National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

16-16

The Atmospheric Infrared Sounder (AIRS) Retrieval, Revisited

Noel Cressie, Rui Wang, and Ben Maloney

National Institute for Applied Statistics Research Australia, University of Wollongong,
Wollongong NSW 2522, Australia Phone +61 2 4221 5435, Fax +61 2 4221 4845.
Email: karink@uow.edu.au

The Atmospheric Infrared Sounder (AIRS) Retrieval, Revisited

Noel Cressie*, Rui Wang†, and Ben Maloney*

Abstract

The algorithm used in the retrieval of geophysical quantities from measurements made by the AIRS instrument depends on two fundamental components. The first is a cost function that is the sum of squares of the differences between radiances and their corresponding forward-model terms. The second is the minimization of this cost function by the method of Vanishing Partial Derivatives (VPD). In this note, we show that the VPD component of the algorithm is identical to a coordinate descent method with Newton-Raphson updates, which allows the AIRS algorithm to be put in context with other optimization algorithms. We also show that the AIRS cost function is a limiting case of the cost function used in Optimal Estimation, which demonstrates how the uncertainty in the AIRS retrieval can be quantified.

1 Introduction

The AIRS instrument flies on NASA's Aqua satellite, which was launched into orbit on May 4, 2002. AIRS has 2378 channels, but only a small selected subset (on the order of 40) is used in a retrieval. It retrieves the following geophysical quantities: T (temperature), q (water vapor), O_3 (ozone), and CO_2 (carbon dioxide). Careful spectroscopy leads to a relatively small dimensional radiance vector and error vector, of dimension $n_\varepsilon = 43$, and the state vector $\mathbf{x}$ consists of just $n_\alpha = 4$ elements that scale the *a priori* column profiles of T, q, O_3, and CO_2, respectively.

Let $\mathbf{y}$ be the n_ε-dimensional vector of measured radiances. The AIRS retrieval assumes the forward model,

$$\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\varepsilon}, \qquad (1)$$

*School of Mathematics and Applied Statistics, University of Wollongong, Australia. Email: ncressie@uow.edu.au

†Columbus, Ohio, USA

where $\boldsymbol{\varepsilon}$ is an n_ε-dimensional error vector that captures imperfections in the given forward function $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_{n_\varepsilon}(\mathbf{x}))'$, and $\mathbf{x}$ is an n_α-dimensional state vector. Write

$$\mathbf{x} = (x_T, x_q, x_O, x_C)', \tag{2}$$

and note that while an AIRS retrieval gives the four geophysical quantities at 100 pressure levels, the *relative* profiles of each quantity are constant throughout the retrieval. Hence, an AIRS retrieval obtains the $n_\alpha = 4$ scaling factors in (2).

The cost function that is minimized with respect to $\mathbf{x}$ is a sum of squares (SS):

$$C_{SS}(\mathbf{x}) \equiv \sum_{i=1}^{n_\varepsilon} (y_i - F_i(\mathbf{x}))^2. \tag{3}$$

This is minimized using the Vanishing Partial Derivative (VPD) method (*Chahine et al.*, 2005), which is reviewed in Section 2. Call this minimized value $\hat{\mathbf{x}}$, and note that it is a function of the radiances $\mathbf{y}$. Since the measurement $\mathbf{y}$ has uncertainty associated with it, so too does the retrieval $\hat{\mathbf{x}}$. Uncertainty quantification of $\hat{\mathbf{x}}$ is an important problem that we address below in Section 3.

In Section 2 of this article, we show that the VPD method is in fact equivalent to a coordinate descent method with Newton-Raphson updates. In Section 3, we observe that $C_{SS}(\mathbf{x})$ given by (3) is a limiting case of the cost function used in Optimal Estimation (*Rodgers*, 2000). Section 4 gives a brief discussion of the potential for retrieval algorithms to be hybrids of that used by AIRS and those based on Optimal Estimation.

2 The VPD Method

For each retrieval, AIRS retrieves a state variable (e.g., CO_2) at a single prespecified pressure level. The state variable is obtained at other pressure levels by matching a column profile so that it agrees at the prespecified pressure level.

The AIRS retrieval uses the Vanishing Partial Derivative (VPD) method to obtain an $\hat{\mathbf{x}}$ that minimizes with respect to $\mathbf{x}$ the cost function:

$$C_{SS}(\mathbf{x}) = (\mathbf{y} - \mathbf{F}(\mathbf{x}))'(\mathbf{y} - \mathbf{F}(\mathbf{x})), \tag{4}$$

which is the same cost function as in (3), but written in vector notation. The VPD method is iterative. At iteration ℓ, define the current state vector to be $\mathbf{x}^{(\ell)}$. The AIRS retrieval updates $x_j^{(\ell)}$, the j-th element of $\mathbf{x}^{(\ell)}$, to obtain $x_j^{(\ell+1)}$: If $x_j^{(\ell)}$ is perturbed using the factor $(1 + \tau_j)$ for small τ_j, we obtain

$$\mathbf{F}(\mathbf{x}^{(\ell)}; \tau_j) = \mathbf{F}(x_1^{(\ell)}, \ldots, x_{j-1}^{(\ell)}, (1 + \tau_j)x_j^{(\ell)}, x_{j+1}^{(\ell)}, \ldots, x_{n_\alpha}^{(\ell)}).$$

Then, by expanding $y_i - F_i(\mathbf{x}^{(\ell)}; \tau_j)$ around $\tau_j = 0$, for each $i = 1, 2, \ldots, n_\varepsilon$, we obtain (to first order),

$$y_i - F_i(\mathbf{x}^{(\ell)}; \tau_j) = \big[y_i - F_i(\mathbf{x}^{(\ell)})\big] - \frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}} x_j^{(\ell)} \cdot \tau_j$$
$$\equiv a_i + \tau_j b_i \ ,$$

where $a_i \equiv [y_i - F_i(\mathbf{x}^{(\ell)})]$ and $b_i \equiv [-\frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}} x_j^{(\ell)}]$. The VPD method uses the simple-linear-regression model (with zero intercept),

$$a_i = -\tau_j b_i + \delta_i \ ,$$

where δ_i is an error term that captures any departure from the straight line. Then an "x–y" line is fitted through the origin to the "(x, y) data," $\{(-b_i, a_i) : i = 1, \ldots, n_\varepsilon\}$.

An Ordinary Least Squares (OLS) fit yields the following estimate for the slope:

$$\widehat{\tau}_j^{(\ell)} = -\frac{\sum_{i=1}^{n_\varepsilon} a_i b_i}{\sum_{i=1}^{n_\varepsilon} b_i^2} = \frac{\sum_{i=1}^{n_\varepsilon} [y_i - F_i(\mathbf{x}^{(\ell)})] \frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}}}{\sum_{i=1}^{n_\varepsilon} \left[\frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}}\right]^2 x_j^{(\ell)}} \ .$$

Then, according to the VPD method, $x_j^{(\ell)}$ is updated to $x_j^{(\ell+1)}$ as follows:

$$x_j^{(\ell+1)} = x_j^{(\ell)} + \widehat{\tau}_j^{(\ell)} x_j^{(\ell)}$$
$$= x_j^{(\ell)} + \frac{\sum_{i=1}^{n_\varepsilon} [y_i - F_i(\mathbf{x}^{(\ell)})] \frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}}}{\sum_{i=1}^{n_\varepsilon} \left[\frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}}\right]^2} \ . \tag{5}$$

For example, suppose the CO_2 variable is retrieved. After updating each of the elements of $\mathbf{x}$ once using the VPD method, the AIRS retrieval changes all elements, except the element for CO_2, back to their initial values. The sequence repeats until there is convergence of the CO_2 value. Thus, the VPD method is a type of regularization, but it looks very different from the Twomey-Tikhonov regularization (*Tikhonov*, 1963; *Twomey*, 1963).

The coordinate descent method (CDM) is often used in optimization of a given criterion with respect to a vector $\mathbf{x}$, where one element of $\mathbf{x}$ is updated at each iteration. The sequence of updates is prespecified and, after all elements are updated in turn, the sequence repeats until convergence. Now suppose we wish to minimize the SS cost function given by (4) and, at each coordinate of $\mathbf{x}$, we use a Newton-Raphson (NR) algorithm (e.g., see *Ypma*, 1995) within a CDM. We now show the equivalence of this familiar method in numerical analysis to the VPD method given by *Chahine et al.* (2005).

Recall the forward model, $\mathbf{y} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\varepsilon}$, given by (1), and consider the coordinate x_j of the state vector $\mathbf{x}$. Let the "initial value" of x_j be $x_j^{(\ell)}$, and fix the other elements of $\mathbf{x}$ at $x_1^{(\ell)}, \ldots, x_{j-1}^{(\ell)}, x_{j+1}^{(\ell)}, \ldots, x_{n_\alpha}^{(\ell)}$. The goal of the CDM here is to minimize the cost function given by (4), where $\mathbf{F}(\mathbf{x})$ is thought of as a function of one variable. Specifically,
$\mathbf{F}(x_1^{(\ell)}, \ldots, x_{j-1}^{(\ell)}, x_j, x_{j+1}^{(\ell)}, \ldots, x_{n_\alpha}^{(\ell)})$ is minimized with respect to x_j.

With a slight abuse of notation, write $\mathbf{F}(x_1^{(\ell)}, \ldots, x_{j-1}^{(\ell)}, x_j, x_{j+1}^{(\ell)}, \ldots, x_{n_\alpha}^{(\ell)})$ as $\mathbf{F}(x_j)$. Taking the derivative of (4) with respect to x_j and putting the result equal to 0 yields:

$$\sum_{i=1}^{n_\varepsilon} [y_i - F_i(x_j)] \frac{dF_i(x_j)}{dx_j} = 0.$$

By writing this equation as $g(x_j) = 0$ and linearizing it around $x_j = x_j^{(\ell)}$, we obtain the approximation, $x_j \simeq x_j^{(\ell)} + g(x^{(\ell)})/g'(x^{(\ell)})$. This motivates the NR algorithm, which updates x_j from initial value $x_j^{(\ell)}$ to $x_j^{(\ell+1)}$ as follows (e.g., see *Fletcher*, 1987):

$$x_j^{(\ell+1)} = x_j^{(\ell)} + \frac{\sum_{i=1}^{n_\varepsilon} [y_i - F_i(\mathbf{x}^{(\ell)})] \frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}}}{\sum_{i=1}^{n_\varepsilon} \left[\frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}} \right]^2}. \quad (6)$$

Inspection of (6) shows that this NR update in a CDM yields an expression that is identical to the VPD update given by (5).

Although the AIRS retrieval does something different with the "initial values," in this section we have shown that each Vanishing Partial Derivative update is equivalent to a Newton-Raphson update in a coordinate descent method. In Section 4, we discuss briefly how this equivalence suggests hybrid retrieval algorithms.

3 A regularization term in the cost function

In this section, we show that the SS cost function given by (4) is a limiting case of the cost function used in retrievals of the state $\mathbf{x}$ using Optimal Estimation (OE). The OE cost function is written as

$$C_{OE}(\mathbf{x}) = (\mathbf{y} - \mathbf{F}(\mathbf{x}))' \mathbf{S}_\varepsilon^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x})) + (\mathbf{x} - \mathbf{x}_\alpha)' \mathbf{S}_\alpha^{-1} (\mathbf{x} - \mathbf{x}_\alpha), \quad (7)$$

where $\mathbf{S}_\varepsilon$, $\mathbf{x}_\alpha$, and $\mathbf{S}_\alpha$ are the measurement-error variance of $\mathbf{y}$, the prior mean of $\mathbf{x}$, and the prior covariance of $\mathbf{x}$, respectively (*Rodgers*, 2000). There are at least two interpretations of (7); one is that the first term ("fidelity" to the data) is regularized with the addition of the second term ("smoothness" of the state). A second interpretation is that, up to an additive constant, (7) is minus twice the log of the posterior distribution of $\mathbf{x}$ given $\mathbf{y}$, where a joint multivariate Gaussian distribution for $\boldsymbol{\varepsilon}$ and $\mathbf{x}$ is assumed. Under this interpretation, minimizing (7) is equivalent to finding the mode of the posterior distribution.

Using the Gauss-Newton method that drops the second derivatives, the basic OE iteration scheme is (*Rodgers*, 2000):

$$\begin{aligned}\mathbf{x}^{(\ell+1)} = \mathbf{x}^{(\ell)} &+ \{\mathbf{S}_\alpha^{-1} + \mathbf{K}^{(\ell)\prime}\mathbf{S}_\varepsilon^{-1}\mathbf{K}^{(\ell)}\}^{-1} \\ &\cdot [\mathbf{K}^{(\ell)\prime}\mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x}^{(\ell)})) - \mathbf{S}_\alpha^{-1}(\mathbf{x}^{(\ell)} - \mathbf{x}_\alpha)]\ ,\end{aligned} \tag{8}$$

where $\mathbf{K}^{(\ell)} \equiv \mathbf{K}(\mathbf{x}^{(\ell)})$ and $\mathbf{K}(\mathbf{x}) \equiv \partial\mathbf{F}(\mathbf{x})/\partial\mathbf{x}$ is the Jacobian matrix. As the Gauss-Newton iteration scheme can be unstable, many retrieval algorithms try to resolve this by using a Levenberg-Marquardt (LM) modification (*Levenberg*, 1944; *Marquardt*, 1963) within the Gauss-Newton algorithm, which replaces the first $\mathbf{S}_\alpha^{-1}$ in (8) (but not the second) with the iteration-dependent term, $(1 + \gamma^{(\ell)})\mathbf{S}_\alpha^{-1}$.

Clearly, if we put $\mathbf{S}_\varepsilon = \sigma_\varepsilon^2\mathbf{I}$ and $\mathbf{S}_\alpha^{-1} = \mathbf{0}$, the OE cost function in (7) is identical, up to a scaling constant, to the SS cost function given by (4). The measurement-error matrix $\mathbf{S}_\varepsilon$ can be considered to be diagonal when there is little or no cross-dependence between the selected channels. Then write $\mathbf{S}_\varepsilon = \text{diag}(\sigma_1^2, \ldots, \sigma_{n_\varepsilon}^2)$, where $\sigma_i^2 \equiv \text{var}(\varepsilon_i)$, for $i = 1, \ldots, n_\varepsilon$, and put

$$\tilde{y}_i \equiv y_i/\sigma_i \text{ and } \tilde{F}_i(\mathbf{x}) \equiv F_i(\mathbf{x})/\sigma_i.$$

Hence, the first term of $C_{OE}(\mathbf{x})$ is,

$$(\mathbf{y} - \mathbf{F}(\mathbf{x}))'\mathbf{S}_\varepsilon^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{x})) = (\tilde{\mathbf{y}} - \tilde{\mathbf{F}}(\mathbf{x}))'(\tilde{\mathbf{y}} - \tilde{\mathbf{F}}(\mathbf{x})); \tag{9}$$

that is, after a simple re-scaling by known $\{\sigma_i : i = 1, \ldots, n_\varepsilon\}$, the SS cost function in (4) is obtained.

Our claim that the SS cost function used by AIRS is a limiting case of the OE cost function can now be established. To show this, write the prior covariance matrix in (7) as

$$\mathbf{S}_\alpha = \sigma_\alpha^2\mathbf{I}.$$

If the AIRS instrument has approximately equal measurement-error variances in all n_ε channels, then $\mathbf{S}_\varepsilon = \sigma_\varepsilon^2\mathbf{I}$, where σ_ε^2 is this common measurement-error variance. We start with this assumption (and later generalize our results to any positive-definite matrix $\mathbf{S}_\varepsilon$). Then

$$C_{OE}(\mathbf{x}) = (1/\sigma_\varepsilon^2)(\mathbf{y} - \mathbf{F}(\mathbf{x}))'(\mathbf{y} - \mathbf{F}(\mathbf{x})) + (1/\sigma_\alpha^2)(\mathbf{x} - \mathbf{x}_\alpha)'(\mathbf{x} - \mathbf{x}_\alpha)$$
$$\propto (\mathbf{y} - \mathbf{F}(\mathbf{x}))'(\mathbf{y} - \mathbf{F}(\mathbf{x})) + R^{-1}(\mathbf{x} - \mathbf{x}_\alpha)'(\mathbf{x} - \mathbf{x}_\alpha),$$

where $R \equiv \sigma_\alpha^2/\sigma_\varepsilon^2$ can be interpreted as the signal-to-noise ratio.

Mathematically, as R tends to infinity, $C_{OE}(\mathbf{x})$ tends to $C_{SS}(\mathbf{x})$, the cost function used by AIRS. Our strategy in quantifying the uncertainty in the AIRS retrieval is to stay back from the limit, where we can apply OE's uncertainty quantification. Then we take the limit as R tends to infinity, of the results obtained from OE, to yield uncertainty quantification for the AIRS retrieval.

Rodgers (2000) developed OE's uncertainty quantification based on linear-approximation theory (called the "delta method" in the statistics literature). That is, he obtained the following approximation to the mean squared prediction error (MSPE) matrix:

$$E(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})' \simeq \hat{\mathbf{S}} \equiv (\mathbf{S}_\alpha^{-1} + \mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K})^{-1}.$$

Hence, for $\mathbf{S}_\varepsilon = \sigma_\varepsilon^2\mathbf{I}$, and $\mathbf{S}_\alpha = \sigma_\alpha^2\mathbf{I}$, we obtain

$$\hat{\mathbf{S}} = \sigma_\varepsilon^2\{(1/R)\mathbf{I} + \mathbf{K}'\mathbf{K}\}^{-1}. \quad (10)$$

Now let R tend to infinity (equivalently, $1/R$ tend to 0), so that the MSPE matrix is,

$$\hat{\mathbf{S}} = \sigma_\varepsilon^2(\mathbf{K}'\mathbf{K})^{-1}, \quad (11)$$

provided $\mathbf{K}'\mathbf{K}$ is invertible. Should $1/R$ not be zero and a specification of it becomes available, it can be substituted into (10), resulting in a matrix that is always invertible.

If $\mathbf{S}_\varepsilon$ is more generally a positive-definite matrix and is known, then (9) is obtained by using the general rescaling, $\tilde{\mathbf{y}} = \mathbf{S}_\varepsilon^{-1/2}\mathbf{y}$, $\tilde{\mathbf{F}}(\mathbf{x}) = \mathbf{S}_\varepsilon^{-1/2}\mathbf{F}(\mathbf{x})$. Thus, minimizing (9) is obtained by replacing $\mathbf{y}$ with $\tilde{\mathbf{y}}$, $\mathbf{F}(\mathbf{x})$ with $\tilde{\mathbf{F}}(\mathbf{x})$, and $\mathbf{K}$ with $\tilde{\mathbf{K}} \equiv \partial\tilde{\mathbf{F}}/\partial\mathbf{x} = \mathbf{S}_\varepsilon^{-1/2}\mathbf{K}$. Consequently, the MSPE matrix is

$$\hat{\mathbf{S}} = (\tilde{\mathbf{K}}'\tilde{\mathbf{K}})^{-1} = (\mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K})^{-1}. \quad (12)$$

Now consider the case where $\mathbf{S}_\varepsilon$ is generally a positive-definite matrix but the AIRS algorithm still minimizes $C_{SS}(\mathbf{x})$ given by (4). Then it is straightforward to show that the MSPE matrix is

$$\hat{\mathbf{S}} = (\mathbf{K}'\mathbf{K})^{-1}(\mathbf{K}'\mathbf{S}_\varepsilon\mathbf{K})(\mathbf{K}'\mathbf{K})^{-1}. \quad (13)$$

Notice that when $\mathbf{S}_\varepsilon = \sigma_\varepsilon^2\mathbf{I}$ in (13), the expression (11) is obtained, as expected. Consequently, for a retrieval based on minimizing the sum-of-squares cost function,

$C_{SS}(\mathbf{x})$, the expression (13) for $\hat{\mathbf{S}}$ should always be used, since it is always valid. Note that it is straightforward to show that even when $\mathbf{S}_\alpha^{-1}$ is not $\mathbf{0}$, a retrieval that minimizes $C_{SS}(\mathbf{x})$ continues to have its uncertainty quantified by the MSPE matrix (13).

Since the Jacobian matrix represents sensitivity of radiances to changes in the elements of the state vector, it is possible that changes in different elements lead to indistinguishable sensitivities that would result in near singularity for $\mathbf{K}'\mathbf{K}$ or $\mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K}$. This possibility has been addressed by *Ramanathan et al.* (2016) although, for the AIRS retrieval, full rank of $\mathbf{K}$ and hence of $\mathbf{K}'\mathbf{K}$ is maintained because the four state-space elements are defined respectively for four different geophysical quantities.

In this section, we have shown that the SS cost function used in the AIRS retrieval can be approximated by an OE cost function with equal measurement-error variances and very large signal-to-noise ratio. The implication of this is that the uncertainty-quantification equations at the disposal of OE-based retrievals, are also at the disposal of the AIRS mission. Specifically, (13) is the (approximate) mean-squared-prediction-error matrix of the AIRS retrieval vector.

4 Discussion

We have seen that the VPD method in the AIRS retrieval is, at its core, a series of Newton-Raphson updates within a coordinate descent method. It is possible to show that the iterative procedure in the AIRS retrieval can be characterized by highly diverse choices of the Levenberg-Marquardt inflation factors. This alternate way to look at VPD is very useful. It suggests how a generic retrieval algorithm, including that of AIRS, might be modified should measurement-error variances be highly different, should the Jacobian matrix have some uncertainty associated with it, or should parts of the state vector come naturally as blocks of variables (e.g., CO_2 values at 20 pressure levels in an atmospheric column would form one block). We have also seen that OE's uncertainty quantification can be implemented on AIRS retrievals, and we give a general result for the (approximate) mean-squared-prediction-error matrix associated with any AIRS retrieval.

Acknowledgments

We would like to thank Ed Olsen and Hai Nguyen at the Jet Propulsion Laboratory for helpful discussions about the AIRS retrieval and Anand Ramanathan of Goddard Space Flight Center for sharing his co-authored manuscript. This research was supported by NASA grant NNH11-ZDA001N-OCO2, the AIRS mission at the

Jet Propulsion Laboratory, Pasadena, CA, and a 2015-2017 Australian Research Council Discovery Project, DP 150104576.

References

Chahine, M., C. Barnet, E. T. Olsen, L. Chen, and E. Maddy. (2005). On the determination of atmospheric minor gases by the method of vanishing partial derivatives with application to CO_2. *Geophysical Research Letters*, *32*, L22, 803, doi:10.1029/2005GL024165.

Fletcher, R. (1987). *Practical Methods of Optimization, 2nd edn.* John Wiley & Sons: New York, NY.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, **2**, 164-168.

Marquardt, D. (1963). An algorithm for least-squares estimation of non-linear parameters. *SIAM Journal on Applied Mathematics*, **11**, 431-441.

Ramanathan, A. K., X. Sun, J. Mao, J. B. Abshire, H. M. Nguyen, J. M. Hobbs, and A. J. Braverman. (2016). Retrieving vertical information in the trace gas distribution using a non-Bayesian Singular Value Decomposition method applied to column absorption measurements. Manuscript.

Rodgers, C. D. (2000). *Inverse Methods for Atmospheric Sounding.* World Scientific Publishing: Singapore.

Tikhonov, A. N. (1963). On the solution of incorrectly stated problems and a method of regularization. *Doklady Academii Nauk SSSR*, **151**, 501-504.

Twomey, S. (1963). On the numerical solution of Fredholm integral equation of the first kind by the inversion of the linear system produced by quadrature. *Journal of the Association of Computing Machinery*, **10**, 97-101.

Ypma, T. J. (1995). Historical development of the Newton-Raphson method. *SIAM Review*, **37**, 531-551.