

Inverse theory: Week 3

Atmospheric Profiles

Hugh C. Pumphrey

2nd October 2007

1 Introduction.

1.1 What these lectures are about.

Remote sensing of the atmosphere is in a sense a harder problem than sensing of the surface because the atmosphere is three-dimensional. The most obvious thing that we might want to know about the atmosphere is its temperature. We would want to know this as a function of height and this is the big difference between sensing the atmospheric temperature and the sea-surface temperature. With the sea surface, the temperature in one “pixel” is a single value – one piece of information. With the atmosphere, it is a function, with a different value at each height – an infinite number of pieces of information.

Many other atmospheric quantities are also functions of altitude. Typical examples are the wind velocity and the mixing ratio of ozone and various other trace molecules. We call the graph of any such quantity as a function of height a *profile*. The bulk of this course is about how we go about estimating profile quantities from remotely-sensed data.

1.2 Notation.

I shall attempt to stick fairly rigidly to using the following typefaces:

- Italics for scalars, like this: x , or like this: X .
- Lower case bold for vectors, like this: \mathbf{x} .
- Upper case bold for matrices, like this: \mathbf{X} .

2 The physics of sensing a profile.

There are several ways to measure the atmosphere from space. We can use a range of wavelengths from the microwave to the ultra-violet. We can use limb-sounding or nadir-sounding geometries. The radiation we detect may be thermal emission from the air, or it may be radiation from the sun which is being absorbed or scattered by the air. Details of these measurement methods are taught in *Fundamentals of Remote Sensing* and in *Radiative Transfer*, but we will review a few of them here. As our main example we will consider measuring a temperature profile by nadir sounding.

2.1 Nadir sounding of temperature.

The spectral radiance at emerging at the top of the atmosphere at frequency ν is given by

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

We have ignored the radiation coming from the ground - we suppose for the moment that the atmosphere absorbs all of this. The vertical co-ordinate, z , may be geometric height or a co-ordinate related to pressure. Now, we pick a set of m frequencies ν_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_ν is approximately the same for all frequencies. Recall that the transmittance is the fraction of the radiation at some position which reaches some other position. We might try to choose frequencies so that some of the transmittances looked like Figure 1.

For a given frequency, the radiation reaching space does not come from very low in the atmosphere because all the radiation emitted there is absorbed again. Nor does it come from very high in the atmosphere, where the air does not absorb at all at this frequency and so it does not emit either. The radiation comes mostly from the region in the middle, where τ is changing with altitude. A suitable part of the spectrum will have an absorption coefficient which varies rapidly with frequency and will also have most of that absorption due to a gas which has a well-known and constant mixing ratio. One commonly used spectral region is 13-15 μm in the thermal infrared; this is used by the HIRS instrument on the TIROS series of polar orbiting weather satellites. The absorption in this region is mainly due to Carbon Dioxide. Another region which is becoming popular is 50-65 GHz, which is in a spin-rotation band of the Oxygen molecule. This region is used by the MSU (Microwave Sounding Unit) on older TIROS satellites and by AMSU (Advanced MSU) on brand-new ones. (The first AMSU was launched in May 1998).

Now, let us re-write Equation 1 for our m chosen frequencies like this:

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

where $K_i(z) = \frac{d\tau}{dz}$ at the i th frequency. The radiance as a continuous function of frequency, N_ν has been replaced by the measured radiance y_i in the i th of our

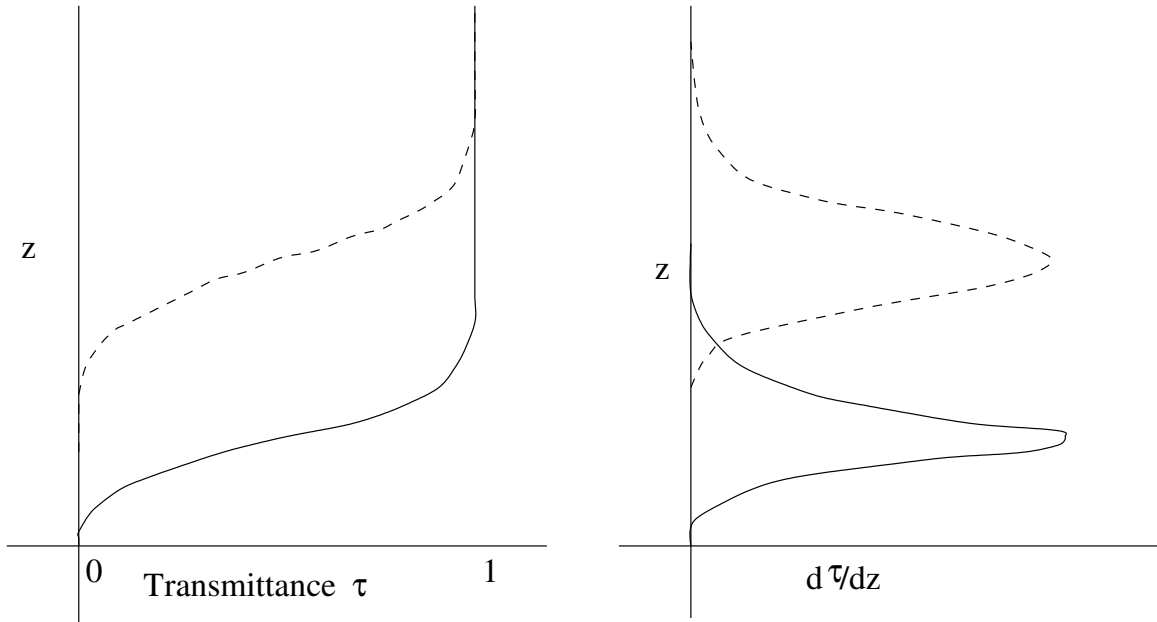


Figure 1: (left) Transmittance from altitude z to space for two channels in an idealised nadir temperature sounder. The dashed line represents a channel where the air absorbs more strongly than in the channel represented by the solid line. (right) Weighting functions $K_i(z) = \frac{d\tau}{dz}$ for the same two channels.

chosen frequencies. (I use y throughout these notes to represent things we can measure, in a general sense.) 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. A function like equation 2 which calculates measurable quantities based on the state of the atmosphere is called a *forward function* and a computer program that implements such a function is called a *forward model*. 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. The question now is: how are we going to dig out the profile of B ? We can't possibly find it exactly because there are infinitely many possibilities and we have only m pieces of information.

2.2 Limb sounding.

Not all instruments for sounding the atmosphere look vertically downwards. For sounding the stratosphere and mesosphere in particular it is advantageous to use a limb-sounding¹ geometry, as shown in figure 2.

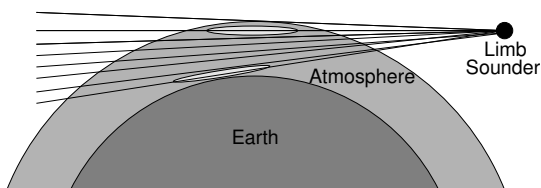


Figure 2: Sketch to show limb sounding geometry

¹“Limb” is a term astronomers use for the edge of the disc of the sun or a planet as it appears in a telescope.

The instrument is scanned across the atmosphere. Because the density of the atmosphere decreases exponentially with height, most of the radiation received at a given scan position comes from altitudes close to the lowest point on the line of sight – this is called the tangent point. This effective source region is several hundred kilometres long, making it possible to detect very weak emitted signals. The price paid for this is that the horizontal resolution is very poor compared to that of a nadir sounder. A typical limb sounder may make measurements at several tens of tangent heights during one scan, and may also make measurements at more than one frequency. This means that the list of measurements y_i from which we will attempt to estimate profiles of temperature (and perhaps composition) usually contains a few hundred elements, many more than is common for a traditional nadir sounder. Limb sounders have been built which use both the infra-red and microwave regions of the spectrum.

2.3 Occultation sounding.

This is variation on limb sounding which uses light from the sun (or a star), as shown in Figure 3.

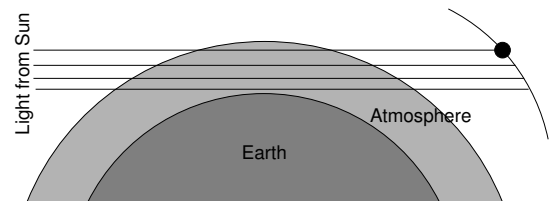


Figure 3: Sketch showing occultation sounding.

The instrument measures how much of the sun's light is absorbed by the atmosphere. By watching the sun as it sets, the instrument detects uncontaminated solar radiation, followed by radiation which has passed

through greater and greater depths of the atmosphere. All measurements are divided by the uncontaminated one to remove any effects due to the source. Occultation instruments may record over 100 measurements during a single sunset, at tangent heights spaced 1 km apart, or less.

3 First Attempts.

3.1 Setting up the problem.

All the examples have the same fundamental difficulty: we are trying to estimate a continuous function from a finite number of measurements. The usual fix is to replace the continuous function with an array of closely spaced points. This clearly introduces some error if you want to know the value of the function between two of the points, but if this causes a problem, you can always use more points.

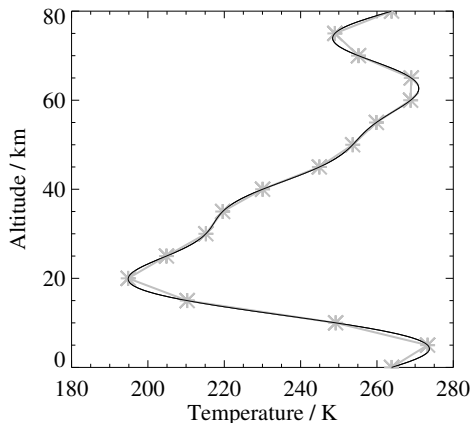


Figure 4: A continuous function of altitude (Temperature in this case, shown in black) can be represented by values at a finite number of heights (shown in grey).

For the nadir temperature sounder, by choosing to replace the continuous function with a discrete one, we have changed the forward function from something like this²:

$$y_i = \int_0^{\infty} x(z)K_i(z)dz \quad (3)$$

which relates a list of numbers y_i to a continuous function $x(z)$ to something like this:

$$y_i = \sum_{j=1}^n K_{ij}x_j$$

which relates two lists of numbers. By considering these numbers as vectors, we can write this as

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

This is instantly recognisable as a set of linear simultaneous equations. We generally call the vector \mathbf{y} the

²This is just the forward function for the nadir temperature sounding case (Equation 2) with the profile of the Planck function replaced by $x(z)$.

measurement vector, while the vector \mathbf{x} which describes the state of the atmosphere is called the *state vector*. The only remaining problem that we might have to consider is how many equations we have, and how many variables. We note in passing that the forward function for limb sounding measurements (and indeed most other remote sounding problems) is not linear like this. However, in all cases, we can write the forward function as some general function of \mathbf{x} , like this:

$$\mathbf{y} = F(\mathbf{x}),$$

expand it as a Taylor series, like this:

$$\mathbf{y} \approx F(\mathbf{x}_L) + \mathbf{K}(\mathbf{x} - \mathbf{x}_L) + \dots$$

where the matrix \mathbf{K} has elements

$$K_{ij} = \frac{d(F(\mathbf{x}))_i}{dx_j}.$$

We can then choose new variables $\mathbf{x} \leftarrow \mathbf{x} - \mathbf{x}_L$ and $\mathbf{y} \leftarrow \mathbf{y} - F(\mathbf{x}_L)$, casting the problem in the form $\mathbf{y} = \mathbf{K}\mathbf{x}$. Thus, the idea of reducing the problem to a set of simultaneous equations is quite general, although the example in Equation 3. only applies to the nadir sounding of temperature. The rows of \mathbf{K} are often called *influence functions* in the remote sensing literature.

3.2 How many equations in how many variables?

The number of simultaneous equations that we have is just the same as the number of elements in \mathbf{y} , which we shall call m . This is generally fixed by the design of the instrument. The number of unknown variables, n , is the number of elements we chose to have in \mathbf{x} . We have three choices:

- $m = n$. This makes the simultaneous equations solvable unless \mathbf{K} turns out to be singular.
- $m > n$. This gives us more equations than unknowns. We can try to find a least-squares solution.
- $m < n$. This gives us less equations than unknowns. There may be an infinite number of solutions, although it is also possible that there is no solution.

In order to try any of these methods out, we need a test case. We will consider an imaginary instrument and construct a \mathbf{K} matrix for it. We will then be able to take a true profile \mathbf{x}_t , use it to calculate some imaginary measurements $\mathbf{y} = \mathbf{K}\mathbf{x}_t + \varepsilon$, where ε is the measurement noise. This is a sample from a random variable and we generate it with a random number generator, so it will be different each time we do the calculation, but we give it a known covariance matrix \mathbf{S}_y . We can then take our calculated “measurements” and attempt to estimate or *retrieve* a profile of \mathbf{x} from them. Figure 5 sets out this testing cycle, which we will be applying to all of the retrieval formulae that we encounter over the next few weeks. Note that by doing this, we are

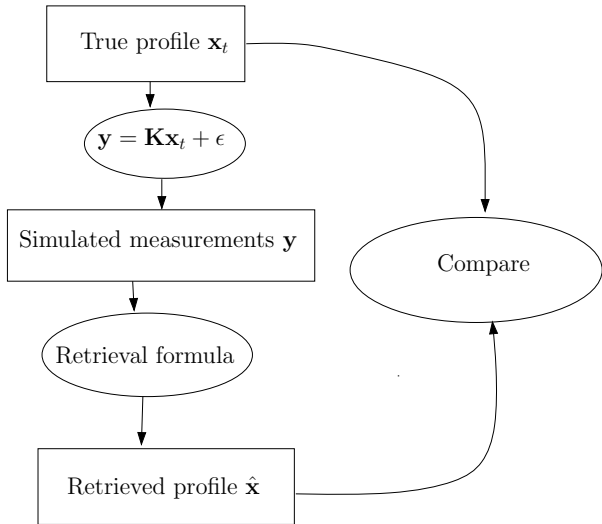


Figure 5: Flowchart of the testing procedure for a retrieval formula.

removing any problems to do with how well our forward model represents the real atmosphere. We are concerned purely with the nature of the simultaneous equations.

3.3 Nadir sounding example.

We consider an imaginary instrument with 11 channels, closely spaced, near a frequency of 56 GHz. The transmissivity and the influence functions are shown in Figures 6 and 7, which should be compared with Figure 1.

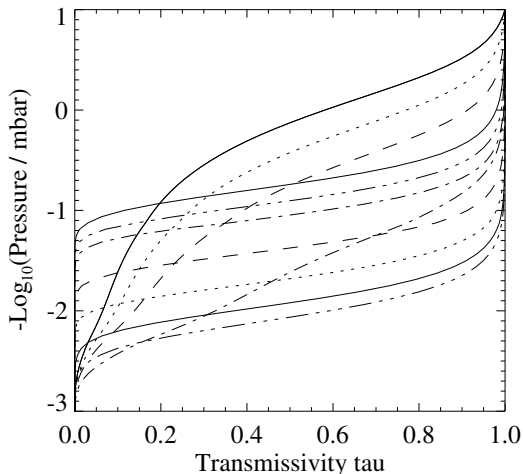


Figure 6: Transmissivity τ as a function of height for our imaginary instrument. Each line represents a different channel of the instrument. The vertical coordinate is approximately proportional to height and covers a range of 0-64 km.

As we noted earlier, the radiance in each channel is a weighted mean of the temperature profile, i.e. it is the temperature of the region of the atmosphere surrounding the peak of the influence function for that channel. We can see this to some extent in Figure 8, in which the true temperature profile \mathbf{x}_t is plotted. I have also

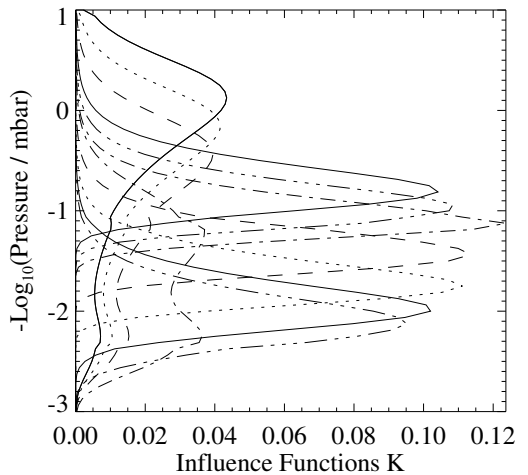


Figure 7: The influence functions $K_i = \frac{d\tau}{dz}$ as a function of height for our imaginary instrument. Each line represents a different channel of the instrument.

put the value of the radiance in each channel (in units of Kelvin) plotted at the altitude of the peak of the influence function for that channel.

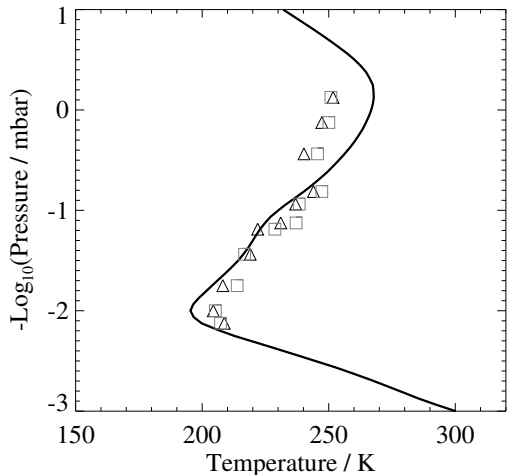


Figure 8: Test profile of temperature (line) and “measured” radiances (symbols). The triangles are radiances with no noise added, the squares have 3K of random noise.

3.3.1 Case for which $m = n$

The obvious way to solve this problem is to choose \mathbf{x} so that it has the same number of elements as \mathbf{y} . As with subsequent cases we write $\hat{\mathbf{x}} = \mathbf{D}\mathbf{y}$, where in this case $\mathbf{D} = \mathbf{K}^{-1}$. The true profile \mathbf{x}_t and the estimated profile $\hat{\mathbf{x}}$ are shown in Figure 9.

At first sight, this looks as if it will be a reasonable retrieval formula. We repeat the test, and, instead of setting ϵ to zero, we let each element of ϵ be a random variable with a standard deviation of 3 K. The retrieved profile (also shown in Figure 9) does not have errors of 3 K, instead they are about 100 K. A small error in the measurements leads to a huge error in the retrieved

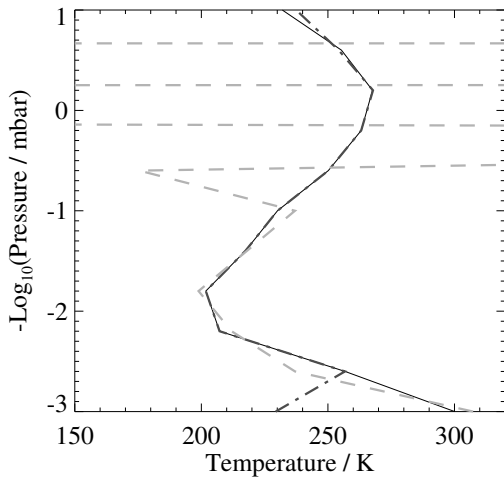


Figure 9: Attempt to retrieve profile by inverting \mathbf{K} . The thin line is the true profile, the dot-dash line is the estimated profile when no noise is added to \mathbf{y} (ε has a standard deviation of 0) and the grey dashed line is the case where ε has a standard deviation of 3 K.

profile.

To see why this happens it is worth taking another look at the situation we looked at in the first lecture, in which we had two measurements and two unknowns as we can look at this situation graphically.

Recall that we considered the following simultaneous equations:

$$y_1 = k_{11}x_1 + k_{12}x_2 \quad (4)$$

$$y_2 = k_{21}x_1 + k_{22}x_2 \quad (5)$$

which we can write as $\mathbf{y} = \mathbf{K}\mathbf{x}$. However you go about solving these equations it is formally the same as writing $\mathbf{x} = \mathbf{K}^{-1}\mathbf{y}$. Graphically, the solution is where the two lines cross in Figure 10. If $k_{21} = k_{11}$ and $k_{22} = k_{12}$ then the two equations are exactly the same. In this case, we have only one line on our graph and we cannot locate a unique solution. This is the same thing as saying that the matrix \mathbf{K} is singular *i.e.* that we cannot calculate its inverse, \mathbf{K}^{-1} . In real remote sensing problems we often have the situation that the matrix is nearly singular but not quite. The example sketched is like this. Formally, there is a solution, but it may be hard to find in practice. If there is some measurement noise, then the situation is as sketched in the lower part of Figure 10. Note how the fact that \mathbf{K} is nearly singular and hence that our two lines are nearly the same line means that a small amount of noise can really mess up our chances of making a good estimate of \mathbf{x} .

Another way to look at this is to note that $\hat{\mathbf{x}} = \mathbf{D}\mathbf{y} = \mathbf{D}(\mathbf{K}\mathbf{x}_t + \varepsilon) = \mathbf{x}_t + \mathbf{D}\varepsilon$. If \mathbf{D} has any very large elements in (which it will have if \mathbf{K} is nearly singular) then small values of ε will produce large errors in $\hat{\mathbf{x}}$.

Going back to our attempt to estimate a profile from a nadir sounder, we need to ask why the matrix \mathbf{K} is nearly singular in most cases. We get a reasonable clue by looking at the weighting functions in Figure 7. We can see that they overlap to a large extent and therefore if we have a sufficient number of channels

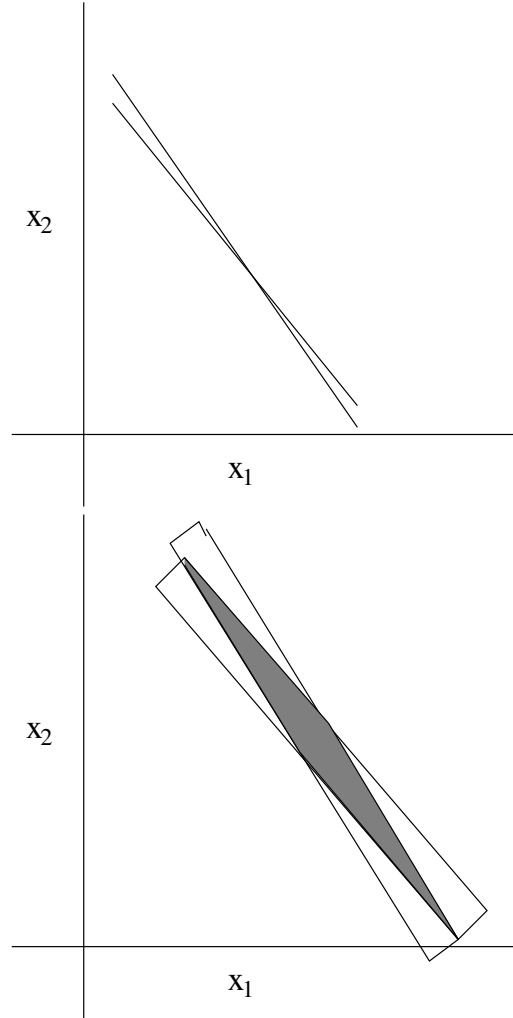


Figure 10: A problem which is well posed but only just. The solution is where the two lines cross. A problem which is well-posed but only just and in which there is measurement noise. Graphically the noise makes our lines have a finite thickness. It only takes a little bit of noise to ensure that the solution could have a large range of values for x_1 and x_2 .

which are sufficiently closely spaced, then one channel tells us nearly the same thing as the ones on either side. How are we going to get round this problem?

3.3.2 Case where $m > n$

One possibility is to use an even smaller number of elements in \mathbf{x} and to make a least-squares fit solution. As discussed in the earlier lectures, this means $\hat{\mathbf{x}} = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{y}$, *i.e.* $\mathbf{D} = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T$. Results are shown in Figure 11, Again, it works fine if there is no measurement error. The effect of measurement error is less than it was with $m = n$, but still much larger than we would like, and we have obtained the improvement at the cost of worse vertical resolution. Besides, how do we know that we have chosen a sensible number of levels n and the best heights to put our levels at?

3.3.3 Case where $m < n$

If we let $m < n$ then our choice of levels matters less. However, there will be infinitely many solutions

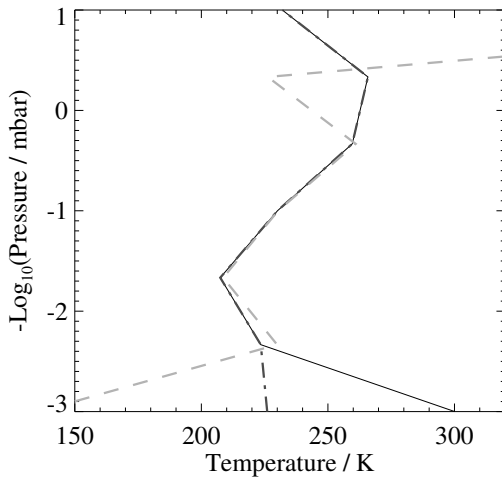


Figure 11: Attempt to retrieve profile by a least-squares fit. The thin line is the true profile, the dot-dash line is the estimated profile when no noise is added to \mathbf{y} (ε has a standard deviation of 0) and the grey dashed line is the case where ε has a standard deviation of 3 K.

– infinitely many profiles for which $\mathbf{y} = \mathbf{K}\mathbf{x}$. We know that we can find one of these using the formula $\mathbf{D} = \mathbf{K}^T(\mathbf{K}\mathbf{K}^T)^{-1}$. It is shown in Figure 12. It is

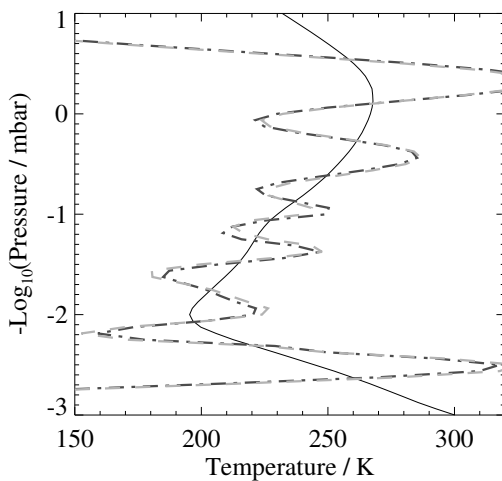


Figure 12: The exact solution $\hat{\mathbf{x}} = \mathbf{K}^T(\mathbf{K}\mathbf{K}^T)^{-1}\mathbf{y}$. The thin line is the true profile, the dot-dash line is the estimated profile when no noise is added to \mathbf{y} (ε has a standard deviation of 0) and the grey dashed line is the case where ε has a standard deviation of 3 K.

hardly surprising that of the infinite number of possible solutions, we have not picked a useful one. In the next lecture, we try harder, by seeing if we can make a better choice of how we represent the profile.

4 Points to remember:

- By representing the profile as a vector of values closely spaced in height, we reduce the problem of estimating the profile from remotely sensed data

to solving a set of simultaneous equations.

- If we choose to have enough values in our profile to represent a continuous function well, we have more unknowns than equations.
- If we choose to represent our profile so we have the same number of unknowns as we have equations, then we can find an exact solution.
- That exact solution is often useless because the matrix \mathbf{K} we invert is almost singular – this means that a tiny measurement error gives a huge error in the retrieved profile.
- The reason that \mathbf{K} is nearly singular is that the influence functions overlap, so that one channel is telling us nearly the same thing as another channel.