

Inverse Theory Week 5: Profiles – how good is our estimate?

Hugh C. Pumphrey

November 5, 2008

1 Understanding the Covariance matrix.

We saw in the previous lecture that for a profile retrieval problem, a Maximum A-posteriori Probability (MAP) formula can give us a useful estimate of the true profile. In this lecture we look at how good our estimate is. We continue using our nadir temperature-sounding example but we let the *a priori* covariance matrix \mathbf{S}_a be a diagonal matrix – this is not necessarily the best choice, but it is the simplest. As before, we let the measurements \mathbf{y} be related to the true profile \mathbf{x}_t by $\mathbf{y} = \mathbf{K}\mathbf{x}_t + \varepsilon$, where the experimental error ε is a random vector with zero mean and covariance matrix \mathbf{S}_y . Our best estimate $\hat{\mathbf{x}}$ of the true profile is then given by

$$\hat{\mathbf{x}} = (\mathbf{S}_a^{-1} + \mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K})^{-1} (\mathbf{S}_a^{-1} \mathbf{x}_a + \mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{y}) \quad (1)$$

and its covariance matrix $\hat{\mathbf{S}}$ is given by:

$$\hat{\mathbf{S}} = (\mathbf{S}_a^{-1} + \mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K})^{-1}. \quad (2)$$

This matrix contains all that we know about how good $\hat{\mathbf{x}}$ is – in this section we look in detail at what it tells us. The simplest part of the information is contained in the diagonal elements of $\hat{\mathbf{S}}$. These are the variances of the elements of the profile - we can take their square roots to find the standard deviation of $\hat{\mathbf{x}}$. As in the previous lecture, we have taken a profile \mathbf{x}_t which we regard as true, generated some radiances \mathbf{y} from it using $\mathbf{y} = \mathbf{K}\mathbf{x}_t + \varepsilon$, and then used equations 1 and 2 to find the best estimate that we can make of the profile, given the radiances. For this example I have assumed that the measurement error is 0.1 K and that \mathbf{S}_y is diagonal. The *a priori* profile \mathbf{x}_a is 250 K at all altitudes. The diagonal elements of \mathbf{S}_a are $(80 \text{ K})^2$ and the off-diagonal elements are exactly zero. The retrieval is successful in that the retrieved profile is less than 5 K from the true one over much of the profile, while the *a priori* is up to 50 K from the true profile. On examining $\hat{\mathbf{S}}$, we find that the diagonal elements are no smaller than $(70 \text{ K})^2$, not much of a reduction on the $(80 \text{ K})^2$ of the *a priori*. The off-diagonal elements, however, have changed a great deal. We can make a good guess at what is

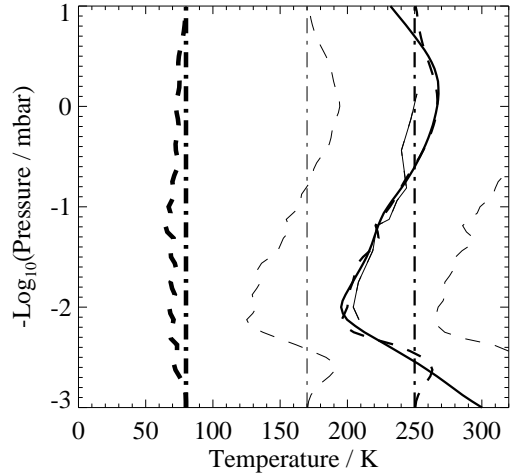


Figure 1: Errors in the Maximum A-posteriori Probability solution. The solid line is the true profile, the dashed line near it is the Maximum A-posteriori Probability solution and the dot-dash line is the *a priori*. The thin lines are the square roots of the diagonal elements of $\hat{\mathbf{S}}$ and \mathbf{S}_a , used as error bars. The very thick lines on the left of the figure are the same things plotted directly.

happening here. The radiances contain information about the large-scale structure of the profile, but not about any small-scale features. As an example, I have re-run the test, with a large wiggle added to the true profile - the results are shown in Figure 2. The large diagonal elements in $\hat{\mathbf{S}}$ are clearly telling us that there might be wiggles like this which the instrument cannot detect. In order to understand how good our solution is, we will have to take the off-diagonal elements of $\hat{\mathbf{S}}$ into account.

One way to do this is to take the eigenvalues λ_i and eigenvectors \mathbf{z}_i of $\hat{\mathbf{S}}$. These are defined by the equation $\hat{\mathbf{S}}\mathbf{z}_i = \lambda_i\mathbf{z}_i$. If $\hat{\mathbf{S}}$ is symmetric (as covariance matrices all are) then a standard result in matrix algebra states that λ_i are real and that \mathbf{z}_i are orthogonal to each other. We can therefore write $\hat{\mathbf{S}}$ as

$$\hat{\mathbf{S}} = \sum_{i=1}^n \lambda_i \mathbf{z}_i \mathbf{z}_i^T = \sum_{i=1}^n \mathbf{l}_i \mathbf{l}_i^T$$

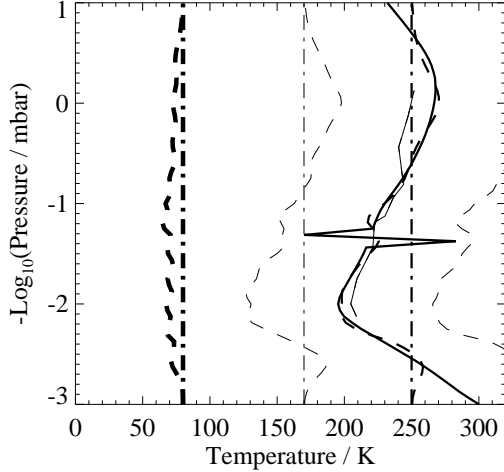


Figure 2: Exactly as Figure 1 except for the wiggle on the true profile. The instrument's limited ability to detect such features means that this wiggle does not appear fully in the retrieved profile.

where $\mathbf{l}_i = \lambda_i^{-\frac{1}{2}} \mathbf{z}_i$. This in turn means that in any one retrieval, the difference between the true and retrieved profiles is given by

$$\hat{\mathbf{x}} - \mathbf{x}_t = \sum_{n=1}^n a_n \mathbf{l}_n$$

where the a_i are random numbers with means of 0 and variances of 1. Figure 3 shows the eigenvalues and eigenvectors for our example retrieval. Note that the first 11 eigenvalues have square roots ranging from 0.2 to 10 and all the other eigenvalues have square roots of 80. What is happening here is that the error is made of n orthogonal components. Our 11 measurements have given 11 of these small to medium errors, the others all have the same large errors they had before we made the measurements. The shapes of the eigenvectors tell us what things we know well about the profile and what things we don't know well. Note that the first few eigenvalues are less than $(0.5 \text{ K})^2$ and that the corresponding eigenvectors are very smooth. This means that we know the smooth features of the profile to less than $\pm 0.5 \text{ K}$. The 10th and 11th eigenvalues are $(5.9 \text{ K})^2$ and $(10 \text{ K})^2$ and the corresponding eigenvectors have wiggles with a wavelength of 8 or 10 km at the top and bottom of the profile. We therefore know about this sort of feature in the profile only to an accuracy of $\pm 6 \text{ K}$ or 10 K . The remaining eigenvectors have very wiggly shapes: the instrument cannot tell us about features with these shapes in the profile.

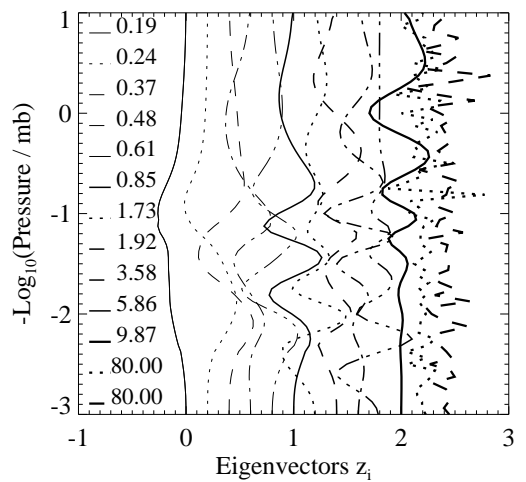
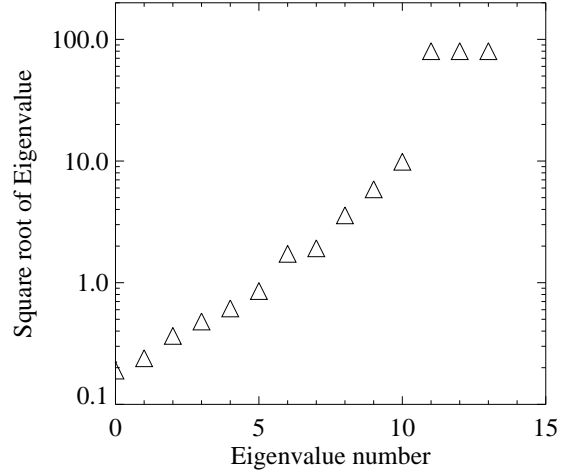


Figure 3: Upper figure shows the square roots of the 14 smallest eigenvalues of $\hat{\mathbf{S}}$, lower one shows their eigenvectors.

2 Resolution and the Averaging Kernels

The covariance matrix of the retrieved profile tells us about how precise we can expect the profile to be. In this section, we consider how much resemblance there is between the true profile and the retrieved one. We know that for a particular measurement $\mathbf{y} = \mathbf{K}\mathbf{x}_t + \varepsilon$, where ε is a random vector with zero mean and with covariance matrix \mathbf{S}_y . We also know that that $\hat{\mathbf{x}} = \mathbf{x}_a + \mathbf{D}(\mathbf{y} - \mathbf{K}\mathbf{x}_a)$. Putting these together, we find that

$$\hat{\mathbf{x}} = \mathbf{x}_a + \mathbf{D}\mathbf{K}(\mathbf{x}_t - \mathbf{x}_a) + \mathbf{D}\varepsilon. \quad (3)$$

This makes it clear why it is undesirable for the contribution matrix \mathbf{D} to have very large elements: the retrieved profile contains the measurement error multiplied by \mathbf{D} . The matrix $\mathbf{D}\mathbf{K} = \frac{d\hat{\mathbf{x}}}{d\mathbf{x}_t} = \mathbf{A}$ is called the averaging kernel matrix. The rows of \mathbf{A} are called averaging kernels and can be thought of

as smoothing functions: in the absence of noise the scalar product of the j th row of \mathbf{A} with \mathbf{x}_t gives you the j th element of $\hat{\mathbf{x}}$. The j th averaging kernel should functions have a peak at the j th element – the width of this peak is an indication of the vertical resolution of the instrument. Figure 4 shows the averaging kernels for our example retrieval. The width

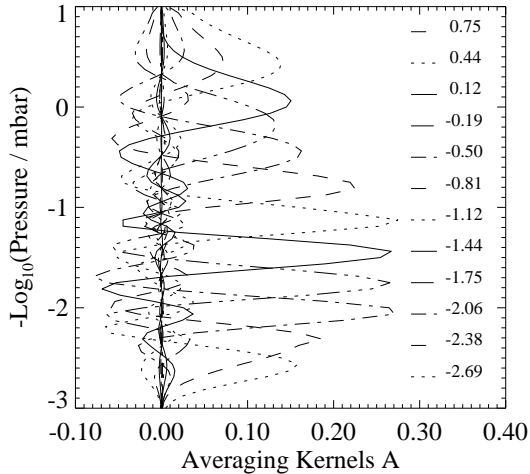


Figure 4: Averaging kernels for our example retrieval. There is one of these functions for every element in the profile. Only the kernel for every fifth element of \mathbf{x} (i.e. every fifth row of \mathbf{A}) is shown in the figure.

of the averaging kernels is a measure of the vertical resolution of the instrument. Features in the profile which are much broader than the averaging kernel width will be measured well while features much narrower than the averaging kernel width will be smoothed out in the retrieved profile. The width of the kernel can be measured in various ways. A simple measure is the full width of the peak at half of its height (often abbreviated as FWHH). This measure does not take account of the ripples on either side of the main peak. Another way to calculate the width of a function is to use the Backus-Gilbert spread. If we regard a single averaging kernel as a function $a(z)$ where z is altitude, the spread of the function about a height z_0 is defined as:

$$s(z_0) = \frac{12 \int (z - z_0)^2 a(z)^2 dz}{(\int a(z) dz)^2}. \quad (4)$$

The factor of 12 assures that the spread of a “top-hat” or “boxcar” function is equal to its width. Figures 5 and 6 show the spread and the FWHH of two simple functions. When calculating the spread of averaging kernels we take z_0 to be the height to which the averaging kernel refers. For functions like Gaussians and boxcars, the B-G spread is similar to the FWHH. For functions which oscillate and go negative, the B-G spread tends to be much larger than the FWHH. Figure 7 shows both spreads for

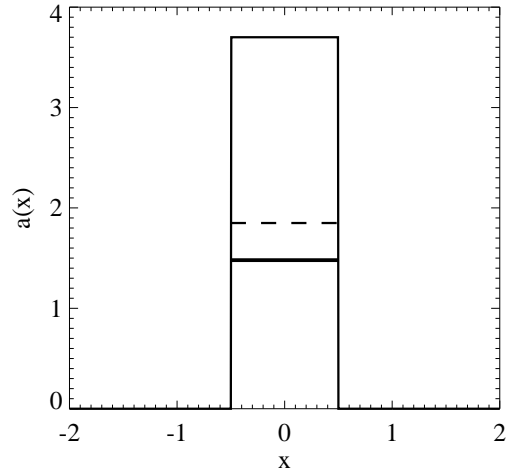


Figure 5: A top-hat function. The FWHH (thin dashed line) and B-G spread (thick line) are the same.

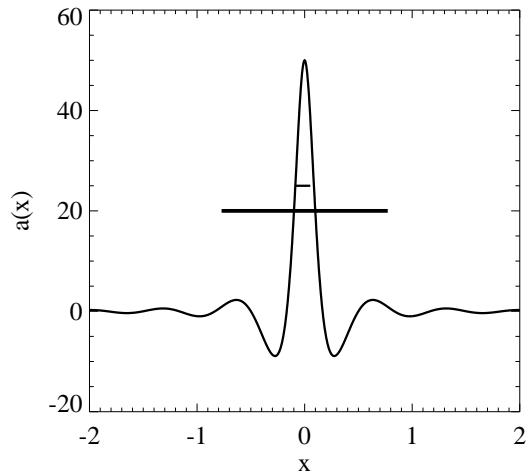


Figure 6: A wiggly function. Now the B-G spread (thick line) is much larger than the FWHH.

the averaging kernels of Figure 4. The FWHH is

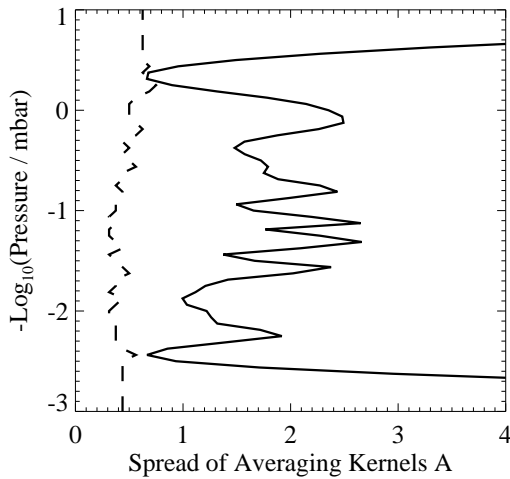


Figure 7: The width of the averaging kernels in Figure 4 as a function of height. The solid line is the Backus-Gilbert spread and the dashed line is the full width at half height.

a rather optimistic estimate of the vertical resolution. The B-G spread, on the other hand, is clearly a rather pessimistic estimate. I include it to demonstrate that when someone tells you how wide a peak is, you should make sure that you know what they mean by the width.

Equation 3 shows that the retrieved profile is the sum of:

1. The *a priori*.
2. The difference between the *a priori* and the true profile, smoothed by the averaging kernel matrix \mathbf{A} .
3. The measurement errors, multiplied by the contribution function matrix \mathbf{D} .

Writing the profile like this helps us to see the various sources of error. It becomes clearer if we rewrite Equation 3 as:

$$\hat{\mathbf{x}} - \mathbf{x}_t = (\mathbf{I} - \mathbf{A})(\mathbf{x}_a - \mathbf{x}_t) + \mathbf{D}\boldsymbol{\varepsilon}.$$

This shows that the difference between the true and retrieved profile can be broken down into two separate terms. The second term is the measurement error, the first is often termed the smoothing error and accounts for the things which the instrument cannot see.

In early instruments, which had noisy detectors and few channels, these two terms were the only ones it was important to consider. In more modern instruments the measurement noise may be small and techniques like limb sounding which have high vertical resolution, can make the smoothing error small as well. In this case, we may need to consider

other sources of error such as inaccuracies in the forward model.

We have now looked at the Maximum A-posteriori Probability solution in great detail. We have seen that it is in some sense the best solution and have examined how good the answer it gives is. We have also seen that it requires an *a priori* profile and its covariance matrix. In the next two lectures we will consider where you might obtain *a priori* information from. We will also look at some of the alternative methods that can be used for retrieving profiles, some of which do not require an *a priori*.