2017

The Atmospheric Infrared Sounder Retrieval, Revisited

Noel A. Cressie
University of Wollongong, ncressie@uow.edu.au

Rui Wang

Benjamin Maloney
University of Wollongong, bkm611@uowmail.edu.au

The Atmospheric Infrared Sounder Retrieval, Revisited

Abstract

The algorithm used in the retrieval of geophysical quantities from the Atmospheric Infrared Sounder (AIRS) instrument depends on two fundamental components. The first is a cost function that is the sum of squares of the differences between cloud-cleared radiances and their corresponding forward-model terms. The second is the minimization of this cost function. For the retrieval of carbon dioxide, the minimization is further improved using the method of vanishing partial derivatives (VPDs). In this letter, we show that this VPD component is identical to a coordinate descent method with Newton-Raphson updates, which allows it 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 uncertainty quantification in the AIRS retrieval can be implemented.

Disciplines

Engineering | Science and Technology Studies

Publication Details

Cressie, N., Wang, R. & Maloney, B. (2017). The Atmospheric Infrared Sounder Retrieval, Revisited. IEEE Geoscience and Remote Sensing Letters, 14 (9), 1504-1507.

The Atmospheric Infrared Sounder (AIRS) Retrieval, Revisited

Noel Cressie[1], Rui Wang[2], and Ben Maloney[1]

***Abstract*—The algorithm used in the retrieval of geophysical quantities from the AIRS instrument depends on two fundamental components. The first is a cost function that is the sum of squares of the differences between cloud-cleared radiances and their corresponding forward-model terms. The second is the minimization of this cost function. For the retrieval of carbon dioxide, the minimization is further improved using the method of Vanishing Partial Derivatives (VPD). In this article, we show that this VPD component is identical to a coordinate descent method with Newton-Raphson updates, which allows it 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 uncertainty quantification in the AIRS retrieval can be implemented.**

I. Introduction

The AIRS instrument flies on NASA's Aqua satellite, which was launched into orbit on May 4, 2002. It is a rich resource of atmospheric data with global coverage over a continuous period of approximately 15 years. 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, 43-dimensional cloud-cleared radiance vector and error vector, and the state vector **x** consists of just four elements that scale the *a priori* column profiles of T, q, O_3, and CO_2, respectively.

There is typically highly variable cloud absorption and scattering in the instrument field-of-view (FOV). AIRS deals with cloudy infra-red spectra by "cloud clearing" them, which simplifies profile calculations as the forward model does not then need to incorporate scattering or absorption by clouds.

Let **y** be the n_ε-dimensional vector of cloud-cleared radiances ([1], [9]). The AIRS retrieval assumes the forward model,

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

where $\boldsymbol{\varepsilon}$ is an n_ε-dimensional error vector that captures imperfections in the forward function, $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), \ldots, F_{n_\varepsilon}(\mathbf{x}))'$, and **x** is an n_α-dimensional state vector. The Level 1b measured radiances are cloud-cleared, resulting in Level 2 cloud-cleared radiances. Then $\mathbf{F}(\cdot)$ is not purely a radiance transfer function but also has a component due to the cloud clearing that extrapolates the 3×3 set of measured radiances (within its associated AMSU footprint) to cloud-free conditions.

[1] School of Mathematics and Applied Statistics, University of Wollongong, Australia. Corresponding email: ncressie@uow.edu.au
[2] Columbus, Ohio, USA

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 cloud-clearing process generates radiances that would have been seen if the 3×3 fields upon which a retrieval is performed were entirely cloud free. From the technical document [6]: For each channel i, the observations R_{ij} are potentially affected by clouds in Field-Of-View (FOV) j ($j = 1, \ldots, 9$). Then the cloud-cleared radiance is derived from the original nine radiances according to

$$y_i = \overline{R}_i + (1/9)\sum_{j=1}^{9} \eta_j (R_{ij} - \overline{R}_i)\,,$$

where $\overline{R}_i$ is the mean of R_{ij} over the nine FOVs, and η_j is the cloud fraction in FOV j (obtained as part of the retrieval process); see also [9]. This describes the cloud-clearing used in the past and current (AIRS-V6) retrieval algorithms.

The cost function that is minimized with respect to **x** is a sum of squares (SS):

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

The AIRS algorithm is based on minimizing (3); see Susskind et al. [10] for a detailed description and references. The retrieval of the CO_2 product has an additional component that minimizes (3) using the Vanishing Partial Derivative (VPD) method (see Chahine et al. [2]), which is reviewed in Section II. Call this minimized value $\hat{\mathbf{x}}$, and note that it is a function of the cloud-cleared radiances **y**. Since **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 III.

In Section II of this article, we show that the VPD method is in fact equivalent to a coordinate descent method with Newton-Raphson updates. In Section III, we observe that $C_{SS}(\mathbf{x})$ given by (3) is a limiting case of the cost function used in Optimal Estimation (see Rodgers [8]). Section IV gives a brief discussion of the potential for retrieval algorithms to be hybrids of that used by AIRS and those based on Optimal Estimation.

II. 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 CO_2 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 of CO_2 updates $x_j^{(\ell)}$, where j is the CO_2 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, then define

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

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),

$$\begin{aligned} y_i - F_i(\mathbf{x}^{(\ell)};\tau_j) &= \left[y_i - F_i(\mathbf{x}^{(\ell)})\right] - \frac{\partial F_i(\mathbf{x}^{(\ell)})}{\partial x_j^{(\ell)}} x_j^{(\ell)} \cdot \tau_j \\ &\equiv a_i + \tau_j b_i \ , \end{aligned}$$

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:

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

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

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

Recall that j denotes the CO_2 variable. 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 ([11], [12]).

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 [13]) within a CDM. We now show the equivalence of this familiar method in numerical analysis to the VPD method given by Chahine et al. in [2].

Recall the forward model, $\mathbf{y} = \mathbf{F}(\mathbf{x}) + \varepsilon$, 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 [3]):

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

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

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

III. 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

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

where $\mathbf{S}_\varepsilon$, $\mathbf{x}_\alpha$, and $\mathbf{S}_\alpha$ are the covariance matrix of $\mathbf{y}$, the prior mean vector of $\mathbf{x}$, and the prior covariance matrix of $\mathbf{x}$, respectively ([8]). 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 ε and $\mathbf{x}$ is assumed. Under this interpretation, minimizing (7) is equivalent to finding the mode of the posterior distribution.

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).

Using the Gauss-Newton method that drops the second derivatives, the basic OE iteration scheme is ([8]):

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

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 ([4], [5]) within the Gauss-Newton algorithm, which replaces the first appearance of the term $\mathbf{S}_\alpha^{-1}$ in (8) (but not the second) with the iteration-dependent term, $(1+\gamma^{(\ell)})\mathbf{S}_\alpha^{-1}$.

The covariance matrix $\mathbf{S}_\varepsilon$ associated with the forward model is often considered to be diagonal, in which case we 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, if the inhomogeneous variances are properly accounted for, the first term of $C_{OE}(\mathbf{x})$ should become,

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

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

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

Assume for the moment that the forward-model errors have approximately equal variances in all n_ε channels and zero covariances, so that $\mathbf{S}_\varepsilon = \sigma_\varepsilon^2\mathbf{I}$, where σ_ε^2 is this common variance. We start with this equal-variance assumption (and later generalize our results to any positive-definite matrix $\mathbf{S}_\varepsilon$). Then

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

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 [8] developed OE's uncertainty quantification based on linear-approximation theory (which is called the "delta method" in the statistics literature). It relies on the following approximation to the prediction error:

$$\hat{\mathbf{x}} - \mathbf{x} \simeq (\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}_\alpha) + \mathbf{G}\boldsymbol{\varepsilon} \,, \quad (10)$$

where $\mathbf{G} \equiv (\mathbf{S}_\alpha^{-1} + \mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K})^{-1}\mathbf{K}'\mathbf{S}_\varepsilon^{-1}$ is the gain matrix, and $\mathbf{A} \equiv \mathbf{G}\mathbf{K}$ is the averaging-kernel matrix. Consequently, Rodgers in [8] obtained an approximation to the first two moments of the prediction error:

$$\text{E}(\hat{\mathbf{x}} - \mathbf{x}) \simeq (\mathbf{A} - \mathbf{I})\text{E}(\mathbf{x} - \mathbf{x}_\alpha) + \mathbf{G}\text{E}(\boldsymbol{\varepsilon}) = \mathbf{0} \,, \quad (11)$$

and because $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are independent,

$$\begin{aligned} \text{E}(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})' &\simeq (\mathbf{A} - \mathbf{I})\text{cov}(\mathbf{x})(\mathbf{A} - \mathbf{I})' + \mathbf{G}\text{cov}(\boldsymbol{\varepsilon})\mathbf{G}' \\ &= (\mathbf{A} - \mathbf{I})\mathbf{S}_\alpha(\mathbf{A} - \mathbf{I})' + \mathbf{G}\mathbf{S}_\varepsilon\mathbf{G}' \,. \end{aligned} \quad (12)$$

There are several equivalent ways to write the mean squared prediction error (MSPE) matrix (12). Here, we choose,

$$\text{E}(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})' \simeq \hat{\mathbf{S}} \equiv (\mathbf{S}_\alpha^{-1} + \mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K})^{-1} \,. \quad (13)$$

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

$$\mathbf{G} = \left(R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K}\right)^{-1}\mathbf{K}' \text{ and } \mathbf{A} = \left(R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K}\right)^{-1}\mathbf{K}'\mathbf{K} \,, \quad (14)$$

and from (13),

$$\hat{\mathbf{S}} = \sigma_\varepsilon^2\left(R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K}\right)^{-1} \,. \quad (15)$$

Now let R tend to infinity (equivalently, let R^{-1} tend to zero), so that from (14),

$$\mathbf{G} = (\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}' \text{ and } \mathbf{A} = \mathbf{I} \,; \quad (16)$$

and from (15), the MSPE matrix is approximately

$$\hat{\mathbf{S}} = \sigma_\varepsilon^2(\mathbf{K}'\mathbf{K})^{-1} \,, \quad (17)$$

provided $\mathbf{K}'\mathbf{K}$ is invertible.

More generally, assume $\mathbf{S}_\varepsilon$ is a known positive-definite matrix and is used in the cost function given by (9). Define $\tilde{\mathbf{y}} = \mathbf{S}_\varepsilon^{-1/2}\mathbf{y}$, $\tilde{\mathbf{F}}(\mathbf{x}) = \mathbf{S}_\varepsilon^{-1/2}\mathbf{F}(\mathbf{x})$. 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, from (17), the MSPE matrix is approximately

$$\hat{\mathbf{S}} = (\tilde{\mathbf{K}}'\tilde{\mathbf{K}})^{-1} = (\mathbf{K}'\mathbf{S}_\varepsilon^{-1}\mathbf{K})^{-1}. \quad (18)$$

Now consider the actual situation where the AIRS algorithm does not rescale using $\mathbf{S}_\varepsilon^{-1/2}$ and so keeps the cost function $C_{SS}(\mathbf{x})$ given by (4). Then from (12) and (16),

$$\begin{aligned} \hat{\mathbf{S}} &= \mathbf{0} + \mathbf{G}\mathbf{S}_\varepsilon\mathbf{G}' \\ &= (\mathbf{K}'\mathbf{K})^{-1}(\mathbf{K}'\mathbf{S}_\varepsilon\mathbf{K})(\mathbf{K}'\mathbf{K})^{-1}. \end{aligned} \quad (19)$$

Notice that when $\mathbf{S}_\varepsilon = \sigma_\varepsilon^2\mathbf{I}$ in (19), the expression (17) is obtained, as expected. The expression (19) is our most general result. In (12)-(19), we recommend evaluating $\mathbf{K}$, $\mathbf{G}$, and $\mathbf{A}$ at $\mathbf{x} = \mathbf{x}_\alpha$ to maintain validity of the approximations given in (10)-(13).

For a retrieval based on minimizing the cost function, $C_{SS}(\mathbf{x})$, the expression (19) for $\hat{\mathbf{S}}$ should always be used, since it is always valid. It is shown in the Appendix 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 matrix (19).

Since the Jacobian matrix represents sensitivity of (cloud-cleared) 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. [7] although, for the AIRS retrieval, full rank of $\mathbf{K}$ and hence of $\mathbf{K}'\mathbf{K}$ is generally 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 is the limit of an OE cost function. 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.

In conclusion, (19) is always the (approximate) MSPE matrix of the AIRS retrieval vector for any true prior covariance matrix $\mathbf{S}_\alpha$, where $\mathbf{S}_\varepsilon$ is the true covariance matrix of the forward-model errors.

IV. Discussion

In Section II, we saw that the VPD method is, at its core, a series of Newton-Raphson updates within a coordinate descent method. 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 forward-model-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).

In Section III, we saw that OE's uncertainty quantification can be implemented on AIRS retrievals, and we gave a general result, expression (19), for the (approximate) mean-squared-prediction-error matrix associated with any AIRS retrieval.

It is now clear what is needed for uncertainty quantification of the AIRS retrieval: $\text{var}(\mathbf{y}) = \mathbf{S}_\varepsilon$ and $\mathbf{K} = \partial\mathbf{F}/\partial\mathbf{x}$, where $\mathbf{F}$ is the forward function of the cloud-cleared radiances $\mathbf{y}$. The first quantity could be obtained using an empirical covariance matrix calculated from a sample of cloud-cleared radiances taken under similar atmospheric conditions. The second quantity could be obtained numerically by perturbing $\mathbf{F}(\mathbf{x})$ about $\mathbf{x} = \mathbf{x}_\alpha$ in each of the four components of the state vector and approximating the derivative with a difference.

Appendix

From Section III,

$$C_{SS}(\mathbf{x}) = \lim_{R\to\infty} \{(\mathbf{y}-\mathbf{F}(\mathbf{x}))'(\mathbf{y}-\mathbf{F}(\mathbf{x}))+R^{-1}(\mathbf{x}-\mathbf{x}_\alpha)'(\mathbf{x}-\mathbf{x}_\alpha)\}\,.$$

From (10), (14), and for R large,

$$\begin{aligned}\hat{\mathbf{x}} - \mathbf{x} \simeq &\{(R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{K} - \mathbf{I}\}(\mathbf{x} - \mathbf{x}_\alpha) \\ &+ (R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\boldsymbol{\varepsilon}\,. \end{aligned} \tag{20}$$

An implicit distributional assumption behind using $C_{SS}(\mathbf{x})$ is $\text{cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2\mathbf{I}$ and $\text{cov}(\mathbf{x}) = R\sigma_\varepsilon^2\mathbf{I}$, for R large. However, the true forward-model-error process may not be homoscedastic, and the true state of the atmosphere will not be infinitely disperse. Hence, we consider the most general case that $\text{cov}(\boldsymbol{\varepsilon}) = \mathbf{S}_\varepsilon$ and $\text{cov}(\mathbf{x}) = \mathbf{S}_\alpha$, which are only assumed to be positive-definite matrices.

From (12), the approximate MSPE matrix consists of two terms. The first term is, for R large,

$$\left\{\left(R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K}\right)^{-1}\mathbf{K}'\mathbf{K} - \mathbf{I}\right\}\mathbf{S}_\alpha\left\{\left(R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K}\right)^{-1}\mathbf{K}'\mathbf{K} - \mathbf{I}\right\}\,. \tag{21}$$

The second term is, for R large,

$$\left(R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K}\right)^{-1}\mathbf{K}'\mathbf{S}_\varepsilon\mathbf{K}\left(R^{-1}\mathbf{I} + \mathbf{K}'\mathbf{K}\right)\,. \tag{22}$$

Now let $R \to \infty$; then (21) converges to the zero matrix, regardless of the value of $\mathbf{S}_\alpha$. Further, (22) converges to (19). Adding the two terms taken to the limit as $R \to \infty$, establishes the desired result.

Acknowledgments

We would like to thank Ed Olsen, Tom Pagano, Bill Irion, and Hai Nguyen at the Jet Propulsion Laboratory for helpful discussions about the AIRS retrieval, an anonymous referee for clarifying the role of cloud-clearing and the use of the VPD algorithm in 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

[1] Chahine, M. (1974). Remote sounding of cloudy atmospheres. I. The single cloud layer. *Journal of the Atmospheric Sciences*, **31**, 233-243.

[2] 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.

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

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

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

[6] Olsen, E. T. (ed.), J. Susskind, J. Blaisdell, and L. Iredell. (2013). AIRS version 6 Level 2 cloud cleared radiances. Technical Document. Jet Propulsion Laboratory. `https://disc.sci.gsfc.nasa.gov/AIRS/documentation/`

[7] 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.

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

[9] Susskind, J., C. D. Barnet, and J. M. Blaisdell. (2003). Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. *IEEE Transactions of Geoscience and Remote Sensing*, **41**, 390-409.

[10] Susskind, J., J. M. Blaisdell, and L. Iredell. (2014). Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm. *Journal of Applied Remote Sensing*, **8**, 084994-084994.

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

[12] 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.

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