

An efficient inversion algorithm for atmospheric remote sensing with application to UV limb observations

Adrian Doicu*, Franz Schreier, Siegfried Hilgers, Albrecht von Bargaen,
Sander Slijkhuis, Michael Hess, Bernd Aberle

DLR, German Aerospace Center, Remote Sensing Technology Institute, Oberpfaffenhofen, 82234 Weßling, Germany

Received 4 May 2006; accepted 6 May 2006

Abstract

In this paper we present a retrieval algorithm for atmospheric remote sensing. The algorithm combines Tikhonov regularization and the iteratively regularized Gauss–Newton method and is devoted to the solution of multi-parameter inverse problems with simple bounds on the variables. The basic features of the algorithm: the solution of the bound-constrained minimization problem, the selection of the optimal regularization parameter, the derivation of the global regularization matrix and the characterization of the solution (error analysis) are discussed in detailed. The inversion algorithm is applied to ozone retrieval from SCIAMACHY limb scatter measurements in the ultraviolet spectral range.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Inverse problems; Nonlinear least squares; Regularization; Atmospheric spectroscopy; Remote sensing; UV limb observations

1. Introduction

Several satellite instruments measure radiation scattered or emitted from the Earth's limb to monitor vertical profiles of atmospheric gases or temperature. A sequence of observations (limb scans) corresponding to different tangent altitudes are used for atmospheric retrieval. Two recently launched instruments employing this technique are the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) and the SCanning Imaging Absorption SpectroMeter for Atmospheric ChartographY (SCIAMACHY) aboard the ENVISAT satellite. The MIPAS instrument [1] observes thermal infrared (IR) emission and provides information about temperature, ozone (O₃), nitrogen family (NO₂, HNO₃, N₂O, etc.) and dynamic tracers (H₂O, CH₄). The SCIAMACHY instrument [2] measures scattered solar radiation in the ultraviolet (UV) to near IR range and reveals information about the Earth's atmospheric composition with respect to O₃ and other minor constituents such as NO₂ and BrO.

Inverse problems arising in atmospheric remote sensing are usually nonlinear and ill posed. Moreover, we are frequently dealing with multi-parameter problems, i.e., several atmospheric profiles are retrieved together

*Corresponding author. Tel.: +49 8153 283015; fax: +49 8153 281446.
E-mail address: adrian.doicu@dlr.de (A. Doicu).

with a set of auxiliary parameters. Quadratic and inequality constraints have to be imposed on the state vector in order to obtain realistic solutions.

The benefits of using a bound-constrained algorithm stem from two observations about most practical problems. Firstly, the restriction on the expected size of the variables is frequently encountered in atmospheric remote sensing. Secondly, even if no bounds are active at the solution, their presence can prevent the objective function from being evaluated at nonphysical points during the iterative process. This kind of inverse problems can be solved by using a multi-parameter regularization method with simple bounds on the variables.

In this paper we present the main features of an inversion algorithm for atmospheric remote sensing [3–5]. The algorithm incorporates two regularization methods: Tikhonov regularization [6] and the iteratively regularized Gauss–Newton method [7] and has been employed for MIPAS [8] and SCIAMACHY limb radiance measurements. The data model is semi-stochastic in contrast to the optimal estimation method (or the Bayesian approach) which assumes a stochastic data model [9].

The organization of our paper is as follows. Section 2 describes the basics of the algorithm and gives all key formulas. The application of the technique to SCIAMACHY limb measurements is discussed in Section 3. Section 4 summarizes our results.

2. Inversion algorithm

The main steps of the retrieval algorithm are: the derivation of the discrete data model, the formulation and the solution of the bound-constrained minimization problem, the selection of the regularization parameter, the derivation of the global regularization matrix for multi-parameter problems and the error analysis depending on the setting in which the problem is treated.

2.1. Discrete data model

The discretization of the radiative transfer equation leads to the data model:

$$\begin{aligned} \mathbf{y} &= \mathbf{F}(\mathbf{x}), \\ \mathbf{y}^\delta &= \mathbf{y} + \boldsymbol{\delta}, \end{aligned} \quad (1)$$

where the mapping $\mathbf{F}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ represents the forward model, $\mathbf{y} \in \mathbb{R}^m$ is the exact data vector, $\mathbf{x} \in \mathbb{R}^n$ is the state vector to be estimated, $\mathbf{y}^\delta \in \mathbb{R}^m$ is the noisy data vector and $\boldsymbol{\delta} \in \mathbb{R}^m$ is the measurement error.

The inverse problem $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{y})$ is solved in the least-squares sense by imposing simple bounds on the variables to obtain a physically meaningful solution, i.e.,

$$\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}. \quad (2)$$

For a multi-parameter problem, the state vector \mathbf{x} has several components \mathbf{z}_i (with not necessarily equal lengths),

$$\mathbf{x} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_P]^T, \quad (3)$$

where \mathbf{z}_i stands for a discretized atmospheric profile or the set of auxiliary parameters, and P is the number of components, cf. Section 2.4.

In our analysis we consider a semi-stochastic data model in the sense that the exact solution $\hat{\mathbf{x}}$ is deterministic but the measurement error $\boldsymbol{\delta}$ is stochastic with zero mean and the covariance matrix $\mathbf{S}_\delta = \mathcal{E}\{\boldsymbol{\delta} \cdot \boldsymbol{\delta}^T\} = (1/m)\mathbf{l}_m$, where \mathcal{E} is the expected value operator and \mathbf{l}_m is the identity matrix (of rank m). In general, if the measurement error is described by a symmetric and positive definite covariance matrix \mathbf{S}_δ , we can obtain a “normalized” data model with identity covariance matrix by using the prewhitening technique [9].

2.2. Bound-constrained minimization problem

The ill-posed problem is solved in the least-squares sense by means of regularization methods. The corresponding bound-constrained minimization problem is of the form:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathcal{F}(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \{ \|\mathbf{F}(\mathbf{x}) - \mathbf{y}^\delta\|^2 + \lambda^2 A(\mathbf{x}) \}$$

subject to the simple bounds $\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}$,

(4)

where λ is the regularization parameter, A is the global regularization term and the symbol $\|\cdot\|$ denotes the Euclidian norm. The regularization term adds a priori knowledge about the solution to the information coming from the measurement. In general, A is a quadratic term,

$$A(\mathbf{x}) = \|\mathbf{H}(\mathbf{x} - \mathbf{x}_a)\|^2,$$
(5)

where \mathbf{x}_a is the a priori state vector, the best beforehand estimator of $\hat{\mathbf{x}}$, and \mathbf{H} is the global regularization matrix. The derivation of the global regularization matrix for multi-parameter problems is discussed in Section 2.4.

The bound-constrained minimization problem can be formulated as a standard least-squares problem $\mathcal{F}(\mathbf{x}) = \frac{1}{2} \|\mathbf{f}(\mathbf{x})\|^2$, where \mathbf{f} is the generalized residual vector

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} \mathbf{F}(\mathbf{x}) - \mathbf{y}^\delta \\ \lambda \mathbf{H}(\mathbf{x} - \mathbf{x}_a) \end{bmatrix}.$$
(6)

Iterative methods for nonlinear least squares can then be employed to solve the minimization problem (cf., e.g., [10]). A Gauss–Newton method with a simplified version of an active-set algorithm [11] and a trust-region method with a local active-set strategy to select the step [12,13] are implemented in our code.

In the Gauss–Newton algorithm, the variables remain free during the iterative process, while in the trust-region algorithm, the variables can be fixed on their bounds. Note that the second algorithm is more sophisticated but more efficient than the first algorithm.

2.2.1. Gauss–Newton method

For this algorithm, the process of finding the new iterate can be summarized as follows:

(1) compute the step direction \mathbf{p}_k by solving the unconstrained problem

$$\min_{\mathbf{p} \in \mathbb{R}^n} \left\{ \mathbf{f}^T(\mathbf{x}_{k\lambda}^\delta) \mathbf{K}_f(\mathbf{x}_{k\lambda}^\delta) \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{K}_f(\mathbf{x}_{k\lambda}^\delta)^T \mathbf{K}_f(\mathbf{x}_{k\lambda}^\delta) \mathbf{p} \right\},$$
(7)

where k is the iteration step, $\mathbf{K}_f(\mathbf{x}) = \mathbf{f}'(\mathbf{x})$ denotes the Jacobian matrix of \mathbf{f} evaluated at \mathbf{x} and $\mathbf{x}_{k\lambda}^\delta$ is the actual iterate,

(2) compute \bar{a} , the maximum positive feasible step along \mathbf{p}_k , i.e., $\mathbf{l} \leq \mathbf{x}_{k\lambda}^\delta + a\mathbf{p}_k \leq \mathbf{u}$ for all a with $0 < a \leq \bar{a}$,

(3) determine a positive step-length a_k for which it holds that

$$\mathcal{F}(\mathbf{x}_{k\lambda}^\delta + a_k \mathbf{p}_k) < \mathcal{F}(\mathbf{x}_{k\lambda}^\delta)$$

and $a_k \leq \zeta \bar{a}$, where $\zeta < 1$, and take $\mathbf{x}_{k+1,\lambda}^\delta = \mathbf{x}_{k\lambda}^\delta + a_k \mathbf{p}_k$.

2.2.2. Trust-region method

For the trust-region method, we take $\mathbf{x}_{k+1,\lambda}^\delta = \mathbf{x}_{k\lambda}^\delta + \mathbf{p}_k$, where \mathbf{p}_k is the solution of the trust-region problem

$$\min_{\mathbf{p} \in \mathbb{R}^n} \left\{ \mathbf{f}^T(\mathbf{x}_{k\lambda}^\delta) \mathbf{K}_f(\mathbf{x}_{k\lambda}^\delta) \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{G}(\mathbf{x}_{k\lambda}^\delta) \mathbf{p} \right\}$$

subject to the simple bounds $\mathbf{l} \leq \mathbf{x} \leq \mathbf{u}$

and the trust-region constraint $\|\mathbf{p}\| \leq \rho_k$.

(8)

Here, ρ_k is the radius of the trust region and $\mathbf{G}(\mathbf{x}_{k\lambda}^\delta)$ is some approximation to the Hessian of \mathcal{F} computed at $\mathbf{x}_{k\lambda}^\delta$. \mathbf{G} can be computed in the framework of a Gauss–Newton model as $\mathbf{G} = \mathbf{K}_f^T \mathbf{K}_f$, or in the framework of a quasi-Newton model as $\mathbf{G} = \mathbf{K}_f^T \mathbf{K}_f + \mathbf{S}$, where \mathbf{S} is a secant approximation to the second-order part of the least-squares Hessian [13]. The algorithm starts with $\mathbf{S}_0 = 0$. With this choice, the first iteration is equivalent to an iteration of the Gauss–Newton method. If the Gauss–Newton step is too long the trust region is shrunk repeatedly to obtain a feasible $\mathbf{x}_{k+1,\lambda}^\delta$.

2.3. Regularization parameter selection

The Gauss–Newton step is equivalent to the minimization of the regularized function

$$\mathcal{F}_k^{\text{linear}}(\mathbf{x}) = \|\mathbf{F}(\mathbf{x}_{k\lambda}^\delta) - \mathbf{y}^\delta + \mathbf{K}(\mathbf{x}_{k\lambda}^\delta)(\mathbf{x} - \mathbf{x}_{k\lambda}^\delta)\|^2 + \lambda^2 \|\mathbf{H}(\mathbf{x} - \mathbf{x}_a)\|^2, \quad (9)$$

for the linear subproblem

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{x}_{k\lambda}^\delta) + \mathbf{K}(\mathbf{x}_{k\lambda}^\delta)(\mathbf{x} - \mathbf{x}_{k\lambda}^\delta) + \mathbf{R}(\mathbf{x}_{k\lambda}^\delta, \mathbf{x}), \quad (10)$$

where $\mathbf{K}(\mathbf{x}) = \mathbf{F}'(\mathbf{x})$ is the Jacobian matrix of \mathbf{F} evaluated at \mathbf{x} and $\mathbf{R}(\mathbf{x}_{k\lambda}^\delta, \mathbf{x})$ is the linearization error. The regularized function for the corresponding linear subproblem can be written in standard form as

$$\mathcal{F}_k^{\text{linear}}(\mathbf{u}) = \|\mathbf{K}_k \mathbf{u} - \mathbf{w}_k^\delta\|^2 + \lambda^2 \|\mathbf{H} \mathbf{u}\|^2, \quad (11)$$

where $\mathbf{u} = \mathbf{x} - \mathbf{x}_a$, $\mathbf{K}_k = \mathbf{K}(\mathbf{x}_{k\lambda}^\delta)$ and $\mathbf{w}_k^\delta = \mathbf{K}_k(\mathbf{x}_{k\lambda}^\delta - \mathbf{x}_a) - \mathbf{r}_{k\lambda}^\delta$. The residual vector at the iteration step k is given by $\mathbf{r}_{k\lambda}^\delta = \mathbf{F}(\mathbf{x}_{k\lambda}^\delta) - \mathbf{y}^\delta$.

Because at each iteration we are dealing with a linear problem, we summarize the most important parameter-choice methods for linear problems. These parameter-choice methods will be extended to the nonlinear case in Section 2.3.6. We begin by recalling the main quantities of interest. The Tikhonov solution to the minimization problem is $\mathbf{u}_{k\lambda}^\delta = \mathbf{K}_{k\lambda}^\dagger \mathbf{w}_k^\delta$, where $\mathbf{K}_{k\lambda}^\dagger$ is the regularized (generalized) inverse or the gain matrix and $\mathbf{x}_{k+1,\lambda}^\delta = \mathbf{u}_{k\lambda}^\delta + \mathbf{x}_a$. The linearized residual is given by $\mathbf{r}_{k\lambda}^{\text{linear},\delta} = \mathbf{K}_k \mathbf{u}_{k\lambda}^\delta - \mathbf{w}_k^\delta$, while the linearized constraint vector is $\mathbf{s}_{k\lambda}^{\text{linear},\delta} = \mathbf{H} \mathbf{u}_{k\lambda}^\delta$. The linearized predictive error is defined as $\mathbf{p}_{k\lambda}^{\text{linear},\delta} = \mathbf{K}_k \mathbf{u}_{k\lambda}^\delta - \mathbf{w}_k$, where $\mathbf{w}_k = \mathbf{K}_k(\mathbf{x}_{k\lambda}^\delta - \mathbf{x}_a) - \mathbf{r}_{k\lambda}$ and $\mathbf{r}_{k\lambda}$ is the residual vector at the iteration step k in the noise-free case given by $\mathbf{r}_{k\lambda} = \mathbf{F}(\mathbf{x}_{k\lambda}^\delta) - \mathbf{y}$. The linearized total error gives the discrepancy between the regularized solution and the exact solution and is defined as $\mathbf{e}_{k\lambda,\text{total}}^{\text{linear},\delta} = \mathbf{u}_{k\lambda}^\delta - \hat{\mathbf{u}}$, where $\hat{\mathbf{u}} = \hat{\mathbf{x}} - \mathbf{x}_a$.

Assuming that the linearity relation (10) holds at $\mathbf{x} = \hat{\mathbf{x}}$, we have

$$\mathbf{p}_{k\lambda}^{\text{linear},\delta} = (\hat{\mathbf{A}}_{k\lambda} - \mathbf{I}_m) \mathbf{K}_k (\hat{\mathbf{x}} - \mathbf{x}_a) + \hat{\mathbf{A}}_{k\lambda} \boldsymbol{\delta}, \quad (12)$$

and

$$\mathbf{u}_{k\lambda}^\delta - \hat{\mathbf{u}} = (\mathbf{A}_{k\lambda} - \mathbf{I}_n) \hat{\mathbf{u}} + \mathbf{K}_{k\lambda}^\dagger \boldsymbol{\delta}, \quad (13)$$

where $\hat{\mathbf{A}}_{k\lambda}$ is the influence matrix and $\mathbf{A}_{k\lambda}$ is the averaging kernel,

$$\begin{aligned} \hat{\mathbf{A}}_{k\lambda} &= \mathbf{K}_k \mathbf{K}_{k\lambda}^\dagger, \\ \mathbf{A}_{k\lambda} &= \mathbf{K}_{k\lambda}^\dagger \mathbf{K}_k. \end{aligned}$$

The error in $\hat{\mathbf{u}}$ can be expressed as a sum of the linearized smoothing error and linearized noise error,

$$\begin{aligned} \mathbf{e}_{k\lambda,\text{total}}^{\text{linear},\delta} &= (\mathbf{A}_{k\lambda} - \mathbf{I}_n) \hat{\mathbf{u}} + \mathbf{K}_{k\lambda}^\dagger \boldsymbol{\delta} \\ &= \mathbf{e}_{k\lambda,\text{smooth}}^{\text{linear}} + \mathbf{e}_{k\lambda,\text{noise}}^{\text{linear},\delta}. \end{aligned} \quad (14)$$

Because the smoothing error increases with increasing λ and the noise error decreases with increasing λ , the total error has a minimum for an optimal value of λ . Due to the linearity relation $\mathbf{p}_{k\lambda}^{\text{linear},\delta} = \mathbf{K}_k \mathbf{e}_{k\lambda,\text{total}}^{\text{linear},\delta}$, we may assume that the linearized predictive error also possesses a minimum for an optimal value of λ . Except for the L-curve selection criterium, the parameter-choice methods discussed in this subsection select the regularization parameter by analyzing the linearized predictive error $\mathbf{p}_{k\lambda}^{\text{linear},\delta}$ or the linearized total error $\mathbf{e}_{k\lambda,\text{total}}^{\text{linear},\delta}$. In the first

case the analysis is performed in the data space (\mathbf{y}), while in the second case, the analysis is carried out in the state space (\mathbf{x}).

The following parameter-choice methods are implemented in our code.

2.3.1. L-curve method

The L-curve method consists in the analysis of the graph of the curve obtained by plotting the norm of the constraint vector $\|s_{k\lambda}^{\text{linear},\delta}\|$ versus the norm of the residual $\|r_{k\lambda}^{\text{linear},\delta}\|$ in log–log scale [14]. This curve exhibits a typical “L” shape, and the optimal value of the regularization parameter is at the corner of the “L”. The solution corresponding to the corner of the L-curve balances the constraint and residual norms. Moreover, a solution near the L-curve’s corner tends to balance the smoothing and noise errors [14]. The corner of the L-curve is defined as the point on the graph with maximal curvature. Because the residual and constraint norms can be computed analytically, function minimization routines can be used to compute the maximum of the curvature [15].

In the optimal estimation method, the expected values of the constraint and residual norms are known as the degree of freedom for signal and noise, respectively [9]:

$$d_s = \mathcal{E}\{\|s_{k\lambda}^{\text{linear},\delta}\|^2\},$$

$$d_n = \mathcal{E}\{\|r_{k\lambda}^{\text{linear},\delta}\|^2\}.$$

The degree of freedom for signal d_s is a decreasing function of the regularization parameter, while the degree of freedom for noise d_n is an increasing function of the regularization parameter.

2.3.2. Generalized cross-validation method

The generalized cross-validation (GCV) [16,17] is a predictive method which seeks to minimize the linearized predictive risk $\|p_{k\lambda}^{\text{linear},\delta}\|^2$. Since $p_{k\lambda}^{\text{linear},\delta}$ is unknown, the GCV method works instead with the GCV function

$$V(\lambda) = \frac{\|r_{k\lambda}^{\text{linear},\delta}\|^2}{\text{trace}^2\{I_m - \widehat{A}_{k\lambda}\}}. \tag{15}$$

The linearized residual is an increasing function of λ and the GCV method seeks to locate the transition point where the linearized residual changes from a very slowly varying function of λ to a more rapidly increasing function. The denominator in the above equation is a monotonically increasing function of λ , such that V has a minimum. The term $\text{trace}\{I_m - \widehat{A}_{k\lambda}\}$ can be regarded as an estimate of the degree of freedom for noise (Eq. (12)) and V is an estimate of λ which balances the degree of freedom for noise against model fit. Unfortunately, the unique minimum of the GCV function can be very flat, thus leading to numerical difficulties in computing the minimum of V .

2.3.3. Unbiased predictive risk estimator method

In analogy to the GCV method, the unbiased predictive risk estimator (UPRE) method seeks to minimize the linearized predictive risk [18]. Instead of using the GCV function, we consider the linearized UPRE $U_{k\lambda}^{\text{linear},\delta}$,

$$U_{k\lambda}^{\text{linear},\delta}(\lambda) = \|r_{k\lambda}^{\text{linear},\delta}\|^2 + \frac{2}{m} \text{trace}\{\widehat{A}_{k\lambda}\} - 1. \tag{16}$$

Since

$$\mathcal{E}\{\|p_{k\lambda}^{\text{linear},\delta}\|^2\} = \mathcal{E}\{U_{k\lambda}^{\text{linear},\delta}\},$$

we see that $U_{k\lambda}^{\text{linear},\delta}$ is an unbiased estimator for the expected value of the linearized predictive risk. The linearized residual is an increasing function of λ , while the second term $\text{trace}\{\widehat{A}_{k\lambda}\}$ (which can be regarded as an estimate of the degree of freedom for signal, Eq. (12)) is a decreasing function of λ . The UPRE possesses a global minimum and the optimal value of the regularization parameter is chosen as the minimizer of $U_{k\lambda}^{\text{linear},\delta}$.

2.3.4. Minimum bound method

Squaring the total error Eq. (14) and applying the expected value operator yields

$$\mathcal{E}\{\|\boldsymbol{\varepsilon}_{k\lambda,\text{total}}^{\text{linear},\delta}\|^2\} \leq 2(\|\boldsymbol{\varepsilon}_{k\lambda,\text{smooth}}^{\text{linear}}\|^2 + \mathcal{E}\{\|\boldsymbol{\varepsilon}_{k\lambda,\text{noise}}^{\text{linear},\delta}\|^2\}). \quad (17)$$

Note that the smoothing error is a deterministic quantity, while the noise error is a stochastic quantity. The expected value of the linearized noise error is given by

$$\mathcal{E}\{\|\boldsymbol{\varepsilon}_{k\lambda,\text{noise}}^{\text{linear},\delta}\|^2\} = \frac{1}{m} \text{trace}\{\mathbf{K}_{k\lambda}^\dagger \mathbf{K}_{k\lambda}^{\dagger\text{T}}\}, \quad (18)$$

while an estimator of the linearized smoothing error vector is

$$\boldsymbol{\varepsilon}_{k\lambda,\text{smooth}}^{\text{linear},\delta} = (\mathbf{K}_{k\lambda}^\dagger - \mathbf{K}_k^\dagger) \mathbf{w}_k^\delta, \quad (19)$$

with $\mathbf{K}_k^\dagger = (\mathbf{K}_k^\text{T} \mathbf{K}_k)^{-1} \mathbf{K}_k^\text{T}$ being the generalized inverse. The optimal value of the regularization parameter will be determined by minimizing the estimator of the bound of the total error [19,20]. Two aspects are relevant for our analysis:

1. In order to simplify our presentation we neglected the model parameter error. The uncertainties of the forward model can be regarded as an additional “measurement” error. This error together with the “true” measurement error determine the total observation uncertainty. Denoting by \mathbf{b} the model parameters comprising those quantities which influence the measurement and by \mathbf{b}_a the best estimate of the model parameters, we define the “effective noise” caused by the model parameters as $\boldsymbol{\delta}_b = \mathbf{K}_{k\text{b}}(\mathbf{b} - \mathbf{b}_a)$ with $\mathbf{K}_{k\text{b}} = \partial \mathbf{F} / \partial \mathbf{b}$, and the total measurement error as $\boldsymbol{\delta}_t = \boldsymbol{\delta}_b + \boldsymbol{\delta}$. Assuming that $\mathbf{b} - \mathbf{b}_a$ is a stochastic quantity with zero mean and known covariance matrix

$$\mathbf{S}(\mathbf{b}) = \mathcal{E}\{(\mathbf{b} - \mathbf{b}_a)(\mathbf{b} - \mathbf{b}_a)^\text{T}\},$$

we see that the total measurement error is stochastic with zero mean and covariance matrix

$$\mathbf{S}(\boldsymbol{\delta}_t) = \mathbf{K}_{k\text{b}} \mathbf{S}(\mathbf{b}) \mathbf{K}_{k\text{b}}^\text{T} + (1/m) \mathbf{I}_m.$$

Consequently, for the linearized noise error $\boldsymbol{\varepsilon}_{k\lambda,\text{noise}}^{\text{linear},\delta} = \mathbf{K}_{k\lambda}^\dagger \boldsymbol{\delta}_t$, we have

$$\mathcal{E}\{\|\boldsymbol{\varepsilon}_{k\lambda,\text{noise}}^{\text{linear},\delta}\|^2\} = \text{trace}\{\mathbf{K}_{k\lambda}^\dagger \mathbf{S}(\boldsymbol{\delta}_t) \mathbf{K}_{k\lambda}^{\dagger\text{T}}\}. \quad (20)$$

2. The estimator of the linearized smoothing error depends on the noisy data vector and for real data with significant noise level, the linearized smoothing error norm given by Eq. (19) is overestimated. In this case we can estimate the variation of the linearized smoothing error by computing a “theoretical” bound of the linearized smoothing error, i.e.,

$$\|\boldsymbol{\varepsilon}_{k\lambda,\text{smooth}}^{\text{linear},\text{th}}\| \leq \|\mathbf{A}_{k\lambda} - \mathbf{I}_n\|_F \|\widehat{\mathbf{x}} - \mathbf{x}_a\|, \quad (21)$$

where $\|\cdot\|_F$ is the Frobenius norm and $\|\widehat{\mathbf{x}} - \mathbf{x}_a\| \leq \rho \|\mathbf{x}_a\|$. The scalar ρ determines the radius of the sphere centered at \mathbf{x}_a within the exact solution is assumed to lie. Information about the influence of the noise level on the solution can be obtained from the discrete Picard condition [14]. Assuming $\mathbf{H} = \mathbf{I}_n$ and considering the singular value decomposition (SVD) of the matrix \mathbf{K}_k , $\mathbf{K}_k = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\text{T}$, we can express $\mathbf{u}_{k\lambda}^\delta$ as

$$\mathbf{u}_{k\lambda}^\delta = \sum_{i=1}^n f_{i\lambda} \frac{\langle \mathbf{u}_i, \mathbf{w}_k^\delta \rangle}{\sigma_i} \mathbf{v}_i, \quad (22)$$

where $f_{i\lambda} = \sigma_i^2 / (\sigma_i^2 + \lambda^2)$ are the filter factors, σ_i are the singular values and \mathbf{u}_i and \mathbf{v}_i are the left and right singular vectors of \mathbf{K}_k . The singular values σ_i decrease monotonically (by definition), while the absolute value of the Fourier coefficients $|\langle \mathbf{u}_i, \mathbf{w}_k^\delta \rangle|$ also decay, on the average, until they settle at a level τ_w determined by the errors in \mathbf{w}_k^δ . If the coefficients $|\langle \mathbf{u}_i, \mathbf{w}_k^\delta \rangle|$ level off at τ_w for $i > i_w$ before the σ_i level off, we can roughly recover the first i_w SVD components of the solution, while the remaining $n - i_w$ components are dominated by the errors. For data with significant noise level, the level off index i_w is small. Note that in order to guarantee small regularization errors, the absolute value of the Fourier coefficients must decay faster than the singular values (discrete Picard condition). For a regularization matrix $\mathbf{H} \neq \mathbf{I}_n$, the singular values σ_i are replaced by the generalized singular values γ_i and the plots of the generalized singular values γ_i , Fourier coefficients $|\langle \mathbf{u}_i, \mathbf{w}_k^\delta \rangle|$,

and Picard coefficients $|\langle \mathbf{u}_i, \mathbf{w}_k^\delta \rangle|/\gamma_i$ indicate if Tikhonov regularization is able to produce a useful regularized solution.

2.3.5. Noise error criterium

In contrast to the smoothing error, the noise error can be computed accurately and a simple parameter-choice method can be formulated. The idea is simply to choose the regularization parameter $\lambda = \lambda(\Delta)$ such that the noise error is equal to an a priori upper bound, i.e.,

$$\sqrt{\mathcal{E}\{\|\mathbf{e}_{k\lambda, \text{noise}}^{\text{linear}, \delta}\|^2\}} = \Delta \|\mathbf{x}_{k\lambda}^\delta\|, \tag{23}$$

where typically $\Delta = 0.05, \dots, 0.1$.

2.3.6. Extension to the nonlinear case

The expressions of the predictive and total errors have been derived under a strong linearity assumption. Therefore, the regularization parameter computed at the first iteration can be used for the subsequent iterations only for nearly linear problems. In this specific case, the above selection criteria appear as a priori parameter-choice methods. For nonlinear problems we have two options: the regularization parameter can be kept constant or it can be variable during the iterative process. In the first case, the regularization method is the Tikhonov regularization [6], while in the second case, the regularization method is the iteratively regularized Gauss–Newton method [7].

2.3.6.1. Tikhonov regularization. For Tikhonov regularization, the above criteria are reformulated as a posteriori parameter-choice methods. The nonlinear minimization problem is solved for a set of regularization parameters λ , and for each λ , the convergence of the sequence of iterates $(\mathbf{x}_{k\lambda}^\delta)$ is used as stopping rule. Denoting by $\mathbf{x}_\lambda^\delta$ the limit of the sequence of iterates, i.e., $\mathbf{x}_{k\lambda}^\delta \rightarrow \mathbf{x}_\lambda^\delta$ as $k \rightarrow \infty$, we may compute the optimal value of the regularization parameter by using the above parameter choice methods with $\mathbf{x}_\lambda^\delta$ in place of $\mathbf{x}_{k\lambda}^\delta$ and $\mathbf{r}_\lambda^\delta$ in place of $\mathbf{r}_{k\lambda}^{\text{linear}, \delta}$, where $\mathbf{r}_\lambda^\delta = \mathbf{F}(\mathbf{x}_\lambda^\delta) - \mathbf{y}^\delta$. It should be mentioned that the nonlinear L-curve method has been introduced by Eriksson [21], the GCV method has been extended to the nonlinear case by O’Sullivan and Wahba [17], while the minimum bound method has been originally formulated in a deterministic setting by Raus [19] and in a discrete, semi-stochastic setting by Lukas [20]. Because the numerical realization of Tikhonov regularization with an a posteriori parameter-choice method requires to solve the nonlinear minimization problem several times for different regularization parameters, this strategy is time consuming and is not appropriate for the operational usage of a retrieval processor.

2.3.6.2. Iteratively regularized Gauss–Newton method. This method can be regarded as a Tikhonov regularization with a variable regularization parameter, i.e., (λ_k) is a monotonically decreasing sequence. The following recursive procedures are used to construct the sequence of iterates [3]:

$$\lambda_k = \beta \lambda_k^{\text{opt}} + (1 - \beta) \lambda_{k-1} \quad \text{with } 0 \leq \beta \leq 1, \tag{24}$$

or

$$\lambda_k = \beta_k \lambda_{\min} + (1 - \beta_k) \lambda_{k-1} \quad \text{with } \beta_k = \frac{r_{k,\lambda}^\delta}{r_{k-1,\lambda}^\delta} \text{ and } r = \|\mathbf{r}\|. \tag{25}$$

In the first case, the parameter β controls the decay rate of the sequence of iterates and λ_k^{opt} is the optimal value of the regularization parameter computed with an appropriate parameter-choice method. In the second case, the upper and lower bounds of the sequence of regularization parameters λ_{\max} and λ_{\min} have to be chosen in advance and the decay rate is given by the ratio of the residuals at two consecutive steps. For $\beta_k \approx 0$ (large reduction of the residual), the new value of the regularization parameter is close to the previous value, while for $\beta_k \approx 1$ (small reduction of the residual), the regularization parameter is considerably decreased towards λ_{\min} . This parameter-choice method allows enough regularization to be applied at the beginning of the iterations and then to be gradually decreased. A specific choice of λ_{\max} and λ_{\min} is discussed in Section 3.

Convergence results for solving nonlinear ill-posed problems by using the iteratively regularized Gauss–Newton method have been given by Blaschke et al. [22], Hohage [23], and Deuffhard et al. [24].

2.4. Global regularization matrix

Due to the multi-parameter character of the problem, specific constraints have to be imposed on each component \mathbf{z}_i of the state vector \mathbf{x} , cf., Eq. (3). The global regularization term is expressed as a linear combination of regularization terms corresponding to each component, i.e.,

$$\Lambda(\mathbf{x}) = \sum_{i=1}^P v_i \|\mathbf{L}_i(\mathbf{z}_i - \mathbf{z}_{ai})\|^2 = \|\mathbf{H}(\mathbf{x} - \mathbf{x}_a)\|^2, \quad (26)$$

where \mathbf{L}_i is the i th regularization matrix corresponding to \mathbf{z}_i and the global regularization matrix \mathbf{H} has the block structure

$$\mathbf{H} = \begin{bmatrix} \sqrt{v_1}\mathbf{L}_1 & 0 & \dots & 0 \\ 0 & \sqrt{v_2}\mathbf{L}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \sqrt{v_P}\mathbf{L}_P \end{bmatrix}. \quad (27)$$

The regularization matrices \mathbf{L}_i are chosen in accordance with the peculiarities of the solutions \mathbf{z}_i . For atmospheric profiles, regularization matrices accounting for the smoothness of the solution (discrete approximations of the first and second derivatives or a priori covariance matrices) are used, while for auxiliary parameters, diagonal matrices controlling the magnitude of the solution are appropriate.

The weighting factors v_i satisfy the normalization condition $\sum_i v_i = 1$ and give the contribution of each regularization matrix \mathbf{L}_i to the global regularization matrix \mathbf{H} . They are computed at the first iteration. For this purpose, we consider the set of one-parameter regularized function

$$\mathcal{F}_{0i}^{\text{linear}}(\mathbf{x}) = \|\mathbf{F}(\mathbf{x}_0) - \mathbf{y}^\delta + \mathbf{K}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|^2 + \lambda_i^2 \|\mathbf{H}_i(\mathbf{x} - \mathbf{x}_a)\|^2, \quad (28)$$

with $i = 1, 2, \dots, P$. Here, \mathbf{H}_i is the i th extended regularization matrix

$$\mathbf{H}_i = \begin{bmatrix} \delta_{i1}\mathbf{L}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \delta_{iP}\mathbf{L}_P \end{bmatrix}, \quad (29)$$

δ_{ik} is the Kronecker symbol and λ_i is the regularization parameter for the one-parameter regularized function $\mathcal{F}_{0i}^{\text{linear}}$. In fact, each one-parameter regularized function contains a regularization term corresponding to one component of the state vector. A possible choice of the weighting factors is

$$v_i = \frac{(\lambda_i^{\text{opt}})^2}{\sum_{j=1}^P (\lambda_j^{\text{opt}})^2}, \quad (30)$$

where λ_i^{opt} is the optimal regularization parameters for the one-parameter regularized function $\mathcal{F}_{0i}^{\text{linear}}$. In practice, λ_i^{opt} can be computed by using the GCV method or the L-curve method. Note that the GCV method has been used for the linear multi-parameter case by Brezinski et al. [25]. If the weighting factors are specified and the global regularization matrix \mathbf{H} is derived, the minimization algorithm can be used to compute the approximate solution.

2.5. Error analysis

The accuracy of a retrieval method can be characterized by the discrepancy between the approximate solution $\mathbf{x}_\lambda^\delta$ and the exact solution $\hat{\mathbf{x}}$. The approximate solution $\mathbf{x}_\lambda^\delta$ is defined as the limit of the sequence of iterates whether the regularization parameter is kept constant or varies during the iterative process.

Considering the linearization around $\mathbf{x}_\lambda^\delta$ and assuming that the linearity relation holds at $\mathbf{x} = \hat{\mathbf{x}}$, yields the total error as a combination of smoothing and noise error

$$\begin{aligned}\mathbf{e}_{\lambda,\text{total}}^\delta &= \mathbf{x}_\lambda^\delta - \hat{\mathbf{x}} \\ &= (\mathbf{A}_\lambda - \mathbf{I}_n)(\hat{\mathbf{x}} - \mathbf{x}_a) + \mathbf{K}_\lambda^\dagger \boldsymbol{\delta} \\ &= \mathbf{e}_{\lambda,\text{smooth}} + \mathbf{e}_{\lambda,\text{noise}}^\delta.\end{aligned}\quad (31)$$

Since the smoothing error is a deterministic quantity, $\mathbf{x}_\lambda^\delta$ is always a biased estimator of $\hat{\mathbf{x}}$ [14]. An estimator for the smoothing error vector is

$$\mathbf{e}_{\lambda,\text{smooth}}^\delta = \mathbf{K}_\lambda^\dagger \mathbf{r}_\lambda^\delta \quad (32)$$

while the covariance matrix of the noise error is

$$\mathbf{S}(\mathbf{e}_{\lambda,\text{noise}}^\delta) = (1/m)\mathbf{K}_\lambda^\dagger \mathbf{K}_\lambda^{\dagger T}.$$

For data with significant noise level,

$$\mathbf{e}_{\lambda,\text{smooth}}^\delta = (\mathbf{A}_\lambda - \mathbf{I}_n)(\mathbf{x}_\lambda^\delta - \mathbf{x}_a)$$

can be used as an alternative estimator of the smoothing error vector.

3. Ozone retrieval from SCIAMACHY limb measurements

3.1. Observed data and retrieval setup

In this section we consider the retrieval of O_3 profiles from UV limb scatter measurements made with the SCIAMACHY grating instrument. The forward model for SCIAMACHY limb radiance simulation is a single-scattering code, where the multiple scattering effect is taken into account by using look-up table corrections [26]. The model does not include the boundary condition at the Earth's surface and the rotational Raman scattering. For satellite-based UV-visible measurements, rotational Raman scattering contributes to the Ring effect [27]. The Ring effect is taken into account by adding a Ring reference spectrum to the simulated radiance. Under these circumstances, the surface albedo and the Ring parameters (multiplicative factors of the Ring reference spectra) are auxiliary parameters which have to be retrieved together with the atmospheric O_3 profile.

SCIAMACHY measurements recorded on 5 February 2004 at latitude 48° south and longitude 45° east are used for the retrieval. These observations consist of 17 limb spectra covering tangent heights from 10.35 to 62.76 km in steps of about 3 km. The zenith and azimuthal angles of the line of sight (at top of atmosphere) are about 63° and 192° , respectively, while the solar zenith and azimuthal angles are about 48° and 64° , respectively.

For our retrieval, only 15 spectra for tangent altitudes between 13.65 and 59.55 km have been used. An altitude retrieval grid with 21 grid points between 14 and 100 km is considered. The grid is uniform between 14 and 59 km with 3 km steps. The spectral domain of analysis is a window of 323.4–333.4 nm in channel 2 of the SCIAMACHY instrument.

The a priori and initial gas profiles were assumed to be identical and were chosen from the US standard atmosphere. The initial value of the surface albedo is 0.3, while the initial Ring parameters are chosen such that the average Ring spectrum (over each scan) is 15% from the simulated radiance. The deviation of the atmospheric profile with respect to the a priori profile is assumed to be smaller than 30%, the surface albedo is considered to be bounded by the reference values 0.2 and 0.8, and we impose that the Ring spectrum will not exceed 30% from the simulated radiance.

The regularization matrix for ozone is the Choleski factor of an a priori covariance matrix, while the regularization matrix for the auxiliary parameters is a diagonal matrix. Considering the one-parameter regularized functions (28) and computing the optimal values of the regularization parameters by the L-curve method yields $\nu_{\text{O}_3} = 0.98$ and $\nu_{\text{aux}} = 0.02$. The global regularization matrix is then computed according to Eq. (27).

3.2. Results

In Fig. 1, we compare the regularization parameter choice methods at the first iteration. The L-curve (a) has a pronounced corner that can be efficiently detected by an algorithm based on the maximization of its curvature. The optimal value of the regularization parameter is $\lambda_0^{\text{LCR}} = 8.04 \times 10^{-1}$. On the contrary, the GCV function (b) is very flat and its minimum cannot be correctly localized. The optimal value of the regularization parameter predicted by the L-curve method is close to the transition point where the GCV function (15) changes from a very slowly varying function of λ_0 to a more rapidly increasing function.

The failure of the GCV method is accompanied by the failure of the UPRE method (c) and the minimum bound method (Fig. 1d). The linearized UPRE (16) has a similar behavior as the GCV function. The curve shows that the influence of the noise is significant, i.e., the linearized residual (which is an increasing function of λ) dominates the degree of freedom for signal (a decreasing function of λ). The linearized smoothing error norm (19) is strongly overestimated and the linearized total error does not possess a global minimum. In fact, the estimator of the linearized smoothing error (expressed in terms of the linearized residual w_0^δ) is inaccurate.

The influence of the noise on the solution can be analyzed by plotting the generalized singular values, the Fourier and the Picard coefficients ($|\langle u_i, w_k^\delta \rangle|/\gamma_i$) at the first iteration. The results are shown in Fig. 2. The

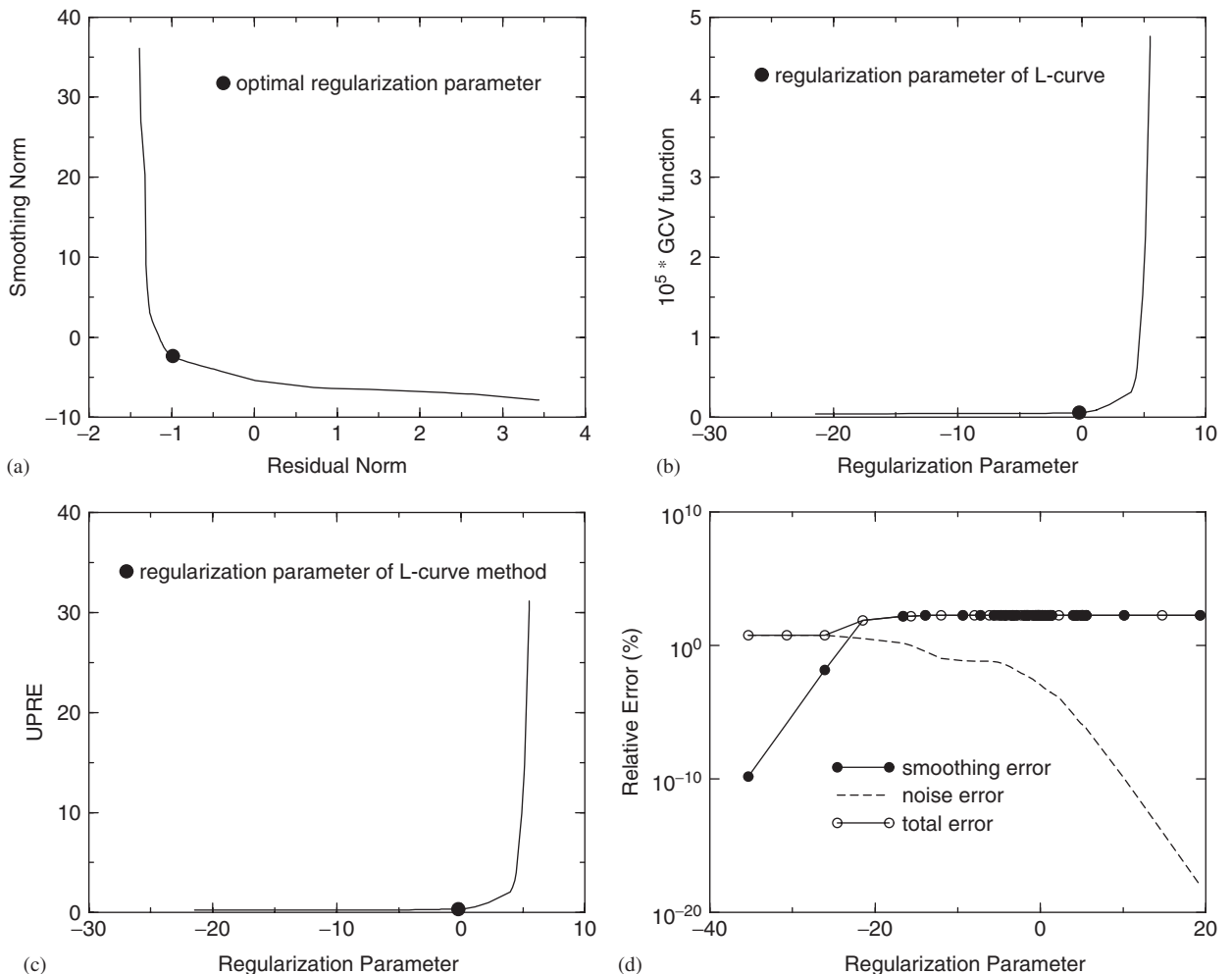


Fig. 1. Regularization parameter choice methods: (a) L-curve, (b) GCV function, (c) linearized unbiased predictive risk estimator (UPRE) and (d) linearized error curves for minimum bound method.

Fourier coefficients level off at $i_w \approx 15$, i.e., we cannot recover the generalized singular value components of the solution for $i > 15$ (more than 75%) because the Fourier coefficients are dominated by noise. If the regularization parameter decreases, more singular values are effectively included in the regularized solution. As a consequence, the noise error increases and we expect that the noise errors will dominate the regularized solution for $\lambda \approx \gamma_{15} = 2.5 \times 10^{-2}$.

In Fig. 3 we plot the theoretical bound of the linearized smoothing error given by Eq. (21) and the corresponding linearized total error. For the radius ρ we considered the values 0.02 and 0.05. In both cases, the linearized total error possesses a minimum which balances the smoothing and noise errors. Because the computation of the smoothing error is inaccurate we compute the regularization parameter according to the noise error criterium (23). For $\Delta = 0.1$ we found $\lambda_0^{\text{noise}}(0.1) = 6.02 \times 10^{-6}$, while for $\Delta = 0.05$, $\lambda_0^{\text{noise}}(0.05) = 5.52 \times 10^{-2}$.

The above analysis shows that the GCV method, UPRE method and minimum bound method are not able to predict a reliable value of the regularization parameter. The total error is too large and these parameter-choice methods fail. Therefore, we choose the L-curve method and the noise error criterium as parameter-choice methods and consider three schemes for regularization parameter selection. In the first scheme, the regularization parameter is constant during the iterative process and is given by $\lambda = \lambda_0^{\text{LCR}}$. In the second and third schemes, the regularization parameter is variable and is computed according to Eq. (25). We choose $\lambda_{\text{max}} = \lambda_0^{\text{LCR}}$. For the second scheme we set $\lambda_{\text{min}} = \lambda_0^{\text{noise}}(0.05)$, while for the third scheme we choose $\lambda_{\text{min}} = \lambda_0^{\text{noise}}(0.1)$. In view of the Picard analysis, we expect that the third solution will be dominated by noise errors. Setting $\lambda_{\text{min}} = \lambda_0^{\text{noise}}$ we intend to guarantee that the noise error remains bounded at the (approximate) solution. The errors at the solution are shown in Fig. 4. Below 50 km, the smoothing errors are smaller than 7×10^{-2} for the first scheme, 3.5×10^{-2} for the second scheme and 5×10^{-4} for the third scheme. For the same altitude range, the noise errors are of about 2.2×10^{-3} in the first case, 5×10^{-2} in the second case and 5×10^{-1} in the third case. It should be observed that the noise errors reach a maximal value which exceeds the tolerance Δ . Note that Δ is the tolerance of the norm of the noise error vector for all components of the state vector (including auxiliary parameters). Certainly, the regularization parameter $\lambda_0^{\text{noise}}(\Delta)$ can be computed for a specific atmospheric profile or a user required altitude range.

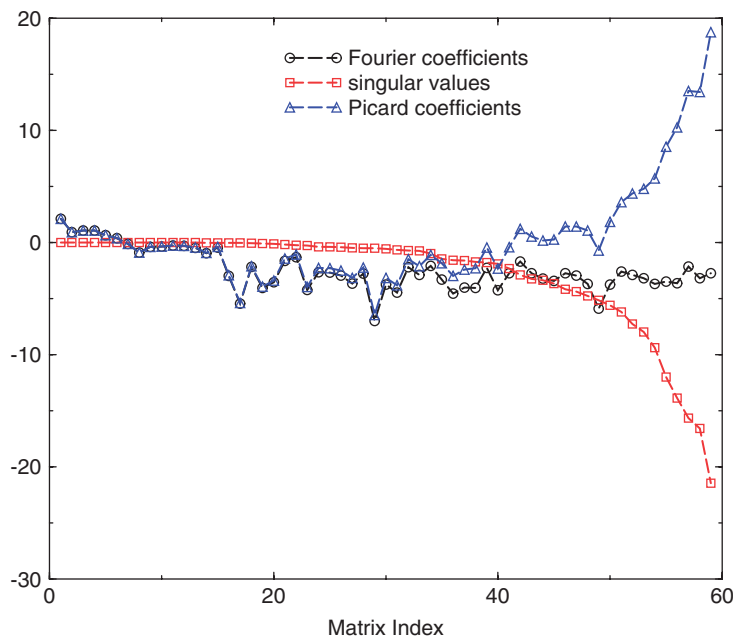


Fig. 2. Generalized singular values, Fourier and Picard coefficients.

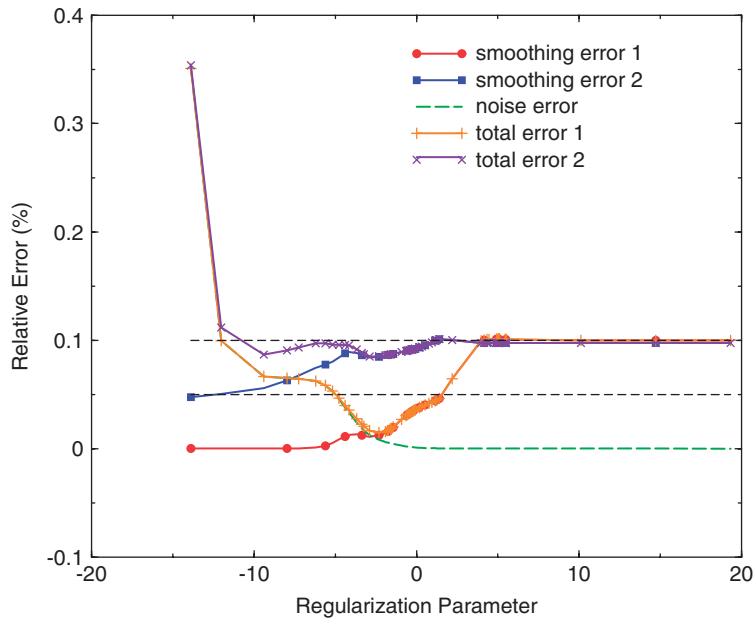


Fig. 3. Theoretical error curves for $\rho = 0.02$ (curve 1) and $\rho = 0.05$ (curve 2).

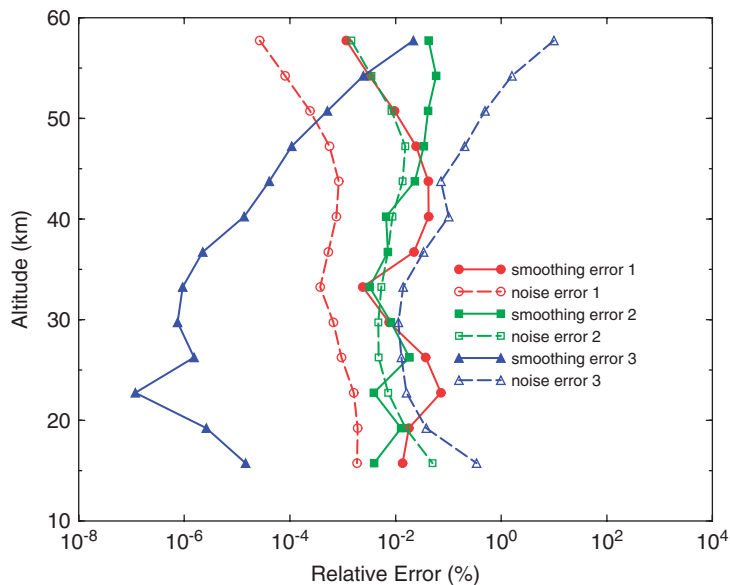


Fig. 4. Errors at the solution for the regularization methods with constant regularization parameter, $\lambda = \lambda_0^{\text{LCR}}$ (curve 1), variable regularization parameter, $\lambda_{\min} = \lambda_0^{\text{noise}}$ (0.05) (curve 2) and variable regularization parameter, $\lambda_{\min} = \lambda_0^{\text{noise}}$ (0.1) (curve 3).

The best trade-off between the smoothing and noise errors appears to be given by the second scheme. For the first scheme, the solution is oversmoothed and is dominated by smoothing errors, while for the third scheme, the solution is underregularized and is dominated by noise errors. The averaging kernels plotted in Fig. 5 give an impression of the smoothing errors.

The retrieved profiles are plotted in Fig. 6. Clearly, the solution given by the first scheme is closer to the a priori than the other two solutions. It should be observed that the solution obtained by the third scheme is

smooth and no oscillations are visible. This peculiarity of the solution is typical for the iteratively regularized Gauss–Newton method that avoids oscillating solutions by decreasing the regularization parameter in a safe way.

In contrast, Tikhonov regularization with a constant regularization parameter may lead to unsmoothed solutions. Small oscillations are visible in Fig. 7 for the Tikhonov solution with $\lambda = \lambda_0^{\text{noise}}(0.1)$. Because the averaging kernel and the regularized generalized inverse are almost insensitive to small variations of the state vector, these solutions are characterized by the same resolution (averaging kernels), smoothing and noise errors. In this case, the smoothness of the solution, defined as the constraint norm $\|\mathbf{L}_2(\mathbf{x}_{\lambda, \text{O}_3}^{\delta} - \mathbf{x}_{\text{a}, \text{O}_3})\|$, can be used as an additional diagnostic tool.

The histories of residuals, iterates and regularization parameters reveal some interesting aspects of the iterative process. The results plotted in Fig. 8 correspond to the third regularization scheme. After five iterations, the residual level off at τ_r . At this stage of our analysis we recall an important stopping criterium for regularization methods: due to the inherent instability of ill-posed problems, an iteration method has to be stopped appropriately to guarantee stability of the iterates [6]. This requirement can be achieved by using the discrepancy principle as stopping rule, i.e., the iteration is stopped at the first index for which the residual is smaller than an upper bound of the measurement error $\tau = \sqrt{\mathcal{E}\{\|\delta\|^2\}}$. In practice, other type of errors (model parameter errors, systematic and random errors) contribute to the “total” or “effective” measurement error and $\tau_r > \tau$. Therefore, the convergence of the iterates has been chosen as stopping rule and the ratio τ_r/τ can be regarded as a measure of the degree of “total” errors. For the iteration steps $k = 5, 6, 7$, the changes of the residual and state vector are small and we may conclude that these points belong to the corner region of the nonlinear L-curve. At $k = 8$, the state vector changes significantly, while the residual decreases slowly. In this case, the underregularized region of the nonlinear L-curve (the vertical part) is attained. The small changes of the state vector for $k = 8, 9, 10$ are due to the small variations (convergence) of the regularization parameter. Note that the second regularization scheme selects the solution corresponding to $k = 6$.

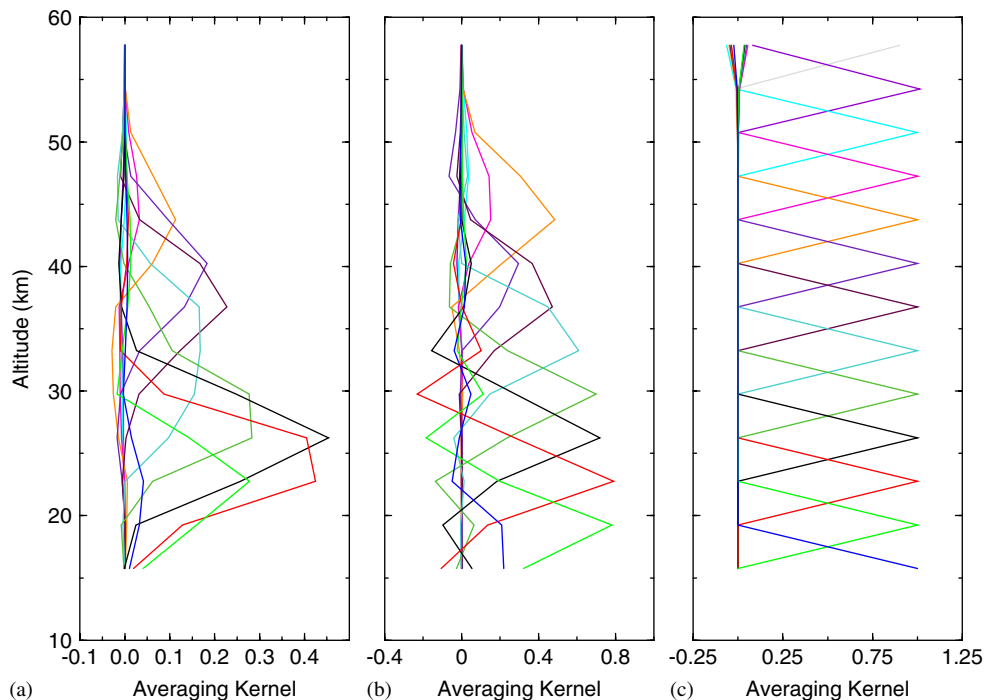


Fig. 5. Averaging kernels for the regularization methods with: (a) constant regularization parameter, $\lambda = \lambda_0^{\text{LCR}}$, (b) variable regularization parameter, $\lambda_{\text{min}} = \lambda_0^{\text{noise}}(0.05)$ and (c) variable regularization parameter, $\lambda_{\text{min}} = \lambda_0^{\text{noise}}(0.1)$.

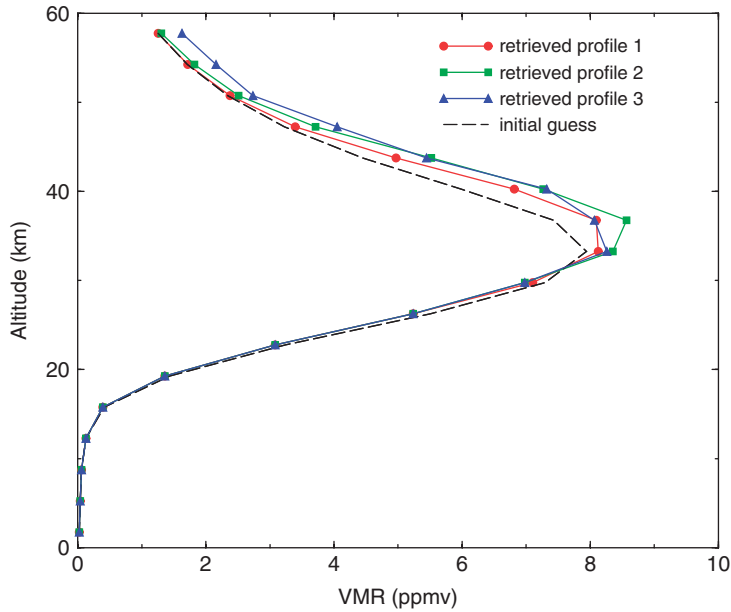


Fig. 6. Retrieved profiles for the regularization methods with constant regularization parameter, $\lambda = \lambda_0^{\text{LCR}}$ (curve 1), variable regularization parameter, $\lambda_{\min} = \lambda_0^{\text{noise}}(0.05)$ (curve 2) and variable regularization parameter, $\lambda_{\min} = \lambda_0^{\text{noise}}(0.1)$ (curve 3).

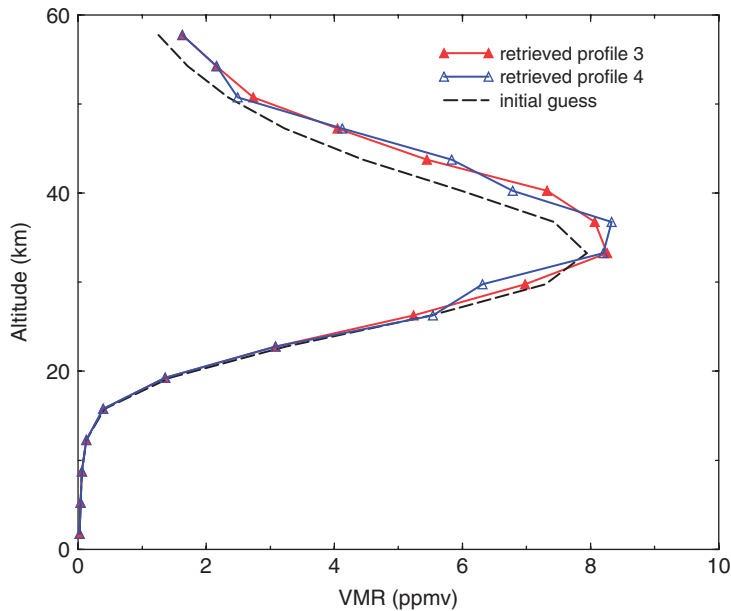


Fig. 7. Retrieved profiles for the regularization methods with variable regularization parameter, $\lambda_{\min} = \lambda_0^{\text{noise}}(0.1)$ (curve 3) and constant regularization parameter, $\lambda = \lambda_0^{\text{noise}}(0.1)$ (curve 4).

4. Conclusions

An inversion algorithm for atmospheric remote sensing has been presented. The algorithm incorporates the Tikhonov regularization and the iteratively regularized Gauss–Newton method into a common shell and is devoted to the solution of multi-parameter and bound-constrained inversion problems. The bound

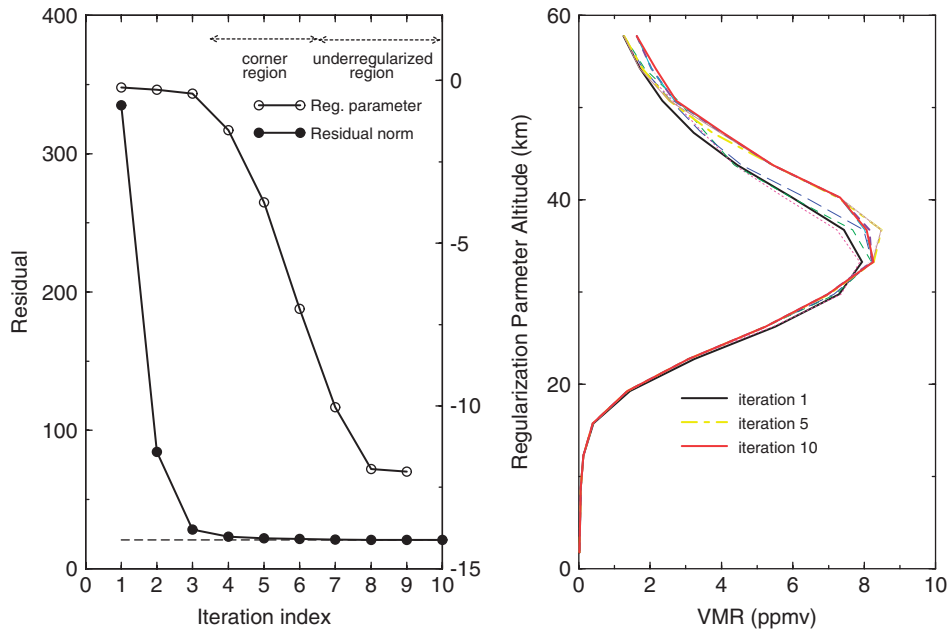


Fig. 8. Histories of residuals, iterates and regularization parameters for the regularization method with variable regularization parameter and $\lambda_{\min} = \lambda_0^{\text{noise}}(0.1)$.

constrained minimization problem can be solved by using a Gauss–Newton method with an active-set algorithm or a trust-region method with a local active-set strategy to select the step. Several parameter-choice methods as for instance the L-curve method, the GCV approach, the UPRE method, the minimum bound method and the noise error criterium are implemented in the code. These methods guarantee an optimal choice of the regularization parameter; for the applications presented in this paper, the L-curve criterion appears to be the superior method. Diagnostic tools such as the discrete Picard condition and the histories of iterates and residuals offer additional information about the iterative process. The error analysis and inversion is performed in a semi-stochastic setting and consists of the computation of the smoothing and noise errors at the (approximate) solution. All these features lead to an efficient and robust inversion algorithm for atmospheric remote sensing.

References

- [1] Fischer H, Oelhaf H. Remote sensing of vertical profiles of atmospheric trace constituents with MIPAS limb–emission spectrometers. *Appl Opt* 1996;35:2787–96.
- [2] Bovensmann H, Burrows JP, Buchwitz M, Frerick J, Noël S, Rozanov VV, Chance KV, Goede AHP. SCIAMACHY: mission objectives and measurements mode. *J Atmos Sci* 1999;56:127–50.
- [3] Doicu A, Schreier F, Hess M. Iteratively regularized Gauss–Newton method for atmospheric remote sensing. *Comp Phys Comm* 2002;148:214–26.
- [4] Doicu A, Schreier F, Hess M. Iteratively regularized Gauss–Newton method for bound–constraint problems in atmosphere remote sensing. *Comp Phys Comm* 2003;153: 59–65.
- [5] Doicu A, Schreier F, Hess M. Iterative regularization methods for atmospheric remote sensing. *JQSRT* 2004;83:47–61.
- [6] Engl HW, Hanke M, Neubauer A. *Regularization of inverse problems*. Dordrecht, NL: Kluwer; 1996.
- [7] Bakushinskii AB. The problem of the convergence of the iteratively regularized Gauss–Newton method. *Comput Math Phys* 1992;32:1353–59.
- [8] von Clarmann T, Ceccherini S, Doicu A, Dudhia A, Funke B, Grabowski U, et al. A blind test retrieval experiment for infrared limb emission spectrometry *J Geophys Res* 2003;108(D23).
- [9] Rodgers CD. *Inverse methods for atmospheric sounding: theory and practise*. Singapore: World Scientific; 2000.
- [10] Bjoerck Å. *Numerical methods for least squares methods*. Philadelphia, PA: SIAM; 1996.
- [11] Gill PE, Murray W, Wright MH. *Practical optimization*. London, UK: Academic Press; 1981.

- [12] Gay DM. A trust-region approach to linearly constrained optimization. In: Griffiths DF, editor. Proceedings of the 10th biennial conference on numerical analysis. Lecture notes in mathematics, vol. 1066. Berlin: Springer; 1984. p. 72–105.
- [13] Dennis Jr JE, Gay DM, Welsch RE. An adaptive nonlinear least-squares algorithm. *ACM Trans Math Soft* 1981;7:348–68.
- [14] Hansen PC. Rank deficient and discrete ill-posed problems: numerical aspects of linear inversion. Philadelphia, PA: SIAM; 1998.
- [15] Kahaner D, Moler C, Nash S. Numerical methods and software. Englewood Cliffs, NJ: Prentice-Hall; 1974.
- [16] Wahba G. Spline models for observational data. Philadelphia, PA: SIAM; 1990.
- [17] O’Sullivan F, Wahba G. A cross validated Bayesian retrieval algorithm for nonlinear remote sensing experiments. *J Comput Phys* 1985;59:441–55.
- [18] Vogel CR. Computational methods for inverse problems. Philadelphia, PA: SIAM; 2002.
- [19] Raus T. On the discrepancy principle for the solution of ill-posed problems. *Acta Commentat Univ Tartuensis Math* 1984;672:16–26 [in Russian].
- [20] Lukas MA. Asymptotic behavior of the minimum bound method for choosing the regularization parameter. *Inverse Problems* 1998;14:149–59.
- [21] Eriksson J. Optimization and regularization of nonlinear least squares problems. Ph.D. thesis, Department of Computing Science, Umea University, Sweden, 1996.
- [22] Blaschke B, Neubauer A, Scherzer O. On convergence rates for the iteratively regularized Gauss–Newton method. *IMA J Numer Anal* 1997;17: 421–36.
- [23] Hohage T. Logarithmic convergence rates of the iteratively regularized Gauss–Newton method for an inverse potential and an inverse scattering problem. *Inverse Problems* 1997;13:1279–300.
- [24] Deuffhard P, Engl HW, Scherzer O. A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinity invariant conditions. *Inverse Problems* 1998;14:1081–106.
- [25] Brezinski C, Redivo-Zaglia M, Rodriguez G, Seatzu S. Multi-parameter regularization techniques for ill-conditioned linear systems. *Numer Math* 2001;94:203–28.
- [26] Flittner DE, Bhartia PK, Herman BM. O₃ profiles retrieved from limb scatter measurements: theory. *Geophys Res Lett* 200;27:2601–04.
- [27] Vountas M, Rozanov VV, Burrows JP. Ring effect: impact of rotational Raman scattering on radiative transfer in Earth’s atmosphere. *JQSRT* 1998;60:943–61.