Scaled Linear Retrieval

Anu Dudhia
Atmospheric, Oceanic and Planetary Physics, Oxford University, UK

12 March 2020

1 Linear Retrieval Concept

In a linear retrieval the measured spectrum, expressed as a (column) vector $\mathbf{y}$ of m spectral points, can be thought of as the sum of a set of n spectral components, represented by the columns of a $(m \times n)$ matrix $\mathbf{K}$:

$$\mathbf{y} = \mathbf{K}\mathbf{x}$$

where $\mathbf{x}$ is a vector of n scale factors for each component. $\mathbf{x}$ can be obtained through the inverse operation

$$\mathbf{x} = \mathbf{G}\mathbf{y}$$

where the $(n \times m)$ matrix $\mathbf{G}$ is given by the standard least squares fit solution

$$\mathbf{G} = \left(\mathbf{K}^T\mathbf{K}\right)^{-1}\mathbf{K}^T$$

Note that $\mathbf{G}$ is a *pseudo-inverse* of the $\mathbf{K}$ matrix: $\mathbf{GK} = \mathbf{I}_n$ whereas, generally, $\mathbf{KG} \neq \mathbf{I}_m$.

2 Jacobian Spectra

One method of constructing the columns $\mathbf{k}_i$ of the $\mathbf{K}$ matrix would be to evaluate the *Jacobian* spectra, representing the expected sensitivity of the measured spectrum to perturbations in physical parameters such as surface temperature or water vapour concentration.

$$\mathbf{k}_i = \frac{\partial \mathbf{y}}{\partial x_i}$$

Such a calculation would normally be performed using a radiative transfer model, calculating the *change* in radiance, $\Delta\mathbf{y}$, expected for a unit change Δx_i in parameter i.

For a purely linear relationship between $\mathbf{y}$ and x_i the elements of $\mathbf{k}_i$ would be constant, irrespective of the atmospheric state.

For a moderately non-linear relationship it might still be possible to linearise about some suitable mean value $\mathbf{x}_0$ of the retrieved parameters

$$\begin{aligned} \mathbf{y} - \mathbf{y}_0 &= \mathbf{K}_0(\mathbf{x} - \mathbf{x}_0) \\ \Rightarrow \quad \mathbf{x} - \mathbf{x}_0 &= \mathbf{G}_0(\mathbf{y} - \mathbf{y}_0) \end{aligned}$$

where $\mathbf{y}_0, \mathbf{K}_0$ are the spectra and Jacobians calculated assume the state $\mathbf{x}_0$, and $\mathbf{G}_0$ is the pseudo-inverse of $\mathbf{K}_0$ evaluated as before.

For example, if y represented a measured radiance, which depends non-linearly (essentially as the Planck function) on the surface temperature x, we would evaluate y_0 and $\mathrm{d}y/\mathrm{d}x$ about some typical temperature, say $x_0 = 280$ K, and then apply the linearised equations to the departure of y from y_0 to obtain the departure of x from x_0.

The advantage of using a Jacobian representation of $\mathbf{K}$ is that we can derive physical parameters $\mathbf{x}$ explicitly.

The disadvantage is that we need to ensure $\mathbf{x}$ contains *all* the variables required to characterise the spectra (so, for example, not just the total water vapour column amount, but information on the vertical distribution of water vapour).

3 Singular Vectors

An alternative to the explicit Jacobian construction of $\mathbf{K}$ is to use a set of vectors representing statistically-determined patterns of variability within the measured spectra.

Using a radiative transfer model, spectra are calculated for a large number N (~100s) of atmospheric conditions. These are arranged in a $(m \times N)$ matrix $\mathbf{A}$.

Using a standard technique of *Singular Vector Decomposition* any matrix $\mathbf{A}$ (where $N > m$) can be decomposed into

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\boldsymbol{\Lambda}\mathbf{V} \\ (m \times N) &= (m \times m)(m \times m)(m \times N) \end{aligned}$$

Where $\mathbf{U}$ and $\mathbf{V}$ are matrices of orthonormal vectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix of *singular values.*

The important point here is that the singular values can be ordered so that only the largest l (typically $l \ll m$) are retained, thus compressing the information:

$$\begin{aligned} \mathbf{A} &\simeq \mathbf{U}'\boldsymbol{\Lambda}'\mathbf{V}' \\ (m \times N) &= (m \times l)(l \times l)(l \times N) \end{aligned}$$

In this application, each *row* of $\mathbf{U}$ represents a mutually orthogonal component of all the spectral patterns in $\mathbf{A}$, while the corresponding singular values in $\mathbf{\Lambda}$ represent the associated magnitude.

Thus, returning to our original linear retrieval equation, we can use $\mathbf{U}'$ as an alternative to Jacobian representation of the matrix $\mathbf{K}$

$$\mathbf{y} = \mathbf{U}'\mathbf{x}$$

where $\mathbf{x}$ now represents the set of coefficients required to fit each spectral pattern (column of $\mathbf{U}'$) to the measurements, rather than explicit physical parameters.

The advantage of such a representation of $\mathbf{K}$ is that, provided our original ensemble of N profiles covers the full range of atmospheres likely to be encountered (at high vertical resolution and considering a wide range of molecules), we do not have to decide on which parameters to retrieve. Nor do we need to concern ourself with linearity issues since the SVD is, by definition, a linear decomposition.

The disadvantage, of course, is also that we just retrieve a set of coefficients for each of our singular vectors rather than physically meaningful parameters.

4 Combined Approach

The SLR uses a retrieval scheme which combines the two different approaches.

We assume we are only interested in retrieving the concentration of one molecule x_m explicitly from its spectral signature $\mathbf{k}_m$, while all the other features in the same spectral range are accounted for by fitting singular vectors $\mathbf{U}'$.

The measured spectrum is therefore modelled as the sum of the molecular contribution $\mathbf{k}_m x_m$ and the singular vectors' contribution $\mathbf{U}'\mathbf{x}'$:

$$\mathbf{y} = \mathbf{K}\mathbf{x} \quad \Rightarrow \quad \mathbf{y} = (\mathbf{k}_m, \mathbf{U}') \begin{pmatrix} x_m \\ \mathbf{x}' \end{pmatrix}$$

The *pseudo inverse* $\mathbf{G}$ is evaluated as before, but can be similarly decomposed

$$\mathbf{x} = \mathbf{G}\mathbf{y} \quad \Rightarrow \quad \begin{pmatrix} x_m \\ \mathbf{x}' \end{pmatrix} = \begin{pmatrix} \mathbf{g}_m^T \\ \mathbf{G}' \end{pmatrix} \mathbf{y}$$

where $\mathbf{g}_m^T$ represents the top row of the $\mathbf{G}$ matrix.

Since we are *only* interested in x_m and not the fitted coefficients of the singular vectors, this can be simplified

$$x_m = \mathbf{g}_m^T \mathbf{y}$$

Thus the retrieval is effectively the dot product of the spectrum with a precomputed vector. Note that (unlike $\mathbf{k}_m$) $\mathbf{g}_m^T$ still depends on the set of 'implicitly' retrieved parameters that define the rest of the $\mathbf{x}$ vector.

The set of spectra computed for the ensemble of N atmospheres used to derive $\mathbf{U}'$ *exclude* the target molecule, otherwise its signature would be included within $\mathbf{U}'$.

While all the individual singular vectors within $\mathbf{U}'$ are, by definition, othogonal, there is just the *assumption* that the target molecule Jacobian $\mathbf{k}_m$ is distinguishable from all of these. If we take a fairly wide spectral range ($m \gg l$) this improves the chances that the signatures will all be distinct and avoid aliasing between the target molecule and the singular vectors.

The Jacobian $\mathbf{k}_m$ is taken from the difference in spectra computed with and without some standard amount of the target molecule for a 'standard' atmosphere. This is an issue which will be discussed next.

5 Jacobian variability

The SLR retrieval assumes a fixed Jacobian vector $\mathbf{k}_m$ derived from a standard atmospheric profile. The retrieval therefore assumes a fixed, linear relationship between the amplitude of the spectral signature and the amount of the target molecule in the atmosphere.

In the infrared, while the *shape* of the Jacobian may not vary too much, the *magnitude* fundamentally depends not only on the absorber amount but also on the thermal contrast between the surface and atmosphere.

Usually the surface is warmer than the troposphere so that molecular lines appear as absorption features. If the temperature of the atmosphere were the same as the surface, any spectral signatures would vanish altogether, so the 'retrieved' amount would be zero, irrespective of the actual concentration. In certain circumstances the surface may be cooler than the atmosphere, causing absorption lines to invert and become emission lines, resulting in the retrieval of negative concentrations.

It's not just the temperature structure; water vapour is highly variable and strongly absorbing so has some spectral overlap with almost all other molecules.

If we knew the atmospheric state (temperature, water vapour) we could recompute the target molecule Jacobian each time, but this would be time-consuming.

Instead we attempt to characterise the change in Jacobian amplitude by deriving a scaling factor from the spectral information.

5.1 Scaled Retrievals

It is assumed that the amplitudes of the singular vectors which are co-retrieved with the target molecule also contain enough information to characterise the change in Jacobian amplitude.

For example, the ensemble of 100s of profiles used for the SVD includes many cases where the atmosphere itself is fixed while just the surface temperature is varied, so the singular vectors should contain a significant component linked to surface temperature. Similarly if there are lines of a well-mixed molecule (i.e., near constant concentration), such as CO_2, N_2O or CH_4, these will provide some information on atmospheric temperature.

However, rather than attempting to derive an explicit relationship a statistical approach is used.

The first step is to create a large number of simulated spectra for atmospheres with a known amount of the target molecule. In fact was the same set of profiles as used for the SVD, but this time with the target molecule added.

The linear retrieval is run on each spectrum, although this time the full retrieval vector $\mathbf{x}$ is retained, including the fit coefficients. Now we have a set of N ($\sim$100s) retrievals for each of which we know

- (a) The correct absorber amount, x_t
- (b) The linearly retrieved absorber amount x_m
- (c) The fitted singular vector coefficients x_j

Noting that x_m effectively scales as $1/\mathbf{k}_m$ we can define a scale factor s which can be applied to x_m to obtain the scaled retrieval x_s, which is the best estimate of x_t

$$s = \frac{x_m}{x_s}$$

We assume the scale factor can be modelled as a linear combination of the fit coefficients x_j

$$s = \sum_j a_j x_j$$

where the coefficients a_j are obtained from a linear regression from the results from the N simulated spectra.

The linear regression will provide a better fit for large magnitudes of s than small values. For this reason the scale factor is chosen as x_m/x_s rather than its reciprocal: large values of $|s|$ correspond to cases where atmospheric conditions *enhance* the sensitivity to the target molecule (e.g., large atmosphere-surface temperature contrasts, positive or negative) and therefore where the retrieval should be more accurate. These are cases where the 'true' Jacobian would be larger than the standard Jacobian.

6 Error Covariance Matrix

For a scalar quantity x the random error is represented by a variance σ_x^2. If x itself is a simple multiple of another variable $x = gy$ then the variances in x and y are related by

$$\sigma_x^2 = g^2 \sigma_y^2$$

If we have a vector $\mathbf{x}$ then we can represent the variance of each element in the same way but we also should allow for error correlations between elements. This is most conveniently handled with a *covariance matrix* $\mathbf{S}_x$, with elements defined by

$$S_{ij} = \langle \varepsilon_i \varepsilon_j \rangle$$

where ε_i denotes the random error on element i, and $\langle \ldots \rangle$ denotes expectation value over a large number of samples. Thus the diagonal elements of $\mathbf{S}_x$ are the variances of each element of $\mathbf{x}$ while the off-diagonal elements hold the correlations (covariance matrices are necessarily symmetric, and with positive diagonal elements).

If the linear retrieval is represented by

$$\mathbf{x} = \mathbf{G}\mathbf{y}$$

the appropriate covariance matrix transform is

$$\mathbf{S}_x = \mathbf{G}\mathbf{S}_y\mathbf{G}^T$$

Comparing this with the scalar transformation of variances, note that g^2 is replaced by $\mathbf{G} \ldots \mathbf{G}^T$, the form required to ensure dimensional conformity for matrix multiplication.

7 Retrieval Error

For IASI we assume that the noise error on each spectral point is constant and uncorrelated. Neither is strictly true, but it is an adequate approximation. Thus we approximate $\mathbf{S}_y \simeq \sigma_y^2 \mathbf{I}_m$ where σ_y^2 is the noise variance on each spectral point.

Applying the covariance transform to map this into a retrieval error

$$\mathbf{S}_x = \mathbf{G}\mathbf{S}_y\mathbf{G}^T = \sigma_y^2 \left(\mathbf{K}^T\mathbf{K}\right)^{-1}$$

However if we are only interested in the variance of the target molecule retrieval this simplifies to

$$\sigma_m^2 \equiv S_{x11} = \sigma_y^2 \mathbf{g}^T \mathbf{g}$$

which is a constant.

Including the (variable) scale factor ($x_s = x_m/s$)

$$\sigma_s^2 = \frac{\sigma_m^2}{s^2}$$

Since we are dealing with scalar variances, we can also define a retrieval 'standard deviation' by taking the square roots

$$\sigma_s = \frac{\sigma_m}{s}$$

The retrieval output contains x_s and σ_s for each spectrum. However the file also contains the (constant) value of σ_m, representing the retrieval SD error that would be obtained for a scale factor $s = 1$. Thus the actual scale factor used for each retrieval can be obtained from $s = \sigma_m/\sigma_s$ — the larger the scale factor s, the smaller σ_s, and the more precise the retrieval.