SCIAMACHY Offline Processor Level1b-2 ATBD

Algorithm Theoretical Baseline Document (SGP OL Version 7)

Issue 3

prepared	institute	date	signature
G. Lichtenberg, S. Hrechanyy	DLR-IMF	16.04.18	Günter Lichtenberg
checked			
SQWG			
copies to			
ESA			

Digital signiert von Günter Lichtenberg
AN: CN=Günter Lichtenberg, O=DLR, OU=IMF-ATP, E=guenter.lichtenberg@dlr.de, C=DE
Grund: Ich bin der Verfasser dieses Dokuments
Ort: Oberpfaffenhofen
Datum: 23-11-2018 11:29:09
Foxit Reader Version: 9.1.0

Contents

List of Figures

List of Tables

Changelog

Change	Page	Section
Issue 1 (10.05.2010)	all new	all new
Issue 1A (26.01.2011)		
Changed: SACURA 3rd bit default value from 'full' to 'no convergence'	-	-
Issue 2 (2.09.2014)		
Spelling errors corrected	all	all
Updated AAIA algorithm description	22	4.2.5
Added SPICI to cloud fraction algorithm description	19	3.2.3
Clarified background database explanation	30	5.5.2
Added section on HCHO	33	5.8
Added section on CHOCHO	35	5.9
Added retrieval settings table to CO section	57	7.4.2
Added section on CH_4	57	7.5
Updated Limb Ozone Profile retrieval settings	115	8.6
Added section on NLC detection	122	9.3.3
Corrected the bit notation to big endian	-	-
Added CH_4 to format description	-	-
Issue 2 A (27.03.2015)		
Added Section on the Retrieval of tropospheric NO_2	126	10
Issue 2 B (13.05.2015)		
Added remark about trop. $NO_2$to MDS description	13	2.3
Added SCODA setting table	123	9.4
Issue 3 (16.04.18)		
Limb Cloud settings updated	117	9
Section about tropospheric BrO added	129	11
Corrected averaging kernel definition and description for calculation of VMR for Limb	112	8.4
Updates to SWIR chapter	40	7
Removed Appendix with L2 structure and referrences thereto, no longer needed		

Part I.

Introduction

1. About this document

1.1. Purpose and scope

SCIAMACHY is a joint project of Germany, the Netherlands and Belgium for atmospheric measurements. SCIAMACHY has been selected by the European Space Agency (ESA) for inclusion in the list of instruments for Earth observation research for the ENVISAT polar platform, which has been launched in 2002. After 10 years in space, the contact to the ENVISAT was lost on the 8th April 2012. On the 9th May 2012, ESA declared the official end of the ENVISAT Mission. The SCIAMACHY programme was under the supervision of the SCIAMACHY science team (SSAG), headed by the Principal Investigators Professor J. P. Burrows (University of Bremen, Germany), Professor I.A.A. Aben (SRON, The Netherlands) and Dr. C. Muller (BIRA, Belgium). The Quality Working Group has been installed in 2007 to intensify the development and implementation of the Algorithm Baseline for the operational data processing system of SCIAMACHY. Current members of the QWG are the University of Bremen (IUP-B) (Lead), BIRA, DLR, and SRON. The expertise of KNMI is brought in via an association with SRON.

This document describes the algorithms used in the Level 1b-2 processing of SCIAMACHY data. Focus is the mathematical description and not the technical implementation.

1.2. Historical Background

The SCIAMACHY instrument was conceived in the mid-1980's in recognition of the need to monitor chemically important trace species on a global basis using passive remote sensing devices. The importance of monitoring global ozone distributions was demonstrated by the TOMS (Total Ozone Monitoring Spectrometer) instrument on board Nimbus 7, especially after the discovery of the ozone hole in 1985. The SCIAMACHY measurement strategy and mission objectives were based in part on the successful extension in the early 1980s of differential spectroscopic methods to retrieve atmospheric trace gas amounts from ground-based measurements using moderate-resolution spectrometers. Following an approach to ESA (European Space Agency) an initial study was commissioned and the results published in 1988. The scientific consortium was then invited to carry out a detailed Phase A feasibility study, which was finished in 1991. During this time, a scaled-down version of SCIAMACHY was commissioned and accepted for inclusion on the ERS-2 satellite. This "mini-SCIAMACHY" became the GOME (Global Ozone Monitoring Experiment) instrument, which has been functioning successfully since the 1995 launch of ERS-2. SCIAMACHY was accepted as an AO (Announcement of Opportunity) instrument to be included on the ENVISAT satellite. For the Phase C/D completion, the flight model was built in 1998, and calibrated in early 1999. ENVISAT was launched in March 2002 and after the commissioning phase, SCIAMACHY was in operational mode between July 2002 and April 2012, when the contact to the ENVISAT was lost.

1.3. Acronyms and Abbreviations

AMC-DOAS	Air Mass Corrected Differential Optical Absorption Spectroscopy
AMF	Air Mass Factor
AO	Announcement of Opportunity
ATBD	Algorithm Theoretical Baseline Document

BIRA	Belgisch Instituut voor Ruimte-Aëronmie
BIRRA	Beer InfraRed Retrieval Algorithm
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Centre)
DOAS	Differential Optical Absorption Spectroscopy
DRACULA	aDvanced Retrieval of the Atmosphere with Constrained and Unconstrained Least squares Algorithms
ECMWF	European Centre for Medium-Range Weather Forecasts
ENVISAT	ENVIronmental SATellite
ESA	European Space Agency
GOME	Global Ozone Monitoring Experiment
IUP-B	Institut für Umweltphysik Bremen (Institute for Environmental Physics)
MDS	Measurement Data Set
PMD	Polarisation Measurement Device
RTM	Radiative Transfer Model
SGP	SCIAMACHY Ground Processor
SCIAMACHY	Scanning Imaging Absorption spectroMeter for Atmospheric CHartographY
SRON	Space Research Organisation Netherlands
TOMS	Total Ozone Monitoring Spectrometer

1.4. Document Overview

The document is divided into three parts plus appendices:

1. Introduction (this part)
2. Nadir retrieval algorithms: Cloud parameters, AAI and total columns of trace gases
3. Limb retrieval algorithms: Cloud parameters and profiles of trace gases
4. Appendices describing data structure etc.

Part two and three start with a general description of the retrieval method. After that the settings for the individual retrievals are described as well as additional or deviating methods used in the retrieval.

2. The Data Processor

2.1. Processing flow overview

The processing from Level 1b to Level 2 products is separated into the following steps (generally performed on individual fitting windows):

1. Calibration of data, i.e. generation of *internal* Level 1c data, using the same algorithms as the scial1c applicator according to the calibration settings in the configuration file
2. Calculation of the ratio Sun/Earth (note that this is not necessarily the reflectance, since not radiometric calibrated data can be used)
3. Climatological pre-processing:
   a) Calculation of Cloud Parameters (fraction, optical depth, top height)
   b) Initialisation of data bases
4. Nadir- and limb retrievals
5. Writing of Level 2 MDS and offline product

2.2. Components of the Data Processor

The SGP consists of three main components

1. The database server
2. The nadir retrieval server
3. The limb retrieval server

Other components taking care of communication between the different running processes, data delivery to and from the SGP_12OL (also referred to simply as SGP or offline processor in this document) and interfaces to the main host running the ENVISAT processors are left out here. A more technical view of the SGP_12OL, including the outward interfaces is given in the Architectural Design Document (Kretschel, 2006). The primary input to the SGP_12OL is a SCIAMACHY Level 1b file, processed by the Level 0-1b processor (Slijkhuis and Lichtenberg, 2014). Within the ground segment the SGP only has one Auxiliary file as input, the initialisation file (Lichtenberg and Kretschel, 2009). All other auxiliary inputs like topographic databases or pressure profiles needed for the trace gas retrieval are handled within the SGP itself, i.e. those are integral part of the offline processor itself. Generally, the processing is done state wise: Each state from an Level 1b input file (usually encompassing a complete orbit) is relegated to one instance of the Nadir or Limb retrieval server. The first step in the processing is the calibration of the Earthshine spectra to radiances. Note, that no Level 1c product is written in the Level 1b to 2 processing, the calibrated spectra solely exist as internal data within the processor. However, the scial1c tool (Scherbakov, 2008) is based on the same algorithms as the calibration in the SGP and can be used to generate Level 1c products. Both deliver the same results which makes it possible to easily verify the SGP w.r.t. scientific reference algorithms.

More than one retrieval server can run at a given time. When the last state of the input product is processed, the results are collected and used to generate the Level 2 product.

2.2.1. Database Server

The database server is a separate process, which handles the auxiliary input for the retrievals. It is split up into the GeoDataServer which provides topographical data, the RefSpecDataServer which provides the reference spectra for the fits of the measured spectrum and CommonDataServer for the remaining data like pressure or temperature profiles. Upon triggering by the retrieval server, the database server not only reads in the requested data, but also does simple tasks like interpolating the correct wavelength grid. It also caches data that are needed by more than one retrieval to accelerate the processing. The server is set up in a generic way making it easy to add or exchange data needed for a given retrieval.

2.2.2. Nadir Retrieval Server

Before the nadir measurements are processed, the cloud parameters and AAI are retrieved in the climatological pre-processing. The cloud parameters are used in the subsequent retrieval of the trace gases. For the Nadir DOAS retrievals the kernel of the GDP was modified to accept SCIAMACHY input. The retrieval algorithms themselves were not modified meaning that the SGP products have the same quality as the GOME products albeit on a higher spatial resolution. The basis for the SGP DOAS implementation is the SDOAS scheme developed by BIRA, which is closely based on the GDOAS developed for GOME.

2.2.3. Limb Retrieval Server

The Limb retrieval is done by the DLR developed package DRACULA. While various models are implemented in DRACULA, the SGP uses the Iterative Regularized Gauss-Newton Method. It includes sophisticated techniques to select the iteration and regularisation parameters. A polynomial is fitted to radiances to remove broad spectral features and ratioed radiances are constructed for each tangent height. Finally the logarithm of the ratios determines the basis for the retrieval.

2.3. Overview of MDS

The so called measurement data sets (MDSs) contain the measurements of the atmosphere. Four MDS's exist:

Cloud & Aerosol MDS Contains the cloud parameters measured in nadir geometry: cloud fraction, cloud top height and cloud optical thickness. The algorithms are described in chapter 3.

Nadir MDS Contains the trace gas columns: O_3, NO_2, HCHO, SO_2, BrO, H_2O, CHOCHO, OClO, CO and CH_4. The algorithms are explained in chapters 5, 6 and 7

Limb MDS Contains the trace gas profiles of $O_{3,}$ NO_2 and BrO. The algorithms are explained in chapter 8.

Limb Cloud MDS Contains height resolved indicators for cloud presence and type (water clouds, PSCs and NLCs). The algorithm is explained in chapter 9.

Part II.

Nadir Retrieval Algorithms

3. Cloud Parameters

3.1. Cloud Top Height and Cloud Optical Thickness

3.1.1. Overview

Clouds play an important role in the Earth climate system (Kondratyev and Binenko, 1984) . The amount of radiation reflected by the Earth-atmosphere system into outer space depends not only on the cloud cover and the total amount of condensed water in the Earth atmosphere but also the size of droplets a_{ef} and their thermodynamic state is also of importance.

The information about microphysical properties, cloud top height and spatial distributions of terrestrial clouds on a global scale can be obtained only with satellite remote sensing systems. Different spectrometers and radiometers , deployed on space-based platforms, measure the angular and spectral distribution of intensity and polarization of reflected solar light (see e.g. Deschamps et al., 1994). Generally, the measured values depend both on geometrical and microphysical characteristics of clouds. Thus, the inherent properties of clouds can be retrieved (at least in principle) by the solution of the inverse problem. The accuracy of the retrieved values depends on the accuracy of measurements and the accuracy of the forward radiative transfer model.

In particular, it is often assumed that clouds can be represented by homogeneous and infinitely extended in the horizontal direction plane-parallel slabs (Kokhanovsky et al., 2003; Rossow, 1989). The range of applicability of such an assumption for real clouds is very limited as is shown by observations of light from the sky on a cloudy day. For example, the retrieved cloud optical thickness τ is apparently dependent on the viewing geometry (Loeb and Davies, 1996). This, of course, would not be the case for an idealized plane-parallel cloud layer. However, both the state-of-art radiative transfer theory and computer technology are not capable to incorporate 3-D effects into operational satellite retrieval schemes. As a result, cloud parameters retrieved should be considered as a rather coarse approximation to reality.

However, even such limited tools produce valuable information on terrestrial clouds properties. For example, it was confirmed by satellite measurements that droplets in clouds over oceans are usually larger than those over land . This feature, for instance, is of importance for the simulation of the Earth's climate .

Concerning trace gas retrievals in the UV/VIS, clouds are considered as "contamination". The part of the column, which is below the top of the clouds, cannot be detected by the satellite. This ghost vertical column has to be estimated from climatological vertical profiles and is added to the vertical column retrieved. It is determined by integrating the profile from surface up to the *cloud top height*. Partial cloudiness within the field of view can be taken into account using fractional cloud cover.

The Semi-Analytical CloUd Retrieval Algorithm (SACURA) is used in the SGP OL1b-2 to retrieve the cloud top height and the cloud optical thickness. Cloud top height is derived from measurements in the oxygen A absorption band as recommended by (Yamomoto and Wark, 1961). In addition, we retrieve the cloud geometrical thickness from the fit of the measured spectrum in the oxygen A-band and optical thickness using the wavelength 755 nm outside of the band.

3.1.2. Detailed Description

3.1.2.1. Cloud Top Height

The determination of the cloud top height h using SACURA is based on the measurements of the top-of-atmosphere (TOA) reflection function R in the oxygen A-band (Rozanov and Kokhanovsky, 2004). The cloud

reflection function is extremely sensitive to the cloud top height in the centre of the oxygen absorption band. To find the value of h, we first assume that the TOA reflectance R can be presented in the form of a Taylor expansion around the assumed value of cloud top height equal to h_0 :

$$R(h) = R(h_0) + \sum_{i=1}^{\infty} a_i (h - h_0)^i \tag{3.1}$$

where $a_i = R^{(i)}(h_o)/I$. Here $R^{(i)}(h_0)$ is the i^{th} derivative of R at the point h_0. The next step is the linearisation, which is a standard technique in the inversion procedures (Rodgers, 2000; Rozanov et al., 1998). We found that the function $R(h)$ is close to a linear one in a broad interval of the argument change (Kokhanovsky and Rozanov, 2004). Therefore, we neglect non-linear terms in equation (3.1). Then it follows:

$$R = R(h_0) + R'(h_o) \cdot (h - h_0) \tag{3.2}$$

where $R' = \frac{dR}{dh}$. We assume that R is measured at several wavelengths in the oxygen A-band. Then instead of the scalar quantity R we can introduce the vector $\underline{R}_{mes}$ with components $R(\lambda_1), R(\lambda_2), \cdots R(\lambda_n)$. The same applies to other scalars in equation (3.1). Therefore, it can be written in the following vector form:

$$\underline{y} = \underline{a}x \tag{3.3}$$

where $\underline{y} = \underline{R}_{mes} - \underline{R}(h_0)$, $\underline{a} = \underline{R}'(h_0)$, and $x = h - h_0$. Note that both measurement and model errors are contained in equation (3.3). The solution $\hat{x}$ of the inverse problem is obtained by the minimizing the following cost function:

$$\Phi = \|\underline{y} - \underline{a}x\|^2 \tag{3.4}$$

where $\|\cdots\|$ means the norm in the Euclid space of the correspondent dimension. The value of $\hat{x}$, where the function Φ has a minimum can be represented as the scalar product in the Euclid space

$$\hat{x} = \frac{\sum_{i=1}^{n} a_i y_i}{\sum_{i=1}^{n} a_i^2} \tag{3.5}$$

with n being the number of spectral channels where the reflection function is measured.

Therefore, from known values of the measured spectral reflection function R_{mes} (and also values of the calculated reflection function R and its derivative R' at $h = h_0$) at several at wavelengths, the value of the cloud top height can be found from equation (3.5) and equality: $h = \hat{x} + h_0$. The value of h_0 can be taken equal to 1.0 km, which is a typical value for low level clouds. The main assumption in our derivation is that the dependence of R on h can be presented by a linear function on the interval x (Kokhanovsky and Rozanov, 2004). An error of 0.25 km can be expected for full convergence; otherwise 0.5 km.

3.1.2.2. Cloud Optical Thickness

Cloud optical thickness τ, together with the cloud fraction c, has a significant impact on the transfer of solar and infra-red radiation through a cloudy atmosphere. τ is defined as the cloud extinction coefficient k_{ext} integrated across the cloud vertical extent. For a vertically homogeneous cloud with geometrical thickness L, $\tau = k_{ext}L$.

For optically thick clouds, the optical thickness in the visible is estimated from the following equation (Kokhanovsky et al., 2003):

$$R(\mu, \mu_0, \phi) = R_{0\infty}(\mu, \mu_0, \phi) - t_{cloud} K_0(\mu) K_0(\mu_0), \tag{3.6}$$

where $R(\mu, \mu_0, \phi) = \pi I^{\uparrow}(\mu, \mu_0, \phi)/\mu_0 F_0$ is the cloud reflection function as measured from the satellite, μ_0 and μ are the cosines of the solar and observation zenith angles, ϕ is the relative azimuth, F_0 is the top-of-atmosphere irradiance on the plane perpendicular to the incident light beam, and $I^{\uparrow}$ is the intensity of the reflected light, $t_{cloud} = 1/(a + b\tau)$. Here t_{cloud} is the cloud spherical transmittance, $a = 1.07$ and $b = 0.75(1 - g)$.

Assuming a given cloud model, e.g., spherical particles and a polydisperse distribution with a given effective radius of droplets, a_{ef}, or predefined ice crystals shapes and size distributions, the escape function $K_0(\mu)$,

and the reflection function for a semi-infinite layer $R_{0\infty}$ are pre-calculated and stored in look-up-tables (LUTs). Approximate equations for these functions can be used as well. In particular, a good approximation for $K_0(\mu)$ is:

$$K_0(\mu) = \frac{3}{7}(1 + 2\mu). \tag{3.7}$$

One derives:

$$\tau = \frac{1}{b}\left[\frac{K_0(\mu)K_0(\mu_0)}{R_\infty(\mu, \mu_0, \phi) - R(\mu, \mu_0, \phi)} - a\right]. \tag{3.8}$$

This equation is used in SACURA for cloud optical thickness retrievals. It follows from this equation that retrievals of τ for very thick clouds ($R \to R_\infty$) are highly uncertain and small errors (e.g., calibration errors) in the measured reflection function will lead to large errors in the retrieved cloud optical thickness. Often, a limiting value of cloud optical thickness is used in the retrieval process e.g., 100, as most clouds, but certainly not all have optical thickness below 100.

3.2. Cloud Fraction

3.2.1. Introduction

The basic idea in OCRA (Optical Cloud Recognition Algorithm Loyola and Ruppert, 1998) is to decompose optical sensor measurements into two components: a cloud-free background and a remainder expressing the influence of clouds, i.e.

$$R = R_{CF} + R_C \tag{3.9}$$

where R is the reflectance and the indices CF and C indicate the cloud-free component and the cloudy component respectively. The reflectance for a given location (x, y), and wavelength λ can be defined as

$$R(x, y, \lambda) \quad = \quad \frac{I(x, y, \lambda)}{I_0(\lambda) \cdot \cos\theta_0 \cdot \cos\theta} \tag{3.10}$$

with

I	Upwelling radiance into the field of view
I_0	Solar irradiance
θ	Viewing Zenith angle
θ_0	Solar Zenith Angle

The key to the algorithm is the construction of the cloud-free composite that is invariant with respect to the atmosphere, and to topography and solar and viewing angles. The comparison of the spatially resolved cloud-free scenes and the measurement gives then the cloud fraction for a given scene. As for the GOME processor, in SCIAMACHY PMD data are used to determine the cloud fraction. These have the advantage that they have a higher spatial resolution than the spectrally resolved measurements ($\approx$ 7 x 30 km). The steps for the calculation of the cloud fraction are

1. Generation of a database with the cloud-free composite by
   a) Extracting PMD signals and calculating a relative reflectance
   b) Find cloud-free scenes

c) Grid data

2. Calculation of cloud fraction

   a) Determine offset and scaling factor

   b) For each measurement, sum up signal of all PMDs and compare with cloud-free composite to determine the cloud fraction

   c) For each measurement, average over subpixels

3.2.2. Generation of the cloud-free composite

In previous versions of the SCIAMACHY data processor the GOME cloud-free composite database was used. Starting from the version 5 of the processor PMD measurements of SCIAMACHY were used to build a database of cloud-free measurements. First the PMD values of the Earth and Sun measurements were extracted from Level 1b data (version 6.03, the latest available version) in the time range from 2003 to 2007. Earth measurements are downsampled to 32 Hz in the Level 1b from the original 40 Hz while the sun measurements still have the original sampling. Additionally, the degradation of the instrument has to be corrected:

$$R_{PMD} = \frac{PMD_{Earth}}{PMD_{Sun}} \cdot \frac{m_{PMD}}{m_{PMDcal}} \cdot \frac{32Hz}{40Hz} \tag{3.11}$$

with $PMD_{Earth/Sun}$ as the signal of PMD Earth and Sun measurements and $m_{PMD(cal)}$ as the degradation correction factor for Earth and Sun PMD measurements. For version 6 of the processor the new m-factors derived from the scan mirror model (Slijkhuis and Lichtenberg (2014)) were used. The above equation holds for all PMDs. In order to determine if the observed scene is cloud-free, the saturation S of the PMD reflectances is calculated :

$$\begin{aligned} S &= \frac{\max(R_{PMD2}, R_{PMD3}.R_{PMD4}) - \min(R_{PMD2}, R_{PMD3}, R_{PMD4})}{\max(R_{PMD2},\, R_{PMD3},\, R_{PMD4})} \\ &= 1 - \frac{\min(R_{PMD2}, R_{PMD3}, R_{PMD4})}{\max(R_{PMD2}, R_{PMD3}, R_{PMD4})} \end{aligned} \tag{3.12}$$

with R_{PMDn} as the signal in PMD number *n*. The saturation was preferred over the colour distance (that is used for the GOME cloud-free composite), because it can be easily expanded for a future separation of cloudy scenes from ice or snow covered surfaces. Both methods are completely equivalent, since they are a measure for the difference between a coloured scene and achromatic (white/grey/black) scene. Totally clouded scenes have a low saturation near 0, because they are white in the wavelength range measured by PMDs 2-4 and the signals in all of them will be similar. A cloud-free scene is indicated by a large saturation.

The database with the cloud-free composite is build on a grid with a resolution of $0.36° \times 0.36°$ (latitude x longitude). Each grid point is initialised with a saturation value of 0 (fully clouded). The database contains the following values:

1. Latitude and longitude of the grid point
2. Centre latitude and longitude of the measurement with the highest saturation so far
3. Reflectance values for all PMDs divided by the solar zenith angle
4. Solar Zenith angle
5. Saturation
6. Colour distance from the white point (see Loyola, 2000)

For each month in the considered time range the saturation values of the observations is sequentially calculated. If the measured saturation is larger than the one on the nearest grid point in the database and thus indicating that the observed scene is less cloudy than the one currently in the database, all values in the database for this grid point are replaced with those of the current measurements. The resulting monthly databases are combined to one database after all data are processed. In future versions of the processor the monthly databases can be used to build a seasonally dependent cloud-free composite to decrease the impact of colour variations of the surface over time.

3.2.3. Determination of the cloud fraction

For a given PMD observation the cloud fraction c_{frac} can now be determined as

$$c_{frac} = \sqrt{\sum_{i=2}^{4} \left[\left(R_{PMDi} - R^{cf}_{PMDi} \right)^2 - \alpha_i \right] \cdot \beta_i} \tag{3.13}$$

with

R_{PMDi}	Observed reflectance of PMD i
R^{cf}_{PMDi}	Reflectance for the cloud-free case from the database
α_i	Offset for PMD i
β_i	Scaling factor for PMD i

The offsets correct for the fact that the measured reflectances can be *smaller* than the cloud-free reflectance in the database. Possible reasons can be seasonal change of vegetation or ocean colour. Since the sign of the difference $R_{PMDi} - R^{cf}_{PMDi}$ is lost, cloud-free areas with lower brightness will give the same cloud fraction as slightly clouded scenes. By setting the offset to a large enough value the cloud fraction can be corrected for this effect. Using a small subset of the reflectance data set, the offset α_i was set to the largest negative value of the reflectance difference found in the set. Note that a future extension of the database to include seasonal variations can minimise the influence of the colour changes.

The scaling factors β_i correct for the fact that the albedo of observed clouds is not uniform but depends the illumination geometry, the thickness of the cloud, the size of the water droplets and other parameters. In effect this leads to different reflectances for scenes with the same cloud cover and eventually to different cloud fractions. In the current algorithm, only one scaling factor is available for each PMD, independent of season and geographical coordinates. In order to minimise errors in the calculation of the cloud fraction, the scaling factors were determined by a comparison with FRESCO+ (Wang et al., 2008).

Version 6 also contains the cloud ice discrimination SPICI using the algorithm as described in Krijger et al. (2005). With this algorithm cloud free subpixels above ice are now properly detected[1]. SPICI employs a similar method as OCRA to determine the colour of the sub-pixel. It uses additionally PMD D (800 -900 nm) to discriminate clouds from ice.

3.2.4. Future improvements

Since the calculation and extraction of SCIAMACHY data from several years is rather time consuming, possible future algorithm improvements were already considered during the generation of the cloud-free composite. These improvements include:

- ➢ Determination of the offset using the complete data set and not only a subset
- ➢ Independent determination of the scaling factors using a composite of totally clouded scenes
- ➢ Introduction of seasonal and geographically dependent offsets and scaling factors

[1] ice and cloud were not distinguishable by their colour using only PMDs in the visible range

4. Absorbing Aerosol Index

4.1. Overview

The Absorbing Aerosol Index (AAI) indicates the presence of elevated absorbing aerosols in the Earth's atmosphere. The aerosol types that are mostly seen in the AAI are desert dust and biomass burning aerosols. The AAIs described in this ATBD are derived from the reflectances measured by SCIAMACHY at 340 and 380 nm.

4.1.1. Retrieval Settings

The wavelength pair used is $[340, 380]$ nm. Both wavelengths originate from cluster 9 of the SCIAMACHY spectrum. The associated reflectances at these two (centre) wavelengths are determined by taking the average of all the reflectances that fall inside a 1-nm wavelength bin around the respective centre wavelengths.

4.2. Detailed Description

4.2.1. Definition of used geometry

Figure 4.1 defines, in a graphical way, the angles that are used to specify the viewing and solar directions. An imaginary volume element that is responsible for the scattering of incident sunlight is located in the origin of the sphere. The solar direction is described by the solar zenith angle θ_0 and the solar azimuth angle ϕ_0. The viewing direction is defined by the viewing zenith angle θ and the viewing azimuth angle ϕ.

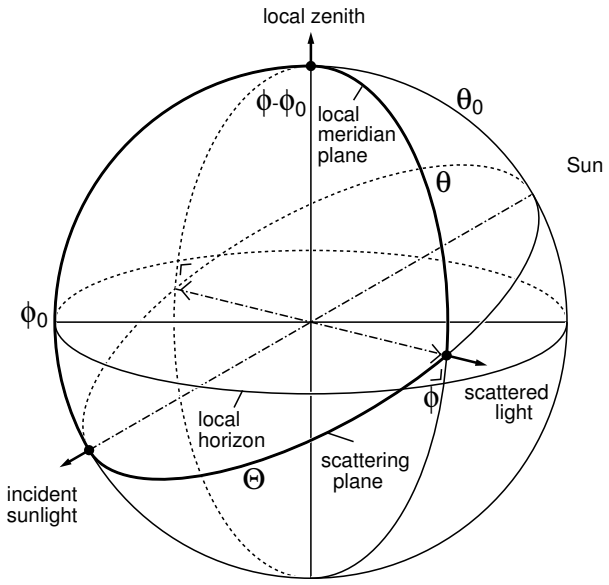

Figure 4.1.: *Definition of viewing and solar angles.*

We make use of a right-handed coordinate frame, meaning that $\phi - \phi_0$ as sketched in figure 4.1 is positive. We followed the same definition of ϕ and ϕ_0 as is present in the SCIAMACHY data. However, there is no need to specify it because only the azimuth difference $\phi - \phi_0$ is of relevance. For completeness, we mention that the single scattering angle, under this definition of angles, can be calculated from

$$\cos\theta = -\cos\theta\cos\theta_0 + \sin\theta\sin\theta_0\cos(\phi - \phi_0) \tag{4.1}$$

4.2.2. Definition of the Residue

The Absorbing Aerosol Index (AAI) is a dimensionless quantity which was introduced to provide information about the presence of UV-absorbing aerosols in the Earth's atmosphere. It is derived directly from another quantity, the residue, which is defined by Herman et al. (1997) as

$$r = -100 \cdot \{^{10}\log(\frac{R_\lambda}{R_{\lambda_0}})^{obs} - ^{10}\log(\frac{R_\lambda}{R_{\lambda_0}})^{Ray}\} \tag{4.2}$$

In this equation, R_λ denotes the Earth's reflectance at wavelength λ. The superscript obs refers to reflectances which are measured by, in this case, SCIAMACHY, while the superscript Ray refers to modelled reflectances. These modelled reflectances are calculated for a cloud-free and aerosol-free atmosphere in which only Rayleigh scattering, absorption by molecules, Lambertian surface reflection as well as surface absorption can take place. The reflectance is defined as

$$R = \frac{\pi I}{\mu_0 E} \tag{4.3}$$

where I is the radiance reflected by the Earth atmosphere (in $Wm^{-2}nm^{-1}sr^{-1}$), E is the incident solar irradiance at the top of the atmosphere perpendicular to the solar beam (in $Wm^{-2}nm^{-1}$), and μ_0 is the cosine of the solar zenith angle θ_0. As for the wavelengths involved, the wavelengths λ and λ_0 must lie in the UV, and were set to 340 and 380 nm, respectively, for SCIAMACHY.

The Rayleigh atmosphere in the simulations is bounded below by a Lambertian surface having a wavelength independent surface albedo A_s, which is not meant to represent the actual ground albedo. It is obtained from requiring that the simulated reflectance equals the measured reflectance at $\lambda_0 = 380$ nm:

$$R^{obs}_{\lambda_0} = R^{Ray}_{\lambda_0}(A_s) \tag{4.4}$$

The surface albedo found in this way is used to calculate R^{Ray}_λ, so one assumes that the surface albedo is constant in the wavelength interval $[\lambda, \lambda_0]$, which is true for most cases. Note that equation (4.2) can now be reduced to

$$r = -100 \cdot ^{10}\log\left(\frac{R^{obs}_\lambda}{R^{Ray}_\lambda}\right) \tag{4.5}$$

4.2.3. Definition of the AAI

The importance of the residue, as defined above, lies in its ability to effectively detect the presence of absorbing aerosols even in the presence of clouds. When a positive residue ($r > 0$) is found, absorbing aerosols were detected. Negative or zero residues on the other hand ($r \leq 0$), suggest an absence of absorbing aerosols. For that reason, the AAI is defined as equal to the residue r where the residue is positive, and it is not defined where the residue has a negative value.

4.2.4. Calculation of the Residue

The first step in the calculation of the residue involves finding the artificial surface albedo A_s for which the measured SCIAMACHY reflectance at the reference wavelength $\lambda_0 = 380$ nm equals the simulated reflectance at the same wavelength (cf. equation 4.4). The simulated reflectances are found from a collection of Look-Up Tables (LUTs). These LUTs are described in section 4.2.5. The simulations basically describe a cloud-free, homogeneous atmosphere which is bounded below by a Lambertian surface. The contribution of the surface to the TOA reflectance may be separated from that of the atmosphere according to the following formula (Chandrasekhar, 1960):

$$R(\mu, \mu_o, \phi - \phi_0, A_s) = R^0(\mu, \mu_o, \phi - \phi_0) + \frac{A_s t(\mu) t(\mu_o)}{1 - A_s s^*} \tag{4.6}$$

In this equation, the first term R^0 is the path reflectance, which is the atmospheric contribution to the reflectance. The second term is the contribution of the surface with an albedo A_s. The parameter t is the total atmospheric transmission for the given zenith angle, s^* is the spherical albedo of the atmosphere for illumination from below, μ is the cosine of the viewing zenith angle θ, and μ is the cosine of the solar zenith angle θ_0 defined earlier. Using equation (4.6) and by demanding that the simulated reflectance $R^{Ray}_{\lambda_0}$ equals the measured reflectance $R^{obs}_{\lambda_0}$ at 380 nm, we find the following expression for the surface albedo A_s:

$$A_s = \frac{R^{obs}_{\lambda_0} - R^0_{\lambda_0}}{t(\mu)t(\mu_0) + s^*(R^{obs}_{\lambda_0} - R^0_{\lambda_0})} \tag{4.7}$$

In this equation, $R^0_{\lambda_0}$ denotes the simulated reflectance for the actual atmosphere, but without surface reflection. R^0 can be expanded in a Fourier series. In our case, with a Rayleigh atmosphere, the expansion is exact with only three terms in $\phi - \phi_0$. We have

$$R^0 = a_0 + \sum_{i=1}^{2} 2a_i(\mu, \mu_0) \cos i(\phi - \phi_0) \tag{4.8}$$

The idea is that with a proper LUT of a_i, $t(\mu)$ and s^*, we can easily calculate R_0, A_s and R_λ. Some interpolation is necessary. In this case we have to interpolate over μ and μ_0 and over the surface height h_s.

4.2.5. Look-up Tables

Using the radiative transfer code DAK (Doubling-Adding KNMI, de Haan et al., 1987), Look-Up Tables (LUTs) were created. The LUTs contain the parameters a_i, $t(\mu)$ and s^* as well as the total Ozone column, as defined in section 4.2.4. The definition of the coefficients a_i is such that the path reflectance R_0 may be reconstructed from them according to

$$R_0 = a_0 + 2a_1 \cos(\phi - \phi_0) + 2a_2 \cos 2(\phi - \phi_0) \tag{4.9}$$

Adding the surface contribution requires equation (4.6) and the parameters $t(\mu)$ and s^* . The simulations were done for cloud-free situations. Aerosols were not included either. The LUTs were prepared as a function of μ and μ_0, surface height h_s and total ozone column. The coefficients were calculated for 42 Gaussian distributed μ and μ_0 points, and for 9 surface heights h_s ranging from 0 to 8 km in 1 km steps. The ozone total column retrieved by SCIAMACHY for the given observation is used for the look-up of the correct values in the LUT.

4.2.6. Degradation Correction

Since the algorithm relies on absolute values for the reflectance, the degradation has to be corrected. The residue increased by, on average, more than 2.5 index points over the last years, and an analysis indicated

that the increase could be completely explained by the instrument degradation which was recorded by instrument Light Path Monitoring (LPM) measurements (which are taken on a regular basis). In the processor, the degradation is now corrected during the internal Level 1b to 1c step . Details of the degradation correction can be found in Slijkhuis and Lichtenberg (2014) and Krijger (2011).

5. SDOAS Algorithm General Description

The vertical column retrieval algorithm GDOAS was developed at BIRA-IASB specifically for GOME on the basis of the well-established and widely used DOAS method (Platt and Stutz, 2008). GDOAS was implemented into the GOME data processor (GDP) version 4. The GDOAS adaptation to the SCIAMACHY instrument was renamed SDOAS and has also been implemented in the SCIAMACHY Ground Processor L12 (from version 3) to have an accuracy similar to the one obtained with GDP4.

This section presents a brief description of the DOAS method that involves two steps: the slant column retrieval and the conversion into vertical columns using calculated air mass factors. The detailed description of GDOAS is provided in (Spurr et al., 2004; Van Roozendael et al., 2006; Lerot et al., 2009).

5.1. Slant column retrieval

The Lambert-Beer extinction law applied to the earthshine measurements is the basis for the slant column retrieval from satellite measurements. A direct solar irradiance measurement void of atmospheric absorption structures is used as the background spectrum. The validity of this law is limited to the weak absorption, which is generally true in case of the atmospheric trace gas absorbers. However, in very particular conditions, the DOAS approximation may be not verified (e.g. in the UV for high solar zenith angles due to the strong ozone absorption or in case of high SO_2 concentrations during volcanic eruptions). In SDOAS it is generally the logarithms of intensities (so-called "optical densities") that are being used. The effects of the broadband molecular and aerosol scattering are represented in the form of the low-order polynomials (typically of the 2nd or 3rd degree). The main equation used for the DOAS retrieval is then

$$\begin{aligned} Y(\lambda) &= \ln\left(\frac{I_\lambda(\Theta)}{I^0_\lambda(\Theta)}\right) \\ &= -\sum_g E_g(\Theta)\cdot\sigma_g(\lambda) - \sum_{j=0}^{n}\alpha_j\left(\lambda-\lambda^*\right)^j - \alpha_R\sigma_R(\lambda) \end{aligned} \tag{5.1}$$

with

$I_\lambda(\Theta)$	Earth radiance at wavelength λ for the geometrical path Θ
I^0_λ	Solar irradiance
E_g	Effective slant column density of gas g
σ_g	Differential absorption cross-section of gas g
λ^*	Reference wavelength for closure polynomial
σ_R	Ring reference spectrum
α_R	Scaling parameter for Ring spectrum

The DOAS retrieval consists in a minimization of the weighted least squares difference between measured and calculated optical densities *Y(λ)* in the spectral interval appropriately selected for the trace gas under study. To calculate optical density, reference absorption cross-sections are provided by independent measurements or are pre-calculated. For instance, the Ring effect caused by the inelastic Raman scattering is treated as a

pseudo-absorber. The corresponding Ring spectrum is usually defined as a (logarithmic) ratio between the radiance spectra with and without inelastic Raman scattering (Chance and Spurr, 1997).

Two measurements of the solar irradiance are realized daily using the two diffusers (ESM and ASM) mounted on the instrument. The algorithm uses the most recent update of the sun spectrum measured by one of the two diffusers, depending on the trace gas to retrieve. The wavelength calibration of the irradiance spectrum is further improved using a cross-correlation procedure involving a high-resolution reference solar spectrum. In addition, to correct for possible distortion of the wavelength grid (e.g. due to the Doppler effect caused by the satellite movement), the wavelength grid of the earthshine spectrum is allowed to be shifted in the DOAS procedure.

The retrievals performed in the channel 2 of SCIAMACHY (BrO, SO_2, $OClO$) revealed a problem, which is due to the presence of a strong polarization-related anomaly occurring right in the middle of the channel. This anomaly interferes with the absorption structures of the retrieved gases that leads to the non-physical results. To reduce this polarization-related artefact, a polarization vector η (often called ETA), extracted from the SCIAMACHY key data set, is included into the SDOAS fit procedure for some of the trace gases. Also, to account for possible intensity offsets in the spectra, a supplementary cross-section calculated as the inverse earthshine spectrum is fitted in the DOAS procedure. This term also corrects for some limitations of the Ring spectrum, e.g. over bright surfaces or in regions where vibrational Raman scattering in ocean water is relevant and is used for most minor trace species.

5.2. Vertical column retrieval

The conversion from slant to vertical column amounts is done via the AMF computation.

The AMF definition used in the SGP is the common one

$$A = \frac{\ln\left(\frac{I_{nog}}{I_g}\right)}{\tau_{vert}} \tag{5.2}$$

with

A	Air mass factor
I_{nog}	Simulated backscatter radiances without main gas
I_g	Simulated backscatter radiance with main gas
τ_{vert}	Vertical optical depth of the main gas

These AMFs are directly calculated on the fly with the multiple scattering radiative transfer model LIDORT (Linearized Discrete Ordinate Radiative Transfer), without the need for large look-up tables. LIDORT assumes the atmosphere as a stratified structure with a given number of optically homogeneous layers. For more information about LIDORT see Spurr et al. (2001).

Since the AMF depends not only on viewing geometry and solar position but also on the wavelength and the concentration vertical profile shape, it is calculated for each retrieved gas individually. The concentration profile shape is generally provided by an external climatology. AMF calculation demands also an information about a surface albedo. In the SGP it is taken from the TOMS albedo data base. Two types of calculations are possible with the data processor:

1. An iterative approach when the total column is one of the input parameters for the selection of the profile within the climatology. Currently, this method is only used for ozone (see below) as it is only needed for strong absorbers.
2. A standard approach when the climatology is only based on seasonal and geographical parameters.

For strongly varying absorbers such as SO_2, the use of climatologies leads to large errors. Therefore, in the current processor two standard AMF are calculated, one for a volcanic eruption scenario and one for a boundary layer pollution situation. Both results are provided to the user who has to decide which value is more appropriate for his application.

For OClO, no vertical column is calculated at all as rapid photochemistry and the change of solar zenith angle along the light path through the stratosphere at low sun make application of the concept of vertical columns difficult.

In general, two types of AMF are calculated: for clear sky (A_{clear}; from the top of the atmosphere to the ground) and for cloudy (A_{cloud}; from the top of the atmosphere to the cloud-top level) scenarios. The cloudy air mass factor calculation is based on cloud parameters generated off-line by dedicated algorithms.

The vertical column is then calculated as

$$V = \frac{\frac{E}{M_R} + c_{frac} \cdot G \cdot A_{cloud}}{(1 - c_{frac}) \cdot A_{clear} + c_{frac} \cdot A_{cloud}} \tag{5.3}$$

with

V	Vertical column density
E	Slant column density of main gas
M_R	Molecular Ring correction scale factor
c_{frac}	Intensity weighted cloud fraction
G	“Ghost column” of main gas below cloud top height
A_{cloud}	AMF to the cloud top level
A_{clear}	AMF to the ground level

The molecular Ring correction M_R is specific to total ozone retrievals (see below)). For other trace gases, this factor is fixed to 1.

The "ghost column" is the part of the column hidden below the cloud and thus not detectable by the satellite instrument. Therefore, for the cloudy pixels, a climatological gas column comprised between the ground and the cloud top needs to be added to the retrieved one. The "ghost column" concept works reasonably well for trace gases with more or less constant tropospheric distributions and prevents the cloud-induced "jumps" in retrieved vertical column fields. For strongly varying species such as SO_2, no ghost column is applied.

5.3. Nadir Ozone Total Column Retrieval

5.3.1. Motivation

The role of ozone in the Earth’s atmosphere is very important since it absorbs solar UV photons, protecting the biosphere from these harmful radiations. Also, the radiation absorption by the high concentrations of ozone in the stratosphere leads to an heating of this part of the atmosphere. This temperature increase in the stratosphere makes this region very stable and influences the dynamic of the whole atmosphere. The discovery of the ozone hole above Antarctica during spring time was one of the most obvious changes observed in Earth’s atmosphere. Although the Montreal Protocol and its amendments have regulated the production and release of ozone depleting substances, the ozone recovery expected to occur in the next decades has to be monitored on a long-term basis. Furthermore, climate change will have an impact on the atmospheric dynamics and chemistry which influence the ozone concentrations. Global measurements of ozone columns can help to better understand these interactions and consequently to better predict the timing of ozone recovery.

Although 90% of ozone is in the stratosphere, tropospheric ozone concentrations can locally be important. In the troposphere, ozone is a pollutant and a constituent of smog. It is also a powerful oxidizing agent and an important greenhouse gas. Total ozone measurements may be combined to stratospheric ozone measurements to derive information on tropospheric ozone content.

5.3.2. Retrieval Settings

Ozone absorption is strongly temperature-dependent in the UV Huggins bands, and in this spectral range an effective temperature is derived from the DOAS fitting in addition to the ozone slant column itself.

Level 1b-1c Settings	
Calibration	All calibrations applied
SMR	D0 (Sun over ESM diffuser)
DOAS Main Settings	
Fitting Interval	325 - 335 nm
Polynomial Degree	3rd order
Wavelength Shift	Wavelength calibration of sun reference optimized over fitting interval by NLLS adjustment on pre-convolved NEWKPNO atlas
Absorption Cross Sections/Fitted Curves	
O_3	Bogumil et al. (2003)@243 K, shifted by 0.02nm, scaled by 1.03
NO_2	Bogumil et al. (2003)@243 K
O_3 *Difference Spectrum*	Difference $\sigma_{O3}^{243K} - \sigma_{O3}^{223K}$, shift allowed, scaled by 1.03
Ring Spectrum	Calculated by convolution of the Chance and Spurr (1997) solar atlas with RRS cross-sections of molecular N2 and O2.
Total Column Calculation: Profiles/AMF	
AMF ref. Wavelength	325.5 nm
O_3 *Profile*	Column-classified ozone profile climatology TOMS V.8
Radiative Transfer Model	LIDORT
Cloud Top Height	SACURA
Cloud Fraction	OCRA

5.3.3. Additional Details

5.3.3.1. Molecular Ring correction

In addition to the filling-in of Fraunhofer absorption solar lines, rotational Raman scattering also causes the filling-in of atmospheric absorption features. This has to be considered for precise total ozone retrievals. It is shown in Van Roozendael et al. (2006) that a simple correction for this molecular Ring effect can be realized by dividing the slant column by a scale factor M_R calculated with the following equation

$$M_R = 1 + \alpha_R \overline{\sigma_R} \left(1 - \frac{\sec(\theta_0)}{(1 - c_{frac}) \cdot A_{clear} + c_{frac} \cdot A_{cloud}} \right) \tag{5.4}$$

with symbols as in equation 5.3 and

α_R	fractional intensity of Raman light (from DOAS fit)
$\overline{\sigma_R}$	average Ring cross section calculated over the fitting interval
θ_0	Solar zenith angle

It is calculated “on the fly” during the vertical column retrieval and usually not available separately.

Following a request from users, the slant column provided in the SGP product is the Ring corrected slant column (i.e. E/M_R).

5.3.3.2. Vertical column with iterative approach

For ozone, the total column is a good proxy for the ozone profile, and column-classified ozone profile climatology is useful for selecting profiles required for the radiative transfer. In GDOAS, the ozone AMF is adjusted iteratively to reflect the actual column content as expressed by the slant column.

In the first iteration, an initial AMF A_0 is computed for a given initial guess of vertical column. The vertical column V_0 is obtained, which serves as the second guess for the vertical column. This process runs until the convergence criterion is reached, i.e. until the relative difference between consequent iterations is less than some predefined small number.

5.4. Nadir NO_2 Total Column Retrieval

5.4.1. Motivation

The main source of NO_2 in the stratosphere is microbiological soil activity. The diurnal, seasonal, and latitudinal variation of NO_2 is dominated by the equilibrium between NO_x and the reservoir substances (mainly N_2O_5, HNO_3, $ClONO_2$). In the troposphere, the main sources of NO_2 are anthropogenic in origin, e.g. industry, power plants, traffic and forced biomass burning. Other, but slightly less important origins comprise natural biomass burning, lightning and microbiological soil activity. NO_2 emissions have increased by more than a factor of 6 since pre-industrial times with concentrations being highest in large urban areas.

Global monitoring of NO_2 emissions is a crucial task since nitrogen oxides play a central role in atmospheric chemistry. For example, it catalyses ozone production, contributes to acidification and also adds to radiative forcing. The high spatial resolution of SCIAMACHY allows us performing very detailed observations of polluted regions.

5.4.2. Retrieval Settings

Level 1b-1c Settings	
Calibration	All calibrations except radiometric
SMR	A0 (Sun over ASM diffuser without radiometric calibration)
DOAS Main Settings	
Fitting Interval	426.5 - 451.5 nm
Polynomial Degree	2nd order
Wavelength Shift	Wavelength calibration of sun reference optimized over fitting interval by NLLS adjustment on pre-convolved NEWKPNO atlas
Absorption Cross Sections/Fitted Curves	
NO_2	Bogumil et al. (2003)@243 K
O_3	Bogumil et al. (2003)@243 K, shifted 0.025 nm
O_2 - O_2	Greenblatt et al. (1990); Wavelength axis corrected by Burkholder
H_2O	Generated from HITRAN database
Ring Spectrum	Calculated by convolution of the Chance and Spurr (1997) solar atlas with RRS cross-sections of molecular N_2 and O_2.
Empirical Functions	
Offset and Slope Correction	Inverse Earthshine spectrum (radiance)
Total Column Calculation: Profiles/AMF	
AMF ref. Wavelength	439 nm
NO_2 Profile	HALOE (Lambert et al., 2000)
Radiative Transfer Model	LIDORT
Cloud Top Height	SACURA
Cloud Fraction	OCRA

5.5. Nadir SO_2 Total Column Retrieval

5.5.1. Retrieval Settings

Level 1b-1c Settings	
Calibration	All calibrations except radiometric
SMR	A0 (Sun over ASM diffuser without radiometric calibration)
DOAS Main Settings	
Fitting Interval	315 - 327 nm
Polynomial Degree	3rd order
Absorption Cross Sections/Fitted Curves	
SO_2	Vandaele et al. (1994)
O_3	Bogumil et al. (2003)@243 K
O_3 Difference Spectrum	Difference $\sigma_{O3}^{243K} - \sigma_{O3}^{223K}$
Ring Spectrum	Vountas et al. (1998)
Empirical Functions	
Offset and Slope Correction	Inverse Earthshine spectrum (radiance)
Background Correction	SCD from reference sector over the Pacific (180° - 220°)
Total Column Calculation: Profiles/AMF	
AMF ref. Wavelength	315 nm
SO_2 Profile Anthropogenic	Pollution scenario: 1 DU present from surface to 1 km height
SO_2 Profile Volcanic	Eruption scenario: 10 DU present in layer between 10 and 11 km
Radiative Transfer Model	LIDORT
Cloud Top Height	Not applied, ghost column set to 0
Cloud Fraction	Not applied

5.5.2. Additional Details

For both, the slant column computation and the total column retrieval the standard method described above was adjusted.

5.5.2.1. Slant Column

In the slant column determination, a spectral interference at large ozone absorption (high solar zenith angle, large ozone column) leads to unrealistic slant columns in the SO_2 retrieval. To correct for this offset, the SO_2 columns from a reference sector over Pacific ($180°-220°$ longitude) are subtracted from the retrieved SO_2 slant columns. For determination of the offset it is assumed that no SO_2 is present in the so called reference sector over the Pacific and that the interfering effect (mainly ozone absorption) does not depend strongly on longitude. The values from the reference sector (i.e. the assumed offsets) are stored in a database (BG_DB) as background values.

Database Filling and Content

For optimum correction, the BG_DB should contain those measurements over the reference sector which are closest in time to the actual measurement analysed at other longitudes. It should also have no gaps, should not be affected by volcanic SO_2 plumes or outliers and should be available in NRT.

The orbit duration of ENVISAT is roughly 100 minutes leading to 14 orbits per day. Of these 14 orbits typically two cross the background reference sector near the equator. For higher latitudes more crossings occur. However, as result of the gaps during limb measurements and from missing data, not all of the reference sector is covered on each day. The BG_DB contains the nadir SO_2 slant columns retrieved by the off-line processor. It is continuously updated during processing. The values are retrieved with the same algorithm as the "regular" slant columns, their only difference is the geolocation of the ground-pixel. New values are written to the database if

- The centre longitude is in the reference sector (currently all longitudes between 180° and 220°)
- The ground-pixel is in the descending node of the orbit (i.e. on the sunlit side of the orbit).
- The RMS of the retrieval is below a limit (currently 0.007)
- The fractional cloud cover is below a limit (currently 0.5)
- At least one latitudinal bin with quality > 0 was found
- The orbit was not already used to add data to the BG_DB (duplicate entries would distort the quality value and the averaged value)

All values obtained on the same day for a given latitude bin are averaged, weighted with their quality. Each latitudinal bin spans 5° leading to 36 bins per day. In order to judge the quality of the background value, the data base also contains a quality flag that can have values between 0 (no data) and 255 (best quality). The quality is determined from the number of data points available for each bin and the standard deviation of the SO_2 slant columns. The quality of the background correction is written to the SCD FITFLAG. If the Background correction is used, bit #8 is set . The quality itself is coded into three bits (#9-11), resulting into 8 values : the value 0 means no background correction was performed, the value 7 marks the best possible correction and values in between represent intermediate quality.

Application of the Background Values

After the SO_2 retrieval the slant columns are written to the MDS of the product with the offset from BG_DB applied. The offsets depend on latitude and time. Since the database is filled continuously during the processing, at the start of the processing there will always be measurements that have no corresponding (in time and latitude) background value. In this case the following method is used to obtain a background correction:

1. Search the database alternately backward and forward in time, until values from the neighbouring latitude bins are found.
2. Interpolate between the 2 neighbouring SCD offsets to the current latitude.
3. Interpolate between the 2 corresponding quality flags; a latitude near a missing latitude bin thus will automatically result in an SCD offset and a quality of or near 0.
4. Subtract the offset from the measurement.

5.5.2.2. Vertical Column

As already mentioned, VCD calculation was adjusted as well. Since SO_2 is present only sporadically, the use of climatological values for the vertical profiles is not possible. Therefore, the approach was taken to assume a constant profile shape for two typical scenarios:

- a profile with 1 DU of SO_2 between a 0 and 1 km simulating an anthropogenic pollution event;
- a profile with 10 DU of SO_2 between a 10 and 11 km simulating a volcanic eruption.

Accordingly, two types of SO_2 vertical columns - *anthropogenic* and *volcanic* - are computed. Both results are written into two different MDSs of the level 2 product.

The second adjustment concerns the treatment of clouds in the retrieval. The ghost column approach that corrects for column parts hidden below clouds (see section 5.2) can lead to meaningless results in case of SO_2. Contrary to O_3, NO_2 and BrO, the vertical atmospheric SO_2 distribution is neither known nor predictable.

Both of the main sources of SO_2, the volcanic eruptions and the anthropogenic pollution events, are sporadic. Therefore, assuming that SO_2 is present below every cloud will give incorrect results. Consequently, no ghost column is added to retrieved SO_2 vertical columns. In order to stay consistent with this approach, for the AMF factor always A_{clear} is used. The users should be advised that in some measurements SO_2 might be hidden below a cloud resulting in too low SO_2 vertical columns.

5.6. Nadir BrO Total Column Retrieval

5.6.1. Motivation

Bromine monoxide (BrO) is a key atmospheric trace gas playing a major role as catalyst of the ozone destruction in both the stratosphere and the troposphere. The main sources of bromine in the stratosphere are long-lived and short-lived brominated organic compounds of natural and anthropogenic origin (CH_3Br and halons). In the troposphere, large emissions of inorganic bromine have been observed in the polar boundary layer when frost flowers on the surface of the ocean vanish at the end of the winter. These bromine explosion events also result in efficient tropospheric ozone depletion. Finally, an other source of inorganic bromine in the troposphere and possibly in the stratosphere is the release by active volcanoes.

5.6.2. Retrieval Settings

Level 1b-1c Settings	
Calibration	All calibrations except radiometric
SMR	A0 (Sun over ASM diffuser without radiometric calibration)
DOAS Main Settings	
Fitting Interval	336 - 351 nm
Wavelength Shift	Wavelength calibration of sun reference optimized over fitting interval by NLLS adjustment on pre-convolved NEWKPNO atlas
Polynomial Degree	3rd order
Absorption Cross Sections/Fitted Curves	
NO_2	Bogumil et al. (2003)@ 243 K
O_3	Bogumil et al. (2003)@ 243 K, shifted
O_2-O_2	Greenblatt et al. (1990), wavelength axis corrected by Burkholder
BrO	Fleischmann et al. (2004)@223 K
O_3 Difference Spectrum	Difference $\sigma_{O3}^{243K} - \sigma_{O3}^{223K}$, shifted
Ring Spectrum	Vountas et al. (1998)
Empirical Functions	
Polarisation Feature Correction	η Nadir included in the fit
Offset and Slope Correction	Inverse Earthshine spectrum
Total Column Calculation: Profiles/AMF	
AMF ref. Wavelength	343.5 nm
BrO Profile	Climatology based on the 3-D CTM BASCOE from BIRA (Theys et al., 2008)
Radiative Transfer Model	LIDORT
Cloud Top Height	SACURA
Cloud Fraction	OCRA

5.7. Nadir OClO Slant Column Retrieval

5.7.1. Retrieval Settings

Level 1b-1c Settings	
Calibration	All calibrations except radiometric
SMR	A0 (Sun over ASM diffuser without radiometric calibration)
DOAS Main Settings	
Fitting Interval	365 - 389 nm
Polynomial Degree	4th order
Absorption Cross Sections/Fitted Curves	
NO_2	Bogumil et al. (2003)@223 K
O_4	Hermans et al. (1999)
OClO	Kromminga et al. (2003)
Ring Spectrum	Vountas et al. (1998)
Undersampling	Constant Undersamlping Spectrum calculated by IUP
Empirical Functions	
Offset and Slope Correction	Inverse Earthshine spectrum (radiance)
Polarisation Feature Correction	η Nadir included in the fit
Intensity Correction Ratio	Ratio of cloudy and cloud-free measurements calculated by IUP

5.7.2. Additional Details

The OClO retrieval in SDOAS is also changed with respect to the standard approach: an interference, which so far could not be explained, can lead to unrealistically large OClO columns over bright surfaces and clouds. Possible reasons are insufficient correction of the Ring effect of problems in the lv1 data, e.g. from instrument stray light, detector non-linearities or detector memory effects. To correct for this artefact, an empirical function (the ratio between two measurements - one over a bright cloud and another from a close clear pixel) is included into the fit procedure. Experience shows, that inclusion of this function can reduce artefacts over bright surfaces to a high degree. This empirical intensity correction ratio spectrum is sometimes referred to as "magic ratio".

5.8. Nadir HCHO Total Column Retrieval

5.8.1. Motivation

Formaldehyde (H_2CO) and glyoxal (CHOCHO) are formed during the oxidation of the volatile organic compounds (VOCs) emitted by plants, fossil fuel combustion, and biomass burning. Due to a rather short lifetime of formaldehyde and glyoxal, their distribution maps represent the emission fields of their precursors, VOCs. Additionally, HCHO could also be used as a tracer for carbonaceous aerosols. This information will contribute to understanding of biogenic emissions, biomass burning, and pollution.

5.8.2. Retrieval Settings

Level 1b-1c Settings	
Calibration	All calibrations except radiometric
SMR	A0 (Sun over ASM diffuser without radiometric calibration)
DOAS Main Settings	
Fitting Interval	328.5 - 346 nm
Polynomial Degree	5th order
Absorption Cross Sections/Fitted Curves	
NO_2	Vandaele et al. (1998)
O_3	Brion et al. (1998)@228 & 243 K
HCHO	Meller and Moortgat (2000)@298 K
BrO	Fleischmann et al. (2004)@223 K
OClO	Bogumil et al. (2003)@293 K
Ring Spectrum	SCIAMACHY irradiance of 20030329, KPNO solar spectrum, Gaussian slit function
Undersampling	Constant Undersamlping Spectrum calcculated by BIRA
Empirical Functions	
Polarisation Feature Correction	η and ς Nadir included in the fit
Offset and Slope Correction	Inverse Earthshine spectrum (radiance)
Background Correction	SCD from reference sector over the Pacific (180° - 220°)
Total Column Calculation: Profiles/AMF	
AMF ref. Wavelength	340 nm
HCHO Profile	IMAGES (Müller and Brasseur (1995))
Radiative Transfer Model	LIDORT
Cloud Top Height	SACURA
Cloud Fraction	OCRA

5.8.3. Additional Details

Three successive steps are needed to retrieve HCHO total columns from SCIAMACHY nadir spectra: the slant column retrieval, the air mass factor calculation and a post-correction based on the slant columns retrieved over the Pacific Ocean and on the a-priori HCHO profiles.

For the AMF calculation, a-priori concentrations profiles are taken from the CTM IMAGES (resolved monthly and 2° X 1.25°).

Zonally and seasonally dependent artefacts due to interferences with BrO and ozone are strongly reduced by the reference sector method, which is applied in the same way as for the SO_2. After the reference sector correction slant columns are converted into a vertical column

$$VCD = \frac{SCD - SCD_0}{AMF} + VCD_0 \tag{5.5}$$

VCD_0 represents the climatological HCHO vertical column in the reference sector at the latitude of the measurement. The climatological vertical column database is built by integrating the appropriate climatological concentration profiles within the reference sector. In this database, the different climatological columns coming

from different profiles with different longitudes for the same latitude are averaged. VCD_0 is obtained via an interpolation scheme in the a-priori column database (as a function of the latitude and time).

Contrary to a majority of other nadir products formaldehyde measurements are not provided for pixels with solar zenith angle larger than 60°.

5.8.4. User warning

Formaldehyde vertical columns from all pixels are written into the product, regardless of their cloud fraction. It is strongly advised to apply a cloud filtering to the formaldehyde vertical columns. It is to be emphasized that vertical columns of pixels with cloud fraction $\leq$ 40% represent actual concentration of formaldehyde. For larger cloud fractions the information content below the cloud altitude is weak because the column is dominated by a climatological information.

5.9. Nadir CHOCHO Total Column Retrieval

5.9.1. Retrieval Settings

Level 1b-1c Settings	
Calibration	Uncalibrated data: only dark current (leakage), memory effect and etalon correction applied
SMR	A0 (Sun over ASM diffuser without radiometric calibration)
DOAS Main Settings	
Fitting Interval	435 - 457 nm
Polynomial Degree	4th order
Absorption Cross Sections/Fitted Curves	
CHOCHO	Volkamer et al. (2005)
NO_2	Bogumil et al. (2003)@223 K
O_3	Bogumil et al. (2003)@273 K
O_4	Greenblatt et al. (1990)
H_2O *liquid*	Pope and Fry (1997)
Ring Spectrum	Vountas et al. (1998)
Empirical Functions	
Offset and Slope Correction	Inverse Earthshine spectrum (radiance)
Background Correction	SCD from reference sector over the Pacific (180° - 200°)
Total Column Calculation: Profiles/AMF	
AMF ref. Wavelength	446 nm
CHOCHO Profile	IMAGES (Müller and Brasseur (1995))
Radiative Transfer Model	LIDORT
Cloud Top Height	SACURA
Cloud Fraction	OCRA

5.9.2. Additional Details

For the retrieval of glyoxal total columns, the DOAS method is used including an offset correction by means of the reference sector method (as in cases of sulphur dioxide and formaldehyde total columns). The only difference is that the offset correction value for glyoxal does not depend on latitude. Another important peculiarity of the glyoxal post-correction scheme is the size of the reference sector: whereas it extents from pole to pole and between 140°W and 180°W for SO_2 and HCHO, it extents only between 160°W and 180°W for glyoxal.

Vertical profiles of glyoxal for AMF calculations are taken from the IMAGES climatology. This climatology contains profiles of formaldehyde as well as those of glyoxal with a temporal resolution of 1 month and a spatial resolution of 2°lat x 1.25°lon.

5.9.3. User warning

For glyoxal the same data policy as for formaldehyde is applied: vertical columns from all pixels are written into the product. In case of glyoxal it is recommended to use the vertical columns of pixels with cloud fraction $\leq$ 20%. For larger cloud fractions the information content is dominated by a climatological information.

6. Water Vapour Retrieval

This chapter describes the Air Mass Corrected Differential Optical Absorption Spectroscopy (AMC-DOAS) algorithm used for the retrieval of total water vapour columns from measurements SCIAMACHY. The algorithm has also been successfully applied to measurements of the GOME and GOME-2 instruments.

6.1. Inversion Algorithm

The AMC-DOAS algorithm (Noël et al., 1999) is based on the well-known Differential Optical Absorption Spectroscopy (DOAS) approach (Platt, 1994) which has been modified to handle effects arising from the strong differential absorption structures of water vapour. The general features of this modified DOAS method are that

1. saturation effects arising from highly structured differential spectral features which are not resolved by the measuring instrument are accounted for, and
2. O_2 absorption features are fitted in combination with H_2O to determine a so-called air mass factor (AMF) correction which compensates to some degree for insufficient knowledge of the background atmospheric and topographic characteristics, like surface elevation and clouds.

The main equation of the Air Mass Corrected DOAS method is given by:

$$\ln\left(\frac{I}{I_0}\right) = P - a\left(\tau_{O_2} + c \cdot C_V^b\right) \tag{6.1}$$

with

I, I_0	Earthshine radiance and solar irradiance
P	Polynomial to correct for broadband contributions (resulting e.g. from Rayleigh and Mie scattering or surface albedo)
τ_{O_2}	Optical depth of O_2
C_V	Vertical column amount of water vapour
b, c	Spectral quantities describing saturation effect and absorption

c contains the effective reference absorption cross section and the air mass factor. The scalar parameter a is the above mentioned AMF correction factor. The quantities τ_{O_2} , b, and c are determined from radiative transfer calculations performed for different atmospheric conditions and solar zenith angles (see Section 6.2). C_V and a are then derived from a non-linear fit. The error of the vertical column is calculated from the covariance matrix also resulting from the fit. A typical fit result is shown in Figure 6.1

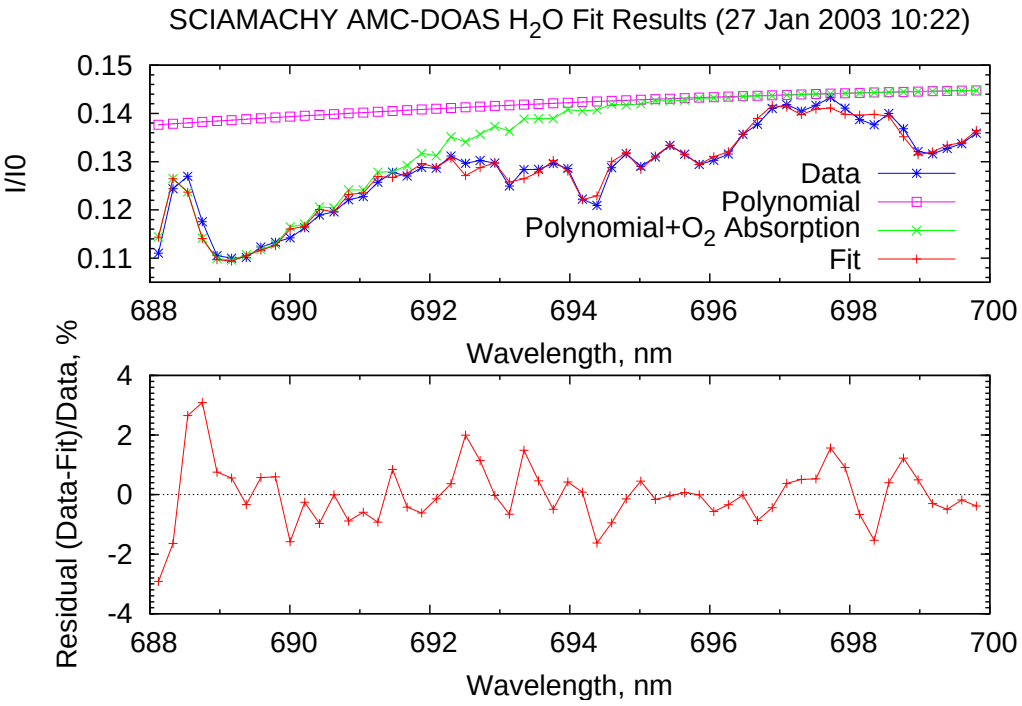

Figure 6.1.: *Example for AMC-DOAS fit results. Top: Spectral fit. Bottom: Residual*

The AMC-DOAS method is applied to the spectral region between 688 nm and 700 nm because both O_2 and water vapour absorb in this region. They are the main absorbers in this spectral range, having slant optical depths of similar strength (see Figure 6.2).

The main purpose of the AMF correction factor is to correct the retrieved water vapour column, but beside this the AMF correction factor can be used as an inherent quality check for the retrieved data. The AMC-DOAS retrieval assumes a cloud-free tropical background atmosphere and does not consider different surface elevations. If the derived AMF correction is too large, this is an indication that these assumptions are not valid (most likely because the observed scene is too cloudy or contains a high mountain area). Experience has shown that retrieved water vapour columns for an AMF correction factor of 0.8 or higher are reliable; only these data are distributed.

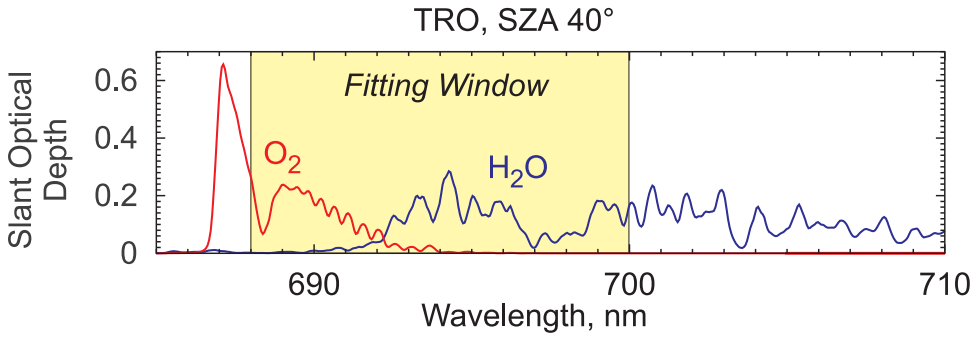

Figure 6.2.: *Typical slant optical depths of water vapour and O_2 in the spectral region around the AMC-DOAS fitting window.*

6.2. Forward Model and Databases

The τ_{O_2} , b, and c parameters are spectral quantities which are taken from a precalculated data base. This data base has been derived using the radiative transfer model (RTM) SCIATRAN (see e.g. Rozanov et al., 2002) as a forward model. The term τ_{O_2} is determined using radiative transfer calculations with and without O. The parameters b and c are determined from radiative transfer calculations assuming different water vapour columns C_V (see Noël et al., 1999, for further details).

The spectra for τ_{O_2}, b and c have been calculated for a set of solar zenith angles (SZAs), namely 0°, 20°, 40°, 50°, 60°, 70°, 80°, 85°, and 88°. Based on this data set the τ_{O_2} , b, and c spectra are then interpolated during the retrieval for the actual SZA. 88° is therefore the maximum SZA for which the retrieval produces reliable results.

The following assumptions have been made for the radiative transfer calculations:

- tropical background atmosphere (MODTRAN profile)
- no clouds
- surface elevation 0 km

An example for τ_{O_2}, b and c is shown in Figure 6.3. Deviations of the real conditions from these assumptions are handled via the retrieved AMF correction factor a.

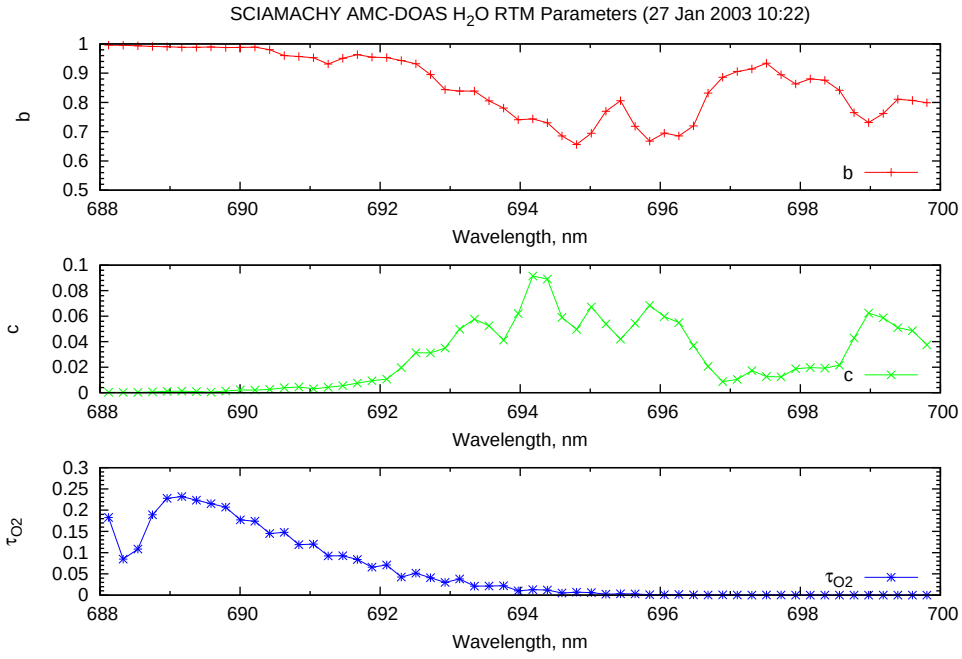

Figure 6.3.: *Example for RTM parameters τ_{O_2}, b and c.*

6.3. Auxiliary Data

Except for the RTM data base described in Section 6.2, the AMC-DOAS method does not rely on any other external data, e.g. calibration factors derived from comparisons with ground based radio sonde measurements as it is the case for e.g. Special Sensor Microwave Imager (SSM/I) data. The retrieved water vapour columns therefore provide a completely independent data set.

7. Infra-red Retrieval using BIRRA

7.1. Introduction

Nadir sounding of vertical column densities of atmospheric gases is well established in atmospheric remote sensing. For UV instruments such as SCIAMACHY (Bovensmann et al., 1999) the analysis is traditionally based on a DOAS–type methodology. This approach has also been successfully applied for analysis of SCIAMACHY (Scanning Imaging Absorption spectroMeter for Atmospheric CHartograghY) near infra-red (IR) channels (see e.g. Buchwitz et al., 2007; Gloudemans et al., 2005; Frankenberg et al., 2005) and was the basis of the BIAS (Basic Infra-red Absorption Spectroscopy) non-linear least squares algorithm (Spurr, 1998). In addition to the column scale factors for the molecular densities a one-parameter "closure term" is fitted to account for any other effects such as (single or multiple) scattering, surface reflection or instrumental artefacts. However, recent sensitivity studies have shown the importance of adequate modelling the instrumental slit function; in particular an under- or overestimate of the slit function half width can have a significant impact on the retrieved columns. Furthermore, in BIAS molecular absorption is taken into account using look–up tables that have been calculated for a coarse altitude grid version of the US standard atmosphere (Anderson et al., 1986) and a dated set of spectroscopic line parameters. Likewise the slant path optical depth is interpolated from look-up tables generated for a finite set of path geometries.

In order to gain greater flexibility in the forward modelling and a more efficient and robust least squares inversion, a "Beer Infra-Red Retrieval Algorithm" (BIRRA) has been implemented Gimeno García et al. (2011) :

- Forward model based on GARLIC (Generic Atmospheric Radiative Transfer Line-by-Line Infrared Code, (Schreier et al., 2014) describing radiance according to Beer's law;
- Non-linear least squares exploiting special structure of the forward model;
- Flexible choice of additional fit parameters besides molecular columns;
- Option for 'online' line-by-line absorption cross section computation and continua.

7.1.1. A first glimpse at near IR nadir observations

Carbon monoxide is an important atmospheric trace gas, highly variable in space and time, that affects air quality and climate. About half of the CO emissions come from anthropogenic sources (e.g. fossil fuel combustion), and further significant contributions are due to biomass burning. CO is a target species of several space-borne instruments, e.g. AIRS, MOPITT, and TES from NASA's EOS satellite series, and SCIAMACHY.

However, carbon monoxide retrieval from SCIAMACHY nadir observations is rather challenging: Only channel 8 from 2259 to 2386 nm features CO absorption signatures, albeit very weak and superposed by stronger absorption lines of concurrent gases, i.e. H_2O and CH_4, see Figure7.1. Additionally, an ice layer on the detector modifies the measured signal. Even worse, degradation of the detector increasingly reduces the number of reliable pixels, i.e. only about 50 of 1024 pixels in channels 8 are useful for CO retrieval, when using the WFM-DOAS bad/dead pixel mask. The "default" atmospheric model used in the original BIAS code is based on the US standard atmosphere (Anderson et al. (1986)). The influence of the atmospheric model on the vertical path transmission is shown in 7.2. Aerosols and clouds can be important in NIR radiative transfer modelling, cf. 7.3.

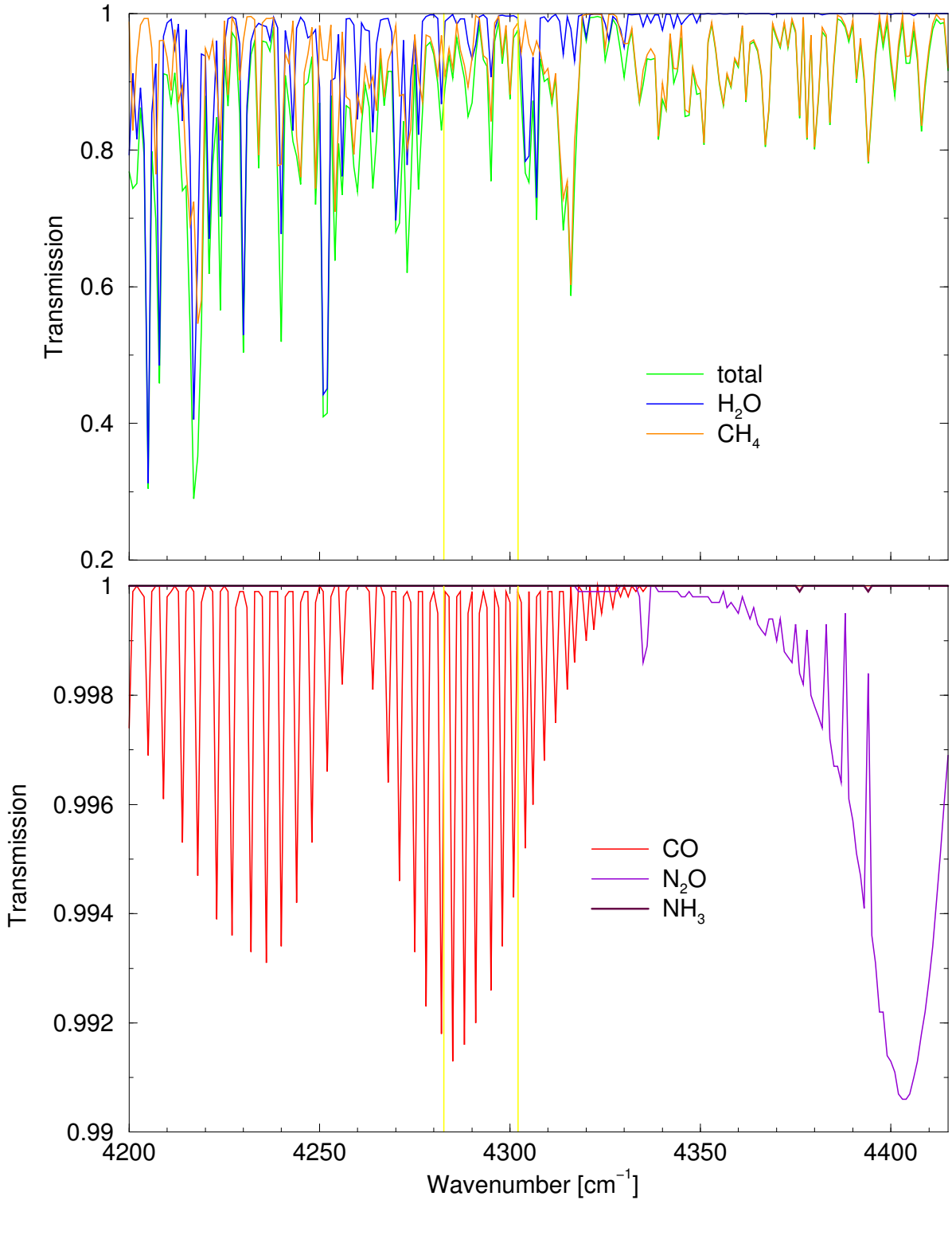

Figure 7.1.: *Atmospheric transmission in channel 8. MODTRAN band model, US standard atmosphere, vertical path 0 – 100km, no aerosols. The yellow bars indicate the micro-window 4282.686 – 4302.131 cm−1 as defined by the WFM–DOAS bad/dead pixel mask.*

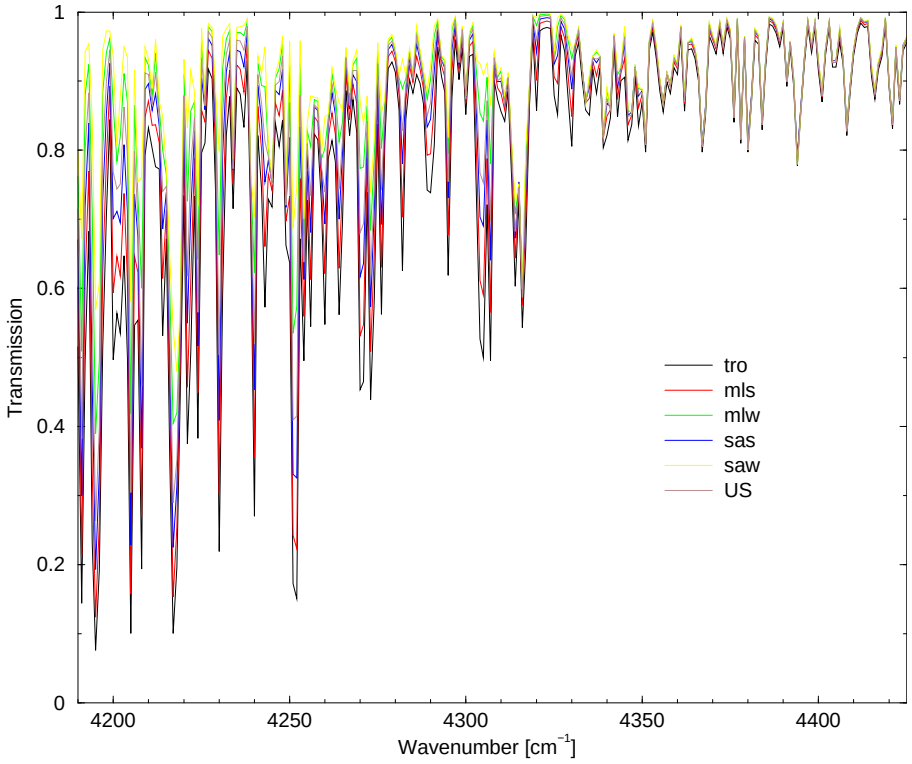

Figure 7.2.: *Comparison of vertical transmission (0 – 100km) assuming various atmospheric models.*

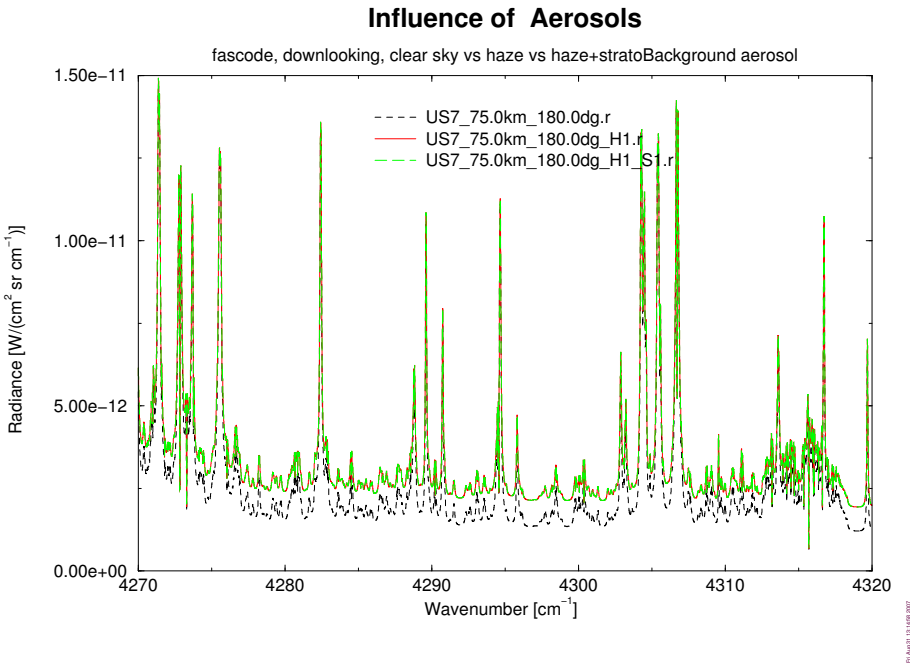

Figure 7.3.: *Comparison of vertical transmission assuming various aerosol models.*

7.2. Forward Modelling

7.2.1. Radiative Transfer

7.2.1.1. General Considerations: Schwarzschild and Beer

For an arbitrary slant path, the intensity (radiance) I at wavenumber ν received by an instrument at position $s = 0$ is given by (neglecting scattering and assuming local thermodynamic equilibrium) the (integral form of the) Schwarzschild equation(Liou, 1980):

$$I(\nu) \;=\; I_{\mathrm{b}}(\nu)\,\mathcal{T}\,(\nu, s_{\mathrm{b}}) \;-\; \int_0^{s_{\mathrm{b}}} B\,(\nu, T(s'))\;\frac{\partial \mathcal{T}\,(\nu, s')}{\partial s'}\,ds', \tag{7.1}$$

where B is the Planck function at temperature T, and I_b is the background contribution at position s_b. The monochromatic transmission $\mathcal{T}$ is given according to Beer's law by

$$\mathcal{T}\,(\nu, s) \;=\; \mathrm{e}^{-\tau(\nu,s)} \quad = \quad \exp\left(-\int_0^{s} \alpha(\nu, s')\,\mathrm{d}s'\right) \tag{7.2}$$

$$= \quad \exp\left(-\int_0^{s} \mathrm{d}s' \sum_m k_m\,(\nu, p(s'), T(s'))\;n_m(s')\;,\right) \tag{7.3}$$

where τ is the optical depth, α is the absorption coefficient, p is the atmospheric pressure, n_m is the number density of molecule m and k_m is its absorption cross section. In high resolution line–by–line models the absorption cross section is obtained by summing over the contributions from many lines,

$$k(\nu, z) \;=\; \sum_l S_l(T(z))\;g_l(\nu;\,\hat{\nu}_l, \gamma_l(p(z), T(z))) \tag{7.4}$$

where each individual line is described by the product of the temperature–dependent line strength S_l and a normalized line shape function g describing the broadening mechanism (for simplicity corrections due to continuum effects (Clough et al., 1989) are neglected here). For the infra-red and microwave the combined effect of pressure broadening (corresponding to a Lorentzian line shape) and Doppler broadening (corresponding to a Gaussian line shape) can be represented by a Voigt line profile.

To compute the spectral radiance and transmission (or equivalently, the optical depth) the integrals in equations 7.1 and 7.2 have to be discretized. Note that in general the atmospheric pressure, temperature, and concentration profiles are given only as a finite sets of data points with a typical altitude range up to about 100 km (see e.g. Anderson et al., 1986).

A common approach is to subdivide the atmosphere in a series of homogeneous layers, each described by appropriate layer mean values for pressure, temperature, and concentrations; then transmission $\mathcal{T}$ is given by the product of the layer transmissions, and the radiance can be calculated recursively (Clough et al., 1988; Edwards, 1988) ("Curtis–Godson approach"). Obviously the integrals in Equations 7.1 and 7.2 can also be evaluated easily by application of standard quadrature schemes (Kahaner et al., 1989).

7.2.1.2. Radiative transfer modelling of nadir near infra-red observations

In the near infra-red contributions from reflected sunlight become important, whereas thermal emission is negligible. Furthermore scattering can be neglected for clear sky (cloud free) observations. Thus the radiative transfer equation simplifies to

$$I(\nu) \;=\; r\;I_{\mathrm{sun}}(\nu)\;\mathcal{T}_{\uparrow}(\nu)\,\mathcal{T}_{\downarrow}(\nu) \tag{7.5}$$

where r is the reflection factor and $\mathcal{T}_{\downarrow}$ and $\mathcal{T}_{\uparrow}$ denote the transmission between reflection point (e.g. Earth surface) and observer and between sun and reflection point, respectively.

The monochromatic transmission $\mathcal{T}$ (relative to the observer) is given according to Beer's law by

$$\mathcal{T}(\nu;s) = \exp[-\int_0^s \alpha(\nu,s')\,\mathrm{d}s']\,, \tag{7.6}$$

$$\alpha(\nu;s) = \sum_m k_m(\nu,s)\,n_m(s) + \alpha^{(c)}(\nu,s) \tag{7.7}$$

where α is the volume absorption coefficient, k_m and n_m are the absorption cross section and density of molecule m, and $\alpha^{(c)}$ the continuum absorption coefficient. Note that the absorption cross section is a function of (altitude dependant) pressure and temperature, i.e. $k(\nu,z) = k\big(\nu,p(z),T(z)\big)$.

In the infra-red and microwave spectral range molecular absorption is due to radiative transitions between rotational and vibrational states of the molecules. A single spectral line is characterized by its position $\hat{\nu}$, line strength S, and line width γ, where the transition wavenumber (or frequency) is determined by the energies E_i, E_f of the initial and final state, $|i\rangle$, $|f\rangle$,

$$\hat{\nu} = \frac{1}{hc}\,(E_\mathrm{f} - E_\mathrm{i}) \tag{7.8}$$

7.2.2. Molecular Absorption

7.2.2.1. Line strength and partition functions

The monochromatic absorption cross section for a single line is defined as the product of the line strength S and a normalized line profile function g essentially determined by line broadening,

$$k(\nu;\hat{\nu},S,\gamma) = S\,\cdot\,g(\nu;\hat{\nu},\gamma) \qquad \text{with} \quad \int_{-\infty}^{+\infty} g\,\mathrm{d}\nu = 1\,. \tag{7.9}$$

For electric dipole transitions the line strength is determined by the square of the temperature dependent matrix element of the electric dipole moment and by further factors accounting for the partition function, Boltzmann-distribution, and stimulated emission,

$$S(T) = \frac{8\pi^3}{3hc}\,\frac{g_i I_a}{Q(T)}\,\hat{\nu}\,e^{-E_i/kT}[1-e^{-hc\hat{\nu}/kT}]\,R_{if}\cdot 10^{-36} \tag{7.10}$$

here g_i is the degeneracy of the nuclear spin of the lower energy state, I_a is the relative abundance of the isotope[1], $Q(T)$ is the total partition sum, R_{if} is the transition probability given by the matrix element of the electric dipole operator $R_{if} = \left|\langle f|\vec{D}|i\rangle\right|^2$. A similar expression is found for the line strength of magnetic quadrupole transitions. In both cases the ratio of line strength at two different temperatures is given by

$$S(T) = S(T_0)\,\times\,\frac{Q(T_0)}{Q(T)}\,\frac{\exp\left(-E_i/kT\right)}{\exp\left(-E_i/kT_0\right)}\,\frac{1\,-\,\exp\left(-hc\hat{\nu}/kT\right)}{1\,-\,\exp\left(-hc\hat{\nu}/kT_0\right)}\,. \tag{7.11}$$

$Q(T)$ is the product of the rotational and vibrational partition functions, $Q = Q_\mathrm{rot}\cdot Q_\mathrm{vib}$, whose temperature dependence are calculated from

$$Q_\mathrm{rot}(T) = Q_\mathrm{rot}(T_0)\left(\frac{T}{T_0}\right)^\beta\,, \tag{7.12}$$

$$Q_\mathrm{vib}(T) = \prod_{i=1}^{N}[1-\exp(-hc\omega_i/kT)]^{-d_i}\,, \tag{7.13}$$

where β is the temperature coefficient of the rotational partition function, and N is the number of vibrational

[1] In the HITRAN– and GEISA databases the abundances of the Earth atmosphere are used.

modes with wavenumbers ω_i and degeneracies d_i. Data required to calculate the vibrational partition sums have been taken from Norton and Rinsland (1991).

7.2.2.2. Line broadening

Even with an ideal spectrometer with full spectral resolution no infinite narrow "delta function" absorption line could be observed. However, the natural line width which results from the finite lifetime of the excited state and is described by a Lorentz profile, can be neglected in atmospheric spectroscopy.

Pressure (collision) broadening - Lorentz profile In case of pure pressure broadening the cross section for a single radiative transition is essentially given by a Lorentzian line profile

$$g_{\mathrm{L}}(\nu) \;=\; \frac{\gamma_{\mathrm{L}}/\pi}{(\nu-\hat{\nu})^2+\gamma_{\mathrm{L}}^2}\;. \tag{7.14}$$

The Lorentz half width (at half maximum, HWHM) γ_{L} is proportional to pressure p and decreases with increasing temperature. In case of a gas mixture with total pressure p and partial pressure p_s of the absorber molecule the total width is given by the sum of a self broadening contribution due to collisions between the absorber molecules and an air-broadening contribution due to collisions with other molecules,

$$\gamma_{\mathrm{L}}(p,p_s,T) \;=\; \left(\gamma_{\mathrm{L}}^{(0,\mathrm{air})}\;\frac{p-p_{\mathrm{S}}}{p_0}\;+\;\gamma_{\mathrm{L}}^{(0,\mathrm{self})}\;\frac{p_{\mathrm{S}}}{p_0}\right)\;\times\;\left(\frac{T_0}{T}\right)^n\;. \tag{7.15}$$

The exponent n specifying the dependence of temperature is so far known for only a few transitions of the most important molecules. The kinetic theory of gases (collision of hard spheres) yields the classical value $n=\frac{1}{2}$. The self-broadening coefficient $\gamma_{\mathrm{L}}^{(\mathrm{self})}$ is so far known for only a few transitions and will otherwise be set to the air-broadening coefficient $\gamma_{\mathrm{L}}^{(\mathrm{air})}$ (mostly specified for N_2 and/or O_2), i.e.

$$\gamma_{\mathrm{L}}(p,T) \;=\; \gamma_{\mathrm{L}}^{(\mathrm{air})}\;\frac{p}{p_0}\;\times\;\left(\frac{T_0}{T}\right)^n \tag{7.16}$$

Typical values of air-broadening coefficients are $\gamma_{\mathrm{L}}\approx 0.1p\,[\mathrm{cm/atm}]$(see Table 2 in Rothman et al., 1987).

Doppler broadening The thermal motion of the molecules leads to Doppler broadening of the spectral lines, which is described by a Gaussian line shape

$$g_{\mathrm{D}}(\nu) \;=\; \frac{1}{\gamma_{\mathrm{D}}}\left(\frac{\ln 2}{\pi}\right)^{1/2}\cdot\exp\left[-\ln 2\left(\frac{\nu-\hat{\nu}}{\gamma_{\mathrm{D}}}\right)^2\right]\;. \tag{7.17}$$

The half width (HWHM) is essentially determined by the line position $\hat{\nu}$, the temperature T, and the molecular mass m,

$$\gamma_{\mathrm{D}} \;=\; \hat{\nu}\,\sqrt{\frac{2\ln 2\;kT}{mc^2}}\;. \tag{7.18}$$

For a typical atmospheric molecule one finds

$$\gamma_{\mathrm{D}}\approx 6\cdot 10^{-8}\,\hat{\nu}\sqrt{T\,[K]}\qquad\text{for}\quad m\approx 36\,\mathrm{amu}.$$

Combined Pressure and Doppler broadening

$$g_V(\nu) \;=\; g_{\mathrm{L}}\otimes g_{\mathrm{D}} \tag{7.19}$$

Several empirical approximations for the half width (HWHM) of a Voigt line (defined by $g_V(\hat{\nu} \pm \gamma_V) = \frac{1}{2} g_V(\hat{\nu})$) have been developed (Olivero and Longbothum, 1977). For the approximation

$$\gamma_V = \frac{1}{2}\left(c_1\gamma_\mathrm{L} + \sqrt{c_2\gamma_\mathrm{L}^2 + 4\gamma_D^2}\right) \qquad \text{with} \quad c_1 = 1.0692,\ c_2 = 0.86639 \tag{7.20}$$

a accuracy of 0.02% has been specified, with $c_1 = c_2 = 1$ the accuracy is in the order of one percent. A comparison of Lorentzian, Doppler, and Voigt half width is given in Figure 7.4.

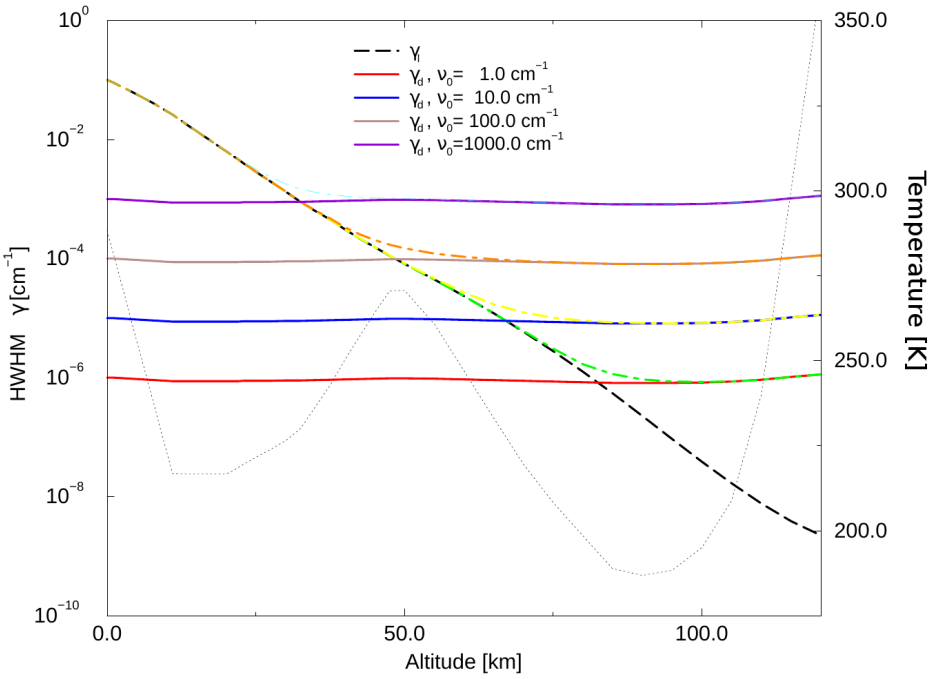

Figure 7.4.: *Half widths (HWHM) for Lorentz-, Doppler- and Voigt-Profile as a function of altitude for a variety of line positions ν. The Lorentz width is essentially proportional to pressure and hence decays approximately exponentially with altitude. In contrast the Doppler width is only weakly altitude dependent. In the troposphere lines are generally pressure broadened, the transition to the Doppler regime depends on the spectral region. The dotted line indicated atmospheric temperature. (Pressure and temperature: US Standard atmosphere, molecular mass 36 amu).*

Numerical Aspects — The Voigt function It is convenient to introduce the Voigt function $K(x, y)$ (normalized to $\sqrt{\pi}$) defined by

$$g_\mathrm{V}(\nu;\, \hat{\nu}, \gamma_\mathrm{L}, \gamma_\mathrm{D}) = \frac{\sqrt{\ln 2/\pi}}{\gamma_\mathrm{D}}\, K(x, y)\,, \tag{7.21}$$

$$K(x, y) = \frac{y}{\pi}\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-t^2}}{(x-t)^2 + y^2}\,\mathrm{d}t\,, \tag{7.22}$$

where the dimensionless variables $x,\ y$ are defined in terms of the distance from the line centre, $\nu - \hat{\nu}_0$, and the Lorentzian and Doppler half–widths $\gamma_\mathrm{L},\ \gamma_\mathrm{D}$:

$$x = \sqrt{\ln 2}\,\frac{\nu - \hat{\nu}_0}{\gamma_\mathrm{D}} \qquad \text{and} \qquad y = \sqrt{\ln 2}\,\frac{\gamma_\mathrm{L}}{\gamma_\mathrm{D}}\,. \tag{7.23}$$

The Voigt function represents the real part of the complex function $W(z)$ with $z = x + \mathrm{i}y$ that, for $y > 0$, is

identical to the complex error function (probability function) defined by Abramowitz and Stegun (1964)

$$w(z) \;=\; \mathrm{e}^{-z^2}\left(1 \;+\; \frac{2\mathrm{i}}{\sqrt{\pi}}\int_0^z \mathrm{e}^{t^2}\,\mathrm{d}t\right) \;=\; \mathrm{e}^{-z^2}\Big(1 \;-\mathrm{erf}(-\mathrm{i}z)\Big) \;=\; \mathrm{e}^{-z^2}\,\mathrm{erfc}(-\mathrm{i}z)\;. \tag{7.24}$$

The complex error function satisfies the differential equation

$$w'(z) \;=\; -2z\cdot w(z) \;+\; \frac{2\mathrm{i}}{\sqrt{\pi}} \tag{7.25}$$

and the series and asymptotic expansions

$$w(z) = \sum_{n=0}^{\infty} \frac{(\mathrm{i}z)^n}{\Gamma\left(\frac{n}{2}+1\right)} \tag{7.26}$$

$$w(z) = \frac{\mathrm{i}}{\pi}\sum_{k=0}^{\infty}\frac{\Gamma\left(k+\frac{1}{2}\right)}{z^{2k+1}} \;=\; \frac{\mathrm{i}}{\pi}\left(\frac{\sqrt{\pi}}{z}+\ldots\right) \tag{7.27}$$

For vanishing arguments x or y one has $K(0,y) = e^{y^2}\Big(1-\mathrm{erf}(y)\Big)$ and $K(x,0) = e^{-x^2}$ respectively, where erf is the error function. Truncating the asymptotic expansion of the complex error function readily shows that the wing of the Voigt profile is approximated by a Lorentzian. Further mathematical properties and relationships of the Voigt function and complex error function can be found in the extensive review by Armstrong (1967) or in Abramowitz and Stegun (1964).

Approximation of an arbitrary function by a rational function, i.e. the quotient P_M/Q_N of two polynomials of degree M and N is generally superior to polynomial approximations (Ralston and Rabinowitz, 1978).

For the computation of the Voigt function GARLIC and BIRRA uses a combination of the Humlíček (1982) and the Weideman (1994) rational approximations (Schreier, 2011).

Numerical Aspects — Computational Challenges The computational challenge of high resolution atmospheric radiative transfer modelling is due to several facts. The summation in Equation 7.4 has to include all relevant lines contributing to the spectral interval considered. In many line-by-line codes a cut-off wavenumber of $25\,\mathrm{cm}^{-1}$ from line centre is frequently employed for truncation of line wings. Note that the widely used HITRAN and GEISA spectroscopic databases (Rothman et al., 2003; Jacquinet-Husson et al., 2008) list more than a million lines of about 40 molecules in the microwave, infra-red, to ultraviolet regime, whereas the JPL spectral line catalogue (Pickett et al., 1998) covering the submillimetre, millimetre, and microwave only has almost 2 million entries.

Furthermore the wavenumber grid has to be set in accordance with the line widths γ, i.e. the grid spacing is typically chosen in the order of $\delta\nu \approx \gamma/4$. Typical line widths due to pressure broadening are in the order of $\gamma(p) \approx (p/p_0)\,0.1\mathrm{cm}$ with $p_0 = 1013\,\mathrm{mb}$. In the atmosphere the pressure decays approximately exponentially with altitude z, and the line width decreases accordingly until Doppler broadening (proportional to line position and the square root of the temperature over molecular mass ratio) becomes dominant (cf. Figure 7.4). For a spectral interval of width $\Delta\nu = 10\mathrm{cm}$ in the region of the CO_2 ν_2 band around 500 cm the number of spectral grid points is in the order of 10^5.

A variety of approaches has been developed to speed–up the calculation and an essential difference between different line-by-line codes is the choice of the line profile approximation, wavenumber grid, and interpolation. Some of the algorithms are specifically designed for the individual functions to be calculated, e.g. the (Clough and Kneizys, 1979) algorithm used in FASCODE (Clough et al., 1988): The Lorentzian (or Voigt function) is decomposed using three or four even quartic functions, each of them is then calculated on its individual grid (a similar technique using quadratic functions has been developed by Uchiyama, 1992). GENLN2 (Edwards, 1988) performs the line–by–line calculation in two stages, i.e., the entire spectral interval of interest is first split in a sequence of “wide meshes”; contributions of lines with their centre in the current wide mesh interval are computed on a fine mesh, and the contribution of other lines is computed on the wide mesh. Fomin (1995) defines a series of grids and evaluates line wing segments of larger distance to the line centre on increasingly

coarse grids. Sparks (1997) also uses a series of grids with $2^k + 1$ grid points ($k = 1, 2, \ldots$, where the coarsest grid with 3 points spans the entire region) and uses a function decomposition similar to ours.

7.2.3. Geometry

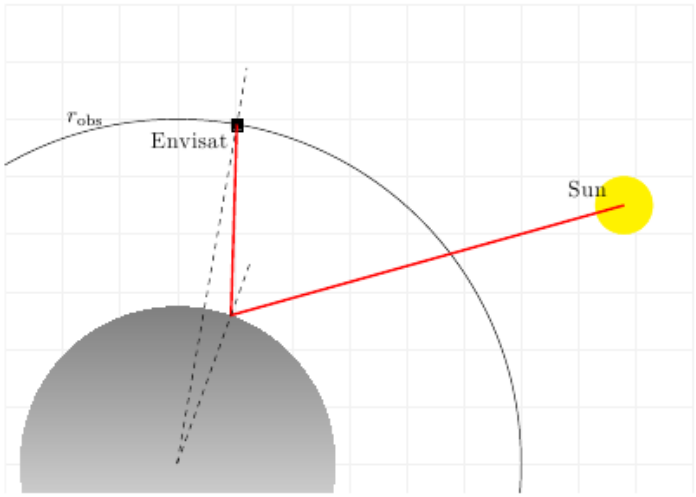

Figure 7.5.: *Observation geometry for SCIAMACHY nadir.*

The line–of–sight (LoS) in an one-dimensional spherical atmosphere is usually defined by the observer position (altitude) and the viewing angle (zenith angle), with $\alpha = 0$ for a vertical uplooking, and $\alpha = 180°$ for a vertical donwlooking (nadir) path.

For calculation of the path geometry it is more appropriate to use radii instead of altitudes, e.g.

$$r_{\mathrm{obs}} = r_{\mathrm{Earth}} + z_{\mathrm{obs}} \tag{7.28}$$

$$r_{\mathrm{end}} = r_{\mathrm{Earth}} + z_{\mathrm{end}} \tag{7.29}$$

for the observer point and the path end point. For all observation geometries the radius (altitude) of the LoS tangent point to the observer is given by

$$r_{\mathrm{t}} = r_{\mathrm{obs}} \cdot \sin\alpha \tag{7.30}$$

7.2.3.1. Uplooking

By convention, uplooking paths are characterized by zenith angles $\alpha < 90°$, see Figure 7.6. The distance of the tangent point to the observer and to the path end point (usually at top-of-atmosphere, ToA) are given by

$$t = r_{\mathrm{obs}} \cdot \cos\alpha \tag{7.31}$$

$$s = \sqrt{r_{\mathrm{end}}^2 - r_{\mathrm{t}}^2} - t \tag{7.32}$$

The angle at the path end point and the earth centred angle are

$$\beta = \arcsin\frac{r_{\mathrm{t}}}{r_{\mathrm{end}}} \tag{7.33}$$

$$\psi = \arccos\frac{r_{\mathrm{t}}}{r_{\mathrm{end}}} + \alpha - \frac{\pi}{2} \tag{7.34}$$

Alternatively the angles can be computed according to the sine theorem

$$\frac{r_{\mathrm{obs}}}{\sin\beta} = \frac{r_{\mathrm{end}}}{\sin(\pi-\alpha)} = \frac{s}{\sin\psi} \tag{7.35}$$

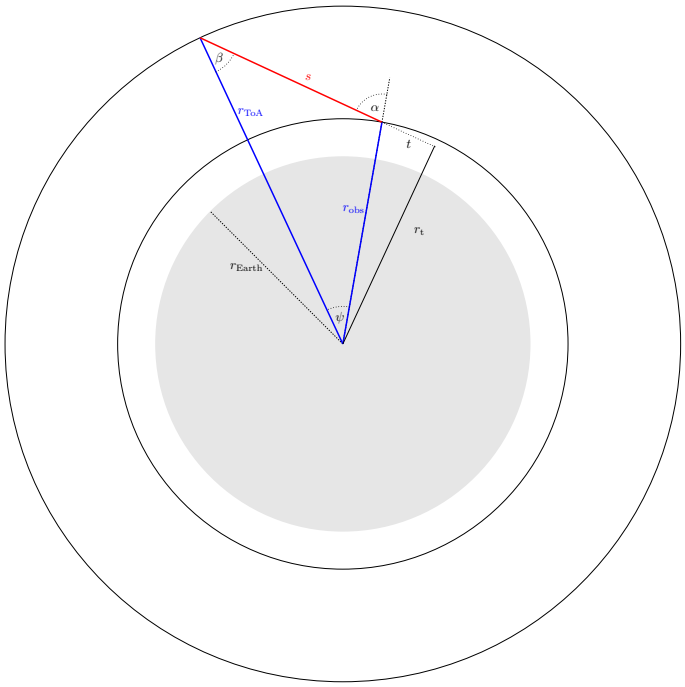

Figure 7.6.: *Geometry of uplooking path.*

7.2.3.2. Downlooking

Downlooking paths are defined by zenith angles $\alpha > 90°$ and a tangent point altitude $h_{\mathrm{t}} < 0$, see Figure 7.7. The length of the path is given by

$$s = \sqrt{r_{\mathrm{obs}}^2 - r_{\mathrm{t}}^2} - t \tag{7.36}$$

with

$$t = r_{\mathrm{obs}} \cdot \cos\bar{\alpha} \tag{7.37}$$

and angles are given by

$$\beta = \pi - \arcsin\frac{r_{\mathrm{t}}}{r_{\mathrm{end}}} \tag{7.38}$$

$$\psi = \arccos\frac{r_{\mathrm{t}}}{r_{\mathrm{obs}}} - \arccos\frac{r_{\mathrm{t}}}{r_{\mathrm{end}}} \tag{7.39}$$

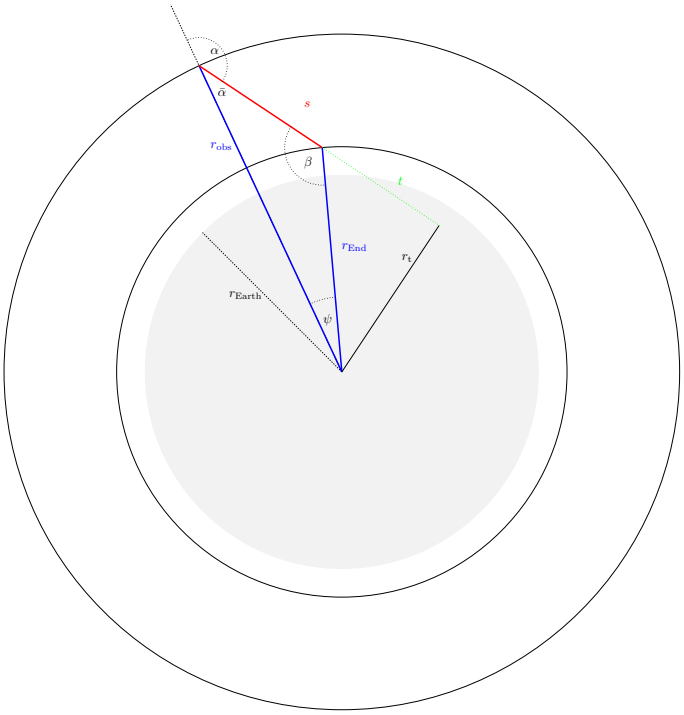

Figure 7.7.: *Geometry of downlooking path.*

7.2.3.3. Refraction

Because of refraction the line of sight is curved towards the Earth centre. For a atmospheric path defined by observer position and viewing angle (w.r.t. zenith) this results in

- ➢ a lower tangent point of the path
- ➢ an increased path length

Path refraction is characterized by an refractive index η different from 1.0 (vacuum). Assuming a layered atmosphere with constant values of pressure, temperature, densities, and hence refractive index in each layer, the path geometry can be deduced from

- ➢ Snell's law

$$\eta_l \sin \beta_l \;=\; \eta_{l+1} \sin \alpha_{l+1} \tag{7.40}$$

- ➢ Triangle geometry

$$\frac{r_l}{\sin \beta_l} \;=\; \frac{r_{l+1}}{\sin \alpha_l} \tag{7.41}$$

- ➢ The refractive index is calculated according to the recipe given for the MIPAS NRT processor (Carlotti et al., 1998)

$$\eta(n) - 1 = 0.000272632 * n/n_0 \qquad \text{with} \quad n_0 = 2.54683 \cdot 10^{19}\,\mathrm{molec/cm^3} \tag{7.42}$$

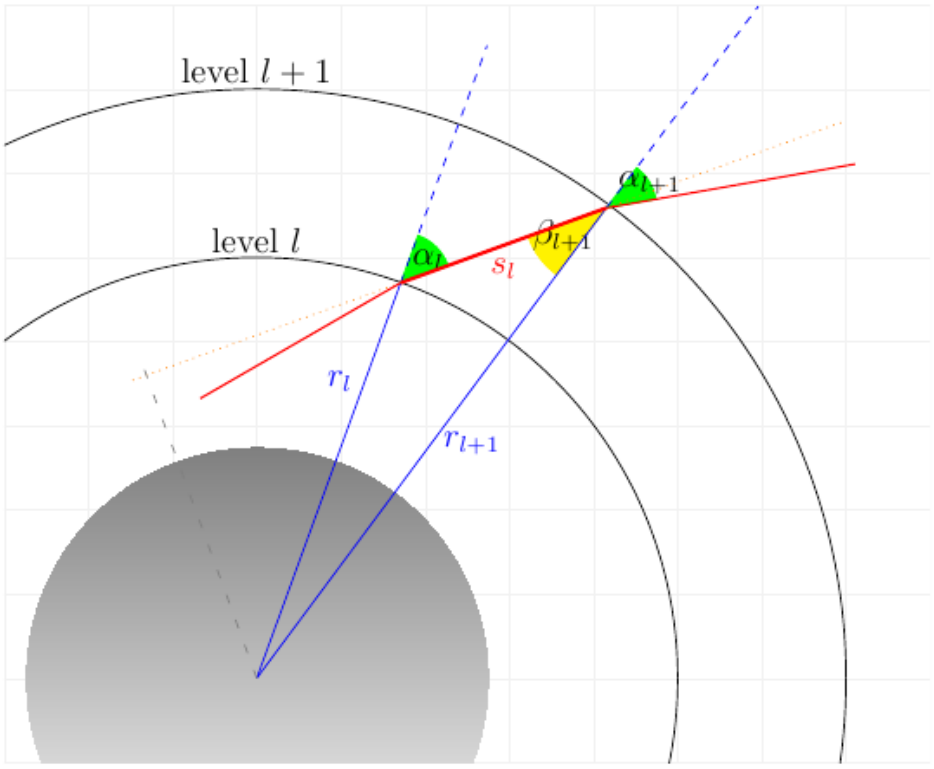

Figure 7.8.: *Refraction for an uplooking path*

For an uplooking path starting at an altitude $h_0 = r_0 - r_{\text{Earth}}$ with zenith angle α_0, the refracted path is calculated in a step-wise fashion, cf. Figure 7.8. First the path incident angle at the top of the current layer, i.e. just below next level is obtained from

$$\beta_l = \arcsin\left(\frac{r_l}{r_{l+1}} \sin \alpha_l\right)$$

The length of the path segment between altitudes h_l and h_{l+1} is given by

$$s_l = \sqrt{(r_{l+1} - r_l \sin \alpha_l) \cdot (r_{l+1} + r_l \sin \alpha_l)} - r_l \cos \alpha_l$$

Using Snell's law (Equation 7.40) the next zenith angle is computed as

$$\alpha_{l+1} = \arcsin\left(\frac{\bar{\eta}_{l,l+1}}{\bar{\eta}_{l+1,l+2}} \cdot \sin \beta_{l+1}\right)$$

where $\bar{\eta}$ is the mean refractive index of the atmospheric layer between r_l and r_{l+1}.

For SCIAMACHY nadir observations, refraction is only taken into account for the Sun — Earth surface path element, that can be modelled as an uplooking path for a ground based "observer" with viewing angle defined by the solar zenith angle (SZA), see Figure 7.5. For the "downlooking" path segment (satellite — Earth surface) with viewing angles $\leq 30°$ refraction can be neglected.

7.2.4. Instrument - Spectral Response

The instrumental response is taken into account by convolution of the monochromatic intensity spectrum (7.1) with a spectral response function $\mathcal{S}$ (a.k.a. instrumental line shape function ILS)

$$\widehat{I(\nu)} \equiv (I \otimes \mathcal{S})(\nu) = \int_{-\infty}^{\infty} I(\nu) \times \mathcal{S}(\nu - \nu')\, \mathrm{d}\nu' \tag{7.43}$$

(in general a further convolution will be required to account for the finite field of view, however, this is usually negligible for nadir viewing). For SCIAMACHY NIR measurements Gaussian, hyperbolic or Voigt profiles are commonly used,

$$\mathcal{S}_G(\nu,\gamma) = \frac{1}{\gamma}\left(\frac{\ln 2}{\pi}\right)^{1/2} \cdot \exp\left[-\ln 2\left(\frac{\nu}{\gamma_D}\right)^2\right] \tag{7.44}$$

$$\mathcal{S}_H(\nu,\gamma) = \frac{\sqrt{2}\gamma^3/\pi}{[\nu^4 + \gamma^4]} \tag{7.45}$$

$$\mathcal{S}_V(\nu,\gamma_L,\gamma_G) = \mathcal{S}_L(\nu,\gamma_L) \otimes \mathcal{S}_G(\nu,\gamma_G) \tag{7.46}$$

where the Voigt profile 7.19 is a convolution of the Gaussian with a Lorentzian profile

$$\mathcal{S}_L(\nu,\gamma) = \gamma/\left[\pi(\nu^2 + \gamma^2)\right] .$$

Note that all profiles defined here are normalized to unity, i.e. $\int_{-\infty}^{\infty} \mathcal{S}(\nu)\,\mathrm{d}\nu = 1$.

7.2.5. Input data for forward model

Modelling high resolution infra-red radiative transfer is usually done by means of line-by-line models reading molecular spectroscopic data from databases such as Hitran (Rothman et al., 2009; Gordon et al., 2017) or Geisa (Jacquinet-Husson et al., 2016), see Figure 7.11. Whereas spectroscopic data of methane and carbon monoxide have not been changed in recent versions of these databases, water spectroscopic data have been updated recently, see Figure 7.9.

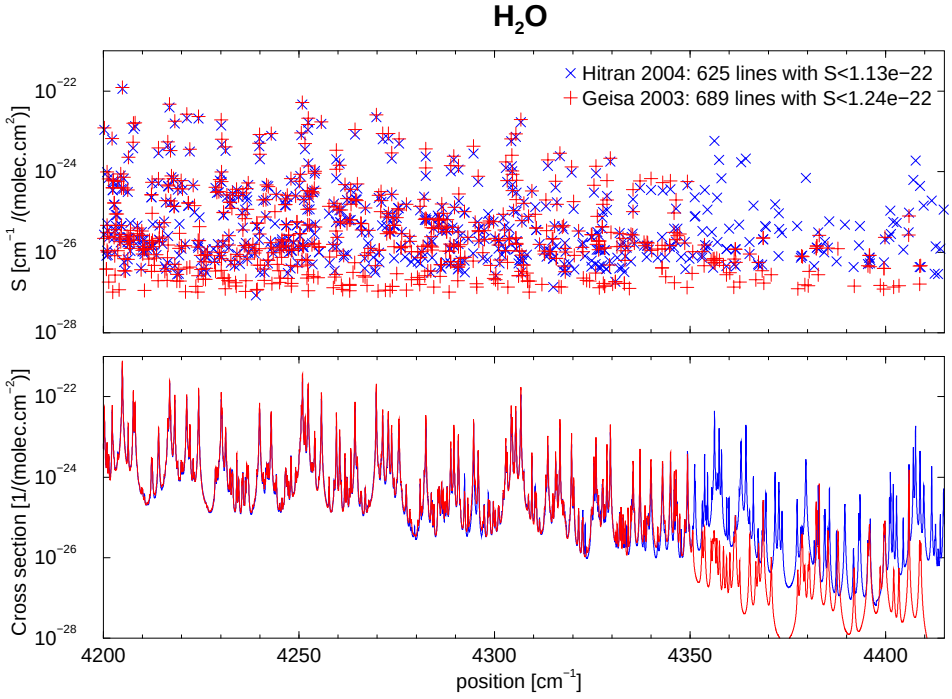

Figure 7.9.: *Water lines in SCIAMACHY channel 8 — Inter-comparison of recent Hitran 2004 and Geisa 2003 line data and corresponding cross sections. (H_2O data in Hitran 2000 are approximately identical to Geisa).*

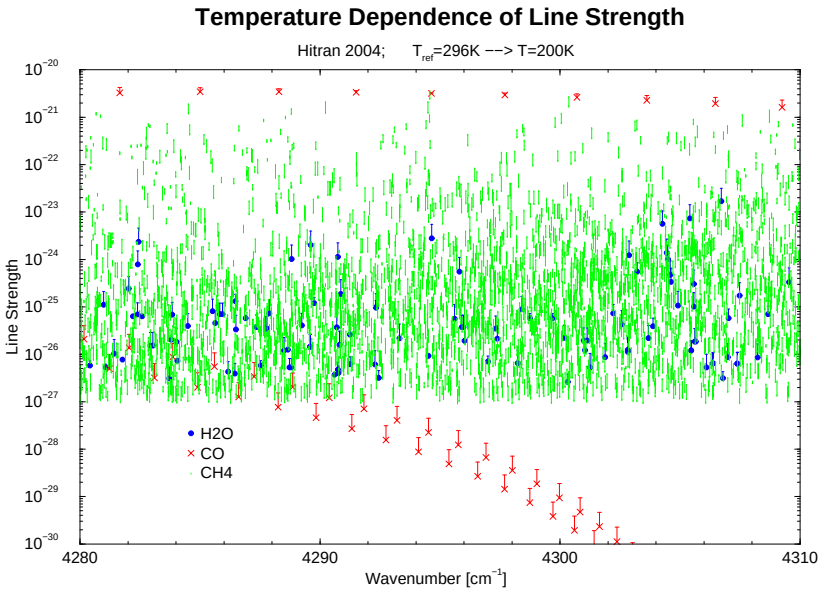

Figure 7.10.: *Temperature dependence of molecular spectroscopic lines in SCIAMACHY channel 8 (Hitran 2004 database) indicated by the length of the vertical "error" bars.*

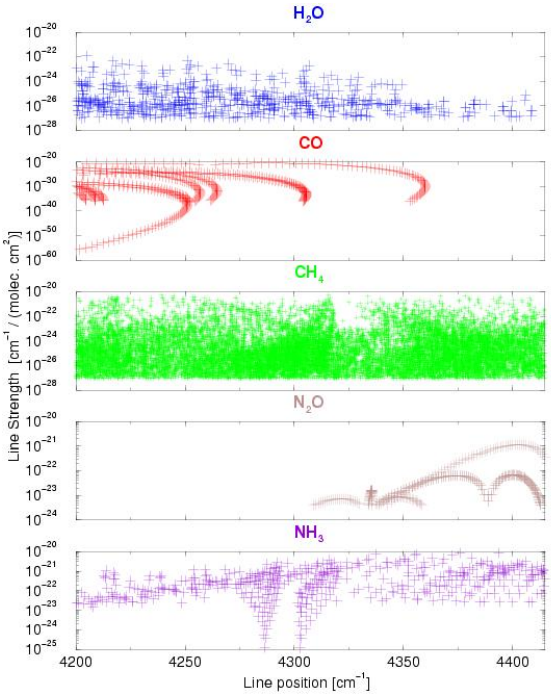

Figure 7.11.: *Molecular spectroscopic lines in SCIAMACHY channel 8 according to the Geisa database.*

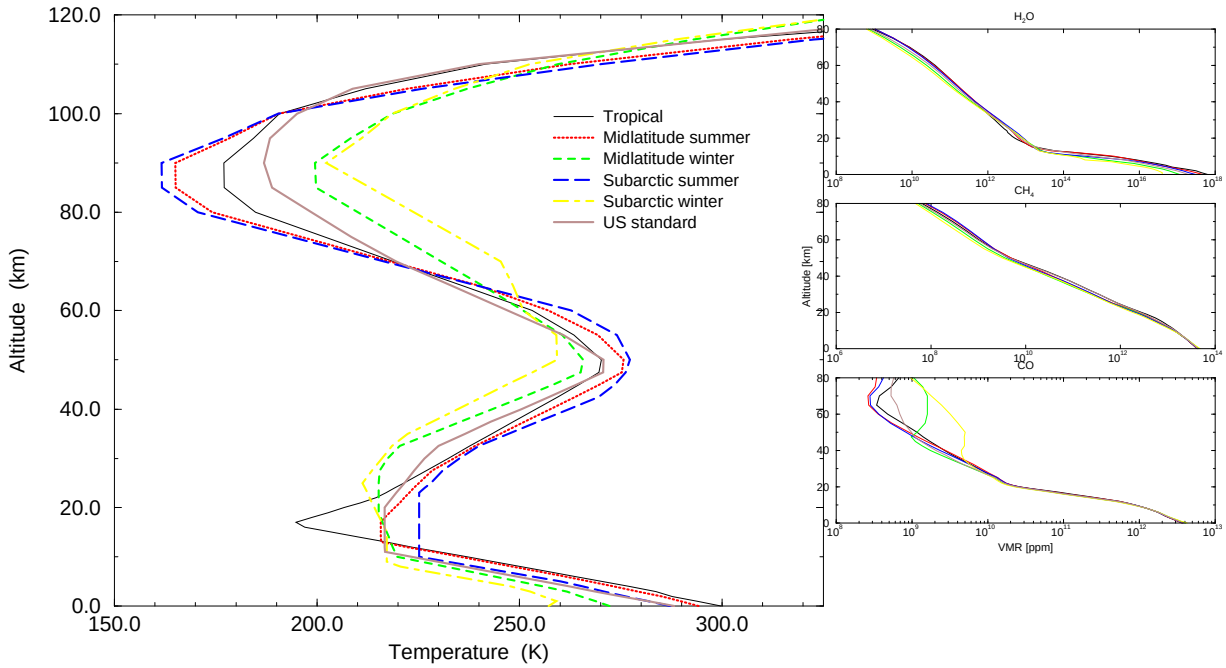

Figure 7.12.: Temperature profiles (left) and volume mixing ratios (right) listed in the AFGL model atmosphere data set.

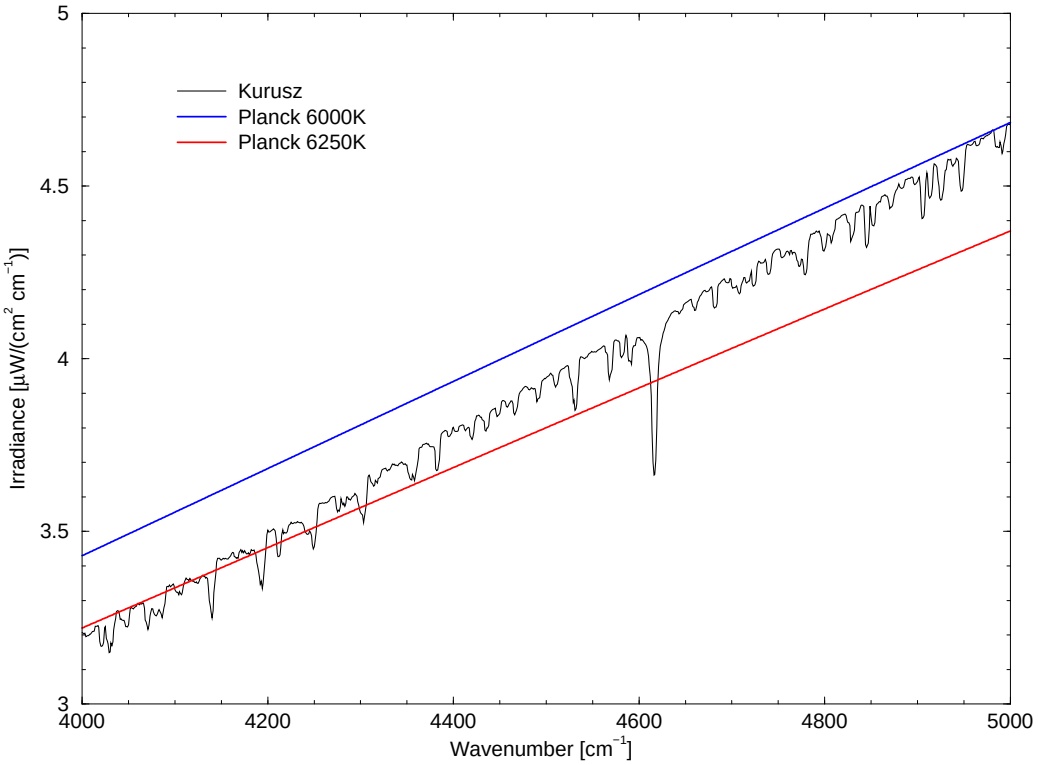

Figure 7.13.: *Solar irradiance*

7.3. Inversion

7.3.1. Non-linear least squares

The objective of SCIAMACHY near infra-red measurements in nadir viewing geometry is to retrieve information on the vertical distribution of trace gases such as N_2O, CH_4 or CO, e.g. the volume mixing ratio $q_X(z)$ or number density $n_X(z) = q_X(z) \cdot n_{\rm air}(z)$ of molecule X. The standard approach to estimate the desired quantities $\vec{x}$ from a measurement $\vec{y}$ (a vector of m components) relies on (in general non-linear) least squares fit

$$\min_x \left\| \vec{y} - \vec{F}(\vec{x}) \right\|^2 \ . \tag{7.47}$$

Here $\vec{F}$ denotes the forward model $\mathbb{R}^n \to \mathbb{R}^m$ essentially given by the radiative transfer and instrument model. Because of the ill-posed nature of vertical sounding inverse problems, it is customary to retrieve column densities

$$N_X(z_0) \equiv \int_{z_0}^{\infty} n_x(z)\,{\rm d}z \ , \tag{7.48}$$

where z_0 is the ground elevation (surface altitude).

Denoting by α_m the scale factors to be estimated and by $\bar{n}_m(z)$ the reference (e.g. climatological) densities of molecule m, the up-welling monochromatic radiance 7.5 can be written as

$$\begin{aligned} I(\nu) = r\mu_\odot\, I_{\rm sun}(\nu) \exp\left[-\int_0^\infty \frac{{\rm d}z'}{\mu} \sum_m \alpha_m \bar{n}_m(z')\, k_m(\nu, z') \right] \\ \exp\left[-\int_0^\infty \frac{{\rm d}z''}{\mu_\odot} \sum_m \alpha_m \bar{n}_m(z'')\, k_m(\nu, z'') \right] \end{aligned} \tag{7.49}$$

where for simplicity we have used the plane–parallel approximation with $\mu \equiv \cos\theta$ for an observer zenith angle θ and $\mu_\odot$ for the solar zenith angle $\theta_\odot$; furthermore continuum is neglected here. Introducing the total optical depth of molecule m (for the reference profiles and the entire path),

$$\begin{aligned} \vec{F}(\vec{x}) &\equiv \widehat{I\,(\nu)} \\ &= r\mu_\odot\, I_{\rm sun}(\nu)\, {\rm e}^{-\sum_m \alpha_m \tau_m(\nu)} \otimes \mathcal{S}(\nu, \gamma) \ + \ b \ , \end{aligned} \tag{7.50}$$

where the state vector $\vec{x}$ is comprised of geophysical and instrumental parameters $\vec{\alpha}, \gamma, r$ and an optional baseline correction b (generally a polynomial).

For the solution of the non-linear least squares problem 7.47 BIRRA uses solvers provided in the PORT Optimization Library(Dennis, Jr. et al., 1981) based on a scaled trust region strategy. BIRRA provides the option to use a non-linear least squares with simple bounds (e.g. non-negativity) to avoid non-physical results, e.g.

$$\min_{x>0} \left\| \vec{y} - \vec{F}(\vec{x}) \right\|^2 \ . \tag{7.51}$$

7.3.2. Separable non-linear least squares

Note that the surface reflectivity r and the baseline correction b enter the forward model $\vec{F} \equiv \widehat{I(\nu; \ldots)}$, 7.50, linearly and the least squares problem 7.47 can be reduced to a separable non-linear least squares problem (Golub and Pereyra, 2003). Splitting the vector $\vec{x}$ of parameters to be fitted into a vector $\vec{\alpha}$ of non-linear parameters and a vector $\vec{\beta}$ of linear parameters, i.e.

$$\vec{x} \longrightarrow (\vec{\alpha}, \vec{\beta}) \qquad \text{with} \quad \vec{x} \in \mathbb{R}^n,\ \vec{\alpha} \in \mathbb{R}^p,\ \vec{\beta} \in \mathbb{R}^q \text{ and } n = p + q, \tag{7.52}$$

the forward model can be written as

$$\vec{F}(\vec{x}) \longrightarrow \sum_{l=1}^{q} \beta_l \vec{f}_l(\vec{\alpha}) \,. \tag{7.53}$$

where $\vec{F} \in \mathbb{R}^m$ and $\vec{f}_l \in \mathbb{R}^m$ for $l = 1, \ldots, q$. Combining these functions in a matrix

$$\mathfrak{A}(\alpha) \equiv \left(\vec{f}_1(\vec{\alpha}), \vec{f}_2(\vec{\alpha}), \ldots, \vec{f}_2(\vec{\alpha})\right) \qquad \text{with} \quad \mathfrak{A} \in \mathbb{R}^{m \times q} \tag{7.54}$$

7.47 is a linear least squares problem $\min_\beta \left\| \vec{y} - \mathfrak{A}\vec{\beta} \right\|^2$ for the vector $\vec{\beta}$, that is (formally) solved by

$$\vec{\beta} \;=\; \left(\mathfrak{A}^T \mathfrak{A}\right)^{-1} \mathfrak{A}^{\mathsf{T}} \vec{y} \,. \tag{7.55}$$

Inserting this solution in 7.53, the original least squares problem 7.47 becomes

$$\min_{\alpha} \left\| \vec{y} - \sum_{l} \left(\left(\mathfrak{A}^{\mathsf{T}} \mathfrak{A}\right)^{-1} \mathfrak{A}^{\mathsf{T}} \vec{y} \right)_l \vec{f}_l(\vec{\alpha}) \right\|^2 \,. \tag{7.56}$$

This is a non-linear least squares problem for $\vec{\alpha}$ independent of $\vec{\beta}$ and can be solved in the usual way by means of Gauss–Newton or Levenberg–Marquardt algorithms. Once the optimum $\vec{\alpha}$ is found, the linear parameter vector $\vec{\beta}$ is obtained from 7.55.

The main advantages of this approach are

- ➢ The non-linear least squares solver has to iterate only for a reduced fit vector $\vec{\alpha}$
- ➢ no initial guess is required for the linear parameters $\vec{\beta}$
- ➢ The size of the Jacobian matrix is reduced

7.4. CO Retrieval

7.4.1. Detailed Description

Carbon monoxide (CO) is an important trace gas affecting air quality and climate. It is highly variable in space and time. About half of the CO comes from anthropogenic sources (e.g. fuel combustion), and further significant contributions are due to biomass burning. CO is a target species of several space-borne instruments, i.e. for AIRS, MOPITT, and TES from NASA's EOS satellite series, and MIPAS and SCIAMACHY on ESA's ENVISAT. For the retrieval of carbon monoxide the spectral interval 4282.68615 to 4302.13102 cm^{-1}, i.e. 2.3244 µm to 2.3350 µm was used. Pressure, temperature, and H2O profiles were taken from the NCEP reanalysis providing profiles four times per day with a latitude / longitude resolution of 2.5 dg (Kistler et al., 2001). An US standard atmosphere was assumed for the molecular density profiles. Surface reflectivity was modelled with a second order polynomial, baseline was ignored. Correction factors for CO, CH_4 and H_2O are simultaneously retrieved. The product contains the correction factors, the resulting columns from the multiplication of the correction factors with the starting values and a "CH_4 corrected" value of the total CO column. The underlying assumption for the latter is that CH_4 is homogeneously distributed compared to CO. The division of the correction factor for CO by the correction factor for CH_4 corrects approximately for remaining instrument effects, clouds in the FoV, etc:

$$xCO = VCD_{CO,ref} \cdot \frac{\alpha_{CO}}{\alpha_{CH4}} \tag{7.57}$$

Additionally the product also contains the directly retrieved CO column:

$$CO = VCD_{CO,ref} \cdot \alpha_{CO} \tag{7.58}$$

CO is a very weak absorber. Additionally, bad pixels and a growing ice layer in channel 8 hamper the retrieval. Thus, the most likely use case for the CO retrieved values are averages over a larger data set (e.g. monthly means). The quality flags in the products should be examined.

7.4.2. Retrival Settings

Level 1b-1c Settings	
Calibration	All calibrations except polarisation and radiometric
SMR	A0 (Sun over ASM diffuser without radiometric calibration)
Main Settings	
Fitting Interval	2324.4 - 2335.0 nm
Absorbers Fitted	CO, CH_4, H_2O
Polynomial Degree Albedo	2
Slitfunction	Gaussian
Proxy for xCO	CH_4

7.5. CH_4 Retrieval

7.5.1. Detailed Description

The CH_4 retrieval uses 2 spectral windows in channel 6. As a proxy correction to take into account the effect of clouds and transmission changes CO_2 is used, since its variations are small compared to methane. As for CO two columns can be found in the product, the total column

$$CH_4 = VCD_{CH4ref} \cdot \alpha_{CH4} \tag{7.59}$$

and the total column corrected by the CO_2 proxy

$$xCH_4 = VCD_{CH4ref} \cdot \frac{\alpha_{CH4}}{\alpha_{CO2}} \tag{7.60}$$

7.5.2. Retrieval Settings

Level 1b-1c Settings	
Calibration	All calibrations except polarisation and radiometric
SMR	A0 (Sun over ASM diffuser without radiometric calibration)
Main Settings	
Fitting Interval	1557.18 - 1594.13 nm & 1628.93 - 1670.56 nm
Absorbers Fitted	CO_2, CH_4, H_2O
Polynomial Degree Albedo	2
Slitfunction	Gaussian
Proxy for xCH_4	CO_2

Part III.

Limb Retrieval Algorithms

8. Limb Algorithm General Description

8.1. Summary of Retrieval Steps

Before we describe the mathematical background of the retrieval in detail we give a short description of the processing and relate the steps to the following sections.

Input for the retrieval are fully calibrated Level 1c data ratioed with a measured spectrum at a reference height.

Geometry of the problem: Section 8.2.2. In order to calculate the radiation field for the forward model, we first formulate the line-of-sight geometry and the coordinate system.

8.2. Forward model radiative transfer

8.2.1. General considerations

The radiative transfer equation for the diffuse radiation is:

$$\frac{\mathrm{d}I}{\mathrm{d}s}\left(\mathbf{r},\Omega\right)=-\sigma_{\mathtt{ext}}\left(\mathbf{r}\right)I\left(\mathbf{r},\Omega\right)+J\left(\mathbf{r},\Omega\right),$$

where $I\left(\mathbf{r},\Omega\right)$ is the radiance at the point $\mathbf{r}$ in the direction characterized by the unit vector Ω, ds is the differential path length in the direction Ω, $\sigma_{\mathtt{ext}}$ is the extinction coefficient and $J\left(\mathbf{r},\Omega\right)$ is the source function. In the following discussion we omit for simplicity specific reference to wavelength dependence. The direction Ω is characterized by the zenith and azimuth angles θ and ϕ, respectively, and we indicate this dependency by writing $\Omega=\Omega\left(\theta,\phi\right)$. The source function can be decomposed into the single and multiple scattering source terms

$$J\left(\mathbf{r},\Omega\right)=J_{\mathtt{ss}}\left(\mathbf{r},\Omega\right)+J_{\mathtt{ms}}\left(\mathbf{r},\Omega\right).$$

The single scattering source function contains the solar pseudo-source term and the thermal emission as determined by a Planck function $B\left(T\right)$,

$$J_{\mathtt{ss}}\left(\mathbf{r},\Omega\right)=\frac{\sigma_{\mathtt{scat}}\left(\mathbf{r}\right)}{4\pi}P\left(\mathbf{r},\Omega,\Omega_{\mathtt{sun}}\right)F_{\mathtt{sun}}\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(\left|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}\left(\mathbf{r}\right)\right|\right)}+\sigma_{\mathtt{abs}}\left(\mathbf{r}\right)B\left(T\left(\mathbf{r}\right)\right),$$

while the multiple scattering source function is given by

$$J_{\mathtt{ms}}\left(\mathbf{r},\Omega\right)=\frac{\sigma_{\mathtt{scat}}\left(\mathbf{r}\right)}{4\pi}\int_{4\pi}P\left(\mathbf{r},\Omega,\Omega'\right)I\left(\mathbf{r},\Omega'\right)\,\mathrm{d}\Omega'.$$

In the above relations, $\Omega_{\mathtt{sun}}=\Omega\left(\theta_{\mathtt{sun}},\phi_{\mathtt{sun}}\right)$ is the unit vector in the sun direction (or the solar direction),

$$\tau_{\mathtt{ext}}\left(\left|\mathbf{r}_1-\mathbf{r}_2\right|\right)=\int_{\left|\mathbf{r}_1-\mathbf{r}_2\right|}\sigma_{\mathtt{ext}}\left(\mathbf{r}'\right)\,\mathrm{d}s',$$

is the extinction optical depth between the points $\mathbf{r}_1$ and $\mathbf{r}_2$, $\mathbf{r}_{\mathtt{TOA}}\left(\mathbf{r}\right)$ is the point at the top of the atmosphere corresponding to $\mathbf{r}$, $\mathbf{r}_{\mathtt{TOA}}=\mathbf{r}-\left|\mathbf{r}_{\mathtt{TOA}}-\mathbf{r}\right|\Omega_{\mathtt{sun}}$, $F_{\mathtt{sun}}$ is the incident solar flux, $\sigma_{\mathtt{scat}}$ is the scattering coefficient and $P\left(\mathbf{r},\Omega,\Omega'\right)$ is the effective scattering phase function with Ω and Ω' being the incident and scattering directions, respectively.

In practical applications it is more convenient to introduce the function

$$\mathcal{S}\left(r,\Omega,\Omega'\right)=\sigma_{\mathtt{scat}}\left(r\right)P\left(r,\Omega,\Omega'\right)$$

and to express the single scattering contribution as

$$J_{\mathtt{ss}}\left(\mathbf{r},\Omega\right)=\frac{F_{\mathtt{sun}}}{4\pi}\mathcal{S}\left(\mathbf{r},\Omega,\Omega_{\mathtt{sun}}\right)\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(\left|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}(\mathbf{r})\right|\right)}+\sigma_{\mathtt{abs}}\left(r\right)B\left(T\left(\mathbf{r}\right)\right)$$

and the multiple scattering term as

$$J_{\mathtt{ms}}\left(\mathbf{r},\Omega\right)=\frac{1}{4\pi}\int_{4\pi}\mathcal{S}\left(\mathbf{r},\Omega,\Omega'\right)I\left(\mathbf{r},\Omega'\right)\,\mathrm{d}\Omega'.$$

Beginning at a reference point $\mathbf{r}_{\mathtt{r}}$ and performing an integration along the path $|\mathbf{r}-\mathbf{r}_{\mathtt{r}}|$, the following integral form of the radiative transfer equation can be obtained:

$$\begin{aligned}I\left(\mathbf{r},\Omega\right) &= I\left(\mathbf{r}_{\mathtt{r}},\Omega\right)\mathrm{e}^{-\tau\left(\left|\mathbf{r}-\mathbf{r}_{\mathtt{r}}\right|\right)}\\ &\quad+\int_{\left|\mathbf{r}-\mathbf{r}_{\mathtt{r}}\right|}J\left(\mathbf{r}',\Omega\right)\mathrm{e}^{-\tau\left(\left|\mathbf{r}-\mathbf{r}'\right|\right)}\mathrm{d}s'.\end{aligned}$$

At this stage of our presentation we consider a discretization of the line of sight as shown in Figure 8.1. To compute the limb radiance at the top of the atmosphere on the near side of the tangent point we use the integral form of the radiative transfer equation and derive a recurrence relation for the diffuse radiance at a set of discrete points along the line of sight. With $\{\mathbf{r}_p\}_{p=1}^{N_p}$ being the set of intersection points of the line of sight with a sequence of spherical surfaces of radii r_p, $i=1,2,...,N_p$, we apply the integral form of the radiative transfer equation to a layer p, with boundary points $\mathbf{r}_p$ and $\mathbf{r}_{p+1}$, and obtain

$$\begin{aligned}I_p\left(\Omega_{\mathtt{LOS}}\right) &= I_{p+1}\left(\Omega_{\mathtt{LOS}}\right)\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},p}\triangle_p}\\ &\quad+\int_{\left|\mathbf{r}_p-\mathbf{r}_{p+1}\right|}J\left(\mathbf{r}',\Omega_{\mathtt{LOS}}\right)\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},p}\left|\mathbf{r}_p-\mathbf{r}'\right|}\mathrm{d}s'.\end{aligned}\tag{8.1}$$

In the above relation $I_p\left(\Omega_{\mathtt{LOS}}\right)=I\left(\mathbf{r}_p,\Omega_{\mathtt{LOS}}\right)$ is the intensity at the point M_p, $\triangle_p=|\mathbf{r}_p-\mathbf{r}_{p+1}|$ is the limb/nadir path and $\bar{\sigma}_{\mathtt{ext},p}$ is the extinction coefficient on the layer p. To transform (8.1) into a computational expression we have to describe more precisely the scattering geometry, to characterize the optical properties of the atmosphere and to evaluate the source terms.

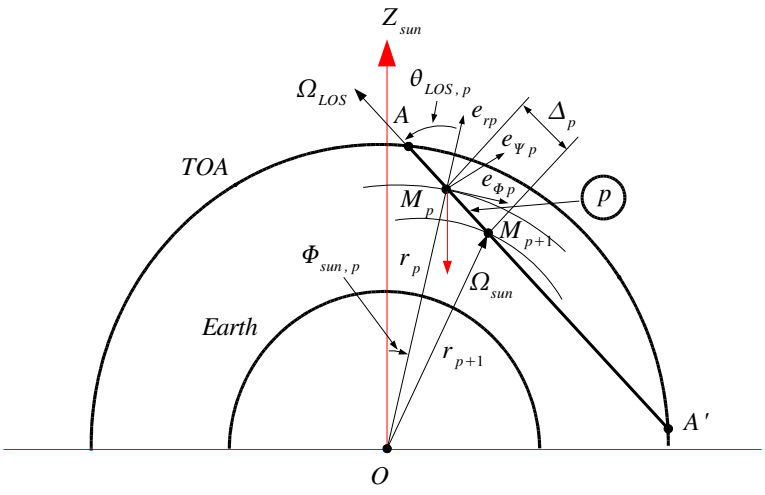

Figure 8.1.: *Integration along the line of sight*

8.2.2. Geometry of the scattering problem

In this section we derive some geometrical parameters under the assumption that the input data are

1. the zenith angle of the line of sight $\theta_{\mathtt{LOS}}$,
2. the zenith angle of the sun direction $\theta_{\mathtt{sun}}$ and
3. the relative azimuthal angle between the line of sight and the sun direction φ_0.

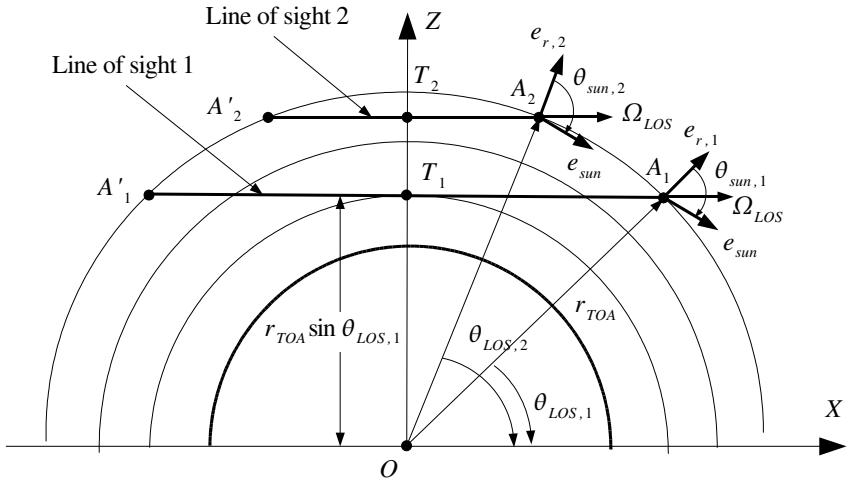

Figure 8.2.: *For a sequence of limb scans characterized by different tangent levels, the solar zenith angle and the zenith angle of the line of sight are scan dependent.*

By convention, we assume that these angles are specified at the top of the atmosphere in a coordinate system attached to the near point of the line of sights. We choose the Earth coordinate system such that the X-axis is parallel to the line of sight $\Omega_{\mathtt{LOS}}$, that is,

$$\Omega_{\mathtt{LOS}} = \mathbf{e}_{\mathtt{x}}.$$

The spherical coordinate system $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\varphi)$ correspond to the near point A on the line of sight, while the spherical coordinate system $(\mathbf{e}_r, \mathbf{e}_{\theta 1}, \mathbf{e}_{\varphi 1})$ is obtained by a plane rotation of the coordinate system $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\varphi)$ with the angle φ_0. Note that for a sequence of limb scans characterized by different tangent levels, the solar zenith angle $\theta_{\mathtt{sun}}$, the zenith and azimuthal angles of the line of sight $\theta_{\mathtt{LOS}}$ and φ_0, respectively, are scan dependent. The spherical unit vectors are given by the relations:

$$\begin{aligned}
\mathbf{e}_r &= \cos\theta_{\mathtt{LOS}}\mathbf{e}_{\mathtt{x}} + \sin\theta_{\mathtt{LOS}}\mathbf{e}_{\mathtt{z}}, \\
\mathbf{e}_\theta &= \sin\theta_{\mathtt{LOS}}\mathbf{e}_{\mathtt{x}} - \cos\theta_{\mathtt{LOS}}\mathbf{e}_{\mathtt{z}}, \\
\mathbf{e}_\varphi &= \mathbf{e}_{\mathtt{y}},
\end{aligned}$$

and

$$\mathbf{e}_{\theta 1} = \cos\varphi_0 \sin\theta_{\mathtt{LOS}}\mathbf{e}_{\mathtt{x}} + \sin\varphi_0\mathbf{e}_{\mathtt{y}} - \cos\varphi_0 \cos\theta_{\mathtt{LOS}}\mathbf{e}_{\mathtt{z}} \quad (8.2)$$

$$\mathbf{e}_{\varphi 1} = -\sin\varphi_0 \sin\theta_{\mathtt{LOS}}\mathbf{e}_{\mathtt{x}} + \cos\varphi_0\mathbf{e}_{\mathtt{y}} + \sin\varphi_0 \cos\theta_{\mathtt{LOS}}\mathbf{e}_{\mathtt{z}} \quad (8.3)$$

while the unit vector of the sun direction $\mathbf{e}_{\mathrm{sun}} = -\Omega_{\mathrm{sun}}$, read as

$$\begin{aligned}\mathbf{e}_{\mathrm{sun}} &= \cos\theta_{\mathrm{sun}}\mathbf{e}_r + \sin\theta_{\mathrm{sun}}\mathbf{e}_{\theta 1} \\ &= \left(\sin\theta_{\mathrm{sun}}\cos\varphi_0\sin\theta_{\mathrm{LOS}} + \cos\theta_{\mathrm{sun}}\cos\theta_{\mathrm{LOS}}\right)\mathbf{e}_{\mathrm{x}} + \sin\theta_{\mathrm{sun}}\sin\varphi_0\mathbf{e}_{\mathrm{y}} \\ &\quad - \left(\sin\theta_{\mathrm{sun}}\cos\varphi_0\cos\theta_{\mathrm{LOS}} - \cos\theta_{\mathrm{sun}}\sin\theta_{LOS}\right)\mathbf{e}_{\mathrm{z}} \end{aligned} \tag{8.4}$$

We define the solar coordinate system such that the Z_{sun}-axis is parallel to the sun direction $\mathbf{e}_{\mathrm{sun}}$ and the Y_{sun}-axis is parallel to the azimuthal unit vector $\mathbf{e}_{\varphi 1}$, i.e.

$$\mathbf{e}_{\mathrm{z,sun}} = \mathbf{e}_{\mathrm{sun}}, \quad \mathbf{e}_{\mathrm{y,sun}} = \mathbf{e}_{\varphi 1}, \quad \mathbf{e}_{\mathrm{x,sun}} = \mathbf{e}_{\mathrm{y,sun}} \times \mathbf{e}_{\mathrm{z,sun}}.$$

In view of (8.3) and (8.4) it is apparent that the Cartesian unit vectors $\mathbf{e}_{\mathrm{x,sun}}$, $\mathbf{e}_{\mathrm{y,sun}}$ and $\mathbf{e}_{\mathrm{z,sun}}$ can be expressed in terms of the Cartesian unit vectors $\mathbf{e}_{\mathrm{x}}$, $\mathbf{e}_{\mathrm{y}}$ and $\mathbf{e}_{\mathrm{z}}$. The unit vector of the line of sight Ω_{LOS} can be computed in the solar coordinate system as

$$\Omega_{\mathrm{LOS}} = \mathbf{e}_{\mathrm{x}} = n_{\mathrm{x,sun}}\mathbf{e}_{\mathrm{x,sun}} + n_{\mathrm{y,sun}}\mathbf{e}_{\mathrm{y,sun}} + n_{\mathrm{z,sun}}\mathbf{e}_{\mathrm{z,sun}} \tag{8.5}$$

with the coordinates being given by

$$\begin{aligned} n_{\mathrm{x,sun}} &= \mathbf{e}_{\mathrm{x}} \cdot \mathbf{e}_{\mathrm{x,sun}} = -\sin\theta_{\mathrm{sun}}\cos\theta_{\mathrm{LOS}} + \cos\theta_{\mathrm{sun}}\sin\theta_{\mathrm{LOS}}\cos\varphi_0 \\ n_{\mathrm{y,sun}} &= \mathbf{e}_{\mathrm{x}} \cdot \mathbf{e}_{\mathrm{y,sun}} = \sin\varphi_0\sin\theta_{\mathrm{LOS}} \\ n_{\mathrm{z,sun}} &= \mathbf{e}_{\mathrm{x}} \cdot \mathbf{e}_{\mathrm{z,sun}} = \sin\theta_{\mathrm{sun}}\sin\theta_{\mathrm{LOS}}\cos\varphi_0 + \cos\theta_{\mathrm{sun}}\cos\theta_{\mathrm{LOS}} \end{aligned}$$

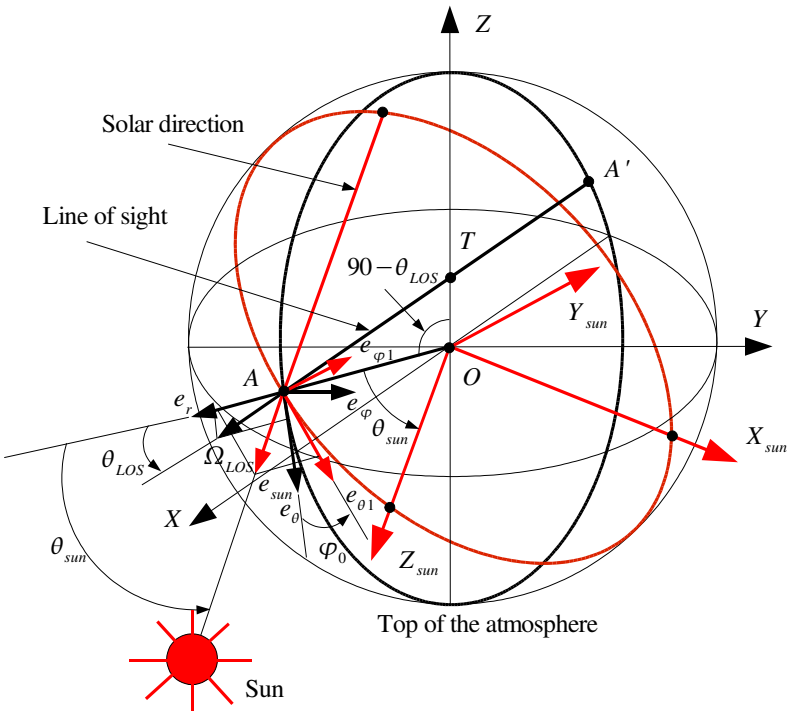

Figure 8.3.: *The Earth coordinate system is such that the X-axis is parallel to the line of sight Ω_{LOS}. The solar coordinate system is such that the Z_{sun}-axis is parallel to the sun direction $\mathbf{e}_{\mathrm{sun}} = -\Omega_{\mathrm{sun}}$ and the Y_{sun}-axis is parallel to the azimuthal unit vector $\mathbf{e}_{\varphi 1}$. The spherical coordinate systems $(\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\varphi)$ and $(\mathbf{e}_r, \mathbf{e}_{\theta 1}, \mathbf{e}_{\varphi 1})$ are rotated by φ_0.*

Let M_p be a point on the line of sight situated at the distance s_p from the near point A. The Cartesian

coordinates of the near point A in the solar coordinate system are given by

$$x_{A,\mathrm{sun}} = -r_1 \sin\theta_{\mathrm{sun}}, \quad y_{A,\mathrm{sun}} = 0, \quad z_{A,\mathrm{sun}} = r_1 \cos\theta_{\mathrm{sun}}. \tag{8.6}$$

In this context we use the representation

$$\mathbf{r}_A = x_{A,\mathrm{sun}}\mathbf{e}_{\mathrm{x,sun}} + y_{A,\mathrm{sun}}\mathbf{e}_{\mathrm{y,sun}} + z_{A,\mathrm{sun}}\mathbf{e}_{\mathrm{z,sun}},$$

to express the position vector of the point M_p on the line of sight as

$$\mathbf{r}_{M_p} = \mathbf{r}_A - s_p \Omega_{\mathrm{LOS}}. \tag{8.7}$$

In view of (8.5) - (8.7) it is apparent that we can compute the Cartesian coordinates of the point M_p in the solar coordinate system: $x_{M_p,\mathrm{sun}}$, $y_{M_p,\mathrm{sun}}$ and $z_{M_p,\mathrm{sun}}$, while a standard transformation routine enable us to calculate the polar coordinates $\left(r_{\mathrm{lev}(p)}, \Phi_{\mathrm{sun},p}, \Psi_{\mathrm{sun},p}\right)$ from the Cartesian coordinates $\left(x_{M_p,\mathrm{sun}}, y_{M_p,\mathrm{sun}}, z_{M_p,\mathrm{sun}}\right)$. The significance of $r_{\mathrm{lev}(p)}$ will be clarified latter.

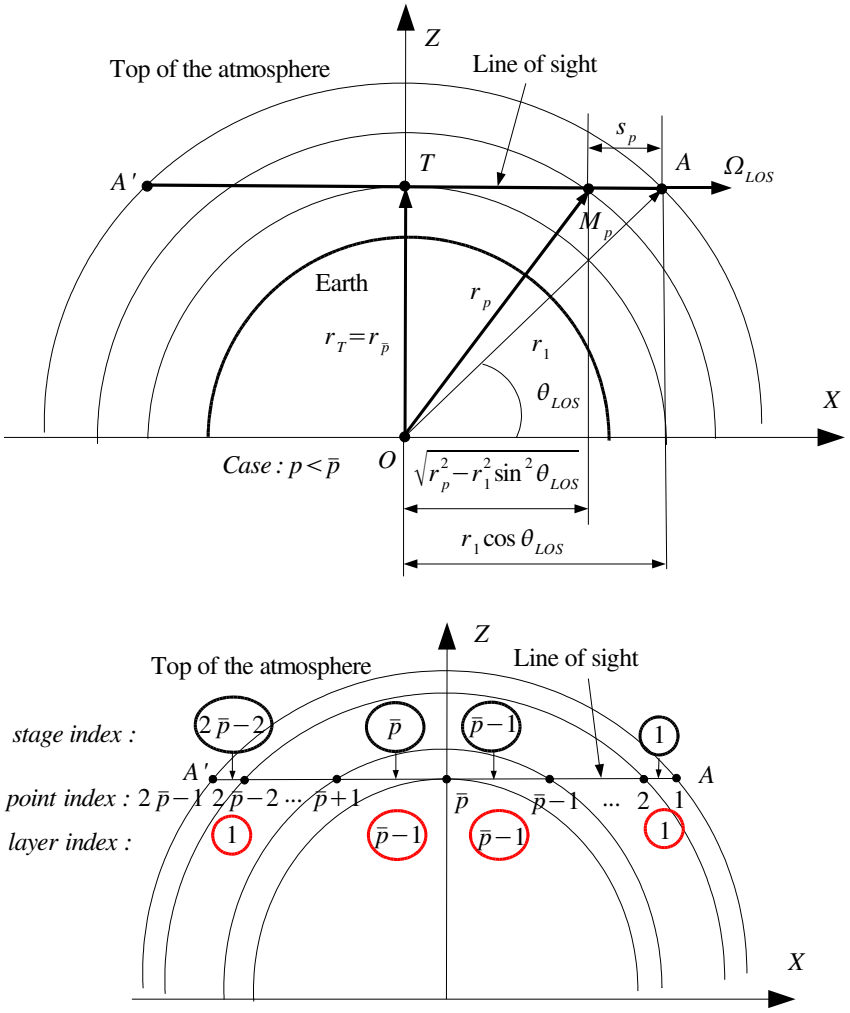

Figure 8.4.: *Limb viewing geometry. Top: The point M_p is situated at a distance s_p from the point A. The depicted situation corresponds to the case $p < \bar{p}$, where $\bar{p}$ is the tangent level index. Bottom: The points on the line of sight are $1, ..., 2\bar{p} - 1$, the stages on the line of sight are $1, ..., 2\bar{p} - 2$ and the traversed limb layers are $1, ..., \bar{p} - 1$.*

In the radiative transfer calculation we are interested in the calculation of the spherical coordinates $\theta_{\mathtt{LOS},p}$ and $\varphi_{\mathtt{LOS},p}$ of the line of sight vector $\Omega_{\mathtt{LOS}}$ in the spherical coordinate system attached to the point M_p. Using the representation (8.5), we have

$$\begin{aligned}\Omega_{\mathtt{LOS}} &\equiv n_{\mathtt{x,sun}}\mathbf{e}_{\mathtt{x,sun}} + n_{\mathtt{y,sun}}\mathbf{e}_{\mathtt{y,sun}} + n_{\mathtt{z,sun}}\mathbf{e}_{\mathtt{z,sun}} \\ &= m_{\mathtt{LOS},r_p}\mathbf{e}_{r_p} + m_{\mathtt{LOS},\Phi_p}\mathbf{e}_{\Phi_p} + m_{\mathtt{LOS},\Psi_p}\mathbf{e}_{\Psi_p},\end{aligned}$$

which gives

$$\begin{aligned}m_{\mathtt{LOS},r_p} &= -\sin\Psi_{\mathtt{sun},p}n_{\mathtt{x,sun}} + \cos\Psi_{\mathtt{sun},p}n_{\mathtt{y,sun}} \\ m_{\mathtt{LOS},\Phi_p} &= \cos\Phi_{\mathtt{sun},p}\cos\Psi_{\mathtt{sun},p}n_{\mathtt{x,sun}} + \cos\Phi_{\mathtt{sun},p}\sin\Psi_{\mathtt{sun},p}n_{\mathtt{y,sun}} - \sin\Phi_{\mathtt{sun},p}n_{\mathtt{z,sun}} \\ m_{\mathtt{LOS},\Psi_p} &= \sin\Phi_{\mathtt{sun},p}\cos\Psi_{\mathtt{sun},p}n_{\mathtt{x,sun}} + \sin\Phi_{\mathtt{sun},p}\sin\Psi_{\mathtt{sun},p}n_{\mathtt{y,sun}} + \cos\Phi_{\mathtt{sun},p}n_{\mathtt{z,sun}}\end{aligned}$$

Finally we pass from the Cartesian coordinates $\left(m_{\mathtt{LOS},r_p}, m_{\mathtt{LOS},\Phi_p}, m_{\mathtt{LOS},\Psi_p}\right)$ to the spherical coordinates $(1, \theta_{\mathtt{LOS},p}, \varphi_{\mathtt{LOS},p})$ by using a standard transformation routine. The cosine of the scattering angle between the line of sight and the sun direction is given by

$$\cos\Theta_{\mathtt{sun}} = \Omega_{\mathtt{sun}} \cdot \mathbf{e}_{\mathtt{LOS}} = -\mathbf{e}_{\mathtt{sun}} \cdot \mathbf{e}_{\mathtt{LOS}} = \sin\Phi_{\mathtt{sun},p}m_{\mathtt{LOS},\Phi_p} - \cos\Phi_{\mathtt{sun},p}m_{\mathtt{LOS},r_p}.$$

Obviously, $\cos\Theta_{\mathtt{sun}}$ should have the same values at all points on the line of sight and this property should be used as an internal check of the code.

To define some parameters of calculation we first introduce the tangent level $\bar{p}$ as that level for which it holds that $r_{\bar{p}} = r_{\mathtt{tg}}$, where $r_{\mathtt{tg}} = r_{\mathtt{earth}} + h_{\mathtt{tg}}$, with $h_{\mathtt{tg}}$ being the tangent altitude. In this context, we define

1. number of points on the limb path, $N_{\mathtt{LOS}}^{\mathtt{point}} = 2\bar{p} - 1$;
2. number of stages on the limb path, $N_{\mathtt{LOS}}^{\mathtt{stage}} = N_{\mathtt{LOS}}^{\mathtt{point}} - 1 = 2\bar{p} - 2$;
3. number of levels above and including the limb path $N_{\mathtt{LOS}}^{\mathtt{level}} = \bar{p}$.
4. number of layers above the limb path, $N_{\mathtt{LOS}}^{\mathtt{layer}} = N_{\mathtt{LOS}}^{\mathtt{level}} - 1 = \bar{p} - 1$;
5. the limb paths computed in the earth coordinate system at all stages,

   $$\triangle_p = s_p - s_{p+1}, \;\; p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{stage}},$$

   where the distances from a generic point M_p on the limb path to the near point A are computed by

   $$\begin{aligned}s_1 &= 0 \quad , \quad s_{2\bar{p}-1} = 2r_1\cos\theta_{\mathtt{LOS}}, \\ s_p &= r_1\cos\theta_{\mathtt{LOS}} - \sqrt{r_p^2 - r_1^2\sin^2\theta_{\mathtt{LOS}}}, \;\; p = 2, ..., \bar{p} - 1, \\ s_{2\bar{p}-p} &= r_1\cos\theta_{\mathtt{LOS}} + \sqrt{r_p^2 - r_1^2\sin^2\theta_{\mathtt{LOS}}}, \;\; p = 2, ..., \bar{p} - 1, \\ s_{\bar{p}} &= r_1\cos\theta_{\mathtt{LOS}};\end{aligned}$$

6. the limb zenith angles of the point M_p, $\theta_{\mathtt{LOS},p}$, $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{point}}$;
7. the limb azimuthal angles of the point M_p, $\varphi_{\mathtt{LOS},p}$, $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{point}}$;
8. the solar zenith angles of the point M_p, $\Phi_{\mathtt{sun},p}$, $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{point}}$;
9. the cosine of the solar scattering angle $\cos\Theta_{\mathtt{sun}}$.

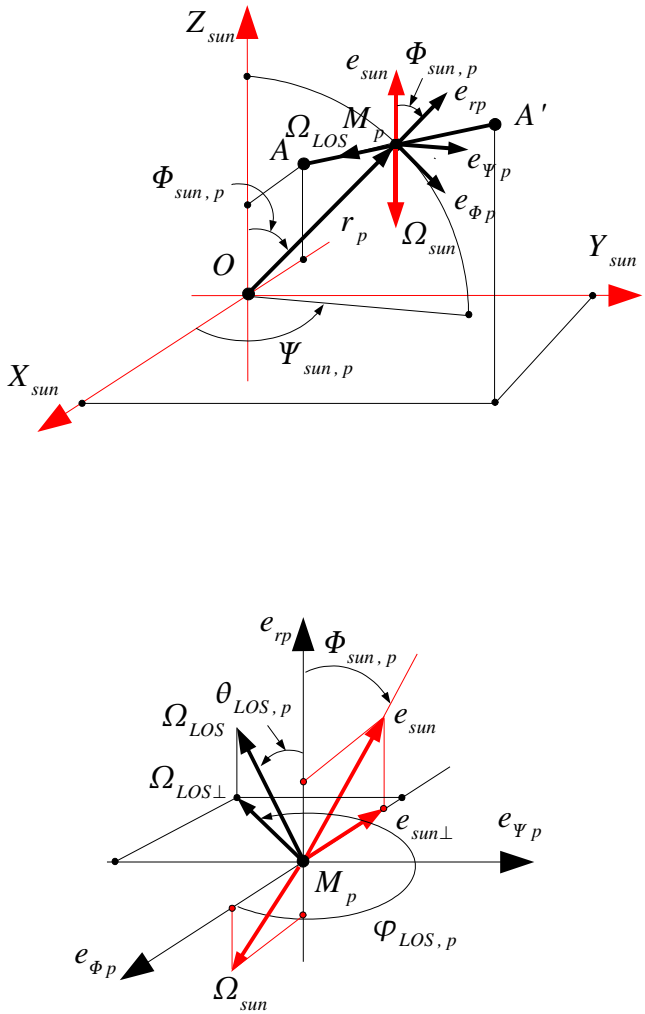

Figure 8.5.: *Top: The spherical coordinates of the generic point M_p in the solar coordinate system $\left(r_{\mathtt{lev}(p)}, \Phi_{\mathtt{sun},p}, \Psi_{\mathtt{sun},p}\right)$. Note that the near point A is situated in the $X_{\mathtt{sun}}Z_{\mathtt{sun}}$-plane. Bottom: Spherical angles $\theta_{\mathtt{LOS},p}$ and $\varphi_{\mathtt{LOS},p}$ of $\Omega_{\mathtt{LOS}}$ in the spherical coordinate system attached to M_p.*

Due to the peculiarities of the limb scattering geometry, two important maps can be defined:

1. the map $\mathrm{lev}\,(p)$ which map the point index p, $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{point}}$, into the level index

$$\mathrm{lev}\,(p) = \begin{cases} p, & 1 \le p \le \bar{p} \\ 2\bar{p} - p, & \bar{p} \le p \le 2\bar{p} - 1 \end{cases};$$

2. the map $\mathrm{lay}\,(p)$ which map the stage index p, $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{stage}}$, into the layer index

$$\mathrm{lay}\,(p) = \begin{cases} p, & 1 \le p \le \bar{p} - 1 \\ 2\bar{p} - p - 1, & \bar{p} \le p \le 2\bar{p} - 2 \end{cases}.$$

The map $\mathrm{lev}\,(p)$ will be used when computing the solar paths $\triangle_{\mathtt{sun},p,s}$ and the multiple scattering term (see below), while the map $\mathrm{lay}\,(p)$ will be used when deriving a recurrence relation for intensities and Jacobian.

When computing the single scattering term we need to evaluate the optical depth between a point M_p and the top of the atmosphere (along the sun direction)

$$\tau^{\mathtt{sun}}_{\mathtt{ext},p} = \int_{|\mathbf{r}_p - \mathbf{r}_{\mathtt{TOA}}|} \sigma_{\mathtt{ext}}\left(\mathbf{r}'\right)\, \mathrm{d}s'.$$

The computational relation is

$$\tau^{\mathtt{sun}}_{\mathtt{ext},p} = \sum_{u=1}^{N^{\mathtt{trav}}_{\mathtt{sun},p}} \bar{\sigma}_{\mathtt{ext},u} \triangle_{\mathtt{sun},p,u},$$

where $N^{\mathtt{trav}}_{\mathtt{sun},p}$ is the effective number of solar traversed layers at the point M_p and $\bar{\sigma}_{\mathtt{ext},u}$ is extinction coefficient on the layer u. We are now concern with the calculation of the solar paths $\triangle_{\mathtt{sun},p,u}$ at all points M_p on the line of sight, that is, for all $u = 1, ..., N^{\mathtt{trav}}_{\mathtt{sun},p}$ and $p = 1, ..., N^{\mathtt{point}}_{\mathtt{LOS}}$. Considering the point M_p in the solar coordinate system we are faced with 2 situations which we have to consider:

1. $\Phi_{\mathtt{sun},p} < \pi/2$. In this case no layers below M_p appears in the calculation of the optical depth and the number of solar traversed layer is
$$N^{\mathtt{trav}}_{\mathtt{sun},p} = \mathrm{lev}\left(p\right) - 1.$$

2. $\Phi_{\mathtt{sun},p} > \pi/2$. In this case additional layers appear below M_p and we compute the index t as the first index for which the inequality $r_{\mathtt{lev}(p)+t+1} < r_{\mathtt{lev}(p)} \sin \Phi_{\mathtt{sun},p}$ holds true. The number of of solar traversed layer is
$$N^{\mathtt{trav}}_{\mathtt{sun},p} = \mathrm{lev}\left(p\right) + t,$$

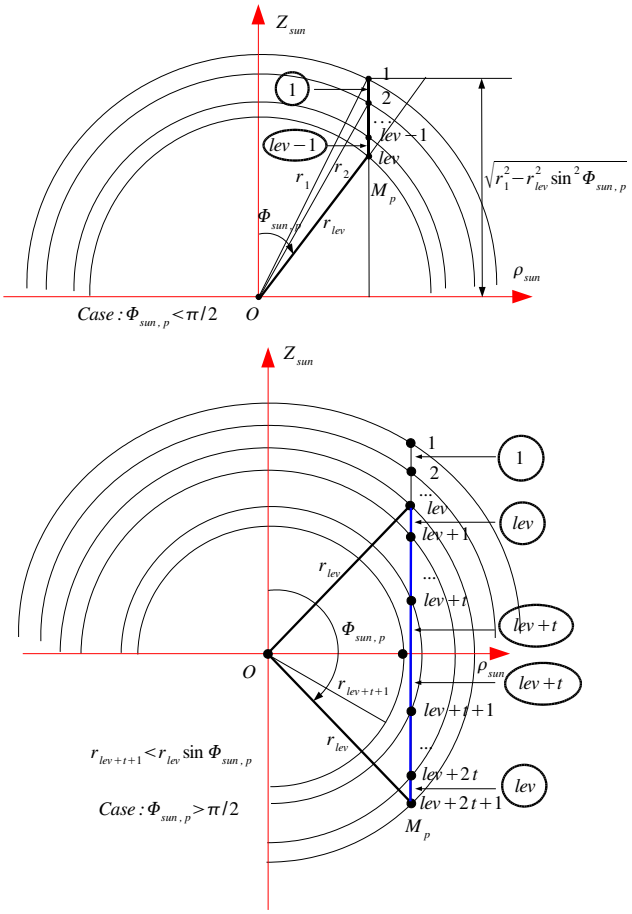

Figure 8.6.: *Solar traversed layers in the cases $\Phi_{\mathtt{sun},p} < \pi/2$ (top) and $\Phi_{\mathtt{sun},p} > \pi/2$ (bottom).*

8.2.3. Optical properties

We consider a discretization of the atmosphere in $N_{\mathtt{lev}}$ levels: $r_1 > r_2 > ... > r_{N_{\mathtt{lev}}}$ with the convention $r_1 = r_{\mathtt{TOA}}$ and $r_{N_{\mathtt{lev}}} = r_{\mathtt{s}}$. A layer j is bounded at above by the level r_j and below by the level r_{j+1}; the number of layers is given by $N_{\mathtt{lay}} = N_{\mathtt{lev}} - 1$.

The gas law for air molecules states that

$$p = nkT,$$

or equivalently that

$$p = n\frac{R}{N_A}T,$$

with $k = R/N_A$. The hydrostatic equilibrium law for air molecules is

$$\mathrm{d}p = -\frac{gpM}{RT}\mathrm{d}r,$$

whence accounting of the gas law we see that

$$\mathrm{d}p = -g\frac{M}{N_A}n\left(r\right)\mathrm{d}r.$$

In terms of finite variations, the hydrostatic equilibrium law for air molecules takes the form

$$n\left(r\right)\triangle r = \frac{N_A}{gM}\triangle p$$

and further

$$n\left(r\right)\triangle r\left[\frac{\mathrm{molec}}{\mathrm{cm}^2}\right] = 2.120156\cdot 10^{22}\cdot\triangle p\left[\mathrm{mb}\right]$$

8.2.3.1. Partial columns

The partial column of the gas g on the layer j is given by

$$\bar{X}_{g,j} = \frac{kT_0}{p_0}\int_{r_{j+1}}^{r_j} n_g\left(r\right)\,\mathrm{d}r = \frac{kT_0}{p_0}\int_{r_{j+1}}^{r_j} VMR_g\left(r\right)n\left(r\right)\,\mathrm{d}r = \frac{T_0}{p_0}\int_{r_{j+1}}^{r_j} VMR_g\left(r\right)\frac{p\left(r\right)}{T\left(r\right)}\,\mathrm{d}r, \tag{8.8}$$

where the gas law $p = nkT$ and the expression relating the number density of the gas g to the air number density n,

$$n_g = VMR_g\cdot n,$$

have been employed. The computational relation then takes the form

$$\bar{X}_{g,j}\left[10^{-3}\mathrm{cm}\right] = 2.69578\cdot 10^7\overline{\left(\frac{VMR_g[.]p\left[\mathrm{mb}\right]}{T\left[\mathrm{K}\right]}\left(r\right)\right)}_j\triangle r_j,$$

with

$$VMR_g\left[.\right] = 10^{-6}\cdot VMR_g\left[\mathrm{ppm}\right].$$

Another relation can be derived if we assume that the hydrostatic equilibrium law holds true. In this case we obtain

$$\begin{aligned}\bar{X}_{g,j} &= \frac{kT_0}{p_0}\int_{r_{j+1}}^{r_j} VMR_g\left(r\right)n\left(r\right)\,\mathrm{d}r = \frac{kT_0}{p_0}\overline{VMR}_{g,j}\int_{r_{j+1}}^{r_j} n\left(r\right)\,\mathrm{d}r \\ &= \frac{kT_0}{p_0}\overline{VMR}_{g,j}\bar{n}_j\triangle r_j = \frac{kT_0N_A}{p_0gM}\overline{VMR}_{g,j}\triangle p_j\end{aligned}$$

and the computational relation read as

$$\bar{X}_{g,j}\left[10^{-3}\mathrm{cm}\right] = 0.789087 \cdot 10^{6} \cdot \overline{VMR}_{g,j}\left[.\right] \triangle p_j \left[\mathrm{mb}\right]. \tag{8.9}$$

8.2.3.2. Number density and VMR

Because the retrieved quantities are the partial columns it is important to relate the partial columns to the number density and VMR. From the relation (cf.(8.8))

$$\bar{X}_{g,j} = \frac{kT_0}{p_0}\int_{r_{j+1}}^{r_j} n_g\left(r\right)\,\mathrm{d}r = \frac{kT_0}{p_0}\bar{n}_{g,j}\triangle r_j,$$

we get

$$\bar{n}_{g,j}\left[\frac{\mathrm{molec}}{\mathrm{cm}^3}\right] = 2.6868 \cdot 10^{11}\frac{\bar{X}_{g,j}\left[10^{-3}\mathrm{cm}\right]}{\triangle r_j\left[\mathrm{km}\right]}.$$

To compute the VMR we use again (8.8) which yields

$$\bar{X}_{g,j} = \frac{T_0}{p_0}\int_{r_{j+1}}^{r_j} VMR_g\left(r\right)\frac{p\left(r\right)}{T\left(r\right)}\,\mathrm{d}r = \frac{T_0}{p_0}\overline{VMR}_{g,j}\overline{\left(\frac{p}{T}\left(r\right)\right)}_j\triangle r_j$$

and further

$$\overline{VMR}_{g,j}\left[.\right] = \frac{\bar{X}_{g,j}\left[10^{-3}\mathrm{cm}\right]}{2.69578 \cdot 10^{7} \cdot \overline{\left(\frac{p[\mathrm{mb}]}{T[\mathrm{K}]}\left(r\right)\right)}_j\triangle r_j\left[\mathrm{km}\right]}.$$

When the hydrostatic equilibrium law is assumed to holds true, the computational relation is (8.9) and the result is,

$$\overline{VMR}_{g,j}\left[.\right] = \frac{\bar{X}_{g,j}\left[10^{-3}\mathrm{cm}\right]}{0.789087 \cdot 10^{6} \cdot \triangle p_j\left[\mathrm{mb}\right]}.$$

8.2.3.3. Scattering optical depth of air molecules

The scattering optical depth of air molecules on the layer j is defined by

$$\bar{\tau}^{\mathtt{mol}}_{\mathtt{scat},j}\left(\lambda\right) = C_{\mathtt{scat}}\left(\lambda\right)\int_{r_{j+1}}^{r_j} n\left(r\right)\,\mathrm{d}r = C_{\mathtt{scat}}\left(\lambda\right)\int_{0}^{\triangle r_j}\frac{p\left(r\right)}{kT\left(r\right)}\,dr$$

and the computational expression is

$$\bar{\tau}^{\mathtt{mol}}_{\mathtt{scat},j}\left(\lambda\right) = 0.724311 \cdot 10^{24} \cdot C_{\mathtt{scat}}\left(\lambda\right)\left[\mathrm{cm}^2\right]\overline{\left(\frac{p\left[\mathrm{mb}\right]}{T\left[\mathrm{K}\right]}\left(r\right)\right)}_j\triangle r_j\left[\mathrm{km}\right]$$

Assuming that the the hydrostatic equilibrium law holds true, we obtain

$$\bar{\tau}^{\mathtt{mol}}_{\mathtt{scat},j}\left(\lambda\right) = C_{\mathtt{scat}}\left(\lambda\right)\int_{r_{j+1}}^{r_j} n\left(r\right)\,\mathrm{d}r = C_{\mathtt{scat}}\left(\lambda\right)\bar{n}_j\triangle r_j = C_{\mathtt{scat}}\left(\lambda\right)\frac{N_A}{gM}\triangle p_j$$

and further

$$\bar{\tau}^{\mathtt{mol}}_{\mathtt{scat},j}\left(\lambda\right) = 2.120156 \cdot 10^{22} \cdot C_{\mathtt{scat}}\left(\lambda\right)\left[\mathrm{cm}^2\right]\triangle p_j\left[\mathrm{mb}\right]$$

For the scattering coefficient we get

$$\bar{\sigma}^{\mathtt{mol}}_{\mathtt{scat},j}\left(\lambda\right) = \frac{\bar{\tau}^{\mathtt{mol}}_{\mathtt{scat},j}\left(\lambda\right)}{\triangle r_j}$$

The Rayleigh scattering coefficient $C_{\mathtt{scat}}(\lambda)$ is computed by using the relation

$$C_{\mathtt{scat}}(\lambda)\left[\mathrm{cm}^2\right] = \frac{3.99\cdot 10^{-4}\,(1/\lambda)^4}{1-1.06\cdot 10^{-2}\,(1/\lambda)^2 - 6.68\cdot 10^{-5}\,(1/\lambda)^4}\cdot 10^{-24}$$

with $\lambda\,[\mu\mathrm{m}]$.

8.2.3.4. Absorption optical depth of gas molecules

The absorption optical depth of gas molecules on the layer j is a sum over all constituents

$$\bar{\tau}^{\mathtt{mol}}_{\mathtt{abs},j}(\lambda) = \sum_g \bar{\tau}^{\mathtt{mol}}_{\mathtt{abs},g,j}(\lambda)\,,$$

where the optical depth of the gas component g is given by

$$\bar{\tau}^{\mathtt{mol}}_{\mathtt{abs},g,j}(\lambda) = \bar{C}_{\mathtt{abs},g,j}(\lambda)\int_{r_{j+1}}^{r_j} n_g(r)\,\mathrm{d}r = \bar{C}_{\mathtt{abs},g,j}(\lambda)\,\frac{p_0}{kT_0}\bar{X}_{g,j}$$

or equivalently, by

$$\bar{\tau}^{\mathtt{mol}}_{\mathtt{abs},g,j}(\lambda) = 2.6868\cdot 10^{16}\cdot \bar{C}_{\mathtt{abs},g,j}(\lambda)\left[\mathrm{cm}^2\right]\bar{X}_{g,j}\left[10^{-3}\mathrm{cm}\right].$$

The absorption coefficient is then computed accordingly to the relation

$$\bar{\sigma}^{\mathtt{mol}}_{\mathtt{abs},j}(\lambda) = \frac{\bar{\tau}^{\mathtt{mol}}_{\mathtt{abs},j}(\lambda)}{\triangle r_j}.$$

8.2.3.5. Scattering optical depth of aerosols

The scattering optical depth of aerosols on the layer j depends on the aerosol optical thickness $t_{\mathtt{aer},j}$ and the wavelength λ,

$$\bar{\tau}^{\mathtt{aer}}_{\mathtt{scat},j}(\lambda) = t_{\mathtt{aer},j}\left[\bar{a}^{\mathtt{aer}}_{3,j} + (\lambda-\lambda_{\mathtt{ref}})\,\bar{a}^{\mathtt{aer}}_{4,j}\right],$$

For the extinction optical depth we have a similar relation

$$\bar{\tau}^{\mathtt{aer}}_{\mathtt{ext},j}(\lambda) = t_{\mathtt{aer},j}\left[\bar{a}^{\mathtt{aer}}_{1,j} + (\lambda-\lambda_{\mathtt{ref}})\,\bar{a}^{\mathtt{aer}}_{2,j}\right].$$

The scattering and extinction coefficients are then given by

$$\bar{\sigma}^{\mathtt{aer}}_{\mathtt{scat},j}(\lambda) = \frac{\bar{\tau}^{\mathtt{aer}}_{\mathtt{scat},j}(\lambda)}{\triangle r_j}$$

and

$$\bar{\sigma}^{\mathtt{aer}}_{\mathtt{ext},j}(\lambda) = \frac{\bar{\tau}^{\mathtt{aer}}_{\mathtt{ext},j}(\lambda)}{\triangle r_j},$$

respectively.

8.2.3.6. Total extinction optical depth

Since

$$\begin{aligned}\text{extinction} \;=\; &\text{scattering by air molecules} + \text{absorption by gas molecules}\\ &+ \text{extinction by aerosols}\end{aligned}$$

the total extinction optical depth on the layer j is computed as

$$\bar{\tau}_{\mathtt{ext},j}(\lambda) = \left[\bar{\tau}^{\mathtt{mol}}_{\mathtt{scat},j}(\lambda) + \bar{\tau}^{\mathtt{mol}}_{\mathtt{abs},j}(\lambda)\right] + \bar{\tau}^{\mathtt{aer}}_{\mathtt{ext},j}(\lambda),$$

while the extinction coefficient is given by

$$\bar{\sigma}_{\mathtt{ext},j}(\lambda) = \frac{\bar{\tau}_{\mathtt{ext},j}(\lambda)}{\triangle r_j}.$$

Because only the absorption of gas molecules depends on partial columns, the partial derivative of the extinction and absorption coefficients with respect to the partial columns of the gas g are given by

$$\frac{\partial \bar{\sigma}_{\mathtt{ext},j}}{\partial \bar{X}_{g,j}}(\lambda) = \frac{\partial \bar{\sigma}_{\mathtt{abs},j}}{\partial \bar{X}_{g,j}}(\lambda) = \frac{\partial \bar{\sigma}^{\mathtt{mol}}_{\mathtt{abs},j}}{\partial \bar{X}_{g,j}}(\lambda) = 2.6868 \cdot 10^{16} \cdot \frac{1}{\triangle r_j} \bar{C}_{\mathtt{abs},g,j}(\lambda)$$

while the partial derivative of the extinction and scattering coefficients with respect to the aerosol optical thickness read as

$$\frac{\partial \bar{\sigma}_{\mathtt{ext},j}}{\partial t_{\mathtt{aer},j}}(\lambda) = \frac{1}{\triangle r_j}\left[\bar{a}^{\mathtt{aer}}_{1,j} + (\lambda - \lambda_{\mathtt{ref}})\,\bar{a}^{\mathtt{aer}}_{2,j}\right]$$

and

$$\frac{\partial \bar{\sigma}_{\mathtt{scat},j}}{\partial t_{\mathtt{aer},j}}(\lambda) = \frac{1}{\triangle r_j}\left[\bar{a}^{\mathtt{aer}}_{3,j} + (\lambda - \lambda_{\mathtt{ref}})\,\bar{a}^{\mathtt{aer}}_{4,j}\right]$$

8.2.3.7. Phase functions

For a scattering atmosphere, the effective scattering phase function accounts of Rayleigh scattering by air molecules (molecular scattering) P and Mie scattering by aerosols (particle scattering),

$$\begin{aligned}\mathcal{S}(r,\Omega,\Omega',\lambda) &= \sigma_{\mathtt{scat}}(r,\lambda)\,P(r,\Omega,\Omega',\lambda)\\ &= \sigma^{\mathtt{mol}}_{\mathtt{scat}}(r,\lambda)\,P_{\mathtt{Ray}}(\Omega,\Omega',\lambda) + \sigma^{\mathtt{aer}}_{\mathtt{scat}}(r,\lambda)\,P_{\mathtt{Mie}}(r,\Omega,\Omega',\lambda).\end{aligned}$$

If $\cos\Theta = \Omega\cdot\Omega'$ is the cosine of the angle Θ between the incident and scattering directions, we have the simplified representation

$$\begin{aligned}\mathcal{S}(r,\cos\Theta,\lambda) &= \sigma_{\mathtt{scat}}(r,\lambda)\,P(r,\cos\Theta,\lambda)\\ &= \sigma^{\mathtt{mol}}_{\mathtt{scat}}(r,\lambda)\,P_{\mathtt{Ray}}(\cos\Theta,\lambda) + \sigma^{\mathtt{aer}}_{\mathtt{scat}}(r,\lambda)\,P_{\mathtt{Mie}}(r,\cos\Theta,\lambda).\end{aligned}$$

In practical applications it is important to compute

$$P_{\mathtt{Ray}}(\cos\Theta_{\mathtt{sun}},\lambda)$$

and

$$\bar{P}_{\mathtt{Mie},p}(\cos\Theta_{\mathtt{sun}},\lambda)$$

for all stages $p = 1,...,N^{\mathtt{stage}}_{\mathtt{LOS}}$. The Rayleigh phase function possesses the representation

$$P_{\mathtt{Ray}}(\cos\Theta,\lambda) = A(\lambda) + B(\lambda)\cos^2\Theta$$

with

$$A(\lambda) = \frac{3+3\rho(\lambda)}{4+2\rho(\lambda)},\quad B(\lambda) = \frac{3-3\rho(\lambda)}{4+2\rho(\lambda)}.$$

The depolarization ratio is computed as

$$\rho(\lambda) = \frac{6F_K(\lambda)-6}{7F_K(\lambda)+3},$$

while the King factor is given by

$$F_K(\lambda) = 1.04695 + 3.25031 \cdot 10^{-4} \left(\frac{1}{\lambda}\right)^2 + 3.86228 \cdot 10^{-5} \left(\frac{1}{\lambda}\right)^4$$

with λ expressed in $\mu\mathrm{m}$. The Mie scattering phase function takes the form

$$\bar{P}_{\mathtt{Mie},\mathtt{lay}(p)}(\cos\Theta,\lambda) = \frac{1-\bar{g}^2_{\mathtt{lay}(p)}(\lambda)}{\left(1+\bar{g}^2_{\mathtt{lay}(p)}(\lambda) - 2\bar{g}_{\mathtt{lay}(p)}(\lambda)\cos\Theta\right)^{3/2}},$$

where the asymmetry factor $\bar{g}_{\mathtt{lay}(p)}$ on the layer $\mathrm{lay}(p)$ is computed as

$$\bar{g}_{\mathtt{lay}(p)}(\lambda) = \bar{a}^{\mathtt{aer}}_{4,\mathtt{lay}(p)} + (\lambda - \lambda_{\mathtt{ref}})\,\bar{a}^{\mathtt{aer}}_{5,\mathtt{lay}(p)}.$$

8.2.4. Recurrence relation for limb radiance and Jacobian

We are now well prepared to transform equation (8.1) into a computational expression. For two points on the line of sight M_p and M_{p+1} characterized by the position vectors $\mathbf{r}_p$ and $\mathbf{r}_{p+1}$ and being the boundaries of the layer $\mathrm{lay}(p)$, the integral form of the radiative transfer equation takes the form

$$\begin{aligned} I_p(\Omega_{\mathtt{LOS}}) &= I_{p+1}(\Omega_{\mathtt{LOS}})\,\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}\triangle_p} \\ &\quad + \int_{|\mathbf{r}_p-\mathbf{r}_{p+1}|} J(\mathbf{r}',\Omega_{\mathtt{LOS}})\,\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}|\mathbf{r}_p-\mathbf{r}'|}\mathrm{d}s', \end{aligned}$$

with $I_p(\Omega_{\mathtt{LOS}}) = I(\mathbf{r}_p,\Omega_{\mathtt{LOS}})$ and $\triangle_p = |\mathbf{r}_p - \mathbf{r}_{p+1}|$. Note that the position vector of the point M_p is $\mathbf{r}_p$, while the radial distance is $r(M_p)$, that is, in the solar coordinate system we have the representation $\mathbf{r}_p = (r(M_p), \Phi_{\mathtt{sun},p}, \Psi_{\mathtt{sun},p})$. The downward recurrence relation starts at the point index $N^{\mathtt{point}}_{\mathtt{LOS}}$, where for limb viewing geometries,

$$I_{N^{\mathtt{point}}_{\mathtt{LOS}}}(\Omega_{\mathtt{LOS}}) = 0.$$

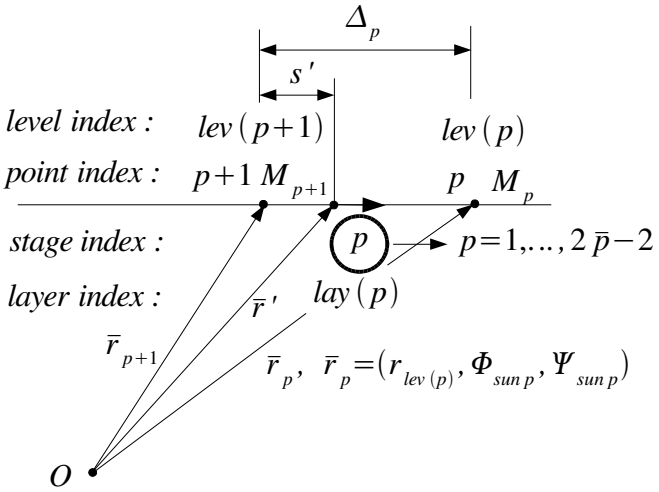

Figure 8.7.: *Stage index p, point indices p and $p+1$, layer index $lay(p)$ and the level indices $lev(p)$ and $lev(p+1)$.*

8.2.4.1. Recurrence relation for limb radiance

In this Section we are concerned with the single and multiple scattering term integrations.

Single scattering term integration

On the layer $\mathrm{lay}\,(p)$, where p is the stage index ranging downward from $N_{\mathtt{LOS}}^{\mathtt{stage}}$ to 1, the single scattering term can be approximated by

$$J_{\mathtt{ss}}\left(\mathbf{r}',\Omega_{\mathtt{LOS}}\right)=\frac{1}{4\pi}\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)}_{p}F_{\mathtt{sun}}\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(\left|\mathbf{r}'-\mathbf{r}_{\mathtt{TOA}}\left(\mathbf{r}'\right)\right|\right)}+\bar{\sigma}_{\mathtt{abs},\mathtt{lay}(p)}\bar{B}_{\mathtt{lay}(p)},$$

where

$$\bar{B}_{\mathtt{lay}(p)}=\frac{1}{2}\left[B\left(T_{p}\right)+B\left(T_{p+1}\right)\right],$$

with T_p and T_{p+1} being the intensities at the points M_p and M_{p+1}, respectively. We define the solar optical depths at the boundary points M_p and M_{p+1} by

$$\begin{aligned}\tau_{\mathtt{ext},p}^{\mathtt{sun}} &= \int_{\left|\mathbf{r}_{p}-\mathbf{r}_{\mathtt{TOA},p}\right|}\sigma_{\mathtt{ext}}\left(\mathbf{r}'\right)\mathrm{d}s'=\sum_{q=1}^{N_{\mathtt{sun},p}^{\mathtt{trav}}}\bar{\sigma}_{\mathtt{ext},q}\triangle_{\mathtt{sun},p,q},\\ \tau_{\mathtt{ext},p+1}^{\mathtt{sun}} &= \int_{\left|\mathbf{r}_{p+1}-\mathbf{r}_{\mathtt{TOA},p+1}\right|}\sigma_{\mathtt{ext}}\left(\mathbf{r}'\right)\mathrm{d}s'=\sum_{q=1}^{N_{\mathtt{sun},p+1}^{\mathtt{trav}}}\bar{\sigma}_{\mathtt{ext},q}\triangle_{\mathtt{sun},p+1,q},\end{aligned}$$

where the $\triangle_{\mathtt{sun},p,q}$ are the solar optical depths at the point M_p for all $q=1,...,N_{\mathtt{sun},p}^{\mathtt{trav}}$ and the $\triangle_{\mathtt{sun},p+1,q}$ are the solar optical depths at the point M_{p+1} for all $q=1,...,N_{\mathtt{sun},p+1}^{\mathtt{trav}}$. We then use the linear approximations

$$\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(\left|\mathbf{r}'-\mathbf{r}_{\mathtt{TOA}}\left(\mathbf{r}'\right)\right|\right)=\frac{s}{\triangle_{p}}\tau_{\mathtt{ext},p}^{\mathtt{sun}}+\left(1-\frac{s}{\triangle_{p}}\right)\tau_{\mathtt{ext},p+1}^{\mathtt{sun}}$$

to obtain

$$\begin{aligned}&\int_{\left|\mathbf{r}_{p}-\mathbf{r}_{p+1}\right|}J_{\mathtt{ss}}\left(\mathbf{r}',\Omega_{\mathtt{LOS}}\right)\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}\left|\mathbf{r}_{p}-\mathbf{r}'\right|}\mathrm{d}s'\\ &=\frac{F_{\mathtt{sun}}}{4\pi}\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)}_{p}\int_{0}^{\triangle_{p}}\mathrm{e}^{-\left[\frac{s}{\triangle_{p}}\tau_{\mathtt{ext},p}^{\mathtt{sun}}+\left(1-\frac{s}{\triangle_{p}}\right)\tau_{\mathtt{ext},p+1}^{\mathtt{sun}}\right]}\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}\left(\triangle_{p}-s\right)}\,\mathrm{d}s\\ &+\bar{\sigma}_{\mathtt{abs},\mathtt{lay}(p)}\bar{B}_{\mathtt{lay}(p)}\int_{0}^{\triangle_{p}}\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}\left(\triangle_{p}-s\right)}\,\mathrm{d}s.\end{aligned}$$

The result of integration is

$$\begin{aligned}I_{\mathtt{ss}}\left(\mathbf{r}_{p},\mathbf{r}_{p+1},\Omega_{\mathtt{LOS}}\right)&=\int_{\left|\mathbf{r}_{p}-\mathbf{r}_{p+1}\right|}J_{\mathtt{ss}}\left(\mathbf{r}',\Omega_{\mathtt{LOS}}\right)\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}\left|\mathbf{r}_{p}-\mathbf{r}'\right|}\mathrm{d}s'\\ &=\frac{F_{\mathtt{sun}}}{4\pi}\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)}_{p}\triangle_{p}b\left(\tau_{p}\right)+a_{0}\left(\tau_{p}\right)\triangle_{p}\bar{\sigma}_{\mathtt{abs},\mathtt{lay}(p)}\bar{B}_{\mathtt{lay}(p)}\end{aligned}$$

where

$$\begin{aligned}a_{0}\left(x\right) &= \int_{0}^{1}\mathrm{e}^{-x(1-\xi)}d\xi=\frac{1-\mathrm{e}^{-x}}{x}\\ b\left(x\right) &= \int_{0}^{1}\mathrm{e}^{-\left[\xi\tau_{\mathtt{ext},p}^{\mathtt{sun}}+(1-\xi)\tau_{\mathtt{ext},p+1}^{\mathtt{sun}}\right]}\mathrm{e}^{-x(1-\xi)}\,\mathrm{d}\xi=\frac{\mathrm{e}^{-\left(\tau_{\mathtt{ext},p+1}^{\mathtt{sun}}+x\right)}-\mathrm{e}^{-\tau_{\mathtt{ext},p}^{\mathtt{sun}}}}{\tau_{\mathtt{ext},p}^{\mathtt{sun}}-\tau_{\mathtt{ext},p+1}^{\mathtt{sun}}-x}\end{aligned}$$

and

$$\tau_{p}=\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}\triangle_{p}.$$

The term $\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)_p}$ is computed by using the relation

$$\overline{\mathcal{S}\left(r,\Omega,\Omega_{\mathtt{sun}}\right)_p}=\bar{\sigma}^{\mathtt{mol}}_{\mathtt{scat},\mathtt{lay}(p)}P_{\mathtt{Ray}}\left(\overline{\cos\Theta}_{\mathtt{sun},p}\right)+\bar{\sigma}^{\mathtt{aer}}_{\mathtt{scat},\mathtt{lay}(p)}\bar{P}_{\mathtt{Mie},\mathtt{lay}(p)}\left(\overline{\cos\Theta}_{\mathtt{sun},p}\right)$$

where the stage values $P_{\mathtt{Ray}}\left(\overline{\cos\Theta}_{\mathtt{sun},p}\right)$ and $\bar{P}_{\mathtt{Mie},\mathtt{lay}(p)}\left(\overline{\cos\Theta}_{\mathtt{sun},p}\right)$ are calculated by the optical module for all $p=1,...,N^{\mathtt{stage}}_{\mathtt{LOS}}$.

Multiple scattering term integration

On the layer $\mathrm{lay}\,(p)$, where p is the stage index ranging downward from $N^{\mathtt{stage}}_{\mathtt{LOS}}$ to 1, the multiple scattering term read as

$$J_{\mathtt{ms}}\left(\mathbf{r}',\Omega_{\mathtt{LOS}}\right)=\frac{1}{4\pi}\int_{4\pi}\bar{\mathcal{S}}_p\left(\Omega_{\mathtt{LOS}},\Omega'\right)I\left(\mathbf{r}',\Omega'\right)\,\mathrm{d}\Omega'$$

and note that the calculation of $\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)_p}$ and $\bar{\mathcal{S}}_p\left(\Omega_{\mathtt{LOS}},\Omega'\right)$ is different. Using the linear approximation

$$I\left(\mathbf{r}',\Omega'\right)=\frac{s}{\triangle_p}I_p\left(\Omega'\right)+\left(1-\frac{s}{\triangle_p}\right)I_{p+1}\left(\Omega'\right),$$

where I_p and I_{p+1} are the intensities at the points M_p and M_{p+1}, respectively, we obtain

$$\begin{aligned}&\int_{|\mathbf{r}_p-\mathbf{r}_{p+1}|}J_{\mathtt{ms}}\left(\mathbf{r}',\Omega_{\mathtt{LOS}}\right)\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}|\mathbf{r}_p-\mathbf{r}'|}\mathrm{d}s'\\&=\left[\int_0^{\triangle_p}\frac{s}{\triangle_p}\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}(\triangle_p-s)}\,\mathrm{d}s\right]\bar{J}^1_{\mathtt{ms},p}\\&+\left[\int_0^{\triangle_p}\left(1-\frac{s}{\triangle_p}\right)\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}(\triangle_p-s)}\,\mathrm{d}s\right]\bar{J}^2_{\mathtt{ms},p}\end{aligned}$$

with

$$\begin{aligned}\bar{J}_{\mathtt{ms},p}&=\frac{1}{4\pi}\int_{4\pi}\bar{\mathcal{S}}_p\left(\Omega_{\mathtt{LOS}},\Omega'\right)I_p\left(\Omega'\right)\,\mathrm{d}\Omega'\\\bar{J}_{\mathtt{ms},p+1}&=\frac{1}{4\pi}\int_{4\pi}\bar{\mathcal{S}}_p\left(\Omega_{\mathtt{LOS}},\Omega'\right)I_{p+1}\left(\Omega'\right)\,\mathrm{d}\Omega'\end{aligned}$$

To compute $\bar{J}_{\mathtt{ms},p}$ we use the relation

$$\begin{aligned}\bar{J}_{\mathtt{ms},p}&=\frac{1}{4\pi}\int_{4\pi}\bar{\mathcal{S}}_p\left(\Omega_{\mathtt{LOS}},\Omega'\right)I_p\left(\Omega'\right)\,\mathrm{d}\Omega'\\&=\frac{1}{2}\sum_{m=0}^{2M-1}\left[\int_{-1}^{1}s_m\left(\bar{\sigma}_{\mathtt{scat},\mathtt{lay}(p)},\mu_{\mathtt{LOS},p},\mu'\right)I_m\left(r\left(M_p\right),\mu'\right)\,\mathrm{d}\mu'\right]\cos m\varphi_{\mathtt{LOS},p}\\&=\frac{1}{2}\sum_{m=0}^{2M-1}\left[\sum_{k=1}^{M}w^+_k s_m\left(\bar{\sigma}_{\mathtt{scat},\mathtt{lay}(p)},\mu_{\mathtt{LOS},p},\mu^+_k\right)I_m\left(r\left(M_p\right),\mu^+_k\right)\right.\\&\left.+\sum_{k=1}^{M}w^-_k s_m\left(\bar{\sigma}_{\mathtt{scat},\mathtt{lay}(p)},\mu_{\mathtt{LOS},p},\mu^-_k\right)I_m\left(r\left(M_p\right),\mu^-_k\right)\right]\cos m\varphi_{\mathtt{LOS},p},\end{aligned}$$

where $\theta_{\mathtt{LOS},p}$ and $\varphi_{\mathtt{LOS},p}$ are the zenith and azimuthal angles of the point M_p, $\mu_{\mathtt{LOS},p}=\cos\theta_{\mathtt{LOS},p}$ and

$$\begin{aligned}s_m\left(\bar{\sigma}_{\mathtt{scat},\mathtt{lay}(p)},\mu_{\mathtt{LOS},p},\mu'\right)&=\sum_{n=m}^{2M-1}\xi_n\left(\bar{\sigma}_{\mathtt{scat},\mathtt{lay}(p)}\right)P^m_n\left(\mu_{\mathtt{LOS},p}\right)P^m_n\left(\mu'\right),\\\xi_n\left(\bar{\sigma}_{\mathtt{scat},\mathtt{lay}(p)}\right)&=\bar{\sigma}^{\mathtt{mol}}_{\mathtt{scat},\mathtt{lay}(p)}\chi^{\mathtt{mol}}_n+\bar{\sigma}^{\mathtt{aer}}_{\mathtt{scat},\mathtt{lay}(p)}\bar{\chi}^{\mathtt{aer}}_{n,\mathtt{lay}(p)}.\end{aligned}$$

The $I_m\left(r\left(M_p\right),\theta\right)$ are the azimuthal components of the intensity at the the point M_p

$$I_p\left(\Omega\right) = I\left(r\left(M_p\right),\theta,\varphi\right) = \sum_{m=0}^{2M-1} I_m\left(r\left(M_p\right),\theta\right)\cos m\varphi$$

and are computed at a set of discrete angles $\theta_k^{\pm}$, $k=1,...,M$ by the multiple scattering module. For $\bar{J}_{\mathtt{ms},p+1}$ we have a similar relation

$$\begin{aligned}\bar{J}_{\mathtt{ms},p+1} &= \frac{1}{4\pi}\int_{4\pi}\bar{\mathcal{S}}_p\left(\Omega_{\mathtt{LOS}},\Omega'\right)I_{p+1}\left(\Omega'\right)\,\mathrm{d}\Omega' \\ &\quad \frac{1}{2}\sum_{m=0}^{2M-1}\left[\sum_{k=1}^{M} w_k^+ s_m\left(\bar{\sigma}_{\mathtt{scat,lay}(p)},\mu_{\mathtt{LOS},p+1},\mu_k^+\right)I_m\left(r\left(M_{p+1}\right)\mu_k^+\right)\right. \\ &\quad \left.+\sum_{k=1}^{M} w_k^- s_m\left(\bar{\sigma}_{\mathtt{scat,lay}(p)},\mu_{\mathtt{LOS},p+1},\mu_k^-\right)I_m\left(r\left(M_{p+1}\right),\mu_k^-\right)\right]\cos m\varphi_{\mathtt{LOS},p+1},\end{aligned}$$

where $I_m\left(r\left(M_{p+1}\right),\theta\right)$ are the azimuthal components of the intensity at the the point M_{p+1} and

$$s_m\left(\bar{\sigma}_{\mathtt{scat,lay}(p)},\mu_{\mathtt{LOS},p+1},\mu'\right) = \sum_{n=m}^{2M-1}\xi_n\left(\bar{\sigma}_{\mathtt{scat,lay}(p)}\right)P_n^m\left(\mu_{\mathtt{LOS},p+1}\right)P_n^m\left(\mu'\right).$$

Performing the integrals we arrive at

$$\begin{aligned}I_{\mathtt{ms}}\left(\mathbf{r}_p,\mathbf{r}_{p+1},\Omega_{\mathtt{LOS}}\right) &= \int_{|\mathbf{r}_p-\mathbf{r}_{p+1}|} J_{\mathtt{ms}}\left(\mathbf{r}',\Omega_{\mathtt{LOS}}\right)\mathrm{e}^{-\bar{\sigma}_{\mathtt{ext,lay}(p)}|\mathbf{r}_p-\mathbf{r}'|}\mathrm{d}s' \\ &= a_2\left(\tau_p\right)\triangle_p\bar{J}_{\mathtt{ms},p} + a_1\left(\tau_p\right)\triangle_p\bar{J}_{\mathtt{ms},p+1}\end{aligned}$$

with

$$\begin{aligned}a_1(x) &= \int_0^1(1-\xi)\mathrm{e}^{-x(1-\xi)}d\xi = \frac{1}{x^2}\left[1-(1+x)\,\mathrm{e}^{-x}\right] \\ a_2(x) &= \int_0^1\xi\mathrm{e}^{-x(1-\xi)}d\xi = \frac{1}{x^2}\left(x-1+\mathrm{e}^{-x}\right).\end{aligned}$$

Collecting all results we are led to the recurrence relation

$$\begin{aligned}I_p\left(\Omega_{\mathtt{LOS}}\right) &= I_{p+1}\left(\Omega_{\mathtt{LOS}}\right)\mathrm{e}^{-\tau_p} + \frac{F_{\mathtt{sun}}}{4\pi}\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)_p}\triangle_p b\left(\tau_p\right) \\ &\quad +a_0\left(\tau_p\right)\triangle_p\bar{\sigma}_{\mathtt{abs,lay}(p)}\bar{B}_{\mathtt{lay}(p)} + a_2\left(\tau_p\right)\triangle_p\bar{J}_{\mathtt{ms},p} + a_1\left(\tau_p\right)\triangle_p\bar{J}_{\mathtt{ms},p+1}\end{aligned}$$

8.2.4.2. Recurrence relation for the Jacobian matrix

The derivative of the radiance at the top of the atmosphere $I_p\left(\Omega_{\mathtt{LOS}}\right)$ with respect to the partial columns $\bar{X}_{g,i}$, with $i=1,...,N_{\mathtt{d}}$ and $g=1,...,N_{\mathtt{gas}}$

$$\frac{\partial I_p}{\partial\bar{X}_{g,i}}\left(\Omega_{\mathtt{LOS}}\right),$$

can be computed in terms of the derivatives with respect to the extinction coefficient

$$\frac{\partial I_p}{\partial\bar{\sigma}_{\mathtt{ext},i}}\left(\Omega_{\mathtt{LOS}}\right)$$

by using the relation

$$\frac{\partial I_m}{\partial \bar{X}_{g,i}}\left(\Omega_{\mathtt{LOS}}\right) = \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(\Omega_{\mathtt{LOS}}\right)\frac{\partial \bar{\sigma}_{\mathtt{ext},i}}{\partial \bar{X}_{g,i}}$$

with

$$\frac{\partial \bar{\sigma}_{\mathtt{ext},i}}{\partial \bar{X}_{g,i}} = \frac{\partial \bar{\sigma}^{\mathtt{mol}}_{\mathtt{abs},i}}{\partial \bar{X}_{g,j}} \cdot 2.6868 \cdot 10^{16} \cdot \frac{1}{\triangle r_i}\bar{C}_{\mathtt{abs},g,i}.$$

For the partial derivatives with respect to the extinction coefficient we have the recurrence relation

$$\begin{aligned}
&\frac{\partial I_p}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(\Omega_{\mathtt{LOS}}\right)\\
&= I_{p+1}\left(\Omega_{\mathtt{LOS}}\right)\frac{\partial}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left\{\mathrm{e}^{-\tau_p}\right\} + \mathrm{e}^{-\tau_p}\frac{\partial I_{p+1}}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(\Omega_{\mathtt{LOS}}\right)\\
&+\frac{F_{\mathtt{sun}}}{4\pi}\triangle_p b\left(\tau_p\right)\frac{\partial}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left\{\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)_p}\right\} + \frac{F_{\mathtt{sun}}}{4\pi}\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)_p}\triangle_p\frac{\partial b\left(\tau_p\right)}{\partial \bar{\sigma}_{\mathtt{ext},i}}\\
&+\triangle_p\bar{\sigma}_{\mathtt{abs},\mathtt{lay}(p)}\bar{B}_{\mathtt{lay}(p)}\frac{\partial}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left\{a_0\left(\tau_p\right)\right\} + a_0\left(\tau_p\right)\triangle_p\bar{B}_{\mathtt{lay}(p)}\frac{\partial \bar{\sigma}_{\mathtt{abs},\mathtt{lay}(p)}}{\partial \bar{\sigma}_{\mathtt{ext},i}}\\
&+a_2\left(\tau_p\right)\triangle_p\frac{\partial \bar{J}_{\mathtt{ms},p}}{\partial \bar{\sigma}_{\mathtt{ext},i}} + \triangle_p\bar{J}_{\mathtt{ms},p}\frac{\partial}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left\{a_2\left(\tau_p\right)\right\}\\
&+a_1\left(\tau_p\right)\triangle_p\frac{\partial \bar{J}_{\mathtt{ms},p+1}}{\partial \bar{\sigma}_{\mathtt{ext},i}} + \triangle_p\bar{J}_{\mathtt{ms},p+1}\frac{\partial}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left\{a_1\left(\tau_p\right)\right\}
\end{aligned}$$

For our calculation it is important to observe that the extinction coefficient $\bar{\sigma}_{\mathtt{ext},i}$ which is computed as

$$\bar{\sigma}_{\mathtt{ext},i} = \left(\bar{\sigma}^{\mathtt{mol}}_{\mathtt{scat},i} + \bar{\sigma}^{\mathtt{mol}}_{\mathtt{abs},i}\right) + \bar{\sigma}^{\mathtt{aer}}_{\mathtt{ext},i}$$

(note that $\bar{\sigma}^{\mathtt{aer}}_{\mathtt{scat},i}$ does not appear explicitly) enters:

1. explicitly,
2. in the expression of $\tau_{\mathtt{sun}}$ and
3. in the expression of $\bar{\sigma}_{\mathtt{abs},\mathtt{lay}(p)}$ multiplying the Planck function B,

$$\bar{\sigma}_{\mathtt{abs},\mathtt{lay}(p)} = \bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)} - \bar{\sigma}_{\mathtt{scat},\mathtt{lay}(p)},$$

but does not enter in the expression of $\mathcal{S}$ (or s_m). The partial derivative terms are computed as follows:

1. The partial derivative of $\mathrm{e}^{-\tau_p}$ is given by

$$\frac{\partial}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left\{\mathrm{e}^{-\tau_p}\right\} = -\triangle_p\mathrm{e}^{-\tau_p}\frac{\partial \bar{\sigma}_{\mathtt{ext},\mathtt{lay}(p)}}{\partial \bar{\sigma}_{\mathtt{ext},i}} = -\triangle_p\mathrm{e}^{-\tau_p}\delta_{\mathtt{lay}(p),i}$$

2. Since $\mathcal{S}$ does not depend on $\bar{\sigma}_{\mathtt{ext},i}$, it is apparent that

$$\frac{\partial}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left\{\overline{\mathcal{S}\left(r,\Omega_{\mathtt{LOS}},\Omega_{\mathtt{sun}}\right)_p}\right\} = 0.$$

3. To compute the term

$$\frac{\partial b\left(\tau_p\right)}{\partial \bar{\sigma}_{\mathtt{ext},i}},$$

we first compute

$$\begin{aligned}\frac{\partial\tau^{\mathtt{sun}}_{\mathtt{ext},p}}{\partial\bar{\sigma}_{\mathtt{ext},i}} &= \frac{\partial}{\partial\bar{\sigma}_{\mathtt{ext},i}}\left\{\sum_{q=1}^{N^{\mathtt{trav}}_{\mathtt{sun},p}}\bar{\sigma}_{\mathtt{ext},q}\triangle_{\mathtt{sun},p,q}\right\}\\ &= \begin{cases}\triangle_{\mathtt{sun},p,i}, & i\leq N^{\mathtt{trav}}_{\mathtt{sun},p}\\ 0, & i>N^{\mathtt{trav}}_{\mathtt{sun},p}\end{cases}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\tau^{\mathtt{sun}}_{\mathtt{ext},p+1}}{\partial\bar{\sigma}_{\mathtt{ext},i}} &= \frac{\partial}{\partial\bar{\sigma}_{\mathtt{ext},i}}\left\{\sum_{q=1}^{N^{\mathtt{trav}}_{\mathtt{sun},p+1}}\bar{\sigma}_{\mathtt{ext},q}\triangle_{\mathtt{sun},p+1,q}\right\}\\ &= \begin{cases}\triangle_{\mathtt{sun},p+1,i}, & i\leq N^{\mathtt{trav}}_{\mathtt{sun},p+1}\\ 0, & i>N^{\mathtt{trav}}_{\mathtt{sun},p+1}\end{cases}.\end{aligned}$$

Then we have

$$\begin{aligned}\frac{\partial b(\tau_p)}{\partial\bar{\sigma}_{\mathtt{ext},i}} &= -b(\tau_p)\left\{\frac{\partial\tau^{\mathtt{sun}}_{\mathtt{ext},p+1}}{\partial\bar{\sigma}_{\mathtt{ext},i}}\right.\\ &+\frac{1}{\tau^{\mathtt{sun}}_{\mathtt{ext},p}-\tau^{\mathtt{sun}}_{\mathtt{ext},p+1}-\tau_p}\left[\left(\frac{\partial\tau^{\mathtt{sun}}_{\mathtt{ext},p}}{\partial\bar{\sigma}_{\mathtt{ext},i}}-\frac{\partial\tau^{\mathtt{sun}}_{\mathtt{ext},p+1}}{\partial\bar{\sigma}_{\mathtt{ext},i}}\right)\right.\\ &\left.\left.-\triangle_p\delta_{\mathtt{lay}(p),i}\right]\right\}\\ &+\frac{1}{\tau^{\mathtt{sun}}_{\mathtt{ext},p}-\tau^{\mathtt{sun}}_{\mathtt{ext},p+1}-\tau_p}\left[\mathrm{e}^{-\tau^{\mathtt{sun}}_{\mathtt{ext},p+1}}\frac{\partial}{\partial\bar{\sigma}_{\mathtt{ext},i}}\left\{\mathrm{e}^{-\tau_p}\right\}\right.\\ &\left.+\mathrm{e}^{-\tau^{\mathtt{sun}}_{\mathtt{ext},p}}\left(\frac{\partial\tau^{\mathtt{sun}}_{\mathtt{ext},p}}{\partial\bar{\sigma}_{\mathtt{ext},i}}-\frac{\partial\tau^{\mathtt{sun}}_{\mathtt{ext},p+1}}{\partial\bar{\sigma}_{\mathtt{ext},i}}\right)\right]\end{aligned}$$

4. The partial derivative of $a_0(\tau_p)$ takes the form

$$\frac{\partial}{\partial\bar{\sigma}_{\mathtt{ext},i}}\left\{a_0(\tau_p)\right\}=a_0'(\tau_p)\triangle_p\delta_{\mathtt{lay}(p),i},$$

where

$$a_0'(x)=\frac{\mathrm{e}^{-x}x-(1-\mathrm{e}^{-x})}{x^2}=\frac{x\mathrm{e}^{-x}+\mathrm{e}^{-x}-1}{x^2}.$$

Similarly, the partial derivatives of $a_2(\tau_p)$ and $a_1(\tau_p)$ are given by

$$\frac{\partial}{\partial\bar{\sigma}_{\mathtt{ext},i}}\left\{a_2(\tau_p)\right\}=a_2'(\tau_p)\triangle_p\delta_{\mathtt{lay}(p),i}$$

with

$$\begin{aligned}a_2'(x) &= \frac{(1-\mathrm{e}^{-x})x^2-2x(x-1+\mathrm{e}^{-x})}{x^4}\\ &= \frac{x^2-x^2\mathrm{e}^{-x}-2x^2+2x-2x\mathrm{e}^{-x}}{x^4}\\ &= \frac{-x^2\mathrm{e}^{-x}-x^2+2x-2x\mathrm{e}^{-x}}{x^4}\\ &= -\frac{(x-2)+\mathrm{e}^{-x}(x+2)}{x^3},\end{aligned}$$

and

$$\frac{\partial}{\partial \bar{\sigma}_{\mathtt{ext},i}} \left\{ a_1 \left(\tau_p \right) \right\} = a_1' \left(\tau_p \right) \triangle_p \delta_{\mathtt{lay}(p),i}$$

with

$$a_1'(x) = a_0' \left(x \right) - a_2'(x).$$

5. The partial derivative of the absorption coefficient is identical to the partial derivative of the extinction coefficient

$$\frac{\partial \bar{\sigma}_{\mathtt{abs,lay}(p)}}{\partial \bar{\sigma}_{\mathtt{ext},i}} = \frac{\partial \bar{\sigma}_{\mathtt{ext,lay}(p)}}{\partial \bar{\sigma}_{\mathtt{ext},i}} = \delta_{\mathtt{lay}(p),i}.$$

6. To compute the partial derivatives of the multiple scattering terms $\bar{J}_{\mathtt{ms},p}$ and $\bar{J}_{\mathtt{ms},p+1}$ we take into account that the s_m does not depend on $\bar{\sigma}_{\mathtt{ext},i}$ and derive

$$\begin{aligned} & \frac{\partial \bar{J}_{\mathtt{ms},p}}{\partial \bar{\sigma}_{\mathtt{ext},i}} \\ = \; & \frac{1}{2} \sum_{m=0}^{2M-1} \left[\sum_{k=1}^{M} w_k^+ s_m \left(\bar{\sigma}_{\mathtt{scat,lay}(p)}, \mu_{\mathtt{LOS},p}, \mu_k^+ \right) \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}} \left(r \left(M_p \right), \mu_k^+ \right) \right. \\ & \left. + \sum_{k=1}^{M} w_k^- s_m \left(\bar{\sigma}_{\mathtt{scat,lay}(p)}, \mu_{\mathtt{LOS},p}, \mu_k^- \right) \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}} \left(r \left(M_p \right), \mu_k^- \right) \right] \cos m\varphi_{\mathtt{LOS},p} \end{aligned}$$

and

$$\begin{aligned} & \frac{\partial \bar{J}_{\mathtt{ms},p+1}}{\partial \bar{\sigma}_{\mathtt{ext},i}} \\ = \; & \frac{1}{2} \sum_{m=0}^{2M-1} \left[\sum_{k=1}^{M} w_k^+ s_m \left(\bar{\sigma}_{\mathtt{scat,lay}(p)}, \mu_{\mathtt{LOS},p+1}, \mu_k^+ \right) \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}} \left(r \left(M_{p+1} \right), \mu_k^+ \right) \right. \\ & \left. + \sum_{k=1}^{M} w_k^- s_m \left(\bar{\sigma}_{\mathtt{scat,lay}(p)}, \mu_{\mathtt{LOS},p+1}, \mu_k^- \right) \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}} \left(r \left(M_{p+1} \right), \mu_k^- \right) \right] \cos m\varphi_{\mathtt{LOS},p+1} \end{aligned}$$

8.2.4.3. Input parameters for limb radiance and Jacobian calculation

The input parameters for radiance and Jacobian calculation are listed below:

1. the extinction coefficient $\bar{\sigma}_{\mathtt{ext},j} \left(\lambda \right)$, the molecular and aerosol scattering coefficients $\bar{\sigma}^{\mathtt{mol}}_{\mathtt{scat},j} \left(\lambda \right)$ and $\bar{\sigma}^{\mathtt{aer}}_{\mathtt{scat},j} \left(\lambda \right)$, and the molecular absorption coefficient $\bar{\sigma}^{\mathtt{mol}}_{\mathtt{abs},j} \left(\lambda \right)$ on all layers, that is, for all $j = 1, ..., N_{\mathtt{lay}}$;
2. the partial derivative of the extinction coefficient

$$\frac{\partial \bar{\sigma}_{\mathtt{ext},j}}{\partial \bar{X}_{g,j}},$$

   for all gases and all on all layers, that is, for all $g = 1, ..., N_{\mathtt{gas}}$ and $j = 1, ..., N_{\mathtt{lay}}$;
3. the Rayleigh and Mie phase functions $P_{\mathtt{Ray}} \left(\overline{\cos \Theta}_{\mathtt{sun},p} \right)$ and $\bar{P}_{\mathtt{Mie},p} \left(\overline{\cos \Theta}_{\mathtt{sun},p} \right)$ on all stages on the line of sight, that is, for all $p = 1, ..., N^{\mathtt{stage}}_{\mathtt{LOS}}$;
4. the limb paths $\triangle_p$ of all stages on the line of sight, that is, for all $p = 1, ..., N^{\mathtt{stage}}_{\mathtt{LOS}}$;
5. the number of solar traversed layers $N^{\mathtt{trav}}_{\mathtt{sun},p}$ at all points M_p on the line of sight, that is, for all $p = 1, ..., N^{\mathtt{point}}_{\mathtt{LOS}}$;
6. the solar optical depths $\triangle_{\mathtt{sun},p,q}$ at all points M_p on the line of sight, that is, for all $q = 1, ..., N^{\mathtt{trav}}_{\mathtt{sun},p}$ and $p = 1, ..., N^{\mathtt{point}}_{\mathtt{LOS}}$;

7. the multiple scattering terms $\bar{J}_{\mathtt{ms},p}$ and $\bar{J}_{\mathtt{ms},p+1}$ on all stages on the line of sight, that is for all $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{stage}}$;
8. the partial derivatives of the multiple scattering terms

$$\frac{\partial \bar{J}_{\mathtt{ms},p}}{\partial \bar{\sigma}_{\mathtt{ext},i}} \text{ and } \frac{\partial \bar{J}_{\mathtt{ms},p+1}}{\partial \bar{\sigma}_{\mathtt{ext},i}},$$

for all stages on the line of sight and all layers, that is, for all $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{stage}}$ and $i = 1, ..., N_{\mathtt{lay}}$.

8.2.5. Pseudo-spherical discrete ordinate radiative transfer equation

The discrete ordinate method converts the radiative transfer equation into a linear system of differential equations by discretizing the angular variation of the phase function and radiance. In the conventional eigenvalue approach, the general solution of the linear system of differential equations consists of a linear combination of all the homogeneous solutions plus the particular solutions for the assumed sources. For a multi-layered medium, the expansion coefficients of the homogeneous solutions are the unknowns of the discretized radiative transfer problem and are computed by imposing the continuity condition for the radiances across the layer interfaces. In the matrix-exponential formalism, the linear system of differential equations is treated as a boundary value problem. For each layer this classical mathematical procedure yields a so called layer equation which relates the level values of the radiance. The discretized radiative transfer problem then reduces to a system of linear algebraic equations for the unknown level values of the radiance.

The purpose of this Section is to present a stable discrete ordinate algorithm for vertically inhomogeneous layered media using the matrix-exponential solution. Conceptually, the algorithm is similar to a finite-element algorithm and involves the following steps:

- ➢ discretization of the atmosphere into a number of distinct but vertically uniform layers,
- ➢ derivation of the layer equation using the matrix-exponential solution,
- ➢ assemblage of the layer equation into the system matrix of the entire atmosphere,
- ➢ solution of the assembled system of equations.

8.2.5.1. Requirement for a pseudo-spherical model

The calculation of the multiple scattering terms $\bar{J}_{\mathtt{ms},p}$ and $\bar{J}_{\mathtt{ms},p+1}$ and of their partial derivatives $\frac{\partial \bar{J}_{\mathtt{ms},p}}{\partial \bar{\sigma}_{\mathtt{ext},i}}$ and $\frac{\partial \bar{J}_{\mathtt{ms},p+1}}{\partial \bar{\sigma}_{\mathtt{ext},i}}$ on the stage index p, with $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{stage}}$ require the knowledge of the azimuthal components of the intensities and their derivatives at the boundary points M_p and M_{p+1}, that is,

$$I_m\left(r\left(M_p\right), \theta_k^{\pm}\right) \text{ and } \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(r\left(M_p\right), \mu_k^{\pm}\right)$$

and

$$I_m\left(r\left(M_{p+1}\right), \theta_k^{\pm}\right) \text{ and } \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(r\left(M_{p+1}\right), \mu_k^{\pm}\right)$$

for all $k = 1, ..., M$, $m = 0, ..., 2M-1$ and $i = 1, ..., N_{\mathtt{d}}$. The above requirements are equivalent with the knowledge of

$$I_m\left(r\left(M_p\right), \theta_k^{\pm}\right) \text{ and } \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(r\left(M_p\right), \mu_k^{\pm}\right)$$

at all points M_p on the line of sight, that is for all $k = 1, ..., M$, $m = 0, ..., 2M-1$, $i = 1, ..., N_{\mathtt{d}}$ and $p = 1, ..., N_{\mathtt{LOS}}^{\mathtt{point}}$. Because the point M_p is situated on a radial direction which encloses the angle $\Phi_{\mathtt{sun},p}$ with the $Z_{\mathtt{sun}}$-axis and assuming that for a solar zenith angle

$$\theta_{\mathtt{sun}} = \pi - \Phi_{\mathtt{sun},p},$$

we are able to compute the intensities

$$I_m\left(r_j,\theta_k^{\pm}\right) \text{ and } \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(r_j,\mu_k^{\pm}\right)$$

for all $k=1,...,M$, $m=0,...,2M-1$, $i=1,...,N_{\mathtt{d}}$ and $j=1,...,N$ by a pseudo-spherical model. To achieve that we may use a simple interpolation procedure and derive

$$\begin{aligned} I_m\left(r\left(M_p\right),\theta_k^{\pm}\right) &= \frac{s}{\triangle r_j} I_m\left(r_{\bar{j}},\theta_k^{\pm}\right) + \left(1-\frac{s}{\triangle r_j}\right) I_m\left(r_{\bar{j}+1},\theta_k^{\pm}\right), \\ \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(r\left(M_p\right),\mu_k^{\pm}\right) &= \frac{s}{\triangle r_j} \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(r_{\bar{j}},\theta_k^{\pm}\right) + \left(1-\frac{s}{\triangle r_j}\right) \frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(r_{\bar{j}+1},\theta_k^{\pm}\right) \end{aligned}$$

where the index $\bar{j}$ is such that $r_{\bar{j}+1} \le r_{\mathtt{lev}(p)} \le r_{\bar{j}}$, $s = r_{\mathtt{lev}(p)} - r_{\bar{j}+1}$ and $\triangle r_j = r_{\bar{j}} - r_{\bar{j}+1}$. It is important to note that $I_m\left(r_j,\theta_k^{\pm}\right)$ is a 3-dimensional array, while $\frac{\partial I_m}{\partial \bar{\sigma}_{\mathtt{ext},i}}\left(r_j,\mu_k^{\pm}\right)$ is a 4-dimensional array.

8.2.5.2. General considerations

In a pseudo-spherical atmosphere the boundary-value problem for the diffuse radiance consists of the inhomogeneous differential equation

$$\mu \frac{\mathrm{d}I}{\mathrm{d}r}\left(r,\mathbf{\Omega}\right) = -\sigma_{\mathtt{ext}}\left(r\right) I\left(r,\mathbf{\Omega}\right) + J_{\mathtt{ss}}\left(r,\mathbf{\Omega}\right) + J_{\mathtt{ms}}\left(r,\mathbf{\Omega}\right), \tag{8.10}$$

the top-of-atmosphere boundary condition ($r = r_{\mathtt{TOA}}$)

$$I\left(r_{\mathtt{TOA}},\mathbf{\Omega}^{-}\right) = 0 \tag{8.11}$$

and the surface boundary condition ($r = r_{\mathtt{s}}$)

$$\begin{aligned} I\left(r_{\mathtt{s}},\mathbf{\Omega}^{+}\right) &= \epsilon\left(\Omega^{+}\right) B\left(r_{\mathtt{s}}\right) + F_{\mathtt{sun}} \frac{A}{\pi} \left|\mu_{\mathtt{sun}}\right| \rho_{\mathtt{norm}}\left(\mathbf{\Omega}^{+},\mathbf{\Omega}_{\mathtt{sun}}\right) \mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(\left|\mathbf{r}_{\mathtt{s}}-\mathbf{r}_{\mathtt{TOA}}\right|\right)} \\ &\quad + \frac{A}{\pi} \int_{2\pi} I\left(r_{\mathtt{s}},\mathbf{\Omega}^{-}\right) \left|\mu^{-}\right| \rho_{\mathtt{norm}}\left(\mathbf{\Omega}^{+},\mathbf{\Omega}^{-}\right) \mathrm{d}\Omega^{-} \end{aligned} \tag{8.12}$$

Assuming the azimuthal expansion of the diffuse radiance I_m ($\varphi_{\mathrm{sun}} = 0$)

$$I\left(r,\Omega\right) = I\left(r,\mu,\varphi\right) = \sum_{m=0}^{2M-1} I_m\left(r,\mu\right) \cos m\varphi,$$

the radiative transfer equation can be expressed in the discrete ordinate space as:

$$\begin{aligned} &\mu_k^{\pm} \frac{\mathrm{d}I_m}{\mathrm{d}r}\left(r,\mu_k^{\pm}\right) \\ &= -\sigma_{\mathtt{ext}}(r) I_m\left(r,\mu_k^{\pm}\right) + \delta_{m0}\sigma_{\mathtt{abs}}(r) B\left(r\right) \\ &+ \left(2-\delta_{m0}\right) \frac{F_{\mathtt{sun}}}{4\pi} \sigma_{\mathtt{scat}}\left(r\right) p_m\left(r,\mu_k^{\pm},\mu_{\mathtt{sun}}\right) \mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(\left|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}\right|\right)} \\ &+ \frac{1}{2} \sum_{l=1}^{M} w_l \sigma_{\mathtt{scat}}\left(r\right) \left[p_m\left(r,\mu_k^{\pm},\mu_l^{+}\right) I_m\left(r,\mu_l^{+}\right) + p_m\left(r,\mu_k^{\pm},\mu_l^{-}\right) I_m\left(r,\mu_l^{-}\right)\right], \end{aligned} \tag{8.13}$$

where p_m are the azimuthal expansion coefficients of the scattering phase function

$$P\left(r,\Omega,\Omega'\right) = \sum_{m=0}^{2M-1} \left(2-\delta_{m0}\right) p_m\left(r,\mu,\mu'\right) \cos\left[m\left(\varphi-\varphi'\right)\right].$$

In equation (8.13) μ is the cosine of the zenith angle; our convention is that $\mu = 1$ for upward radiation and $\mu = -1$ for downward radiation. The set $\left\{\mu_k^+, w_k\right\}_{k=\overline{1,M}}$ is the set of Gauss-Legendre quadrature points and weights in the interval $(0,1)$, while $\left\{\mu_k^-, w_k\right\}_{k=\overline{1,M}}$ with $\mu_k^- = -\mu_k^+$ is the set of quadrature points and weights in the interval $(-1,0)$. Defining the radiance vector function in the discrete ordinate space by

$$\mathbf{i}_m(r) = \begin{bmatrix} \mathbf{i}_m^+(r) \\ \mathbf{i}_m^-(r) \end{bmatrix}$$

with $\left[\mathbf{i}_m^{\pm}(r)\right]_k = I_m\left(r, \mu_k^{\pm}\right)$, $k = 1, ..., M$ we are led to the linear system of differential equations

$$\frac{\mathrm{d}\mathbf{i}_m}{\mathrm{d}r}(r) = \mathbf{A}_m(r)\,\mathbf{i}_m(r) + \mathbf{b}_m(r). \tag{8.14}$$

The entries of the layer matrix

$$\mathbf{A}_m(r) = \begin{bmatrix} \mathbf{A}_m^{11}(r) & \mathbf{A}_m^{12}(r) \\ \mathbf{A}_m^{21}(r) & \mathbf{A}_m^{22}(r) \end{bmatrix}$$

are given by

$$\begin{aligned}
\left[\mathbf{A}_m^{11}(r)\right]_{kl} &= \frac{1}{2\mu_k^+}\left[w_l\sigma_{\mathtt{scat}}(r)\,p_m\left(r,\mu_k^+,\mu_l^+\right) - 2\sigma_{\mathtt{ext}}(r)\delta_{kl}\right], \\
\left[\mathbf{A}_m^{12}(r)\right]_{kl} &= \frac{1}{2\mu_k^+}w_l\sigma_{\mathtt{scat}}(r)\,p_m\left(r,\mu_k^+,\mu_l^-\right), \\
\left[\mathbf{A}_m^{21}(r)\right]_{kl} &= \frac{1}{2\mu_k^-}w_l\sigma_{\mathtt{scat}}(r)\,p_m\left(r,\mu_k^-,\mu_l^+\right), \\
\left[\mathbf{A}_m^{22}(r)\right]_{kl} &= \frac{1}{2\mu_k^-}\left[w_l\sigma_{\mathtt{scat}}(r)\,p_m\left(r,\mu_k^-,\mu_l^-\right) - 2\sigma_{\mathtt{ext}}(r)\delta_{kl}\right]
\end{aligned}$$

for all $k, l = 1, ..., M$. The source vector is decomposed into a solar and a thermal contribution

$$\mathbf{b}_m(r) = \mathbf{b}_{\mathtt{sun},m}(r)\,\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}(|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}|)} + \mathbf{b}_{\mathtt{th},m}(r),$$

where

$$\left[\mathbf{b}_{\mathtt{sun},m}^{\pm}(r)\right]_k = \frac{1}{\mu_k^{\pm}}(2-\delta_{m0})\,\frac{F_{\mathtt{sun}}}{4\pi}\sigma_{\mathtt{scat}}(r)\,p_m\left(r,\mu_k^{\pm},\mu_{\mathtt{sun}}\right)$$

and

$$\left[\mathbf{b}_{\mathtt{th},m}^{\pm}(r)\right]_k = \frac{1}{\mu_k^{\pm}}\delta_{m0}\sigma_{\mathtt{abs}}(r)B(r)$$

for all $k = 1, ..., M$.

8.2.5.3. Integral form of the layer equation

Let us consider a discretization of the atmosphere in N levels: $r_1 > r_2 > ... > r_N$ with the convention $r_1 = r_{\mathtt{TOA}}$ and $r_N = r_{\mathtt{s}}$, where $r_{\mathtt{s}}$ is the lowest (surface) point of the atmosphere. A layer j is bounded above by the level r_j and below by the level r_{j+1}; the number of layers is $N-1$. The optical coefficients and the phase function are assumed to be constant within each layer and for the layer j with geometrical thickness $\triangle r_j = r_j - r_{j+1}$ we denote by $\bar{\mathbf{A}}_{m,j}$ the average value of $\mathbf{A}_m(r)$. In this regard equation (8.14) reduces to a linear system of differential equations with constant coefficients

$$\frac{\mathrm{d}\mathbf{i}_m}{\mathrm{d}r}(\rho) = \bar{\mathbf{A}}_{m,j}\mathbf{i}_m(\rho) + \mathbf{b}_m(\rho), \tag{8.15}$$

where the layer coordinate ρ is defined by $\rho = r - r_{j+1}$ and $0 \leq \rho \leq \triangle r_j$. Solving equation (8.15) with the boundary condition

$$\mathbf{i}_m(0) = \mathbf{i}_{m,j+1}$$

yields

$$\mathbf{i}_{m,j} = \mathrm{e}^{\bar{\mathbf{A}}_{m,j}\triangle r_j}\mathbf{i}_{m,j+1} + \int_0^{\triangle r_j} \mathrm{e}^{\bar{\mathbf{A}}_{m,j}(\triangle r_j - \rho)}\mathbf{b}_m(\rho)\,\mathrm{d}\rho \tag{8.16}$$

with $\mathbf{i}_{m,j}$ and $\mathbf{i}_{m,j+1}$ being the radiances at the boundary levels j and $j+1$, respectively. Using the property of the matrix exponential: $\mathrm{e}^{a\mathbf{A}}\mathrm{e}^{b\mathbf{A}} = \mathrm{e}^{(a+b)\mathbf{A}}$, which holds true for two arbitrary scalars a and b, we left multiply the above equation by the matrix $\mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\triangle r_j}$ and obtain the integral form of the layer equation

$$\mathbf{i}_{m,j+1} = \mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\triangle r_j}\mathbf{i}_{m,j} - \int_0^{\triangle r_j} \mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\rho}\mathbf{b}_m(\rho)\,\mathrm{d}\rho. \tag{8.17}$$

8.2.5.4. Layer equation

In principle, the exponential of a matrix can be computed by methods involving the matrix eigenvalues, approximation theory, differential equations and the matrix characteristic polynomial. In the present analysis we concentrate on the matrix eigenvalue method and the Padé approximation. Furthermore, the computation of the integral of the matrix exponential in equation (8.17) requires a parametrization of the source term within each layer. In this regard, the thermal layer vector $\mathbf{b}_{\mathtt{th},m}(\rho)$ is linearly interpolated between the level values $\bar{\mathbf{b}}_{\mathtt{th},m,j}$ and $\bar{\mathbf{b}}_{\mathtt{th},m,j+1}$,

$$\mathbf{b}_{\mathtt{th},m}(\rho) = \frac{\rho}{\triangle r_j}\bar{\mathbf{b}}_{\mathtt{th},m,j} + \left(1 - \frac{\rho}{\triangle r_j}\right)\bar{\mathbf{b}}_{\mathtt{th},m,j+1},$$

although a higher-order polynomial approximation can also be considered. The solar layer vector is computed accordingly to the average secant approximation, that is,

$$\mathbf{b}_{\mathtt{sun},m}(\rho)\,\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}(|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}|)} = \bar{\mathbf{b}}_{\mathtt{sun},m,j}\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}(\rho)}$$

and

$$\tau_{\mathtt{ext}}^{\mathtt{sun}}(\rho) = \frac{\rho}{\triangle r_j}\tau_{\mathtt{ext},j}^{\mathtt{sun}} + \left(1 - \frac{\rho}{\triangle r_j}\right)\tau_{\mathtt{ext},j+1}^{\mathtt{sun}},$$

where $\bar{\mathbf{b}}_{\mathtt{sun},m,j}$ is a layer quantity, $\tau_{\mathtt{ext},j}^{\mathtt{sun}}$ and $\tau_{\mathtt{ext},j+1}^{\mathtt{sun}}$ are the solar optical depths at the boundary levels j and $j+1$. For most practical applications encountered in atmospheric remote sensing, the accuracy of the average secant approximation is reasonable; for extreme situations involving optically thick layers and high solar zenith angles, improved exponential-polynomial parametrization can also be employed.

8.2.5.5. Matrix eigenvalue method

In the eigenvalue solution method the matrix exponential and the integral terms involving matrix exponentials are computed by using the spectral decomposition of the matrix $\bar{\mathbf{A}}_{m,j}$. This method is especially efficient due to the special structure of the matrix $\bar{\mathbf{A}}_{m,j}$, written for convenience as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ -\mathbf{A}^{12} & -\mathbf{A}^{11} \end{bmatrix}.$$

The eigenvalues of $\mathbf{A}$ are real and occur in pairs, in which case the order of the algebraic eigenvalue problem can be reduced by a factor of 2. The steps of computing the eigenstate of the matrix $\mathbf{A}$ can be summarized as follows:

1. Compute $\mathcal{A}^+ = \mathbf{A}^-\mathbf{A}^+$, where $\mathbf{A}^+ = \mathbf{A}^{11} + \mathbf{A}^{12}$ and $\mathbf{A}^- = \mathbf{A}^{11} - \mathbf{A}^{12}$, and determine the eigenstate $\left\{\mu_k, \mathbf{w}_k^+\right\}_{k=\overline{1,M}}$ of the matrix $\mathcal{A}^+$.
2. Normalize the vectors $\mathbf{w}_k^+$ for $k = 1, ..., M$.

3. Compute the eigenvectors of the matrix $\mathcal{A}^- = \mathbf{A}^+\mathbf{A}^-$

$$\mathbf{w}_k^- = \frac{1}{\sqrt{\mu_k}}\mathbf{A}^+\mathbf{w}_k^+, \;\; k = 1, ..., M. \tag{8.18}$$

4. Set

$$\mathbf{v}_k^+ = \frac{1}{2}\left(\mathbf{w}_k^+ + \mathbf{w}_k^-\right), \;\; \mathbf{v}_k^- = \frac{1}{2}\left(\mathbf{w}_k^+ - \mathbf{w}_k^-\right), \;\; k = 1, ..., M.$$

5. Construct the eigenvectors of $\mathbf{A}$ as

$$\bar{\mathbf{v}}_k^+ = \begin{bmatrix} \mathbf{v}_k^+ \\ \mathbf{v}_k^- \end{bmatrix}, \;\; \bar{\mathbf{v}}_k^- = \begin{bmatrix} \mathbf{v}_k^- \\ \mathbf{v}_k^+ \end{bmatrix}, \;\; k = 1, ..., M.$$

It should be remarked that if $\mathbf{A}\bar{\mathbf{v}}_k^+ = \lambda_k^+\bar{\mathbf{v}}_k^+$ with

$$\lambda_k^+ = \sqrt{\mu_k}, \;\; k = 1, ..., M,$$

that is,

$$\begin{bmatrix} \mathbf{A}^{11}\mathbf{v}_k^+ + \mathbf{A}^{12}\mathbf{v}_k^- \\ -\mathbf{A}^{12}\mathbf{v}_k^+ - \mathbf{A}^{11}\mathbf{v}_k^- \end{bmatrix} = \lambda_k^+ \begin{bmatrix} \mathbf{v}_k^+ \\ \mathbf{v}_k^- \end{bmatrix},$$

then there holds

$$\mathbf{A}\bar{\mathbf{v}}_k^- = \begin{bmatrix} \mathbf{A}^{11}\mathbf{v}_k^- + \mathbf{A}^{12}\mathbf{v}_k^+ \\ -\mathbf{A}^{12}\mathbf{v}_k^- - \mathbf{A}^{11}\mathbf{v}_k^+ \end{bmatrix} = -\lambda_k^+ \begin{bmatrix} \mathbf{v}_k^- \\ \mathbf{v}_k^+ \end{bmatrix} = -\lambda_k^+\bar{\mathbf{v}}_k^-.$$

Also note that the normalization condition imposed on $\mathbf{w}_k^+$ is required for an explicit analytic determination of the partial derivatives of the radiance field. The spectral decomposition of the matrix $\mathbf{A}$ is then

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}$$

with

$$\mathbf{V} = \left[\bar{\mathbf{v}}_1^+, ..., \bar{\mathbf{v}}_M^+, \bar{\mathbf{v}}_1^-, ..., \bar{\mathbf{v}}_M^-\right]$$

and

$$\Lambda = \begin{bmatrix} \lambda_1^+ & ... & 0 & 0 & ... & 0 \\ & ... & & & & \\ 0 & ... & \lambda_M^+ & 0 & ... & 0 \\ 0 & ... & 0 & -\lambda_1^+ & ... & 0 \\ & & & & ... & \\ 0 & ... & 0 & 0 & ... & -\lambda_M^+ \end{bmatrix} \stackrel{\text{def}}{=} \text{diag}\left[\lambda_k^+; -\lambda_k^+\right].$$

Returning to our conventional notation, that is, setting $\bar{\mathbf{A}}_{m,j}$ for $\mathbf{A}$ and $\mathbf{V}_{m,j}$ for $\mathbf{V}$, we express the matrix exponential in equation (8.17) as

$$\mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\triangle r_j} = \mathbf{V}_{m,j}\Lambda_{m,j}^0\mathbf{V}_{m,j}^{-1}, \tag{8.19}$$

where the diagonal matrix $\Lambda_{m,j}^0$ is given by

$$\Lambda_{m,j}^0 = \text{diag}\left[a_0\left(\lambda_k^+\triangle r_j\right); a_0\left(-\lambda_k^+\triangle r_j\right)\right]$$

with

$$a_0\left(x\right) = \mathrm{e}^{-x}.$$

To compute the integrals involving the source term, we use the basic result

$$\int \mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\rho} f\left(\rho\right) \mathrm{d}\rho = \mathbf{V}_{m,j}\left\{\int \text{diag}\left[a_0\left(\lambda_k^+\rho\right); a_0\left(-\lambda_k^+\rho\right)\right] f\left(\rho\right) \mathrm{d}\rho\right\}\mathbf{V}_{m,j}^{-1}$$

with f being some scalar function of ρ and obtain

$$\begin{aligned}&\int_0^{\triangle r_j} \mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\rho}\mathbf{b}_m\left(\rho\right)\,\mathrm{d}\rho\\&=\mathbf{V}_{m,j}\Lambda^1_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{th},m,j}\right)+\mathbf{V}_{m,j}\Lambda^2_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{th},m,j+1}\right)\\&+\mathbf{V}_{m,j}\Lambda^3_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{sun},m,j}\right).\end{aligned}\tag{8.20}$$

The diagonal matrices entering in equation (8.20) are given by

$$\begin{aligned}\Lambda^1_{m,j}&=\mathrm{diag}\left[a_1\left(\lambda^+_k\triangle r_j\right);a_1\left(-\lambda^+_k\triangle r_j\right)\right],\\\Lambda^2_{m,j}&=\mathrm{diag}\left[a_2\left(\lambda^+_k\triangle r_j\right);a_2\left(-\lambda^+_k\triangle r_j\right)\right],\\\Lambda^3_{m,j}&=\mathrm{diag}\left[b_1\left(\lambda^+_k\triangle r_j\right);b_1\left(-\lambda^+_k\triangle r_j\right)\right],\end{aligned}\tag{8.21}$$

while the interpolation functions read as

$$\begin{aligned}a_1\left(x\right)&=\int_0^1\xi\mathrm{e}^{-x\xi}\,\mathrm{d}\xi=\frac{1-\left(1+x\right)\mathrm{e}^{-x}}{x^2},\\a_2\left(x\right)&=\int_0^1\left(1-\xi\right)\mathrm{e}^{-x\xi}\,\mathrm{d}\xi=\frac{x-1+\mathrm{e}^{-x}}{x^2},\\b_1\left(x\right)&=\int_0^1\mathrm{e}^{-\left[\xi\tau^{\mathtt{sun}}_{\mathtt{ext},j}+(1-\xi)\tau^{\mathtt{sun}}_{\mathtt{ext},j+1}\right]}\mathrm{e}^{-x\xi}\,\mathrm{d}\xi=\frac{\mathrm{e}^{-\left(\tau^{\mathtt{sun}}_{\mathtt{ext},j}+x\right)}-\mathrm{e}^{-\tau^{\mathtt{sun}}_{\mathtt{ext},j+1}}}{\tau^{\mathtt{sun}}_{\mathtt{ext},j+1}-\tau^{\mathtt{sun}}_{\mathtt{ext},j}-x}.\end{aligned}$$

Inserting equations (8.19) and (8.20) into equation (8.17), we establish a first representation of the layer equation

$$\begin{aligned}\mathbf{i}_{m,j+1}&=\mathbf{V}_{m,j}\Lambda^0_{m,j}\mathbf{V}^{-1}_{m,j}\mathbf{i}_{m,j}-\mathbf{V}_{m,j}\Lambda^1_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{th},m,j}\right)\\&-\mathbf{V}_{m,j}\Lambda^2_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{th},m,j+1}\right)-\mathbf{V}_{m,j}\Lambda^3_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{sun},m,j}\right).\end{aligned}\tag{8.22}$$

For negative eigenvalues of the matrix $\bar{\mathbf{A}}_{m,j}$ the arguments of the exponential functions in equation (8.21) are positive and as a result the layer equation (8.22) is numerically unstable. This instability can be circumvented if we left multiply equation (8.22) by $\mathbf{V}^{-1}_{m,j}$ and scale the resulting equation by the diagonal matrix

$$\mathbf{D}_{m,j}=\mathrm{diag}\left[1;a_0\left(\lambda^+_k\triangle r_j\right)\right].$$

We then obtain

$$\begin{aligned}&\mathbf{D}_{m,j}\mathbf{V}^{-1}_{m,j}\mathbf{i}_{m,j+1}\\&=\bar{\Lambda}^0_{m,j}\mathbf{V}^{-1}_{m,j}\mathbf{i}_{m,j}-\bar{\Lambda}^1_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{th},m,j}\right)\\&-\bar{\Lambda}^2_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{th},m,j+1}\right)-\bar{\Lambda}^3_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j\bar{\mathbf{b}}_{\mathtt{sun},m,j}\right)\end{aligned}\tag{8.23}$$

with $\bar{\Lambda}^p_{m,j}=\mathbf{D}_{m,j}\Lambda^p_{m,j}$ for all $p=0,...,3$. Accounting of the identities

$$\mathrm{e}^{-x}a_1\left(-x\right)=a_2\left(x\right),\;\;\mathrm{e}^{-x}a_2\left(-x\right)=a_1\left(x\right),$$

we find the following numerically stable representations of the diagonal matrices in equation (8.23):

$$\begin{aligned}\bar{\Lambda}^0_{m,j}&=\mathrm{diag}\left[a_0\left(\lambda^+_k\triangle r_j\right);1\right],\\\bar{\Lambda}^1_{m,j}&=\mathrm{diag}\left[a_1\left(\lambda^+_k\triangle r_j\right);a_2\left(\lambda^+_k\triangle r_j\right)\right],\\\bar{\Lambda}^2_{m,j}&=\mathrm{diag}\left[a_2\left(\lambda^+_k\triangle r_j\right);a_1\left(\lambda^+_k\triangle r_j\right)\right],\\\bar{\Lambda}^3_{m,j}&=\mathrm{diag}\left[b_1\left(\lambda^+_k\triangle r_j\right);b_2\left(\lambda^+_k\triangle r_j\right)\right]\end{aligned}\tag{8.24}$$

with

$$b_2(x) = \mathrm{e}^{-x} b_1(-x) = \frac{\mathrm{e}^{-\left(\tau^{\mathtt{sun}}_{\mathtt{ext},j+1}+x\right)} - \mathrm{e}^{-\tau^{\mathtt{sun}}_{\mathtt{ext},j}}}{\tau^{\mathtt{sun}}_{\mathtt{ext},j} - \tau^{\mathtt{sun}}_{\mathtt{ext},j+1} - x}.$$

In compact notations the scaled layer equation (8.23) can be expressed as

$$\mathfrak{A}^1_{m,j}\mathbf{i}_{m,j} + \mathfrak{A}^2_{m,j}\mathbf{i}_{m,j+1} = \mathfrak{b}_{m,j},$$

where the layer quantities are given by

$$\begin{aligned}
\mathfrak{A}^1_{m,j} &= \bar{\Lambda}^0_{m,j}\mathbf{V}^{-1}_{m,j}, \\
\mathfrak{A}^2_{m,j} &= -\mathbf{D}_{m,j}\mathbf{V}^{-1}_{m,j}, \\
\mathfrak{b}_{m,j} &= \bar{\Lambda}^1_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j \bar{\mathbf{b}}_{\mathtt{th},m,j}\right) + \bar{\Lambda}^2_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j \bar{\mathbf{b}}_{\mathtt{th},m,j+1}\right) \\
&\quad + \bar{\Lambda}^3_{m,j}\mathbf{V}^{-1}_{m,j}\left(\triangle r_j \bar{\mathbf{b}}_{\mathtt{sun},m,j}\right).
\end{aligned}$$

It should be remarked that the scaling process has a symmetry effect on the layer matrices $\mathfrak{A}^1_{m,j}$ and $\mathfrak{A}^2_{m,j}$: the matrices $\bar{\Lambda}^0_{m,j}$ and $\mathbf{D}_{m,j}$ have a "complementary" diagonal structure.

8.2.5.6. Padé approximation

For optically thin layers the Padé approximation to the matrix exponential simplifies the algorithm implementation. Essentially, the p^{th} diagonal Padé approximation to the exponential of the matrix $\mathbf{A}$ is given by

$$\mathbf{Q}_p^{-1}(-\mathbf{A})\,\mathbf{Q}_p(\mathbf{A}),$$

where $\mathbf{Q}_p(\mathbf{X})$ is a polynomial in $\mathbf{X}$ of degree p defined by

$$\mathbf{Q}_p(\mathbf{X}) = \sum_{k=0}^{p} c_k \mathbf{X}^k, \quad c_k = \frac{(2p-k)!p!}{(2p)!k!(p-k)!}.$$

For an efficient application of the Padé approximation and under the assumption of optically thin layers, we suppose a linear approximation of the source term

$$\mathbf{b}_m(\rho) = \frac{\rho}{\triangle r_j}\bar{\mathbf{b}}_{m,j} + \left(1 - \frac{\rho}{\triangle r_j}\right)\bar{\mathbf{b}}_{m,j+1}$$

with the radiance level values being given by

$$\begin{aligned}
\bar{\mathbf{b}}_{m,j} &= \bar{\mathbf{b}}_{\mathtt{th},m,j} + \bar{\mathbf{b}}_{\mathtt{sun},m,j}\mathrm{e}^{-\tau^{\mathtt{sun}}_{\mathtt{ext},j}}, \\
\bar{\mathbf{b}}_{m,j+1} &= \bar{\mathbf{b}}_{\mathtt{th},m,j+1} + \bar{\mathbf{b}}_{\mathtt{sun},m,j}\mathrm{e}^{-\tau^{\mathtt{sun}}_{\mathtt{ext},j+1}}.
\end{aligned}$$

In this context the first-order Padé approximation

$$\mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\triangle r_j} \approx \left(\mathbf{I} + \frac{\triangle r_j}{2}\bar{\mathbf{A}}_{m,j}\right)^{-1}\left(\mathbf{I} - \frac{\triangle r_j}{2}\bar{\mathbf{A}}_{m,j}\right)$$

together with integral approximations

$$\frac{1}{\triangle r_j}\int_0^{\triangle r_j} \frac{\rho}{\triangle r_j}\mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\rho}\,\mathrm{d}\rho \approx \frac{1}{2}\mathbf{I} - \frac{\triangle r_j}{3}\bar{\mathbf{A}}_{m,j} + \frac{\triangle r_j^2}{8}\bar{\mathbf{A}}^2_{m,j} \tag{8.25}$$

and

$$\frac{1}{\triangle r_j}\int_0^{\triangle r_j}\left(1 - \frac{\rho}{\triangle r_j}\right)\mathrm{e}^{-\bar{\mathbf{A}}_{m,j}\rho}\,\mathrm{d}\rho \approx \frac{1}{2}\mathbf{I} - \frac{\triangle r_j}{6}\bar{\mathbf{A}}_{m,j} + \frac{\triangle r_j^2}{24}\bar{\mathbf{A}}^2_{m,j} \tag{8.26}$$

give the layer equation

$$\mathfrak{A}^1_{m,j}\mathbf{i}_{m,j} + \mathfrak{A}^2_{m,j}\mathbf{i}_{m,j+1} = \mathfrak{b}_{m,j}$$

with

$$\begin{aligned}
\mathfrak{A}^1_{m,j} &= \mathbf{I} - \frac{\triangle r_j}{2}\bar{\mathbf{A}}_{m,j}, \\
\mathfrak{A}^2_{m,j} &= -\left(\mathbf{I} + \frac{\triangle r_j}{2}\bar{\mathbf{A}}_{m,j}\right), \\
\mathfrak{b}_{m,j} &= \left(\frac{1}{2}\mathbf{I} - \frac{\triangle r_j}{12}\bar{\mathbf{A}}_{m,j}\right)\left(\triangle r_j \bar{\mathbf{b}}_{m,j}\right) \\
&\quad + \left(\frac{1}{2}\mathbf{I} + \frac{\triangle r_j}{12}\bar{\mathbf{A}}_{m,j}\right)\left(\triangle r_j \bar{\mathbf{b}}_{m,j+1}\right).
\end{aligned}$$

The integral approximations (8.25) and (8.26) have been obtained by considering a Taylor expansion of the matrix exponential function, while the layer equation has been derived by left multiplying equation (8.17) with $\mathbf{I} + (\triangle r_j/2)\,\bar{\mathbf{A}}_{m,j}$. As in the case of the matrix eigenvalue method, this process leads to a symmetric form of the layer matrices $\mathfrak{A}^1_{m,j}$ and $\mathfrak{A}^2_{m,j}$.

The second-order Padé approximation

$$\exp\left\{-\bar{\mathbf{A}}_{m,j}\triangle r_j\right\} \approx \left(\mathbf{I} + \frac{\triangle r_j}{2}\bar{\mathbf{A}}_{m,j} + \frac{\triangle r_j^2}{6}\bar{\mathbf{A}}^2_{m,j}\right)^{-1}\left(\mathbf{I} - \frac{\triangle r_j}{2}\bar{\mathbf{A}}_{m,j} + \frac{\triangle r_j^2}{6}\bar{\mathbf{A}}^2_{m,j}\right)$$

in conjunction with the integral approximations (8.25) and (8.26) yield a similar equation with the layer quantities

$$\begin{aligned}
\mathfrak{A}^1_{m,j} &= \mathbf{I} - \frac{\triangle r_j}{2}\bar{\mathbf{A}}_{m,j} + \frac{\triangle r_j^2}{6}\bar{\mathbf{A}}^2_{m,j} \\
\mathfrak{A}^2_{m,j} &= -\left(\mathbf{I} + \frac{\triangle r_j}{2}\bar{\mathbf{A}}_{m,j} + \frac{\triangle r_j^2}{6}\bar{\mathbf{A}}^2_{m,j}\right), \\
\mathfrak{b}_{m,j} &= \left(\frac{1}{2}\mathbf{I} - \frac{\triangle r_j}{12}\bar{\mathbf{A}}_{m,j} + \frac{\triangle r_j^2}{24}\bar{\mathbf{A}}^2_{m,j}\right)\left(\triangle r_j \bar{\mathbf{b}}_{m,j}\right) \\
&\quad + \left(\frac{1}{2}\mathbf{I} + \frac{\triangle r_j}{12}\bar{\mathbf{A}}_{m,j} + \frac{\triangle r_j^2}{24}\bar{\mathbf{A}}^2_{m,j}\right)\left(\triangle r_j \bar{\mathbf{b}}_{m,j+1}\right).
\end{aligned}$$

The algorithm based on the first-order Padé approximation is essentially equivalent to the finite-difference method.

8.2.5.7. System matrix of the entire atmosphere

The radiative transfer problem must be solved subject to the boundary conditions at the top of the atmosphere and at the lower surface of the atmosphere. At the top of the atmosphere the down-welling diffuse radiation vanishes and the boundary condition can be expressed in matrix form as

$$[\mathbf{0}, \mathbf{I}]\,\mathbf{i}_{m,1} = \mathbf{0}, \tag{8.27}$$

where $\dim([\mathbf{0}, \mathbf{I}]) = M \times 2M$. The surface boundary condition also possesses a matrix representation

$$[\mathbf{I}, \mathbf{R}_m]\,\mathbf{i}_{m,N} = \mathbf{r}_m, \tag{8.28}$$

where the entries of the reflection matrix and of the reflection vector are given by

$$[\mathbf{R}_m]_{kl} = -2Aw_l\left|\mu_l^-\right|\rho_m^{\mathtt{norm}}\left(\mu_k^+, \mu_l^-\right) \tag{8.29}$$

and

$$[\mathbf{r}_m]_k = \delta_{m0}\epsilon\left(\mu_k^+\right)B\left(r_{\mathtt{s}}\right) + \left(2-\delta_{m0}\right)A\frac{F_{\mathtt{sun}}}{\pi}\left|\mu_{\mathtt{sun}}\right|\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(\left|r_{\mathtt{s}}-r_{\mathtt{TOA}}\right|\right)}\rho_m^{\mathtt{norm}}\left(\mu_k^+,\mu_{\mathtt{sun}}\right), \tag{8.30}$$

respectively. In equations (8.29) and (8.30) ϵ is the surface emissivity, A is the surface albedo and $\rho_m^{\mathtt{norm}}$ are azimuthal expansion coefficients of the normalized bi-directional reflection function.

Assembling the layer equations together with the boundary conditions (8.27) and (8.28) into the global matrix of the entire atmosphere, leads to the matrix equation

$$\mathfrak{A}_m\mathbf{i}_m = \mathfrak{b}_m$$

with $\dim\left(\mathfrak{A}_m\right) = 2MN \times 2MN$. The organization of the system matrix $\mathfrak{A}_m$ and of the source vector $\mathfrak{b}_m$ is shown in Table 1. As in the conventional approach the matrix $\mathfrak{A}_m$ has $3M-1$ sub- and super-diagonals and it may be compressed into band-storage and then inverted using standard methods.

	$\mathfrak{A}_m$						$\mathfrak{b}_m$
row\column	$2M$	$2M$	$2M$	...	$2M$	$2M$	
M	$[\mathbf{0},\mathbf{I}]$	$\mathbf{0}$	$\mathbf{0}$	...	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$
$2M$	$\mathfrak{A}_{m,1}^1$	$\mathfrak{A}_{m,1}^2$	$\mathbf{0}$	...	$\mathbf{0}$	$\mathbf{0}$	$\mathfrak{b}_{m,1}$
$2M$	$\mathbf{0}$	$\mathfrak{A}_{m,2}^1$	$\mathfrak{A}_{m,2}^2$	...	$\mathbf{0}$	$\mathbf{0}$	$\mathfrak{b}_{m,2}$
...	...	...	...	..	...	...	...
$2M$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	...	$\mathfrak{A}_{m,N-1}^1$	$\mathfrak{A}_{m,N-1}^2$	$\mathfrak{b}_{m,N-1}$
M	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	...	$\mathbf{0}$	$[\mathbf{I},\mathbf{R}_m]$	$\mathbf{r}_m$

Table 8.1.: *Organization of the system matrix and of the source vector for the entire atmosphere.*

8.2.5.8. Weighting functions calculation

The retrieval of atmospheric constituents from satellite measurements requires the knowledge of the weighting functions, i.e. the partial derivatives of the measured radiance with respect to the atmospheric parameters being retrieved. The process of obtaining the set of partial derivatives which constitute the matrix of weighting functions is commonly referred to as linearisation analysis.

The radiance measured by a satellite instrument can be expressed in the framework of the source integration technique in terms of the solution of the radiative transfer equation In this regard the derivatives calculation can be performed by linearising the radiative transfer equation with respect to the desired parameters.

The signal measured by a satellite instrument can be modelled by integrating the radiative transfer equation along the line of sight. For a nadir viewing geometry with the line of sight bounded by the surface point S and the point at the top of the atmosphere A, the diffuse radiance in the measurement direction $\mathbf{\Omega}_{\mathtt{m}}$ can be expressed as

$$I\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right) = I_{\mathtt{ss}}\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right) + I_{\mathtt{ms}}\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right), \tag{8.31}$$

where

$$I_{\mathtt{ss}}\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right) = I_{\mathtt{ss}}\left(\mathbf{r}_{\mathtt{s}},\mathbf{\Omega}_{\mathtt{m}}\right)\mathrm{e}^{-\tau_{\mathtt{ext}}\left(\left|\mathbf{r}_A-\mathbf{r}_{\mathtt{s}}\right|\right)} + \int_{\left|\mathbf{r}_A-\mathbf{r}_{\mathtt{s}}\right|} J_{\mathtt{ss}}\left(\mathbf{r},\mathbf{\Omega}_{\mathtt{m}}\right)\mathrm{e}^{-\tau_{\mathtt{ext}}\left(\left|\mathbf{r}_A-\mathbf{r}\right|\right)}\mathrm{d}s \tag{8.32}$$

and

$$I_{\mathtt{ms}}\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right) = I_{\mathtt{ms}}\left(\mathbf{r}_{\mathtt{s}},\mathbf{\Omega}_{\mathtt{m}}\right)\mathrm{e}^{-\tau_{\mathtt{ext}}\left(\left|\mathbf{r}_A-\mathbf{r}_{\mathtt{s}}\right|\right)} + \int_{\left|\mathbf{r}_A-\mathbf{r}_{\mathtt{s}}\right|} J_{\mathtt{ms}}\left(\mathbf{r},\mathbf{\Omega}_{\mathtt{m}}\right)\mathrm{e}^{-\tau_{\mathtt{ext}}\left(\left|\mathbf{r}_A-\mathbf{r}\right|\right)}\mathrm{d}s \tag{8.33}$$

are the single and the multiple scattering contributions, respectively, and

$$\tau_{\mathtt{ext}}\left(|\mathbf{r}_1-\mathbf{r}_2|\right)=\int_{|\mathbf{r}_1-\mathbf{r}_2|}\sigma_{\mathtt{ext}}\left(\mathbf{r}'\right)\,\mathrm{d}s'$$

is the extinction optical depth between the points $\mathbf{r}_1$ and $\mathbf{r}_2$. The single and the multiple scattering source functions are defined by

$$J_{\mathtt{ss}}\left(\mathbf{r},\mathbf{\Omega}\right) = F_{\mathtt{sun}}\frac{\sigma_{\mathtt{scat}}\left(\mathbf{r}\right)}{4\pi}P\left(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}_{\mathtt{sun}}\right)\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(|\mathbf{r}-\mathbf{r}_{\mathtt{TOA}}|\right)}, \tag{8.34}$$

$$J_{\mathtt{ms}}\left(\mathbf{r},\mathbf{\Omega}\right) = \frac{\sigma_{\mathtt{scat}}\left(\mathbf{r}\right)}{4\pi}\int_{4\pi}P\left(\mathbf{r},\mathbf{\Omega},\mathbf{\Omega}'\right)I\left(\mathbf{r},\mathbf{\Omega}'\right)\mathrm{d}\Omega', \tag{8.35}$$

while the surface values of the radiance fields are given by

$$I_{\mathtt{ss}}\left(\mathbf{r}_{\mathtt{s}},\mathbf{\Omega}^{+}\right) = F_{\mathtt{sun}}\frac{A}{\pi}\left|\mu_{\mathtt{sun}}\right|\rho_{\mathtt{norm}}\left(\mathbf{\Omega}^{+},\mathbf{\Omega}_{\mathtt{sun}}\right)\mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(|\mathbf{r}_{\mathtt{s}}-\mathbf{r}_{\mathtt{TOA}}|\right)}, \tag{8.36}$$

$$I_{\mathtt{ms}}\left(\mathbf{r}_{\mathtt{s}},\mathbf{\Omega}^{+}\right) = \frac{A}{\pi}\int_{2\pi}I\left(\mathbf{r}_{\mathtt{s}},\mathbf{\Omega}^{-}\right)\left|\mu^{-}\right|\rho_{\mathtt{norm}}\left(\mathbf{\Omega}^{+},\mathbf{\Omega}^{-}\right)\,\mathrm{d}\Omega^{-}. \tag{8.37}$$

For a limb viewing geometry the boundary conditions (8.36) and (8.37) have to be replaced by homogeneous boundary conditions.

For a numerical computation of the path integrals in equations (8.32) and (8.33) the atmosphere is discretized in homogeneous layers. In this context the measured radiance $I\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right)$ can be regarded as a function of some vector parameters, e.g. the layer values of the extinction and the scattering coefficients or of some scalar parameters, e.g. the surface albedo. If ς_i is an atmospheric parameter in the layer i, then equation (8.31) yields

$$\frac{\partial I}{\partial\varsigma_i}\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right)=\frac{\partial I_{\mathtt{ss}}}{\partial\varsigma_i}\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right)+\frac{\partial I_{\mathtt{ms}}}{\partial\varsigma_i}\left(\mathbf{r}_A,\mathbf{\Omega}_{\mathtt{m}}\right).$$

The single scattering radiance satisfies the radiative transfer equation with the source function (8.34) and the boundary condition (8.36). Integrating the differential equation along all paths of the line of sight bounded by adjacent layers, we are led to a recurrence relation for the single scattering radiance. This recurrence relation depends analytically on the layer values of the optical coefficients and on the surface albedo. As a result a recurrence relation for the partial derivative of the single scattering radiance can be derived in a straightforward manner.

In view of equations (8.33), (8.35) and (8.37) the multiple scattering radiance can be computed on a recursive basis, which in turn leads to a recurrence relation for the partial derivatives $\partial I_{\mathtt{ms}}/\partial\varsigma_i$ along the boundary points of the line of sight. This recursion requires the knowledge of the partial derivative of the radiance field in a plane-parallel atmosphere and at all levels j, $\partial I_j/\partial\varsigma_i$. To compute the level quantities $\partial I_j/\partial\varsigma_i$ we will use the radiance solution computed in the framework of the discrete ordinate method with matrix exponential.

The matrix exponential formalism operates with the concept of the layer equation

$$\mathfrak{A}^1_{m,j}\mathbf{i}_{m,j}+\mathfrak{A}^2_{m,j}\mathbf{i}_{m,j+1}=\mathfrak{b}_{m,j}, \tag{8.38}$$

which relates the level values of the radiance field

$$\mathbf{i}_{m,j}=\begin{bmatrix}\mathbf{i}^{+}_{m,j}\\ \mathbf{i}^{-}_{m,j}\end{bmatrix}$$

with $\left[\mathbf{i}^{\pm}_{m,j}\left(r\right)\right]_k=I_m\left(r_j,\pm\mu_k\right)$, $k=1,...,M$.

The layer equation together with the boundary conditions at the top and the bottom of the atmosphere are assembled into the global matrix of the entire atmosphere and the solution of the resulting system of equations yields the level values of the radiance field.

To compute the partial derivative with respect to the atmospheric parameter ς_i we linearise the layer equation (8.38) and obtain

$$\mathfrak{A}^1_{m,j}\frac{\partial \mathbf{i}_{m,j}}{\partial \varsigma_i} + \mathfrak{A}^2_{m,j}\frac{\partial \mathbf{i}_{m,j+1}}{\partial \varsigma_i} = \frac{\partial \mathfrak{b}_{m,j}}{\partial \varsigma_i} - \frac{\partial \mathfrak{A}^1_{m,j}}{\partial \varsigma_i}\mathbf{i}_{m,j} - \frac{\partial \mathfrak{A}^2_{m,j}}{\partial \varsigma_i}\mathbf{i}_{m,j+1}. \tag{8.39}$$

As for radiance calculation all sets of derivative layer equations (8.39) are assembled into a global system of equations for the entire atmosphere. It is worth to notice, that the system matrix for derivative calculation coincides with the system matrix for radiance calculation; only the right-hand sides are different. In order to increase the efficiency of the method we compute the partial derivatives with respect to all atmospheric parameters of interest ς_i, $i = 1, ..., N_{\mathtt{d}}$, that is, we solve a system of equations with multiple right-hand sides. To compute the partial derivatives of the layer quantities $\mathfrak{A}^1_{m,j}$, $\mathfrak{A}^2_{m,j}$ and $\mathfrak{b}_{m,j}$ we apply the chain rule. In the case of Padé approximation, the derivatives calculation is trivial, but in the case of the matrix eigenvalue method we are faced with the calculation of the partial derivatives of the eigenvector matrix $\mathbf{V}^{-1}_{m,j}$ and of the eigenvalues λ_k.

To compute $\partial \mathbf{V}^{-1}_{m,j}/\partial \varsigma_i$ and $\partial \lambda_k/\partial \varsigma_i$ we consider the eigenvalue problem for the matrix $\mathcal{A}^+$

$$\mathcal{A}^+ \mathbf{w}^+_k = \mu_k \mathbf{w}^+_k.$$

For a fixed discrete ordinate index k we take the derivative with respect to ς_i and obtain

$$\frac{\partial \mathcal{A}^+}{\partial \varsigma_i}\mathbf{w}^+_k + \mathcal{A}^+\frac{\partial \mathbf{w}^+_k}{\partial \varsigma_i} = \frac{\partial \mu_k}{\partial \varsigma_i}\mathbf{w}^+_k + \mu_k\frac{\partial \mathbf{w}^+_k}{\partial \varsigma_i}. \tag{8.40}$$

Equation (8.40) is a system of M equations with $M+1$ unknowns: the scalar $\partial \mu_k/\partial \varsigma_i$ and the vector $\partial \mathbf{w}^+_k/\partial \varsigma_i$. Since the eigenvectors $\mathbf{w}^+_k$ are normalized, we derive an additional equation

$$\mathbf{w}^{+T}_k \frac{\partial \mathbf{w}^+_k}{\partial \varsigma_i} = 0, \tag{8.41}$$

which implies the compatibility of the system of equations. Using equations (8.40) and (8.41) the resulting system of equations can be written in matrix form as

$$\begin{bmatrix} \mathbf{w}^+_k & \mu_k\mathbf{I} - \mathcal{A}^+ \\ 0 & \mathbf{w}^{+T}_k \end{bmatrix} \begin{bmatrix} \frac{\partial \mu_k}{\partial \varsigma_i} \\ \frac{\partial \mathbf{w}^+_k}{\partial \varsigma_i} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathcal{A}^+}{\partial \varsigma_i}\mathbf{w}^+_k \\ 0 \end{bmatrix}.$$

It is important to observe that we can solve the above system of equations for all atmospheric parameters ς_i, $i = 1, ..., N_{\mathtt{d}}$, that is, we can solve the matrix equation

$$\begin{bmatrix} \mathbf{w}^+_k & \mu_k\mathbf{I} - \mathcal{A}^+ \\ 0 & \mathbf{w}^{+T}_k \end{bmatrix} \begin{bmatrix} \frac{\partial \mu_k}{\partial \varsigma_1} & \cdots & \frac{\partial \mu_k}{\partial \varsigma_{N_{\mathtt{d}}}} \\ \frac{\partial \mathbf{w}^+_k}{\partial \varsigma_1} & \cdots & \frac{\partial \mathbf{w}^+_k}{\partial \varsigma_{N_{\mathtt{d}}}} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathcal{A}^+}{\partial \varsigma_1}\mathbf{w}^+_k & \cdots & \frac{\partial \mathcal{A}^+}{\partial \varsigma_{N_{\mathtt{d}}}}\mathbf{w}^+_k \\ 0 & \cdots & 0 \end{bmatrix}.$$

If $\partial \mu_k/\partial \varsigma_i$ is known, the partial derivative of $\lambda^+_k = \sqrt{\mu_k}$ with respect to ς_i follows immediately

$$\frac{\partial \lambda^+_k}{\partial \varsigma_i} = \frac{1}{2\lambda^+_k}\frac{\partial \mu_k}{\partial \varsigma_i}.$$

To compute the partial derivative of $\mathbf{w}^-_k$ we use the definition (8.18) and apply the chain rule to obtain

$$\frac{\partial \mathbf{w}^-_k}{\partial \varsigma_i} = \left(-\frac{1}{\lambda^{+2}_k}\frac{\partial \lambda_k}{\partial \varsigma_i}\right)\mathbf{A}^+\mathbf{w}^+_k + \frac{1}{\lambda^+_k}\frac{\partial \mathbf{A}^+}{\partial \varsigma_i}\mathbf{w}^+_k + \frac{1}{\lambda^+_k}\mathbf{A}^+\frac{\partial \mathbf{w}^+_k}{\partial \varsigma_i}.$$

Further calculations give

$$\frac{\partial \mathbf{v}^+_k}{\partial \varsigma_i} = \frac{1}{2}\left(\frac{\partial \mathbf{w}^+_k}{\partial \varsigma_i} + \frac{\partial \mathbf{w}^-_k}{\partial \varsigma_i}\right), \quad \frac{\partial \mathbf{v}^-_k}{\partial \varsigma_i} = \frac{1}{2}\left(\frac{\partial \mathbf{w}^+_k}{\partial \varsigma_i} - \frac{\partial \mathbf{w}^-_k}{\partial \varsigma_i}\right)$$

and

$$\frac{\partial \bar{\mathbf{v}}_k^+}{\partial \varsigma_i} = \begin{bmatrix} \frac{\partial \mathbf{v}_k^+}{\partial \varsigma_i} \\ \frac{\partial \mathbf{v}_k^-}{\partial \varsigma_i} \end{bmatrix}, \quad \frac{\partial \bar{\mathbf{v}}_k^-}{\partial \varsigma_i} = \begin{bmatrix} \frac{\partial \mathbf{v}_k^-}{\partial \varsigma_i} \\ \frac{\partial \mathbf{v}_k^+}{\partial \varsigma_i} \end{bmatrix}.$$

Using now the definition of the $\mathbf{V}_{m,j}$ matrix we obtain

$$\frac{\partial \mathbf{V}_{m,j}}{\partial \varsigma_i} = \left[\frac{\partial \bar{\mathbf{v}}_1^+}{\partial \varsigma_i}, ..., \frac{\partial \bar{\mathbf{v}}_M^+}{\partial \varsigma_i}, \frac{\partial \bar{\mathbf{v}}_1^-}{\partial \varsigma_i}, ..., \frac{\partial \bar{\mathbf{v}}_M^-}{\partial \varsigma_i} \right];$$

whence, taking into account that $\mathbf{V}_{m,j}\mathbf{V}_{m,j}^{-1} = \mathbf{I}$, we end up with

$$\frac{\partial \mathbf{V}_{m,j}^{-1}}{\partial \varsigma_i} = -\mathbf{V}_{m,j}^{-1} \frac{\partial \mathbf{V}_{m,j}}{\partial \varsigma_i} \mathbf{V}_{m,j}^{-1}.$$

The calculation of the partial derivatives for $N_{\mathtt{d}}$ atmospheric parameters and a specific solar zenith angle requires the solution of a system of equations with $N_{\mathtt{d}} + 1$ right-hand sides; while, in the matrix eigenvalue method, a system of equations with $M + 1$ unknowns has to be solved additionally for each layer.

8.2.6. Picard iteration

Different methods have been developed to solve the multiple scattering problem in a spherical atmosphere. These include an order of scattering solution method, finite difference method in conjunction with a combined differential-integral approach and the Monte Carlo method. In this Section we analyse several versions of the Picard iteration for solving the radiative transfer equation in a spherical atmosphere.

To compute the limb radiance at the top of the atmosphere we use the integral form of the radiative transfer equation

$$I(\mathbf{r}, \Omega) = I(\mathbf{r}_{\mathtt{r}}, \Omega) \, \mathrm{e}^{-\tau_{\mathtt{ext}}(|\mathbf{r}-\mathbf{r}_{\mathtt{r}}|)} + I_{\mathtt{ss}}(\mathbf{r}, \mathbf{r}_{\mathtt{r}}, \Omega) + I_{\mathtt{ms}}(\mathbf{r}, \mathbf{r}_{\mathtt{r}}, \Omega), \tag{8.42}$$

where $\mathbf{r}_{\mathtt{r}}$ is a reference point and derive a recurrence relation for the diffuse radiance at a set of discrete points along the line of sight. The recurrence relation then takes the form

$$I(\mathbf{r}_p, \Omega) = I(\mathbf{r}_{p+1}, \Omega) \exp(-\tau_p) + I_{\mathtt{ss}}(\mathbf{r}_p, \mathbf{r}_{p+1}, \Omega) + I_{\mathtt{ms}}(\mathbf{r}_p, \mathbf{r}_{p+1}, \Omega), \tag{8.43}$$

where $\{\mathbf{r}_p\}_{p=1}^{N_p}$ is the set of intersection points of the line of sight with a sequence of spherical surfaces. In a single scattering model the multiple scattering contribution $I_{\mathtt{ms}}$ is neglected, while in a multiple scattering model this term has to be included in the computation. Our further analysis is focused on the computation of the multiple scattering contribution $I_{\mathtt{ms}}$, which in turns requires the computation of the diffuse radiance at the limb points $\mathbf{r}_p$, $p = 1, 2, ..., N_p$.

We choose a global coordinate system by directing the Z-axis along the solar direction, since this choice leads to an axis-symmetric radiation field. At each point on the line of sight the local coordinate system is chosen as the local spherical coordinate system. The polar angles θ and φ of the direction Ω are specified in the local coordinate system. By convention the first quadrant of the local coordinate system corresponds to $\theta \in (0, \pi/2)$ and $\varphi \in (-\pi/2, \pi/2)$, the second quadrant to $\theta \in (0, \pi/2)$ and $\varphi \in (\pi/2, 3\pi/2)$, the third quadrant to $\theta \in (\pi/2, \pi)$ and $\varphi \in (\pi/2, 3\pi/2)$ and finally, the fourth quadrant to $\theta \in (\pi/2, \pi)$ and $\varphi \in (-\pi/2, \pi/2)$.

Because the problem is axis-symmetric, i.e. $I(r, \Theta, \Psi, \Omega) = I(r, \Theta, \Omega)$, a two-dimensional Picard iteration can be used to compute the radiances at a set of discrete points in the azimuthal plane $\Psi = 0$. The domain of analysis is shown in Fig. 8.8 and is given by:

$$D = \{(r, \Theta, \Psi) / r \in [r_{\mathtt{s}}, r_{\mathtt{TOA}}], \Theta \in [0, \Theta_{\mathtt{shd}}], \Psi = 0\},$$

where $r_{\mathtt{s}}$ is the lowest (surface) point of the atmosphere, $r_{\mathtt{TOA}}$ is the radius at the top of the atmosphere and $\Theta_{\mathtt{shd}}$ is the polar angle at which the shadow region begins. Along the radial line $\Theta = 0$ we consider axis-symmetric boundary conditions

$$I(r, \Theta = 0, \theta, \varphi) = I(r, \Theta = 0, \theta, \varphi + \pi), \varphi \in (-\pi/2, \pi/2), \tag{8.44}$$

while along the radial line $\Theta = \Theta_{\mathtt{shd}}$ we assume homogeneous boundary conditions (for the radiances pointing into the domain)

$$I\left(r, \Theta = \Theta_{\mathtt{shd}}, \theta, \varphi\right) = 0, \; \varphi \in (\pi/2, 3\pi/2)\,. \tag{8.45}$$

The domain of analysis is discretized in N_r optically homogeneous spherical shells with radii r_i in decreasing order, i.e. $r_1 = r_{\mathtt{TOA}} > r_2 > ... > r_{N_r} = r_{\mathtt{s}}$, while the zenith direction is discretized in N_Φ equidistant radial lines. The local direction of the radiance in each grid point is discretized in $N_\theta N_\varphi$ discrete ordinates, where N_θ is the number of zenith directions and N_φ is the number of azimuthal directions.

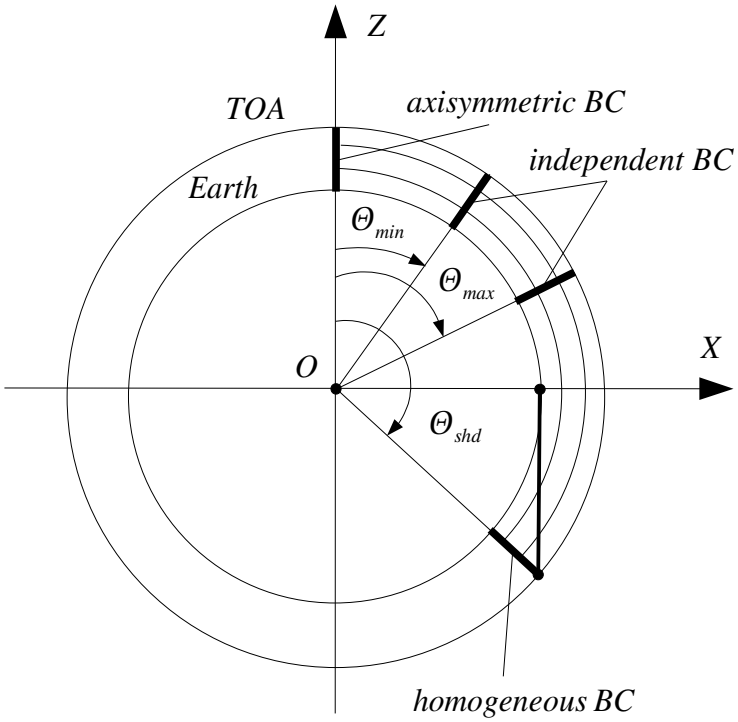

Figure 8.8.: *Domain of analysis and boundary conditions.*

The radiative transfer equation

$$\frac{\mathrm{d}I}{\mathrm{d}s}\left(\mathbf{r}, \Omega\right) = -\sigma_{\mathtt{ext}}\left(\mathbf{r}\right) I\left(\mathbf{r}, \Omega\right) + J_{\mathtt{ss}}\left(\mathbf{r}, \Omega\right) + J_{\mathtt{ms}}\left(\mathbf{r}, \Omega\right)$$

with

$$\begin{aligned} J_{\mathtt{ss}}\left(\mathbf{r}, \Omega\right) &= \frac{F_{\mathtt{sun}}}{4\pi} \mathcal{S}\left(\mathbf{r}, \Omega, \Omega_{\mathtt{sun}}\right) \mathrm{e}^{-\tau_{\mathtt{ext}}^{\mathtt{sun}}\left(|\mathbf{r} - \mathbf{r}_{\mathtt{TOA}}(\mathbf{r})|\right)}, \\ J_{\mathtt{ms}}\left(\mathbf{r}, \Omega\right) &= \frac{1}{4\pi} \int_{4\pi} \mathcal{S}\left(\mathbf{r}, \Omega, \Omega'\right) I\left(\mathbf{r}, \Omega'\right) \mathrm{d}\Omega' \end{aligned}$$

is of the form $I = f\left(I\right)$ and the Picard iteration technique based on a fixed-point iteration is appropriate for its numerical solution: if the sequence $I^{(n+1)} = f\left(I^{(n)}\right)$ converges to I and the function f is continuous, then there holds $f\left(I^{(n)}\right) \to f\left(I\right)$, which, in turn, yields $I = f\left(I\right)$. Thus, at the iteration step n the recurrence relation for computing the radiance at the generic point $\mathbf{r}$ and along the characteristic Ω read as

$$I^{(n+1)}\left(\mathbf{r}, \Omega\right) = I^{(n+1)}\left(\mathbf{r}_{\mathtt{r}}, \Omega\right) \mathrm{e}^{-\tau_{\mathtt{ext}}\left(|\mathbf{r} - \mathbf{r}_{\mathtt{r}}|\right)} + I_{\mathtt{ss}}\left(\mathbf{r}, \mathbf{r}_{\mathtt{r}}, \Omega\right) + I_{\mathtt{ms}}^{(n)}\left(\mathbf{r}, \mathbf{r}_{\mathtt{r}}, \Omega\right), \tag{8.46}$$

where essentially $\mathbf{r}$ stands for the grid point $\mathbf{r}_{ij} = (r_i, \Theta_j, \Psi = 0)$ with $i = 1, ..., N_r$ and $j = 1, ..., N_\Phi$ and Ω for the discrete ordinate direction $\Omega_{kl} = (\theta_k, \varphi_l)$ with $k = 1, ..., N_\theta$ and $l = 1, ..., N_\varphi$.

In the long characteristic method the reference point is the intersection point of the characteristic with the model boundary D (Fig. 8.9). The radiance at the generic grid point is computed on a recursive basis by considering the intersection of the characteristic with a set of spherical surfaces $\{\mathbf{r}_q\}_{q=1}^{N_q}$, for which it holds true that $r_1 = r_\mathbf{r} > r_2 > ... > r_{N_q} = r$. The recurrence relation is as in equation (8.43), but with p replaced by q. The radiance at the reference point $I^{(n+1)}(\mathbf{r}_\mathbf{r}, \Omega)$ is used to initialize the recursion and is specified by the model boundary radiance. The multiple scattering radiance $I_{\mathtt{ms}}^{(n)}(\mathbf{r}_q, \Omega)$ at the characteristic point $\mathbf{r}_q$ is computed by a linear interpolation in the spatial and the discrete-ordinate domains using the grid point values $I_{\mathtt{ms}}^{(n)}(\mathbf{r}_{ij}, \Omega_{kl})$. The single scattering contribution $I_{\mathtt{ss}}(\mathbf{r}_q, \mathbf{r}_{q+1}, \Omega)$ is calculated at all stages of the characteristic without any interpolation.

In the short characteristic method the reference point is the intersection point of the characteristic with the cell boundary (Fig. 8.9). The radiance at the generic point is computed accordingly to equation (8.46), where the radiance at the reference point $I^{(n+1)}(\mathbf{r}_\mathbf{r}, \Omega)$ is determined via a linear interpolation of the grid point values of the face pierced by the characteristic. As before the multiple scattering radiance $I_{\mathtt{ms}}^{(n)}(\mathbf{r}_\mathbf{r}, \Omega)$ is computed by linear interpolation, while the single scattering contribution $I_{\mathtt{ss}}(\mathbf{r}, \mathbf{r}_\mathbf{r}, \Omega)$ is calculated in an exact manner.

The number of iterations required to attain convergence can be reduced by an appropriate choice of the initial estimate. For each radial line $\Theta_j \leq \Theta_{\mathtt{lim}}$, where $\Theta_{\mathtt{lim}} \lessapprox 90°$, we compute the initial radiance field by using a one-dimensional model. Because the calculations are performed individually for each radial line (or solar zenith angle), this model is equivalent to the independent pixel approximation for three-dimensional radiative transfer in clouds.

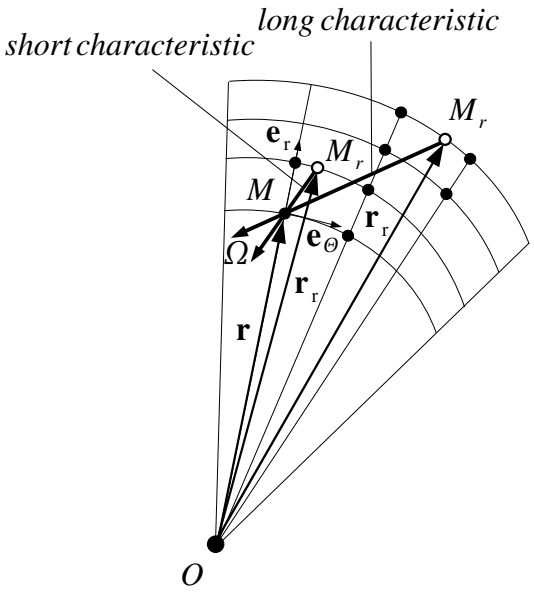

Figure 8.9.: *Long and short characteristics.*

The short and the long characteristic Picard iteration involves the following steps:

1. Consider a two-dimensional discretization of the domain of analysis D.
2. Compute the initial estimate $I(\mathbf{r}_{ij}, \Omega_{kl})$ at all radial lines $\Theta_j \leq \Theta_{\mathtt{lim}}$ by using the independent pixel approximation and initialize the radiances at all radial lines $\Theta_{\mathtt{shd}} \geq \Theta_j > \Theta_{\mathtt{lim}}$ to zero.
3. Determine the single scattering contribution $I_{\mathtt{ss}}$ at all grid points and in all discrete ordinate directions.
4. Calculate the grid values of the multiple scattering radiance $I_{\mathtt{ms}}(\mathbf{r}_{ij}, \Omega_{kl})$.

5. Beginning downward for each layer $i = 1, ..., N_r - 1$ compute the radiances at the lower nodes for all discrete ordinates in the third and fourth quadrant and at all radial lines. For the boundary cells take into account the boundary conditions (8.44) and (8.45).
6. Compute the upward radiances at the bottom of the domain by using the boundary conditions for a Lambertian surface.
7. Working upwards repeat the calculations of step 5, but compute the radiances at the upper nodes for all discrete ordinates in the first and second quadrant.
8. Interpolate the radiances at each point $\mathbf{r}_p$ of the line of sight by using the grid radiance values of the cell containing the point.
9. Compute the multiple scattering contribution $I_{\mathtt{ms}}(\mathbf{r}_p, \mathbf{r}_{p+1}, \Omega)$ at all stages of the line of sight.
10. Repeat steps 4-9 until the maximum change in the multiple scattering contributions is smaller than some preassigned error values at all stages of the line of sight.

Steps 4-7 represent one Picard iteration. A specific feature of the above method is that the grid values of the multiple scattering radiance $I_{\mathtt{ms}}(\mathbf{r}_{ij}, \Omega_{kl})$ are computed before starting the downward and the upward recurrences, and are implicitly determined by the grid radiance values $I(\mathbf{r}_{ij}, \Omega_{kl})$ at the previous Picard iteration. This technique works well for the long characteristic method, but yields a slow convergence rate for the short characteristic method. To remedy this drawback we use an iterative scheme, in which we update the multiple scattering radiances at each layer calculation. For example, in the case of the downward recurrence we first compute the multiple scattering radiances at the top of the layer and then proceed to calculate the radiances at the bottom of the layer. The multiple scattering radiances then include the values of the grid radiances in the third and fourth quadrant at the actual Picard iteration. By convention this version of the Picard iteration will be called the accelerated short characteristic method.

A further reduction of the computation time can be achieved by using the following approximate model: instead of solving the radiative transfer problem on the domain D, we solve the problem on the reduced domain

$$D_{\mathtt{r}} = \{(r, \Theta, \Psi) / r \in [r_{\mathtt{s}}, r_{\mathtt{TOA}}], \Theta \in [\Theta_{\mathtt{min}}, \Theta_{\mathtt{max}}], \Psi = 0\}$$

with $\Theta_{\mathtt{min}} < \Theta_p < \Theta_{\mathtt{max}}$ for all $p = 1, 2, ..., N_p$. The reduced domain of analysis is illustrated in Fig. 8.8. The values of the diffuse radiance pointing into the domain (the boundary values along the radial lines $\Theta = \Theta_{\mathtt{min}}$ and $\Theta = \Theta_{\mathtt{max}}$) are computed by using the pseudo-spherical model and remain unchanged during the iterative process. Thus, the boundary lines become independent of the interior cells and the radiance at their grid points serves as the boundary radiance for the interior points. This type of boundary conditions is known as open boundaries. Therefore this method will be referred to as the Picard iteration method with open boundaries.

8.3. Inversion methods

Most of the inverse problems arising in atmospheric remote sensing are non-linear. In this chapter we discuss the practical aspects of Tikhonov regularization for solving the non-linear problem

$$\mathbf{F}(\mathbf{x}) = \mathbf{y}, \tag{8.47}$$

where, due to the complexity of the radiative transfer, the forward model is computed by using a numerical model. The non-linear system of equations (8.47) is called a discrete ill-posed problem because the underlying continuous problem is ill-posed. If we accept a characterization of ill-posedness via linearisation, the condition number of the Jacobian matrix $\mathbf{K}(\mathbf{x}) = \mathbf{F}'(\mathbf{x})$ may serve as a quantification of ill-posedness. The right-hand side $\mathbf{y}$ of (8.47) is supposed to be contaminated by measurement errors and we have the representation

$$\mathbf{y}^{\delta} = \mathbf{y} + \delta,$$

where $\mathbf{y}^{\delta}$ is the noisy data and δ is the noise vector. In a deterministic setting the error is characterized by the noise level Δ, while in a semi-stochastic setting δ is assumed to be a discrete white noise vector with the covariance matrix $\mathbf{C}_{\delta} = \sigma^2 \mathbf{I}_m$.

8.3.1. Tikhonov regularization

The formulation of Tikhonov regularization for non-linear problems is straightforward: the equation (8.47) is replaced by a minimization problem involving the objective function

$$\mathcal{F}(\mathbf{x}) = \frac{1}{2}\left[\left\|\mathbf{F}(\mathbf{x}) - \mathbf{y}^{\delta}\right\|^2 + \alpha \left\|\mathbf{L}(\mathbf{x} - \mathbf{x}_{\mathtt{a}})\right\|^2\right]. \tag{8.48}$$

For a positive regularization parameter minimizers of the Tikhonov function always exists (but are not unique) and a global minimizer $\mathbf{x}_{\alpha}^{\delta}$ is called a regularized solution.

In this Section we review appropriate optimization methods for minimizing the Tikhonov function, discuss practical algorithms for computing the new iterate and characterize the solution error. Finally, we analyse the performances of the Tikhonov regularization with a priori, a posteriori and error-free parameter choice methods.

8.3.1.1. Inversion models

For retrieval problems in the visible spectral region we consider two inversion models. The first inversion model is the radiance model

$$R_{\mathtt{meas}}(\lambda, h) \approx P_{\mathtt{ampl}}\left(\lambda, \mathbf{p}_{\mathtt{ampl}}, h\right) R_{\mathtt{sim}}(\lambda, \mathbf{x}, h), \tag{8.49}$$

where λ is the wavelength, $P_{\mathtt{ampl}}$ is a polynomial of low order with coefficients $\mathbf{p}_{\mathtt{ampl}}$ and R stands for the "scan-ratioed" radiance ratio, that is, the radiance profile normalized with respect to a reference tangent height

$$R(\cdot, h) = \frac{I(\cdot, h)}{I(\cdot, h_{\mathtt{ref}})}. \tag{8.50}$$

The normalization procedure removes the need to measure the exoatmospheric solar irradiance and the absolute instrument calibration. This is similar to the self-calibration feature of the limb occultation method, where the unattenuated solar radiation is measured outside the atmosphere. In addition, there is a reduction in the effect of the surface reflectance and clouds that can control the diffuse radiation even at high altitudes. The normalization procedure does not completely remove the effect of the surface albedo, but does reduce the accuracy to which the algorithm must model this effect. The closure term $P_{\mathtt{ampl}}$ is intended to account on the contribution from aerosols with smooth spectral signature. The second inversion model is the differential radiance model

$$\ln \bar{R}_{\mathtt{meas}}(\lambda, h) \approx \ln \bar{R}_{\mathtt{sim}}(\lambda, \mathbf{x}, h) \tag{8.51}$$

with

$$\ln \bar{R}_{\mathtt{sim}}(\lambda, \mathbf{x}, h) = \ln R_{\mathtt{sim}}(\lambda, \mathbf{x}, h) - P_{\mathtt{sim}}(\lambda, \mathbf{p}_{\mathtt{sim}}, h)$$

and

$$\ln \bar{R}_{\mathtt{meas}}(\lambda, h) = \ln R_{\mathtt{meas}}(\lambda, h) - P_{\mathtt{meas}}(\lambda, \mathbf{p}_{\mathtt{meas}}, h).$$

For any $\mathbf{x}$ the simulated smoothing polynomial $P_{\mathtt{sim}}$ of coefficients $\mathbf{p}_{\mathtt{sim}} = \mathbf{p}_{\mathtt{sim}}(\mathbf{x})$ and the measurement smoothing polynomial $P_{\mathtt{meas}}$ of coefficients $\mathbf{p}_{\mathtt{meas}}$, are defined by

$$\mathbf{p}_{\mathtt{sim}} = \arg\min_{\mathbf{p}} \left\|\ln R_{\mathtt{sim}}(\cdot, \mathbf{x}, h) - P_{\mathtt{sim}}(\cdot, \mathbf{p}, h)\right\|^2$$

and

$$\mathbf{p}_{\mathtt{meas}} = \arg\min_{\mathbf{p}} \left\|\ln R_{\mathtt{meas}}(\cdot, h) - P_{\mathtt{meas}}(\cdot, \mathbf{p}, h)\right\|^2,$$

respectively. In general,a smoothing polynomial is assumed to account on the low frequency structure due to the scattering mechanism, in which case $\ln \bar{R}$ will mainly reflects the absorption process due to the gas molecules.

The choice of the inversion model is a very important task of the retrieval process, because an appropriate formulation may considerably reduce the non-linearity of the problem. In a stochastic framework the degree of non-linearity can be examined by comparing the forward model with its linearisation within the a priori variability.

For this purpose we assume that $\mathbf{x}$ is a random vector characterized by a Gaussian a priori probability density function with mean $\mathbf{x}_\mathrm{a}$ and covariance matrix $\mathbf{C}_\mathrm{x}$. In the $\mathbf{x}$-space the ellipsoid

$$(\mathbf{x}-\mathbf{x}_\mathrm{a})^T\,\mathbf{C}_\mathrm{x}^{-1}\,(\mathbf{x}-\mathbf{x}_\mathrm{a})=1$$

represents the contour of the a priori covariance, outlining the region within which the state vector is likely to lie. Considering the linear transformation

$$\mathbf{z}=\Sigma_\mathrm{x}^{-1/2}\mathbf{V}_\mathrm{x}^T\,(\mathbf{x}-\mathbf{x}_\mathrm{a})\,,$$

where $\mathbf{C}_\mathrm{x}=\mathbf{V}_\mathrm{x}\Sigma_\mathrm{x}\mathbf{V}_\mathrm{x}^T$, we observe that in the $\mathbf{z}$-space the contour of the a priori covariance is a sphere of radius 1 centred at the origin, that is, $\mathbf{z}^T\mathbf{z}=1$. The points $\mathbf{z}_k^{\pm}=\left[\begin{array}{ccccccc}0, & ..., & \pm 1, & ..., & 0\end{array}\right]^T$ represents the intersection of the sphere with the coordinate axes and delimit the region to which the state vector belongs. In the $\mathbf{x}$-space these boundary points are given by

$$\mathbf{x}_k^{\pm}=\mathbf{x}_\mathrm{a}+\mathbf{V}_\mathrm{x}\Sigma_\mathrm{x}^{1/2}\mathbf{z}_k^{\pm}=\mathbf{x}_\mathrm{a}\pm\mathbf{c}_k,$$

where the vectors $\mathbf{c}_k$, defined by the decomposition $\mathbf{V}_\mathrm{x}\Sigma_\mathrm{x}^{1/2}=[\mathbf{c}_1,...,\mathbf{c}_n]$, represent the error patterns for the covariance matrix $\mathbf{C}_\mathrm{x}$. The size of the linearisation error

$$\mathbf{R}\,(\mathbf{x})=\mathbf{F}\,(\mathbf{x})-\mathbf{F}\,(\mathbf{x}_\mathrm{a})-\mathbf{K}\,(\mathbf{x}_\mathrm{a})\,(\mathbf{x}-\mathbf{x}_\mathrm{a})$$

can be evaluated through the quantity

$$\varepsilon_{\mathtt{lin}k}=\frac{1}{m\sigma^2}\left\|\mathbf{R}\,(\mathbf{x}_\mathrm{a}\pm\mathbf{c}_k)\right\|^2.$$

If $\varepsilon_{\mathtt{lin}k}\leq 1$ for all k, then the inverse problem is said to be linear to the accuracy of the measurements within the normal range of variation of the state. Numerical simulations have shown that the differential radiance model (8.51) yields a smaller linearisation error than the radiance model (8.49). For this reason the differential radiance model will be adopted in our simulations.

8.3.1.2. Optimization methods for the Tikhonov function

In the framework of Tikhonov regularization the regularized solution $\mathbf{x}_\alpha^\delta$ is computed by minimizing the function

$$\mathcal{F}\,(\mathbf{x})=\frac{1}{2}\left[\left\|\mathbf{F}\,(\mathbf{x})-\mathbf{y}^\delta\right\|^2+\alpha\left\|\mathbf{L}\,(\mathbf{x}-\mathbf{x}_\mathrm{a})\right\|^2\right], \tag{8.52}$$

where the factor $1/2$ has been included in order to avoid the appearance of a factor two in the derivatives. The general minimization problem can be formulated as the least-squares problem

$$\mathcal{F}\,(\mathbf{x})=\frac{1}{2}\left\|\mathbf{f}\,(\mathbf{x})\right\|^2, \tag{8.53}$$

where the augmented vector $\mathbf{f}$ is given by

$$\mathbf{f}\,(\mathbf{x})=\left[\begin{array}{c}\mathbf{F}\,(\mathbf{x})-\mathbf{y}^\delta\\ \sqrt{\alpha}\mathbf{L}\,(\mathbf{x}-\mathbf{x}_\mathrm{a})\end{array}\right].$$

The regularized solution can be computed by using optimization methods for unconstrained minimization problems. Essentially, optimization tools are iterative methods, which use the Taylor expansion to compute approximations to the objective function at all points in the immediate neighbourhood to the current iterate. For Newton–type methods the quadratic model

$$\mathcal{M}\,(\mathbf{p})=\mathcal{F}\,(\mathbf{x})+\mathbf{g}^T\,(\mathbf{x})\,\mathbf{p}+\frac{1}{2}\mathbf{p}^T\mathbf{G}\,(\mathbf{x})\,\mathbf{p} \tag{8.54}$$

is used as a reasonable approximation to the objective function. In (8.54) $\mathbf{g}$ and $\mathbf{G}$ are the gradient and the Hessian of $\mathcal{F}$, that is,

$$\mathbf{g}(\mathbf{x}) = \nabla\mathcal{F}(\mathbf{x}) = \mathbf{K}_{\mathbf{f}}^{T}(\mathbf{x})\,\mathbf{f}(\mathbf{x})$$

and

$$\mathbf{G}(\mathbf{x}) = \nabla^2\mathcal{F}(\mathbf{x}) = \mathbf{K}_{\mathbf{f}}^{T}(\mathbf{x})\,\mathbf{K}_{\mathbf{f}}(\mathbf{x}) + \mathbf{Q}(\mathbf{x}),$$

where

$$\mathbf{K}_{\mathbf{f}}(\mathbf{x}) = \mathbf{f}'(\mathbf{x}) = \begin{bmatrix} \mathbf{K}(\mathbf{x}) \\ \sqrt{\alpha}\mathbf{L} \end{bmatrix}$$

is the Jacobian matrix of $\mathbf{f}$ and

$$\mathbf{Q}(\mathbf{x}) = \sum_i f_i(\mathbf{x})\,\mathbf{G}_i(\mathbf{x})$$

with $\mathbf{G}_i = \nabla^2 f_i$ being the Hessian of f_i is the second–order derivative term. Although the objective function (8.53) can be minimized by a general method, in most circumstances the special forms of the gradient and the Hessian make it worthwhile to use methods designed specifically for least-squares problems.

Non-linear optimization methods can be categorized into two broad classes: “step-length-based methods” and “trust-region methods”.

Step–length methods

For an iterative method it is important to have a measure of progress in order to decide whether a new iterate $\mathbf{x}_{k+1\alpha}^{\delta}$ is better than the current iterate $\mathbf{x}_{k\alpha}^{\delta}$. A natural measure of progress is to require a decrease of $\mathcal{F}$ at every iteration and to impose the descent condition

$$\mathcal{F}\left(\mathbf{x}_{k+1\alpha}^{\delta}\right) < \mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right).$$

A method that impose this condition is termed a descent method. A step–length–based algorithm requires the computation of a vector $\mathbf{p}_{k\alpha}^{\delta}$ called the search direction and the calculation of a positive scalar τ_k, the step length, for which it holds that

$$\mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta} + \tau_k\mathbf{p}_{k\alpha}^{\delta}\right) < \mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right).$$

To guarantee that the objective function $\mathcal{F}$ can be reduced at the k^{th} iteration step, the search direction $\mathbf{p}_{k\alpha}^{\delta}$ should be a descent direction at $\mathbf{x}_{k\alpha}^{\delta}$, that is, the inequality

$$\mathbf{g}^{T}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\mathbf{p}_{k\alpha}^{\delta} < 0$$

should hold true.

In the steepest-descent method for general optimization the objective function is approximated by a linear model and the search direction is taken as

$$\mathbf{p}_{k\alpha}^{\delta} = -\mathbf{g}\left(\mathbf{x}_{k\alpha}^{\delta}\right).$$

The negative gradient $-\mathbf{g}\left(\mathbf{x}_{k\alpha}^{\delta}\right)$ is termed the direction of steepest descent and, evidently, the steepest-descent direction is a descent direction (unless the gradient vanishes) since

$$\mathbf{g}^{T}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\mathbf{p}_{k\alpha}^{\delta} = -\left\|\mathbf{g}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\right\|^2 < 0.$$

The convergence rate of the steepest-descent method is linear and a method yielding a super-linear convergence rate is the Newton method.

In the Newton method the objective function is approximated by the quadratic model (8.54) and the search direction $\mathbf{p}_{k\alpha}^{\delta}$, which minimizes the quadratic function, solves the so called Newton equation

$$\mathbf{G}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\mathbf{p} = -\mathbf{g}\left(\mathbf{x}_{k\alpha}^{\delta}\right). \tag{8.55}$$

If the Hessian $\mathbf{G}\left(\mathbf{x}_{k\alpha}^{\delta}\right)$ is positive definite, only one iteration is required to reach the minimum of the model function (8.54) from any starting point and we expect good convergence from Newton’s method when the

quadratic model is accurate. For a general non-linear function Newton's method converges quadratically to the minimizer $\mathbf{x}_\alpha^\delta$ if the initial guess is sufficiently close to $\mathbf{x}_\alpha^\delta$, the Hessian matrix is positive definite at $\mathbf{x}_\alpha^\delta$ and the step lengths $\{\tau_k\}$ converges to unity. Note that when $\mathbf{G}\left(\mathbf{x}_\alpha^\delta\right)$ is positive definite, the solution of (8.55) is a descent direction, since

$$\mathbf{g}^T\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{p}_{k\alpha}^\delta = -\mathbf{g}^T\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{G}^{-1}\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{g}\left(\mathbf{x}_{k\alpha}^\delta\right) < 0$$

holds true.

In the Gauss-Newton method for least–squares problems it is assumed that the first-order term $\mathbf{K}_\mathbf{f}^T\mathbf{K}_\mathbf{f}$ in the expression of the Hessian dominates the second-order term $\mathbf{Q}$. This assumption is not justified when the residuals at the solution are very large, i.e. roughly speaking, when the residual $\left\|\mathbf{f}\left(\mathbf{x}_\alpha^\delta\right)\right\|$ is comparable to the largest eigenvalue of $\mathbf{K}_\mathbf{f}^T\left(\mathbf{x}_\alpha^\delta\right)\mathbf{K}_\mathbf{f}\left(\mathbf{x}_\alpha^\delta\right)$. For small residual problems the search direction solves the equation

$$\mathbf{K}_\mathbf{f}^T\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{K}_\mathbf{f}\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{p} = -\mathbf{K}_\mathbf{f}^T\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{f}\left(\mathbf{x}_{k\alpha}^\delta\right) \tag{8.56}$$

and possesses the variational characterization

$$\mathbf{p}_{k\alpha}^\delta = \arg\min_\mathbf{p}\left\|\mathbf{K}_\mathbf{f}\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{p} + \mathbf{f}\left(\mathbf{x}_{k\alpha}^\delta\right)\right\|^2. \tag{8.57}$$

The vector that solves (8.56) or (8.64) is called the Gauss-Newton direction and if $\mathbf{K}_\mathbf{f}$ is of full column rank, then the Gauss-Newton direction is uniquely and approaches the Newton direction as $\left\|\mathbf{Q}\left(\mathbf{x}_\alpha^\delta\right)\right\|$ tends to zero. Consequently, if $\left\|\mathbf{f}\left(\mathbf{x}_\alpha^\delta\right)\right\|$ is zero and the columns of $\mathbf{K}\left(\mathbf{x}_\alpha^\delta\right)$ are linearly independent, the Gauss-Newton method can ultimately achieve a quadratic rate of convergence, despite the fact that only first derivatives are used to compute $\mathbf{p}_{k\alpha}^\delta$.

For large–residual problems the term $\left\|\mathbf{f}\left(\mathbf{x}_\alpha^\delta\right)\right\|$ is not small and the second–order term $\mathbf{Q}$ cannot be neglected. In fact, a large-residual problem is one in which the optimal residual $\left\|\mathbf{f}\left(\mathbf{x}_\alpha^\delta\right)\right\|$ is large relative to the small eigenvalues of $\mathbf{K}_\mathbf{f}^T\left(\mathbf{x}_\alpha^\delta\right)\mathbf{K}_\mathbf{f}\left(\mathbf{x}_\alpha^\delta\right)$, but not with respect to its largest eigenvalue. One possible strategy for large-residual problems is to include a quasi-Newton approximation $\bar{\mathbf{Q}}\left(\mathbf{x}_{k\alpha}^\delta\right)$ to the second–order derivative term $\mathbf{Q}\left(\mathbf{x}_{k\alpha}^\delta\right)$ and to compute the search direction by solving the equation

$$\left[\mathbf{K}_\mathbf{f}^T\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{K}_\mathbf{f}\left(\mathbf{x}_{k\alpha}^\delta\right) + \bar{\mathbf{Q}}\left(\mathbf{x}_{k\alpha}^\delta\right)\right]\mathbf{p} = -\mathbf{K}_\mathbf{f}^T\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{f}\left(\mathbf{x}_{k\alpha}^\delta\right). \tag{8.58}$$

Quasi-Newton methods are based on the idea of building up curvature information as the iterations proceed using the observed behaviour of the objective function and of the gradient. The initial approximation of the second–order derivative term is usually taken as zero and, with this choice, the first iteration of the quasi-Newton method is equivalent to an iteration of the Gauss–Newton method. After $\mathbf{x}_{k+1\alpha}^\delta$ has been computed, a new approximation of $\bar{\mathbf{Q}}\left(\mathbf{x}_{k+1\alpha}^\delta\right)$ is obtained by updating $\bar{\mathbf{Q}}\left(\mathbf{x}_{k\alpha}^\delta\right)$ to take into account the newly-acquired curvature information. An update formula read as

$$\bar{\mathbf{Q}}\left(\mathbf{x}_{k+1\alpha}^\delta\right) = \bar{\mathbf{Q}}\left(\mathbf{x}_{k\alpha}^\delta\right) + \mathbf{U}_k,$$

where the update matrix $\mathbf{U}_k$ is usually chosen as a rank–one matrix. The standard condition for updating $\bar{\mathbf{Q}}$ is known as the quasi–Newton condition and requires that the Hessian should approximate the curvature of the objective function along the change in $\mathbf{x}$ during the k^{th} iteration. The most widely used quasi-Newton scheme, which satisfies the quasi–Newton condition and possesses the property of hereditary symmetry, is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update

$$\bar{\mathbf{Q}}\left(\mathbf{x}_{k+1\alpha}^\delta\right) = \bar{\mathbf{Q}}\left(\mathbf{x}_{k\alpha}^\delta\right) - \frac{1}{\mathbf{s}_k^T\mathbf{W}\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{s}_k}\mathbf{W}\left(\mathbf{x}_{k\alpha}^\delta\right)\mathbf{s}_k\mathbf{s}_k^T\mathbf{W}\left(\mathbf{x}_{k\alpha}^\delta\right) + \frac{1}{\mathbf{h}_k^T\mathbf{s}_k}\mathbf{h}_k\mathbf{h}_k^T, \tag{8.59}$$

where

$$\mathbf{s}_k = \mathbf{x}_{k+1\alpha}^\delta - \mathbf{x}_{k\alpha}^\delta$$

is the change in $\mathbf{x}$ during the k^{th} iteration

$$\mathbf{h}_k = \mathbf{g}\left(\mathbf{x}_{k+1\alpha}^\delta\right) - \mathbf{g}\left(\mathbf{x}_{k\alpha}^\delta\right)$$

is the change in gradient and

$$\mathbf{W}\left(\mathbf{x}_{k\alpha}^{\delta}\right)=\mathbf{K}_{\mathtt{f}}^{T}\left(\mathbf{x}_{k+1\alpha}^{\delta}\right)\mathbf{K}_{\mathtt{f}}\left(\mathbf{x}_{k+1\alpha}^{\delta}\right)+\bar{\mathbf{Q}}\left(\mathbf{x}_{k\alpha}^{\delta}\right).$$

A step-length procedure is frequently included in Newton–type methods because a step of unity along the Newton direction will not necessarily reduce the objective function, even though it is the step to the minimum of the function. The main requirements of a step-length algorithm can be summarized as follows: if $\mathbf{x}$ and $\mathbf{p}$ denote the actual iterate and the search direction, respectively, then

1. the average rate of decrease from $\mathcal{F}(\mathbf{x})$ to $\mathcal{F}(\mathbf{x}+\tau\mathbf{p})$ should be at least some prescribed fraction of the initial rate of decrease in that direction

$$\mathcal{F}(\mathbf{x}+\tau\mathbf{p})\leq\mathcal{F}(\mathbf{x})+\varepsilon_{\mathtt{f}}\tau\mathbf{g}^{T}(\mathbf{x})\,\mathbf{p},\ \ \varepsilon_{\mathtt{f}}>0;$$

2. the rate of decrease of $\mathcal{F}$ in the direction $\mathbf{p}$ at $\mathbf{x}+\tau\mathbf{p}$ should be larger than some prescribed fraction of the rate of decrease in the direction $\mathbf{p}$ at $\mathbf{x}$

$$\mathbf{g}^{T}(\mathbf{x}+\tau\mathbf{p})\,\mathbf{p}\geq\varepsilon_{\mathtt{g}}\mathbf{g}^{T}(\mathbf{x})\,\mathbf{p},\ \ \varepsilon_{\mathtt{g}}>0.$$

The first condition guarantees a sufficient decreases in $\mathcal{F}$ values relative to the length of the step, while the second condition avoids too small steps relative to the initial rate of decrease of $\mathcal{F}$. The condition $\varepsilon_{\mathtt{g}}>\varepsilon_{\mathtt{f}}$ implies that both conditions can be satisfied simultaneously. In practice, the second condition is not needed because the use of a backtracking strategy avoids excessively small steps.

Since computational experience has shown the importance of taking a full step whenever possible, the modern strategy of a step-length algorithm is to start with $\tau=1$ and then, if $\mathbf{x}+\mathbf{p}$ is not acceptable, "backtrack" (reduce τ) until an acceptable $\mathbf{x}+\tau\mathbf{p}$ is found. The implemented backtracking step-length algorithm uses only condition (1) and is based on quadratic and cubic interpolation. On the first backtracking the new step length is selected as the minimizer of the quadratic interpolation function $m_{\mathtt{q}}(\tau)$, satisfying the conditions

$$m_{\mathtt{q}}(0)=\mathcal{F}(\mathbf{x}),\ \ m_{\mathtt{q}}'(0)=\mathbf{g}^{T}(\mathbf{x})\,\mathbf{p},\ \ m_{\mathtt{q}}(1)=\mathcal{F}(\mathbf{x}+\mathbf{p})$$

but being constrained to be larger than $\varepsilon_1=0.1$ of the old step length. On all subsequent backtracks the new step length is chosen by using the values of the objective function at the last two values of the step length. Essentially, if τ and $\tau_{\mathtt{prev}}$ are the last two values of the step length, the new step length is computed as the minimizer of the cubic interpolation function $m_{\mathtt{c}}(\tau)$, satisfying the conditions

$$m_{\mathtt{c}}(0)=\mathcal{F}(\mathbf{x}),\ m_{\mathtt{c}}'(0)=\mathbf{g}^{T}(\mathbf{x})\,\mathbf{p},$$

and

$$m_{\mathtt{c}}(\tau)=\mathcal{F}(\mathbf{x}+\tau\mathbf{p}),\ m_{\mathtt{c}}\left(\tau_{\mathtt{prev}}\right)=\mathcal{F}\left(\mathbf{x}+\tau_{\mathtt{prev}}\mathbf{p}\right),$$

but being constrained to be larger than $\varepsilon_1=0.1$ and smaller than $\varepsilon_2=0.5$ of the old step length.

Trust–region methods

In a trust region method the step length τ_k is taken as unity, so that the new iterate is defined by

$$\mathbf{x}_{k+1\alpha}^{\delta}=\mathbf{x}_{k\alpha}^{\delta}+\mathbf{p}_{k\alpha}^{\delta}.$$

For this reason the term "step" is often used to designate the search direction $\mathbf{p}_{k\alpha}^{\delta}$. In order to ensure that the descent condition holds, it must be necessary to compute several trial vectors before finding a satisfactory $\mathbf{p}_{k\alpha}^{\delta}$. The most common mathematical formulation of this idea computes the trial step $\mathbf{p}_{k\alpha}^{\delta}$ by solving the constrained minimization problem

$$\min_{\mathbf{p}}\mathcal{M}_k(\mathbf{p})\ \text{ subject to }\ \|\mathbf{p}\|\leq\Gamma_k,\tag{8.60}$$

where $\mathcal{M}_k$ is the quadratic model (8.54) at the current iterate $\mathbf{x}_{k\alpha}^{\delta}$ and Γ_k is the trust region radius. Assuming that the solution occurs on the boundary of the constrained region, the minimization of the Lagrangian

$$\mathcal{L}(\mathbf{p},\lambda) = \mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) + \mathbf{g}^T\left(\mathbf{x}_{k\alpha}^{\delta}\right)\mathbf{p} + \frac{1}{2}\mathbf{p}^T\mathbf{G}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\mathbf{p} + \frac{1}{2}\lambda\left(\|\mathbf{p}\|^2 - \Gamma_k^2\right)$$

yields the first–order optimality conditions

$$\left[\mathbf{G}\left(\mathbf{x}_{k\alpha}^{\delta}\right) + \lambda\mathbf{I}_n\right]\mathbf{p}_\lambda = -\mathbf{g}\left(\mathbf{x}_{k\alpha}^{\delta}\right) \tag{8.61}$$

and

$$\|\mathbf{p}_\lambda\|^2 = \Gamma_k^2. \tag{8.62}$$

Particularizing the trust-region method for general minimization to least-squares problems with a Gauss–Newton Hessian approximation, we deduce that the search direction solves the equation

$$\left[\mathbf{K}_{\mathtt{f}}^T\left(\mathbf{x}_{k\alpha}^{\delta}\right)\mathbf{K}_{\mathtt{f}}\left(\mathbf{x}_{k\alpha}^{\delta}\right) + \lambda\mathbf{I}_n\right]\mathbf{p}_\lambda = -\mathbf{K}_{\mathtt{f}}^T\left(\mathbf{x}_{k\alpha}^{\delta}\right)\mathbf{f}\left(\mathbf{x}_{k\alpha}^{\delta}\right), \tag{8.63}$$

while the Lagrange parameter λ solves the equation (8.62). For comparison with the Gauss-Newton method we note that the solution of (8.63) is a solution of the linear regularized least-squares problem

$$\mathbf{p}_{k\alpha}^{\delta} = \arg\min_{\mathbf{p}}\left[\left\|\mathbf{K}_{\mathtt{f}}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\mathbf{p} + \mathbf{f}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\right\|^2 + \lambda\|\mathbf{p}\|^2\right]. \tag{8.64}$$

If λ is zero, $\mathbf{p}_{k\alpha}^{\delta}$ is the Gauss-Newton direction as $\lambda \to \infty$, $\mathbf{p}_{k\alpha}^{\delta}$ becomes parallel to the steepest-descent direction $-\mathbf{g}\left(\mathbf{x}_{k\alpha}^{\delta}\right)$.

Generally, a trust-region algorithm uses the predictive reduction in the linearised model (8.54)

$$\Delta\mathcal{F}_{\mathtt{pred}k} = \mathcal{M}_k(0) - \mathcal{M}_k\left(\mathbf{p}_{k\alpha}^{\delta}\right) \tag{8.65}$$

and the actual reduction in the objective function

$$\Delta\mathcal{F}_k = \mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta} + \mathbf{p}_{k\alpha}^{\delta}\right) \tag{8.66}$$

to decide whether the trial step $\mathbf{p}_{k\alpha}^{\delta}$ is acceptable and how the next trust-region radius is chosen. The heuristic to update the size of the trust region usually depends on the ratio of the expected change in $\mathcal{F}$ to the predicted change

$$r_k = \frac{\Delta\mathcal{F}_k}{\Delta\mathcal{F}_{\mathtt{pred}k}}.$$

The trust-region algorithm finds a new iterate and produces a trust-region radius for the next global iteration. The algorithm starts with the calculation of the trial step $\mathbf{p}$ for the actual trust-region radius (c.f. equations (8.62) and (8.63)) and with the computation of the prospective iterate $\mathbf{x}_{\mathtt{new}} = \mathbf{x} + \mathbf{p}$ and of the objective function $F(\mathbf{x}_{\mathtt{new}})$. Then, depending on the average rate of decrease of the objective function, the following situations may appear:

1. If $\mathcal{F}(\mathbf{x}_{\mathtt{new}}) \geq \mathcal{F}(\mathbf{x}) + \varepsilon_{\mathtt{f}}\mathbf{g}^T(\mathbf{x})\mathbf{p}$, then the step is unacceptable. In this case, if the trust-region radius is too small, the algorithm terminates with $\mathbf{x}_{\mathtt{new}} = \mathbf{x}$. If not, the step length $\tau_{\min}$ is computed as the minimizer of the quadratic interpolation function $m_{\mathtt{q}}(\tau)$, satisfying the conditions

   $$m_{\mathtt{q}}(0) = \mathcal{F}(\mathbf{x}), \quad m_{\mathtt{q}}'(0) = \mathbf{g}^T(\mathbf{x})\mathbf{p}, \quad m_{\mathtt{q}}(1) = \mathcal{F}(\mathbf{x}+\mathbf{p})$$

   and the new radius is chosen as $\tau_{\min}\|\mathbf{p}\|$ but constrained to be between $\varepsilon_{1\Gamma} = 0.1$ and $\varepsilon_{2\Gamma} = 0.5$ of the old radius.

2. If $\mathcal{F}(\mathbf{x}_{\mathtt{new}}) < \mathcal{F}(\mathbf{x}) + \varepsilon_{\mathtt{f}}\mathbf{g}^T(\mathbf{x})\mathbf{p}$, then the step is acceptable and the reduction of the objective function predicted by the quadratic model

   $$\triangle\mathcal{F}_{\mathtt{pred}} = -\mathbf{g}^T(\mathbf{x})\mathbf{p} - 0.5\|\mathbf{K}_{\mathtt{f}}(\mathbf{x})\mathbf{p}\|^2$$

   is computed. If $\Delta\mathcal{F}$ and $\triangle\mathcal{F}_{\mathtt{pred}}$ agree to within a prescribed tolerance or negative curvature is indicated,

then the trust-region radius is increased and the while-iteration loop is continued. If not, $\mathbf{x}_{\mathtt{new}}$ is accepted as the new iterate and the trust-region radius is updated for the next global iteration.

Stopping criteria

In a deterministic setting the standard stopping criteria for unconstrained minimization are:

1. the relative gradient test

$$\max_i \left| \frac{\left[\mathbf{g}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\right]_i \max_i \left\{\left|\left[\mathbf{x}_{k\alpha}^{\delta}\right]_i\right|, \mathrm{typ}\left[\mathbf{x}\right]_i\right\}}{\max\left\{\left|\mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right)\right|, \mathrm{typ}\mathcal{F}\right\}} \right| \leq \varepsilon_{\mathtt{g}},$$

2. the absolute function convergence test

$$\mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) < \varepsilon_{\mathtt{fa}},$$

3. the relative function convergence test

$$\frac{\mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathcal{F}\left(\mathbf{x}_{k+1\alpha}^{\delta}\right)}{\mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right)} \leq \varepsilon_{\mathtt{fr}},$$

4. the X-convergence test

$$\frac{\max_i \left\{\left|\left[\mathbf{x}_{k+1\alpha}^{\delta}\right]_i - \left[\mathbf{x}_{k\alpha}^{\delta}\right]_i\right|\right\}}{\max_i \left\{\left|\left[\mathbf{x}_{k+1\alpha}^{\delta}\right]_i\right|, \mathrm{typ}\left[\mathbf{x}\right]_i\right\}} \leq \varepsilon_{\mathtt{x}}.$$

It should be mentioned that the problem of measuring relative change when the argument z is near zero is addressed by substituting z with $\max\left\{|z|, \mathrm{typ}z\right\}$, where $\mathrm{typ}z$ is an user's estimate of a typical magnitude of z. Also note that in the PORT optimization routines the X-convergence test is expressed as

$$\frac{\max_i \left\{\left|\left[\mathbf{x}_{k+1\alpha}^{\delta}\right]_i - \left[\mathbf{x}_{k\alpha}^{\delta}\right]_i\right|\right\}}{\max_i \left\{\left|\left[\mathbf{x}_{k+1\alpha}^{\delta}\right]_i\right| + \left|\left[\mathbf{x}_{k\alpha}^{\delta}\right]_i\right|\right\}} \leq \varepsilon_{\mathtt{x}}. \tag{8.67}$$

In a stochastic framework the absolute and relative function convergence test are formulated as

$$\chi^2\left(\mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right) \leq m$$

and

$$\frac{\chi^2\left(\mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right) - \chi^2\left(\mathbf{F}\left(\mathbf{x}_{k+1\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right)}{\chi^2\left(\mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right)} \leq \varepsilon_{\mathtt{fr}},$$

respectively, while the X-convergence test read as

$$\frac{\left(\mathbf{x}_{k+1\alpha}^{\delta} - \mathbf{x}_{k\alpha}^{\delta}\right)^T \widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\left(\mathbf{x}_{k+1\alpha}^{\delta} - \mathbf{x}_{k\alpha}^{\delta}\right)}{n} \leq \varepsilon_{\mathtt{x}},$$

where

$$\chi^2\left(\mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right) = \left[\mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right]^T \mathbf{C}_{\delta}^{-1}\left[\mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right],$$

$\mathbf{C}_{\delta}$ is the measurement error covariance and $\widehat{\mathbf{C}}_{\mathbf{x}}$ is the a posteriori covariance. If $\mathbf{C}_{\delta} = \sigma^2\mathbf{I}_m$, then the absolute function convergence test is equivalent to the discrepancy principle for the square residual norm

$$\left\|\mathbf{r}_{k\alpha}^{\delta}\right\|^2 = \left\|\mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right\|^2.$$

Because in practical applications the estimation of the covariance matrices is problematic, we suggest to use the X-convergence test as in (8.67) and the relative function convergence test as in a deterministic setting, but with the residual norm $\left\|\mathbf{r}_{k\alpha}^{\delta}\right\|^2$ in place of the Tikhonov function $\mathcal{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right)$. For a data model with white noise this test will be also valid in a stochastic setting.

8.3.1.3. Practical methods for computing the new iterate

A step-length based method for minimizing the Tikhonov function is of the form of the following model algorithm:

1. compute the search direction;
2. compute the step length (see 8.3.1.2);
3. terminate the iterative process accordingly to the X-convergence criterion or the relative function convergence test.

The step-length procedure is optional, but our experience demonstrates that this step improves the stability of the method and reduce the number of iterations. In this Section we concern with the computation of the search direction $\mathbf{p}_{k\alpha}^{\delta}$ or, more precisely, with the computation of the new iterate $\mathbf{x}_{k+1\alpha}^{\delta} = \mathbf{x}_{k\alpha}^{\delta} + \mathbf{p}_{k\alpha}^{\delta}$. Certainly, if a step-length procedure is a part of the inversion algorithm, then $\mathbf{x}_{k+1\alpha}^{\delta}$ is the prospective iterate, but we prefer to use the term "new iterate" because it is frequently encountered in remote sensing community.

Using the explicit expressions of the augmented vector $\mathbf{f}$ and of its Jacobian $\mathbf{K}_{\mathbf{f}}$, we deduce that the Gauss-Newton step $\mathbf{p}_{k\alpha}^{\delta}$ solves the equation (cf. (8.56))

$$\left(\mathbf{K}_{k\alpha}^{T}\mathbf{K}_{k\alpha} + \alpha\mathbf{L}^{T}\mathbf{L}\right)\mathbf{p} = -\mathbf{K}_{k\alpha}^{T}\left[\mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{y}^{\delta}\right] - \alpha\mathbf{L}^{T}\mathbf{L}\left(\mathbf{x}_{k\alpha}^{\delta} - \mathbf{x}_{\mathsf{a}}\right)$$

with $\mathbf{K}_{k\alpha} = \mathbf{K}\left(\mathbf{x}_{k\alpha}^{\delta}\right)$. The new iterate then takes the form

$$\mathbf{x}_{k+1\alpha}^{\delta} = \mathbf{x}_{k\alpha}^{\delta} + \mathbf{p}_{k\alpha}^{\delta} = \mathbf{x}_{\mathsf{a}} + \mathbf{K}_{k\alpha}^{\dagger}\mathbf{y}_{k}^{\delta}, \tag{8.68}$$

where

$$\mathbf{K}_{k\alpha}^{\dagger} = \left(\mathbf{K}_{k\alpha}^{T}\mathbf{K}_{k\alpha} + \alpha\mathbf{L}^{T}\mathbf{L}\right)^{-1}\mathbf{K}_{k\alpha}^{T}$$

and

$$\mathbf{y}_{k}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) + \mathbf{K}_{k\alpha}\left(\mathbf{x}_{k\alpha}^{\delta} - \mathbf{x}_{\mathsf{a}}\right) \tag{8.69}$$

are the generalized inverse and the noisy data at the iteration step k, respectively.

In order to give a more practical interpretation of the Gauss-Newton iterate (8.68), we consider a linearisation of $\mathbf{F}$ about $\mathbf{x}_{k\alpha}^{\delta}$

$$\mathbf{F}\left(\mathbf{x}\right) = \mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) + \mathbf{K}_{k\alpha}\left(\mathbf{x} - \mathbf{x}_{k\alpha}^{\delta}\right) + \mathbf{R}\left(\mathbf{x}, \mathbf{x}_{k\alpha}^{\delta}\right).$$

If $\mathbf{x}^{\dagger}$ is a solution of the equation $\mathbf{F}\left(\mathbf{x}\right) = \mathbf{y}$, then $\mathbf{x}^{\dagger}$ also solves the equation

$$\mathbf{K}_{k\alpha}\left(\mathbf{x}^{\dagger} - \mathbf{x}_{\mathsf{a}}\right) = \mathbf{y}_{k},$$

where

$$\mathbf{y}_{k} = \mathbf{y} - \mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) + \mathbf{K}_{k\alpha}\left(\mathbf{x}_{k\alpha}^{\delta} - \mathbf{x}_{\mathsf{a}}\right) - \mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{k\alpha}^{\delta}\right)$$

is the exact data at the iteration step k. Because $\mathbf{y}_{k}$ is unknown, we consider the equation

$$\mathbf{K}_{k\alpha}\left(\mathbf{x} - \mathbf{x}_{\mathsf{a}}\right) = \mathbf{y}_{k}^{\delta} \tag{8.70}$$

with $\mathbf{y}_{k}^{\delta}$ being given by (8.69). Evidently, the noise in the data is due to the measurement noise and the linearisation error and we have the representation

$$\mathbf{y}_{k}^{\delta} - \mathbf{y}_{k} = \delta + \mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{k\alpha}^{\delta}\right).$$

Because the non-linear problem is ill-posed, its linearisation is also ill-posed and we solve the linear equation (8.70) by means of Tikhonov regularization with a penalty term of the form $\Omega\left(\mathbf{x}\right) = \left\|\mathbf{L}\left(\mathbf{x} - \mathbf{x}_{\mathsf{a}}\right)\right\|$. The Tikhonov function for the linear sub-problem takes the form

$$\mathcal{F}_{1k}\left(\mathbf{x}\right) = \left\|\mathbf{y}_{k}^{\delta} - \mathbf{K}_{k\alpha}\left(\mathbf{x} - \mathbf{x}_{\mathsf{a}}\right)\right\|^{2} + \alpha\left\|\mathbf{L}\left(\mathbf{x} - \mathbf{x}_{\mathsf{a}}\right)\right\|^{2}$$

and as in (8.68) the new iterate is given by $\mathbf{x}_{k+1\alpha}^{\delta} = \mathbf{x}_{\mathsf{a}} + \mathbf{K}_{k\alpha}^{\dagger}\mathbf{y}_{k}^{\delta}$. Thus, the solution of a non-linear ill–posed problem by means of Tikhonov regularization is equivalent to the solution of a sequence of ill–posed linearisation of the forward model about the current iterate.

The new iterate can be computed by using the GSVD of the matrix pair $(\mathbf{K}_{k\alpha}, \mathbf{L})$. Although the GSVD is of great theoretical interest for analysing general-form regularization problems, it is of computational interest only for small- and medium-sized problems. The reason is that the computation of the GSVD of the matrix pair $(\mathbf{K}_{k\alpha}, \mathbf{L})$ is quite demanding; the conventional implementation requires about $2m^2n+15n^3$ operations. For practical solutions of large-scale regularization problems it is much simpler to deal with standard-from problems in which $\mathbf{L} = \mathbf{I}_n$. The regularization in standard form relies on the solution of the linear equation

$$\bar{\mathbf{K}}_{k\alpha} \triangle \bar{\mathbf{x}} = \mathbf{y}_k^\delta \tag{8.71}$$

with $\bar{\mathbf{K}}_{k\alpha} = \mathbf{K}_{k\alpha}\mathbf{L}^{-1}$ and

$$\triangle \mathbf{x} = \mathbf{x} - \mathbf{x}_\mathbf{a} = \mathbf{L}^{-1} \triangle \bar{\mathbf{x}}$$

by means of Tikhonov regularization with $\mathbf{L} = \mathbf{I}_n$. Considering the SVD of the Jacobian matrix $\bar{\mathbf{K}}_{k\alpha} = \mathbf{U}\Sigma\mathbf{V}^T$, the solution of the standard-form problem

$$\left(\bar{\mathbf{K}}_{k\alpha}^T\bar{\mathbf{K}}_{k\alpha} + \alpha\mathbf{I}_n\right) \triangle \bar{\mathbf{x}} = \bar{\mathbf{K}}_{k\alpha}^T\mathbf{y}_k^\delta \tag{8.72}$$

is given by

$$\triangle\bar{\mathbf{x}}_{k+1\alpha}^\delta = \sum_{i=1}^{n} \frac{\sigma_i}{\sigma_i^2 + \alpha} \left(\mathbf{u}_i^T\mathbf{y}_k^\delta\right) \mathbf{v}_i.$$

An efficient implementation of Tikhonov regularization for large-scale problems, which also take into account that we wish to solve (8.72) several times for various regularization parameters, is described below. In this approach the standard-form problem (8.72) is treated as a least-squares problem of the form

$$\left\| \begin{bmatrix} \bar{\mathbf{K}}_{k\alpha} \\ \sqrt{\alpha}\mathbf{I}_n \end{bmatrix} \triangle \bar{\mathbf{x}} - \begin{bmatrix} \mathbf{y}_k^\delta \\ \mathbf{0} \end{bmatrix} \right\|^2.$$

The matrix $\bar{\mathbf{K}}_{k\alpha}$ is transformed into an upper bidiagonal matrix $\mathbf{J}$

$$\bar{\mathbf{K}}_{k\alpha} = \mathbf{U} \begin{bmatrix} \mathbf{J} \\ \mathbf{0} \end{bmatrix} \mathbf{V}^T$$

by means of orthogonal transformations from the left and from the right with $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{J} \in \mathbb{R}^{n\times n}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$. The orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$ are represented by series of orthogonal transformations, which are usually stored in appropriate arrays and later used when matrix-vector multiplications, e.g. $\mathbf{U}^T\mathbf{x}$ and $\mathbf{V}\mathbf{x}$, are needed. Making the changes of variables

$$\xi = \mathbf{V}^T \triangle \bar{\mathbf{x}}$$

and

$$\mathbf{z}^\delta = \mathbf{U}^T\mathbf{y}_k^\delta, \;\; \mathbf{z}^\delta = \begin{bmatrix} \mathbf{z}_1^\delta \\ \mathbf{z}_2^\delta \end{bmatrix}, \;\; \mathbf{z}_1^\delta \in \mathbb{R}^n,$$

we are led to an equivalent minimization problem involving the objective function

$$\left\| \begin{bmatrix} \mathbf{J} \\ \sqrt{\alpha}\mathbf{I}_n \end{bmatrix} \xi - \begin{bmatrix} \mathbf{z}_1^\delta \\ \mathbf{0} \end{bmatrix} \right\|^2.$$

The above minimization problem can be solved very efficiently by means of $O(n)$ operations. Essentially, for each value of the regularization parameter we compute the QR factorization

$$\begin{bmatrix} \mathbf{J} \\ \sqrt{\alpha}\mathbf{I}_n \end{bmatrix} = \mathbf{Q}_\alpha \begin{bmatrix} \mathbf{T}_\alpha \\ \mathbf{0} \end{bmatrix} \tag{8.73}$$

by means of $2n-1$ Givens rotations, where $\mathbf{T}_\alpha \in \mathbb{R}^{n\times n}$ is an upper bidiagonal matrix and $\mathbf{Q}_\alpha \in \mathbb{R}^{2n\times 2n}$ is a

product of Givens rotations. Further, defining the vector

$$\zeta_\alpha^\delta = \mathbf{Q}_\alpha^T \begin{bmatrix} \mathbf{z}_1^\delta \\ \mathbf{0} \end{bmatrix}$$

and partitioning ζ_α^δ as

$$\zeta_\alpha^\delta = \begin{bmatrix} \zeta_{\alpha 1}^\delta \\ \zeta_{\alpha 2}^\delta \end{bmatrix}, \ \zeta_{\alpha 1}^\delta \in \mathbb{R}^n,$$

we obtain

$$\xi_\alpha^\delta = \mathbf{T}_\alpha^{-1} \zeta_{\alpha 1}^\delta$$

and finally,

$$\triangle \bar{\mathbf{x}}_{k+1\alpha}^\delta = \mathbf{V} \xi_\alpha^\delta.$$

The standard-form problem (8.72) can be formulated as the normal equation

$$\bar{\mathbf{K}}_\mathbf{f}^T \bar{\mathbf{K}}_\mathbf{f} \triangle \bar{\mathbf{x}} = \bar{\mathbf{K}}_\mathbf{f}^T \bar{\mathbf{f}} \tag{8.74}$$

with

$$\bar{\mathbf{f}} = \begin{bmatrix} \mathbf{y}_k^\delta \\ \mathbf{0} \end{bmatrix}, \ \bar{\mathbf{K}}_\mathbf{f} = \begin{bmatrix} \bar{\mathbf{K}}_{k\alpha} \\ \sqrt{\alpha}\mathbf{I}_n \end{bmatrix}$$

and iterative methods, as for instance the CGNE and the LSQR algorithms, can be used for computing the new iterate. For large-scale problems the computational efficiency can be increased by using an appropriate pre-conditioner. Note that for the normal equation (8.74) the pre-conditioner $\mathbf{M}$ should be chosen such that the condition number of $\mathbf{M}^T \bar{\mathbf{K}}_\mathbf{f}^T \bar{\mathbf{K}}_\mathbf{f} \mathbf{M}$ is small. A pre-conditioner constructed by using the close connection between the Lanczos algorithm and the conjugate gradient method is usually used. If $\bar{\mathbf{K}}_{k\alpha} = \mathbf{U}\Sigma\mathbf{V}^T$ is a singular value decomposition of the Jacobian matrix, then there holds

$$\bar{\mathbf{K}}_\mathbf{f}^T \bar{\mathbf{K}}_\mathbf{f} = \mathbf{V} \left[\operatorname{diag} \left(\sigma_i^2 + \alpha \right)_{n \times n} \right] \mathbf{V}^T$$

and for a fixed index r the preconditioned matrix can be constructed as

$$\mathbf{M} = \mathbf{V} \begin{bmatrix} \operatorname{diag} \left(\frac{1}{\sqrt{\sigma_i^2 + \alpha}} \right)_{r \times r} & \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(n-r) \times r} & \operatorname{diag} \left(\frac{1}{\sqrt{\alpha}} \right)_{(n-r) \times (n-r)} \end{bmatrix} \mathbf{V}^T.$$

We then obtain

$$\mathbf{M}^T \bar{\mathbf{K}}_\mathbf{f}^T \bar{\mathbf{K}}_\mathbf{f} \mathbf{M} = \mathbf{V} \begin{bmatrix} \operatorname{diag} (1)_{r \times r} & \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(n-r) \times r} & \operatorname{diag} \left(\frac{\sigma_i^2 + \alpha}{\alpha} \right)_{(n-r) \times (n-r)} \end{bmatrix} \mathbf{V}^T$$

and the condition number of $\mathbf{M}^T \mathbf{K}_\mathbf{f}^T \mathbf{K}_\mathbf{f} \mathbf{M}$ is $1 + \sigma_{r+1}^2 / \alpha$. If σ_{r+1}^2 is not much larger than α, then the condition number is small and very few iterations are required to compute the new iterate. Turning now to practical implementation issues we mention that iterative algorithms are coded without explicit reference to $\mathbf{M}$; only the matrix–vector product $\mathbf{Mx}$ is involved. Since

$$\mathbf{Mx} = \frac{1}{\sqrt{\alpha}} \mathbf{x} + \sum_{i=1}^{r} \left(\frac{1}{\sqrt{\sigma_i^2 + \alpha}} - \frac{1}{\sqrt{\alpha}} \right) \left(\mathbf{v}_i^T \mathbf{x} \right) \mathbf{v}_i$$

we observe that the calculation of $\mathbf{Mx}$ requires the knowledge of the first r singular values and vectors of $\bar{\mathbf{K}}_{k\alpha}$ and these quantities can be efficiently computed by the Lanczos algorithm . The steps of computing the r singular values and the right singular vectors of a $m \times n$ matrix $\mathbf{A}$ can be summarized as follows:

1. apply r steps of the Lanczos bi-diagonalization algorithm with Householder orthogonalization to produce a lower $(r+1) \times r$ bidiagonal matrix $\mathbf{B}$, a $n \times r$ matrix $\bar{\mathbf{V}}$ containing the right singular vectors and a

$m \times (r+1)$ matrix $\bar{\mathbf{U}}$ containing the left singular vectors

$$\mathbf{A}\bar{\mathbf{V}} = \bar{\mathbf{U}}\mathbf{B};$$

2. compute the QR factorization of the bidiagonal matrix $\mathbf{B}$

$$\mathbf{B} = \mathbf{Q}\begin{bmatrix}\mathbf{R}\\ \mathbf{0}\end{bmatrix},$$

where $\mathbf{Q}$ is a $(r+1) \times (r+1)$ unitary matrix and $\mathbf{R}$ is an upper $r \times r$ bidiagonal matrix;

3. compute the singular value decomposition of the bidiagonal matrix $\mathbf{R}$

$$\mathbf{R} = \mathbf{U}_{\mathrm{R}}\Sigma\mathbf{V}_{\mathrm{R}}^{T};$$

4. the first r singular values are the diagonal entries of Σ, while the corresponding right singular vectors are the columns of the $n \times r$ matrix

$$\mathbf{V} = \bar{\mathbf{V}}\mathbf{V}_{\mathrm{R}}.$$

The standard-form problem (8.72) can be also expressed as

$$\mathbf{A} \triangle \bar{\mathbf{x}} = \mathbf{b} \tag{8.75}$$

with

$$\mathbf{A} = \bar{\mathbf{K}}_{k\alpha}^{T}\bar{\mathbf{K}}_{k\alpha} + \alpha\mathbf{I}_n$$

and

$$\mathbf{b} = \bar{\mathbf{K}}_{k\alpha}^{T}\mathbf{y}_k^{\delta}.$$

The symmetric system of equations (8.75) can now be solved by using standard iterative solvers, as for instance the Conjugate Gradient Squared (CGS) or the Bi-conjugate Gradient Stabilized (Bi-CGSTAB). A relevant practical aspect is that for iterative methods the matrix $\mathbf{A}$ is never formed explicitly as only matrix-vector products with $\mathbf{A}$ and eventually with $\mathbf{A}^T$ are required. The calculation of the matrix-vector product $\mathbf{Ax}$ demands the calculation of $\bar{\mathbf{K}}_{k\alpha}^{T}\bar{\mathbf{K}}_{k\alpha}\mathbf{x}$ and $\bar{\mathbf{K}}_{k\alpha}^{T}\bar{\mathbf{K}}_{k\alpha}\mathbf{x}$ should be computed as $\bar{\mathbf{K}}_{k\alpha}^{T}\left(\bar{\mathbf{K}}_{k\alpha}\mathbf{x}\right)$ and not by forming the cross-product matrix $\bar{\mathbf{K}}_{k\alpha}^{T}\bar{\mathbf{K}}_{k\alpha}$. The reasons for avoiding explicit formation of the cross-product matrix is the loss of information due to round-off and the excessively large computational time. A right pre-conditioner for the system (8.75), i.e.

$$\mathbf{A}\mathbf{M}_{\mathrm{a}} \triangle \bar{\mathbf{x}}' = \mathbf{b}, \;\; \mathbf{M}_{\mathrm{a}} \triangle \bar{\mathbf{x}}' = \triangle\bar{\mathbf{x}}$$

can also be constructed by using the Lanczos algorithm. The right pre-conditioner is given by

$$\mathbf{M}_{\mathrm{a}} = \mathbf{V}\begin{bmatrix}\mathrm{diag}\left(\frac{1}{\sigma_i^2+\alpha}\right)_{r\times r} & \mathbf{0}_{r\times(n-r)}\\ \mathbf{0}_{(n-r)\times r} & \mathrm{diag}\left(\frac{1}{\alpha}\right)_{(n-r)\times(n-r)}\end{bmatrix}\mathbf{V}^T,$$

in which case the condition number of $\mathbf{AM}_{\mathrm{a}}$ is $1 + \sigma_{r+1}^2/\alpha$ and

$$\mathbf{M}_{\mathrm{a}}\mathbf{x} = \frac{1}{\alpha}\mathbf{x} + \sum_{i=1}^{r}\left(\frac{1}{\sigma_i^2+\alpha} - \frac{1}{\alpha}\right)\left(\mathbf{v}_i^T\mathbf{x}\right)\mathbf{v}_i.$$

8.3.1.4. Error characterisation

An important part of a retrieval is to assess the accuracy of the regularized solution by performing an error analysis. The most commonly used methods to calculate the errors in the non-linear case are based on a linearisation of the forward model and of the gradient of the Tikhonov function.

Linearisation of the forward model function

The Gauss–Newton iterate $\mathbf{x}_{k+1\alpha}^{\delta}$ is the regularized solution of the linear problem (8.70) and its expression is given by (8.68). For the exact data vector $\mathbf{y}$ the Gauss–Newton iterate possesses a similar representation

$$\mathbf{x}_{k+1\alpha} - \mathbf{x}_{\mathrm{a}} = \mathbf{K}_{k\alpha}^{\dagger}\mathbf{y}_k,$$

where, in order to avoid an abundance of notations, $\mathbf{y}_k$ is now given by

$$\mathbf{y}_k = \mathbf{y} - \mathbf{F}\left(\mathbf{x}_{k\alpha}\right) + \mathbf{K}\left(\mathbf{x}_{k\alpha}\right)\left(\mathbf{x}_{k\alpha} - \mathbf{x}_{\mathrm{a}}\right).$$

As in the linear case we consider the decomposition

$$\mathbf{x}^{\dagger} - \mathbf{x}_{k+1\alpha}^{\delta} = \left(\mathbf{x}^{\dagger} - \mathbf{x}_{k+1\alpha}\right) + \left(\mathbf{x}_{k+1\alpha} - \mathbf{x}_{k+1\alpha}^{\delta}\right) \tag{8.76}$$

and try to estimate each term in the right-hand side of (8.76). Using the linearisation about $\mathbf{x}_{k\alpha}$ and $\mathbf{x}_{k\alpha}^{\delta}$

$$\mathbf{y} = \mathbf{F}\left(\mathbf{x}_{k\alpha}\right) + \mathbf{K}\left(\mathbf{x}_{k\alpha}\right)\left(\mathbf{x}^{\dagger} - \mathbf{x}_{k\alpha}\right) + \mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{k\alpha}\right)$$

and

$$\mathbf{y} = \mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) + \mathbf{K}_{k\alpha}\left(\mathbf{x}^{\dagger} - \mathbf{x}_{k\alpha}^{\delta}\right) + \mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{k\alpha}^{\delta}\right),$$

respectively, and assuming that $\mathbf{K}_{k\alpha} \approx \mathbf{K}\left(\mathbf{x}_{k\alpha}\right)$, we express the first term in (8.76) as

$$\mathbf{x}^{\dagger} - \mathbf{x}_{k+1\alpha} = \left(\mathbf{x}^{\dagger} - \mathbf{x}_{\mathrm{a}}\right) - \mathbf{K}_{k\alpha}^{\dagger}\mathbf{y}_k = \left(\mathbf{I}_n - \mathbf{A}_{k\alpha}\right)\left(\mathbf{x}^{\dagger} - \mathbf{x}_{\mathrm{a}}\right) - \mathbf{K}_{k\alpha}^{\dagger}\mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{k\alpha}\right)$$

and the second term as

$$\mathbf{x}_{k+1\alpha} - \mathbf{x}_{k+1\alpha}^{\delta} = \mathbf{K}_{k\alpha}^{\dagger}\left(\mathbf{y}_k - \mathbf{y}_k^{\delta}\right) = -\mathbf{K}_{k\alpha}^{\dagger}\delta - \mathbf{K}_{k\alpha}^{\dagger}\left[\mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{k\alpha}^{\delta}\right) - \mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{k\alpha}\right)\right]$$

with

$$\mathbf{A}_{k\alpha} = \mathbf{K}_{k\alpha}^{\dagger}\mathbf{K}_{k\alpha}$$

being the averaging kernel. Inserting the above relations into (8.76) we find that

$$\mathbf{x}^{\dagger} - \mathbf{x}_{k+1\alpha}^{\delta} = \left(\mathbf{I}_n - \mathbf{A}_{k\alpha}\right)\left(\mathbf{x}^{\dagger} - \mathbf{x}_{\mathrm{a}}\right) - \mathbf{K}_{k\alpha}^{\dagger}\delta - \mathbf{K}_{k\alpha}^{\dagger}\mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{k\alpha}^{\delta}\right). \tag{8.77}$$

Assuming that the sequence $\left\{\mathbf{x}_{k\alpha}^{\delta}\right\}$ converges and denoting by $\mathbf{x}_{\alpha}^{\delta}$ the limit of this sequence, we let $k \to \infty$ in (8.77) and obtain

$$\mathbf{e}_{\alpha}^{\delta} = \mathbf{e}_{\mathrm{s}\alpha} + \mathbf{e}_{\mathrm{n}\alpha}^{\delta} + \mathbf{e}_{\mathrm{l}\alpha}^{\delta}, \tag{8.78}$$

where

$$\mathbf{e}_{\alpha}^{\delta} = \mathbf{x}^{\dagger} - \mathbf{x}_{\alpha}^{\delta}$$

is the error vector at the solution,

$$\mathbf{e}_{\mathrm{s}\alpha} = \left(\mathbf{I}_n - \mathbf{A}_{\alpha}\right)\left(\mathbf{x}^{\dagger} - \mathbf{x}_{\mathrm{a}}\right)$$

is the smoothing error vector,

$$\mathbf{e}_{\mathrm{n}\alpha}^{\delta} = -\mathbf{K}_{\alpha}^{\dagger}\delta$$

is the noise error vector and

$$\mathbf{e}_{\mathrm{l}\alpha}^{\delta} = -\mathbf{K}_{\alpha}^{\dagger}\mathbf{R}\left(\mathbf{x}^{\dagger}, \mathbf{x}_{\alpha}^{\delta}\right)$$

is the non-linearity error vector. Note that the generalized inverse $\mathbf{K}_{\alpha}^{\dagger}$ and the averaging kernel $\mathbf{A}_{\alpha}$, which enter in the expressions of the error components, are evaluated at $\mathbf{x}_{\alpha}^{\delta}$.

The expression of the total error can also be derived by using the fact that $\mathbf{x}_{\alpha}^{\delta}$ is a minimizer of the Tikhonov function. The stationary condition for $\mathcal{F}$ at $\mathbf{x}_{\alpha}^{\delta}$

$$\mathbf{g}\left(\mathbf{x}_{\alpha}^{\delta}\right) = \nabla\mathcal{F}\left(\mathbf{x}_{\alpha}^{\delta}\right) = 0$$

implies that $\mathbf{x}_\alpha^\delta$ satisfies the Euler equation

$$\mathbf{K}_\alpha^T \left[\mathbf{F}\left(\mathbf{x}_\alpha^\delta\right) - \mathbf{y}^\delta\right] + \alpha \mathbf{L}^T\mathbf{L}\left(\mathbf{x}_\alpha^\delta - \mathbf{x}_\mathtt{a}\right) = 0; \tag{8.79}$$

whence considering the linearisation of $\mathbf{F}$ about $\mathbf{x}_\alpha^\delta$

$$\mathbf{y} = \mathbf{F}\left(\mathbf{x}_\alpha^\delta\right) + \mathbf{K}_\alpha\left(\mathbf{x}^\dagger - \mathbf{x}_\alpha^\delta\right) + \mathbf{R}\left(\mathbf{x}^\dagger, \mathbf{x}_\alpha^\delta\right), \tag{8.80}$$

we find that

$$\left(\mathbf{K}_\alpha^T\mathbf{K}_\alpha + \alpha\mathbf{L}^T\mathbf{L}\right)\left(\mathbf{x}^\dagger - \mathbf{x}_\alpha^\delta\right) = \alpha\mathbf{L}^T\mathbf{L}\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right) - \mathbf{K}_\alpha^T\delta - \mathbf{K}_\alpha^T\mathbf{R}\left(\mathbf{x}^\dagger, \mathbf{x}_\alpha^\delta\right).$$

Further, using the identity

$$\left(\mathbf{K}^T\mathbf{K} + \alpha\mathbf{L}^T\mathbf{L}\right)^{-1}\alpha\mathbf{L}^T\mathbf{L} = \mathbf{I}_n - \left(\mathbf{K}^T\mathbf{K} + \alpha\mathbf{L}^T\mathbf{L}\right)^{-1}\mathbf{K}^T\mathbf{K},$$

we obtain

$$\mathbf{x}^\dagger - \mathbf{x}_\alpha^\delta = \left(\mathbf{I}_n - \mathbf{A}_\alpha\right)\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right) - \mathbf{K}_\alpha^\dagger\delta - \mathbf{K}_\alpha^\dagger\mathbf{R}\left(\mathbf{x}^\dagger, \mathbf{x}_\alpha^\delta\right), \tag{8.81}$$

which is the explicit form of (8.78). Thus, the error representations in the non-linear and the linear case are similar, excepting an additional term, which represents the non-linearity error. If the minimizer $\mathbf{x}_\alpha^\delta$ is sufficiently close to the exact solution $\mathbf{x}^\dagger$, the non-linearity error can be neglected and the agreement is perfect.

In a semi–stochastic framework we suppose that $\mathbf{K}_\alpha$ is deterministic and, as a result, the error vector $\mathbf{e}_\alpha^\delta$ is stochastic with mean $\mathbf{e}_{\mathtt{s}\alpha}$ and covariance $\mathbf{C}_\mathtt{en} = \sigma^2\mathbf{K}_\alpha^\dagger\mathbf{K}_\alpha^{\dagger T}$. As in the linear case, we may define the mean square error matrix

$$\begin{aligned}\mathbf{S}_\alpha &= \mathbf{e}_{\mathtt{s}\alpha}\mathbf{e}_{\mathtt{s}\alpha}^T + \mathbf{C}_\mathtt{en} \\ &= \left(\mathbf{I}_n - \mathbf{A}_\alpha\right)\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)^T\left(\mathbf{I}_n - \mathbf{A}_\alpha\right)^T + \sigma^2\mathbf{K}_\alpha^\dagger\mathbf{K}_\alpha^{\dagger T}\end{aligned} \tag{8.82}$$

to quantify the dispersion of the regularized solution $\mathbf{x}_\alpha^\delta$ about the true solution $\mathbf{x}^\dagger$. The one-rank matrix $\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)^T$ can be approximated by

$$\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)^T \approx \left(\mathbf{x}_\alpha^\delta - \mathbf{x}_\mathtt{a}\right)\left(\mathbf{x}_\alpha^\delta - \mathbf{x}_\mathtt{a}\right)^T \tag{8.83}$$

or by

$$\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)^T \approx \frac{\sigma^2}{\alpha}\left(\mathbf{L}^T\mathbf{L}\right)^{-1}. \tag{8.84}$$

The approximation (8.83) yields the so called semi-stochastic representation of $\mathbf{S}_\alpha$, while the approximation (8.84) yields the stochastic representation of $\mathbf{S}_\alpha$, since in this case $\mathbf{S}_\alpha$ coincides with the a posteriori covariance matrix in a statistical inversion.

Accounting on all assumptions made, we deduce that a linearised error analysis can be performed when

1. the sequence of iterates $\left\{\mathbf{x}_{k\alpha}^\delta\right\}$ converges;
2. the forward model can be linearised about $\mathbf{x}_\alpha^\delta$ in the sense that the non-linearity residual $\mathbf{R}\left(\mathbf{x}^\dagger, \mathbf{x}_\alpha^\delta\right)$ is small;
3. the data error model is correct.

If one of these assumptions is violated the error analysis is erroneous. The convergence of iterates can be guaranteed by using an appropriate termination criterion, but the second and third assumptions require more attention.

The linearity assumption can be verified at the boundary of a confidence region for the solution. For this purpose we consider a singular value decomposition of the means square error matrix $\mathbf{S}_\alpha = \mathbf{V}\Sigma\mathbf{V}^T$ and define the normalized error patterns $\mathbf{s}_k$ for $\mathbf{S}_\alpha$ from the decomposition $\mathbf{V}\Sigma^{1/2} = [\mathbf{s}_1, ..., \mathbf{s}_n]$. The linearisation error

$$\mathbf{R}\left(\mathbf{x}\right) = \mathbf{F}\left(\mathbf{x}\right) - \mathbf{F}\left(\mathbf{x}_\alpha^\delta\right) + \mathbf{K}_\alpha\left(\mathbf{x} - \mathbf{x}_\alpha^\delta\right)$$

can be estimated by comparing

$$\varepsilon_{\mathtt{lin}k} = \frac{1}{m\sigma^2} \left\| \mathbf{R} \left(\mathbf{x}_\alpha^\delta \pm \mathbf{s}_k \right) \right\|^2 \approx 1$$

for all $k = 1, ..., n$.

The validity of the data error model assumption is perhaps the most important problem of an error analysis. If the data error $\delta_{\mathtt{y}}$ contains only the noise error component δ and if δ is a Gaussian random vector with zero mean and covariance matrix $\mathbf{C}_\delta = \sigma^2 \mathbf{I}_m$, then the square residual norm $\left\| \mathbf{r}_\alpha^\delta \right\|^2$ is Chi-squared distributed with variance σ^2 and $\mathrm{trace}\left(\mathbf{I}_m - \widehat{\mathbf{A}}_\alpha \right)$ degrees of freedom. In the limit $\alpha \to 0$ $\left\| \mathbf{r}_\alpha^\delta \right\|^2$ is Chi-squared distributed with variance σ^2 and $m - n$ degrees of freedom and it must holds

$$\left\| \mathbf{r}_\alpha^\delta \right\|^2 = \left\| \mathbf{F} \left(\mathbf{x}_\alpha^\delta \right) - \mathbf{y}^\delta \right\|^2 \approx (m - n)\, \sigma^2, \;\; \alpha \to 0.$$

If the contribution of the systematic error $\delta_{\mathtt{sys}}$ in the data error $\delta_{\mathtt{y}}$ is significant, we have instead

$$\left\| \mathbf{F} \left(\mathbf{x}_\alpha^\delta \right) - \mathbf{y}^\delta \right\|^2 \approx (m - n) \left(\frac{1}{m} \left\| \delta_{\mathtt{sys}} \right\|^2 + \sigma^2 \right), \;\; \alpha \to 0.$$

The presence of the systematic errors introduce an additional bias in the solution and to handle with this type of errors we may proceed as in the linear case, that is, we may replace the data error $\delta_{\mathtt{y}}$ by an equivalent white noise $\delta_{\mathtt{e}}$ such that

$$\mathcal{E} \left\{ \left\| \delta_{\mathtt{e}} \right\|^2 \right\} = \mathcal{E} \left\{ \left\| \delta_{\mathtt{y}} \right\|^2 \right\}.$$

The variance of the white noise $\delta_{\mathtt{e}}$ is then given by

$$\sigma_{\mathtt{e}}^2 = \frac{1}{m} \left\| \delta_{\mathtt{sys}} \right\|^2 + \sigma^2$$

and the estimate

$$\sigma_{\mathtt{e}}^2 \approx \frac{1}{m - n} \left\| \mathbf{F} \left(\mathbf{x}_\alpha^\delta \right) - \mathbf{y}^\delta \right\|^2, \;\; \alpha \to 0$$

can be used to perform an error analysis with the equivalent white noise covariance matrix $\mathbf{C}_{\delta_{\mathtt{e}}} = \sigma_{\mathtt{e}}^2 \mathbf{I}_m$. It is apparent that by this equivalence we increase the noise variance and eliminate the bias due to the systematic errors.

Linearisation of the gradient of the Tikhonov function

The error representation (8.81) has been derived by assuming a linearisation of the forward model about the minimizer of the Tikhonov function. This derivation does not make use of the Gauss approximation of the Hessian, since the Euler equation (8.79) holds true for the Gauss-Newton method as well as for the Newton method. An alternative error representation can be derived for a general Hessian matrix by using the linearisation of the gradient of the objective function. Because the objective function depends on the data $\mathbf{y}^\delta$ we will indicate this dependence by writing $\mathcal{F}\left(\mathbf{x}, \mathbf{y}^\delta \right)$ in place of $\mathcal{F}\left(\mathbf{x} \right)$.

Setting $\mathbf{x}^\dagger = \mathbf{x}_\alpha^\delta - \triangle \mathbf{x}_\alpha^\delta$ and using the representation $\mathbf{y} = \mathbf{y}^\delta - \delta$ we expand the gradient at $\left(\mathbf{x}^\dagger, \mathbf{y} \right)$ in a first-order Taylor series about $\left(\mathbf{x}_\alpha^\delta, \mathbf{y}^\delta \right)$

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \mathbf{x}} \left(\mathbf{x}^\dagger, \mathbf{y} \right) &= \frac{\partial \mathcal{F}}{\partial \mathbf{x}} \left(\mathbf{x}_\alpha^\delta - \triangle \mathbf{x}_\alpha^\delta, \mathbf{y}^\delta - \delta \right) \\ &= \frac{\partial \mathcal{F}}{\partial \mathbf{x}} \left(\mathbf{x}_\alpha^\delta, \mathbf{y}^\delta \right) - \frac{\partial^2 \mathcal{F}}{\partial \mathbf{x}^2} \left(\mathbf{x}_\alpha^\delta, \mathbf{y}^\delta \right) \triangle \mathbf{x}_\alpha^\delta \\ &\quad - \frac{\partial^2 \mathcal{F}}{\partial \mathbf{x} \partial \mathbf{y}^\delta} \left(\mathbf{x}_\alpha^\delta, \mathbf{y}^\delta \right) \delta + \mathtt{R} \left(\mathbf{x}^\dagger, \mathbf{x}_\alpha^\delta \right). \end{aligned}$$

At the minimum we have

$$\frac{\partial \mathcal{F}}{\partial \mathbf{x}} \left(\mathbf{x}_\alpha^\delta, \mathbf{y}^\delta \right) = 0, \tag{8.85}$$

whence taking into account that

$$\frac{\partial \mathcal{F}}{\partial \mathbf{x}}\left(\mathbf{x}^{\dagger},\mathbf{y}\right)=\mathbf{g}\left(\mathbf{x}^{\dagger},\mathbf{y}\right)=\alpha\mathbf{L}^{T}\mathbf{L}\left(\mathbf{x}^{\dagger}-\mathbf{x}_{\mathtt{a}}\right)$$

and that

$$\frac{\partial^{2}\mathcal{F}}{\partial \mathbf{x}^{2}}\left(\mathbf{x}_{\alpha}^{\delta},\mathbf{y}^{\delta}\right)=\mathbf{G}\left(\mathbf{x}_{\alpha}^{\delta}\right),\quad \frac{\partial^{2}\mathcal{F}}{\partial \mathbf{x}\partial \mathbf{y}^{\delta}}\left(\mathbf{x}_{\alpha}^{\delta},\mathbf{y}^{\delta}\right)=-\mathbf{K}^{T}\left(\mathbf{x}_{\alpha}^{\delta}\right),$$

we obtain

$$\mathbf{x}^{\dagger}-\mathbf{x}_{\alpha}^{\delta}=\alpha\mathbf{G}_{\alpha}^{-1}\mathbf{L}^{T}\mathbf{L}\left(\mathbf{x}^{\dagger}-\mathbf{x}_{\mathtt{a}}\right)-\mathbf{G}_{\alpha}^{-1}\mathbf{K}_{\alpha}^{T}\delta-\mathbf{G}_{\alpha}^{-1}\mathtt{R}\left(\mathbf{x}^{\dagger},\mathbf{x}_{\alpha}^{\delta}\right) \tag{8.86}$$

with $\mathbf{G}_{\alpha}=\mathbf{G}\left(\mathbf{x}_{\alpha}^{\delta}\right)$. Thus, neglecting the non-linearity error, the smoothing and the noise errors take the forms

$$\mathbf{e}_{\mathtt{s}\alpha}=\alpha\mathbf{G}_{\alpha}^{-1}\mathbf{L}^{T}\mathbf{L}\left(\mathbf{x}^{\dagger}-\mathbf{x}_{\mathtt{a}}\right)$$

and

$$\mathbf{e}_{\mathtt{n}\alpha}^{\delta}=-\mathbf{G}_{\alpha}^{-1}\mathbf{K}_{\alpha}^{T}\delta,$$

respectively. The representations (8.81) and (8.86) coincide only for a Gauss approximation of the Hessian and the difference between them stems from the different linearisation employed. The mean and the covariance matrix of the error vector $\mathbf{e}_{\alpha}^{\delta}$ are given by the smoothing error vector $\mathbf{e}_{\mathtt{s}\alpha}$ and the retrieval noise covariance matrix

$$\mathbf{C}_{\mathtt{en}}=\sigma^{2}\mathbf{G}_{\alpha}^{-1}\mathbf{K}_{\alpha}^{T}\mathbf{K}_{\alpha}\mathbf{G}_{\alpha}^{-1}.$$

When the quasi–Newton method is used to compute the minimizer of Tikhonov function, the Hessian approximation can be used for error estimation. However, for the Gauss–Newton method, an additional computational step involving the calculation of the Hessian at the solution has to be performed. For this purpose we consider the Taylor expansion about the minimizer $\mathbf{x}_{\alpha}^{\delta}$

$$\mathcal{F}\left(\mathbf{x}\right)\approx\mathcal{F}\left(\mathbf{x}_{\alpha}^{\delta}\right)+\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{\alpha}^{\delta}\right)^{T}\mathbf{G}_{\alpha}\left(\mathbf{x}-\mathbf{x}_{\alpha}^{\delta}\right), \tag{8.87}$$

where by definition the entries of the Hessian are given by

$$\left[\mathbf{G}_{\alpha}\right]_{ij}=\frac{\partial^{2}\mathcal{F}}{\partial x_{i}\partial x_{j}}\left(\mathbf{x}_{\alpha}^{\delta}\right) \tag{8.88}$$

with $x_{i}=\left[\mathbf{x}\right]_{i}$. Equations (8.87) and (8.88) suggest that we may use finite differences for approximating $\mathbf{G}_{\alpha}$. Setting for convenience $\mathbf{x}=\mathbf{x}_{\alpha}^{\delta}$ and denoting by Δx_{i} the displacement in the i^{th} component of $\mathbf{x}$, we may compute the diagonal elements by using (8.87)

$$\left[\mathbf{G}_{\alpha}\right]_{ii}=2\frac{\mathcal{F}\left(x_{i}+\Delta x_{i}\right)-\mathcal{F}\left(x_{i}\right)}{\left(\Delta x_{i}\right)^{2}} \tag{8.89}$$

and the off-diagonal elements from (8.88) with central differences

$$\begin{aligned}\left[\mathbf{G}_{\alpha}\right]_{ij} &= \left[\mathcal{F}\left(x_{i}+\Delta x_{i},x_{j}+\Delta x_{j}\right)-\mathcal{F}\left(x_{i}-\Delta x_{i},x_{j}+\Delta x_{j}\right)\right.\\ &\quad \left.-\mathcal{F}\left(x_{i}+\Delta x_{i},x_{j}-\Delta x_{j}\right)+\mathcal{F}\left(x_{i}-\Delta x_{i},x_{j}-\Delta x_{j}\right)\right]/\left(4\Delta x_{i}\Delta x_{j}\right).\end{aligned} \tag{8.90}$$

The calculation of the Hessian by using finite differences requires an adequate choice of the step sizes Δx_{i}. The difficulty associated with the selection of the step size stems from the fact that the objective function may varies slowly in some directions of the $\mathbf{x}$-space and rapidly in other. The implemented iterative algorithm significantly improves the reliability of the Hessian matrix calculation and is based on the following result: if $\mathbf{G}_{\alpha}$ is the exact Hessian and $\mathbf{G}_{\alpha}=\mathbf{V}\Sigma\mathbf{V}^{T}$ with $\Sigma=\left[\mathrm{diag}\left(\sigma_{i}^{2}\right)_{n\times n}\right]$ is a singular value decomposition of $\mathbf{G}_{\alpha}$, then the linear transformation

$$\mathbf{x}=\mathbf{V}\Sigma^{-1/2}\mathbf{z} \tag{8.91}$$

implies that in the $\mathbf{z}$-space the surfaces of constant $\mathcal{F}$ are spheres, i.e.

$$\mathcal{F}(\mathbf{z}) - \mathcal{F}(\mathbf{z}_0) = \frac{1}{2}(\mathbf{z}-\mathbf{z}_0)^T(\mathbf{z}-\mathbf{z}_0). \tag{8.92}$$

The computation of the pseudo-Hessian by finite differences is performed in the $\mathbf{z}$-space by using (8.89) and (8.90) and this process is more stable than a Hessian calculation in the $\mathbf{x}$-space. The step sizes Δz_i are chosen such that, when computing the diagonal elements, the variations $\mathcal{F}(z_i+\Delta z_i) - \mathcal{F}(z_i)$ are approximately one.

8.3.1.5. Parameter choice methods

The choice of the regularization parameter plays an important role in computing a reliable approximation of the solution. In this section we present selection criteria with variable and constant regularization parameters. In the first case the regularization parameter is estimated at each iteration step, while in the second case the minimization of the Tikhonov function is done a few times with different regularization parameters.

A priori parameter choice methods

The expected error estimation can be formulated as an a priori parameter choice method. The idea is to perform a random exploration of a domain to which the solution is suppose to lie and for each state vector realization $\mathbf{x}_i^\dagger$ to compute the optimal regularization parameter for error estimation $\overline{\alpha}_{\mathtt{opt}i} = \arg\min_\alpha \mathcal{E}\left\{\left\|\mathbf{e}_{\alpha i}^\delta\right\|^2\right\}$ and the exponent $p_i = \ln\overline{\alpha}_{\mathtt{opt}i}/\ln\sigma$. The regularization parameter is then chosen as $\alpha_{\mathtt{e}} = \sigma^{\bar{p}}$, where $\bar{p} = (1/N)\sum_{i=1}^N p_i$ is the sample mean exponent and N is the number of events.

The error estimation method can be formulated for non-linear problems by representing the expected error at the solution as

$$\mathcal{E}\left\{\left\|\mathbf{e}_\alpha^\delta\right\|^2\right\} = \left\|\mathbf{e}_{\mathtt{s}\alpha}\right\|^2 + \mathcal{E}\left\{\left\|\mathbf{e}_{\mathtt{n}\alpha}^\delta\right\|^2\right\}$$

with

$$\mathbf{e}_{\mathtt{s}\alpha} = (\mathbf{I}_n - \mathbf{A}_\alpha)\left(\mathbf{x}^\dagger - \mathbf{x}_{\mathtt{a}}\right) = \sum_{i=1}^n \frac{\alpha}{\gamma_i^2+\alpha}\left(\mathbf{x}^\dagger - \mathbf{x}_{\mathtt{a}}\right)^T \mathbf{w}_i^{-1}\mathbf{w}_i \tag{8.93}$$

and

$$\mathcal{E}\left\{\left\|\mathbf{e}_{\mathtt{n}\alpha}^\delta\right\|^2\right\} = \sigma^2\mathrm{trace}\left(\mathbf{K}_\alpha^\dagger\mathbf{K}_\alpha^{\dagger T}\right) = \sigma^2\sum_{i=1}^n\left(\frac{\gamma_i^2}{\gamma_i^2+\alpha}\frac{1}{\sigma_i}\right)^2\left\|\mathbf{w}_i\right\|^2. \tag{8.94}$$

In (8.93) and (8.94) γ_i are the generalized singular values of the matrix pair $(\mathbf{K}_\alpha, \mathbf{L})$, $\mathbf{w}_i$ is the i^{th} column of the non-singular matrix $\mathbf{W}$ and $\mathbf{w}_i^{-1}$ is the i^{th} line of the inverse matrix $\mathbf{W}^{-1}$. Because the Jacobian matrix $\mathbf{K}_\alpha$ is evaluated at the solution, a non-linear minimization method has to be employed for computing $\overline{\alpha}_{\mathtt{opt}} = \arg\min_\alpha \mathcal{E}\left\{\left\|\mathbf{e}_\alpha^\delta\right\|^2\right\}$ for each state vector realization. The resulting algorithm is extremely computational expensive and in order to alleviate this drawback we approximate the Jacobian matrix at the solution by the Jacobian matrix at the a priori state. This is a realistic assumption because many ill-posed problems arising in atmospheric remote sensing are nearly linear (with an appropriate choice of the forward model their degree of non-linearity is of moderate size). The a priori parameter choice method is then equivalent to the expected error estimation method applied to a linearisation of the forward model about the a priori state.

Another version of the expected error estimation method can be designed by assuming a semi-stochastic error representation and by using an iterative process for minimizing the expected error. Two relevant features reduce the performances of the so called iterated expected error estimation method:

1. The semi-stochastic error representation is valid if the Tikhonov solution is an acceptable approximation of the exact solution, that is, if the regularization parameter is in the neighbourhood of the optimal value of the regularization parameter.
2. The minimizer of the expected error is in general larger than the optimal value of the regularization parameter.

Selection criteria with variable regularization parameters

Since the solution of a non-linear ill-posed problem by means of Tikhonov regularization is equivalent to the solution of a sequence of ill-posed linearisation of the forward model about the current iterate, parameter choice methods for the linear problem (8.70) can be used to compute the regularization parameter at each iteration step.

The noise in the right–hand side of the linearised equation (8.70) is due to the measurement noise in the data and the linearisation error. At the iteration step k the noise level can be estimated by the minimum value of the linearised residual corresponding to the smallest singular value $\left\|\mathbf{r}^{\delta}_{\mathtt{1}k\mathtt{min}}\right\|$. In the framework of the discrepancy principle the regularization parameter is then selected as the solution of the equation

$$\left\|\mathbf{r}^{\delta}_{\mathtt{1}k\alpha}\right\|^2 = \tau \left\|\mathbf{r}^{\delta}_{\mathtt{1}k\mathtt{min}}\right\|^2$$

with $\tau > 1$.

Due to the difficulties associated with the noise level estimation, error-free parameter choice methods (based only on information about the noisy data) are more attractive. For the linearised equation (8.70), the following parameter choice methods can be considered:

1. the generalized cross-validation

$$\alpha_{\mathtt{gcv}k} = \arg\min_{\alpha} \upsilon^{\delta}_{k\alpha}$$

with

$$\upsilon^{\delta}_{k\alpha} = \frac{\left\|\mathbf{r}^{\delta}_{\mathtt{1},k\alpha}\right\|^2}{\left[\mathrm{trace}\left(\mathbf{I}_m - \widehat{\mathbf{A}}_{k\alpha}\right)\right]^2}, \tag{8.95}$$

2. the maximum likelihood estimation

$$\alpha_{\mathtt{mle}k} = \arg\min_{\alpha} \lambda^{\delta}_{k\alpha}$$

with

$$\lambda^{\delta}_{k\alpha} = \frac{\mathbf{y}^{\delta T}_{k}\left(\mathbf{I}_m - \widehat{\mathbf{A}}_{k\alpha}\right)\mathbf{y}^{\delta}_{k}}{\sqrt[m]{\det\left(\mathbf{I}_m - \widehat{\mathbf{A}}_{k\alpha}\right)}}, \tag{8.96}$$

3. the L-curve method

$$\alpha_{\mathtt{lc}k} = \arg\max_{\alpha} \kappa^{\delta}_{\mathtt{lc}k\alpha}$$

with

$$\kappa^{\delta}_{\mathtt{lc}k\alpha} = \frac{x''_k(\alpha)\, y'_k(\alpha) - x'_k(\alpha)\, y''_k(\alpha)}{\left[\left(x'_k(\alpha)\right)^2 + \left(y'_k(\alpha)\right)^2\right]^{3/2}}$$

and

$$x_k(\alpha) = \log\left(\left\|\mathbf{r}^{\delta}_{\mathtt{1}k\alpha}\right\|^2\right), \quad y_k(\alpha) = \log\left(\left\|\mathbf{c}^{\delta}_{k\alpha}\right\|^2\right).$$

In the above relations the linearised residual and the constrained vector are given by

$$\mathbf{r}^{\delta}_{\mathtt{1}k\alpha} = \mathbf{y}^{\delta}_{k} - \mathbf{K}_{k\alpha}\left(\mathbf{x}^{\delta}_{k+\mathtt{1}\alpha} - \mathbf{x}_{\mathtt{a}}\right) = \left(\mathbf{I}_m - \widehat{\mathbf{A}}_{k\alpha}\right)\mathbf{y}^{\delta}_{k}$$

and

$$\mathbf{c}^{\delta}_{k\alpha} = \mathbf{L}\left(\mathbf{x}^{\delta}_{k+\mathtt{1}\alpha} - \mathbf{x}_{\mathtt{a}}\right)$$

respectively.

In practice the following recommendations for choosing the regularization parameter has to be taken into account:

1. at the beginning of the iterations large α-values should be used to avoid local minima and to get well-conditioned least squares problems to solve;

2. during the iterations the regularization parameter should be decreased slowly to achieve a stable solution.

Numerical experiments have shown that a brutal use of the regularization parameter $\alpha_{\mathtt{opt}k}$ computed by one of the above parameter selection criteria may lead to an oscillation sequence of α-values. A heuristic formula that deals with this problem can be formulated as follows: the regularization parameter is the weighted sum between the previous α-value and the proposed regularization parameter $\alpha_{\mathtt{opt}k}$

$$\alpha_k = \begin{cases} \xi\alpha_{k-1} + (1-\xi)\,\alpha_{\mathtt{opt}k}, & \alpha_{\mathtt{opt}k} < \alpha_{k-1} \\ \alpha_{k-1}, & \alpha_{\mathtt{opt}k} \geq \alpha_{k-1} \end{cases}.$$

This selection rule guarantees a descending sequence of regularization parameters and the resulting method is very similar to the iteratively regularized Gauss-Newton method.

Selection criteria with constant regularization parameters

The numerical realization of these parameter choice methods requires to solve the non-linear minimization problem several times for different regularization parameters. Each minimization is solved with a regularization parameter α and a solution $\mathbf{x}_\alpha^\delta$ is obtained. If the solution is satisfactory as judged by these selection criteria, then the inverse problem is considered to have been solved. The discrete values of the regularization parameters are chosen as $\alpha_i = \sigma^{p_i}$ with $p_i > 0$.

In the framework of the discrepancy principle the regularization parameter is the solution of the equation

$$\left\|\mathbf{y}^\delta - \mathbf{F}\left(\mathbf{x}_\alpha^\delta\right)\right\|^2 = \tau\Delta^2 \tag{8.97}$$

with $\tau > 1$. Because of non-linear problems, the discrepancy principle equation has only a solution under very strong restrictive assumptions, we use a simplified version of this selection criterion: if the sequence $\{\alpha_i\}$ is in descending order, we choose the largest regularization parameter $\alpha_{i^\star}$ for which the residual norm is below the noise level, that is,

$$\left\|\mathbf{y}^\delta - \mathbf{F}\left(\mathbf{x}_{\alpha_{i^\star}}^\delta\right)\right\|^2 \leq \tau\Delta^2 < \left\|\mathbf{y}^\delta - \mathbf{F}\left(\mathbf{x}_{\alpha_i}^\delta\right)\right\|^2, \;\; 0 \leq i < i^\star.$$

The generalized discrepancy principle can also be formulated as an a posteriori parameter choice method for the non-linear Tikhonov regularization. A heuristics justification of this parameter choice method can be given in a deterministic setting by using the error norm estimate

$$\left\|\mathbf{e}_\alpha^\delta\right\|^2 \leq 2\left(\left\|\mathbf{e}_{\mathtt{s}\alpha}\right\|^2 + \left\|\mathbf{e}_{\mathtt{n}\alpha}^\delta\right\|^2\right)$$

together with the noise error bound

$$\left\|\mathbf{e}_{\mathtt{n}\alpha}^\delta\right\|^2 < \frac{2\tau\Delta^2}{\alpha}, \;\; \tau > 1.$$

To estimate the smoothing error we assume $\mathbf{L} = \mathbf{I}_n$ and consider the unperturbed solution $\mathbf{x}_\alpha$ corresponding to the exact data $\mathbf{y}$. The stationary condition for the Tikhonov function at $\mathbf{x}_\alpha$ yields

$$\mathbf{K}_\alpha^T\left[\mathbf{F}\left(\mathbf{x}_\alpha\right) - \mathbf{y}\right] + \alpha\mathbf{I}_n\left(\mathbf{x}_\alpha - \mathbf{x}_{\mathtt{a}}\right) = 0 \tag{8.98}$$

with $\mathbf{K}_\alpha = \mathbf{K}\left(\mathbf{x}_\alpha\right)$. We then obtain

$$\mathbf{x}^\dagger - \mathbf{x}_\alpha = \left(\mathbf{I}_n - \mathbf{A}_\alpha\right)\left(\mathbf{x}^\dagger - \mathbf{x}_{\mathtt{a}}\right) - \mathbf{K}_\alpha^\dagger\mathbf{R}\left(\mathbf{x}^\dagger, \mathbf{x}_\alpha\right)$$

and further

$$\mathbf{e}_{\mathtt{s}\alpha} = \left(\mathbf{I}_n - \mathbf{A}_\alpha\right)\left(\mathbf{x}^\dagger - \mathbf{x}_{\mathtt{a}}\right).$$

Taking into account that for any $\mathbf{x}$ there holds

$$\left\|\left(\mathbf{I}_n - \mathbf{A}_\alpha\right)\mathbf{x}\right\|^2 = \sum_{i=1}^n \left(\frac{\alpha}{\sigma_i^2 + \alpha}\right)^2 \left(\mathbf{v}_i^T\mathbf{x}\right)^2 \leq \sum_{i=1}^n \left(\mathbf{v}_i^T\mathbf{x}\right)^2 = \left\|\mathbf{x}\right\|^2,$$

we deduce that an upper bound for the error norm is given by

$$M(\alpha) = 4\left(\frac{1}{2}\left\|\mathbf{x}^{\dagger} - \mathbf{x}_{\alpha}\right\|^{2} + \tau\frac{\Delta^{2}}{\alpha}\right).$$

To derive the necessary condition for a minimum of the estimate $M(\alpha)$ we consider the function

$$f(\alpha) = \frac{1}{2}\left\|\mathbf{x}^{\dagger} - \mathbf{x}_{\alpha}\right\|^{2} \tag{8.99}$$

and find that

$$\frac{\mathrm{d}f}{\mathrm{d}\alpha}(\alpha) = -\left(\mathbf{x}^{\dagger} - \mathbf{x}_{\alpha}\right)^{T}\frac{\mathrm{d}\mathbf{x}_{\alpha}}{\mathrm{d}\alpha}. \tag{8.100}$$

Formal differentiation of the Euler equation (8.98) with respect to α yields

$$\frac{\mathrm{d}\mathbf{K}_{\alpha}^{T}}{\mathrm{d}\alpha}\left[\mathbf{F}(\mathbf{x}_{\alpha}) - \mathbf{y}\right] + \mathbf{K}_{\alpha}^{T}\mathbf{K}_{\alpha}\frac{\mathrm{d}\mathbf{x}_{\alpha}}{\mathrm{d}\alpha} + \alpha\frac{\mathrm{d}\mathbf{x}_{\alpha}}{\mathrm{d}\alpha} = -\left(\mathbf{x}_{\alpha} - \mathbf{x}_{\mathtt{a}}\right) \tag{8.101}$$

and the neglect of the first term in (8.101) gives

$$\frac{\mathrm{d}\mathbf{x}_{\alpha}}{\mathrm{d}\alpha} \approx -\left(\mathbf{K}_{\alpha}^{T}\mathbf{K}_{\alpha} + \alpha\mathbf{I}_{n}\right)^{-1}\left(\mathbf{x}_{\alpha} - \mathbf{x}_{\mathtt{a}}\right) = \frac{1}{\alpha}\mathbf{K}_{\alpha}^{\dagger}\left[\mathbf{F}(\mathbf{x}_{\alpha}) - \mathbf{y}\right].$$

The linearisation

$$\mathbf{y} \approx \mathbf{F}(\mathbf{x}_{\alpha}) + \mathbf{K}_{\alpha}\left(\mathbf{x}^{\dagger} - \mathbf{x}_{\alpha}\right)$$

and the matrix identity

$$\left(\mathbf{K}_{\alpha}^{T}\mathbf{K}_{\alpha} + \alpha\mathbf{I}_{n}\right)^{-1}\mathbf{K}_{\alpha}^{T} = \mathbf{K}_{\alpha}^{T}\left(\mathbf{K}_{\alpha}\mathbf{K}_{\alpha}^{T} + \alpha\mathbf{I}_{m}\right)^{-1}$$

then yield

$$\begin{aligned}
\frac{\mathrm{d}f}{\mathrm{d}\alpha}(\alpha) &\approx \frac{1}{\alpha}\left(\mathbf{x}^{\dagger} - \mathbf{x}_{\alpha}\right)^{T}\mathbf{K}_{\alpha}^{\dagger}\left[\mathbf{y} - \mathbf{F}(\mathbf{x}_{\alpha})\right] \\
&= \frac{1}{\alpha}\left[\mathbf{K}_{\alpha}\left(\mathbf{x}^{\dagger} - \mathbf{x}_{\alpha}\right)\right]^{T}\left(\mathbf{K}_{\alpha}\mathbf{K}_{\alpha}^{T} + \alpha\mathbf{I}_{m}\right)^{-1}\left[\mathbf{y} - \mathbf{F}(\mathbf{x}_{\alpha})\right] \\
&\approx \frac{1}{\alpha}\left[\mathbf{y} - \mathbf{F}(\mathbf{x}_{\alpha})\right]^{T}\left(\mathbf{K}_{\alpha}\mathbf{K}_{\alpha}^{T} + \alpha\mathbf{I}_{m}\right)^{-1}\left[\mathbf{y} - \mathbf{F}(\mathbf{x}_{\alpha})\right].
\end{aligned}$$

If we now replace $\mathbf{x}_{\alpha}$ by $\mathbf{x}_{\alpha}^{\delta}$ and $\mathbf{y}$ by $\mathbf{y}^{\delta}$ we obtain the generalized discrepancy principle equation in the form

$$\alpha\left[\mathbf{y}^{\delta} - \mathbf{F}\left(\mathbf{x}_{\alpha}^{\delta}\right)\right]^{T}\left(\mathbf{K}_{\alpha}\mathbf{K}_{\alpha}^{T} + \alpha\mathbf{I}_{m}\right)^{-1}\left[\mathbf{y}^{\delta} - \mathbf{F}\left(\mathbf{x}_{\alpha}^{\delta}\right)\right] = \tau\Delta^{2}.$$

Error-free methods with fixed regularization parameter are natural extensions of the corresponding criteria for linear problems and the most popular are the generalized cross validation, the maximum likelihood estimation and the non-linear L-curve method.

To formulate the generalized cross validation and the maximum likelihood estimation we employ some heuristic arguments. At the iteration step k the generalized cross validation function $\upsilon_{k\alpha}^{\delta}$ and the maximum likelihood function $\lambda_{k\alpha}^{\delta}$, as given by (8.95) and (8.96), respectively, depend on the influence matrix $\widehat{\mathbf{A}}_{k\alpha}$, the linearised residual

$$\mathbf{r}_{1k\alpha}^{\delta} = \mathbf{y}_{k}^{\delta} - \mathbf{K}_{k\alpha}\left(\mathbf{x}_{k+1\alpha}^{\delta} - \mathbf{x}_{\mathtt{a}}\right),$$

and the noisy data vector

$$\mathbf{y}_{k}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}\left(\mathbf{x}_{k\alpha}^{\delta}\right) + \mathbf{K}_{k\alpha}\left(\mathbf{x}_{k\alpha}^{\delta} - \mathbf{x}_{\mathtt{a}}\right).$$

Assuming that the iterates $\mathbf{x}_{k\alpha}^{\delta}$ converge to $\mathbf{x}_{\alpha}^{\delta}$, then $\widehat{\mathbf{A}}_{k\alpha}$ converges to the influence matrix at the solution $\widehat{\mathbf{A}}_{\alpha} = \mathbf{K}_{\alpha}\mathbf{K}_{\alpha}^{\dagger}$, $\mathbf{r}_{1k\alpha}^{\delta}$ to the non-linear residual

$$\mathbf{r}_{\alpha}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}\left(\mathbf{x}_{\alpha}^{\delta}\right)$$

and $\mathbf{y}_{k}^{\delta}$ to

$$\mathbf{y}_{\alpha}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}\left(\mathbf{x}_{\alpha}^{\delta}\right) + \mathbf{K}_{\alpha}\left(\mathbf{x}_{\alpha}^{\delta} - \mathbf{x}_{\mathtt{a}}\right).$$

Thus, in the limit $k \to \infty$ the generalized cross validation function takes the form

$$\upsilon_\alpha^\delta = \frac{\left\|\mathbf{y}^\delta - \mathbf{F}\left(\mathbf{x}_\alpha^\delta\right)\right\|^2}{\left[\operatorname{trace}\left(\mathbf{I}_m - \widehat{\mathbf{A}}_\alpha\right)\right]^2}$$

and the maximum likelihood function read as

$$\lambda_\alpha = \frac{\mathbf{y}_\alpha^{\delta T}\left(\mathbf{I}_m - \widehat{\mathbf{A}}_\alpha\right)\mathbf{y}_\alpha^\delta}{\sqrt[m]{\det\left(\mathbf{I}_m - \widehat{\mathbf{A}}_\alpha\right)}} = \frac{\sum_{i=1}^m \left(\mathbf{u}_i^T \mathbf{y}_\alpha^\delta\right)^2 / \left(\gamma_i^2 + \alpha\right)}{\sqrt[m]{\prod_{i=1}^m 1/\left(\gamma_i^2 + \alpha\right)}},$$

where $(\gamma_i; \mathbf{w}_i, \mathbf{u}_i, \mathbf{v}_i)$ is a generalized singular system of the matrix pair $(\mathbf{K}_\alpha, \mathbf{L})$.

The non-linear L-curve is the plot of the constraint norm $\left\|\mathbf{c}_\alpha^\delta\right\|^2 = \left\|\mathbf{L}\left(\mathbf{x}_\alpha^\delta - \mathbf{x}_\mathtt{a}\right)\right\|^2$ against the residual norm $\left\|\mathbf{r}_\alpha^\delta\right\|^2 = \left\|\mathbf{y}^\delta - \mathbf{F}\left(\mathbf{x}_\alpha^\delta\right)\right\|^2$ for a range of values of regularization parameter α. This curve is monotonically decreasing and convex. In a computational sense the non-linear L-curve consists of a number of discrete points corresponding to the different values of the regularization parameter and, in practice, the following techniques can be used for choosing the regularization parameter:

1. As for linear iterative regularization methods, we fit a cubic spline curve to the discrete points of the L-curve $\left(x\left(\alpha_i\right), y\left(\alpha_i\right)\right)$ with $x\left(\alpha\right) = 2\log\left\|\mathbf{r}_\alpha^\delta\right\|$ and $y\left(\alpha\right) = 2\log\left\|\mathbf{c}_\alpha^\delta\right\|$, and determine the point on the original discrete curve that is closest to the spline curve's corner.
2. In the framework of the minimum distance function approach we compute

   $$\alpha_{\mathtt{lc}} = \arg\min_i d^2\left(\alpha_i\right)$$

   with

   $$d^2\left(\alpha\right) = \left[x\left(\alpha\right) - x_0\right]^2 + \left[y\left(\alpha\right) - y_0\right]^2,$$

   $x_0 = \min_i x\left(\alpha_i\right)$ and $y_0 = \min_i y\left(\alpha_i\right)$.
3. Relying on the definition of the corner of the L-curve as given by Reginska, we determine the regularization parameter as

   $$\alpha_{\mathtt{lc}} = \arg\min_i \left\{x\left(\alpha_i\right) + y\left(\alpha_i\right)\right\},$$

   that is, we detect the minimum of the logarithmic L-curve rotated by $\pi/4$ radians.

8.4. Relation Number Density and VMR

The following definitions are used for the Limb MDS

- Number of elements of the retrieval grid: n_{main} (= number of profile entries and number of layers)
- State vector: x
- Element index per layer: $k = 1 \ldots n_{main}$
- State vector elements assigned to partial columns per layer: x_k, $k = 1 \ldots n_{main}$ (vertical columns per layer = partial columns per layer)
- Retrieval grid:
  - Ordering: top to down
  - Top of atmosphere is fixed to 100 km and used for the height grid
  - Information at TOA is used for height and pressure grid
  - Height at the lower boundary of each layer k : z_k

- Pressure at the lower boundary of each layer k : p_k
- Altitude grid element: $\Delta z_k = z_{k-1} - z_k$ with $z_0 = 100$ km
- Pressure grid element: $\Delta p_k = p_{k-1} - p_k$
- Pressure is read from climatology with p_0 at top of atmosphere

➢ Number density per layer: n with $n_k, k = 1 \ldots n_{main}$

➢ Volume mixing ratio per layer: vmr with vmr_k , $k = 1 \ldots n_{main}$

Since it is very often desirable to represent the retrieved profile in number density or volume mixing ratio, the relations with respect to the chosen state vector representation (partial columns per layer) are given in the following. Note that the conversion is related to quantities which describe the result for a given profile layer. Additionally, note that the information provided for pressure and height is implicitly extended by the entries at TOA which are not given in the product.Because the retrieved quantities are the partial columns it is important to relate the partial columns to the number density and VMR.

Using

$$n_k \left[\frac{\text{molec}}{\text{cm}^3}\right] = 2.6868 \cdot 10^{11} \frac{x_k \left[10^{-3}\text{cm}\right]}{\triangle z_k \left[\text{km}\right]}.$$

and

$$vmr_k \left[.\right] = \frac{x_k \left[10^{-3}\text{cm}\right]}{0.789087 \cdot 10^6 \cdot \triangle p_k \left[\text{mb}\right]},$$

we express n_k and vmr_k as

$$n_k = s_k^{\mathtt{nd}} x_k$$

and

$$vmr_k = s_k^{\mathtt{vmr}} x_k,$$

respectively, where

$$s_k^{\mathtt{nd}} = \frac{2.6868 \cdot 10^{11}}{\triangle z_k \left[\text{km}\right]},$$
$$s_k^{\mathtt{vmr}} = \frac{1}{0.789087 \cdot 10^6 \cdot \triangle p_k \left[\text{mb}\right]}.$$

If n_k, vmr_k and x_k are the entries of the number density vector $\mathbf{n}$, the VMR vector $\mathbf{vmr}$, and the state vector $\mathbf{x}$, respectively, we express the above relations in matrix form as:

$$\mathbf{n} = \mathbf{S}_{\mathtt{nd}}\mathbf{x}, \quad \mathbf{vmr} = \mathbf{S}_{\mathtt{vmr}}\mathbf{x}.$$

Here, $\mathbf{S}_{\mathtt{nd}}$ and $\mathbf{S}_{\mathtt{vmr}}$ are diagonal matrices defined by $[\mathbf{S}_{\mathtt{nd}}]_{ij} = s_i^{\mathtt{nd}} \cdot \delta_{ij}$ and $[\mathbf{S}_{\mathtt{vmr}}]_{ij} = s_i^{\mathtt{vmr}} \cdot \delta_{ij}$. The conversion can also be applied to the representation of the covariance matrix, the correlation matrix, the relative error and finally the averaging kernel.

The computed covariance matrix is in fact the means square error matrix in a semi-stochastic framework, i.e.,

$$\mathbf{C}(\mathbf{x}) = \mathcal{E}\{\mathbf{x}\mathbf{x}^T\} = (\mathbf{I}_n - \mathbf{A}) \left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right) \left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)^T (\mathbf{I}_n - \mathbf{A})^T + \sigma^2 \mathbf{K}^\dagger \mathbf{K}^{\dagger T},$$

where the one-rank matrix $(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a})(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a})^T$ is approximated by

$$\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)\left(\mathbf{x}^\dagger - \mathbf{x}_\mathtt{a}\right)^T \approx \left(\mathbf{x} - \mathbf{x}_\mathtt{a}\right)\left(\mathbf{x} - \mathbf{x}_\mathtt{a}\right)^T.$$

The covariance matrix $\mathbf{C}(\mathbf{x})$ can be then converted into number density representation by

$$\mathbf{C}(\mathbf{n}) = \mathbf{S}_\mathtt{nd}\mathbf{C}(\mathbf{x})\mathbf{S}_\mathtt{nd}$$

and into VMR representation by

$$\mathbf{C}(\mathbf{vmr}) = \mathbf{S}_\mathtt{vmr}\mathbf{C}(\mathbf{x})\mathbf{S}_\mathtt{vmr}.$$

The relative error ϵ_i at the solution is defined by

$$\epsilon_i = \frac{\sqrt{[\mathbf{C}(\mathbf{x})]_{ii}}}{x_i} \cdot 100 = \frac{\sqrt{[\mathbf{C}(\mathbf{n})]_{ii}}}{n_i} \cdot 100 = \frac{\sqrt{[\mathbf{C}(\mathbf{vmr})]_{ii}}}{vmr_i} \cdot 100,$$

while the off-diagonal elements of the correlation matrix of the fit are given by

$$\rho_{ij} = \frac{[\mathbf{C}(\mathbf{x})]_{ij}}{\sqrt{[\mathbf{C}(\mathbf{x})]_{ii}} \cdot \sqrt{[\mathbf{C}(\mathbf{x})]_{jj}}} = \frac{[\mathbf{C}(\mathbf{n})]_{ij}}{\sqrt{[\mathbf{C}(\mathbf{n})]_{ii}} \cdot \sqrt{[\mathbf{C}(\mathbf{n})]_{jj}}} = \frac{[\mathbf{C}(\mathbf{vmr})]_{ij}}{\sqrt{[\mathbf{C}(\mathbf{vmr})]_{ii}} \cdot \sqrt{[\mathbf{C}(\mathbf{vmr})]_{jj}}}$$

with $i, j = 1 \ldots n_{main}$ and $i \neq j$.

If $x_j^\dagger = [\mathbf{x}^\dagger]_j$ represents the "true" profile in partial columns on layer j, then the averaging kernel is defined by

$$[\mathbf{A}(\mathbf{x})]_{ij} = \frac{\partial x_i}{\partial x_j^\dagger}$$

The averaging kernels for the number density and VMR are then given by

$$[\mathbf{A}(\mathbf{n})]_{ij} = \frac{\partial n_i}{\partial n_j^\dagger} = \frac{s_i^\mathtt{nd}}{s_j^\mathtt{nd}}[\mathbf{A}(\mathbf{x})]_{ij},$$

and

$$[\mathbf{A}(\mathbf{vmr})]_{ij} = \frac{\partial vmr_i}{\partial vmr_j^\dagger} = \frac{s_i^\mathtt{vmr}}{s_j^\mathtt{vmr}}[\mathbf{A}(\mathbf{x})]_{ij},$$

respectively.

8.5. Common Characteristics of the Profile retrieval

For all profile retrievals the results of the Limb cloud detection are taken into account. In order to avoid cloud contamination, the first measurement data used for the profile retrieval is taken at the first *cloud free* height above the configured minimum height, i.e.

$$h_0 \quad = \quad \begin{cases} h_{min} & \text{for } \max(h_{cloud}) < h_{min} \\ h_{i+1} & \text{for } \max(h_{cloud}) > h_{min} \text{ with } i = \text{Index of } \max(h_{cloud}) \end{cases} \tag{8.102}$$

Retrievals below 20 km and above the reference height are expected to have larger errors.

8.6. Limb Ozone Profile Retrieval Settings

Input spectra	
Calibration	All calibrations applied
Retrieval Height Range	13.5 km (or lowest cloud free height) - 65 km
Reference Height	46 km (for 520 - 590 nm); 65 km (for 283 - 310 nm)
Fit Settings	
Fitting Interval	283 - 310 nm and 520 - 590 nm
Polynomial Degree	4th order
Number of Layers	33
Aerosol Model	LOWTRAN
Absorption Cross Sections/Fitted Curves	
O_3	Bogumil et al. (2003)@243 K
NO_2	Bogumil et al. (2003)@243 K
Profiles	
O_3	McLinden et al. (2002)
NO_2	McLinden et al. (2002)

8.7. Limb NO_2 Profile Retrieval Settings

Input spectra	
Calibration	All calibrations applied
Retrieval Height Range	13.5 km (or lowest cloud free height) - 46 km
Reference Height	43 km
Fit Settings	
Fitting Interval	420 - 470 nm
Polynomial Degree	3rd order
Number of Layers	33
Absorption Cross Sections/Fitted Curves	
O_3	Bogumil et al. (2003)@243 K
NO_2	Bogumil et al. (2003)@243 K
Profiles	
O_3	McLinden et al. (2002)
NO_2	McLinden et al. (2002)

8.8. Limb BrO Profile Retrieval Settings

Input spectra	
Calibration	All calibrations applied
Retrieval Height Range	10.5 km (or lowest cloud free height) - 35 km
Reference Height	35 km
Fit Settings	
Fitting Interval	337 - 357 nm
Polynomial Degree	4th order
Number of Layers	33
Absorption Cross Sections/Fitted Curves	
O_3	Bogumil et al. (2003)@243 K
BrO	Bogumil et al. (2003)@243 K
Profiles	
O_3	McLinden et al. (2002)
BrO	McLinden et al. (2002)

9. Limb Cloud Retrieval

9.1. Background

9.1.1. Motivation

In limb mode SCIAMACHY measures light scattered along the line-of-sight. If the line-of-sight intersects a cloud at a certain height, the spectrally resolved measurements differ from cloud free measurements. To eliminate systematic uncertainties and to enhance the sensitivity of the cloud determination, differences and/or ratios of spectral measurements in a certain wavelength region are used. Here we use radiance ratios at 750 nm and 1090 nm (and of 1550 nm and 1685 nm, see Section 9.3) for the detection of clouds. In an ideal Rayleigh scattering atmosphere, populated only by molecules, the difference between radiances at two wavelengths about 300 nm apart is large ($I \sim \lambda^{-4}$), while for larger particles like cloud droplets this difference is reduced (e.g., $I \sim \lambda^{-1}$, Section 9.1.2). The basic geometry and the principle of the SCIAMACHY cloud top height detection from limb measurements is shown in Figure 9.1. The satellite instrument detects different light from the cloud and above, which is then derived as the cloud top height (CTH) above the tangent point.

9.1.2. Theory

In earth's atmosphere radiation is scattered non-isotropically depending on the wavelength of the incident radiation, the atmospheric density as well as on the size and shape of the scattering particles. The distribution of the intensity I of scattered radiation is described by

- *Rayleigh scattering:* wavelength $\gg$ particle size (molecules like nitrogen or oxygen) and/or
- *Mie scattering:* wavelength $\ll$ particle size (aerosol, cloud particles as can be found in tropospheric clouds and polar stratospheric clouds (PSCs))

In the Rayleigh scattering case we find $I \sim \lambda^4$, whereas in the case of Mie scattering I is proportional to $\lambda^{-\alpha}$, with a value of α varying from 0 to 4. Due to the dependence of the spectral signature on the size of the scattering particles it is possible to derive information about the physical properties of clouds from comparisons of radiation at different wavelengths.

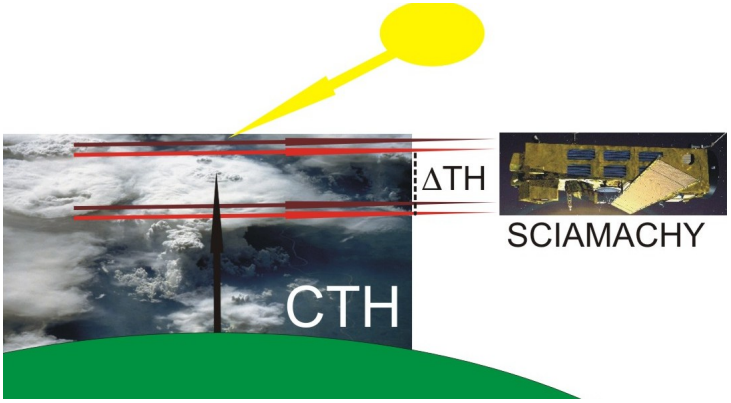

Figure 9.1.: *Illustration of the principle of the cloud top height detection using scattered solar radiation at clouds for two wavelengths in the near infra-red (red and brown darts). SCIAMACHY scans the limb in tangent height steps with a difference ΔTH=3.3 km. The cloud top height CTH is calculated for the tangent height (the vector being perpendicular to the ground and the line-of-sight).*

9.1.3. Product description

Using this method a detection tropospheric clouds (i.e. stratus, cumulus and cirrus), cirrus clouds at the tropical tropopause, stratospheric clouds in polar winters and noctilucent clouds near the polar summer mesopause is straightforward. While for every cloud type the detection is based on the same physical principles, each cloud type detection requires specific details. For example, detecting PSCs appropriately requires knowledge about the stratospheric aerosol content to avoid erroneous detection due to high levels of aerosol.

9.2. Algorithm Description

9.2.1. Physical Considerations

As cloud particles (diameter between 2 μm to 1.5 mm) are larger than the wavelength used in the UV-to-IR measurement range of SCIAMACHY (around 0.8 μm), scattered radiation tends to be scattered not uniformly in all directions, as is the case for the much smaller background air molecules (~0.1 nm). In general, the phase function of scattered radiation from water cloud droplets has a strong forward peak. On the other hand cloud droplets scatter radiation at different wavelengths more uniformly than molecules. Furthermore molecules absorb radiation in certain wavelength region.

To identify wavelength regions, which are suitable for radiance comparisons to detect clouds, certain requirements have to be fulfilled

1. Radiation at wavelengths below 400 nm is not suitable because the atmosphere gets optically thick for Rayleigh scattering in the upper troposphere at about 15-18 km;
2. Spectral windows with strong molecular absorption like for ozone, oxygen or water vapour are not suitable because of the height dependence of the absorption profiles.

For the identification ('flagging') of cloud contaminated pixels in the SCIAMACHY limb data and the retrieval of cloud top heights we choose two wavelength pairs of 750 nm and 1090 nm, and of 1550 nm and 1685 nm. The ratio of 750 nm and 1090 nm is used for the general detection of clouds. Radiance at 750 nm (SCIAMACHY channel 4) is basically free of absorption signatures, as ozone absorption is comparatively small (see Figure 9.2). It is situated between the H_2O absorption band at 725 nm and the O_2-A band at 760 nm. The 1090 nm radiance (SCIAMACHY channel 6) is just outside a H_2O absorption band ranging from 1100 nm to 1170 nm.

The wavelength pair of 1550 nm and 1685 nm is chosen for the detection of ice clouds, as the spectral signatures of ice is more pronounced in this wavelength region Kokhanovsky et al. (2005). The imaginary part of the refractive index is shown in Figure 9.3 taken from Acarreta et al. (2004). The 1550 nm region is outside the strong water vapour absorption band around 1400 nm.

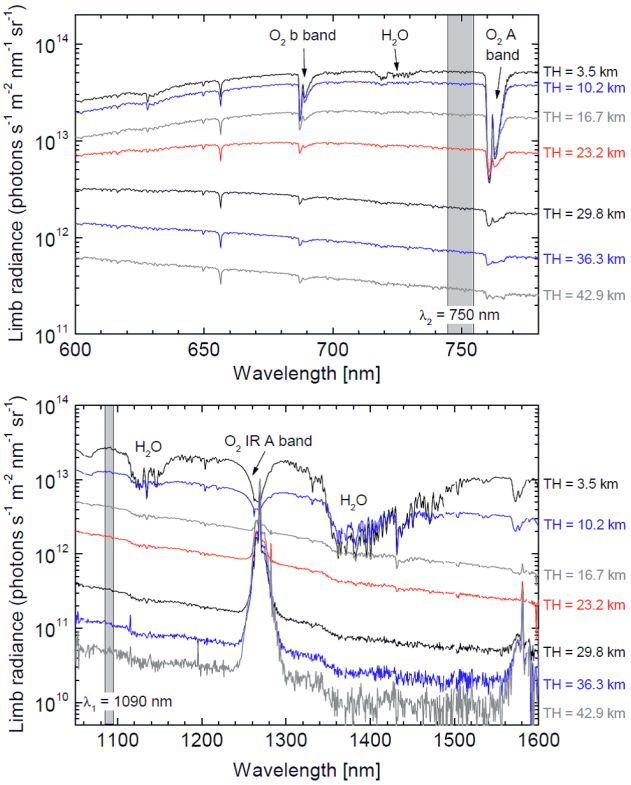

Figure 9.2.: *Calibrated Limb spectra for the SCIAMACHY channel 4 (top) and channel 6 (bottom) for an example measurement (von Savigny et al. (2005)).*

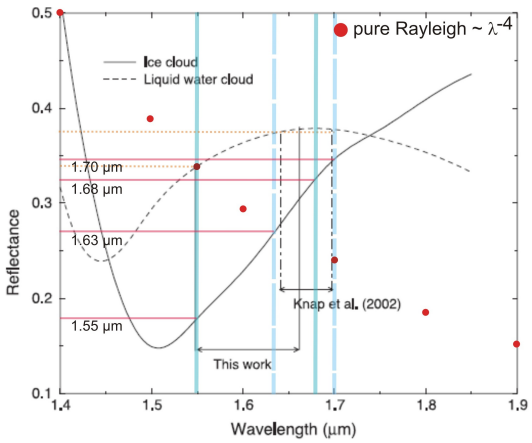

Figure 9.3.: *Taken from (Acarreta et al. (2004))and updated. The blue lines give wavelengths of SCIAMACHY channel 6+ suitable for ice cloud detection. The red dots indicate a spectral behaviour of a pure Rayleigh scattering atmosphere, showing the strong decrease of radiance.*

9.2.2. Mathematical Description of the Algorithm

To improve the signal-to-noise ratio SCIAMACHY measurements are integrated over the following wavelength windows: 750 – 751 nm, 1088 – 1092 nm, 1550 – 1553.2 nm and 1683 – 1687 nm. For the detection of cloudy pixels the height dependent colour index R_c is created from the ratio of the wavelength dependent intensities

$$R_c(TH) = \frac{I(\lambda_1, TH)}{I(\lambda_2, TH)} \tag{9.1}$$

TH denotes the tangent height for the SCIAMACHY line-of-sight, λ the wavelength, and I the intensity. For the determination of the cloud top, the colour index ratio profile is computed as follows

- ➢ Starting with the lowest height, pairs of two adjacent colour indices are divided.
- ➢ The maximum of the colour index ratio will then show the cloud top.

The colour index ratio $\Theta(TH)$ is defined as

$$\Theta(TH) = \frac{R_c(TH)}{R_c(TH + \Delta TH)} \tag{9.2}$$

For SCIAMACHY limb measurements the tangent height difference ΔTH is 3.3 km, which also gives the vertical resolution. Taking this difference into account, a limb measurement at e.g. 20 km tangent height has a horizontal resolution along track of about 400 km and due to the scan cycle of 240 km across track.

9.3. Application and Determination of Cloud Types

9.3.1. Tropospheric Clouds

The main reason for flagging cloudy pixels is to support a successful trace gas retrieval. A fully clouded scene at high altitudes will influence the quality of the retrieved trace gas profiles. For absorbers peaking in the troposphere like water vapour, cloudy scenes will foil a reliable retrieval. Thus it is mandatory to know if there is cloud contamination in the line-of-sight and what the cloud top height is.

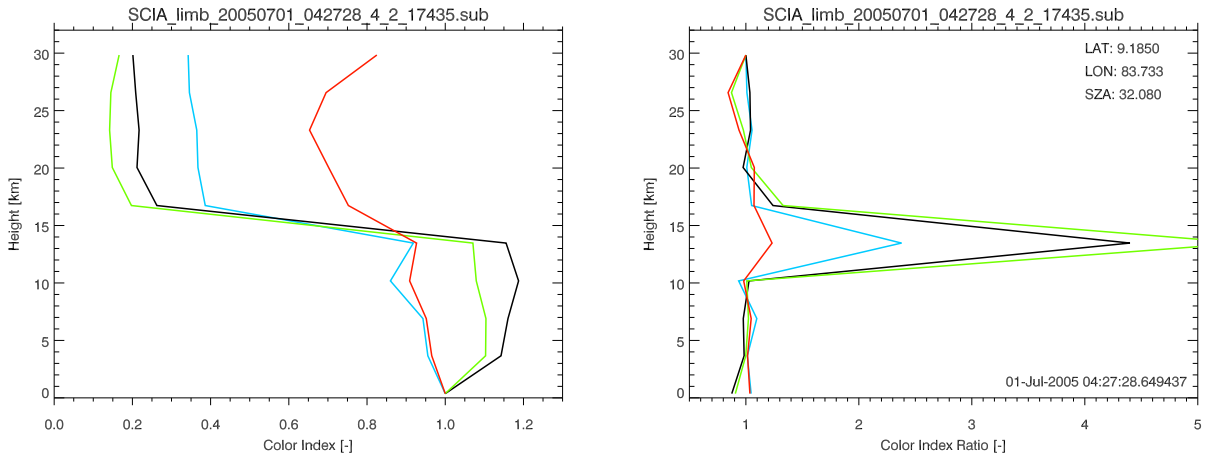

Figure 9.4.: *Colour index profiles of SCIAMACHY Limb radiances (left) and the corresponding colour index ratio profiles (right).*

An example for colour index profiles is shown in Figure 9.4 for a measurement in the tropics at 9.2° N and 83.7° E on 1 July 2005. The depicted lines in the altitude range from 0 to 30 km show different ratios

- ➢ 1090/750 nm (black line),
- ➢ 1685/1550 nm (red line).

The ratios 1550/750 nm (blue line) and 1685/750 nm (green line) are shown only for testing purposes. These ratios are redundant and are not used for the detection of clouds.

The near infra-red colour ratios may be used for the distinction of water and ice clouds. To have the same starting value at the bottom, the radiance profiles are divided by the lowermost radiance measured by SCIAMACHY. The right part of Figure 9.4 shows the colour index ratios for the same measurement. The cloud top can be found at the maximum value at a height of about 14.5 km. In Figure 9.5 the colour index ratios (CIR) along an entire SCIAMACHY orbit are shown. To flag a fully cloudy scene in one CIR profile a value of $\Theta = 2$ is taken (solid circles), while partly cloudy scenes (black rings) are assigned for values between 1.4 and 2. CIR below the threshold of 1.4 are considered as cloud free (depicted as full circles at 0 km altitude). The values are empirical and still preliminary. They will be verified using a larger dataset in the frame of an ongoing verification. In the upper part of the figure the cloud free region in the tropics shows slightly elevated values at 20 km, which could hint to cirrus clouds. This also is supported by the ice cloud detection as shown in the lower plot, where clouds are detected at these altitudes. The NIR gradients tend to have outliers at the upper boundary possibly due to small signal to noise ratios.

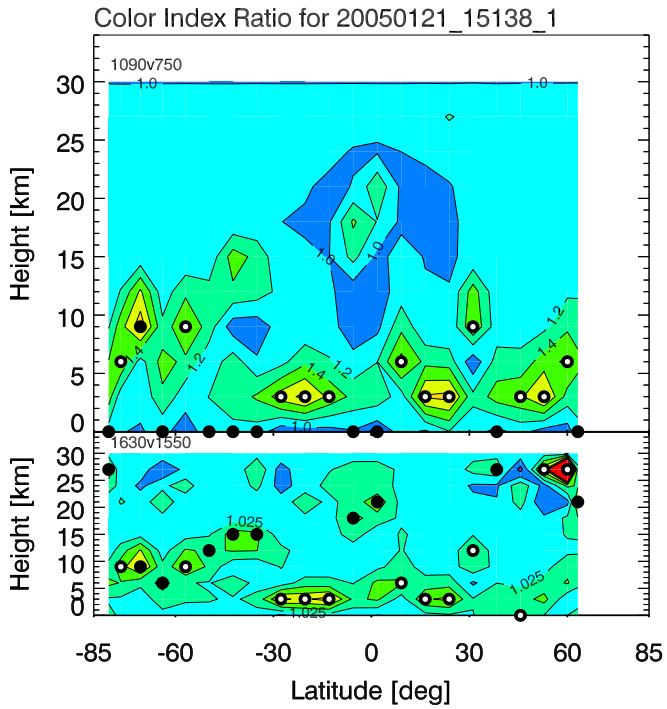

Figure 9.5.: *Colour index ratios for a SCIAMACHY orbit as a function of latitude and tangent height. The circles depict the cloud top heights. Full cloud coverage is given by solid circles above the ground, while partially cloudy scenes are illustrated by the black rings in the upper part of the plot. In the lower part solid circles depict ice phase clouds.*

9.3.2. Polar Stratospheric Clouds (PSCs)

While general tropospheric clouds are easy to detect, more sophisticated detection criteria are needed for PSCs (see Figure 9.4) to reduce false results. Currently the stratospheric aerosol content is low due to the absence of volcanic eruptions for longer time periods. The wavelength ratio of 750 nm and 1090 nm is used for PSC detection. To distinguish between an aerosol loaded stratosphere and an optically thin PSC, a threshold of the colour index ratio of $\Theta(TH) > 1.35$(for the northern hemisphere); > 1.3 (for the southern one) was chosen. This value is a result of model studies of von Savigny et al. (2005). High altitude cirrus clouds can also be confused with PSCs, so as a further constraint, $\Theta(TH)$ should be taken from measurements above the climatological tropopause. A simple constraint is to choose a tangent height between 15 and 30 km and to look for PSCs only above 50° latitude.

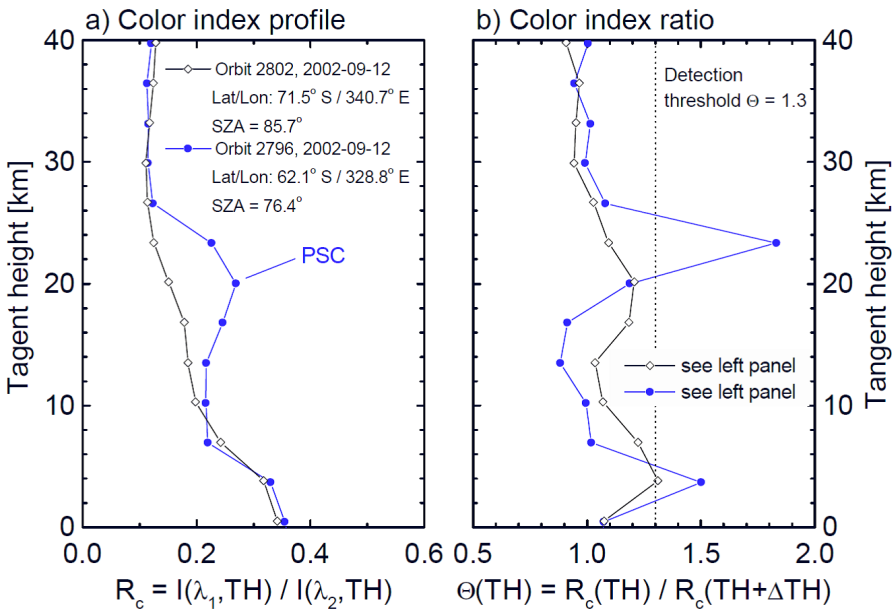

Figure 9.6.: *Colour index profile (a) and colour index ratios (b) in case of a PSC contaminated measurement (blue) and a background profile (black).*

9.3.3. Noctilucent Clouds (NLCs)

Noctilucent clouds (NLCs) are a mesospheric phenomenon (80 and 85 km), occurring at high latitudes during summertime, when the mesopause region is very cold. A typical season starts about 3–4 weeks before the summer solstice and persists for approximately 3 months (Robert et al. (2009)). These clouds can frequently be detected from ground. UV-radiation at 265 nm and 291 nm is used for the detection of clouds at these heights, because multiple scattering and reflection from the ground can be neglected for the limb scattering geometry. Two mechanisms have been implemented to establish an NLC detection method.

1) For stronger NLCs differencing is a feasible way for detection:

$$\Theta_D(TH) = I(TH) - I(TH + \triangle TH) \tag{9.3}$$

A negative difference (Eq.9.3) at tangent heights between 75 km and 90 km indicates NLCs in the line-of-sight (see Fig. 9.7).

2) A division of the radiance is used to get the colour ratio for weaker NLCs:

$$\Theta(TH) = \frac{R_c(TH)}{R_c(TH + \Delta TH)} \tag{9.4}$$

If both wavelengths give ratios $\Theta(TH) > 3$, NLCs are detected (von Savigny et al. (2004)). This is for weak NLCs that do not have a peak in radiance at these altitudes, but only a bump.

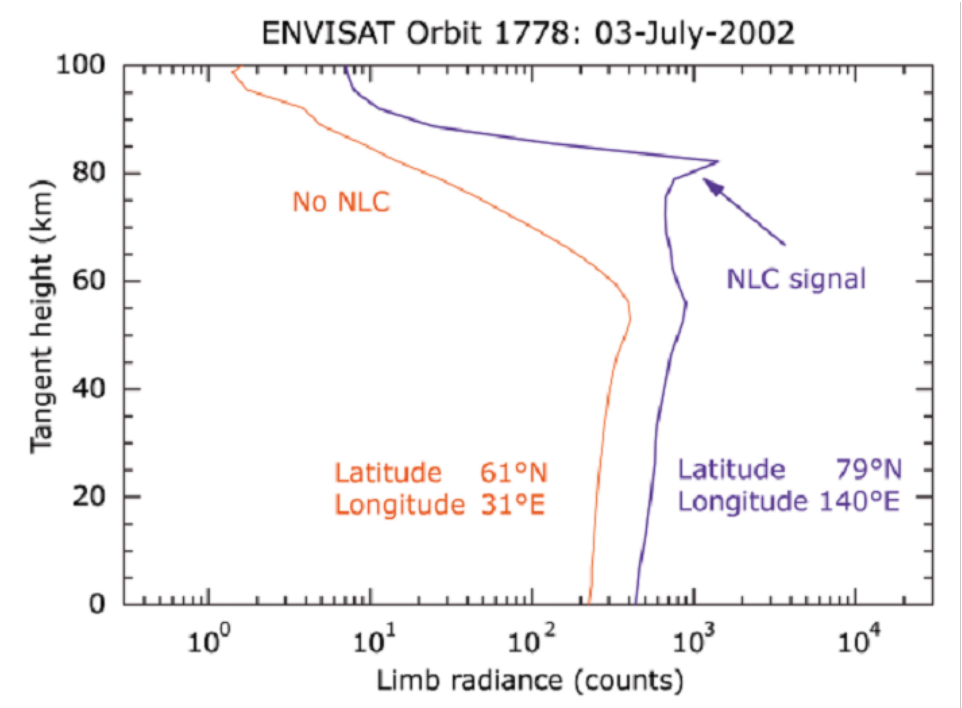

Figure 9.7.: *D*etection of NLC due to the increase in backscattered limb radiance above 80 km (purple) and a background measurement (red).

9.4. Summary and Implementation

Cloud detection can substantially influence the limb results, if a limb profile is sensitive to cloud contamination with respect to the spectral measurement. A possible drawback could be that results may be biased towards an atmosphere that is too dry. This is because clouds need moisture to evolve. This moisture gets transported in higher altitudes. If cloudy scenes are removed the resulting picture of the atmosphere represents a too dry environment.

On average the altitude coverage of profiles may change because, due to cloud, indexing profiles from lower altitudes may be skipped as they are flagged cloud contaminated.

On the other hand, a cloud index may improve results significantly by sorting out cloud contaminated scenes. A cloud index in principle also inherits the possibility to distinguish between different cloud types.

Currently it has not been investigated what influence the coverage of the field of view (FOV) of SCIAMACHY has (i.e. partial coverage, horizontal coverage, vertical coverage).

We recommend to carefully choose the wavelength region for setting up the cloud index. The wavelength regions shown in this ATBD are not the only wavelengths where a detection of clouds is possible but they have proven to work. A final determination of the quality of the detection with this set up is currently work in progress. We generally recommend using a cloud index, because sorting out obstacles from the FOV is essential for limb sounding.

9.4.1. Exception Rules

As the SCIAMACHY data is not perfect, especially in channel 6 at wavelengths near 1685 nm , it is necessary to sort out unexpected, non physical radiances. Thus a few exception rules were implemented to avoid extreme colour index ratios due to high or negative radiances.

1. As it very unusual to have negative radiances in the three wavelength bands (750 nm, 1090 nm, 1550 nm), the retrieval is stopped, if this shall happen.
2. For the band 1685 nm two tests are made:
   a) The wavelengths where negative radiances occur at any tangent height below the maximum height of 35 km are not used for the calculation of the mean.
   b) If unusually high radiances were found within the range, they will also be excluded. This is done by calculating the mean and the standard deviation of the sum of the radiances at the tangent heights up to the maximum height.

Part IV.

Tropospheric Products Algorithms

10. Tropospheric NO_2

10.1. Motivation

Tropospheric NO_2 is an air pollutant negatively affecting ecosystems and human health. The main anthropogenic sources of NO_2 in the troposphere are fossil fuel combustion and biomass burning. There are natural sources as well. These are microbial production in soils, wildfires and lightning.

10.2. Retrieval Settings

Nadir Settings	
	see Section 5.4.2
Limb Settings	
	see Section 8.7
Limb-Nadir Matching Settings	
Tropopause height	ECMWF ERA-Interim re-analysis
Cloud limits	radiance cloud fraction < 50% (corresponds to effective CF < 0.2)
Background Ref. Sector	180 – 220° (Pacific)
AMF Settings	
Stratospheric	LIDORT (HALOE NO_2 Profiles Climatology)
Tropospheric	Look-up table of the box AMFs used in the operational processing of the GOME-2 data Tropospheric NO_2 profiles from the MOZART CTM

The retrieval of tropospheric NO_2 columns from SCIAMACHY measurements is performed in several steps making use of various SCIAMACHY products (NO_2 nadir slant columns, NO_2 limb profiles) as well as several external data sets needed for the calculation of tropospheric air mass factors. Additional normalization is necessary using a background data base of values over the Pacific reference sector. As NO_2 cross-sections depend on temperature and in the total column retrieval temperatures in the stratosphere are assumed, a correction needs to be applied for $AMF_{tropospheric}$ calculation accounting for differing temperature in the troposphere.

In principle, the retrieval approach is simple. Nadir observations provide columns integrated over the full atmosphere down to the ground (note that only almost cloudless nadir pixels - CF < 0.2 - are processed). Limb measurements on the other hand provide a stratospheric profile, which can be integrated to the stratospheric column. The difference of the two measurements provides the tropospheric column. In practice, the subtraction of the stratospheric contribution has to be performed on the slant columns in order to account for the different sensitivity of the nadir measurements in different altitudes. Therefore, the limb profiles are integrated and for each nadir pixel converted to the stratospheric slant columns using $AMF_{stratospheric}$ accounting for the given viewing conditions (SZA, LOS, relative azimuth). Since many more nadir measurements than limb profiles are performed within one state, the limb data need to be interpolated to the position of the nadir pixels. This is

done by first computing all limb columns for a complete orbit, then interpolating them in latitude for each of the 4 viewing angles separately. This yields four VC_{strat} values for every nadir state. These values are then linearly interpolated and extrapolated in viewing angle to the respective nadir position.

A further complication arises from small systematic differences between limb and nadir columns, which need to be corrected for. This is done by computing the average nadir and limb slant columns over the Pacific reference sector at all latitudes and correcting the limb data with the difference between the two values.

The overall retrieval for each measurement can be broken down in the following steps:

1. Computation of latitudinal dependent offset between limb and nadir slant columns
2. Computation of stratospheric nadir slant column using offset-corrected limb profiles
3. Computation of tropospheric slant column using nadir slant columns and calculated stratospheric slant columns
4. Addition of a tropospheric background to account for tropospheric NO_2 in the Pacific reference sector
5. Computation of tropospheric vertical column applying $AMF_{tropospheric}$

The schematic chart visualizing the retrieval steps is depicted in Figure 10.1. L2 input is colored green, databases blue, intermediate and final results yellow and processing steps red.

Figure 10.1.: Flow chart of the tropospheric NO2 retrieval algorithm.

11. Tropospheric BrO

11.1. Motivation

Bromine monoxide plays a significant role in the tropospheric chemistry, especially in the polar regions. It was discovered that during spring time large amounts of bromine species are released in the boundary layer of both polar regions. This phenomenon was named “polar bromine explosion”. So far it is known that bromine is emitted from sea salt during a number of photochemical and heterogeneous reactions. This mechanism, however, is still not completely understood, in particular roles of wind strength or sea ice age. But it is absolutely evident that this phenomenon leads to severe ozone depletion in the polar troposphere.

On a smaller scale tropospheric bromine emissions have been observed also over salt lakes, in the marine boundary layer and in volcanic plumes.

The tropospheric BrO retrieved from the SCIAMACHY measurements provides a very useful data set, which will undoubtedly be helpful in further studies of polar bromine explosions. It will also improve our knowledge about the existence of a tropospheric BrO background at the global scale.

The operational tropospheric BrO retrieval is based on the scientific algorithm developed by BIRA (**?**).

11.2. Retrieval Settings

11.2.1. Splitting of the tropospheric and tropospheric parts

Details of the BrO total column retrieval are described in Section 5.6. Here a splitting of the total column into the tropospheric and the stratospheric parts is explained.

The stratospheric BrO profiles are calculated using the dynamical BrO climatology (Theys et al., 2008). The climatology is based on the output of the 3-D chemical transport model BASCOE. The impact of the atmospheric dynamics on the stratospheric BrO distribution is treated by means of $^{Bry}/_{O3}$ correlations, while photochemical effects are taken into account using stratospheric NO_2 columns as an indicator for the $^{BrO}/_{Bry}$ ratio. In practice, for each pixel, a stratospheric BrO profile is calculated from the climatology using the total O_3 (Nadir product) and stratospheric NO_2vertical columns (intermediate result from the tropospheric NO_2 product). The stratospheric BrO vertical columns (VCDstrato) are derived by integrating climatological stratospheric BrO profiles between the tropopause and the top- of-atmosphere. VCDstrato is than converted into the SCDstrato:

$$SCDstrato = VCDstrato * AMFstrato \tag{11.1}$$

Tropospheric slant column is calculated by subtracting the stratospheric part from the total slant columns:

$$SCDtropo = SCDtotal - SCDstrato \tag{11.2}$$

During the verification phase it has been discovered that there is an offset between the operational and the scientific (BIRA) BrO slant columns. Besides the operational slant columns also demonstrate a time trend not confirmed by any satellite or ground-based sensors. In order to get rid of this offset an equatorial correction was recommended by the scientific product developers. For a given day, all SCDs (from all orbits of this day) in the equatorial latitudinal band (5°S-5°N) are taken and the average SCD in this region is estimated. This average SCDs is then subtracted from all SCDs of the day. Finally, a constant SCD value of 7.5e13 molec/cm² is added to compensate for the equatorial value. By doing so, any constant offset is automatically removed.

The equatorial correction values are pre-calculated from the latest completely processed L2 version. Here, an assumption is made that equatorial BrO SCDs do not change from version to version. This assumption has been confirmed by comparison of the equatorial corrections based on L2v6.01 with those computed from the verification data set of the current L2 version (L2v7).

Finally, the tropospheric column is computed as follows:

$$VCDtropo = \frac{SCDtropo}{AMFtropo} \tag{11.3}$$

Nadir Settings	
	see Section 5.6
Stratospheric Part	
	see Subsection 11.2.1
Limits	
Cloud limit	cloud fraction < 0.4
AMFtropo limit	AMFtropo > 0.5 (to ensure that measurements are sensitive enough to the tropospheric BrO)
SZA limit	80°
Used Look-up Tables	
AMF Stratospheric	Look-up table of the box AMFs used in the BIRA scientific processor Stratospheric BrO profiles from the BASCOE CTM
AMF Tropospheric	Look-up table of the box AMFs used in the BIRA scientific processor Gaussian tropospheric BrO profile with a maximum at 6 km and a full width half maximum of 2 km
Tropopause height	ECMWF ERA-Interim re-analysis

The schematic chart visualizing the retrieval steps is depicted in Figure 11.1. L2 input is colored green, databases blue, intermediate and final results yellow and processing steps red.

This tropospheric BrO will become a part of the current total BrO product. Following the new approach the total BrO column results from the sum of BrOstrato and BrOtropo.

Figure 11.1.: Flow chart of the tropospheric BrO retrieval algorithm.

Bibliography

Abramowitz, M. and Stegun, I.: Handbook of Mathematical Functions, National Bureau of Standards, AMS55, New York, 1964.

Acarreta, J. R., Stammes, P., and Knap, W. H.: First retrieval of cloud phase from SCIAMACHY spectra around 1.6 @mm, Atmospheric Research, 72, 89–105, 2004.

Anderson, G., Clough, S., Kneizys, F., Chetwynd, J., and Shettle, E.: AFGL Atmospheric Constituent Profiles ($0 - 120\,\mathrm{km}$), Tech. Rep. TR-86-0110, AFGL, 1986.

Armstrong, B.: Spectrum Line Profiles: The Voigt Function, J. Quant. Spectrosc. & Radiat. Transfer, 7, 61–88, doi:$10.1016/0022\text{-}4073(67)90057\text{-}\mathrm{X}$, 1967.

Bogumil, K., Orphal, J., Homann, T., Voigt, S., Spietz, P., Fleischmann, O. C., Vogel, A., Hartmann, M., Kromminga, H., Bovensmann, H., Frerick, J., and Burrows, J. P.: Measurements of molecular absorption spectra with the SCIAMACHY pre-flight model: instrument characterization and reference data for atmospheric remote-sensing in the 230-2380 nm region, Journal of Photochemistry and Photobiology, 157, 167–184, 2003.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY: Mission Objectives and Measurement Modes, Journal of Atmospheric Sciences, 56, 127–150, doi:$10.1175/1520\text{-}0469(1999)056$, 1999.

Brion, J., Chakir, A., Charbonnier, J., Daumont, D., Parisse, C., and Malicet, J.: Absorption spectra measurements for the ozone molecule in the 350-830 nm region, J. Atmos. Chem., 30, 291–299, 1998.

Buchwitz, M., Khlystova, I., Bovensmann, H., and Burrows, J. P.: Three years of global carbon monoxide from SCIAMACHY: comparison with MOPITT and first results related to the detection of enhanced CO over cities, 7, 2399–2411, 2007.

Carlotti, M., Höpfner, M., Raspollini, P., Ridolfi, M., and Carli, B.: High level algorithm definition and physical and mathematical optimisations, Tech. Rep. TN-IROE-RSA9601, IROE, Florence, Italy, 1998.

Chance, K. and Spurr, R.: Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum., Applied optics, 36, 5224–5230, 1997.

Chance, K. V. and Spurr, R. J. D.: Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, App. Opt., 36, 5224–5230, doi:$10.1364/\mathrm{AO}.36.005224$, 1997.

Chandrasekhar, S.: Radiative Transfer, Dover, Mineola, New York, 1960.

Clough, S. and Kneizys, F.: Convolution Algorithm for the Lorentz function, Appl. Opt., 18, 2329–2333, 1979.

Clough, S., Kneizys, F., Anderson, G., Shettle, E., Chetwynd, J., Abreu, L., Hall, L., and Worsham, R.: FASCOD3: spectral simulation, in: IRS'88: Current Problems in Atmospheric Radiation, edited by Lenoble, J. and Geleyn, J., pp. 372–375, A. Deepak Publishing, 1988.

Clough, S., Kneizys, F., and Davies, R.: Line Shape and the Water Vapor Continuum, Atmos. Res., 23, 229–241, 1989.

de Haan, J., Bosma, P., and Hovenier, J.: The adding method for multiple scattering calculations of polarised light, Astron. Astrophys., 183, 371–391, 1987.

Dennis, Jr., J., Gay, D., and Welsch, R.: An Adaptive Nonlinear Least–squares Algorithm, 7, 348–368, 1981.

Deschamps, P., Breon, F., Leroy, M., Podaire, A., Bricaud, A., Buriez, J., and Seze, G.: The POLDER mission: instrument characteristics and scientific objectives, IEEE Transactions on Geoscience and Remote Sensing, 32, 598–615, doi:$10.1109/36.297978$, 1994.

Edwards, D.: Atmospheric transmittance and radiance calculations using line–by–line computer models, in: Modelling of the Atmosphere, vol. 928, pp. 94–116, 1988.

Fleischmann, O., Hartmann, M., Burrows, J. P., and Orphal, J.: New ultraviolet absorption cross-sections of BrO at atmospheric temperatures measured by time-windowing Fourier transform spectroscopy, J. Photochem. Photobiol. A, 168, 117–132, 2004.

Fomin, B.: Effective Interpolation technique for line–by–line calculation of radiation absorption in gases, J. Quant. Spectrosc. & Radiat. Transfer, 53, 663–669, 1995.

Frankenberg, C., Platt, U., and Wagner, T.: Iterative maximum a posteriori (IMAP)-DOAS for retrieval of strongly absorbing trace gases: Model studies for CH_4 and CO_2 retrieval from near infrared spectra of SCIAMACHY onboard ENVISAT, 5, 9–22, 2005.

Gimeno García, S., Schreier, F., Lichtenberg, G., and Slijkhuis, S.: Near infrared nadir retrieval of vertical column densities: methodology and application to SCIAMACHY, Atmos. Meas. Tech., 4, 2633–2657, doi: 10.5194/amt-4-2633-2011, 2011.

Gloudemans, A., Schrijver, H., Kleipool, Q., van den Broek, M., Straume, A., Lichtenberg, G., van Hees, R., Aben, I., and Meirink, J.: The impact of SCIAMACHY near-infrared instrument calibration on CH_4 and CO total columns, 5, 2369–2383, 2005.

Golub, G. and Pereyra, V.: Separable nonlinear least squares: the variable projection method and its applications, 19, R1–R26, 2003.

Gordon, I., Rothman, L., et al.: The HITRAN2016 molecular spectroscopic database, J. Quant. Spectrosc. & Radiat. Transfer, 203, 3 – 69, doi:10.1016/j.jqsrt.2017.06.038, 2017.

Greenblatt, G. D., Orlando, J. J., Burkholder, J. B., and Ravishankara, A. R.: Absorption measurements of oxygen between 330 and 1140 nm, J. Geophys. Res., 95, 18 577–18 582, doi:10.1029/JD095iD11p18577, 1990.

Herman, J. R., Bhartia, P. K., Torres, O., Hsu, C., Seftor, C., and Celarier, E.: Global distribution of UV-absorbing aerosols from Nimbus 7/TOMS data, Journal of Geophysical Research, 102, 16 911–16 922, doi: 10.1029/96JD03680, 1997.

Hermans, C., Vandaele, A. C., Carleer, M., Fally, S., Colin, R., Jenouvrier, A., Coquart, B., and Mèrienne, M.-F.: Absorption cross-sections of atmospheric constituents: NO_2, O_2, and H_2O, Environ. Sci. Pollut. R., 6, 151–158, 1999.

Humlíček, J.: Optimized computation of the Voigt and complex probability function, J. Quant. Spectrosc. & Radiat. Transfer, 27, 437–444, doi:10.1016/0022-4073(82)90078-4, 1982.

Jacquinet-Husson, N., Scott, N., Chedin, A., Crepeau, L., Armante, R., Capelle, V., Orphal, J., Coustenis, A., Boonne, C., Poulet-Crovisier, N., Barbe, A., Birk, M., Brown, L., Camy-Peyret, C., Claveau, C., Chance, K., Christidis, N., Clerbaux, C., Coheur, P., Dana, V., Daumont, L., Backer-Barilly, M. D., Lonardo, G. D., Flaud, J., Goldman, A., Hamdouni, A., Hess, M., Hurley, M., Jacquemart, D., Kleiner, I., Kopke, P., Mandin, J., Massie, S., Mikhailenko, S., Nemtchinov, V., Nikitin, A., Newnham, D., Perrin, A., Perevalov, V., Pinnock, S., Regalia-Jarlot, L., Rinsland, C., Rublev, A., Schreier, F., Schult, L., Smith, K., Tashkun, S., Teffo, J., Toth, R., Tyuterev, V., Auwera, J. V., Varanasi, P., and Wagner, G.: The GEISA spectroscopic database: Current and future archive for Earth and planetary atmosphere studies, J. Quant. Spectrosc. & Radiat. Transfer, 109, 1043–1059, doi:10.1016/j.jqsrt.2007.12.015, 2008.

Jacquinet-Husson, N. et al.: The 2015 edition of the GEISA spectroscopic database, 327, 31 – 72, doi:10.1016/j.jms.2016.06.007, 2016.

Kahaner, D., Moler, C., and Nash, S.: Numerical Methods and Software, Prentice–Hall, Englewood Cliffs, NJ, 1989.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation, Bull. Am. Met. Soc., 82, 247–267, doi:10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2, 2001.

Kokhanovsky, A. A. and Rozanov, V. V.: The physical parameterization of the top-of-atmosphere reflection function for a cloudy atmosphere-underlying surface system: the oxygen A-band case study, Journal of Quantitative Spectroscopy and Radiative Transfer, 85, 35–55, doi:10.1016/S0022-4073(03)00193-6, 2004.

Kokhanovsky, A. A., Rozanov, V. V., Zege, E. P., Bovensmann, H., and Burrows, J. P.: A semianalytical cloud retrieval algorithm using backscattered radiation in 0.4-2.4 μm spectral region, Journal of Geophysical Research (Atmospheres), 108, 4008–+, doi:10.1029/2001JD001543, 2003.

Kokhanovsky, A. A., Rozanov, V. V., Burrows, J. P., Eichmann, K.-U., Lotz, W., and Vountas, M.: The SCIAMACHY cloud products: Algorithms and examples from ENVISAT, Advances in Space Research, 36, 789–799, doi:10.1016/j.asr.2005.03.026, 2005.

Kondratyev, K. Y. and Binenko, V.: Impact of Cloudiness on Radiation and Climate, Leningrad: Gidrometeoizdat, 1984.

Kretschel, K.: ENVISAT-1 SCIAMACHY Level 1b to 2 Offline Processing, Architectural Design Document, issue 2 (ENV-IS-DLR-SCI-2200-0005), Tech. rep., DLR, 2006.

Krijger, J. M.: SCIAMACHY degradation model ATBD, SRON-SQWG-TN-2011-001, draft, Tech. rep., SRON, 2011.

Krijger, J. M., Aben, I., and Schrijver, H.: Distinction between clouds and ice/snow covered surfaces in the identification of cloud-free observations using SCIAMACHY PMDs, Atmospheric Chemistry & Physics, 5, 2729–2738, 2005.

Kromminga, H., Orphal, J., Spietz, P., Voigt, S., and Burrows, J. P.: The temperature dependence (213-293 K) of the absorption cross-sections of OClO in the 340-450 nm region measured by Fourier-transform spectroscopy, Journal of Photochemistry and Photobiology A, 157, 149–160, 2003.

Lambert, J.-C., Granville, J., Van Roozendael, M., Müller, J.-F., Goutail, F., Pommereau, J. P., Sarkissian, A., Johnston, P. V., and Russell III, J. M.: Global Behaviour of Atmospheric NO2 as Derived from the Integrated Use of Satellite, Ground-based Network and Balloon Observations, in: Atmospheric Ozone - 19th Quad. Ozone Symp., Sapporo, Japan, edited by NASDA, pp. 201–202, 2000.

Lerot, C., Van Roozendael, M., van Geffen, J., van Gent, J., Fayt, C., Spurr, R., Lichtenberg, G., and von Bargen, A.: Six years of total ozone column measurements from SCIAMACHY nadir observations, Atmospheric Measurement Techniques, 2, 87–98, 2009.

Lichtenberg, G. and Kretschel, K.: SCIAMACHY Level 1b-2 Off-line Data Processing Configuration Management Of Level 1b-2 Auxiliary Data Files, issue 4 (ENV-CMA-DLR-SCIA-0061), Tech. rep., DLR, 2009.

Liou, K.-N.: An Introduction to Atmospheric Radiation, Academic Press, Orlando, 1980.

Loeb, N. G. and Davies, R.: Observational evidence of plane parallel model biases: Apparent dependence of cloud optical depth on solar zenith angle, J. Geophys. Res., 101, 1621–1634, doi:10.1029/95JD03298, 1996.

Loyola, D.: Cloud retrieval for SCIAMACHY, in: Proc. of ERS-ENVISAT Symposium, Gothenburg, Sweden, 16-20 Oct 2000, vol. SP-641, ESA, 2000.

Loyola, D. and Ruppert, T.: A new PMD cloud-recognition algorithm forGOME, ESA Earth Observation Quarterly, 58, 45–47, 1998.

McLinden, C. A., McConnell, J., Griffioen, E., and McElroy, C.: A vector radiative–transfer model for the Odin/OSIRIS project, Can J Phys., 80, 375–93, 2002.

Meller, R. and Moortgat, G. K.: Temperature dependence of the absorption cross section of HCHO between 223 and 323K in the wavelength range 225-375 nm, J. Geophys. Res., 105, 7089–7102, doi:10.1029/1999JD901074, 2000.

Müller, J.-F. and Brasseur, G.: IMAGES: A three-dimensional chemical transport model of the global troposphere, J. Geophys. Res., 100, 16 445–16 490, 1995.

Noël, S., Buchwitz, M., Bovensmann, H., Hoogen, R., and Burrows, J. P.: Atmospheric water vapor amounts retrieved from GOME satellite data, Geophys. Res. Lett., 26, 1841–1844, doi:10.1029/1999GL900437, 1999.

Norton, R. and Rinsland, C.: ATMOS data processing and science analysis methods, Appl. Opt., 30, 389–400, 1991.

Olivero, J. and Longbothum, R.: Empirical Fits to the Voigt Line Width: a brief Review, J. Quant. Spectrosc. & Radiat. Transfer, 17, 233–236, doi:10.1016/0022-4073(77)90161-3, 1977.

Pickett, H., Poynter, R., Cohen, E., Delitsky, M., Pearson, J., and Müller, H.: Submillimeter, millimeter, and microwave spectral line catalog, J. Quant. Spectrosc. & Radiat. Transfer, 60, 883–890, 1998.

Platt, U.: Differential optical absorption spectroscopy (DOAS): in Air Monitoring by Spectroscopic Techniques, John Willey, New York, 1994.

Platt, U. and Stutz, J.: Differential Optical Absorption Spectroscopy: Principles and Applications, Springer Verlag, 2008.

Pope, R. M. and Fry, E. S.: Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurements, Appl. Opt., 36, 8710–8723, 1997.

Ralston, A. and Rabinowitz, P.: A First Course in Numerical Analysis, McGraw–Hill Book Company, second edn., 1978.

Robert, C. E., von Savigny, C., Burrows, J. P., and Baumgarten, G.: Climatology of noctilucent cloud radii and occurrence frequency using SCIAMACHY, Journal of Atmospheric and Solar-Terrestrial Physics, 71, 408–423, doi:10.1016/j.jastp.2008.10.015, 2009.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, Atmos. Oceanic Planet. Phys., World Scientific Publishing, Singapore, 2000.

Rossow, W. B.: Measuring Cloud Properties from Space: A Review., Journal of Climate, 2, 201–213, doi:10.1175/1520-0442(1989)002<0201:MCPFSA>2.0.CO;2, 1989.

Rothman, L., Gamache, R., Goldman, A., Brown, L., Toth, R., Pickett, H., Poynter, P., Flaud, J.-M., Camy-Peyret, C., Barbe, A., Husson, N., Rinsland, C., and Smith, M.: The HITRAN database: 1986 edition, Appl. Opt., 26, 4058, 1987.

Rothman, L., Barbe, A., Benner, D. C., Brown, L., Camy-Peyret, C., Carleer, M., Chance, K., Clerbaux, C., Dana, V., Devi, V., Fayt, A., Flaud, J.-M., Gamache, R., Goldman, A., Jacquemart, D., Jucks, K., Lafferty, W., Mandin, J.-Y., Massie, S., Nemtchinov, V., Newnham, D., Perrin, A., Rinsland, C., Schroeder, J., Smith, K., Smith, M., Tang, K., Toth, R., Auwera, J. V., Varanasi, P., and Yoshino, K.: The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001, J. Quant. Spectrosc. & Radiat. Transfer, 82, 5–44, doi:10.1016/S0022-4073(03)00146-8, 2003.

Rothman, L., Gordon, I., Barbe, A., Benner, D. C., Bernath, P., Birk, M., Boudon, V., Brown, L., Campargue, A., Champion, J.-P., Chance, K., Coudert, L., Dana, V., Devi, V., Fally, S., Flaud, J.-M., Gamache, R., Goldman, A., Jacquemart, D., Kleiner, I., Lacome, N., Lafferty, W., Mandin, J.-Y., Massie, S., Mikhailenko, S., Miller, C., Moazzen-Ahmadi, N., Naumenko, O., Nikitin, A., Orphal, J., Perevalov, V., Perrin, A., Predoi-Cross, A., Rinsland, C., Rotger, M., Simecková, M., Smith, M., Sung, K., Tashkun, S., Tennyson, J., Toth, R., Vandaele, A., and Auwera, J. V.: The HITRAN 2008 molecular spectroscopic database, J. Quant. Spectrosc. & Radiat. Transfer, 110, 533 – 572, doi:10.1016/j.jqsrt.2009.02.013, 2009.

Rozanov, V. V. and Kokhanovsky, A. A.: Semianalytical cloud retrieval algorithm as applied to the cloud top altitude and the cloud geometrical thickness determination from top-of-atmosphere reflectance measurements in the oxygen A band, Journal of Geophysical Research (Atmospheres), 109, 5202–+, doi:10.1029/2003JD004104, 2004.

Rozanov, V. V., Kurosu, T., and Burrows, J. P.: Retrieval of atmospheric constituents in the UV-visible: a new quasi-analytical approach for the calculation of weighting functions., Journal of Quantitative Spectroscopy and Radiative Transfer, 60, 277–299, doi:10.1016/S0022-4073(97)00150-7, 1998.

Rozanov, V. V., Buchwitz, M., Eichmann, K.-U., de Beek, R., and Burrows, J. P.: Sciatran - a new radiative transfer model for geophysical applications in the 240-2400 NM spectral region: the pseudo-spherical version, Advances in Space Research, 29, 1831–1835, doi:10.1016/S0273-1177(02)00095-9, 2002.

Scherbakov, D.: SCIAMACHY Command Line Tool SciaL1c Software User's Manual, issue 2B (ENV-SUM-DLR-SCIA-0071), Tech. rep., DLR, 2008.

Schreier, F.: Optimized Implementations of Rational Approximations for the Voigt and Complex Error Function, J. Quant. Spectrosc. & Radiat. Transfer, 112, 1010–1025, doi:10.1016/j.jqsrt.2010.12.010, 2011.

Schreier, F., Gimeno García, S., Hedelt, P., Hess, M., Mendrok, J., Vasquez, M., and Xu, J.: GARLIC – A General Purpose Atmospheric Radiative Transfer Line-by-Line Infrared-Microwave Code: Implementation and Evaluation, J. Quant. Spectrosc. & Radiat. Transfer, 137, 29–50, doi:10.1016/j.jqsrt.2013.11.018, 2014.

Slijkhuis, S. and Lichtenberg, G.: ENVISAT-1 SCIAMACHY Level 0 to 1c Processing, Algorithm Theoretical Basis Document, issue 6 (ENV-ATB-DLR-SCIA-0041), Tech. rep., DLR, 2014.

Sparks, L.: Efficient Line–by–Line Calculation of Absorption Coefficients to High Numerical Accuracy, J. Quant. Spectrosc. & Radiat. Transfer, 57, 631–650, doi:10.1016/S0022-4073(96)00154-9, 1997.

Spurr, R.: SCIAMACHY Level 1b to 2 Off-line Processing, Tech. Rep. ENV-ATB-SAO-SCIA-2200-0003, DLR–DFD and Harvard–CfA–SAO, 1998.

Spurr, R., Kurosu, T., and Chance, K.: A Linearized discrete Ordinate Radiative Transfer Model for Atmospheric Remote Sensing, Journal of Quantitative Spectroscopy and Radiative Transfer, 68, 689–735, 2001.

Spurr, R., van Roozendael, M., and Loyola, D. G.: Algorithm Theoretical Basis Document for GOME Total Column Densities of Ozone and Nitrogen Dioxide, Tech. rep., DLR, 2004.

Theys, N., van Roozendael, M., Errera, Q., Hendrick, F., Daerden, F., Chabrillat, S., Dorf, M., Pfeilsticker, K., Rozanov, A., Lotz, W., Burrows, J. P., Lambert, J.-C., Goutail, F., Roscoe, H. K., and de Mazière, M.: A global stratospheric bromine monoxide climatology based on the BASCOE chemical transport model, Atmospheric Chemistry & Physics Discussions, 8, 17 581–17 629, 2008.

Uchiyama, A.: Line–by–Line Computation of the Atmospheric Absorption Spectrum Using the Decomposed Voigt Line Shape, J. Quant. Spectrosc. & Radiat. Transfer, 47, 521–532, 1992.

Van Roozendael, M., Loyola, D., Spurr, R., Balis, D., Lambert, J., Livschitz, Y., Valks, P., Ruppert, T., Kenter, P., Fayt, C., et al.: Ten years of GOME/ERS-2 total ozone data – The new GOME data processor (GDP) version 4: 1. Algorithm description, Journal of Geophysical Research-Atmospheres, 111, D14 311, 2006.

Vandaele, A. C., Simon, P. C., Guilmot, J. M., Carleer, M., and Colin, R.: SO_2 absorption cross section measurement in the UV using a Fourier transform spectrometer, J. Geophys. Res., 99, 25 599–25 606, doi:10.1029/94JD02187, 1994.

Vandaele, A. C., Hermans, C., Simon, P. C., Carleer, M., Colin, R., Fally, S., Merienne, M. F., Jenouvrier, A., and Coquart, B.: Measurements of the NO_2 absorption cross-section from 42 000 cm^{-1} to 10 000 cm^{-1} (238-1000 nm) at 220K and 294 K, J. Quant. Spectrosc. Ra., 59, 171–184, 1998.

Volkamer, R., Spietz, P., Burrows, J., and Platt, U.: High-resolution absorption cross-section of glyoxal in the UV–vis and IR spectral ranges, Journal of Photochemistry and Photobiology A: Chemistry, 172, 35–46, 2005.

von Savigny, C., Kokhanovsky, A., Bovensmann, H., Eichmann, K.-U., Kaiser, J. W., Noël, S., Rozanov, A. V., Skupin, J., and Burrows, J. P.: NLC Detection and Particle Size Determination: First Results from SCIAMACHY on ENVISAT, Adv. Space Res., 34, 851–856, 2004.

von Savigny, C., Ulasi, U. P., Eichmann, K.-U., Bovensmann, H., and Burrows, J. P.: Detection and mapping of polar stratospheric clouds using limb scattering observations, Atmos. Chem. Phys., 5, 3071–3079, 2005.

Vountas, M., Rozanov, V. V., and Burrows, J. P.: Ring effect: impact of rotational Raman scattering on radiative transfer in Earth's atmosphere., Journal of Quantitative Spectroscopy and Radiative Transfer, 60, 943–961, doi:10.1016/S0022-4073(97)00186-6, 1998.

Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendael, M.: FRESCO+: an improved O_2 A-band cloud retrieval algorithm for tropospheric trace gas retrievals, Atmospheric Chemistry & Physics, 8, 6565–6576, 2008.

Weideman, J.: Computation of the Complex Error Function, 31, 1497–1518, doi:10.1137/0731077, 1994.

Yamomoto, G. and Wark, D. Q.: Discussion of letter by A. Hanel: Determination of cloud altitude from a satellite, J. Geophys. Res., 66, 3596, 1961.