ELSEVIER

Computer Physics Communications 148 (2002) 214–226

Computer Physics Communications

www.elsevier.com/locate/cpc

Iteratively regularized Gauss–Newton method for atmospheric remote sensing

Adrian Doicu *, Franz Schreier, Michael Hess

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

Received 16 April 2002; received in revised form 2 May 2002

Abstract

In this paper we present an inversion algorithm for nonlinear ill-posed problems arising in atmospheric remote sensing. The proposed method is the iteratively regularized Gauss–Newton method. The dependence of the performance and behaviour of the algorithm on the choice of the regularization matrices and sequences of regularization parameters is studied by means of simulations. A method for improving the accuracy of the solution when the identity matrix is used as regularization matrix is also discussed. Results are presented for atmospheric temperature retrievals from a far infrared spectrum observed by an airborne uplooking heterodyne instrument.

PACS: 02.60.Pn; 02.50; 07.05.Kf; 42.68.W

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

1. Introduction

The Bayesian approach has a dominating role in atmospheric remote sensing where it is well known as "optimal estimation" [1]. A priori information about the atmospheric state is encapsulated in the form of probability distributions, which are independent of the observed data. When such distributions are combined with probabilistic information about data uncertainties (both random and theoretical) it is possible to derive a final (a posteriori) probability distribution assimilating both types of information. However, the construction of the a priori probability distribution is a controversial matter when statistical information about atmospheric variability is poor. In this case regularization methods accounting for deterministic information about the atmospheric state parameters are a pleasant alternative.

One of the best understood regularization methods for nonlinear ill-posed problems is the method of Tikhonov regularization [2–5]. O'Sullivan and Wahba [5] used Tikhonov regularization with an extended form of the generalized cross validation for vertical temperature retrieval. Besides generalized cross validation technique other a posteriori regularization parameter choice methods are the discrepancy principle [6] and the nonlinear L-curve

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

0010-4655/02/$ – see front matter
PII: S0010-4655(02)00555-6

criterion [7]. It must be emphasized that Tikhonov regularization is time consuming, because the selection of the regularization parameter requires the solution of several nonlinear minimization subproblems. A 'cheap' version of Tikhonov regularization for nonlinear problems is the iteratively regularized Gauss–Newton method introduced by Bakushinskii [8]. This method can be regarded as a Tikhonov regularization with a variable regularization parameter. The mathematical foundations of the method were discussed by Blaschke et al. [9], Hohage [10], and Deuflhard et al. [11]. In atmospheric inversion this method was used by Tautenhahn [12] for temperature retrieval. Tautenhahn used the identity matrix as regularization matrix and a parameter choice method based on the noise level.

This paper has several aims: to present the peculiarities of an iteratively regularized Gauss–Newton algorithm and to analyze the influence of different regularization matrices and sequences of regularization parameters on the inversion performances of a vertical temperature profile. From a practical point of view we are also interested in the feasibility of temperature retrieval by airborne far infrared atmospheric spectroscopy.

2. Formulation of the discrete problem

Atmospheric remote sensing in the microwave and infrared spectral regions can be done by measurements of the thermal emission of the atmosphere. From a computational point of view the basic problem is the inversion of the radiative transfer equation [13]. 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 thermodynamical equilibrium)

$$I(\nu) = I_b(\nu)\mathcal{T}(\nu, s_b) - \int_0^{s_b} B\big(\nu, T(s')\big)\frac{\partial \mathcal{T}(\nu, s')}{\partial s'}\,\mathrm{d}s', \tag{1}$$

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

$$\mathcal{T}(\nu, s) = \exp\left[-\int_0^{s} \sum_g k_g\big(\nu, p(s'), T(s')\big)n_g(s')\,\mathrm{d}s'\right], \tag{2}$$

where p is the atmospheric pressure, n_g is the number density of molecule g and k_g is its absorption cross section. In general, the absorption cross section is obtained by summing over the contributions from many lines. For an individual line the spectral absorption cross section is the product of the temperature-dependent line strength and a normalized line shape function describing the broadening mechanism. For the infrared and under atmospheric conditions, 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. The instrumental response is taken into account by convolution of the monochromatic intensity spectrum (1) with an instrumental line shape function, while a further convolution will be required to account for the finite field of view.

Spectroscopic instruments working in the infrared spectral region measure the intensity (spectral radiance) at a finite number of typically equidistant wavenumbers $\nu_i,\ i = 1, \ldots, m$ (equivalent to frequencies or wavelengths). Consequently a collocation method, with collocation points given by the spectral characteristics of the instrument, is used to discretize the integral equation (1) with respect to the left-hand side. In addition, a quadrature approach is employed to approximate the integrals in Eqs. (1) and (2). As there is a unique relation between the path variable s and the altitude h, it is convenient to consider the temperature T or the molecular density profiles n_g at altitudes $h_j,\ j = 1, \ldots, n$, as unknowns of the inverse problem. Note that it is frequently justified to neglect the horizontal (latitudinal and longitudinal) variability of the atmospheric parameters and to retrieve the temperature or molecular density profiles as a function of altitude. In practice, the retrieval grid $h_j,\ j = 1, \ldots, n$, is chosen as the altitude grid specified in the atmospheric profiles data base [14]. The data corresponding to the retrieval grid

are interpolated to the integration grid by using some interpolation methods, as for instance the piecewise cubic Hermite interpolation. In our forward model a trapezoidal quadrature scheme, the method of overlapping parabolas (essentially an optimized version of the AVINT routine from the SLATEC library [15]) and a piecewise cubic Hermite quadrature [16] are implemented. For details concerning the numerical performances of the quadrature schemes we refer to Schreier and Schimpf [17].

The discretization process leads to the nonlinear system of equation

$$z = G(x), \tag{3}$$

where the mapping $G : R^n \to R^m$ representing the forward model is assumed to be continuously differentiable, $z \in R^m$ is the exact data vector (intensity spectrum) and $x \in R^n$ is the state vector containing the atmospheric parameters (temperature or molecular density profiles) to be retrieved. Here R^n stands for the n-dimensional real Euclidean space with the usual inner product $\langle x, y\rangle = x^{\mathrm{T}}y$, while $\|\cdot\|_2$ denotes the l_2 vector norm and the subordinated l_2 matrix norm. In our analysis we assume that the exact data are attainable, i.e. that there exists the exact solution $\hat{x} \in D(G) \subseteq R^n$ such that $z = G(\hat{x})$.

Measurements are made to a finite accuracy and in practice only the noise contaminated data vector z^ε,

$$z^\varepsilon = z + \varepsilon, \tag{4}$$

is available. In the Bayesian approach the state vector is considered to be a random vector and the probability distribution of the state vector that is consistent with the data and the a priori information is computed. In our approach we consider a semi-stochastic data model in the sense that the exact solution $\hat{x}$ is deterministic but the measurement error ε is stochastic with zero mean and the covariance matrix S_ε, $S_\varepsilon = \sigma_\varepsilon^2 \widehat{S}_\varepsilon$, where $\widehat{S}_\varepsilon$ is a symmetric and positive definite matrix (usually a diagonal matrix). The data model is:

$$y = F(\hat{x}) \tag{5}$$

$$y^\delta = y + \delta, \tag{6}$$

where the new variables are $y = \widehat{S}_\varepsilon^{-1/2} z$, $F = \widehat{S}_\varepsilon^{-1/2} G$, $y^\delta = \widehat{S}_\varepsilon^{-1/2} z^\delta$, $\delta = \widehat{S}_\varepsilon^{-1/2}\varepsilon$, and $\widehat{S}_\varepsilon^{-1/2}$ is the inverse of the symmetric square root of $\widehat{S}_\varepsilon$. The measurement error δ has zero mean and the covariance matrix $S_\delta = \mathcal{E}\{\delta \cdot \delta^{\mathrm{T}}\} = \sigma_\varepsilon^2 I$, where $\mathcal{E}$ is the expected value operator.

A discretization scheme consisting in a collocation method and an approximation technique with basis functions of small support, e.g., B-splines, was used by O'Sullivan [5] and Wahba [18]. In this case the regularization term is a quadratic functional of atmospheric parameters. However, due to the complexity of the forward model we prefer to use the quadrature approach. This strategy allows us to implement various regularization matrices in a simple way.

3. Iteratively regularized Gauss–Newton method for the discrete problem

The goal of our analysis is to find the most featureless state vector x^δ consistent with the data and whatever other deterministic information are available. An estimate x^δ can be found by minimizing the so-called output least squares function

$$\mathcal{F}(x) = \tfrac{1}{2}\left\| F(x) - y^\delta \right\|_2^2 \tag{7}$$

possibly by an iterative method. The Gauss–Newton method for the minimization of (7) leads to the formal iterative solution

$$x_{k+1}^\delta = x_k^\delta - \left(F'(x_k^\delta)^{\mathrm{T}} F'(x_k^\delta)\right)^{-1} F'(x_k^\delta)^{\mathrm{T}} \left(F(x_k^\delta) - y^\delta\right), \tag{8}$$

where $F'(x) \in R^{m\times n}$ denotes the Jacobian matrix evaluated at x.

In mathematical models for inverse problems, where $F : D(F) \subseteq X \to Y$ is a smoothing operator between the Hilbert spaces X and Y, the generalized inverse of F'^*F' (where F'^* denotes the adjoint of F') is usually unbounded, so that each iteration would be unstable even if it were well defined. Hence, the term F'^*F' has to be replaced by an operator with a bounded inverse to ensure that each iteration step is well defined. In fact, a priori information like measures of solution magnitude and smoothness can be incorporated in order to stabilize the iterative process. Due to the inherent instability of ill-posed problems, an iteration method has to be stopped appropriately to guarantee stability of the iterates. These requirements can be achieved by the iteratively regularized Gauss–Newton method. In the discrete case this method uses the stabilization term

$$\left(\alpha_k L^{\mathrm{T}} L + F'(x_k^\delta)^{\mathrm{T}} F'(x_k^\delta)\right)^{-1} \alpha_k L^{\mathrm{T}} L\left(x_k^\delta - x_{\mathrm{a}}\right)$$

leading to the iterative solution

$$x_{k+1}^\delta = x_k^\delta - \left(\alpha_k L^{\mathrm{T}} L + F'(x_k^\delta)^{\mathrm{T}} F'(x_k^\delta)\right)^{-1}\left[F'(x_k^\delta)^{\mathrm{T}}\left(F(x_k^\delta) - y^\delta\right) + \alpha_k L^{\mathrm{T}} L(x_k^\delta - x_{\mathrm{a}})\right], \tag{9}$$

where L is some regularization matrix, (α_k) is a monotonically decreasing sequence satisfying

$$\alpha_k > 0, \qquad 1 \leqslant \frac{\alpha_k}{\alpha_{k+1}} \leqslant c, \tag{10}$$

and x_{a} is the a priori state vector, the best beforehand estimate of $\hat{x}$. Note that x_{k+1}^δ has the variational characterization

$$\mathcal{F}_k^l(x) = \left\| F(x_k^\delta) - y^\delta + F'(x_k^\delta)(x - x_k^\delta)\right\|_2^2 + \alpha_k \left\| L(x - x_{\mathrm{a}})\right\|_2^2. \tag{11}$$

The a priori state vector is assumed to satisfy a source condition of the form $x_{\mathrm{a}} - \hat{x} = f(F'(\hat{x})^{\mathrm{T}} F'(\hat{x}))q$, where $\|q\|_2$ is sufficiently small and f is a continuous function satisfying $f(0) = 0$. Bakushinskii [8] proved convergence of the iteratively regularized Gauss–Newton method under source conditions with f of the form $f(t) = t$. This approach was extended by Blaschke et al. [9] to source condition of Hölder-type $f(t) = t^\nu$, $\nu \geqslant 0$, and by Hohage [10] to logarithmic source condition of the form $f(t) = [\ln(\mathrm{e}t_0/t)]^{-p}$ with $0 < t \leqslant t_0$ and $p > 0$. Moreover, they established convergence rates using the discrepancy principle as an a posteriori stopping rule, i.e. the iteration is stopped at the first index $k_* = k_*(\Delta)$ for which

$$\left\| F(x_{k_*}^\delta) - y^\delta \right\|_2 \leqslant \tau\Delta < \left\| F(x_k^\delta) - y^\delta \right\|_2, \quad 0 \leqslant k < k_*, \tag{12}$$

where $\tau > 1$ and Δ is an upper bound for the error, $\|\delta\|_2 \leqslant \Delta$. To prove results about convergence rates Blaschke et al. [9] and Hohage [10] had to assume some nonlinearity conditions on the mapping F. However, Deuflhard et al. [11] had shown that convergence rate results can be obtained without assuming logarithmic or Hölder-type source conditions. In this context the authors proved that under a stronger nonlinearity condition (the Newton–Mysovskii condition) the estimate

$$\left\| x_k^\delta - \hat{x}\right\|_2 = \mathrm{O}\left(g(\alpha_k)\right), \quad 0 \leqslant k \leqslant k_*(\Delta),$$

holds. Here $g : [0, \infty) \to R$ is a monotonically increasing, continuous function satisfying $g(0) = 0$ and some additional closeness conditions. This result together with the asymptotic estimate of the regularization parameter $\alpha_{k_*(\Delta)}$, i.e. $\alpha_{k_*(\Delta)} = \mathrm{O}(\Delta)$ [10], complete the convergence analysis of the iteratively regularized Gauss–Newton method. Although these results were derived for the identity operator, they are valid in the discrete case for an arbitrary regularization matrix (with minor changes of the source and closeness conditions).

4. Selection of regularization matrix and regularization parameters

For practical applications we have to choose the regularization matrix and to construct a monotonically decreasing sequence of regularization parameters. L is typically either the identity matrix ($L = L_0 = I$), a diagonal

weighting matrix or a discrete approximation of a derivative operator. For example, for an equidistant discretization, the matrices

$$L_1 = \begin{bmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{bmatrix} \in R^{(n-1)\times n} \tag{13}$$

and

$$L_2 = \begin{bmatrix} 1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 2 & -1 \end{bmatrix} \in R^{(n-2)\times n} \tag{14}$$

are scaled approximations to the first and second derivative operators $\mathcal{L}_1$ and $\mathcal{L}_2$, where $\mathcal{L}_1 f = f'$ and $\mathcal{L}_2 f = f''$ [19].

If we have some knowledge about the magnitude of the state vector and we want to constraint the solution to be smooth, we can combine several derivative orders (Sobolev norm) and determine the regularization matrix by the Cholesky factorization [20]

$$L^{\mathrm{T}} L = \sum_{k=0}^{2} w_k L_k^{\mathrm{T}} L_k, \tag{15}$$

where $w_k \geqslant 0$ are some weighting factors such that $\sum_{k=0}^{2} w_k = 1$. The weighting factors are chosen in accordance with the peculiarities of the solution. For ill-conditioned linear systems of equations Brezinski et al. [21] recently proposed an automatic choice method for the weighting factors w_k. The regularization matrix can also be constructed by means of statistical information, that is, L can be some approximation of an a priori covariance matrix. For example, in the case of an a priori symmetric and positive definite covariance matrix $S_{\mathrm{a}} = \sigma_{\mathrm{a}}^2 \widehat{S}_{\mathrm{a}}$ with unknown scale factor σ_{a}, we can determine L by using the Cholesky factorization $L^{\mathrm{T}} L = \widehat{S}_{\mathrm{a}}^{-1}$. If the variances $\sigma_{\mathrm{a}} \upsilon_i$ of the a priori profile are known in the sense that υ_i are known (i.e. the variances are known except for a multiplicative constant σ_{a}), we can express the matrix $\widehat{S}_{\mathrm{a}}$ as:

$$\left[\widehat{S}_{\mathrm{a}}\right]_{ij} = \upsilon_i \upsilon_j \exp\left(-\frac{|h_i - h_j|}{l_{\mathrm{cor}}}\right), \quad i, j = 1, \ldots, n, \tag{16}$$

where l_{cor} is a length determining the correlation between the parameters at different altitudes h_i. In the case of an equidistant altitude grid $\Delta h = h_i - h_{i-1}$ and $\upsilon_i = \upsilon$ for all i, we use the expression of $\widehat{S}_{\mathrm{a}}^{-1}$ given by Steck and von Clarmann [22] to represent L as

$$L = \frac{1}{\upsilon\sqrt{1-\exp(-2a)}} \begin{bmatrix} 1 & -\exp(-a) & 0 & \ldots & 0 & 0 \\ 0 & 1 & -\exp(-a) & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 & -\exp(-a) \\ 0 & 0 & 0 & \ldots & 0 & \sqrt{1-\exp(-2a)} \end{bmatrix} \tag{17}$$

where $a = \triangle h / l_{\mathrm{cor}}$. From a deterministic point of view we see that L acts like a smoothing matrix similar to the discrete approximation to the first derivative operator L_1. In fact, for $l_{\mathrm{cor}} \to 0$, $L \to L_0$, while for $l_{\mathrm{cor}} \to \infty$, $L^{\mathrm{T}} L$ behaves like $L_1^{\mathrm{T}} L_1$. Note that if the iterative process is stopped according to the discrepancy principle and we set $\sigma_{\mathrm{a}} = \sigma_\varepsilon / \sqrt{\alpha_{k_*(\Delta)}}$ we are not allowed to interpret $S_{\mathrm{a}} = \sigma_{\mathrm{a}}^2 \widehat{S}_{\mathrm{a}}$ as the a priori covariance matrix. In this case S_{a} will depend on the statistics of the data, and the whole reason for trying to solve the problem from a Bayesian point of view is to have the ability to incorporate probabilistic data-independent information about the solution [23].

We turn now to the problem of regularization parameter selection. The new iterate x_{k+1}^{δ} minimizes the objective function $\mathcal{F}_k^l$, cf. Eq. (11), which may be written as

$$\mathcal{F}_k^l(x) = \left\| F'(x_k^{\delta})x - \left[F'(x_k^{\delta})\hat{x} + \delta + \mathcal{R}(x_k^{\delta},\hat{x})\right]\right\|_2^2 + \alpha_k \left\| L(x - x_{\mathrm{a}})\right\|_2^2, \tag{18}$$

where

$$\mathcal{R}\left(x_k^{\delta},\hat{x}\right) \equiv F(\hat{x}) - F\left(x_k^{\delta}\right) - F'\left(x_k^{\delta}\right)\left(\hat{x} - x_k^{\delta}\right)$$

denotes the Taylor remainder for the linearization around x_k^{δ}. Let us denote by α_k^{LC} the regularization parameter for the linear subproblem chosen by the L-curve criterion [19], and let us assume that the singular values of the matrix $F'(x_k^{\delta})$ do not vary significantly during the iteration process. For growing k-values, $\delta + \mathcal{R}(x_k^{\delta},\hat{x})$ decreases, and consequently α_k^{LC} decreases. This suggests that α_k may be chosen as α_k^{LC}. However, numerical experiments show that a brutal use of α_k^{LC} may lead to oscillating sequences of α_k-values. A heuristic formula leading to a decreasing sequence of (α_k) was proposed by Eriksson [7]:

$$\alpha_k = \beta\alpha_k^{LC} + (1-\beta)\alpha_{k-1}, \quad 0 \leqslant \beta \leqslant 1, \tag{19}$$

where the parameter β can be used to control the decay rate of the sequence (α_k). This regularization parameter choice method will be referred to as the weighted L-curve criterion. Another selection criterion was proposed by Tautenhahn [3,12]:

$$\alpha_k = \frac{\Delta}{\|F(x_k^{\delta}) - y^{\delta}\|_2}\,\alpha_{k-1}. \tag{20}$$

In contrast to criterion (19) this strategy allows the regularization parameters to decrease very fast at the beginning of iterations. This selection criterion will be referred to as the noise level criterion.

5. Iteratively regularized Gauss–Newton algorithm

A numerical robust method for computing the new iterate x_{k+1}^{δ} and the regularization parameter α_k^{LC} relies on the generalized singular value decomposition (GSVD) of the matrix pair $(K,\ L)$, where $K = F'(x_k^{\delta})$ [19]. We recall that if $K \in R^{m\times n}$, $L \in R^{p\times n}$ and $m \geqslant n \geqslant p$, the GSVD of the matrix pair $(K,\ L)$ is given by

$$K = USX^{-1} \quad \text{and} \quad L = VMX^{-1},$$

where $U \in R^{m\times m}$ and $V \in R^{p\times p}$ are unitary matrices, $X \in R^{n\times n}$ is a nonsingular matrix, $S \in R^{m\times n}$ and $M \in R^{p\times n}$ consist of diagonal matrices,

$$S = \begin{bmatrix} \mathrm{diag}(\mu_i) & 0 \\ 0 & I_{n-p} \\ 0 & 0 \end{bmatrix}, \qquad M = \begin{bmatrix} \mathrm{diag}(\eta_i) & 0 \end{bmatrix},$$

$\mathrm{diag}(\mu_i), \mathrm{diag}(\eta_i) \in R^{p\times p}$ and $\gamma_i = \mu_i/\eta_i,\ i = 1, 2, \ldots, p,$ are the generalized singular values. The new iterate is found by first computing the minimizer of the objective function $\mathcal{F}_k^l$, written in compact form as

$$\mathcal{F}_k^l(r) = \left\| \begin{pmatrix} K \\ \sqrt{\alpha_k}L \end{pmatrix} r - \begin{pmatrix} g \\ 0 \end{pmatrix} \right\|_2^2, \tag{21}$$

that is

$$r_{k+1} = \sum_{i=1}^{p} \frac{\gamma_i^2}{\gamma_i^2 + \alpha_k}\,\frac{u_i^{\mathrm{T}} g}{\mu_i}\,x_i + \sum_{i=p+1}^{n} \left(u_i^{\mathrm{T}} g\right)x_i, \tag{22}$$

where $g = K(x_k^\delta - x_a) - (F(x_k^\delta) - y^\delta)$ and u_i and x_i denote the columns of U and X, respectively, $U = [u_1, u_2, \ldots, u_m]$ and $X = [x_1, x_2, \ldots, x_n]$, and then by making a back-transformation $x_{k+1}^\delta = r_{k+1} + x_a$.

The regularization parameter α_k is computed according to the weighted L-curve criterion (19) or the noise level criterion (20). In the first case, the L-curve criterion for the linear subproblem (21) is used to select the regularization parameter α_k^{LC}. We recall that the L-curve is a double logarithmic plot of the residual norm $\|Kr_\alpha - g\|_2$ versus the constraint norm $\|Lr_\alpha\|_2$ with the regularization parameter α as the parameter. This curve has a typically L-shaped corner. According to Hansen [19], the corner of the L-curve, i.e. the point with maximum curvature, appears for regularization parameters close to the optimal parameter α^{LC} that balances the regularization errors and the perturbation errors. Because the residual norm and the constraint norm are given by simple analytical formulas [24] we use Brent's function minimization routine `FMIN` [16] to compute the maximum of the curvature.

The iteratively regularized Gauss–Newton algorithm can be summarized as follows:

0. Initialization: Choose the regularization matrix L. Choose x_a, x_0^δ and α_{-1} (usually, but not necessary, $x_a = x_0^\delta$). Set $k = 0$.
1. Compute $F(x_k^\delta)$, $K = F'(x_k^\delta)$ and the GSVD of $(K,\ L)$.
2. Determine α_k^{LC} and compute α_k according to the weighted L-curve criterion (19) or the noise level criterion (20).
3. Compute the new iterate $x_{k+1}^\delta = r_{k+1} + x_a$ using (22).
4. If the discrepancy principle (12) is fulfilled then stop the iteration, else set $k = k + 1$ and go to step 1.

Since the iteratively regularized Gauss–Newton method is a version of the Tikhonov regularization with variable α, Eriksson [7] used at step 3 a line search algorithm. In this case, step 3 can be reformulated as follows:

3′. Compute the Gauss–Newton direction $p_k = r_{k+1} - (x_k^\delta - x_a)$, determine the step-length a_k so that the objective function

$$\mathcal{F}_k(x) = \tfrac{1}{2}\left[\left\|F(x) - y^\delta\right\|_2^2 + \alpha_k \left\|L(x - x_a)\right\|_2^2\right] \tag{23}$$

is sufficiently reduced and take $x_{k+1}^\delta = x_k^\delta + a_k p_k$.

The discrepancy principle is frequently used for nonlinear ill-posed problems. This criterion requires the knowledge of an upper bound for the error. In practice, this bound can be chosen as the expected value of $\|\delta\|_2$, i.e. $\Delta = \sqrt{\mathcal{E}\{\|\delta\|_2^2\}} = \sigma_\varepsilon\sqrt{m}$. If we do not have information about the noise level, Tikhonov regularization with the generalized cross validation or the nonlinear L-curve criterion can be used. Note that when stability was required, stopping criteria like the convergence of the iterates or of the linearization error were used by Schimpf and Schreier [25] and Gonzáles and Vélez-Reyes [26]. Although the choice of stopping criteria may not be a delicate question as far as stability is concerned, it is essential for computational expense.

6. Numerical simulations

The motivation for these simulations has been twofold: to study the feasibility of temperature retrieval from spectra recorded by the 2.5 THz OH Measurement Airborne Sounder (THOMAS) and to analyze the behaviour of the solution for different regularization matrices and sequences of regularization parameters. THOMAS is a high resolution heterodyne spectrometer measuring the atmospheric thermal emission in the far infrared. The dominant spectral signatures in the observed spectral region are due to the hydroxyl radical (a rotational line triplet at 83.869 cm^{-1}), water vapor (nb. the wing of a strong line at about 84.456 cm^{-1}), and ozone.

Ideally the vertical distribution of OH is the only unknown of the retrieval problem. However, analysis of the spectra recorded during the second CRISTA/MAHRSI validation campaign 1997 (Cryogenic Infrared

Spectrometers and Telescope for the Atmosphere/Middle Atmosphere High Resolution Spectrograph Investigation) [27] has indicated the need to consider H_2O as unknown, too. Although temperature has a significant impact on the spectrum, it is usually assumed to be known (the retrievals reported by Englert et al. [27] used preliminary CRISTA data). Clearly the retrieval of actual temperature profiles could lead to an improved OH analysis.

The synthetic measurement spectrum used in this numerical study largely resembled typical THOMAS observations made aboard the DLR research aircraft FALCON. An observer altitude of 12 km and a pointing angle of 80° from zenith has been assumed. A radiance spectrum (1) of $m = 200$ data points between 83.84 and 83.88 cm^{-1} was simulated using a line-by-line atmospheric radiative transfer code [17]. The exact atmospheric temperature profile $\hat{x}$ as well as pressure, OH, H_2O, and O_3 profiles were taken from the U.S. standard atmosphere. Noise ε described by a Gaussian probability distribution with zero mean and covariance matrix $S_\varepsilon = \sigma_\varepsilon^2 I$ was added to the simulated spectrum. The signal-to-noise ratio, defined as $SNR = \|y^\delta\|_2/(\sigma_\varepsilon\sqrt{m})$, was chosen as 100. Since we are not interested in a statistical analysis we do not perform a Monte Carlo experiment and consider a single observation.

The a priori and initial profile of temperature x_a and x_0^δ, respectively, were assumed to be identical and were chosen as a scaled version of the exact profile, i.e. $x_a = x_0^\delta = 0.85\hat{x}$. A vertical grid with 2.5 km spacing was selected. Thus the unknown temperature profile was represented by a vector of 20 elements in the altitude regime 12.5–60 km.

As indicated by Eqs. (1) and (2) the inverse problem of temperature estimation from an infrared intensity spectrum is nonlinear. Table 1, which serves to indicate the nonlinearity of the problem, shows the one-norm $\|\cdot\|_1$, the Euclidean norm $\|\cdot\|_2$ and the infinity norm $\|\cdot\|_\infty$ of the Jacobian K at two different states, one of which corresponds to the exact solution $\hat{x}$, the other to the initial guess x_0^δ.

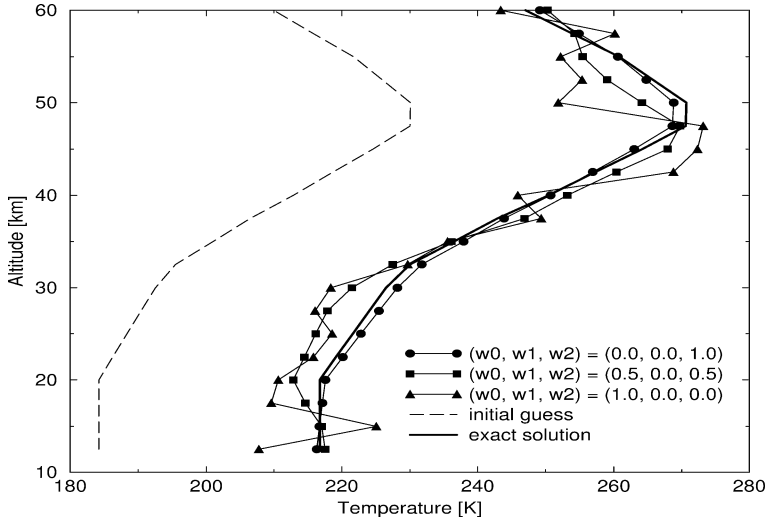
The retrieved profiles for different regularization matrices are shown in Figs. 1 and 2. In Table 2 the corresponding relative errors $\varepsilon = \|\hat{x} - x_{k_*}^\delta\|_2/\|\hat{x}\|_2$ are listed. The sequences of regularization parameters were chosen according to the weighted L-curve criterion (19). The retrieved profiles in Fig. 1a, b show oscillations, that are more pronounced if the contribution of the identity matrix $L_0 = I$ to the global regularization matrix L increases. This behaviour is a consequence of the smoothness of the exact solution, which requires the use of the L_1 or L_2 regularization matrices. The results in Fig. 2 are in agreement with the peculiarities of the regularization matrix described by exponential decay as l_{cor} increases. By increasing the correlation length, the regularization

Table 1
One-norm, Euclidean norm and infinity norm of the Jacobian K at x_0^δ and $\hat{x}$

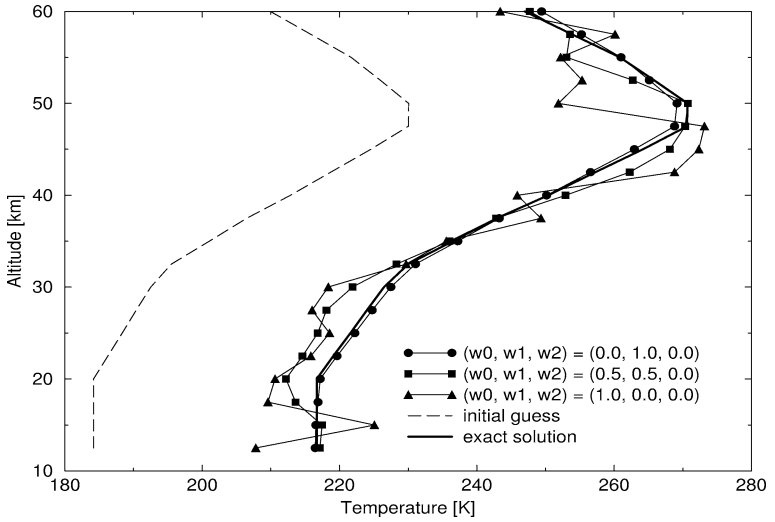
Matrix norms	x_0^δ	$\hat{x}$
$\|K\|_1$	4988.47	4505.90
$\|K\|_2$	420.26	351.15
$\|K\|_\infty$	107.72	57.32

Table 2
Relative errors ϵ for the regularization matrix combining several derivative orders or described by exponential decay, cf. Eqs. (15) and (16)

Regularization matrix	Combination of derivatives $(w_0, w_1, w_2)^T$					Exponential decay l_{cor}		
	1.0	0.5	0.0	0.5	0.0			
	0.0	0.0	0.0	0.5	1.0	5 km	20 km	40 km
	0.0	0.5	1.0	0.0	0.0			
Relative error [%]	3.21	1.64	0.50	1.46	0.41	1.42	0.90	0.23



(a)



(b)

Fig. 1. Results of temperature retrieval for a regularization matrix combining several derivative orders with: (a) $(w_0, w_1, w_2) = (0.0, 0.0, 1.0)$, $(w_0, w_1, w_2) = (0.5, 0.0, 0.5)$, and $(w_0, w_1, w_2) = (1.0, 0.0, 0.0)$, (b) $(w_0, w_1, w_2) = (0.0, 1.0, 0.0)$, $(w_0, w_1, w_2) = (0.5, 0.5, 0.0)$, and $(w_0, w_1, w_2) = (1.0, 0.0, 0.0)$.

matrix behaves like the discrete approximation to the first derivative operator and consequently the relative errors decrease. The best solution was obtained in the case $l_{\text{cor}} = 40$ km.

A technique for improving the inversion performance in the case $L = L_0$ can be given as follows:

1. Let $x_{k_*}^{\delta}$ be the solution obtained by using the discrepancy principle. Smooth the solution by using Tikhonov regularization, that is, determine $x_{\text{smooth}}^{\delta}$ by minimizing the objective function

$$\mathcal{F}(x) = \left\| x - x_{k_*}^{\delta} \right\|_2^2 + \alpha \| L_2 x \|_2^2, \tag{24}$$

where α is chosen according to the L-curve criterion.

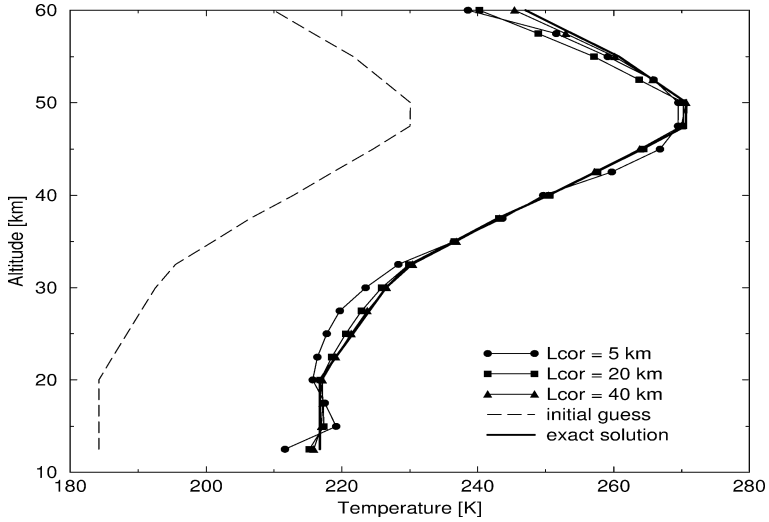


Fig. 2. Result of temperature retrieval for a regularization matrix described by exponential decay with: $l_{\mathrm{cor}} = 5$ km, $l_{\mathrm{cor}} = 20$ km and $l_{\mathrm{cor}} = 40$ km.

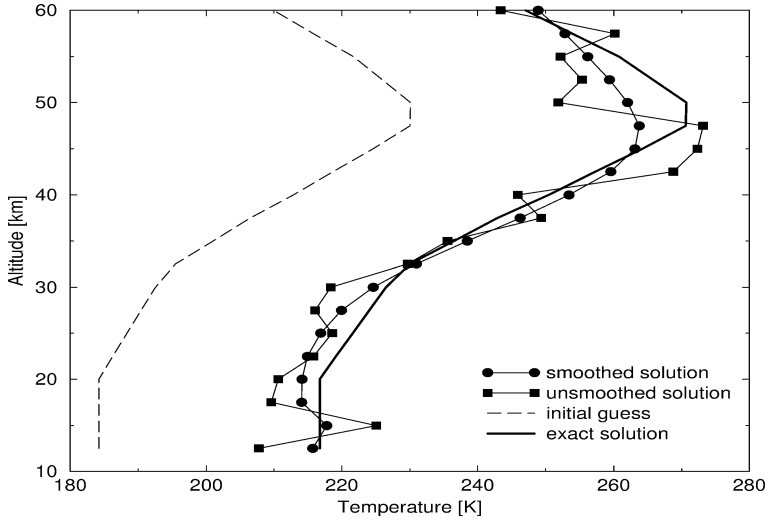


Fig. 3. Result of temperature retrieval using the smoothing procedure.

2. Set $x_0 = x_{\mathrm{a}} = x^{\delta}_{\mathrm{smooth}}$ and restart the iteratively regularized Gauss–Newton algorithm.

In Fig. 3 we show the solution that was obtained by restarting the algorithm. The relative error decreases from 3.21 to 1.53%. Note that this technique can also be used in the Levenberg–Marquardt method, where at each iteration only the magnitude of the solution is controlled.

Next we consider the regularization matrices (15) with $(w_0, w_1, w_2) = (0.0, 0.0, 1.0)$, $(w_0, w_1, w_2) = (0.5, 0.0, 0.5)$, and $(w_0, w_1, w_2) = (1.0, 0.0, 0.0)$ and select the sequence of regularization parameters according to weighted L-curve criterion (19) and noise level criterion (20). For the first selection criterion we choose $\beta = 0.2$, and this choice leads to a slowly decreasing sequence of regularization parameters. The histories of regularization

Table 3
Histories of regularization parameters α_k, relative residuals $\mathcal{F}(x_k^\delta)/(m\sigma_\epsilon^2)$ and relative errors ϵ_k for: (a) $(w_0,\ w_1,\ w_2) = (0.0,\ 0.0,\ 1.0)$, (b) $(w_0,\ w_1, w_2) = (0.5, 0.0, 0.5)$ and (c) $(w_0, w_1, w_2) = (1.0, 0.0, 0.0)$. The columns 2, 3 and 4 correspond to the weighted L-curve criterion, while the columns 5, 6 and 7 correspond to the noise level criterion. The smallest generalized singular values of (K, L) are displayed in column 8

(a)							
k	α_k	$\mathcal{F}(x_k^\delta)/(m\sigma_\varepsilon^2)$	ε_k	α_k	$\mathcal{F}(x_k^\delta)/(m\sigma_\varepsilon^2)$	ε_k	$\gamma_k^{\min}$
0	186.35	211.51	0.15	186.35	211.51	0.15	3.56e−2
1	149.19	3.18	6.75e−2	26.03	3.18	6.75e−2	3.12e−2
2	119.47	1.09	1.43e−2	10.50	1.24	1.62e−2	3.43e−2
3	–	0.45	5.05e−3	–	0.46	7.54e−3	–
(b)							
0	11.56	211.51	0.15	11.56	211.51	0.15	4.91e−2
1	10.19	7.78	6.56e−2	0.74	7.78	6.56e−2	5.07e−2
2	9.67	3.26	2.64e−2	7.02e−2	5.27	0.10	6.03e−2
3	8.58	1.22	1.65e−2	–	negative solution	–	–
4	–	0.45	1.64e−2	–	–	–	–
(c)							
0	8.21	211.51	0.15	8.21	211.51	0.15	0.13
1	7.21	40.68	8.21e−2	0.10	40.68	8.21e−2	0.12
2	6.49	24.61	3.83e−2	–	negative solution	–	–
3	5.87	1.77	2.88e−2	–	–	–	–
4	5.32	0.61	3.02e−2	–	–	–	–
5	–	0.47	3.12e−2	–	–	–	–

parameters α_k, relative residuals $\mathcal{F}(x_k^\delta)/(m\sigma_\varepsilon^2)$ and relative errors ε_k are illustrated in Table 3. The sequences of regularization parameters given by the noise level criterion decrease faster than the sequences given by the weighted L-curve criterion. In the case $(w_0, w_1, w_2) = (0.0, 0.0, 1.0)$ both selection criteria lead to comparable solutions. In the next cases, the noise level criteria leads to negative solutions with large relative errors. It appears that the solution deteriorates significantly when the regularization parameter becomes comparable to the smallest generalized singular values of (K, L). Therefore, we can conclude that allowing the regularization parameter to decrease too fast will result in convergence toward a solution with large relative error with respect to the exact solution. A more suitable strategy is to start with a rather large α and then to decrease α gradually in a safe way.

7. Conclusions

An iteratively regularized Gauss–Newton algorithm for atmospheric retrieval is presented. At each iteration the new iterate and the regularization parameter are computed by using the generalized singular value decomposition of the Jacobian matrix and the regularization matrix.

Our analysis is focused on the choice of the regularization matrix and the sequence of regularization parameters. The combination of several derivative orders into a single regularization matrix as well as the use of a regularization matrix with an exponential decay gives more flexibility to the algorithm and offers the possibility to control simultaneously the magnitude and the smoothness of the solution. Our numerical simulations demonstrate that a regularization matrix with an exponential decay is suitable for temperature retrieval, at least for the particular example considered in Section 6. An appropriate strategy for choosing the regularization parameter relies on the use of the L-curve criterion for each linear subproblem. This technique allows enough regularization to be applied at the beginning of iterations and then to be gradually decreased. The numerical examples indicate that

an algorithm based on the weighted L-curve criterion is more efficient than one based on the noise level criterion, even though its more time consuming. Note that the forward calculations are the most time-consuming part of the retrieval.

A technique for improving the accuracy of the solution in the case $L = I$ is also presented. This method consists in a smoothing procedure based on Tikhonov regularization and a restarted iteratively regularized Gauss–Newton algorithm.

In summary, the iteratively regularized Gauss–Newton method using a regularization matrix with an exponential decay and a parameter choice strategy based on the weighted L-curve criterion clearly emerges as a valuable technique for atmospheric profile retrieval.

Acknowledgement

The authors would like to thank Dr. Birger Schimpf for numerous valuable discussions and suggestions during the preparation of this work.

References

[1] C.D. Rodgers, Inverse Methods for Atmospheric Sounding: Theory and Practise, World Scientific, Singapore, 2000.
[2] H.W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, NL, 1996.
[3] U. Tautenhahn, Error estimates for regularized solutions of nonlinear ill-posed problems, Inverse Problems 10 (1994) 485–500.
[4] A. Neubauer, Tikhonov regularisation for non-linear ill-posed problems: Optimal convergence rates and finite-dimensional approximation, Inverse Problems 5 (1989) 541–557.
[5] F. O'Sullivan, G. Wahba, A cross validated Bayesian retrieval algorithm for nonlinear remote sensing experiments, J. Comput. Phys. 59 (1985) 441–455.
[6] U. Tautenhahn, On a general regularization scheme for nonlinear ill-posed problems, Inverse Problems 13 (1997) 1427–1437.
[7] J. Eriksson, Optimization and regularization of nonlinear least squares problems. PhD thesis, Department of Computing Science, Umea University, Sweden, 1996.
[8] A.B. Bakushinskii, The problem of the convergence of the iteratively regularized Gauss–Newton method, Comput. Math. Phys. 32 (1992) 353–1359.
[9] B. Blaschke, A. Neubauer, O. Scherzer, On convergence rates for the iteratively regularized Gauss–Newton method, IMA J. Numer. Anal. 17 (1997) 421–436.
[10] T. Hohage, Logarithmic convergence rates of the iteratively regularized Gauss–Newton method for an inverse potential and an inverse scattering problem, Inverse Problems 13 (1997) 1279–1300.
[11] P. Deuflhard, H.W. Engl, O. Scherzer, A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions, Inverse Problems 14 (1998) 1081–1106.
[12] U. Tautenhahn, Numerische Vergleiche zwischen Tikhonovscher und stochastischer Regularisierung nichtlinearer Systeme am Beispiel der Auswertung von Satellitenmessdaten, Beitr. Num. Math. 11 (1983) 161–171.
[13] K.-N. Liou, An Introduction to Atmospheric Radiation, Academic Press, Orlando, 1980.
[14] G.P. Anderson, S.A. Clough, F.X. Kneizys, J.H. Chetwynd, E.P. Shettle, AFGL atmospheric constituent profiles (0–120 km), Technical Report TR-86-0110, AFGL, 1986.
[15] SLATEC, Common Mathematical Library (Version 4.1), Available from NetLib http://www.netlib.org/, July 1993.
[16] D. Kahaner, C. Moler, S. Nash, Numerical Methods and Software, Prentice-Hall, Englewood Cliffs, NJ, 1989.
[17] F. Schreier, B. Schimpf, A new efficient line-by-line code for high resolution atmospheric radiation computations incl. derivatives, in: W.L. Smith, Y. Timofeyev (Eds.), IRS 2000: Current Problems in Atmospheric Radiation, A. Deepak, 2001, pp. 381–384.
[18] G. Wahba, Spline Models for Observational Data, SIAM, Philadelphia, PA, 1990.
[19] P.C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion, SIAM, Philadelphia, PA, 1998.
[20] P. Eriksson, Analysis and comparison of two linear regularization methods for passive atmospheric observation, J. Geophys. Res. 105 (2000) 18 157–18 167.
[21] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, S. Seatzu, Multi-parameter regularization technique for ill-conditioned linear systems, in: Applied Inverse Problems: Theoretical and Computational Aspects, Montecatini Terme, Italy, June 2001.
[22] T. Steck, T. von Clarmann, Constrained profile retrieval applied to the observation mode of the Michelson interferometer for passive atmospheric sounding, Appl. Opt. 40 (2001) 3559–3571.

[23] W.P. Gouveia, J.A. Scales, Resolution of seismic waveform inversion: Bayes versus Occam, Inverse Problems 13 (1997) 323–349.
[24] C.R. Vogel, Non-convergence of the L-curve regularization parameter selection method, Inverse Problems 12 (1996) 535–547.
[25] B. Schimpf, F. Schreier, Robust and efficient inversion of vertical sounding atmospheric high-resolution spectra by means of regularization, J. Geophys. Res. 102 (1997) 16 037–16 055.
[26] F.O. Gonzáles, M. Vélez-Reyes, Algorithms for nonlinear retrieval problems in atmospheric remote sensing using regularization methods, in: J.E. Russell (Ed.), Proc. SPIE, in: Satellite Remote Sensing of Clouds and Aerosols III, Vol. 3495, 1998, pp. 110–121.
[27] C. Englert, B. Schimpf, M. Birk, F. Schreier, M. Krocka, R.G. Nitsche, R.U. Titz, M.E. Summers, The 2.5 THz heterodyne spectrometer THOMAS: Measurement of OH in the middle atmosphere and comparison with photochemical model results, J. Geophys. Res. 105 (2000) 22 211–22 223.