# A Commoner's Approach to 1dvar

Larry McMillin March 21, 2005

The intent of this presentation is to make the concept of optimal sounding retrievals as intuitive as possible. To do so, I am starting from simplest model I can think of and building toward the retrieval equation. As considerations are added to make it close to the retrieval equation, it gets increasingly complicated. It does not follow all the way the through and at the one point, I have to start with conventional approach to get to the exact equation. Strictly following this approach only gets you close. Never the less, it does give useful insight that complements the more conventional approach.

Most people understand the concept of averaging to reduce errors. If the errors are equal, it is easy. If one is more accurate, then obviously you want to weight the one that is more accurate more. This is all that both 1dvar retrievals and the data assimilation used in numerical models are. The question is how much more? We start with the Equation for weighted average

$$T_{averaged} + \varepsilon_{averaged} = W_1 \cdot (T_1 + \varepsilon_1) + W_2 \cdot (T_2 + \varepsilon_2) \quad .$$

But we also have to constrain the weights so that

$$W_1 + W_2 = 1.0$$

to keep the units. Otherwise we would start with inches and end Kellicams or some equally obscure unit. Combining equations gives

$$T_{averaged} + \varepsilon_{averaged} = W_1 \cdot (T_1 + \varepsilon_1) + (1. - W_1) \cdot (T_2 + \varepsilon_2)$$

We want to find the weight that gives the minimum variance. To do so, we need to calculate the variance by squaring the equation

$$(T_{averaged} + \varepsilon_{averaged})^2 = W_1^2 \cdot (T_1 + \varepsilon_1)^2 + (1. - 2W_1 + W_1^2) \cdot (T_2 + \varepsilon_2)^2$$

Since we are interested in the variance over a sample, we note that since the errors are random, cross products between $\varepsilon$ and $T$ have an expected value of zero and can be dropped out. A full expansion would still include terms with $T$ but we know they will drop later so we will skip them to keep things simple. Then taking the derivative of $\varepsilon_{averaged}$ with respect to $W_1$ gives

$$\partial(\varepsilon_{averaged}) / \partial W_1 = 2 \cdot W_1 \cdot \varepsilon_1^2 - 2 \cdot \varepsilon_2^2 + 2 \cdot W_1 \cdot \varepsilon_2^2 = 0.$$

Solving for $W_1$ and $W_2$ gives

$$W_1 = \varepsilon_2^2 / (\varepsilon_1^2 + \varepsilon_2^2) \qquad W_2 = \varepsilon_1^2 / (\varepsilon_1^2 + \varepsilon_2^2)$$

and manipulation gives an alternate form

$$W_1 = 1.0 / \varepsilon_1^2 / (1.0 / \varepsilon_1^2 + 1.0 / \varepsilon_2^2) \qquad W_2 = 1.0 / \varepsilon_2^2 / (1.0 / \varepsilon_1^2 + 1.0 / \varepsilon_2^2)$$

It is interesting to note that what the weights really do is scale the quantities so that the scaled variances are equal so that equal variances are added just like the simple case we started with. This is demonstrated below

$$W_2 \cdot \varepsilon_2^2 = W_1 \cdot \varepsilon_1^2 = \varepsilon_1^2 \cdot \varepsilon_2^2 / (\varepsilon_1^2 + \varepsilon_2^2)$$

To make things consistent with retrieval literature, we will rename the errors by calling

$b = \varepsilon_1$ and $y = \varepsilon_2$ to get

$$T_{averaged} + \varepsilon_{averaged} = W_1 \cdot (T_1 + b) + W_2 \cdot (T_2 + y)$$

Up till now, we have been considering like things, as in temperatures with the same units. But if the units were different, one would have to have multiplied by a scaling factor to convert the units. The scaling factor also affects the values of the weights. For soundings, the scaling factor is often a function. But the common methods linearize the function about the point defined by the current estimate of the truth. This is usually the background for the first iteration and the current retrieval for later iterations. In any case, for a given iteration, the function is evaluated and the value is a constant just like any other scaling factor. We will denote the scaling factor by the letter $K$. Including the scaling factor gives

$$T_{averaged} + \varepsilon_{averaged} = W_1 \cdot (T_1 + b) + W_2 \cdot K \cdot (T_2 + y)$$

Note that the errors are also scaled so the scaling factor is included in the weight as well. We will talk more about that later, but first we will subtract a value from both sides. It could be any value, but we will choose the value given by the background. Doing so gives

$$T_{averaged} + \varepsilon_{averaged} - g = W_1 \cdot (T_1 + b) + W_2 \cdot K \cdot (T_2 + y) - g$$

Since the weights sum to 1.0, the guess can be moved inside the brackets to give

$$T_{averaged} + \varepsilon_{averaged} - g = W_1[(T_1 + b) - g] + W_2[K(T_2 + y) - g]$$

Converting the guess to the same units as $T_2$ gives and remembering that we set the guess equal to the value of $T_1$ gives

$$T_{averaged} + \varepsilon_{averaged} - g = W_1[0] + W_2[K(T_2) - g]$$

but note that zero still has an error distribution given by $b$ and the second term has one given by $y$

This equation can also be written as

$$T_{averaged} + \varepsilon_{averaged} = g + W_2 K[T_2 + y - f(g)]$$

Remember that for our scalar case without any units conversion, $W_2$ can be written as

$$W_2 = \varepsilon_2^{-2} \cdot (\varepsilon_1^{-2} + \varepsilon_2^{-2})^{-2}$$

But since we are now using $y$ this becomes

$$W_2 = y_2^{-2} \cdot (y_1^{-2} + y_2^{-2})^{-2}$$

When the units are included it becomes

$$W_2 = (K \cdot y_2^2 \cdot K)^{-1} \cdot (b_1^{-2} + (K \cdot y_2^2 \cdot K)^{-1})^{-1}$$

Which results in

$$T_{averaged} + \varepsilon_{averaged} = g + (b_1^{-2} + (K \cdot y_2^2 \cdot K)^{-1})^{-1} \cdot (K \cdot y_2^2 \cdot K)^{-1} \cdot K \cdot [T_2 - f(g)]$$

which can be simplified to

$$T_{averaged} + \varepsilon_{averaged} = g + (b_1^{-2} + (K \cdot y_2^2 \cdot K)^{-1})^{-1} \cdot (K \cdot y_2^2)^{-1} \cdot [T_2 - f(g)]$$

There is yet another consideration. That is the definition of $y$. Up till now, it has been the error in the measurements. When we subtracted a guess, $y$ remained unchanged. However, when $y$ becomes $f(g)$ the process involves the calculation of radiances from the guess state (i.e. temperature profile). Even though we know guess perfectly (it is just a set of numbers) the process of calculating radiances from a known state is subject to errors due the calculation process. The definition of $y$ is now changed to include errors in the both the measurement and the forward calculation. One of the ramifications of this is that if one of these is known much more accurately than the other, then the largest error becomes the limiting factor. In other words, the resources spent on the instrument and the forward calculation have to be kept in balance. Getting one very accurate while ignoring the other provides no benefit.

So far we have been considering a simple scalar problem and we could have used brightness temperatures. But we want to assimilate radiances. This presents some additional considerations.

One is that radiances are averages of the values we want. Suppose I tell you I have a surface and its area is x and ask you to draw it. You know something because you know the area, but you need some information about the shape. If I slice the object into slabs and give the area of the slabs, you can do better. If I give you information about the smoothness or some other feature of the shape you can do even better. The closest example is the problem of having

a profile of layer temperatures. You can never get a unique solution for the level temperatures.

For soundings this means that the satellite measurements are never provide the complete information for a profile. If a solution is attempted without a constraint, a ridiculous solution is obtained. This comes about for 3 reasons

1. As in the case of the layer temperatures for every profile, there a number of profiles with fine vertical structure that give the same answer. One needs to know something to select the right one.

2. The measurements are not independent. This means the temperature for a given level contributes to the radiances for several channels.

3. The measurements contain errors. Suppose one channel has an error and the others are all perfect. If the measurements are assumed to error free, a profile needs to be found that keeps the radiances for all the channels except the noisy one constant and changes the one for the noisy channel to satisfy the erroneous value. Without the other channels, the noise would correspond to a small temperature change over levels contributing to that channel. However, since the values for the other channels must be kept constant, the only way to do this is to make a large change for some of the levels. Then since the other channels also see this level, other levels have to be changed to cancel the effect on the other channels. The net effect

of all this is to produce one of the noisy solutions that are possible because of condition 1.

This means that some information in addition to the radiances is needed to produce a solution. Examples are:

1. a guess profile
2. a constraint such as smoothness of the profile.

The smoothness constraint will be left as an exercise. But general approach is derived as follows. Suppose we are retrieving the temperatures for a profile. One of the values that might be used as an estimate for one layer is the temperature at a nearby layer. When this is done, the value for a layer enters on one side of the equation for itself and on the other for other layers. When the algebra is done, the equations for layers become connected.

In any case, once a constraint is selected, it has to be applied in the right weights to produce a good solution. Using it either too strongly or too weakly is not optimal. This is what the retrieval process does.

The second consideration is that the measurements are not temperatures or temperatures scaled to another unit, but rather the product of a sum of weighted temperatures with the weights being determined by the atmospheric transmittances. Note that these weights are not the ones that we have been discussing earlier, but rather part of the units conversion factor. The problem is that we have

radiance and we want to get temperatures. There are several ways to do this:

1. convert radiances to brightness temperatures.
    a. but the transmittances become functions of temperature
2. use the radiances in the retrieval and put the conversion in the retrieval system
    a. The conversion factors are then the elements of a Jacobian matrix

This model has some implications. Note that the weights are determined by the error.

These equations have some implications. Suppose we want to let the satellite measurements have the maximum impact on a forecast model. Below are some alternatives and their effects.

1. Have a good guess. This gives a good retrieval. But if the guess is good, then why not just use it. In fact a retrieval produced with a good guess will be pretty much independent of the measurements.
2. Produce a retrieval based on a separate guess from the one used in the model. Since the weight is determined by the combination of the errors in both measurements and the guess, the error used to determine the weights is increased and the measurements have less influence on the model than it would if only one guess were used.

3. Produce a retrieval based on the same guess as the one used in the model. This can be done if the error is counted correctly. The trap is that exactly the same guess with exactly the same error enters twice with weights based on the assumption that there are two guess values with different errors that are independent. Since the errors are absolutely identical, the errors are not weighted correctly if the weights are based on the assumption that they differ. The errors have to be counted correctly.
4. Directly assimilate the radiances. This allows the maximum influence of the radiances on the forecast because only one guess is involved. Step 3 can produce the same result with more work. Remember the 3 laws of thermodynamics. You can't get something for nothing, you can't break even, and you can't even come close. This is similar. Doing better than assimilating the radiance based on a single background is like inventing a perpetual motion machine.

If the solution is iterated it must de done right. Some errors that have been made are:

1. iterate the solution and use the result of the previous solution as the guess for the next iteration. This has the effect increasing the weight given to the satellite measurements at each step. When too much weight is given to the measurements, noise in the measurements and the forward calculation are amplified.

2. Do a retrieval based on one guess and supply the retrieval to a model using a separate guess. This has the effect of down weighting the effect of the measurements because weight is given to two guesses and part of the additional guess weight is taken from the measurement contribution.

Solutions are iterated because the accuracy of the linearization step depends on the accuracy of the retrieval. The way to do the iteration is to

1. use the same guess for the background each time.
2. Iterate the state (profile) used to calculate the Jacobian

## The Penalty Function Approach

Up till now, we have been working with a scalar model. Although one can get close to the final solution by analogy, there are a couple of little details that, at least so far, don't make it to the exact solution. So for matrices using the full solution, the usual penalty function approach is now given following Eyre 1991 as given in his 2002, ECMWF Meteorological Training Course Lecture Series. The penalty function can be written as

$$J(x) = -0.5 \cdot (x - x^b)^T \cdot B^{-1} \cdot (x - x^b) - 0.5(y^m - y\{x\})^T \cdot Y^{-1} \cdot (y^m - y\{x\})$$

differentiating with respect to x and denoting it by $J'(x)$ gives

$$J'(x) = B^{-1} \cdot (x - x^b) - K(x)^T \cdot Y^{-1} \cdot (y^m - y\{x\}) = 0.$$

where $K(x)$ is the Jacobian matrix containing the derivatives $dy\{x\}/d(x)$.

Integrating gives

$$y\{x\} = y\{x^b\} + K \cdot (x - x^b)$$

and noting that for the linear case $K(x) = K(x^b) = K$ and substituting gives

$$x = x^b + (B^{-1} + K^T \cdot Y^{-1} \cdot K)^{-1} \cdot K^T \cdot Y^{-1} \cdot (y^m - y\{x^b\})$$

An equivalent form which can be more stable and more efficient can be obtained using the matrix identity

$$(I + AB)^{-1} = I - A(I + BA)^{-1} B$$

To get

$$x = x^b + B \cdot K^T \cdot (K \cdot B \cdot K^T + Y)^{-1} \cdot (y^m - y\{x^b\})$$

The difference is that in one case the matrix product that occurs in the inverse is singular by itself and the inverse is stable only because of the matrix that is added to it. This is because one form has the number of channels as the inner dimension and the number of temperatures, water vapor amounts, etc. being retrieved as the outer dimension. The second form has the dimensions reversed.

Finally note the similarity between this retrieval and the older versions using brightness temperatures to convert from radiances to temperature

$$x = x^b + B \cdot K^T \cdot (K \cdot B \cdot K^T + Y)^{-1} \cdot (y^m - y\{x^b\})$$

$$x = x^g + S \cdot A^T \cdot (A \cdot S \cdot A^T + Y)^{-1} \cdot (y^m - y\{x^g\})$$

Note that $A$ takes the place of $K$ since the conversion from radiance to temperature is done in the Jacobian. The newer Jacobian is simply a more accurate way to do the conversion. In the form using weighting functions as $A$, the weighting functions ignore the dependence of the weighting functions on the state vectors on the grounds that it is small compared to the change in brightness temperature. While this assumption is true, it slows the convergence and can lead to a slightly different answer. It was a good approximation in its day, but it is not a good approach to use now.