

Information operator approach and iterative regularization methods for atmospheric remote sensing

A. Doicu^{a,*}, S. Hilgers^a, A. von Bargaen^a, A. Rozanov^b, K.-U. Eichmann^b,
C. von Savigny^b, J.P. Burrows^b

^aGerman Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, Germany

^bInstitute of Environmental Physics, University of Bremen, Germany

Received 5 May 2006; accepted 6 May 2006

Abstract

In this study, we present the main features of the information operator approach for solving linear inverse problems arising in atmospheric remote sensing. This method is superior to the stochastic version of the Tikhonov regularization (or the optimal estimation method) due to its capability to filter out the noise-dominated components of the solution generated by an inappropriate choice of the regularization parameter. We extend this approach to iterative methods for nonlinear ill-posed problems and derive the truncated versions of the Gauss–Newton and Levenberg–Marquardt methods. Although the paper mostly focuses on discussing the mathematical details of the inverse method, retrieval results have been provided, which exemplify the performances of the methods. These results correspond to the NO₂ retrieval from SCIAMACHY limb scatter measurements and have been obtained by using the retrieval processors developed at the German Aerospace Center Oberpfaffenhofen and Institute of Environmental Physics of the University of Bremen.

© 2006 Elsevier Ltd. All rights reserved.

1. Introduction

Several methods have been employed to retrieve atmospheric parameters from the measurements of the atmospheric radiation. The most widely used approach is the optimal estimation method which seeks the statistically most likely solution given the measurement and the a priori information [1]. The prior information about the atmospheric state consists in an a priori profile and its covariance matrix and is encapsulated in the form of probability distributions, which are independent of the measurement. When such distributions are combined with probabilistic information about data uncertainties it is possible to derive a final (a posteriori) probability distribution assimilating both types of information.

The optimal estimation method can be regarded as a stochastic version of Tikhonov regularization, whereas the regularization matrix and the strength of regularization are entirely determined by the a priori covariance matrix. Because the strength of regularization (or the regularization parameter) is an a priori constraint commonly kept unchanged during the retrieval process, the Tikhonov solutions may be dominated by noise.

*Corresponding author.

E-mail address: adrian.doicu@dlr.de (A. Doicu).

The noise-dominated components of the solution can be filtered out by analyzing the information content of the measurement. This method is known as the information operator approach and has been proposed by Kozlov in a stochastic framework [2]. Applications of the information operator approach to the retrieval of atmospheric parameters from the scattered radiance measured in the nadir geometry and the transmitted radiance obtained from the solar occultation measurements have been reported by Hoogen et al. [3] and Rozanov [4], respectively.

In the present paper, we will interpret the information operator approach from a deterministic point of view, i.e., as an alternative approach to the stochastic Tikhonov regularization for solving linear inverse problems. This interpretation will enable us to extend the basic idea of the information operator approach to other deterministic iterative methods for nonlinear ill-posed problems.

The organization of our paper is as follows. In Section 2 we present the main features of the information operator approach for the linear case. The basic concepts are then extended to the nonlinear case in Section 3. Section 4 is dedicated to a comparison of retrieval results obtained by using two different processors. These retrieval processors, which use different iterative regularization methods based on the information operator approach, have been developed at the German Aerospace Center Oberpfaffenhofen and Institute of Environmental Physics of the University of Bremen. Conclusions are drawn in Section 5.

2. Linear case

We consider the linear data model

$$y = Kx, \quad (1)$$

where the mapping $K \in \mathbb{R}^{m \times n}$ represents the forward model, $y \in \mathbb{R}^m$ is the exact data vector and $x \in \mathbb{R}^n$ is the state vector containing the atmospheric profile to be retrieved. Measurements are made to a finite accuracy and in practice only the noisy data vector

$$y^\delta = y + \delta, \quad (2)$$

is available. The exact data is assumed to be attainable, i.e., there exists the true solution \hat{x} , such that $y = K\hat{x}$. In our analysis, we consider a semi-stochastic data model in the sense that the true solution \hat{x} is deterministic but the measurement error δ is stochastic with zero mean and the covariance matrix $S_\delta = \mathcal{E}\{\delta \cdot \delta^T\} = I_m$, where \mathcal{E} is the expected value operator and I_m is the identity matrix (of rank m). In general, if the measurement error is described by a symmetric and positive definite covariance matrix S_δ , a “normalized” data model with identity covariance matrix can be obtained by using the prewhitening technique [1].

The inverse ill-posed problem is solved in the least squares sense by means of Tikhonov regularization with a “stochastic” constraint [5,6]. This stochastic version of Tikhonov regularization is equivalent to the optimal estimation method; the only difference is the smoothing error which is a deterministic quantity. Essentially, an approximate solution is computed by minimizing the regularized function

$$\mathcal{F}^{\text{linear}}(x) = \frac{1}{2}[\|Kx - y^\delta\|^2 + (x - x_a)^T S_a^{-1}(x - x_a)], \quad (3)$$

where S_a is the a priori covariance matrix and x_a is the a priori state vector, the best beforehand estimator of \hat{x} . Assuming that the a priori covariance matrix is of the form $S_a = \sigma_a^2 \hat{S}_a$, where σ_a is the a priori standard deviation of the profile, we introduce the stochastic regularization matrix L (of rank n) by the Choleski factorization $\hat{S}_a^{-1} = L^T L$. Thus, $S_a^{-1} = \lambda_a^2 L^T L$, where $\lambda_a = 1/\sigma_a$ is the stochastic regularization parameter. Note that in the framework of the optimal estimation method, the regularized function (3) is interpreted as the logarithm of the a posteriori probability density function of the state when the measurement is given [1]. In practice, the regularized function is expressed in terms of the deviation with respect to the a priori state $\Delta x = x - x_a$, and the solution to the minimization problem is given by

$$\Delta x_{\lambda_a}^\delta = K_{\lambda_a}^\dagger r_a^\delta, \quad (4)$$

while the regularized solution is

$$x_{\lambda_a}^\delta = x_a + \Delta x_{\lambda_a}^\delta, \quad (5)$$

where $K_{\lambda_a}^\dagger = (K^T K + \lambda_a^2 L^T L)^{-1} K^T$ is the regularized (generalized) inverse or the gain matrix and $r_a^\delta = y^\delta - Kx_a$ is the residual vector at the a priori state. A numerical robust method for computing the regularized solution relies on the generalized singular value decomposition (GSVD) of the matrix pair (K, L) . If $K \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{n \times n}$, and $m \geq n$, the GSVD of the matrix pair (K, L) is given by $K = U \Sigma_1 Z^{-1}$ and $L = V \Sigma_2 Z^{-1}$, where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are unitary matrices, $Z \in \mathbb{R}^{n \times n}$ is a non-singular matrix, $\Sigma_1 \in \mathbb{R}^{m \times n}$ and $\Sigma_2 \in \mathbb{R}^{n \times n}$ are block-diagonal matrices,

$$\Sigma_1 = \begin{bmatrix} \text{diag}(\alpha_i)_{n \times n} \\ 0_{(m-n) \times n} \end{bmatrix}, \quad \Sigma_2 = [\text{diag}(\beta_i)_{n \times n}], \quad (6)$$

and $\gamma_i = \alpha_i / \beta_i$, $i = 1, 2, \dots, n$, are the generalized singular values appearing in decreasing order, i.e., $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_n$ [6]. We obtain $K_{\lambda_a}^\dagger = Z \Sigma U^T$, with

$$\Sigma = \begin{bmatrix} \text{diag} \left(\frac{f_{\lambda_a, i}}{\alpha_i} \right)_{n \times n} & 0_{n \times (m-n)} \end{bmatrix}, \quad (7)$$

and

$$\Delta x_{\lambda_a}^\delta = \sum_{i=1}^n f_{\lambda_a, i} \frac{u_i^T r_a^\delta}{\alpha_i} z_i, \quad (8)$$

where

$$f_{\lambda_a, i} = \frac{\gamma_i^2}{\gamma_i^2 + \lambda_a^2} \quad (9)$$

are the filter factors, and u_i and z_i denote the column vectors of $U = [u_1, u_2, \dots, u_m]$ and $Z = [z_1, z_2, \dots, z_n]$, respectively. The filter factors characterizing the damping or filtering of the GSVD components z_i are typically close to 1 for large γ_i and much smaller than 1 for the small γ_i . In this way, the contributions to the regularized solution corresponding to the smaller γ_i are effectively filtered out.

In the optimal estimation method, the information gained by making a measurement can be described by comparing the entropies \mathcal{S} of the a priori and a posteriori probability density functions $\mathcal{P}(x)$ and $\mathcal{P}(x|y^\delta)$, respectively [1]. The information content of a measurement H is the change in entropy consequent on making the measurement, $H = \mathcal{S}[\mathcal{P}(x)] - \mathcal{S}[\mathcal{P}(x|y^\delta)]$, and can be interpreted as the logarithm of a generalization of the signal-to-noise ratio. H can be expressed in terms of the so-called information operator $P = S_a K^T K$, as

$$H = \frac{1}{2} \ln[\det(P + I_n)]. \quad (10)$$

Replacing S_a by $(1/\lambda_a^2)(L^T L)^{-1}$ and using the GSVD of the matrix pair (K, L) we obtain $P = Z \Sigma_p Z^{-1}$, with $\Sigma_p = (1/\lambda_a^2)[\text{diag}(\gamma_i^2)_{n \times n}]$, and we see that the eigenvalues and eigenvectors of P are γ_i^2/λ_a^2 and x_i , respectively. The information content of a measurement then becomes

$$H = \frac{1}{2} \sum_{i=1}^n \ln \left(1 + \frac{\gamma_i^2}{\lambda_a^2} \right). \quad (11)$$

In the information operator approach, only eigenvalues which are greater than unity are considered to give a relevant contribution to the information content and the state space spanned by the eigenvectors associated with the relevant eigenvalues corresponds to the effective state space accessible with the measurement [2–4].

Returning to the stochastic Tikhonov regularization and using this assumption, we choose $\Delta x_{\lambda_a}^{\delta \perp} \in \text{Span}\{z_i\}_{i=1}^{N_{\text{cut}}}$, where N_{cut} is the maximum index i for which $\gamma_i \geq \lambda_a$, and set

$$x_{\lambda_a}^\delta = x_a + \Delta x_{\lambda_a}^{\delta \perp}. \quad (12)$$

The simplest choice of $\Delta x_{\lambda_a}^{\delta \perp}$ is

$$\Delta x_{\lambda_a}^{\delta \perp} = \sum_{i=1}^{N_{\text{cut}}} f_{\lambda_a, i} \frac{u_i^T r_a^\delta}{\alpha_i} z_i, \quad (13)$$

or equivalently,

$$\Delta x_{\lambda_a}^{\delta\perp} = K_{\lambda_a}^{\dagger\perp} r_a^\delta, \tag{14}$$

where $K_{\lambda_a}^{\dagger\perp}$ is the truncated regularized inverse, $K_{\lambda_a}^{\dagger\perp} = Z^\perp \Sigma U^T$, and

$$Z^\perp = [z_1, z_2, \dots, z_{N_{\text{cut}}}, 0, \dots, 0].$$

$\Delta x_{\lambda_a}^{\delta\perp}$ can also be chosen as the projection of $\Delta x_{\lambda_a}^\delta$ onto the linear span $\text{Span}\{z_i\}_{i=1}^{N_{\text{cut}}}$. In this case, the Gram–Schmidt procedure can be used to construct an orthonormal basis $\{z_i^\perp\}_{i=1}^{N_{\text{cut}}}$ [7], and $\Delta x_{\lambda_a}^{\delta\perp}$ is given by

$$\Delta x_{\lambda_a}^{\delta\perp} = \sum_{i=1}^{N_{\text{cut}}} (z_i^{\perp T} \Delta x_{\lambda_a}^\delta) z_i^\perp. \tag{15}$$

From a deterministic point of view, the information operator approach with $\Delta x_{\lambda_a}^{\delta\perp}$ given by Eq. (13) can be regarded as a truncated version of the stochastic Tikhonov regularization. The filter factors can be expressed as

$$f_{\lambda_a, i}^\perp = \begin{cases} f_{\lambda_a, i}, & i \leq N_{\text{cut}}, \\ 0, & i > N_{\text{cut}} \end{cases} \tag{16}$$

and we see that the information operator approach has sharper filter factors than the stochastic Tikhonov regularization. Thus, the information operator approach appears to be superior to the stochastic Tikhonov regularization if the regularization parameter λ_a (determined by statistical considerations) is too low so that noise-dominated components are included in the solution. In fact the information operator approach is a combination between the truncated GSVD with the filter factors $f_i^{\text{TGSVD}} = 1$, for $i \leq N_{\text{cut}}$ and $f_i^{\text{TGSVD}} = 0$, for $i > N_{\text{cut}}$, and the stochastic Tikhonov regularization with the filter factors $f_{\lambda_a, i}$, for $i = 1, 2, \dots, n$. If $\Delta x_{\lambda_a}^{\delta\perp}$ is computed accordingly to Eq. (15), the information operator approach can be regarded as a projected version of the stochastic Tikhonov regularization.

In the framework of Tikhonov regularization, the retrieval is characterized by the averaging kernel $A_{\lambda_a} = K_{\lambda_a}^\dagger K$, which gives informations about the smoothing error, and the square root of the diagonal of the noise covariance matrix $S(\epsilon_{\text{noise}}^\delta) = K_{\lambda_a}^\dagger K_{\lambda_a}^{\dagger T}$, which characterize the noise error. In terms of the GSVD of the matrix pair (K, L) , we have

$$A_{\lambda_a, ij} = \sum_{k=1}^n Z_{ik} f_{\lambda_a, k} (Z^{-1})_{kj} \tag{17}$$

and

$$S(\epsilon_{\text{noise}}^\delta)_{ii} = \sum_{k=1}^n \left(\frac{f_{\lambda_a, k}}{\alpha_k} \right)^2 Z_{ik}^2. \tag{18}$$

For the information operator approach, the expression of the error in \hat{x} can be derived by using Eqs. (2), (12), and (14). We obtain $\epsilon_{\text{total}}^\delta = \epsilon_{\text{smooth}}^\delta + \epsilon_{\text{noise}}^\delta$, where $\epsilon_{\text{total}}^\delta = x^\delta - \hat{x}$ is the total error, $\epsilon_{\text{smooth}}^\delta = (A_{\lambda_a}^\perp - I_n)(\hat{x} - x_a)$ is the smoothing error and $\epsilon_{\text{noise}}^\delta = K_{\lambda_a}^{\dagger\perp} \delta$ is the noise error. The entries of the truncated averaging kernel $A_{\lambda_a}^\perp = K_{\lambda_a}^{\dagger\perp} K$ are given by

$$A_{\lambda_a, ij}^\perp = \sum_{k=1}^{N_{\text{cut}}} Z_{ik} f_{\lambda_a, k} (Z^{-1})_{kj}, \tag{19}$$

and it is apparent that $A_{\lambda_a}^\perp$ deviates more from I_n than A_{λ_a} . The diagonal elements of the truncated noise covariance matrix $S^\perp(\epsilon_{\text{noise}}^\delta) = K_{\lambda_a}^{\dagger\perp} K_{\lambda_a}^{\dagger\perp T}$, are

$$S^\perp(\epsilon_{\text{noise}}^\delta)_{ii} = \sum_{k=1}^{N_{\text{cut}}} \left(\frac{f_{\lambda_a, k}}{\alpha_k} \right)^2 Z_{ik}^2, \tag{20}$$

and evidently, $S^\perp(\varepsilon_{\text{noise}}^\delta)_{ii} < S(\varepsilon_{\text{noise}}^\delta)_{ii}$. Therefore, the regularized solutions computed by using the information operator approach are characterized by larger smoothing errors and smaller noise errors than the solutions given by the stochastic Tikhonov regularization.

3. Nonlinear case

In the framework of the stochastic Tikhonov regularization, the nonlinear problem,

$$\begin{aligned} y &= F(x), \\ y^\delta &= y + \delta, \end{aligned}$$

with $F: \mathbb{R}^n \rightarrow \mathbb{R}^m$ being the forward model, is replaced by the minimization problem $\min_x \mathcal{F}(x)$, where

$$\mathcal{F}(x) = \frac{1}{2} [\|F(x) - y^\delta\|^2 + \lambda_a^2 \|L(x - x_a)\|^2]. \quad (21)$$

As in the linear case, the regularization matrix L and the regularization parameter λ_a are chosen by using the Choleski factorization of the a priori covariance matrix. The minimization problem can be solved iteratively by using the Gauss–Newton method for the augmented residual vector

$$f(x) = \begin{bmatrix} F(x) - y^\delta \\ \lambda_a L(x - x_a) \end{bmatrix},$$

i.e., $\mathcal{F}(x) = \frac{1}{2} \|f(x)\|^2$, while quasi-Newton approximation methods have to be employed for large-residual problems [8]. Essentially, at each iteration step k , the Gauss–Newton method minimizes the regularized function

$$\mathcal{F}_k^{\text{linear}}(x) = \|F(x_{\lambda_a, k}^\delta) - y^\delta + K(x_{\lambda_a, k}^\delta)(x - x_{\lambda_a, k}^\delta)\|^2 + \lambda_a^2 \|L(x - x_a)\|^2, \quad (22)$$

or equivalently,

$$\mathcal{F}_k^{\text{linear}}(\Delta x) = \|K_k \Delta x - w_{\lambda_a, k}^\delta\|^2 + \lambda_a^2 \|L \Delta x\|^2, \quad (23)$$

where $K(x) = F'(x)$ is now the Jacobian matrix of F evaluated at x , $K_k = K(x_{\lambda_a, k}^\delta)$, $\Delta x = x - x_a$ and $w_{\lambda_a, k}^\delta = K_k(x_{\lambda_a, k}^\delta - x_a) + r_{\lambda_a, k}^\delta$. $r_{\lambda_a, k}^\delta$ is the residual vector at the iteration step k and is given by $r_{\lambda_a, k}^\delta = y^\delta - F(x_{\lambda_a, k}^\delta)$. The minimizer of $\mathcal{F}_k^{\text{linear}}$ is given by

$$\Delta x_{\lambda_a, k}^\delta = K_{\lambda_a, k}^\dagger w_{\lambda_a, k}^\delta, \quad (24)$$

while the new iterate is

$$x_{\lambda_a, k+1}^\delta = x_a + \Delta x_{\lambda_a, k}^\delta. \quad (25)$$

In the framework of the information operator approach, the natural extension of the results presented in Section 2 leads to the iterative method

$$\Delta x_{\lambda_a, k}^{\delta \perp} = K_{\lambda_a, k}^{\dagger \perp} w_{\lambda_a, k}^\delta \quad (26)$$

and

$$x_{\lambda_a, k+1}^\delta = x_a + \Delta x_{\lambda_a, k}^{\delta \perp}, \quad (27)$$

where $K_{\lambda_a, k}^{\dagger \perp}$ is now the truncated generalized inverse at $x_{\lambda_a, k}^\delta$. By convention, the iterative regularization method (26) and (27) will be referred to as the truncated regularized Gauss–Newton method.

An alternative Newton type method for the solution of nonlinear inverse problems is the Levenberg–Marquardt method [9,10]. The key idea of the Levenberg–Marquardt method consists in repeatedly linearizing the operator equation $F(x) = y$ around the actual iterate $x_{\lambda_a, k}^\delta$,

$$F(x) \approx F(x_{\lambda_a, k}^\delta) + K(x_{\lambda_a, k}^\delta)(x - x_{\lambda_a, k}^\delta), \quad (28)$$

and then minimizing the Tikhonov function,

$$\mathcal{F}_k^{\text{linear}}(x) = \|F(x_{\lambda_a,k}^\delta) - y^\delta + K(x_{\lambda_a,k}^\delta)(x - x_{\lambda_a,k}^\delta)\|^2 + \lambda_a^2 \|L(x - x_{\lambda_a,k}^\delta)\|^2. \quad (29)$$

$\mathcal{F}_k^{\text{linear}}$ can be written in standard form as

$$\mathcal{F}_k^{\text{linear}}(\Delta x) = \|K_k \Delta x - r_{\lambda_a,k}^\delta\|^2 + \lambda_a^2 \|L \Delta x\|^2, \quad (30)$$

where $\Delta x = x - x_{\lambda_a,k}^\delta$, and it is readily seen that the minimizer of $\mathcal{F}_k^{\text{linear}}$ is

$$\Delta x_{\lambda_a,k}^\delta = K_{\lambda_a,k}^\dagger r_{\lambda_a,k}^\delta. \quad (31)$$

Note that the only difference between (22) and (29) is in the penalty term. The reuse of x_a in the penalty term—besides its use as initial guess—may have a stabilizing effect on the minimizing elements, i.e., the iterates will never diverge to infinity in norm.

We introduce the truncated Levenberg–Marquardt method as a regularization method which computes $\Delta x_{\lambda_a}^{\delta\perp} \in \text{Span}\{z_i\}_{i=1}^{N_{\text{cut}}}$ as

$$\Delta x_{\lambda_a,k}^{\delta\perp} = K_{\lambda_a,k}^{\dagger\perp} r_{\lambda_a,k}^\delta, \quad (32)$$

and the new iterate as

$$x_{\lambda_a,k+1}^\delta = x_{\lambda_a,k}^\delta + \Delta x_{\lambda_a,k}^{\delta\perp}. \quad (33)$$

In the conventional regularizing Levenberg–Marquardt method, the regularization parameter is selected at each iteration from a trust region approach for the errors (in the data space or the state space). In the truncated version, the regularization parameter is constant but the cutting index N_{cut} is variable and is determined by the eigenvalues of the Jacobian matrix at each iteration step.

4. Numerical simulations

Several new satellite instruments such as SCIAMACHY [11] and GOMOS [12] aboard the Envisat satellite, OSIRIS [13] onboard the Odin satellite, as well as SAGE III [14] onboard a Meteor-3M spacecraft measure in limb viewing geometry ultraviolet and visible solar radiation scattered by the Earth's atmosphere to monitor vertical profiles of O₃ and other minor constituents such as NO₂, BrO, and OClO. A sequence of observations (limb scan) corresponding to different tangent altitudes is used for the retrieval of atmospheric parameters.

The goal of our numerical simulations is to demonstrate the advantages of the truncation technique described in the previous section as applied to the retrieval of NO₂ vertical distributions from SCIAMACHY limb measurements. For this purpose, retrieval results obtained by two different processors are intercompared. The first processor developed at the German Aerospace Center (DLR) Oberpfaffenhofen uses the truncated regularized Gauss–Newton method, while the second processor developed at the Institute of Environmental Physics (IUP) of the University of Bremen employs the truncated Levenberg–Marquardt method. The processors of the different groups use their own radiative transfer models to retrieve geophysical parameters from SCIAMACHY data. The radiative transfer model used by the DLR processor is a single-scattering model, while the multiple-scattering effect is taken into account by using look-up table corrections [15]. The IUP processor uses a multiple-scattering model based on the combined differential–integral approach with a Picard iterative scheme [16]. For both processors, the data model is expressed in terms of normalized limb radiances and relies on the differential absorption technique [17], i.e.,

$$\begin{aligned} y^\delta &= \ln(I_{\text{meas}}^n) - P_{\text{meas}}, \\ F &= \ln(I_{\text{sim}}^n) - P_{\text{sim}}. \end{aligned} \quad (34)$$

The normalized vector of limb radiances is given by $I^n = I/I^{\text{ref}}$, where $I^{\text{ref}} = I^{\text{ref}}(\lambda)$ is the reference radiance spectrum for all wavelengths of interest λ and $I = I(s, \lambda)$ is the vector of radiance spectra for all tangent heights s in the selected tangent height range excepting the reference tangent height s_{ref} . As a reference scan a limb measurement at an upper tangent height is commonly used. P is a lower-order polynomial in λ , whose coefficients are obtained by fitting the function $\ln(I^n)$ in the wavelength domain. Note that the spectra

normalization eliminates most of the solar Fraunhofer structure, the instrument response function and the need for an absolute instrument calibration. By subtracting lower-order polynomials from radiance spectra we remove those spectral features that are smoothly varying functions of wavelength, particularly those due to Rayleigh and Mie scattering. The differential absorption type model is very sensitive to weak absorptions and has a moderate degree of nonlinearity (the number of iterations required to achieve convergence does not exceed 4).

The atmospheric parameters are retrieved from measurements in the visible spectral range. The spectral domain of analysis is a window of 420–450 nm in Channel 3 of the SCIAMACHY instrument. In this spectral interval, the main retrieval gases are NO_2 and O_3 . For each gas, the a priori covariance matrix is

$$[S_a]_{ij} = \sigma_a^2 x_{0i} x_{0j} \exp\left(-\frac{|h_i - h_j|}{l_{\text{cor}}}\right), \quad i, j = 1, \dots, n, \quad (35)$$

where x_0 is the initial guess and l_{cor} is a length determining the correlation between the parameters at different altitudes h_i . If not stated otherwise, $\sigma_{a,\text{NO}_2} = 1.0$ and $\sigma_{a,\text{O}_3} = 0.2$, while $l_{\text{cor}} = 3.3$ km for both gases. The a priori and initial gas profiles were assumed to be identical and were chosen from the model of McLinden [18]. The retrieval grid is a rough grid with a step of 3.5 km between 10.5 and 42 km. This choice leads to a small number of unknowns and a low computer time as required by the operational usage of a retrieval processor. The number of limb spectra is 10, and the tangent heights vary from 13 to 43 km. The reference spectrum chosen is the spectrum with 43 km tangent altitude.

In Figs. 1 and 2 we illustrate the retrieved profiles and the relative errors computed with the conventional and the truncated versions of the regularized Gauss–Newton method. The results are calculated with the DLR processor and correspond to SCIAMACHY measurements recorded on January 15, 2004 at latitude 26° south and longitude 164° west. For this example we use a weak regularization by choosing a high value for the standard deviation of the NO_2 profile, i.e., $\sigma_{a,\text{NO}_2} = 5.0$. The curves show that the conventional solution is undersmoothed and the truncated regularized Gauss–Newton method filters out the profile oscillations. Consequently, the conventional solution has smaller smoothing errors and larger noise errors than the truncated solution. The smoothing error is the dominant error source and this component of the total error increases at low and high altitudes.

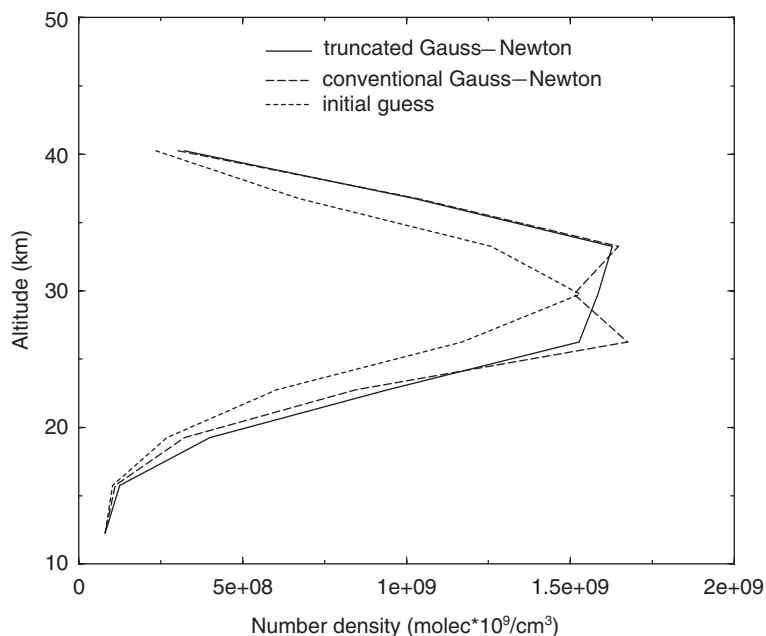


Fig. 1. Retrieved profiles computed with the conventional and truncated versions of the regularized Gauss–Newton method for a weak regularization.

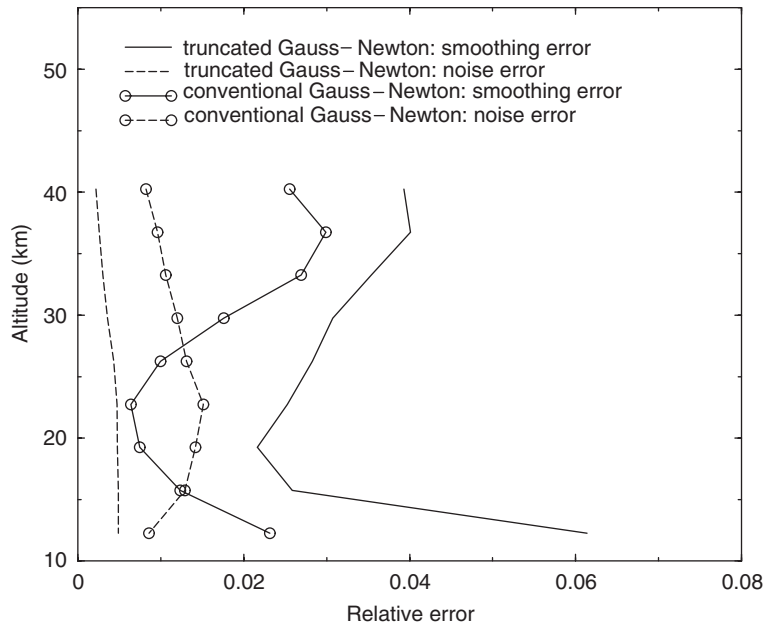


Fig. 2. Smoothing and noise errors of the profiles retrieved with the conventional and truncated versions of the regularized Gauss–Newton method.

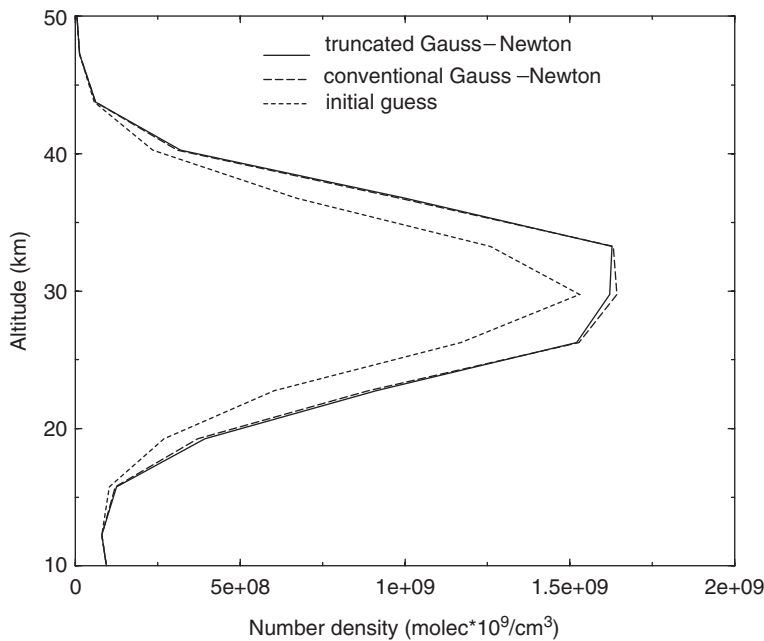


Fig. 3. Retrieved profiles computed with the conventional and truncated versions of the regularized Gauss–Newton method for a strong regularization.

The profiles plotted in Fig. 3 correspond to a stronger regularization, i.e., $\sigma_{a,NO_2} = 1.0$. In this case, both methods give similar results and the additional regularization effect of the truncated regularized Gauss–Newton method does not play a significant role.

In our next example, spectral data from four orbits: 9816 (15 January 2004), 10740 (20 March 2004), 12105 (23 June 2004) and 13379 (20 September 2004) have been analyzed. Exemplary retrieved profiles are shown in

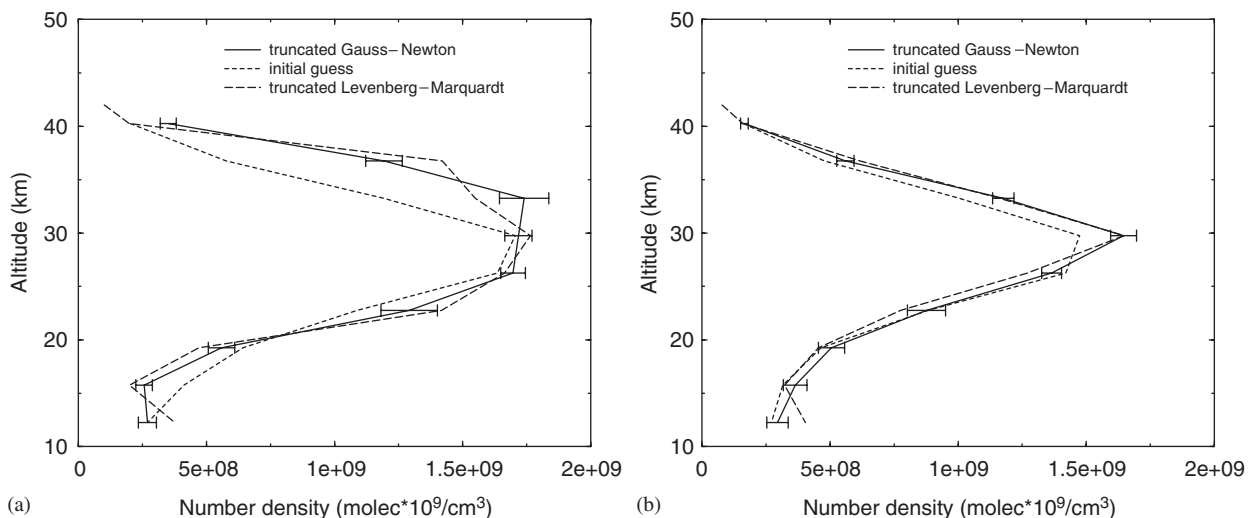


Fig. 4. Exemplary profiles computed with the truncated versions of the Gauss–Newton and Levenberg–Marquardt methods: (a) orbit 9816 and (b) orbit 10740.

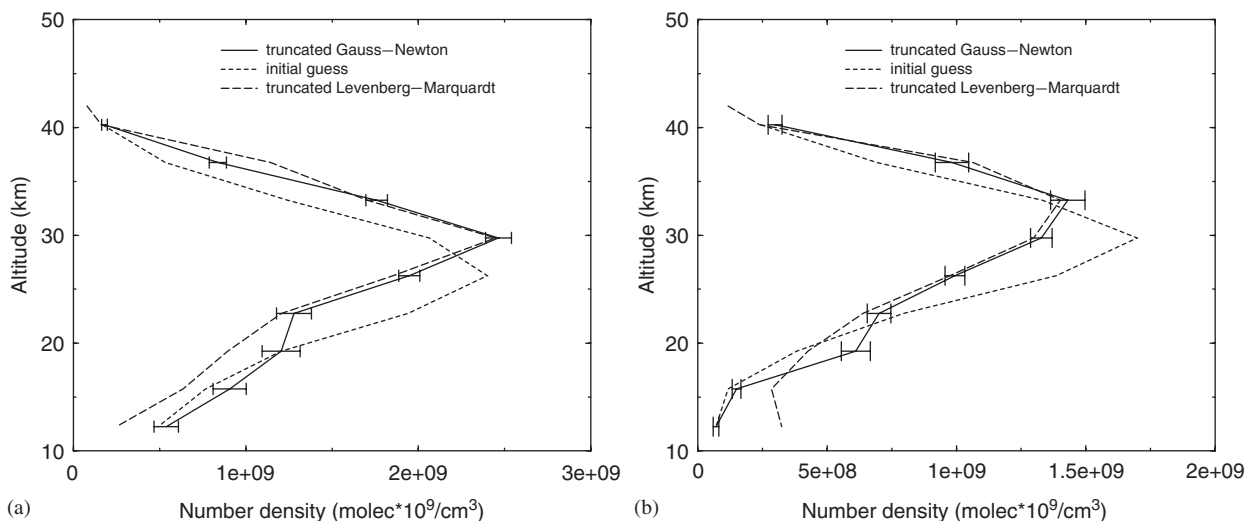


Fig. 5. Exemplary profiles computed with the truncated versions of the Gauss–Newton and Levenberg–Marquardt methods: (a) orbit 12105 and (b) orbit 13379.

Figs. 4 and 5. Even though the truncated regularized Gauss–Newton method contains an a priori dependent regularization term, the retrievals are largely free of the a priori contamination. The truncated Levenberg–Marquardt solutions show small oscillations as compared to the truncated Gauss–Newton solutions. Note that in general, the Levenberg–Marquardt method may lead to undersmoothed solutions as compared to the regularized Gauss–Newton method because the contribution of the regularization term decreases during the iterative process (as $x_{\lambda_a, k}^{\delta}$ converges).

In Fig. 6 we plot the relative differences between the solutions. The differences are averaged over all states of the SCIAMACHY orbits and on average one curve corresponds to 25 profiles. The relative differences are smaller than 10% in the maximum of the NO_2 profile and less than 20% below 20 km. The reason for this discrepancy at low altitudes is that the sensitivity to NO_2 is low and due to the strong regularization, the truncated Gauss–Newton solution is influenced by the a priori information.

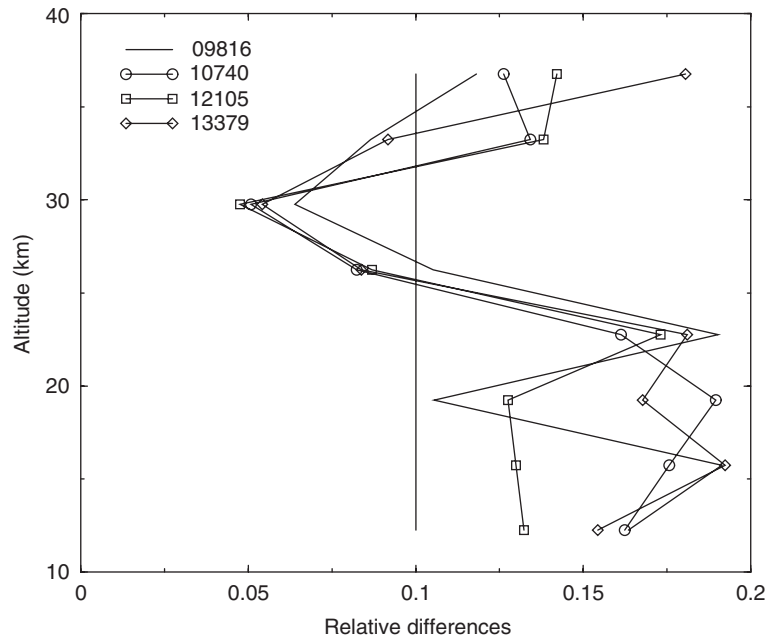


Fig. 6. Average relative differences between the solutions computed with the truncated versions of the Gauss–Newton and Levenberg–Marquardt methods.

5. Conclusion

The main features of the information operator approach for solving linear inverse problems have been presented. Relying on the GSVD of the Jacobian and the regularization matrix we showed that the information operator approach is a truncated version of the stochastic Tikhonov regularization with sharper filter factors. Due to a stronger regularization, the information operator solutions are characterized by larger smoothing errors and smaller noise errors than the Tikhonov solutions. The main feature of this method is the filtering of the noise-dominated components of the Tikhonov solutions generated by an inappropriate choice of the regularization parameter.

The idea of projecting the regularized solution onto the effective state space accessible with the measurement has been extended to the nonlinear case by deriving the truncated versions of the Gauss–Newton and Levenberg–Marquardt methods. The numerical simulations concerning the intercomparison of two different processors demonstrated the applicability of this truncation technique to SCIAMACHY limb measurements. The number densities of NO₂ retrieved with the different methods agree to within 10% at the peak of the profiles.

In our analysis, the regularization parameter together with the regularization matrix are specified by the a priori covariance matrix. In the deterministic version of the Tikhonov regularization, the optimal value of the regularization parameter is computed by using appropriate parameter-choice methods as for instance the L-curve method [6], the generalized cross-validation approach [19] or the minimum bound method [20]. Essentially, the optimal value of the regularization parameter is a trade off between the data and a priori constraint. These parameter choice methods can also be formulated for the information operator approach if we regard this method as a truncated Tikhonov regularization with the filter factors $f_{\lambda,i}^{\perp} = f_{\lambda,i}$ for $i \leq N_{\text{cut}}$, and $f_{\lambda,i}^{\perp} = 0$ for $i > N_{\text{cut}}$, where N_{cut} is the maximum index i for which $\gamma_i \geq \lambda$. Formulations of parameter choice methods for the information operator approach will be the topic of a future paper.

References

- [1] Rodgers CD. Inverse methods for atmospheric sounding: theory and practise. Singapore: World Scientific; 2000.

- [2] Kozlov VP. Design of experiments related to the inverse problem of mathematical physics. In: Ermakov CM, editor. *Mathematical theory of experiment design*. Moscow: Nauka; 1983. p. 216–46.
- [3] Hoogen R, Rozanov VV, Burrows JP. Ozone profiles from GOME satellite data: algorithm description and first validation. *J Geophys Res* 1999;104:8263–80.
- [4] Rozanov A. *Modeling of radiative transfer through a spherical planetary atmosphere: application to atmospheric trace gases retrieval from occultation- and limb-measurements in UV-Vis-NIR*. Germany: Logos Verlag Berlin; 2001.
- [5] Engl HW, Hanke M, Neubauer A. *Regularization of inverse problems*. Dordrecht: Kluwer; 1996.
- [6] Hansen PC. *Rank deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Philadelphia, PA: SIAM; 1998.
- [7] Noble B, Daniel JW. *Applied linear algebra*. Englewood Cliffs, NJ: Prentice-Hall; 1988.
- [8] Tsidu GM. On the accuracy of covariance matrix: Hessian versus Gauss–Newton methods in atmospheric remote sensing with infrared spectroscopy. *J Quant Spectrosc Radiat Transfer* 2005;96:103–21.
- [9] Hanke M. A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Probl* 1997;13:79–95.
- [10] Dennis JE, Schnabel RB. *Numerical methods for unconstrained optimization and nonlinear equations*. Englewood Cliffs, NJ: Prentice-Hall; 1983.
- [11] Bovensmann H, Burrows JP, Buchwitz M, Frerick J, Noel S, Rozanov VV, et al. *SCIAMACHY—Mission objectives and measurements modes*. *J Atmos Sci* 1999;56:127–50.
- [12] Bertaux J, Megie G, Widemann T, Chassefiere E, Pellinen R, Kyrola E, et al. Monitoring of ozone trend by stellar occultations: the GOMOS instrument. *Adv Space Res* 1991;11(3):237–42.
- [13] Llewellyn EJ, Lloyd ND, Degenstein DA, et al. The OSIRIS instrument on the Odin spacecraft. *Can J Phys* 2004;82(6):411–22.
- [14] McCormick MP, Chu WP, Zawodny JM, Mauldin LE, McMaster LR. Stratospheric Aerosol and Gas Experiment III (SAGE III) aerosol and trace gas measurements for Earth Observing System (EOS). *Proceedings of the SPIE International Society of Optical Engineering*, vol. 1491; 1991. p. 125–41.
- [15] Flittner DE, Bhartia PK, Herman BM. O₃ profiles retrieved from limb scatter measurements: theory. *Geophys Res Lett* 2000;27:2601–4.
- [16] Rozanov A, Rozanov V, Burrows JP. A numerical radiative transfer model for a spherical planetary atmosphere: combined differential–integral approach involving the Picard iterative approximation. *J Quant Spectrosc Radiat Transfer* 2001;69:491–512.
- [17] Platt U. *Differential optical absorption spectroscopy (DOAS)*. In: Sigrist M, editor. *Air monitoring by spectroscopic techniques*. Hoboken, NJ: Wiley; 1994. p. 27–84.
- [18] McLinden CA. Routine for generating model atmospheres. Personal Communication; 2000.
- [19] Wahba G. *Spline models for observational data*. Philadelphia, PA: SIAM; 1990.
- [20] Raus T. On the discrepancy principle for the solution of ill-posed problems. *Acta Commentat Univ Tartuensis Math* 1984;672:16–26.