at each frequency. The amount of noise power received by the equipment is therefore proportional to the bandwidth of the receiver. The GRAS instrument samples at a rate of $f_s = 250$ Hz, before downconverting the sample rate to 50 Hz. The most rapidly oscillating signal that can be detected has a frequency of $f_s/2$, which means that the bandwidth is

$$B = \frac{f_s}{2} = \frac{250}{2} = 125 \ \text{ Hz} \tag{8.5}$$

For the GRAS instruments the noise level for occultations with high impact height is 50 dB-Hz [1]. The signal to noise ratio is given by

$$\mathrm{SNR}(dB) = \frac{P}{P_N}(\mathrm{dB}) = 10\log\frac{P(W)}{P_N(W)} = 10\log\frac{P(W)}{N(W/Hz)B(Hz)} \tag{8.6}$$

where N is the noise power per Hz, and B the bandwidth in Hz. We can develop the expression

$$\frac{P}{P_N}(\mathrm{dB}) = 10\log P(W) - 10\log N(W/Hz) - 10\log B(Hz) \tag{8.7}$$

It is the term $-10\log N$ which is 50 dB-Hz. For a signal power of unity (which we shall henceforth use in this section) the noise power then becomes

$$P_N = N \times B = 10^{-5} \times B = 10^{-5} \times 125 = 1.25 \times 10^{-3} \tag{8.8}$$

and the noise amplitude

$$A_N = \sqrt{10^{-5} \times B} = \sqrt{1.25 \times 10^{-3}} \approx 0.0354 \tag{8.9}$$

The amplitude of the noise is supposed to be interpreted as the standard deviation for the amplitude of the complex noise vector at any time instant. The distribution is normal (Gaussian) and the complex noise vector, $\mathcal{A}_N$, can be generated in Matlab using

$$\mathcal{A}_N = A_N(\mathrm{randn}(1) + i \times \mathrm{randn}(1)) \tag{8.10}$$

This complex number is simply added to the original simulated signal that is received in orbit, and all procedures after (unraveling, phase matching) are performed as usual.

8.4.2 Results for case 15

In Fig. 8.15 we can see the signal amplitude in orbit after addition of the noise. Using the unraveling technique of section 6 and the phase matching of section 7 without alteration we get the resulting bending angle curve shown in Fig. 8.16. Clearly the noise magnitude grows with decreasing impact height. this is not strange, for as can be seen in Fig. 8.15 the signal amplitude goes down for low SLTA (i.e. low impact heights), and the noise-to-power ratio goes up.

In Fig. 8.17 a Gaussian filter having $\sigma = 20$ m and $w = 200$ m has been applied. The noise has been successfully suppressed below the limits everywhere except near the super-refractive region. In Fig. 8.18 the region

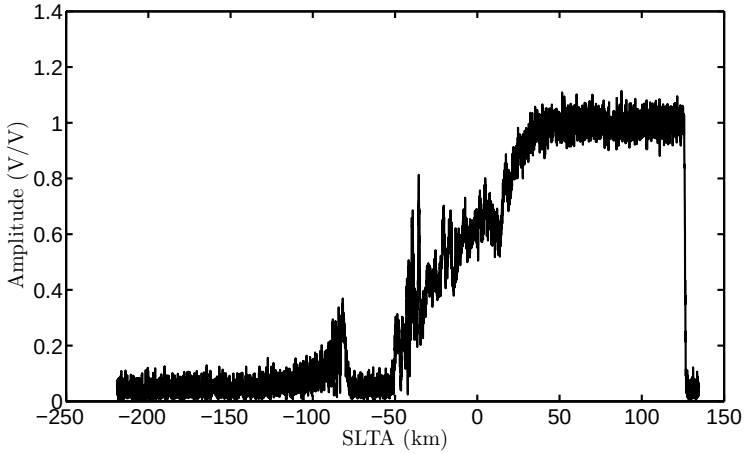

Figure 8.15: The case 15 (class 4) amplitude profile with added noise (50 dB-Hz).

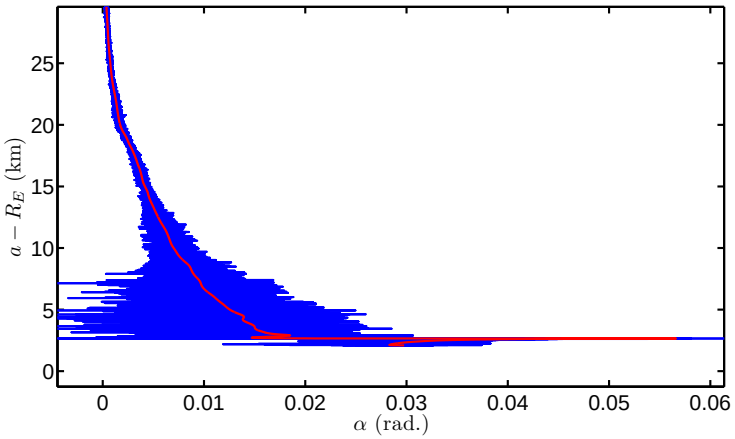

Figure 8.16: The case 15 (class 4) bending angle profile with added noise (50 dB-Hz) (blue curve), vs. the GO bending angle (red curve).

between 10 and 35 km can be seen. It has been filtered using $\sigma = 150$ m and $w = 1500$ m, but the curve does not stay within the boundaries except at low impact heights. Increasing σ and w tends to increase the difference for low heights, and decrease it for higher heights. Finally in Fig. 8.18 the upper segment between 35 and 80 km can be seen. A filter with $\sigma = 600$ m and $w = 6000$ m has been used, and it appears to be no problem keeping the rms difference within the bounds.

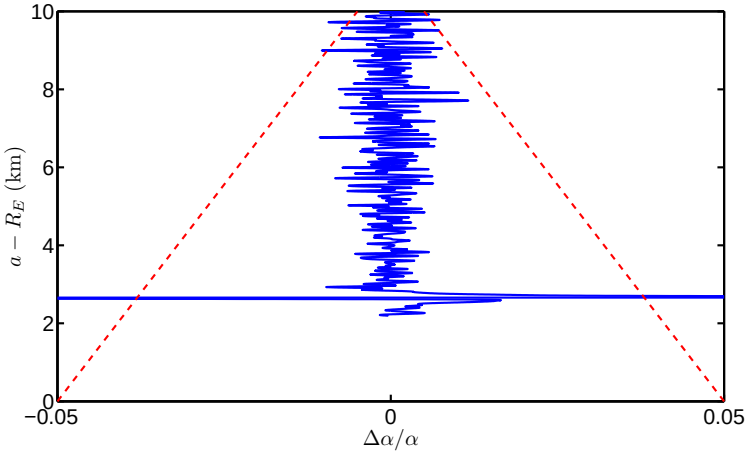

Figure 8.17: The relative difference between simulations and GO for case 15 with added noise and filtered with a Gaussian filter ($\sigma = 20$ m, $w = 200$ m) in the height segment 0-10 km. The relative difference keeps roughly within the bounds everywhere except in the super-refractive regions.

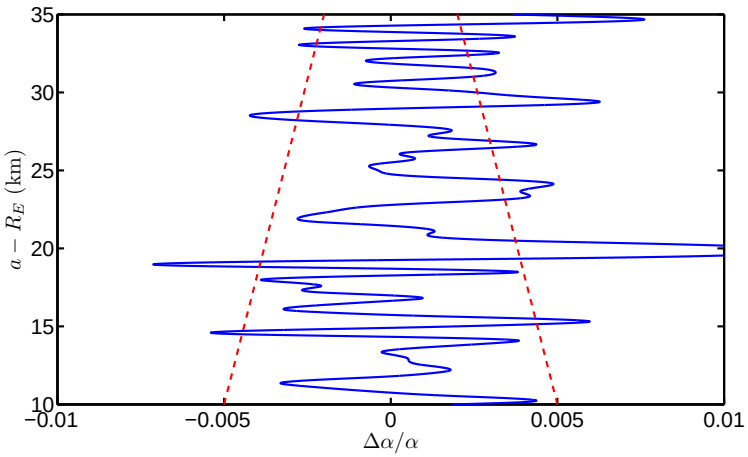

Figure 8.18: Same as Fig. 8.17 but with $\sigma = 150$ m and $w = 1500$ m, and in the height segment 10-35 km. The relative difference does not stay within the bounds.

8.5 Statistical investigation of the 55 reference cases

What we ultimately want to know is whether the methods described in this report are accurate enough to comply with the RUAG standard. To do this

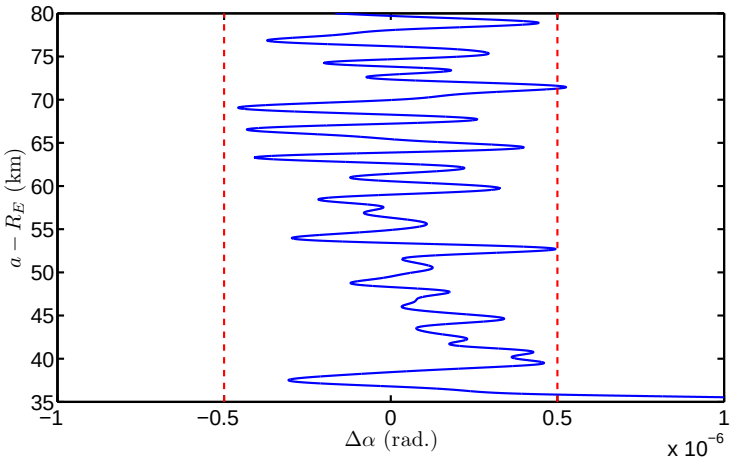

Figure 8.19: Same as Figs. 8.17 and 8.18 but showing the absolute difference using a filter with $\sigma = 600$ m and $w = 6000$ m, in the height segment 35-80 km. The absolute difference stays mainly inside the bounds.

we will have to perform a statistical analysis of the 55 reference cases [27]. It is quite straightforward to find the rms of the difference in bending angle. The resulting profiles from the phase matching procedure have slightly different start-height, but the same end-height, and the same height spacing. This means that none of the resulting a-vectors are the same. And the same goes for the Abel transform results. The first step is therefore to interpolate the simulated and calculated bending angles on a common a-vector that has the same fine height spacing as the vector from the phase matching. After this step is done one can easily calculate the square difference for each profile and add them. To find the average square one has to keep track of at what height each profile starts and stops, so as not to divide with too large a number at the height in question. After the square sum has been divided with the appropriate number at each height one simply takes the root of the vector to find the rms. The results can be seen in Figs. 8.20-8.23.

In Fig. 8.20 we see the lower 10 km's. Clearly we are within the bounds everywhere except in those places corresponding to the super-refractive regions.

In Figs. 8.21 and 8.22 we are in the middle segment between 10 and 35 km. Here the filtering method seems to be inadequate, and something more sophisticated will have to be used. It is possible that a filter having a height dependent variance will perform better, as is indicated by the comparison of Figs. 8.21 and 8.22, where the rms relative difference takes on markedly different values at different heights depending on the value for the variance,

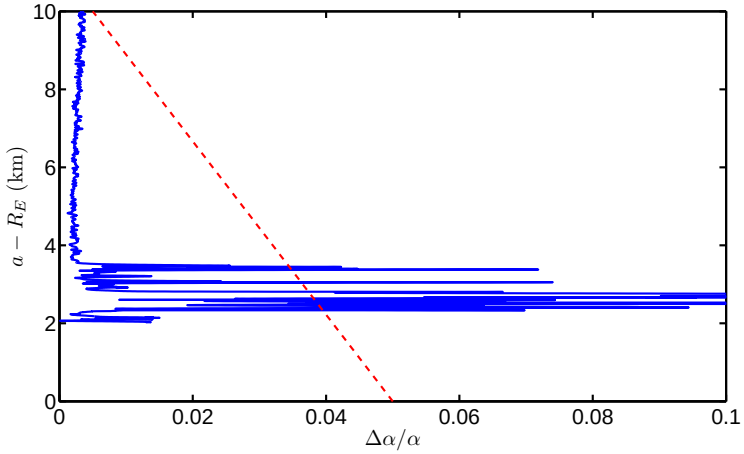

Figure 8.20: The rms of the relative difference between simulations and GO for the 55 cases in the segment between 0 and 10 km. The simulated bending angle was filtered with $\sigma = 25$ m, and $w = 250$ m. The rms relative difference is within the bound (red dashed line) everywhere except in the super-refractive regions.

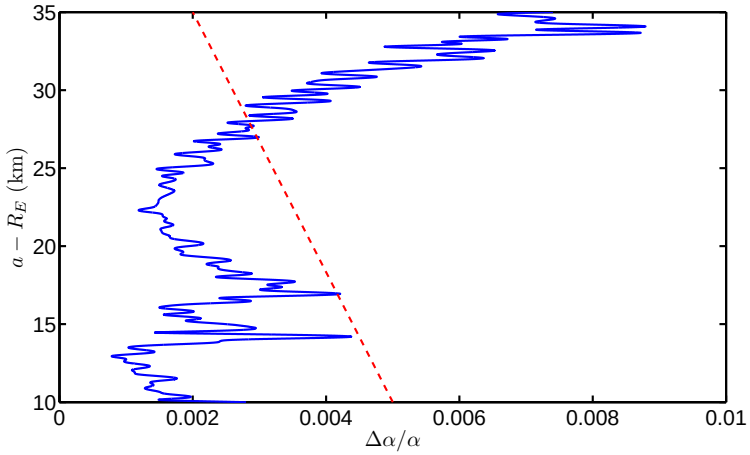

Figure 8.21: Same as Fig. 8.20 but in the segment 10-35 km, and the simulations have been filtered with $\sigma = 200$ m, and $w = 2000$ m. The rms relative difference strays outside the bound for high impact heights.

and $\sigma = 200$ m performs better for low heights, whereas $\sigma = 400$ m performs best at high heights.

In Fig. 8.23 the upper segment between 35 and 80 km can be seen. The rms is well within the bounds everywhere except in the very lowest bit.

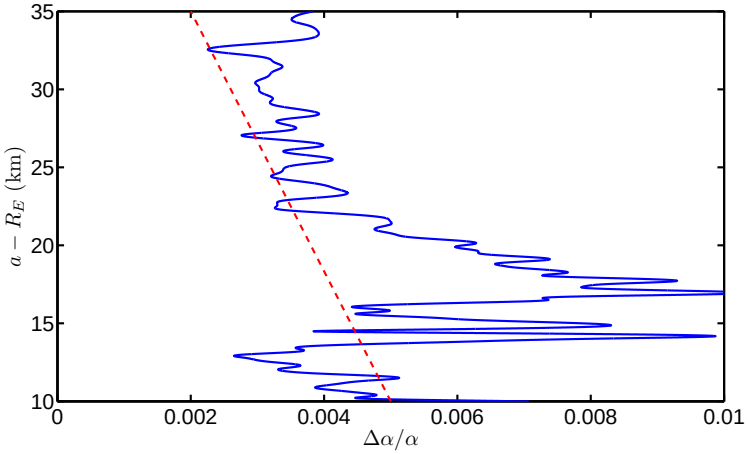

Figure 8.22: Same as Fig. 8.21 but using $\sigma = 400$ m, and $w = 4000$ m. The rms relative difference behaves better for high heights and worse for low as compared to Fig. 8.21.

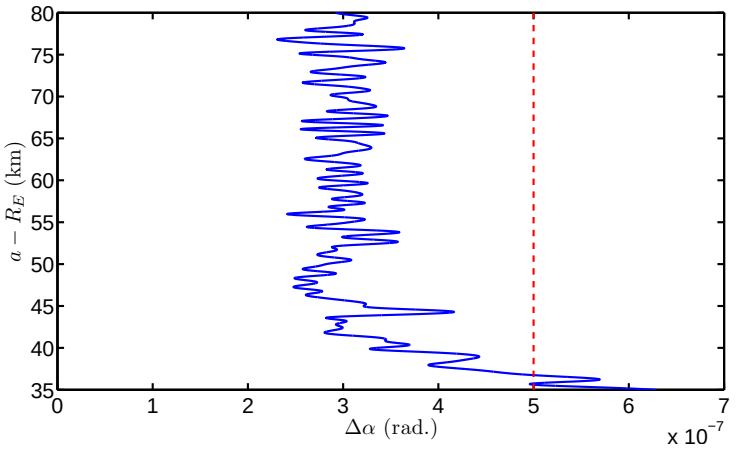

Figure 8.23: The rms of the absolute difference between simulations and GO for the 55 cases in the segment between 35 and 80 km. The simulated bending angle was filtered with $\sigma = 500$ m, and $w = 5000$ m. The rms absolute difference is within the bound (red dashed line) everywhere except at the very lowest impact heights.

8.6 Comparison between simulations with and without noise

One question that arises when looking at the difference between simulations and GO, and comparing the results with and without noise is: how is the

super-refractive peak influenced by the addition of noise, and does the simulations with and without noise give the same results? This is a rather simple matter to investigate. In Fig. 8.24 we can see the bending angle without noise (blue curve), and with noise (red curve). The curves agree very nicely if one disregards the noise. However, if we zoom in around the super-refractive region (Fig. 8.25) we see that the curves are in great contradiction at one specific point. The curve with noise actually reaches negative bending angles, which is unphysical and clearly marks an error. Taking the relative difference between the curves $((\alpha - \alpha_{\text{noise}})/\alpha)$ we get the results shown in Fig. 8.26. It is hard to see anything relevant in this figure, as it is riddled with noise. If we filter both signals with $\sigma = 25$ m, and $w = 250$ m we get the results in Fig. 8.27. The curves are now in nice agreement except in one specific region around 2.7 km. It is the contribution from the erroneous point in the α_{noise}-vector that has been smeared out, and the difference between the simulations with and without noise appear more dramatic than it really is. Finally in Fig. 8.28 we see both curves after filtering. The bending angle with noise (red) reaches a slightly lower value in the super-refractive region than the bending angle without noise.

The super-refractive region in case 15 with noise was only affected at one single position in the height vector. This indicates that we may be dealing with an unlikely coincidental correlation between sharp gradients in the bending angle and an unusually high noise peak, leading to an error in the phase matching routine. To test this hypothesis we need to look at more super-refractive profiles. In Fig. 8.29 the bending angles for case 5 is shown. Case 5 is also a class 4 profile and looks almost identical to case 15 in this height segment. However looking at Fig. 8.30 we see that there is no indication of there being a super-refractive region. This provides strong evidence for our hypothesis.

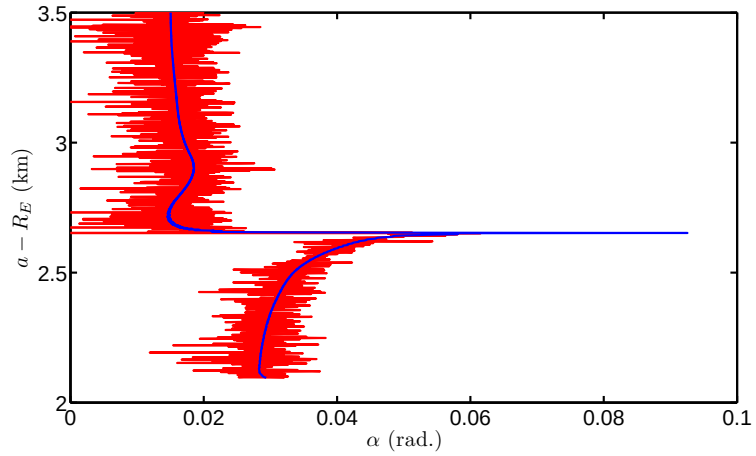

Figure 8.24: Simulated bending angle for case 15 with noise (red curve) and without (blue curve).

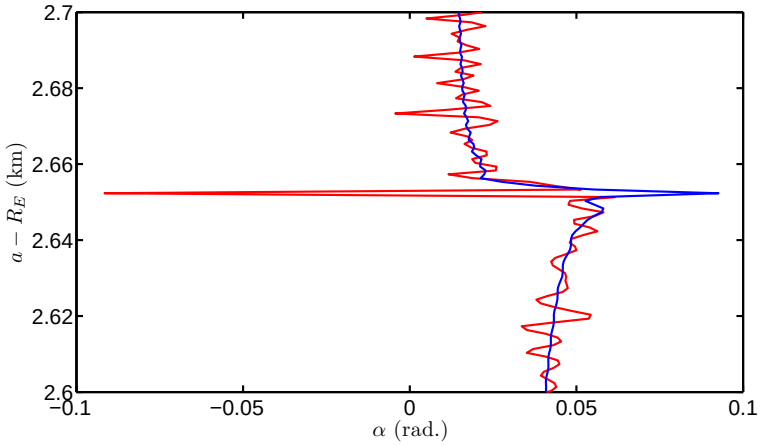

Figure 8.25: Simulated bending angle for case 15 with noise (red curve) and without (blue curve), zoomed in around the super-refractive region. The red curve is obviously erroneous at one point where the bending angle reaches negative values.

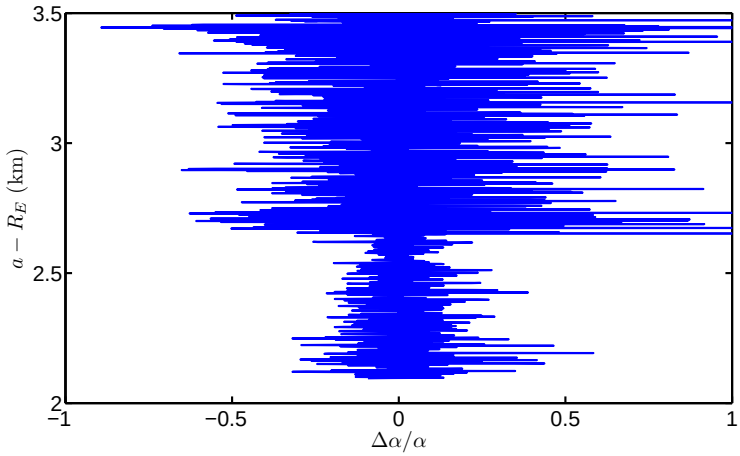

Figure 8.26: The difference between the simulated bending angle for case 15 with and without noise. It is hard to discern anything interesting due to the noise, and the curves obviously need to be filtered before we can inspect them properly.

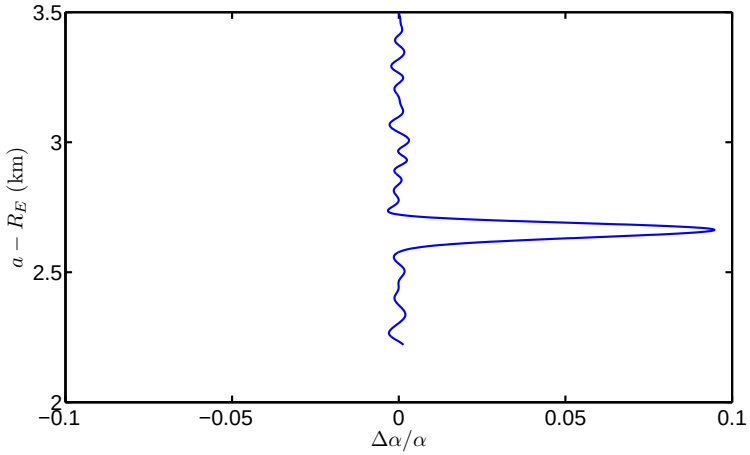

Figure 8.27: Same as Fig. 8.26 but both signals have been filtered using $\sigma = 25$ m, and $w = 250$ m. The relative difference is rather low everywhere except around the point shown in Fig. 8.25.

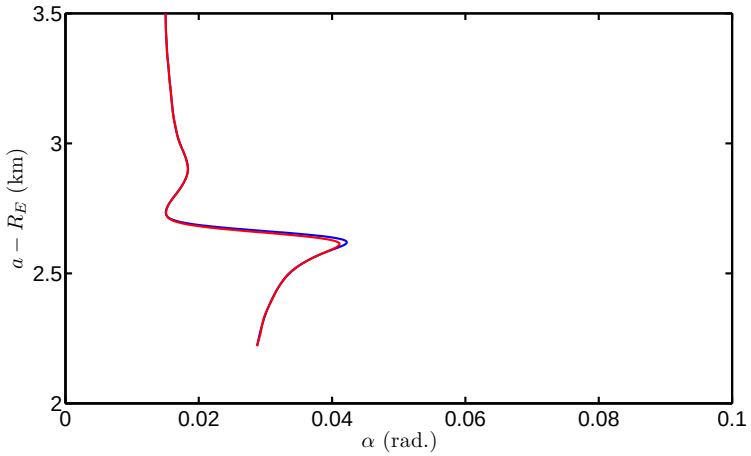

Figure 8.28: Same as Fig. 8.24 but with both signals filtered using $\sigma = 25$ m, and $w = 250$ m. The curves appear to be almost identical, but a small displacement can be seen near the point of the error in the noisy curve.

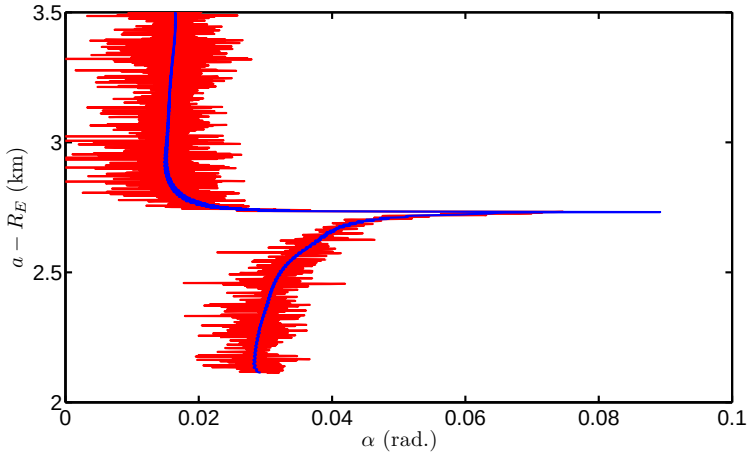

Figure 8.29: Simulated bending angle for case 5 with noise (red curve) and without (blue curve). The overall appearence is very similar to case 15, which motivates a comparison between the two cases.

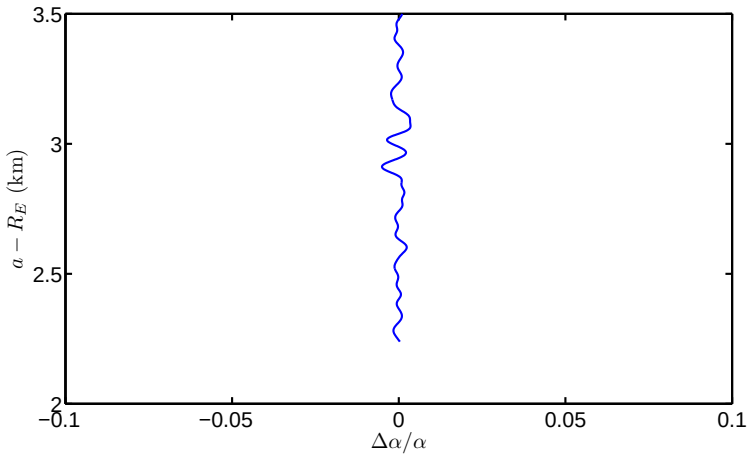

Figure 8.30: The relative difference between the simulated bending angle for case 5 with and without noise. It is almost impossible to tell that there is a super-refractive region. This should be contrasted with Fig. 8.26.

Chapter 9

The Abel transform in the presence of super-refraction

9.1 Introduction

When using the inverse Abel transform to retrieve the refractive index as a function of height, it has been noted that the method introduces a bias when the refractive index exceeds a critical gradient, $dN/dr < -157$ km^{-1}. The basic physical mechanism behind this is the lack of occultation data from certain height segments in the atmosphere, so-called super-refractive layers. The rays simply pass through critical regions and provide little information about them. Below we will go through the theory and demonstrate what magnitude of bias we can expect.

9.2 Ray paths

In an atmosphere where the refractive index only depends on the distance from the Earth's center, and under conditions where geometric optics applies, we can use Bouger's rule to describe the path of a ray [11]

$$n(r)r\sin\phi = \text{const.} \tag{9.1}$$

where n is the refractive index, r the distance from the Earth's center, and ϕ the angle between the ray path tangent and a vector projected from the Earth's center to the ray path, see Fig. 9.1.

A ray launched from space towards Earth starts with $r = r_0$, and $\phi = \phi_0$. In space we have $n = 1$, so

$$\text{const.} = r_0\sin\phi_0 \equiv a \tag{9.2}$$

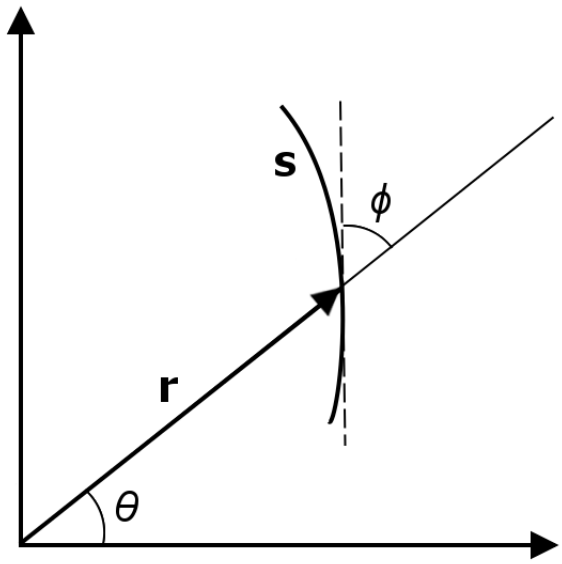

Figure 9.1: The ray path, s, and the vector from Earth's center, r, tracking the ray. The angle ϕ is measured with respect to the ray tangent, and the angle θ with respect to some arbitrary abscissa.

Where a is normally called the impact parameter [12], or the ray asymptote miss distance [28]. The rays are then determined by the conditions at their launch as

$$n(r)r\sin\phi = a \tag{9.3}$$

If the ray enters the atmosphere with an angle that does not lead to impact with the Earth's surface, the ray path will be slightly bent, and have a point of closest approach. hereafter referred to as the tangent height, to the Earth where $r = r_t$, and $\phi = \pi/2$. In that case

$$n(r_t)r_t = a \tag{9.4}$$

In this way, rays launched with different a's will have different tangent heights. This allows us to analyze the structure of the atmosphere by observing the bending of different rays.

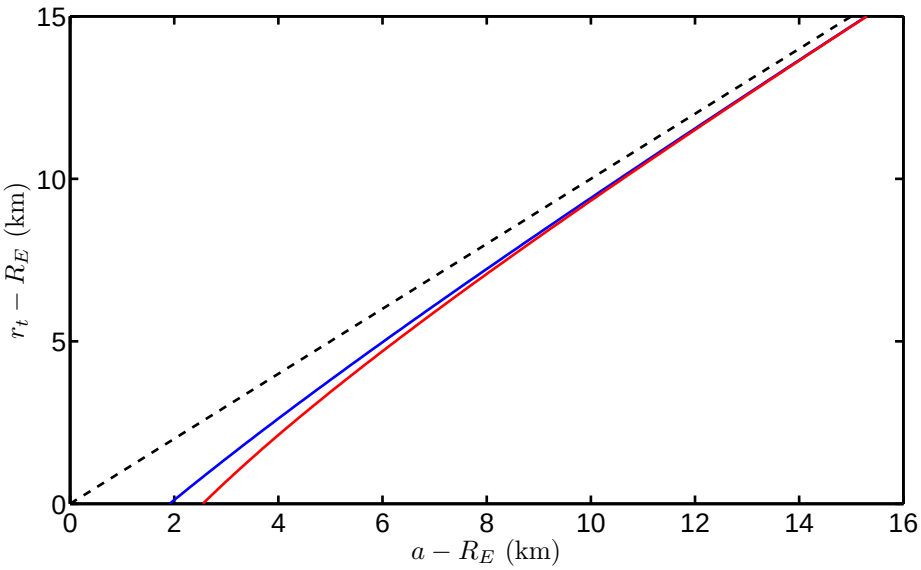

Figure 9.2: The solution to Eq. (9.6) for an atmosphere having $H = 8$ km, $N_E = 300$ (red line), an atmosphere having $H = 6$ km, $N_E = 400$ (black solid line), and one having $n = 1$ (black dashed line). The units on the axis are km's.

In an "ideal" atmosphere we would expect a refractive index that varies expone

$$n(r) = 1 + 10^{-6} N_E \exp\left(-\frac{r - R_E}{H}\right) \tag{9.5}$$

where R_E is the Earth radius (usually taken to be 6371 km), N_E is the refractive index at the Earth's surface, and H is the scale height for a decrease in the refractive index, normally around 6-8 km's [12]. Given such a variation we find

$$r_t(1 + 10^{-6} N_E \exp\left(-\frac{r_t - R_E}{H})\right) = a \tag{9.6}$$

An equation which can be solved using basic numerical techniques. The result are shown in Fig. 9.2 for some different atmospheres. There are three interesting things to note in Fig. 9.2. First of all, for every value of a there is a corresponding value for r_t. Secondly, the lower the value of a, the lower the ray will penetrate the atmosphere. Third, the lowest impact parameter is around 2 km (corresponding to a tangent height equal to the Earth's radius).

Now, Eq. (9.6) represents an ideal atmosphere. In reality there will

be rapid variations in the refractive index as a function of height. This introduces some problems. The solution to Eq. (9.6) gives a monotonic increase in a with increasing r, i.e.

$$\frac{da}{dr_t} > 0 \tag{9.7}$$

When this is not the case, there will be layers of local decrease in a with r. This will result in no rays having tangent points within such layers. To see why this is we need to think about different rays and their tangent points. Rays with large a's have large r_t's, and rays with lower a's have lower r_t's. So a ray will travel to a sufficient depth until it reaches the first point where $rn(r)$ is sufficiently low. It cares not about segments of local increase in $rn(r)$. But on the other hand, in the region of a local increase in $rn(r)$ with decreasing r, all rays having tangent points in that segment will not reach such low altitudes in the first place. Hence, there will be gaps in the information retrieval using this method. These regions of local increase in $rn(r)$ with decreasing r occur when

$$\frac{da}{dr_t} < 0 \tag{9.8}$$

and using Eq. (9.4) we find

$$\frac{dn}{dr_t} < -\frac{n}{r_t} \tag{9.9}$$

Using $r_t \approx R_E \approx 6371$ km, and $n \approx 1$ we find a rough approximation

$$\frac{dn}{dr_t} < -1.57 \times 10^{-4} \ \text{km}^{-1} \tag{9.10}$$

Or using $N = (n-1) * 10^6$

$$\frac{dN}{dr_t} < -157 \ \text{km}^{-1} \tag{9.11}$$

Incidentally, these regions correspond to what is called super-refractive regions. In such regions the rays have a curvature radius which is smaller than that of the Earth. To see this is we can rewrite Eq. (9.3) slightly

$$\sin\phi = \frac{a}{rn(r)} \tag{9.12}$$

Clearly, when $rn(r) > a$ we have $\sin\phi < 1$, and $\phi \in (0, \pi/2)$ for a ray traveling towards Earth. As the ray travels deeper into an ideal atmosphere $rn(r)$ comes closer and closer to a, until eventually $r = r_t$, $rn(r) = a$, and

$\phi = \pi/2$. This is the point where the ray starts to move away from the Earth. In a non-ideal atmosphere there will be areas of local increase in $rn(r)$ with decreasing radius. When a ray enters such a region, $rn(r)$ will momentarily grow, and ϕ will momentarily decrease. This will result in longer ray trajectories, as these layers will have to be traversed before the ray reaches its tangent point. When the curvature of the ray becomes smaller than the curvature of the Earth, the trajectories can become long indeed, and for one particular value of a it will even be infinite.

The curvature, κ, of the ray is given by

$$\kappa = |\frac{d\hat{s}}{ds}| \tag{9.13}$$

where $\hat{s}$ is the unit tangent vector of the ray, and ds a length segment along the ray path. The radius of curvature, ρ, is related to the curvature through

$$\rho = \frac{1}{\kappa} \tag{9.14}$$

From basic geometrical optics [11] it is known that the differential equation describing the ray paths is

$$\frac{d}{ds}(n\hat{s}) = \nabla n \tag{9.15}$$

Which means that

$$\frac{d\hat{s}}{ds} = \frac{1}{n}\left[\nabla n - \frac{dn}{ds}\hat{s}\right] \tag{9.16}$$

Now given the geometry in Fig. 9.1, we have

$$\hat{s} = -\cos\phi\hat{r} - \sin\phi\hat{\theta} \tag{9.17}$$

And in the case when n is only a function of radius we have

$$\frac{dn}{ds} = \frac{dn}{dr}\frac{\partial r}{\partial s} = -\cos\phi\frac{dn}{dr} \tag{9.18}$$

Using $\nabla n = \hat{r}dn/dr$ we find

$$\frac{d\hat{s}}{ds} = \frac{1}{n}\frac{dn}{dr}\left[\hat{r}\sin^2\phi - \hat{\theta}\cos\phi\sin\phi\right] \tag{9.19}$$

Consequently

$$\kappa = |\frac{d\hat{s}}{ds}| = \frac{1}{n}\frac{dn}{dr}\sin\phi \tag{9.20}$$

The most interesting rays are those that are just about to turn when they enter a region of increasing $rn(r)$. In that case $\sin\phi \approx 1$ and

$$\kappa \approx \frac{1}{n}\frac{dn}{dr} \tag{9.21}$$

Such a ray will travel with the same curvature as the Earth when $|1/\kappa| = R_E$, or keeping in mind the geometry of the problem

$$\frac{1}{n}\frac{dn}{dr} \approx -\frac{1}{R_E} \quad \Leftrightarrow \quad \frac{dN}{dr} \approx -157 \ \ \mathrm{km}^{-1} \tag{9.22}$$

And when

$$\frac{dN}{dr} < -157 \ \ \mathrm{km}^{-1} \tag{9.23}$$

the rays will experience a curvature which is stronger than that of the Earth's surface.

9.3 The Abel transform

Returning to Eq. (9.19) we see that the incremental change in direction in the ray after an incremental step is

$$|d\hat{s}| = ds\frac{1}{n}\frac{dn}{dr}\sin\phi \tag{9.24}$$

Using $ds = -dr/\cos\phi$, and Eq. (9.3) we find

$$d\alpha = -\frac{1}{n}\frac{dn}{dr}\frac{a}{\sqrt{r^2n^2 - a^2}}dr \tag{9.25}$$

where $d\alpha$ is the change in angle of the unit vector in the direction of the ray, i.e. $d\alpha \equiv |d\hat{s}|/|\hat{s}|$.

Integrating $d\alpha$ between 0 and α_t, we find the angular change of the ray when it has reached the tangent point

$$\int_0^{\alpha_t} d\alpha = \alpha_t = -\int_{r_0}^{r_t} \frac{1}{n}\frac{dn}{dr}\frac{a}{\sqrt{r^2n^2 - a^2}}dr \tag{9.26}$$

The angular change on the way out of the atmosphere is again α_t, and it is customary to use $-2\alpha_t = \alpha$, in which case

$$\alpha = -2a\int_{r_t}^{r_0} \frac{1}{n}\frac{dn}{dr}\frac{1}{\sqrt{r^2n^2 - a^2}}dr \tag{9.27}$$

Since $n = 1$, and $dn/dr = 0$ when $r = r_0 \to \infty$ we can use

$$\alpha = -2a \int_{r_t}^{\infty} \frac{1}{n} \frac{dn}{dr} \frac{1}{\sqrt{r^2 n^2 - a^2}} dr \tag{9.28}$$

When performing the integral above there will be some numerical problems, and the integration close to the singularity $r_t n(r_t) = a$ will have to be performed with some care. One method of doing this is described in section 9.6.1. But the peculiarities that has to do with the super-refractive regions do not arise solely because of the singularity. We can demonstrate this quite easily by plotting the integrand for a just above and just below the super-refractive layer. First of all, in Fig. 9.3 one can see an example of a refractivity profile that contains a super-refractive layer. When plotting the refractive index, the super-refractive region appears only as a segment where the slope is rather flat (red segment in the left panel of Fig. 9.3). On the other hand, if one plots $rn(r)$ as a function of r (right panel), it becomes clear that something special will happen when a is in the region of the little bump, indicated by the green line. One ray might have an a which is sufficiently large for the integration to stop just before the bump, located at $a = a_B$. Whereas for a ray with a slightly lower a the integration will have to be taken over the entire bump. This will of course result in a sudden jump in the bending angle as a function of a. The jump will occur for all rays having an a lower than a_B, but the magnitude of the jump will certainly depend on the proximity of a to a_B, because of the singularity in the integrand of Eq. (9.28). This is clearly demonstrated in Fig. 9.4, where one can see the effect of the jump on the integrand. The left panel shows the integrand for the case when a is slightly higher than a_B. The integration stops at the singularity, marked by the vertical dashed line. There is nothing surprising about the curve in the left panel, except perhaps a certain jaggedness. This is due to the interpolation technique. Since the data is provided only at discrete locations, marked by asterisks, to achieve a vivid picture the data in between will have to be interpolated. In this case we have used piecewise linear interpolation. The curve on the right panel on the other hand show some interesting features. Due to the jump, and its appearance close to the location of the singularity there is a big bump in the curve of the integrand. The dimensions of the bump is dictated by the product of the derivative of the refractive index, and the term containing the singularity. The last term grows as $1/\sqrt{\Delta}$, where Δ is the distance to the singularity, whereas the first term grows and decays over the super-refractive region. These things combine to produce the strange curved features of the integrand, see Fig. 9.4. Overlaid on this comes the effect of the interpolation, which makes the curve jagged. These features also give rise to the distinct appearance of the bending angle as a function of

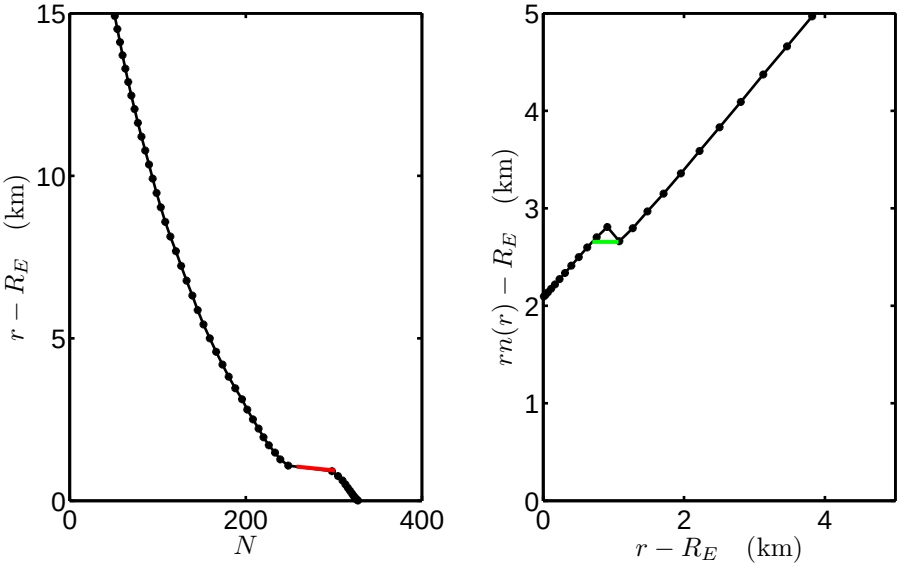

Figure 9.3: Left panel shows the refractivity vs. height for a class 4 profile (case 15). The red segment indicates the super-refractive region. Right panel shows the product of the refractive index and height vs. height for the same data as in the left panel. The green segment indicates the jump that creates the peculiarities in the bending angle profile for a super-refractive profile.

the impact parameter in the case of super-refractivity, as compared to a less extreme refractive index profile, see Figs. 9.5 and 9.6. Looking at Fig. 9.7 it becomes clear that the peculiarities in Fig. 9.4 causes the abrupt changes in the bending angle curve for refractivity curves containing super-refractive regions.

9.4 The inverse Abel transform

By defining $x = rn(r)$ we can use

$$\frac{dn}{dr}dr = \frac{dn}{dx}\frac{dx}{dr}dr = \frac{dn}{dx}dx \tag{9.29}$$

and Eq. (9.28) becomes

$$\alpha(a) = -2a\int_a^\infty \frac{1}{n}\frac{dn}{dx}\frac{1}{\sqrt{x^2 - a^2}}dx \tag{9.30}$$

provided there is a one-to-one relationship between x and r. In the case of super-refraction there is not.

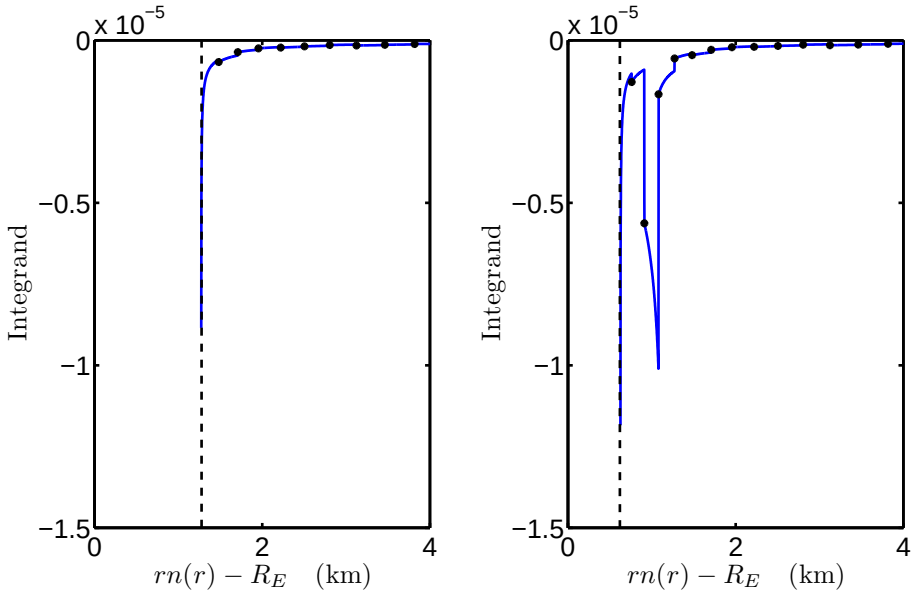

Figure 9.4: Left panel shows the integrand in Eq. (9.28) as a function of r for a ray where a is slightly higher than a_B. The right panel shows the same thing but for an a which is slightly lower than a_B. The dashed lines indicate the singularity that appears when $rn(r) \to a$.

If the relationship is one-to-one we can invert Eq. (9.30). To do this we multiply by $da/\sqrt{a^2 - a_1^2}$ and integrate between a_1 and ∞

$$\int_{a_1}^{\infty} \frac{\alpha(a)}{\sqrt{a^2 - a_1^2}} da = -\int_{a_1}^{\infty} \frac{2a}{\sqrt{a^2 - a_1^2}} \left[\int_a^{\infty} \frac{1}{n(x)} \frac{dn(x)}{dx} \frac{1}{\sqrt{x^2 - a^2}} dx \right] da \tag{9.31}$$

On the right hand side, the order of integration can be rewritten. The integration domain is bounded by $a = a_1$ and $x = a$ (see Fig. 9.8), which means that the integration order can be switched according to

$$\int_{a=a_1}^{a=\infty} \ldots \left[\int_{x=a}^{x=\infty} \ldots dx \right] da = \int_{x=a_1}^{x=\infty} \ldots \left[\int_{a=a_1}^{a=x} \ldots da \right] dx \tag{9.32}$$

We end up with

$$\int_{a_1}^{\infty} \frac{\alpha(a) da}{\sqrt{a^2 - a_1^2}} = -\int_{x=a_1}^{x=\infty} \frac{1}{n(x)} \frac{dn(x)}{dx} \left[\int_{a=a_1}^{a=x} \frac{2a}{\sqrt{a^2 - a_1^2}} \frac{da}{\sqrt{x^2 - a^2}} \right] dx \tag{9.33}$$

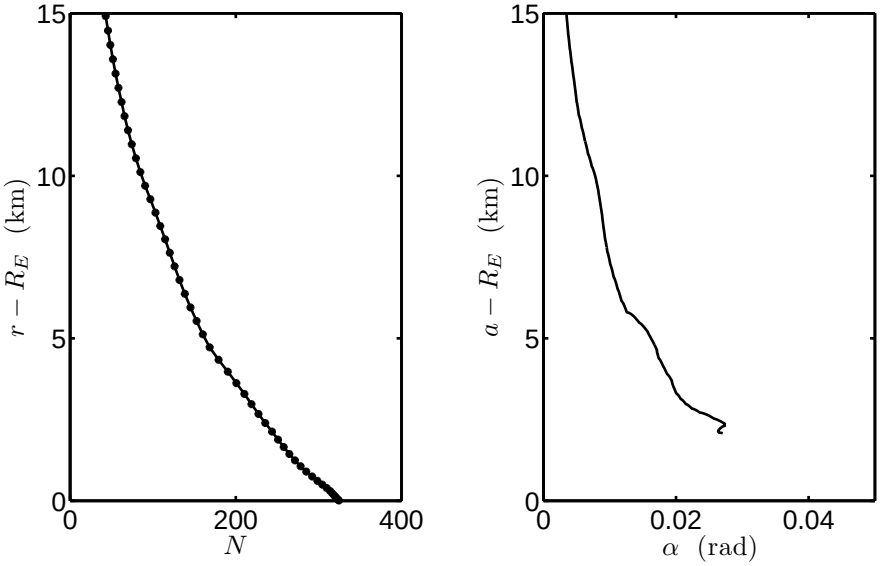

Figure 9.5: Left panel shows the refractivity vs. height for a class 1 profile (case 1). Right panel shows the bending angle (α) vs. the impact parameter a for the same case.

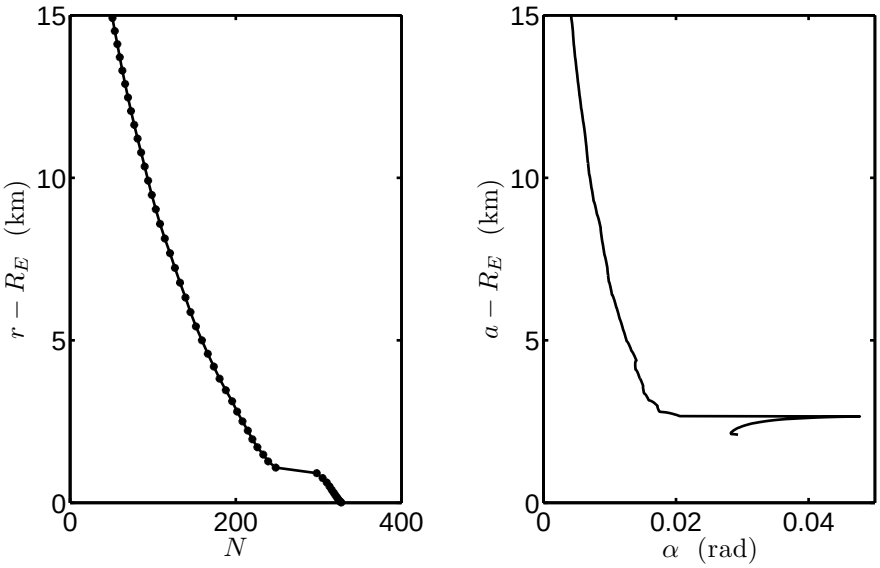

Figure 9.6: Left panel shows the refractivity vs. height for a class 4 profile (case 15). Right panel shows the bending angle (α) vs. the impact parameter a for the same case.

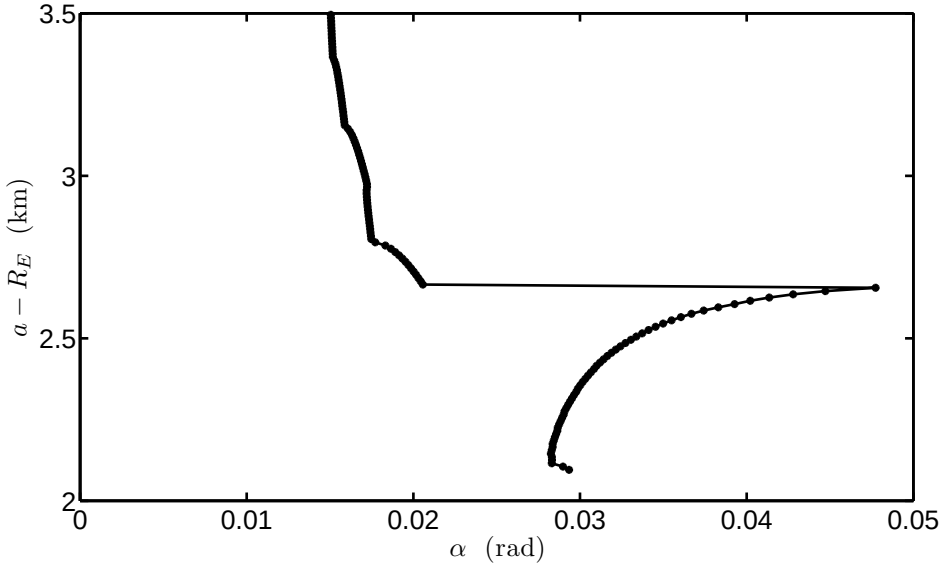

Figure 9.7: A closer look on the region around the super-refractive segment in Fig. 9.6.

The expression within the brackets can be integrated immediately, which gives us

$$\int_{a_1}^{\infty} \frac{\alpha(a)}{\sqrt{a^2 - a_1^2}} da = -\int_{x=a_1}^{x=\infty} \frac{1}{n(x)} \frac{dn(x)}{dx} \left[\arcsin\left(\frac{2a^2 - x^2 - a_1^2}{x^2 - a_1^2}\right) \right]_{a=a_1}^{a=x} dx = \tag{9.34}$$

$$= -\pi \int_{x=a_1}^{x=\infty} \frac{1}{n(x)} \frac{dn(x)}{dx} = \pi \ln n(x = a_1) \tag{9.35}$$

Hence

$$n(a_1) = \exp\left(\frac{1}{\pi} \int_{a_1}^{\infty} \frac{\alpha(a)}{\sqrt{a^2 - a_1^2}} da \right) \tag{9.36}$$

So by performing the integral above we end up with n as a function of a_1. We can then use $r_1 n(r_1) = a_1$ to find the r-vector corresponding to n

$$r_1 = \frac{a_1}{n(a_1)} \tag{9.37}$$

The bias problem arises because there is no data produced by the super-refractive layer. To perform the integral properly in the case of super-refraction, the integral has to be split into four regions according to

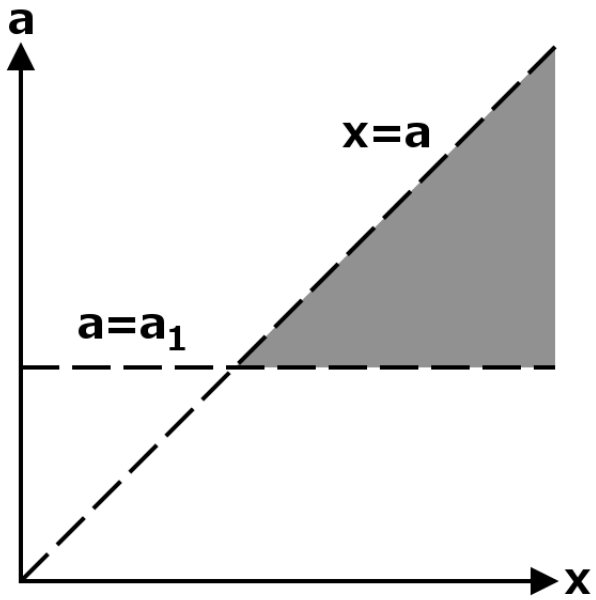

Figure 9.8: The integration domain (grey area) in Eq. (9.32).

$$\int_{a_1}^{\infty} \frac{\alpha(a)}{\sqrt{a^2 - a_1^2}} da \rightarrow \int_{a_1}^{a_a} \ldots da + \int_{a_a}^{a_b} \ldots da + \int_{a_b}^{a_c} \ldots da + \int_{a_c}^{\infty} \ldots da \quad (9.38)$$

where the different a's are illustrated in Fig. 9.9. In the sections delimited by x_a, x_b and x_c, the function x is one-to-one with respect to r. In reality, there is no data for α in the sections between a_a and a_c, unless external data is introduced, either by using other observation methods, numerical simulation, or educated guessing. Failure to perform these integrals correctly will lead to an underestimation of n in the region below r_c. According to Sokolovskiy [29], the magnitude of this error can be as big as 10 %. In Fig. 9.10 one can see the refractivity obtained through Abel transform and inversion compared to the correct refractivity in the case of a super-refractive region (case 15). In Fig. 9.11 one can see the fractional error in the inverted refractivity for all the class 4 cases. It appears that the maximum error is around 14 %, although most profiles never reach that high values.

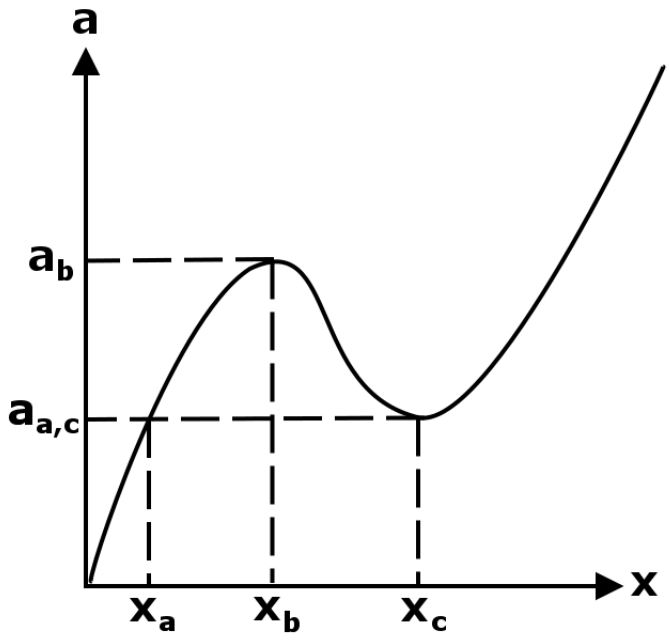

Figure 9.9: The different segments that would be necessary to properly invert the bending angle using the inverse Abel transform and retrieve the refractive index in the case of super-refraction.

9.5 Mitigation

The bias problem was originally reported by Rocken et al. [30], but attributed incorrectly to sub-refraction ($dN/dz > 0$). Kursinski [12, 31] claimed in his thesis that the bias was due to super-refraction, mainly caused by moisture gradients associated with inversion caps in the marine boundary layer. Since then, there have been many publications on the subject [32–37], but no universal method for mitigating the bias problem seems to have been found. The comparison of the different techniques, and the evaluation of the best one is a big task which cannot be performed within the time-frames of the present NRFP project.

9.6 Numerical integration

The presence of the singularities in Eqs. (9.28) and (9.36) makes the numerical integration a bit problematic. However, by making certain approxima-

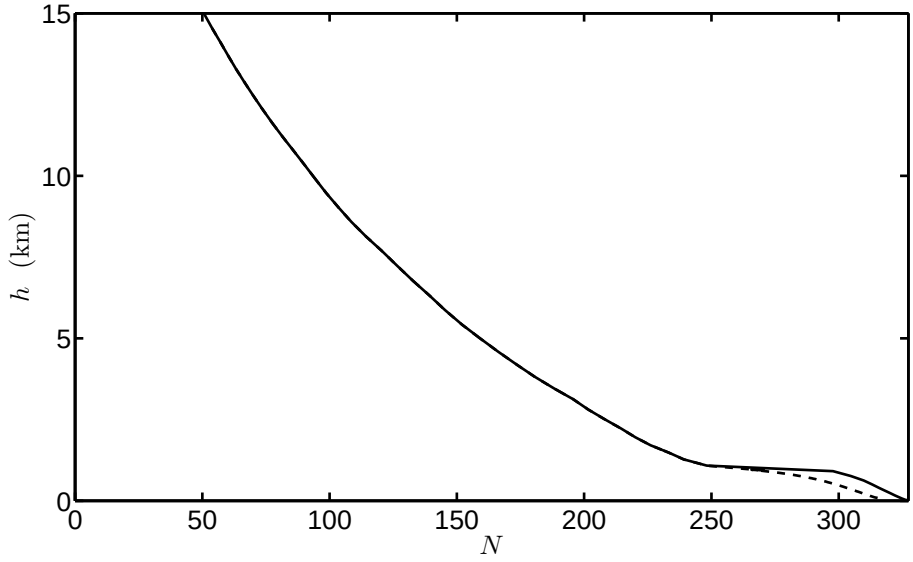

Figure 9.10: The true refractivity (solid line) and the refractivity obtained by Abel inversion (dashed line) for a class 4 profile (case 15). The inverted curve starts to become negatively biased below the critical height.

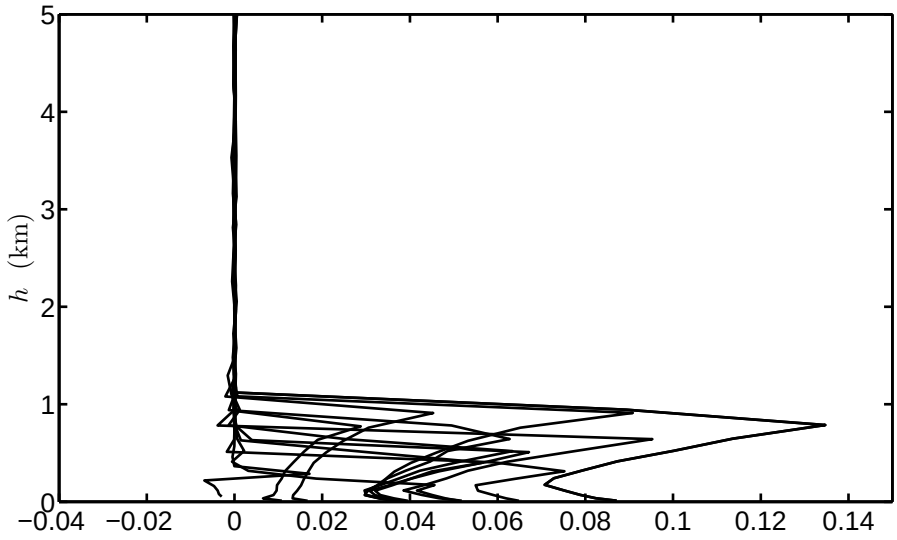

Figure 9.11: The fractional error in the inverted refractivity data, i.e. $(N - N_{inv})/N$ for all class 4 cases. The error never exceeds 14 %.

tions, good results can be achieved with little effort.

9.6.1 Forward integration

We will first look at Eq. (9.28)

$$\alpha = -2a \int_{r_t}^{\infty} \frac{1}{n} \frac{dn}{dr} \frac{1}{\sqrt{r^2 n^2 - a^2}} dr \tag{9.39}$$

Simply using the function trapz in Matlab will give erroneous results because of the sharp gradients close to the singularity $rn(r) = a$. The major problems stem from the part of the integrand containing the square root. We start by changing the integration variable to $x = rn(r)$, and use Eq. (9.30)

$$\alpha = -2a \int_{x=a}^{\infty} \frac{d \ln n}{dx} \frac{1}{\sqrt{x^2 - a^2}} dx \tag{9.40}$$

In this case it is irrelevant that x and r are not one-to-one. There will be no bending angle corresponding to the super-refractive layer, and the bending angle curve we produce with the integral above will be accurate. Now, close to the singularity, over a small interval in x, the fraction in the integrand will grow violently. In fact, the term grows so rapidly due mainly to the change in x that we can approximate everything else as constant. This means using

$$\frac{d \ln n}{dx} \approx \frac{\ln n[i+1] - \ln n[i]}{x[i+1] - x[i]} \tag{9.41}$$

where $[i]$ indicates the i'th position in a vector representing actual discrete data. We get

$$I_i = \frac{\ln n[i+1] - \ln n[i]}{x[i+1] - x[i]} \left[\operatorname{arccosh}\left(\frac{x[i+1]}{a}\right) - \operatorname{arccosh}\left(\frac{x[i]}{a}\right) \right] \tag{9.42}$$

with the bending angle given by

$$\alpha(a) \approx -2a \left[\sum_{i=j}^{M} I_i + A \right] \tag{9.43}$$

where

$$A = \frac{\ln n[j] - \ln n[j-1]}{x[j] - x[j-1]} \left[\operatorname{arccosh}\left(\frac{x[j]}{a}\right) - \operatorname{arccosh}\left(\frac{x[j-1]}{a}\right) \right] \tag{9.44}$$

and $i = j$ is the point on the vector $n[i]r[i]$ which is closest above a, and M is the length of the vector x.

9.6.2 Inverse integration

Now let's look at Eq. (9.36).

$$n(a_1) = \exp\left(\frac{1}{\pi}\int_{a_1}^{\infty}\frac{\alpha(a)}{\sqrt{a^2-a_1^2}}da\right) \tag{9.45}$$

It suffers from the same problems associated with the singularity as Eq. (9.28). One easy way of handling this is to use a linear interpolation for α, and integrate the function accordingly. The problematic integral is

$$I = \int_{a_1}^{\infty}\frac{\alpha(a)}{\sqrt{a^2-a_1^2}}da \tag{9.46}$$

We use

$$\alpha(a_i \le a \le a_{i+1}) \approx \alpha(a_i) + (a-a_i)\frac{\alpha(a_{i+1})-\alpha(a_i)}{a_{i+1}-a_i} = A_i + aB_i \tag{9.47}$$

where

$$A_i = \frac{\alpha(a_i)a_{i+1} - a_i\alpha(a_{i+1})}{a_{i+1}-a_i} \tag{9.48}$$

$$B_i = \frac{\alpha(a_{i+1})-\alpha(a_i)}{a_{i+1}-a_i} \tag{9.49}$$

Then

$$I_i = \int_{a_i}^{a_{i+1}}\frac{A_i + aB_i}{\sqrt{a^2-a_1^2}}da \tag{9.50}$$

which integrates to

$$I_i = A_i\left[\operatorname{arccosh}\left(\frac{a_{i+1}}{a_1}\right) - \operatorname{arccosh}\left(\frac{a_i}{a_1}\right)\right] + B_i\left[\sqrt{a_{i+1}^2-a_1^2} - \sqrt{a_i^2-a_1^2}\right] \tag{9.51}$$

And finally

$$I = \sum_{i=j}^{M-1} I_i + C \tag{9.52}$$

where

$$C = A_j \text{arccosh}\left(\frac{a_j}{a_1}\right) + B_j\sqrt{a_j^2 - a_1^2} \tag{9.53}$$

where $i = j$ is the point on the vector $a[i]$ which is closest above a_1.

Chapter 10

Conclusions

In this report we have gone through the theory behind and the implementation in Matlab of an end-to-end RO simulator including: the propagation of the field through the atmosphere and into orbit using multiple phase screen (MPS) propagation and a diffractive integral; addition of noise to the signal in orbit; and retrieval of the bending angle and impact height through phase matching (PM). We have compared the output of the simulator package with those of geometrical optics (GO), through the use of the Abel transform. We have seen that the entire simulation package delivers reliable high precision data when compared to GO.

In the case of no added noise we were able to determine that there is numerical noise. The exact source of this noise is not yet determined. The simulated data show low level numerical noise appearing in the MPS, and during the propagation to orbit. If this is not the cause, perhaps it is the phase matching which causes it, or as seen in section 4, there are fundamental differences between GO and wave optics that appear in the form of diffraction patterns. In any case, this noise can be removed easily using a Gaussian filter applied on the bending angle data. The remaining discrepancy between simulations and GO are located to the super-refractive regions and are not negligible. The reason for this is most likely that GO is not perfect. When the gradients in the refractive index become very strong we should expect some variations between the simulated results and GO. The bending angle singularity associated with the super-refractive region appears to be located slightly lower for the simulations than for the calculated profiles. Why this is so is unknown at this moment.

When adding white noise to the signal we get bending angle profiles that oscillates randomly around a smoothly varying average, with noise increasing for lower impact heights. In the lower and upper segments the noise can be suppressed under the specified limits using a Gaussian filter, but not in the

middle region (10-35 km). For the filter to work properly it should have some connection to the size of the Fresnel zone. In this middle segment the Fresnel zone undergoes rapid growth with height, and it seems reasonable, and there are results that indicate, that we could achieve much better noise reduction by having a filter that changes parameters as we ascend. This will have to be a task for the future.

In section 8.6 we look at the difference between simulations with and without noise in the super-refractive region. The results indicate that variations between the curves above the noise floor may appear in odd cases. An analysis showed that case 15 had one point in the bending angle vector where the difference became significant. Since the height vector has a resolution of 1 m this indicates that we are dealing with a statistical fluke, where large gradients in α are overlain with unusually strong noise, resulting in an error in the phase matching routine. The same analysis performed on case 5, which is also a class 4 profile, reveled no differences larger than the noise floor. We can probably conclude two things from this fact. First of all, adding noise to the simulations does not perturb the phase matching routine except in extreme cases, when it will lead to large discrepancies between the simulations with and without noise. This hypothesis will have to be verified by studying case 15 with new noise seeds, and by studying all case 4 profiles to see how often the error occurs. Taking such an analysis further could show us at what noise level we will be unable to resolve super-refractive peaks using these techniques, a problem which is interesting also from an engineering point of view, as it is related to the satellite antenna gain. Second, the difference between GO and simulations can be attributed to one or more of three causes. (1) Maybe there is some small error in the way we calculate the GO bending angle that has so far gone unnoticed. (2) Perhaps there is some small calculation error in the MPS routine and the propagation to orbit. (3) But most likely it is due to fundamental differences between GO and the full wave-optical treatment caused by the finite wavelength coupled with sharp gradients in the refractive index.

In order to perform the Abel transform accurately and quickly we also presented a numerical technique to perform the integration, and we investigated the Abel transform in the case of super-refractivity. The study revealed the fact that the technique is flawed. Certain segments of the atmosphere do not produce any received signals, and the inverse Abel transform results in a bias, where the lower parts of the refractivity curve are not retrieved correctly. Several techniques have been reported that are supposed to mitigate the inversion problem, but it is unclear which (if any) technique yields the best results. This is also a topic that could be interesting to study in the future.

Appendix A

Fresnel zone width

A.1 Introduction

When removing noise from the bending angle (see section 8) it is necessary to make sure that the chosen filter width corresponds to the specifications given for the data handling procedure. The width of the filter is determined in relation to the Fresnel zone width for the occultation. Hence it is necessary to find the Fresnel zone width as a function of impact height. In the thesis by Kursinski [31] there is a derivation for the Fresnel zone width in RO based on geometrical optics. Below we will present a slightly different derivation that leads to the same result. We will also plot the Fresnel zone width as a function of height for a generic profile, and find an analytical approximation for the curve.

A.2 Derivation of Fresnel zone width based on Fermat's principle

The Fresnel zone is the region within which adjacent rays differ less than half a wavelength in optical path between source and endpoint. To find the zone radius we should find the perturbation in the impact height a which causes half a wavelength of difference in the optical path length. If $\tau(a)$ is the optical path length we want to find

$$\tau(a + \Delta a_{\lambda/2}) - \tau(a) = \lambda/2 \tag{A.1}$$

And in order to find this value for $\Delta a_{\lambda/2}$ we shall employ Fermat's principle.

Fermat's principle states that "rays of light traverse the path of stationary optical length with respect to variations of the path". In our case we have a

very nice formula for the optical length of the rays

$$\tau(a) = \sqrt{r_0^2 - a^2} + \sqrt{r_r^2 - a^2} + \alpha a - 2\int_a^\infty \sqrt{x^2 - a^2}\frac{d\ln n}{dx}dx \tag{A.2}$$

If we increase a slightly to $a + \Delta a$ we describe a new ray that will strike at a different point in the LEO path, with a resulting optical path length

$$\tau(a+\Delta a) = \sqrt{r_0^2 - (a+\Delta a)^2} + \sqrt{r_r^2 - (a+\Delta a)^2} + \alpha(a+\Delta a)(a+\Delta a) - \tag{A.3}$$
$$- 2\int_{a+\Delta a}^\infty \sqrt{x^2 - (a+\Delta a)^2}\frac{d\ln n}{dx}dx$$

The first three terms in this formula describes a pure geometrical approximation for the physical path of the ray, whereas the last term describes the elongation of the optical path caused by the refractive index of the atmosphere, and the shortening of the optical path due to the slightly shorter physical path taken by the ray, as compared to the path described by the first three terms. Knowing this we should be able to construct an artificial ray that starts and stops at the same point as $\tau(a)$, but when traveling through the atmosphere it takes the path belonging to $\tau(a + \Delta a)$. To make sure that the new ray, let us call it τ_x, ends up at the right point we need to keep θ constant. θ is the angle of separation between the GNSS and LEO satellites, and is given by

$$\theta = \pi + \alpha - \phi_0 - \phi_r \tag{A.4}$$

If we expand to second order in Δa we find the required α_x

$$\alpha_x = \alpha(a) + \Delta a\left(\frac{\partial\phi_0}{\partial a} + \frac{\partial\phi_r}{\partial a}\right) + \frac{1}{2}\Delta a^2\left(\frac{\partial^2\phi_0}{\partial a^2} + \frac{\partial^2\phi_r}{\partial a^2}\right) \tag{A.5}$$

The expansion for the first two terms in Eq. (A.3) can be found by taking derivatives

$$\frac{\partial}{\partial a}(\sqrt{r_{0,r}^2 - a^2}) = -\frac{a}{\sqrt{r_{0,r}^2 - a^2}} \tag{A.6}$$

But since $r_{0,r}\sin\phi_{0,r} = a$ we get

$$\frac{\partial}{\partial a}(\sqrt{r_{0,r}^2 - a^2}) = -a\frac{\partial \phi_{0,r}}{\partial a} \tag{A.7}$$

and

$$\frac{\partial^2}{\partial a^2}(\sqrt{r_{0,r}^2 - a^2}) = -\frac{\partial \phi_{0,r}}{\partial a} - a\frac{\partial^2 \phi_{0,r}}{\partial a^2} \tag{A.8}$$

wherefore

$$\sqrt{r_{0,r}^2 - (a+\Delta a)^2} \approx \sqrt{r_{0,r}^2 - a^2} - a\Delta a\frac{\partial \phi_{0,r}}{\partial a} - \frac{1}{2}\Delta a^2\left(\frac{\partial \phi_{0,r}}{\partial a} + a\frac{\partial^2 \phi_{0,r}}{\partial a^2}\right) \tag{A.9}$$

Finally the expansion for the last term in (A.3) is most easily found taking derivatives as well

$$\frac{\partial}{\partial a}\left(-2\int_a^\infty \sqrt{x^2 - a^2}\frac{d\ln n}{dx}dx\right) = 2a\int_a^\infty \frac{1}{\sqrt{x^2 - a^2}}\frac{d\ln n}{dx}dx + \tag{A.10}$$
$$+ 2\sqrt{x^2 - a^2}\frac{d\ln n}{dx}|_a = -\alpha$$

and

$$\frac{\partial}{\partial a}\left(-2\int_a^\infty \sqrt{x^2 - a^2}\frac{d\ln n}{dx}dx\right) = -\frac{\partial \alpha}{\partial a} \tag{A.11}$$

Hence

$$-2\int_{a+\Delta a}^\infty \sqrt{x^2 - (a+\Delta a)^2}\frac{d\ln n}{dx}dx \approx -2\int_a^\infty \sqrt{x^2 - a^2}\frac{d\ln n}{dx}dx - \alpha\Delta a - \frac{1}{2}\Delta a^2\frac{\partial \alpha}{\partial a} \tag{A.12}$$

We shall now use the formula for the artificial ray τ_x, which is

$$\tau_x = \sqrt{r_0^2 - (a+\Delta a)^2} + \sqrt{r_r^2 - (a+\Delta a)^2} + \alpha_x(a+\Delta a) - \tag{A.13}$$
$$- 2\int_{a+\Delta a}^\infty \sqrt{x^2 - (a+\Delta a)^2}\frac{d\ln n}{dx}dx$$

We insert all the expansions we have found and get (to second order)

$$\tau_x \approx \tau(a) + \frac{1}{2}\Delta a^2\left(\frac{\partial \phi_0}{\partial a} + \frac{\partial \phi_r}{\partial a} - \frac{\partial \alpha}{\partial a}\right) = \tau(a) - \frac{1}{2}\Delta a^2\frac{\partial \theta}{\partial a} \tag{A.14}$$

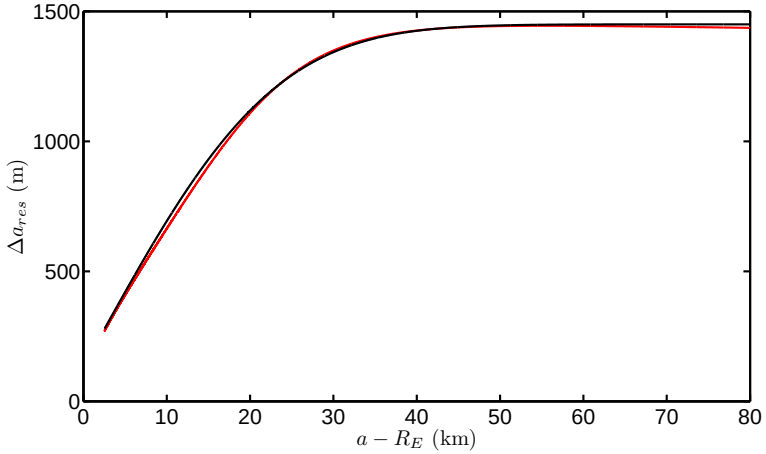

Figure A.1: The Fresnel zone width for a generic refractivity profile ($N = 400\exp(-h/7000)$), red line, and an analytical fit (Eq. A.17), black line.

Hence the Fresnel zone radius is

$$\Delta a_{\lambda/2} = \sqrt{\lambda(\frac{\partial\theta}{\partial a})^{-1}} = \sqrt{\lambda\left(\frac{\partial\phi_r}{\partial a} + \frac{\partial\phi_0}{\partial a} - \frac{\partial\alpha}{\partial a}\right)^{-1}} \tag{A.15}$$

and the diameter, or width of the zone, is $a_{\text{res}} = 2\Delta_{\lambda/2}$.

The expression is easily evaluated since we already have found procedures to calculate $\alpha(a)$, and we can find the two other derivatives through

$$\frac{\partial\phi_{0,r}}{\partial a} = \frac{1}{r_{0,r}\cos\phi_{0,r}} = \frac{1}{\sqrt{r_{0,r}^2 - a^2}} \tag{A.16}$$

In Fig. A.1 we can see the Fresnel zone width for a generic profile, $N = 400\exp(-h/7000)$, and an analytical fit to this curve given by

$$\Delta a_{res} = 280 + 1170\ \text{erf}\left(\frac{a - R_E}{23000}\right) \tag{A.17}$$

Finding the Fresnel zone width as a function of height for realistic complicated profiles is not meaningful. The Fresnel zone width becomes undefined in the case of multipath, and in the case of super-refractivity it goes to zero. Even for refractivity profiles lacking these complications the width will fluctuate wildly, and the limited number of samples used in the refractivity will lead to great uncertainty in the "correct" value for $\partial\alpha/\partial a$. Fortunately this has no importance for the work in this report.

Appendix B

Straight line tangent altitude (SLTA)

B.1 The straight line tangent altitude

In the RO community the straight line tangent altitude (SLTA) is used frequently. On the other hand, it's definition, and how to find it is rarely described. For a circular Earth there is one way to define it without ambiguity - the distance from the surface of the Earth to the point where a line drawn between the GNSS and LEO and a line from the Earth's center are perpendicular. Using this definition we can find a compact formula for the SLTA in the coordinate system used in this report. Using the coordinate system in Fig. 5.2, the coordinates of the GNSS are (Y_0, Z_0), and the coordinates of the LEO are (Y_r, Z_r). The equation for the line between GNSS and LEO is

$$Y(Z) = \left(\frac{Y_r - Y_0}{Z_0 - Z_r}\right) Z + \frac{Y_r Z_0 - Y_0 Z_r}{Z_0 - Z_r} = kZ + m \tag{B.1}$$

The line that is perpendicular to this one and passes through the origin is given by

$$Y_\perp(Z) = -\frac{1}{k} Z = -\left(\frac{Z_0 - Z_r}{Y_r - Y_0}\right) Z \tag{B.2}$$

These lines cross at

$$Y_{x,\perp}(Z_x) = Y_x(Z_x) \quad \Leftrightarrow \quad Z_x = -\frac{mk}{1+k^2}, \quad Y_x = \frac{m}{1+k^2} \tag{B.3}$$

and the SLTA is given by

$$\text{SLTA} = \sqrt{Y_x^2 + Z_x^2} - R_E = \frac{m}{\sqrt{1+k^2}} - R_E \tag{B.4}$$

In the case when the ellipticity of the Earth is taken into account the local radius of the Earth (i.e. directly underneath the crossing point) will have to be used, and one will have to be quite careful in what one means by the SLTA.

Bibliography

[1] M. Bonnedal, J. Christensen, A. Carlström, and A. Berg, "Metop-gras in-orbit instrument performance," *GPS solutions*, vol. 14, no. 1, pp. 109–120, 2010.

[2] V. I. Tatarski, *Wave Propagation in a Turbulent Medium*. Dover Publications Inc., 1967.

[3] M. Levy, *Parabolic Equation Methods for Electromagnetic Wave Propagation*. IET Electromagnetic Waves Series, The Institution of Engineering and Technology, London, 2009.

[4] E. T. Karayel and D. P. Hinson, "Sub-fresnel-scale vertical resolution in atmospheric profiles from radio occultation," *Radio Science*, vol. 32, no. 2, pp. 411–423, 1997.

[5] D. L. Knepp, "Multiple phase-screen calculation of the temporal behavior of stochastic waves," *Proceedings of the IEEE*, vol. 71, no. 6, pp. 722–737, 1983.

[6] J. Martin and S. M. Flatté, "Intensity images and statistics from numerical simulation of wave propagation in 3-d random media," *Applied Optics*, vol. 27, no. 11, pp. 2111–2126, 1988.

[7] S. V. Sokolovskiy, "Modeling and inverting radio occultation signals in the moist troposphere," *Radio Science*, vol. 36, no. 3, pp. 441–458, 2001.

[8] H. H. Benzon, A. S. Nielsen, and L. Olsen, "An atmospheric wave optics propagator - theory and application," 2003. Scientific Report 03-01, Danish Meteorological Institute.

[9] M. E. Gorbunov and A. S. Gurvich, "Microlab-1 experiment: Multipath effects in the lower troposphere," *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 103, no. D12, pp. 13819–13826, 1998.

[10] H. H. Benzon and M. E. Gorbunov, "Description of a wave propagator and radio occultation simulations," 2012. Eumetsat Study: Optimisation of tracking strategies for radio occultations Version 2.2.4.

[11] M. Born and E. Wolf, *Principles of Optics*, vol. 7th expanded edition. Cambridge University Press, Cambridge, 1999.

[12] E. R. Kursinski, G. A. Hajj, S. S. Lery, and B. Herman, "The gps radio occultation technique," *Terrestrial, Atmospheric and Oceanic Sciences, TAO*, vol. 11, no. 1, pp. 53–114, 2000.

[13] M. Gorbunov and K. Lauritsen, "Analysis of wave fields by fourier integral operators and their application for radio occultations," *Radio Science*, vol. 39, no. 4, 2004.

[14] M. Gorbunov and K. Lauritsen, "Linearized zverev transform and its application for modeling radio occultations," *Radio Science*, vol. 42, no. 3, 2007.

[15] A. Sommerfeld, *Optics*, vol. IV of *Lectures on Theoretical Physics*. Academic Press, London, 2009.

[16] J. W. Goodman, *Introduction to Fourier Optics*. McGraw-Hill, London, 1968.

[17] M. E. Gorbunov, A. S. Gurvich, and L. Bengtsson, "Advanced algorithms of inversion of GPS/MET satellite data and their application to reconstruction of temperature and humidity," 1996. rep. 211, Max-Planck-Institute for Meteorology, Hamburg.

[18] A. Sommerfeld, *Partial Differential Equations in Physics*, vol. VI of *Lectures on Theoretical Physics*. Academic Press, London, 2009.

[19] M. D. Mortensen, "The back-propagation method for inversion of radio occultation data," 1998. Scientific Report 98-14, Danish Meteorological Institute.

[20] M. D. Meincke, "Inversion methods for atmospheric profiling with GPS occultations," 1998. Scientific Report 99-11, Danish Meteorological Institute.

[21] M. E. Gorbunov, A. S. Gurvich, and L. Kornblueh, "Comparative analysis of radioholographic methods of processing radio occultation data," *Radio science*, vol. 35, no. 4, pp. 1025–1034, 2000.

[22] M. E. Gorbunov, "Radioholographic methods for processing radio occultation data in multipath regions," 2001. Scientific Report 01-02, Danish Meteorological Institute.

[23] S. Sokolovskiy, W. Schreiner, C. Rocken, and D. Hunt, "Detection of high-altitude ionospheric irregularities with gps/met," *Geophysical research letters*, vol. 29, no. 3, pp. 3–1, 2002.

[24] M. Gorbunov, "Canonical transform method for processing radio occultation data in the lower troposphere," *Radio Science*, vol. 37, no. 5, pp. 9–1, 2002.

[25] A. S. Jensen, M. S. Lohmann, H.-H. Benzon, and A. S. Nielsen, "Full spectrum inversion of radio occultation signals," *Radio Science*, vol. 38, no. 3, 2003.

[26] A. S. Jensen, M. S. Lohmann, A. S. Nielsen, and H.-H. Benzon, "Geometrical optics phase matching of radio occultation signals," *Radio science*, vol. 39, no. 3, 2004.

[27] S. B. Healy, "Optimising tracking strategies for radio occultation. task 1 - the profile dataset," 2012. EUMETSAT/ECMWF Project Report.

[28] G. Fjeldbo and V. Eshleman, "Atmosphere of venus as studied with the mariner 5 dual radio-frequency occultation experiment," *Radio Science*, vol. 4, no. 10, pp. 879–897, 1969.

[29] S. Sokolovskiy, "Effect of superrefraction on inversions of radio occultation signals in the lower troposphere," *Radio Science*, vol. 38, no. 3, 2003.

[30] C. Rocken, R. Anthes, M. Exner, D. Hunt, S. Sokolovskiy, R. Ware, M. Gorbunov, W. Schreiner, D. Feng, B. Herman, *et al.*, "Analysis and validation of GPS/MET data in the neutral atmosphere," *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 102, no. D25, pp. 29849–29866, 1997.

[31] E. R. Kursinski, *The GPS radio occultation concept: theoretical performance and initial results*. PhD thesis, Calif. Inst. of Technol., Pasadena, 1997.

[32] P. I. Palmer, J. Barnett, J. Eyre, and S. Healy, "A nonlinear optimal estimation inverse method for radio occultation measurements of temperature, humidity, and surface pressure," *Journal of Geophysical Re-*

search: Atmospheres (1984–2012), vol. 105, no. D13, pp. 17513–17526, 2000.

[33] X. Zou, B. Wamg, H. Liu, R. Aathes, T. Matsumura, and Y.-J. Zhu, "Use of GPS/MET refraction angles in three-dimensional variational analysis," *Quarterly Journal of the Royal Meteorological Society*, vol. 126, no. 570, pp. 3013–3040, 2000.

[34] C. Ao, "Effect of ducting on radio occultation measurements: An assessment based on high-resolution radiosonde soundings," *Radio Science*, vol. 42, no. 2, 2007.

[35] C. O. Ao, T. Meehan, G. Hajj, A. Mannucci, and G. Beyerle, "Lower troposphere refractivity bias in gps occultation retrievals," *Journal of Geophysical Research: Atmospheres (1984–2012)*, vol. 108, no. D18, 2003.

[36] F. Xie, S. Syndergaard, E. R. Kursinski, and B. M. Herman, "An approach for retrieving marine boundary layer refractivity from gps occultation data in the presence of superrefraction.," *Journal of Atmospheric & Oceanic Technology*, vol. 23, no. 12, 2006.

[37] F. Xie, *Development of a GPS Occultation Retrieval Method for Characterizing the Marine Boundary Layer in the Presence of Super-Refraction*. PhD thesis, University of Arizona, 2006.