Inverse Theory Week 4 part 1
Representation Functions

Hugh C. Pumphrey

9th October 2007

1 Introduction.

In the previous lecture we examined the problem of retrieving a profile and saw that it was fundamentally ill-conditioned in that we want a continuous function of altitude and can only make a finite number of measurements. We reduced the problem to a set of simultaneous equations by representing the profile by a list of values at reasonably closely spaced altitudes. For a typical nadir sounder, if we space our altitudes closely enough to give a smooth-looking profile, we still have many more unknowns than equations. The very naive idea of calculating $\hat{\mathbf{x}} = \mathbf{K}^T(\mathbf{K}\mathbf{K}^T)^{-1}\mathbf{y}$ which is one of an infinite number of possible exact solutions turned out not to be very useful. In this lecture, we try something else. We have m measurements: why not try to represent the solution as a polynomial with m terms, or a Fourier series with m terms, or some other similar representation?

2 Recap of the problem.

As an example, we will re-use the nadir-sounding of temperature. Recall that the radiance emerging at the top of the atmosphere at frequency ν is given by

$$L_\nu = \int_0^\infty B_\nu(T(z))\frac{d\tau}{dz}dz. \quad (1)$$

We pick a set of m frequencies $\nu_i,\ i = 1, m$ for which the transmittance τ varies a great deal. We will consider an instrument that makes measurements of the radiation emitted by the atmosphere at these m frequencies. We choose them to be close together so that B_v is approximately the same for all frequencies. We re-write Equation 1 for our m chosen frequencies like this:

$$y_i = \int_0^\infty B_{\bar{\nu}}(T(z))K_i dz. \quad (2)$$

where $K_i = \frac{d\tau}{dz}$. We have replaced B_ν with $B_{\bar{\nu}}$ where $\bar{\nu}$ is the mean of our chosen frequencies – we will drop the subscript from now on. On the left side of Equation 2 we have m single quantities that we could measure. Buried in the right-hand side is the thing we are looking for, the temperature profile. Because we have chosen closely spaced frequencies, so that B is essentially the same for all our measurements, we would be happy with a profile of B. This is because the Planck function is invertible - if we know T we know B and vice versa. As in the previous lecture, the question is: how are we going to estimate the profile of B?

3 Representation Functions

This time, we will not immediately turn the continuous functions of height into discrete ones. We will write the profile of B as the sum of m different continuous functions of height which we will call $W_j(z)$. These could be sines and cosines, in which case we would be representing the profile by m terms of a Fourier series as shown in Figure 1.

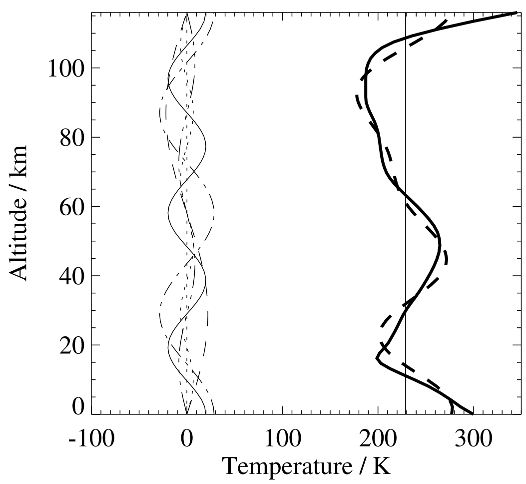

Figure 1: Temperature profile fitted using sines and cosines. The thick solid line is the original profile and the dashed one is the Fourier series. The temperature is represented over the altitude range $0 \leq z \leq Z$ by a function of the form $w_0 + w_1\cos(2\pi z/Z) + w_2\sin(2\pi z/Z) + w_3\cos(4\pi z/Z) + w_4\sin(4\pi z/Z) + \cdots$. The thin lines are the individual terms in this function.

They could be polynomials as shown in Figure 2. They could simply be triangles, which would be equivalent to choosing m levels and joining them up with straight lines as shown in Figure 3. (Note that this is essentially what we did in the last lecture when we chose to have as many points in the profile as we had measurements – we should expect the same sort of trouble.) None of these examples can represent an arbitrary profile exactly but all could produce a useful approximation. We can write B in all of these cases like this:

$$\hat{B}(z) = \sum_{j=1}^{m} w_j W_j(z). \quad (3)$$

(We put a hat on B to emphasise that this is not the true profile, but our estimate of it.) Now all we need are the

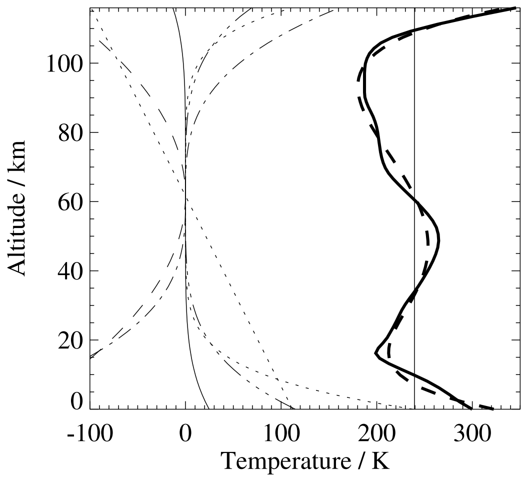

Figure 2: As Figure 1 but with temperature profile fitted using a polynomial. The temperature is represented by a function of the form $w_0 + w_1 x + w_2 x^2 + w_3 x^3 + \cdots$ where $x = z - 50$.

m coefficients w_j which we will try to find from our m measurements. Substituting this into Equation2 gives:

$$y_i = \sum_{j=1}^{m} w_j \int_0^\infty W_j(z) K_i(z) dz = \sum_{j=1}^{m} C_{ij} w_j. \quad (4)$$

We can calculate the elements C_{ij} of the square matrix $\mathbf{C}$ because we know K_i and we chose $W_j(z)$ ourselves. Equation4 is just a set of m simultaneous equations with the w_j being the m things we want to know. We can rewrite it as $\mathbf{y} = \mathbf{Cw}$ where $\mathbf{y}$ is a vector containing the measurements y_i and $\mathbf{w}$ is a vector containing the coefficients we want. We can solve for $\mathbf{w}$ by multiplying by the inverse of the matrix $\mathbf{C}$: $\mathbf{w} = \mathbf{C}^{-1}\mathbf{y}$ and then calculate the profile from Equation 3. You will get a chance to try this in the practical exercise to see how well it works.

In order to apply this method in a practical situation, it will be necessary to represent the profile $\hat{B}(z)$ as a list of numbers at a set of pre-determined heights, just as we did in the previous lecture. We are representing a function as a vector which we will write $\hat{\mathbf{x}}$. We choose this vector to have n elements where n is large enough to make our predetermined heights be closely spaced, but not so large that the vectors fill our computer. We represent the functions W_j as vectors $\mathbf{w}_j$ and the functions K_j as vectors $\mathbf{k}_j$ in the same way. We then make an $m \times n$ matrix $\mathbf{K}$ whose rows are the vectors $\mathbf{k}_j$ and an $n \times m$ matrix $\mathbf{W}$ whose columns are the vectors $\mathbf{w}_j$. Equation 3 becomes $\hat{\mathbf{x}} = \mathbf{Ww}$, $\mathbf{C}$ is given by $\mathbf{C} = \mathbf{KW}$ and Equation 4 becomes $\mathbf{y} = \mathbf{Cw} = \mathbf{K}\hat{\mathbf{x}}$. Our solution is therefore

$$\hat{\mathbf{x}} = \mathbf{W}(\mathbf{KW})^{-1}\mathbf{y} = \mathbf{Dy}. \quad (5)$$

Let's try this to see how well it works.

As before, we take a temperature profile, calculate some radiances from it using the formula $\mathbf{y} = \mathbf{Kx} + \varepsilon$ and then apply Equation 5 to try to recover the original profile from the radiances. Also as before, we will try both with $\varepsilon = 0$

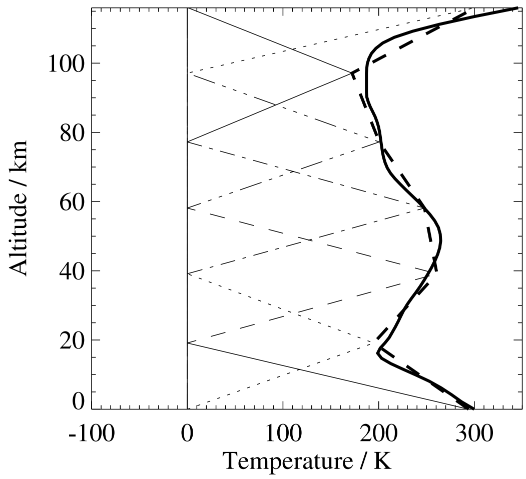

Figure 3: As Figure 1 but with temperature profile fitted using straight line segments (triangular basis functions).

and with ε being a random variable with a known standard deviation. We will use a polynomial representation for $\mathbf{W}$, as in Figure 2. The results are shown in Figure 4. It looks

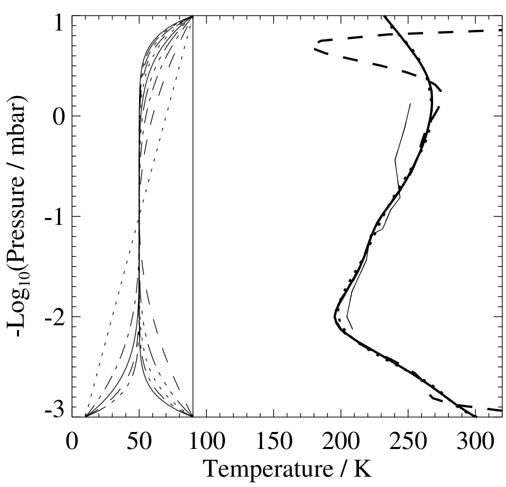

Figure 4: The true temperature profile is a thick solid line. The dotted line is a polynomial fit to this. The thin line is the nine measurements, plotted at the peaks of the influence functions. The dashed line is the polynomial fit which agrees exactly with the measurements, as given by Equation 5. The thin lines on the left are the eleven functions $\mathbf{w}_j$.

quite successful except at the top of the profile. However we have assumed that we could measure the profile exactly. Now let's add a little noise to the measurements - we'll choose a measurement error of 1 K Now our retrieval looks like Figure 5.

The matrix $\mathbf{D}$ is called the contribution matrix and its columns are called contribution functions. It is what you multiply the measurements by to get your estimate of the profile. There are several properties which we would like $\mathbf{D}$ to have:

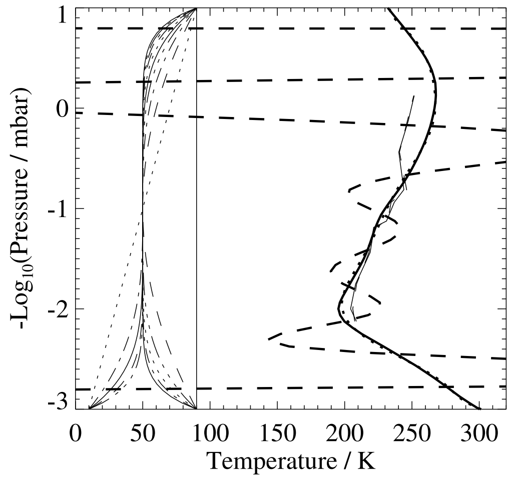

Figure 5: As Figure 4 but we have added noise of 1 K to the measurements. The noisy measurements are shown by the thin dashed line but this is so close to the measurements with no noise that you can't really see it in the figure. Nevertheless, that small amount of noise has made the retrieved profile more or less useless.

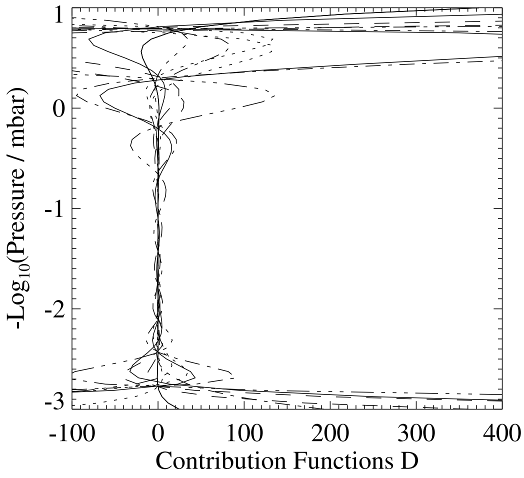

Figure 6: Contribution functions for retrieval by polynomial representation. The very large values at top and bottom mean that small amounts of measurement noise will lead to huge fluctuations in the retrieved profile in those regions.

- The elements of $\mathbf{D}$ should not be too large. (Experimental errors in $\mathbf{y}$ will be multiplied by them).
- Any negative elements should be relatively small. (You don't expect an increase in temperature to be associated with a decrease in radiance.)
- The columns $\mathbf{d}_j$ should be singly-peaked functions. The peak in $\mathbf{d}_j$ should be at the same height as the peak in $\mathbf{k}_j$.

The contribution functions for our example retrieval are shown in Figure 6; it is clear that they do not have the features that we would like.

We (hopefully) chose $\mathbf{W}$ sensibly so the bad behaviour is coming from the $(\mathbf{KW})^{-1}$, which is nearly singular. We saw in the previous lecture how the fact that the influence functions overlap a great deal makes $\mathbf{K}$ nearly singular. It is possible that by making a better choice of $\mathbf{W}$, we can make $\mathbf{KW}$ less nearly singular. This is still not wholly satisfactory, of course, as we don't know if we have made the best choice of $\mathbf{W}$. As an example, we try the cosine and sine functions to see if they are any better than the polynomial. The retrieval test is shown in Figure 1 and the contribution functions in Figure 8. It is clear that the Fourier series representation is a better choice than the polynomials were. The retrieval test behaves less wildly at the top and bottom of the profile and the contribution functions are smaller.

One interesting possibility is to use the influence functions themselves as our representation functions. In this case, $\mathbf{W} = \mathbf{K}^T$, so

$$\hat{\mathbf{x}} = \mathbf{W}(\mathbf{KW})^{-1}\mathbf{y} = \mathbf{K}^T(\mathbf{KK}^T)\mathbf{y} = \mathbf{Dy}. \quad (6)$$

... but this is the exact solution we tried in the previous lecture and which didn't work. We can now see why: At the top and bottom of the profile, all the influence functions are zero, but we are trying to represent a profile which has a value of about 250 K. Why not try using 10 of the 11 influence functions, and a constant? We show the results in Figure 9 and the contribution functions in Figure 10.

This works better than the trig functions, but it is still feels like a bit of an ad-hoc way of solving the problem. Our choice of a representation basis is clearly important and there are many choices we could make. It would be better to avoid having to make that choice.

Now, it is clear that our measurements cannot tell us everything about the profile, but they do tell us something. We never know everything about the profile, but we do know more after we make the measurements than we did before. In the next lecture we consider more carefully what information is carried by the measurements.

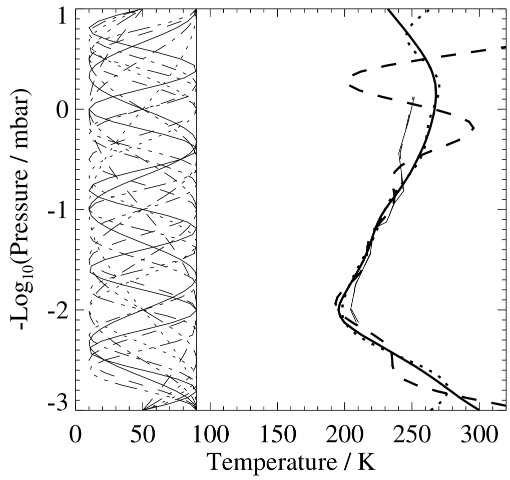

Figure 7: As Figure 5, but we have used a Fourier series to represent the profile.

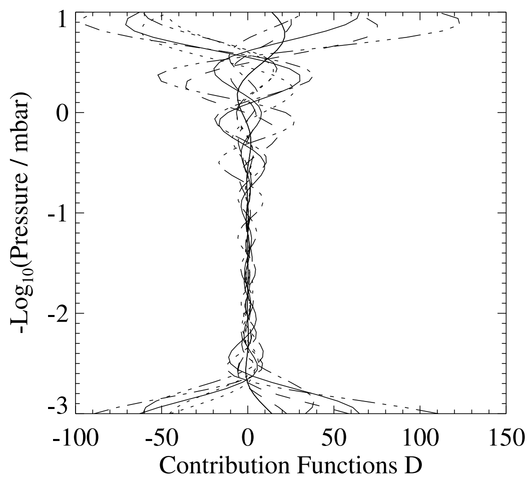

Figure 8: As figure 6, but for the Fourier series representation.

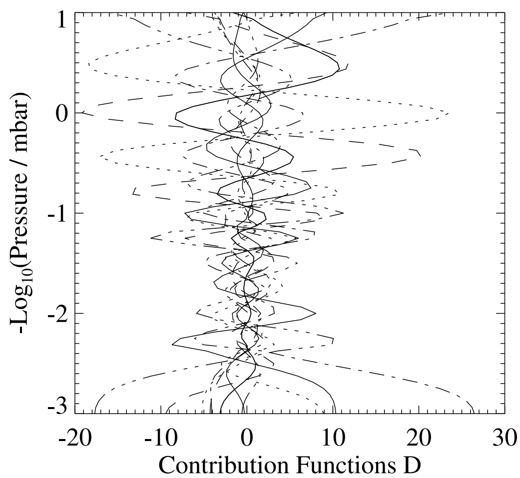

Figure 10: As figure 6, but for the case that uses the influence functions as representation functions.

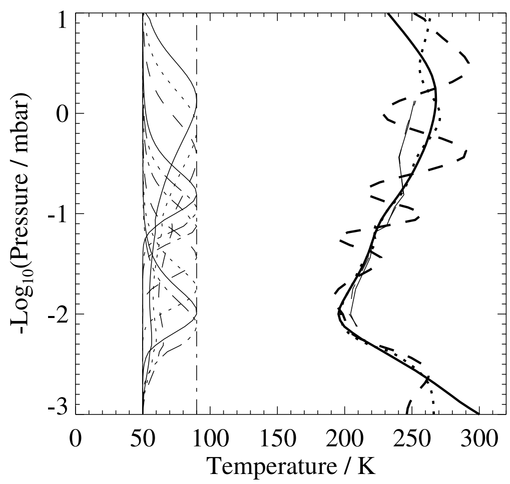

Figure 9: As figure 5, but using the influence functions as representation functions.