Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

Constrained regularization methods for ozone profile retrieval from UV/VIS nadir spectrometers

Adrian Doicu *, Olena Schüssler, Diego Loyola

Remote Sensing Technology Institute, German Aerospace Centre, Oberpfaffenhofen, Wessling, Germany

A B S T R A C T

In this paper we present several constrained regularization methods for ozone profile retrieval from UV/VIS nadir sounding instruments such as GOME, SCIAMACHY, OMI and GOME-2. These methods extend the Tikhonov regularization and the iteratively regularized Gauss–Newton method with equality and inequality constraints imposed on the vertical column. It will be shown that this type of information, which is delivered by an independent algorithm like DOAS or GODFIT, significantly improves the accuracy and stability of the profile retrieval.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Modern space-born UltraViolet and VISible (UV/VIS) nadir atmospheric chemistry instruments like GOME/ERS-2, SCIAMACHY/ENVISAT, OMI/AURA and GOME-2/MetOp measure the backscattered radiation from the Earth-atmosphere system with a relatively high spectral resolution. These measurements are used for the routine retrieval of atmospheric composition quantities such as total column amounts of ozone, NO_2, SO_2, BrO, H_2O, HCHO and OClO. Additionally, it is also possible to derive height-resolved information about the atmospheric ozone concentration. The strongly decreasing ozone absorption cross section in the 270–340 nm wavelength range provides information on the vertical distribution of ozone. Shorter wavelengths in this range contain only information on the ozone distribution for the top layers of the atmosphere, whereas the larger wavelengths also carry ozone information from the lower layers of the atmosphere up to the ground surface. A number of ozone profile retrieval algorithms from GOME measurements have been developed during the last years. The inversion

in the current algorithms is typically performed using the Bayesian approach implemented via optimal estimation [1–4] or Tikhonov regularization [5].

The inverse problems under examination is nonlinear and can be expressed by the data model:

$$\mathbf{y}^\delta = \mathbf{F}(\mathbf{x}) + \boldsymbol{\delta}, \tag{1}$$

where $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^m$ is the forward model, $\mathbf{x} \in \mathbb{R}^n$ is the state vector, $\mathbf{y}^\delta \in \mathbb{R}^m$ is the noisy data vector, and $\boldsymbol{\delta} \in \mathbb{R}^m$ is the measurement error. In the framework of the Bayesian approach, a regularized solution of (1) is computed as the minimizer of the a posteriori potential [6]

$$V(\mathbf{x}|\mathbf{y}^\delta) = [\mathbf{y}^\delta - \mathbf{F}(\mathbf{x})]^T \mathbf{C}_\delta^{-1} [\mathbf{y}^\delta - \mathbf{F}(\mathbf{x})] + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{C}_x^{-1} (\mathbf{x} - \mathbf{x}_a),$$

where $\mathbf{C}_\delta$ and $\mathbf{C}_x$ are the measurement noise and the a priori covariance matrices, respectively, and $\mathbf{x}_a$ is the a priori state vector. Often, due to a lack of a priori information, we use an ad hoc covariance matrix of the form $\mathbf{C}_x = \sigma_x^2 \mathbf{C}_{nx}$, where $\mathbf{C}_{nx}$ is chosen as a Gaussian correlation matrix and the profile variance σ_x^2 is considered to be a free parameter of the retrieval. If our a priori knowledge is encapsulated in the form of an ellipsoid of constant probability of the a priori, then the shape and orientation of this ellipsoid are determined by $\mathbf{C}_{nx}$, while its size is determined by σ_x^2. In statistical

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

inversion theory, the Bayesian approach is analog to the method of Tikhonov from classical regularization theory [7–10]. To put in evidence this similarity we assume that the noise is white with the covariance matrix $\mathbf{C}_\delta = \sigma^2 \mathbf{I}_m$ and compute the Cholesky factor of the inverse of the normalized covariance matrix $\mathbf{C}_{n\mathbf{x}}^{-1} = \mathbf{L}^T \mathbf{L}$, where σ^2 is the noise variance and $\mathbf{L}$ is the regularization matrix. In this context, the Bayesian solution minimizes simultaneously the a posteriori potential and the Tikhonov function

$$\mathcal{F}(\mathbf{x}) = \sigma^2 V(\mathbf{x}|\mathbf{y}^\delta) = \|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x})\|^2 + \alpha \|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|^2,$$

and the regularization parameter $\alpha = \sigma^2 / \sigma_x^2$ can be interpreted as the noise-to-signal ratio. In view of the above equivalence, the Bayesian approach is often referred to as a stochastic version of Tikhonov regularization.

As the regularization matrix reproduces the shape of the Gaussian correlation function, $\mathbf{L}$ controls the smoothness of the profile as the discrete approximations to the first-order ($\mathbf{L}_1$) and the second-order ($\mathbf{L}_2$) derivative operators do. Note that if we intend to control the magnitude of the solution, then $\mathbf{L}$ should be chosen as either the identity matrix ($\mathbf{L}_0 = \mathbf{I}_n$) or a diagonal matrix.

The computation of an appropriate estimate of α, or equivalently, the derivation of a reliable estimator of σ_x^2, is a crucial issue of Tikhonov regularization (see, e.g. [11,12]). With too little regularization, reconstructions deviate significantly from the a priori and the solution is said to be underregularized. With too much regularization, the reconstructions are too close to the a priori and the solution is said to be overregularized. Unfortunately, at the present time, there is no fail-safe regularization parameter choice method which guarantees small solution errors in any circumstance, that is, for any noisy data vector. An amelioration of the problems associated with the regularization parameter selection is achieved in the framework of the so-called iterative regularization methods, of which the iteratively regularized Gauss–Newton method is a relevant example [13–15]. These approaches are less sensitive to overestimations of the regularization parameter, but for underestimations of the regularization parameter, the reconstruction errors of iterative methods and Tikhonov regularization are comparably large and no substantial solution improvement can be achieved.

The goal of this paper is to design a class of regularization methods for ozone profile retrieval which are less susceptible to the selection of the regularization parameter over a large range of values. The idea is to incorporate additional constraints into the classical regularization method which should come into play especially when the regularization parameter is too small and large deviations from the a priori profile are expected. The additional constraints are imposed on the integrated ozone profile, i.e., the vertical column, which can be computed with sufficiently accuracy by an independent algorithm like DOAS [16,17] or GODFIT [18]. Essentially, in this class of regularization methods, we control the smoothness of the profile through the regularization matrix and the magnitude of the profile through the vertical column.

2. Constrained regularization methods

In this section we review the algorithmic implementations of Tikhonov regularization and of the iteratively regularized Gauss–Newton method and describe the modifications required to design the constrained versions of these methods. In order to simplify our presentation we assume that we are dealing with the retrieval of one gas component, i.e., ozone, and that the entry x_i of $\mathbf{x}$ is the partial column on the layer i. The number of layers is n and the vertical column is then given by $\sum_{i=1}^n x_i$. The layer $i = 1$ is situated at the top of the atmosphere, while the layer $i = n$ is situated at the Earth's surface.

2.1. Constrained Tikhonov regularization

The solution of a nonlinear ill-posed problem by means of Tikhonov regularization is equivalent to the solution of a sequence of ill-posed linearizations of the forward model about the current iterate. At the iteration step k, we consider the linearized equation

$$\mathbf{K}_{k\alpha}(\mathbf{x} - \mathbf{x}_a) = \mathbf{y}_k^\delta, \tag{2}$$

where $\mathbf{K}_{k\alpha}$ is the Jacobian matrix at the actual iterate $\mathbf{x}_{k\alpha}^\delta$ and $\mathbf{y}_k^\delta$ is the "linearized" noisy data vector,

$$\mathbf{y}_k^\delta = \mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_{k\alpha}^\delta) + \mathbf{K}_{k\alpha}(\mathbf{x}_{k\alpha}^\delta - \mathbf{x}_a).$$

As the nonlinear problem is ill-posed, its linearization is also ill-posed, and we solve the linearized Eq. (2) by means of Tikhonov regularization with the penalty term $\Omega(\mathbf{x}) = \|\mathbf{L}(\mathbf{x} - \mathbf{x}_a)\|$ and the regularization parameter α. In terms of the profile deviation $\Delta \mathbf{x} = \mathbf{x} - \mathbf{x}_a$, the computational step involves the minimization of the quadratic function

$$\mathcal{Q}(\Delta \mathbf{x}) = \mathbf{g}^T \Delta \mathbf{x} + \tfrac{1}{2} \Delta \mathbf{x}^T \mathbf{G} \Delta \mathbf{x}, \tag{3}$$

with $\mathbf{G} = \mathbf{K}_{k\alpha}^T \mathbf{K}_{k\alpha} + \alpha \mathbf{L}^T \mathbf{L}$ and $\mathbf{g} = -\mathbf{K}_{k\alpha}^T \mathbf{y}_k^\delta$. The new iterate is then computed from $\Delta \mathbf{x}_{k+1\alpha}^\delta = \mathbf{x}_{k+1\alpha}^\delta - \mathbf{x}_a$, where the minimizer $\Delta \mathbf{x}_{k+1\alpha}^\delta$ is given by $\Delta \mathbf{x}_{k+1\alpha}^\delta = -\mathbf{G}^{-1} \mathbf{g}$. A practical realization of the method of Tikhonov regularization is outlined in the Appendix.

2.1.1. Tikhonov regularization with equality constraints

In the equality-constrained version of Tikhonov regularization we focus on the minimization step of the quadratic function (3) and formulate an optimization problem with a linear equality constraint: At the iteration step k, we compute the profile deviation $\Delta \mathbf{x}_{k+1\alpha}^\delta$ by solving the quadratic programming problem

$$\min_{\Delta \mathbf{x}} \quad \mathcal{Q}(\Delta \mathbf{x}) = \mathbf{g}^T \Delta \mathbf{x} + \tfrac{1}{2} \Delta \mathbf{x}^T \mathbf{G} \Delta \mathbf{x}, \tag{4}$$

$$\text{subject to} \quad \sum_{i=1}^n \Delta x_i = c. \tag{5}$$

Here, c is the difference between the vertical columns corresponding to the "exact" state vector and the a priori. By convention, c will be called the relative vertical column (with respect to the a priori), and for the time being, c is assumed to be known.

For solving the quadratic programming problem (4)–(5) the null-space or the range-space methods can be employed [19,20]. In the framework of the null-space method, the matrix $\mathbf{Z} \in \mathbb{R}^{n \times (n-1)}$, whose columns are a basis for the null space of the constraint matrix $\mathbf{A} = [1, 1, \ldots, 1]$ (cf. (5)), plays a significant role. In general, the matrix $\mathbf{Z}$ can be derived by using the QR factorization of $\mathbf{A}^T$, or it can be computed by using the variable-reduction technique [19]. In the present analysis we adopt the variable-reduction technique, in which case, the algorithm involves the following steps:

(1) compute a feasible point satisfying the linear constraint, e.g.,

$$\Delta \overline{\mathbf{x}} = c \Delta \overline{\mathbf{x}}_n \quad \text{with } \Delta \overline{\mathbf{x}}_n = \frac{1}{n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix};$$

(2) compute the gradient of $\mathcal{Q}$ at $\Delta \overline{\mathbf{x}}$,

$$\overline{\mathbf{g}} = c \mathbf{g}_n + \mathbf{g} \quad \text{with } \mathbf{g}_n = \mathbf{G} \Delta \overline{\mathbf{x}}_n,$$

and construct the matrix $\mathbf{Z} \in \mathbb{R}^{n \times (n-1)}$ as

$$\mathbf{Z} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \\ -1 & -1 & \cdots & -1 \end{bmatrix};$$

(3) determine the feasible step

$$\mathbf{p} = -\mathbf{H}\overline{\mathbf{g}} = -c\mathbf{H}\mathbf{g}_n - \mathbf{H}\mathbf{g},$$

where

$$\mathbf{H} = \mathbf{Z}(\mathbf{Z}^T \mathbf{G} \mathbf{Z})^{-1} \mathbf{Z}^T$$

is the reduced inverse Hessian of $\mathcal{Q}$ subject to the constraint;
(4) compute the solution of the minimization problem as

$$\Delta \mathbf{x}_{k+1\alpha}^{\delta}(c) = \Delta \overline{\mathbf{x}} + \mathbf{p} = c(\Delta \overline{\mathbf{x}}_n - \mathbf{H}\mathbf{g}_n) - \mathbf{H}\mathbf{g}. \tag{6}$$

The above solution representation explicitly indicates the dependency on the relative vertical column, and this representation is of beneficial use in practice. The reason is that c is considered as a free parameter of the retrieval ranging in a chosen interval $[c_{min}, c_{max}]$. The problem to be solved is the computation of the strengths of the constraints, or more precisely, of the regularization parameter, which controls the smoothness of the solution, and of the relative vertical column, which controls the magnitude of the solution. Essentially, we must solve a multi-parameter regularization problem. In this case we adopt a simple strategy: we use an a priori chosen regularization parameter but compute the relative vertical column by using an approach which is similar to the L-curve method [21]. Note that the L-curve is the plot of the constraint norm against the residual norm for a range of values of the regularization parameter, and the corner of the L-curve reflects a trade-off between solution smoothness and residual. Two regularization methods

with a dynamical selection criterion for the vertical column can be designed:

(1) *Equality-constrained Tikhonov regularization with an outer loop.* The numerical realization of this method requires to solve the nonlinear constrained minimization problem several times for different values of the vertical column. Each minimization is solved with a relative vertical column c and a solution $\mathbf{x}_{\alpha}^{\delta}(c)$ is obtained. Then, the nonlinear residual

$$R_{\delta}(c) = \|\mathbf{F}(\mathbf{x}_{\alpha}^{\delta}(c)) - \mathbf{y}^{\delta}\|$$

and the constraint norm

$$C_{\delta}(c) = \|\mathbf{L}[\mathbf{x}_{\alpha}^{\delta}(c) - \mathbf{x}_a]\|$$

are computed at the solution. The optimal value of the relative total column is defined as the minimizer of the distance function [22]

$$d(c)^2 = \left[\frac{R_{\delta}(c)}{R_{\delta \max}}\right]^2 + \left[\frac{C_{\delta}(c)}{C_{\delta \max}}\right]^2 \tag{7}$$

where $R_{\delta \max} = \max_c R_{\delta}(c)$, $C_{\delta \max} = \max_c C_{\delta}(c)$ and $c \in [c_{min}, c_{max}]$.
(2) *Equality-constrained Tikhonov regularization with an inner loop.* At each iteration step k, we compute $\Delta \mathbf{x}_{k+1\alpha}^{\delta}(c)$ for a fixed value of c, and evaluate the residual and the constraint norms of the linearized equation

$$R_{\delta}(c) = \|\mathbf{y}_k^{\delta} - \mathbf{K}_{k\alpha} \Delta \mathbf{x}_{k+1\alpha}^{\delta}(c)\|$$

and

$$C_{\delta}(c) = \|\mathbf{L} \Delta \mathbf{x}_{k+1\alpha}^{\delta}(c)\|,$$

respectively. As before, the optimal value of the total column is the minimizer of the distance function (7) over the interval $[c_{min}, c_{max}]$.

Noting that the minimization of the distance function is usually performed by using a discrete search algorithm it is readily seen that the first solution method is more time consuming than the second one. By virtue of (6), the computation of $\Delta \mathbf{x}_{k+1\alpha}^{\delta}$ involves only a scalar-vector multiplication and the summation of two vectors, and as a result, the computational effort of the equality-constrained Tikhonov regularization with an inner loop is not much higher than that of the ordinary method.

The interval of variation of the relative vertical column should be chosen so that the distance function has a minimum for the assumed values of the signal-to-noise ratio. To get some idea of where this interval lies, we may use as guide the value of the total column delivered by an independent retrieval. Evidently, this additional information is for reference only.

2.1.2. Tikhonov regularization with inequality constraints

An inequality-constrained version of Tikhonov regularization can be derived if the total column is known with sufficiently accuracy. The information on the total column should be the result of an independent retrieval, which can be performed in a distinct spectral interval by using appropriate algorithms like DOAS or GODFIT. As opposed to the equality-constrained method, the vertical

column is not a free parameter of the retrieval which is optimized by using an internal selection criterion. The proposed inequality-constrained Tikhonov regularization is of the form of the following model algorithm: At the iteration step k, we compute the profile deviation $\Delta \mathbf{x}_{k+1\alpha}^{\delta}$ by solving the quadratic programming problem

$$\min_{\Delta \mathbf{x}} \quad \mathcal{Q}(\Delta \mathbf{x}) = \mathbf{g}^T \Delta \mathbf{x} + \tfrac{1}{2} \Delta \mathbf{x}^T \mathbf{G} \Delta \mathbf{x}, \tag{8}$$

$$\text{subject to} \quad \sum_{i=1}^{n_t} \Delta x_i \leq c_{\max}, \tag{9}$$

$$\sum_{i=1}^{n} \Delta x_i \geq c_{\min}. \tag{10}$$

Here, the layer $n_t < n$, delimitates the tropospheric from the stratospheric region, while the quantities $c_{\min}$ and $c_{\max}$ represent a lower and an upper bound imposed on the vertical column (for a practical selection of $c_{\min}$ and $c_{\max}$ we refer to Section 3). The reasons for the choice (9)–(10) are the following:

(1) In general, the constraints should be linearly independent since otherwise one of the constraints can be omitted without altering the solution.
(2) As the nadir radiance is less sensitive to profile variations in the troposphere, the condition (9) does not allow large profile deviations in the stratospheric region.
(3) The condition (10) guarantees a sufficiently large deviation of the profile (with respect to the a priori) over the entire altitude range.

The quadratic programming problem (8)–(10) can be solved by using primal and dual active set methods. The dual active set method of Goldfarb and Idnani [23] generates dual-feasible iterates by keeping track of an active set of constraints. The method does not have the possibility of cycling and benefits from always having an easily calculated feasible starting point. An implementation of the method of Goldfarb and Idnani is the routine "solve.qp" from the optimization package "quadprog" [24]. Since this routine is free available through internet we use it as a black box and refer to [23] for an algorithm description.

2.2. Constrained iteratively regularized Gauss–Newton method

The iteratively regularized Gauss–Newton method can be regarded as modified version of Tikhonov regularization with a variable regularization parameter. The regularization parameters can be chosen in advanced as the terms of a decreasing sequence, i.e.,

$$0 < \alpha_{k+1} \leq \alpha_k, \quad \lim_{k \to \infty} \alpha_k = 0. \tag{11}$$

Another important difference is that the iterative process is stopped accordingly to the discrepancy principle, that is, the iterative process terminates after $k^\star$ steps such that

$$\|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_{k^\star}^\delta)\|^2 \leq \tau \varDelta^2 < \|\mathbf{y}^\delta - \mathbf{F}(\mathbf{x}_k^\delta)\|^2, \quad 0 \leq k < k^\star, \tag{12}$$

where $\varDelta$ is the noise level and $\tau > 1$ is chosen sufficiently large. In a semi-stochastic setting and for white noise with variance σ^2, an appropriate estimate of $\varDelta^2$ is the expected value of the noise $\mathcal{E}\{\|\delta\|^2\} = m\sigma^2$.

Although several strategies for selecting the sequence of parameters $\{\alpha_k\}$ can be considered, the simplest and the most robust parameter choice rule assumes that the α_k are the terms of a geometric sequence, i.e., $\alpha_k = q\alpha_{k-1}$ with $0 < q < 1$ [25]. The reduction of the regularization parameter leads to acceptable reconstruction errors especially when a strong regularization is applied at the beginning of the iterative process. In this case, the limiting values of the sequences of the regularization parameters are comparable whatever the initial values of the regularization parameters are. Furthermore, the method of Tikhonov regularization using these limiting values as a priori regularization parameters yields small solution errors. The only drawback of the iteratively regularized Gauss–Newton method is that an increase of the initial regularization strength is accompanied by an increase of the number of iterations.

Any iterative method using the discrepancy principle as stopping rule requires the knowledge of the noise level. Because in many practical problems arising in atmospheric remote sensing, the noise level is an unknown quantity (due to the systematic errors), we use the following stopping rule: We store all iterates $\mathbf{x}_k^\delta$ and require the convergence of the nonlinear residuals $\mathbf{r}_k^\delta$ within a prescribed tolerance. If $\mathbf{r}^\delta$ is the residual at the last iterate, we choose that solution $\mathbf{x}_{k^\star}^\delta$ for which it holds that

$$\|\mathbf{r}_{k^\star}^\delta\|^2 \leq \tau \|\mathbf{r}^\delta\|^2 < \|\mathbf{r}_k^\delta\|^2, \quad 0 \leq k < k^\star.$$

The above heuristic stopping rule does not have a mathematical justification, but works sufficiently well in practice. To our knowledge there is no mathematical literature dealing with this topic, and for the time being, we do not see another viable alternative for practical implementations. An algorithmic implementation of the iteratively regularized Gauss–Newton method is illustrated in the Appendix.

As in the case of Tikhonov regularization, the constrained versions of the iteratively regularized Gauss–Newton method assume the replacement of the unconstrained minimization step by a quadratic programming problem involving the equality constraint (5) or the inequality constraints (9) and (10). Since in general, iterative methods requires more Newton steps than Tikhonov regularization, only the equality-constrained iteratively regularized Gauss–Newton method with an inner loop is appropriate for practical applications. It should be pointed out that for the equality-constrained method, both strengths of the constraints are computed internally: the regularization parameter, which controls the smoothness of the solution, is decreased during the Newton iteration by a constant factor, and the vertical column, which controls the magnitude of the solution, is determined by using the minimum distance function approach.

3. Numerical simulations

To analyze the performances of the constrained regularization methods we consider an ozone retrieval test problem.

The ozone profile is retrieved from nadir simulated data by considering 375 equidistant points in the spectral interval ranging from 290 to 335 nm. In this spectral interval, O_3 and NO_2 are considered as active gases. The atmosphere is discretized with a step of 3.5 km between 0 and 70 km, and a step of 10 km between 70 and 100 km. The total numbers of levels is then 24, while the number of layers is 23. The exact state vector is chosen as a translated and a scaled version of a climatological profile with a translation distance in the vertical direction of 3 km and a scaling factor of 1.3. The exact relative vertical column for ozone is $\bar{c} = 120$ DU, and the bounds c_{min} and c_{max} are chosen as follows:

(1) For equality constraints, we set $c_{min} = 80$ DU and $c_{max} = 125$ DU, and 80 values of the relative vertical column are considered in the interval $[c_{min}, c_{max}]$. The reason for choosing this large interval of variation is that we have to guarantee that the distance function has a minimum for any value of the regularization parameter and for low values of the SNR.
(2) For inequality constraints, we choose $c_{max} = \bar{c} = 120$ DU and $c_{min} = 105$ DU. The choice of the upper bound is reasonable since c_{max} controls only the vertical column above the troposphere. Note that the lower bound corresponds to a relative error of $\varepsilon_c = 12.5\%$, where $c_{min} = \bar{c}(1 - \varepsilon_c)$.

The solar zenith angle is 40°, while the zenith and azimuthal angles of the line of sight are 20° and 90°, respectively. The regularization matrix is the Cholesky factor of the normalized covariance matrix

$$[\mathbf{C}_{nx}]_{ij} = [\mathbf{x}_a]_i [\mathbf{x}_a]_j \exp\left(-\frac{|z_i - z_j|}{l}\right), \quad i, j = 1, \dots, n,$$

with an altitude-independent correlation length $l = 3.5$ km.

In Fig. 1 we plot the solution errors for Tikhonov regularization and its equality-constrained versions. The regularization parameter is chosen as $\alpha = \sigma^p$, where σ is the noise standard deviation and p is a positive exponent. The results show that for large values of p (small values of the regularization parameter), the solution errors for the constrained methods are smaller than the solution errors for the unconstrained method, while for small values of p, the solution errors are comparable. Thus, the equality constraint comes into effect for underestimations of the regularization parameter.

The normalized constraint norm, residual norm and distance function are illustrated in Fig. 2. The dependency of these quantities on the relative vertical column is similar to their dependency on the regularization parameter: For small values of the relative vertical column, the profiles may have oscillatory artifacts around the a priori, so that the mean profiles are essentially close to the a priori. Consequently, the constraint norm is large, while the residual norm is small. However, in contrast to the regularization parameter dependency, the constraint norm is not a
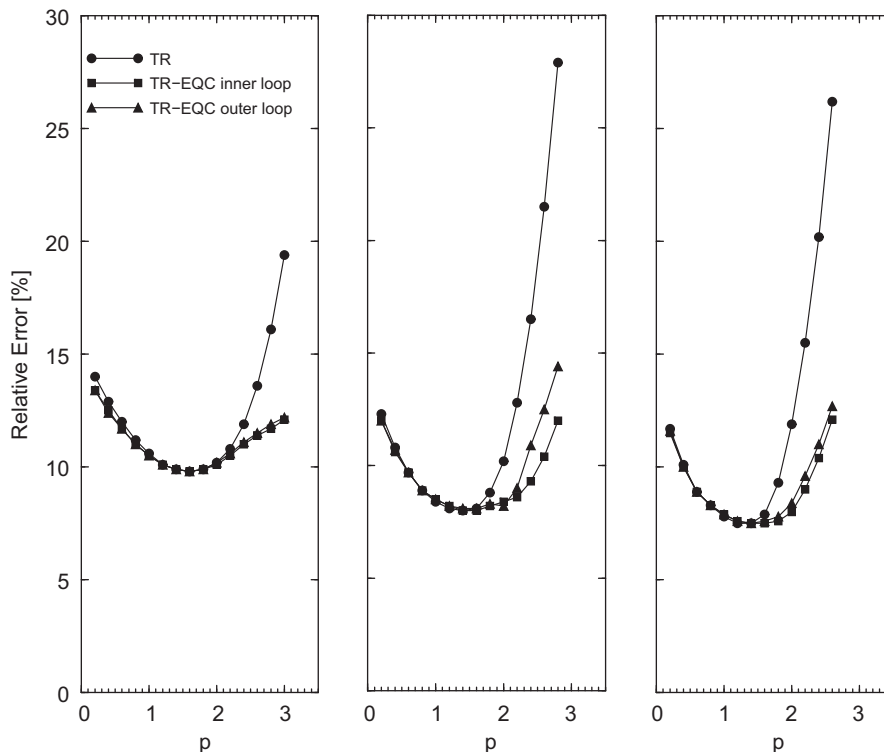
Fig. 1. Relative solution errors for Tikhonov regularization (TR) and the Tikhonov regularization with equality constraints (TR-EQC). The results are computed for the following values of the SNR: 50 (left), 100 (middle) and 150 (right). The regularization parameter is given by $\alpha = \sigma^p$, where σ is the noise standard deviation.
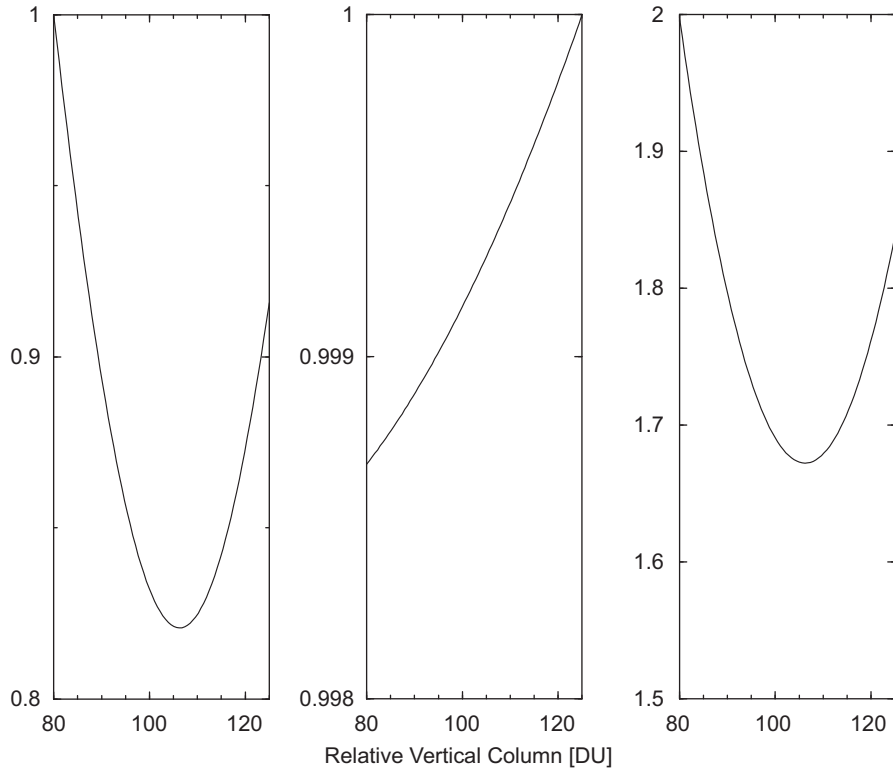
Fig. 2. Normalized constraint norm (left), residual norm (middle) and distance function (right) for different values of the relative vertical column.
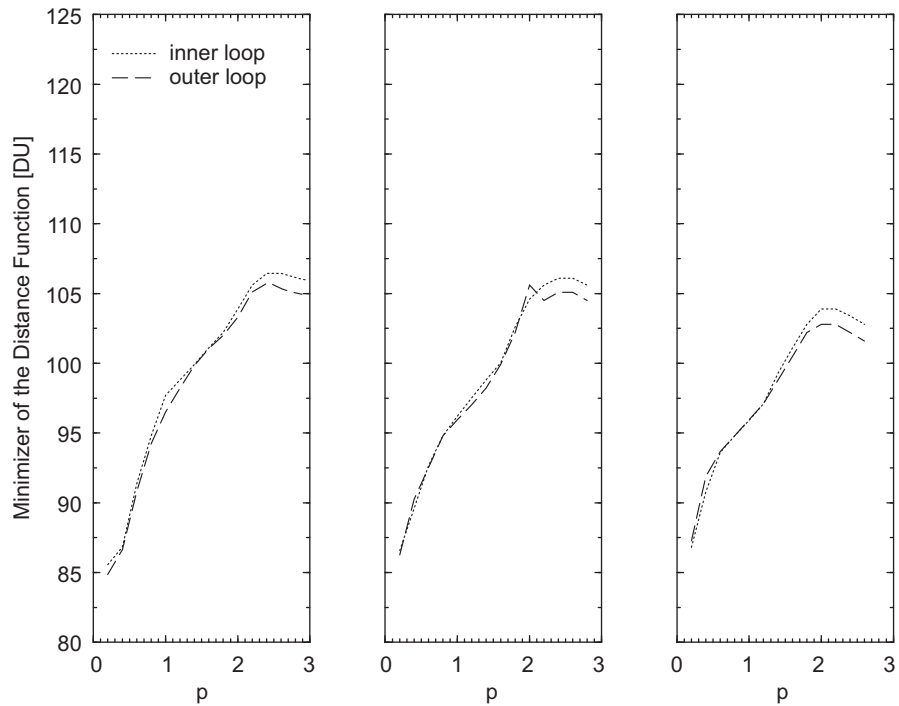

Fig. 3. Minimizers of the distance function for the equality-constrained Tikhonov regularization with an inner and an outer loop. The plots correspond to the following values of the SNR: 50 (left), 100 (middle) and 150 (right).

monotonically decreasing function of the vertical column and possesses a minimum. Since the residual is a monotonically increasing function of the vertical column, the minimizer of the distance function is shifted to the left of the minimizer of the constraint norm.

The minimizers of the distance function are shown in Fig. 3. The results evidence that the minimizers predicted by the two regularization methods are very close.

The retrieval results illustrated in Fig. 4 correspond to a small value of the regularization parameter. The profiles computed by using Tikhonov regularization deviates significantly from the a priori, while the retrieved profiles computed by using the constrained Tikhonov regularizations are smoother and approximate sufficiently well the exact profile (especially in the troposphere).

In Fig. 5 we plot the relative errors in the solutions computed by using Tikhonov regularization and the constrained and unconstrained versions of the iteratively regularized Gauss–Newton method. The plots show that the unconstrained iteratively regularized Gauss–Newton method still yields reliable results for small values of the exponent p, or equivalently, for large initial values of the regularization parameter. Thus, the iteratively regularized Gauss–Newton method is more stable than Tikhonov regularization with respect to overestimations of the regularization parameter, while for underestimations of the regularization parameter, both methods are of comparable accuracies. By contrast, the constrained versions of the iteratively regularized Gauss–Newton methods yield acceptable reconstruction errors over the entire domain of variation of the regularization parameter. At this stage of our presentation we would like to pointed out that as opposed to the equality-constrained method, the inequality-constrained method is sensitive to the selection of the bounds c_{min} and c_{max}. Especially, the selection of the bound c_{min} is critical. Referring to Fig. 3 and taking into account that p is increased during the iterative process, we observe that c_{min} should not be too far from the minimizer of the distance function corresponding to the final value of p. In this regard, a good choice for c_{min} is the plateau at 105 DU; smaller values below this level leads to larger solution errors.

The retrieved profiles computed by using the equality-constrained iteratively regularized Gauss–Newton method and Tikhonov regularization are shown in Fig. 6. For $p = 2.4$, the Tikhonov solution is underregularized, while for $p = 0.2$, the solution is overregularized in the sense that mainly the scaling and less the translation of the a priori profile is reproduced. In both situations, the profiles computed by using the equality-constrained iteratively regularized Gauss–Newton method are better approximations of the exact profile.

The comparison of the numerical effort of the methods can be inferred from Tables 1 and 2. From Table 1, we observe that the computing times of Tikhonov regularization and of its equality-constrained version with an inner loop are almost the same. From Table 2, we see that for $p = 0.2$, the unconstrained Tikhonov regularization is by a factor of 2 faster than the
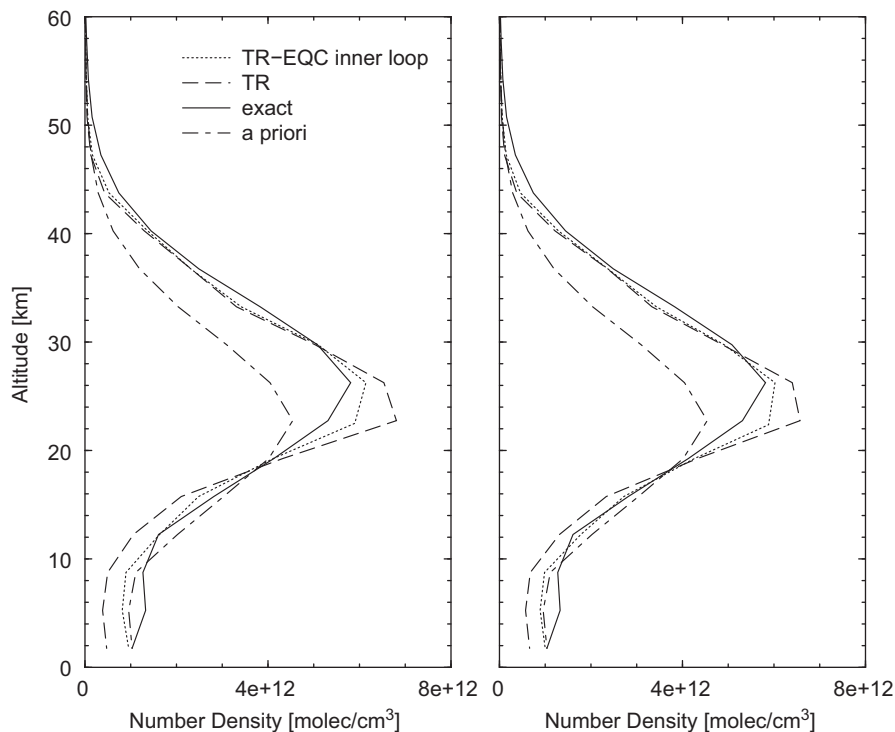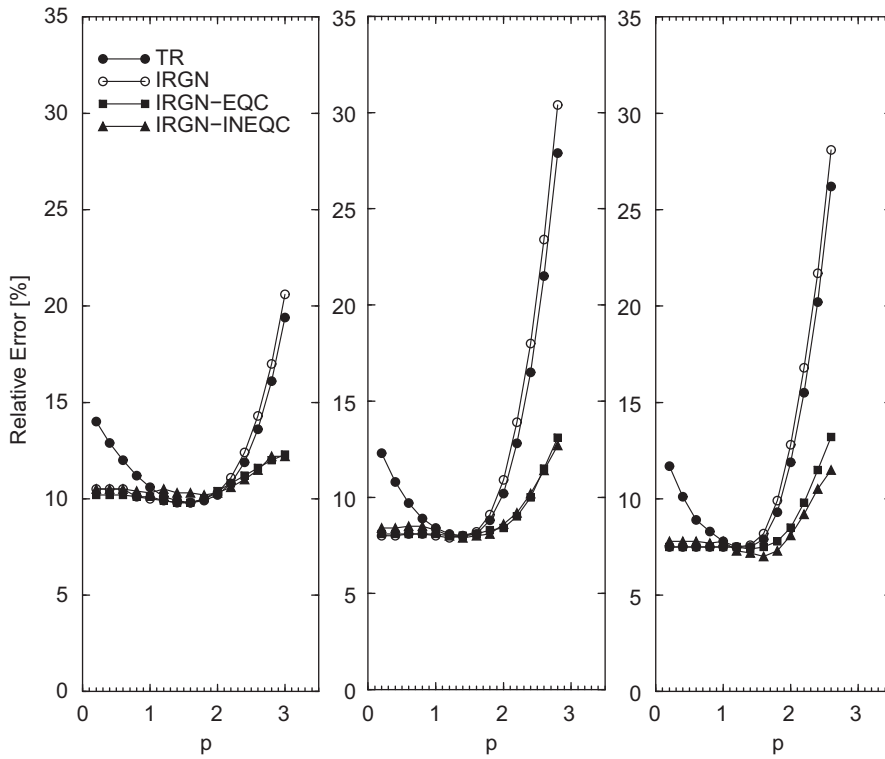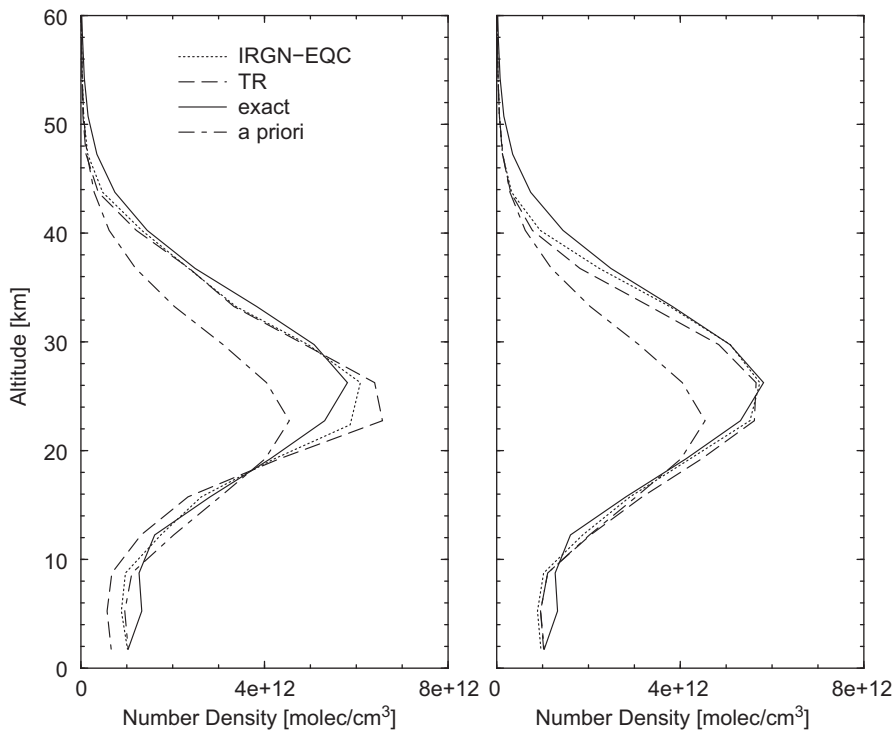

Fig. 4. Retrieved profiles computed by using Tikhonov regularization (TR) and the equality-constrained Tikhonov regularization with an inner loop (TR-EQC) in the case $p = 2.4$. The plots correspond to SNR = 100 (left) and SNR = 150 (right).

Fig. 5. Relative error in the solutions computed by using Tikhonov regularization (TR), the iteratively regularized Gauss–Newton method (IRGN) and the iteratively regularized Gauss–Newton method with equality (IRGN-EQC) and inequality constraints (IRGN-INEQC). The plots correspond to the following values of the SNR: 50 (left), 100 (middle) and 150 (right).

Fig. 6. Retrieved profiles computed by using Tikhonov regularization (TR) and the iteratively regularized Gauss–Newton method with equality constraint (IRGN-EQC) in the case SNR = 100. The plots in the left panel correspond to $p = 2.4$, while the plots in the right panel correspond to $p = 0.2$.

Table 1
Computing time in mm:ss format for Tikhonov regularization and its constrained versions.

p	TR	TR-EQC inner loop	TR-EQC outer loop
2.4	0:20 (4;16.5)	0:21 (4;9.2)	12:28 (4;9.3)
0.2	0:20 (4;12.3)	0:21 (4;12.9)	12:28 (4;12.8)

The numbers in the parentheses represent the number of iterations and the relative solution errors in percent.

Table 2
The same as in Table 1 but for the iteratively regularized Gauss–Newton method and its constrained versions.

p	TR	IRGN	IRGN-EQC	IRGN-INEQC
2.4	0:20 (4;16.5)	0:23 (5;18.0)	0:26 (5;9.8)	0:24 (5;9.9)
0.2	0:20 (4;12.3)	0:39 (12;8.1)	0:50 (12;8.1)	0:42 (12;8.3)

constrained iteratively regularized Gauss–Newton methods, while for $p = 2.4$ their efficiencies are comparable. This increase of computing time is the price that we have to pay for obtaining stable approximations of the solution over a large range of values of the regularization parameter.

4. Conclusions

Novel constrained regularization methods have been designed for ozone retrieval by nadir UV/VIS measurements. The unconstrained minimization step is replaced by a constrained minimization step consisting in additional equality or inequality constraints imposed on the ozone vertical column. The equality-constrained minimization problem is solved in the framework of the null-space approach by using an explicit representation of the solution in terms of the vertical column, while the inequality-constrained minimization problem is solved by using a dual active set algorithm. For equality constraints, the optimal value of the vertical column is computed by employing the minimum distance function criterion and the two schemes for vertical column calculation (with an inner and an outer loop) yield similar results. The main conclusions of our numerical analysis can be summarized as follows:

(1) The constrained versions of Tikhonov regularization and the iteratively regularized Gauss–Newton method are more stable than Tikhonov regularization with respect to underestimations of the regularization parameter and over a large interval of variation of the regularization parameter, respectively.
(2) For equality constraints, the selection of the bounds c_{min} and c_{max} is not critical, and the value of the relative vertical column is only informative. By contrast, for inequality constraints, the information about the relative vertical column should be sufficiently accurate.
(3) The computing time of the equality-constrained Tikhonov regularization with an inner loop resembles that of the ordinary method, while the computing

times of the constrained iteratively regularized Gauss–Newton methods are by a factor of 2 larger than that of Tikhonov regularization.

The application of the constrained regularization methods showed significant improvements in ozone profile retrieval and seems to be a promising tool for ozone retrieval from UV/VIS backscatter measurements, provided by the GOME family of instruments.

Appendix A

Algorithm 1 outlines the method of Tikhonov regularization. An interesting feature of the algorithm is that the regularization parameter is chosen accordingly to the a priori rule $\alpha = \sigma^p$, where the exponent p plays the role of a free parameter. This choice is inspired from the Bayesian approach, which is characterized by the selection criterion $\alpha \propto \sigma^2$. The step-length procedure is optional, but our experience demonstrates that this step improves the stability of the method and reduce the number of iterations.

Algorithm 1. Tikhonov regularization. The step-length procedure computes the scalar τ_k such that the descent condition $\mathcal{F}(\mathbf{x}_{k\alpha}^{\delta} + \tau_k \mathbf{p}_k^{\delta}) < \mathcal{F}(\mathbf{x}_{k\alpha}^{\delta})$ is satisfied.

choose p, $\varepsilon_{\times}$, ε_r, and k_{max} and set $\alpha = \sigma^p$;
$\mathbf{x}_{0\alpha}^{\delta} = \mathbf{x}_a$; compute $\mathbf{F}(\mathbf{x}_{0\alpha}^{\delta})$ and $\mathbf{K}_{0\alpha}$;
$\mathbf{y}_0^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_{0\alpha}^{\delta}) + \mathbf{K}_{0\alpha}(\mathbf{x}_{0\alpha}^{\delta} - \mathbf{x}_a)$; $\mathbf{r}_{0\alpha}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_{0\alpha}^{\delta})$;
$k = 0$;
repeat
 $\mathbf{G} = \mathbf{K}_{k\alpha}^T \mathbf{K}_{k\alpha} + \alpha \mathbf{L}^T \mathbf{L}$, $\mathbf{g} = -\mathbf{K}_{k\alpha}^T \mathbf{y}_k^{\delta}$;
 solve $\min_{\Delta \mathbf{x}} \mathcal{Q}(\Delta \mathbf{x}) = \mathbf{g}^T \Delta \mathbf{x} + \frac{1}{2} \Delta \mathbf{x}^T \mathbf{G} \Delta \mathbf{x}$ for $\Delta \mathbf{x}_{k+1\alpha}^{\delta}$;
 compute the search direction $\mathbf{p}_k^{\delta} = (\mathbf{x}_a + \Delta \mathbf{x}_{k+1\alpha}^{\delta}) - \mathbf{x}_{k\alpha}^{\delta}$;
 compute the step length τ_k for $\mathcal{F}(\mathbf{x}) = \frac{1}{2} \left\| \begin{bmatrix} \mathbf{F}(\mathbf{x}) - \mathbf{y}^{\delta} \\ \sqrt{\alpha} \mathbf{L}(\mathbf{x} - \mathbf{x}_a) \end{bmatrix} \right\|^2$;
 $\mathbf{x}_{k+1\alpha}^{\delta} = \mathbf{x}_{k\alpha}^{\delta} + \tau_k \mathbf{p}_k^{\delta}$; compute $\mathbf{F}(\mathbf{x}_{k+1\alpha}^{\delta})$ and $\mathbf{K}_{k+1\alpha}$;
 $\mathbf{y}_{k+1}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_{k+1\alpha}^{\delta}) + \mathbf{K}_{k+1\alpha}(\mathbf{x}_{k+1\alpha}^{\delta} - \mathbf{x}_a)$; $\mathbf{r}_{k+1\alpha}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_{k+1\alpha}^{\delta})$;
 $k \leftarrow k+1$;
until $(\|\mathbf{x}_{k\alpha}^{\delta} - \mathbf{x}_{k-1\alpha}^{\delta}\| \leq \varepsilon_{\times}.\text{or}.|\|\mathbf{r}_{k\alpha}^{\delta}\| - \|\mathbf{r}_{k-1\alpha}^{\delta}\|| \leq \varepsilon_r \|\mathbf{r}_{k-1\alpha}^{\delta}\|.\text{or}.k > k_{max})$

Algorithm 2. Iteratively regularized Gauss–Newton method

choose p, ε_r, τ, and k_{max} and set $\alpha = \sigma^p$;
$\mathbf{x}_0^{\delta} = \mathbf{x}_a$; compute $\mathbf{F}(\mathbf{x}_0^{\delta})$ and $\mathbf{K}_0$;
$\mathbf{y}_0^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_0^{\delta}) + \mathbf{K}_0(\mathbf{x}_0^{\delta} - \mathbf{x}_a)$; $\mathbf{r}_0^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_0^{\delta})$;
$k = 0$;
repeat
 if $(k > 0)$ $\alpha \leftarrow q\alpha$;
 $\mathbf{G} = \mathbf{K}_k^T \mathbf{K}_k + \alpha \mathbf{L}^T \mathbf{L}$, $\mathbf{g} = -\mathbf{K}_k^T \mathbf{y}_k^{\delta}$;
 solve $\min_{\Delta \mathbf{x}} \mathcal{Q}(\Delta \mathbf{x}) = \mathbf{g}^T \Delta \mathbf{x} + \frac{1}{2} \Delta \mathbf{x}^T \mathbf{G} \Delta \mathbf{x}$ for $\Delta \mathbf{x}_{k+1}^{\delta}$;
 compute the search direction $\mathbf{p}_k^{\delta} = (\mathbf{x}_a + \Delta \mathbf{x}_{k+1}^{\delta}) - \mathbf{x}_k^{\delta}$;
 compute the step length τ_k for $\mathcal{F}(\mathbf{x}) = \frac{1}{2} \left\| \begin{bmatrix} \mathbf{F}(\mathbf{x}) - \mathbf{y}^{\delta} \\ \sqrt{\alpha} \mathbf{L}(\mathbf{x} - \mathbf{x}_a) \end{bmatrix} \right\|^2$;
 $\mathbf{x}_{k+1}^{\delta} = \mathbf{x}_k^{\delta} + \tau_k \mathbf{p}_k^{\delta}$; compute $\mathbf{F}(\mathbf{x}_{k+1}^{\delta})$ and $\mathbf{K}_{k+1}$;
 $\mathbf{y}_{k+1}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_{k+1}^{\delta}) + \mathbf{K}_{k+1}(\mathbf{x}_{k+1}^{\delta} - \mathbf{x}_a)$; $\mathbf{r}_{k+1}^{\delta} = \mathbf{y}^{\delta} - \mathbf{F}(\mathbf{x}_{k+1}^{\delta})$;
 store $\mathbf{x}_{k+1}^{\delta}$ and $\|\mathbf{r}_{k+1}^{\delta}\|$;
 $k \leftarrow k+1$;
until $(|\|\mathbf{r}_k^{\delta}\| - \|\mathbf{r}_{k-1}^{\delta}\|| \leq \varepsilon_r \|\mathbf{r}_{k-1}^{\delta}\|.\text{or}.k > k_{max})$
$\|\mathbf{r}^{\delta}\| = \|\mathbf{r}_k^{\delta}\|$;
choose $\mathbf{x}_{k^*}^{\delta}$ with k^* being the first index for which $\|\mathbf{r}_{k^*}^{\delta}\| \leq \tau \|\mathbf{r}^{\delta}\|$.

Algorithm 2 illustrates the iteratively regularized Gauss–Newton method. In this case, the exponent p determines the initial value of the regularization parameter which is then decreased during the iterative process.

References

[1] Munro R, Siddans R, Reburn WJ, Kerridge BJ. Direct measurement of tropospheric ozone distributions from space. Nature 1998;392:, 168–71.

[2] Hoogen R, Rozanov VV, Burrows JP. Ozone profiles from GOME satellite data: algorithm description and first validation. J Geophys Res 1999;104:, 8263–80.

[3] Van der ARJ, Van Oss RF, Piters AJM, Fortuin JPF, Meijer YJ, Kelder H. Ozone profile retrieval from recalibrated global ozone monitoring experiment data. J Geophys Res 2002 doi:10.1029/2001JD000696.

[4] Liu X, Chance K, Sioris CE, Spurr RJD, Kurosu TP, Martin RV, et al. Ozone profile and tropospheric ozone retrievals from global ozone monitoring experiment: algorithm description and validation. J Geophys Res 2005 doi:10.1029/2005JD006240.

[5] Hasekamp O, Landgraf J. Ozone profile retrieval from backscattered ultraviolet radiances: the inverse problem solved by regularization. J Geophys Res 2001;106:, 8077–88.

[6] Rodgers CD. Inverse methods for atmospheric sounding: theory and practice. Singapore: World Scientific; 2000.

[7] Golub GH, Heath M, Wahba G. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 1979;21:, 215–23.

[8] O'Sullivan F, Wahba G. A cross validated Bayesian retrieval algorithm for nonlinear remote sensing experiments. J Comput Phys 1985;59:, 441–55.

[9] Steck T, von Clarmann T. Constrained profile retrieval applied to the observation mode of the Michelson interferometer for passive atmospheric sounding. Appl Opt 2001;40:, 3559–71.

[10] Vogel CR. Computational methods for inverse problems. Philadelphia: SIAM; 2002.

[11] Schimpf B, Schreier F. Robust and efficient inversion of vertical sounding atmospheric high-resolution spectra by means of regularization. J Geophys Res 1997;102:, 16037–55.

[12] Steck T. Methods for determining regularization for atmospheric retrieval problems. Appl Opt 2002;41:, 1788–97.

[13] Bakushinkii AB. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. USSR Comput Math Math Phys 1984;24:, 181–2.

[14] Engl HW, Hanke M, Neubauer A. Regularization of inverse problems. Dordrecht: Kluwer Academic Publishers; 2000.

[15] Kaltenbacher B, Neubauer A, Scherzer O. Iterative regularization methods for nonlinear ill-posed problems. Radon series on computational and applied mathematics. Berlin: Walter de Gruyter; 2008.

[16] Van Roozendael M, Loyola D, Spurr R, Balis D, Lambert JC, Livschitz Y, et al. Ten years of GOME/ERS-2 total ozone data: the new GOME data processor (GDP) version 4: I. Algorithm description. Journal of Geophysical Research 2006;111:, D14311.

[17] Balis D, Lambert JC, van Roozendael M, Spurr R, Loyola D, Livschitz Y, et al. Ten years of GOME/ERS-2 total ozone data: the new GOME data processor (GDP) version 4: II. Ground-based validation and comparisons with TOMS V7/V8. J Geophys Res 2007;112:, D07307.

[18] Lerot C, van Roozendael M, Lambert JC, Granville J, van Gent J, Loyola D, et al. The GODFIT algorithm a direct fitting approach to improve the accuracy of total ozone measurements from GOME. Int J Remote Sensing 2009, accepted for publication.

[19] Gill PE, Murray W, Wright MH. Practical optimization. London: Elsevier Academic Press; 2004.

[20] Nocedal J, Wright SJ. Numerical optimization. New York: Springer; 2006.

[21] Hansen PC. Rank deficient and discrete ill-posed problems: numerical aspects of linear inversion. Philadelphia: SIAM; 1998.

[22] Belge M, Kilmer ME, Miller EL. Efficient determination of multiple regularization parameters in a generalized L-curve framework. Inverse Problems 2002;18:, 1161–83.

[23] Goldfarb D, Idnani A. A numerically stable dual method for solving strictly convex quadratic programs. Math Programming 1983;27:, 1–33.

[24] Quadprog: functions to solve quadratic programming problems ⟨http://cran.r-project.org/web/packages/quadprog/index.html⟩.

[25] Doicu A, Schreier F, Hess M. Iteratively regularized Gauss–Newton method for atmospheric remote sensing. Comput Phys Commun 2002;148:, 214–26.