Least Squares Method (Kalnay, 5.3)

We start with a toy model example, the two temperatures problem.

We use two methods to solve it, a sequential and a variational approach, and find that they are equivalent: they yield identical results.

The problem is important because the methodology and results carry over to multivariate OI, Kalman filtering, and 3D-Var and 4D-Var assimilation.

If you fully understand the toy model, you should find the more realistic application straightforward.

Statistical estimation

Introduction. Each of you: Guess the temperature in this room right now. How can we get a best estimate of the temperature?

⋆ ⋆ ⋆

The best estimate of the state of the atmosphere is obtained by combining prior information about the atmosphere (background or first guess) with observations.

In order to combine them *optimally*, we also need statistical information about the errors in these pieces of information.

As an introduction to statistical estimation, we consider the simple problem, that we call the two temperatures problem:

Given two independent observations T_1 and T_2, determine the best estimate of the true temperature T_t.

Simple (toy) Example

Let the two observations of temperature be

$$\left.\begin{aligned} T_1 &= T_t + \varepsilon_1 \\ T_2 &= T_t + \varepsilon_2 \end{aligned}\right\}$$

[For example, we might have two *iffy* thermometers].

The observations have errors ε_i, which we don't know.

Let $E(\ )$ represent the expected value, i.e., the average of many similar measurements.

We assume that the measurements T_1 and T_2 are unbiased:

$$E(T_1 - T_t) = 0\,, \qquad E(T_2 - T_t) = 0$$

or equivalently,

$$E(\varepsilon_1) = E(\varepsilon_2) = 0$$

We also assume that we know the variances of the observational errors:

$$E(\varepsilon_1^2) = \sigma_1^2 \qquad E(\varepsilon_2^2) = \sigma_2^2$$

We next assume that the errors of the two measurements are uncorrelated:

$$E(\varepsilon_1 \varepsilon_2) = 0$$

This implies, for example, that there is no systematic tendency for one thermometer to read high ($\varepsilon_2 > 0$) when the other is high ($\varepsilon_2 > 0$).

⋆ ⋆ ⋆

The above equations represent the statistical information that we need about the actual observations.

We estimate T_t as a linear combination of the observations:

$$T_a = a_1 T_1 + a_2 T_2$$

The analysis T_a should be unbiased:

$$E(T_a) = E(T_t)$$

This implies

$$a_1 + a_2 = 1$$

T_a will be the best estimate of T_t if the coefficients are chosen to minimize the mean squared error of T_a:

$$\sigma_a^2 = E[(T_a - T_t)^2] = E\{[a_1(T_1 - T_t) + a_2(T_2 - T_t)]^2\}$$

subject to the constraint $a_1 + a_2 = 1$.

This may be written

$$\sigma_a^2 = E[(a_1\varepsilon_1 + a_2\varepsilon_2)^2]$$

Expanding this expression for σ_a^2, we get

$$\sigma_a^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2$$

To minimize σ_a^2 w.r.t. a_1, we require $\partial\sigma_a^2/\partial a_1 = 0$.

Naïve solution: $\partial\sigma_a^2/\partial a_1 = 2a_1\sigma_1^2 = 0$, so $a_1 = 0$.

Similarly, $\partial\sigma_a^2/\partial a_2 = 0$ implies $a_2 = 0$.

We have forgotten the constraint $a_1 + a_2 = 1$.

So, a_1 and a_2 are not independent.

Substituting $a_2 = 1 - a_1$, we get

$$\sigma_a^2 = a_1^2\sigma_1^2 + (1 - a_1)^2\sigma_2^2$$

Equating the derivative w.r.t. a_1 to zero, $\partial\sigma_a^2/\partial a_1 = 0$, gives

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \qquad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

Thus, we have expressions for the weights a_1 and a_2 in terms of the variances (which are assumed to be known).

We define the precision to be the inverse of the variance. It is a measure of the accuracy of the observations.

Note: The term precision, while a good one, does not have universal currency, so it should be defined when used.

⋆ ⋆ ⋆

Substituting the coefficients in $\sigma_a^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2$, we obtain

$$\sigma_a^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

This can be written in the alternative form:

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1{}^2} + \frac{1}{\sigma_2{}^2}$$

Thus, *if the coefficients are optimal, the precision of the analysis is the sum of the precisions of the measurements.*

Variational approach

We can also obtain the same best estimate of T_t by minimizing a *cost function*.

The cost function is defined as the sum of the squares of the distances of T to the two observations, weighted by their observational error precisions:

$$J(T) = \frac{1}{2}\left[\frac{(T - T_1)^2}{\sigma_1^2} + \frac{(T - T_2)^2}{\sigma_2^2}\right]$$

The minimum of the cost function J is obtained is obtained by requiring $\partial J/\partial T = 0$.

⋆ ⋆ ⋆

Exercise: Prove that $\partial J/\partial T = 0$ gives the same value for T_a as the least squares method.

The control variable for the minimization of J (i.e., the variable with respect to which we are minimizing the cost function) is the temperature.

For the least squares method, the control variables were the weights.

The equivalence between the minimization of the analysis error variance and the variational cost function approach is important.

This equivalence also holds true for multidimensional problems, in which case we use the covariance matrix rather than the scalar variance.

It indicates that OI and 3D-Var are solving the same problem by different means.

⋆ ⋆ ⋆

Example: Suppose $T_1 = 2 \quad \sigma_1 = 2 \quad T_2 = 0 \quad \sigma_2 = 1$.

Show that $T_a = 0.4$ and $\sigma_a = \sqrt{0.8}$.

⋆ ⋆ ⋆

$$\sigma_1^2 + \sigma_2^2 = 5$$

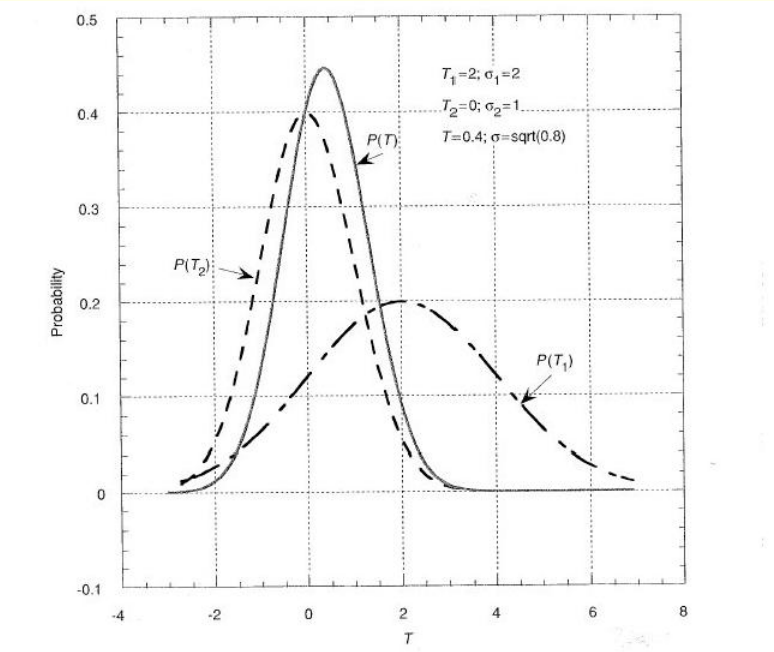
$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{1}{5} \qquad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} = \frac{4}{5}$$

CHECK: $a_1 + a_2 = 1$.

$$T_a = a_1 T_1 + a_2 T_2 = \frac{1}{5} \times 2 + \frac{4}{5} \times 0 = 0.4$$

$$\sigma_a^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{4 \times 1}{4 + 1} = 0.8$$

This solution is illustrated in the next figure.



The probability distribution for a simple case.

The analysis has a pdf with a maximum closer to T_2, and a smaller standard deviation than either observation.

Conclusion of the foregoing.

Simple Sequential Assimilation

We consider the 'toy' example as a prototype of a full multivariate OI.

Recall that we wrote the analysis as a linear combination

$$T_a = a_1 T_1 + a_2 T_2$$

The requirement that the analysis be unbiassed led to $a_1 + a_2 = 1$, so

$$T_a = T_1 + a_2(T_2 - T_1)$$

Assume that one of the two temperatures, say $T_1 = T_b$, is not an observation, but a background value, such as a forecast.

Assume that the other value is an observation, $T_2 = T_o$.

We can write the analysis as

$$T_a = T_b + W(T_o - T_b)$$

where $W = a_2$ can be expressed in terms of the variances.

The least squares method gave us the optimal weight:

$$W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

When the analysis is written as

$$T_a = T_b + W(T_o - T_b)$$

the quantity $(T_o - T_b)$ is called the ***observational innovation***, i.e., the new information brought by the observation.

It is also known as the observational increment (with respect to the background).

The analysis error variance is, as before, given by

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2} \qquad \text{or} \qquad \sigma_a^2 = \frac{\sigma_b^2 \sigma_o^2}{\sigma_b^2 + \sigma_o^2}$$

The analysis variance can be written as

$$\sigma_a^2 = (1 - W)\sigma_b^2$$

⋆ ⋆ ⋆

Exercise: Verify all the foregoing formulæ.

⋆ ⋆ ⋆

We have shown that the simple two-temperatures problem serves as a paradigm for the problem of objective analysis of the atmospheric state.

Collection of Main Equations

We gather the principal equations here:

$$T_a = T_b + W(T_o - T_b)$$

$$W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

$$\sigma_a^2 = \frac{\sigma_b^2 \sigma_o^2}{\sigma_b^2 + \sigma_o^2} = W\sigma_o^2$$

$$\sigma_a^2 = (1 - W)\sigma_b^2$$

These four equations have been derived for the simplest scalar case ...

... but they are important for the problem of data assimilation because they have exactly the same form as more general equations:

The least squares sequential estimation method is used for real multidimensional problems (OI, interpolation, 3D-Var and even Kalman filtering).

Therefore we will interpret these four equations in detail.

The first equation

$$T_a = T_b + W(T_o - T_b)$$

This says:

The analysis is obtained by adding to the background value, or first guess, the innovation (the difference between the observation and first guess), weighted by the optimal weight.

The second equation

$$W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

This says:

The optimal weight is the background error variance multiplied by the inverse of the total error variance (the sum of the background and the observation error variances).

Note that the larger the background error variance, the larger the correction to the first guess.

⋆ ⋆ ⋆

Look at the limits: $\sigma_o^2 = 0$; $\sigma_b^2 = 0$.

The third equation

The variance of the analysis is

$$\sigma_a^2 = \frac{\sigma_b^2 \sigma_o^2}{\sigma_b^2 + \sigma_o^2}$$

This can also be written

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}$$

This says:

The precision of the analysis (inverse of the analysis error variance) is the sum of the precisions of the background and the observation.

The fourth equation

$$\sigma_a^2 = (1 - W)\sigma_b^2$$

This says:

The error variance of the analysis is the error variance of the background, reduced by a factor equal to one minus the optimal weight.

It can also be written

$$\sigma_a^2 = W\sigma_o^2$$

All the above statements are important because they also hold true for sequential data assimilation systems (OI and Kalman filtering) for multidimensional problems.

In these problems, in which T_b and T_a are three-dimensional fields of size order 10^7 and T_o is a set of observations (typically of size 10^5), we have to replace expressions as follows:

- error variance $\Longrightarrow$ error covariance matrix
- optimal weight $\Longrightarrow$ optimal gain matrix.

Note that there is one essential *tuning* parameter in OI:

It is the ratio of the observationalvariance to the background error variance:

$$\left(\frac{\sigma_o}{\sigma_b}\right)^2$$

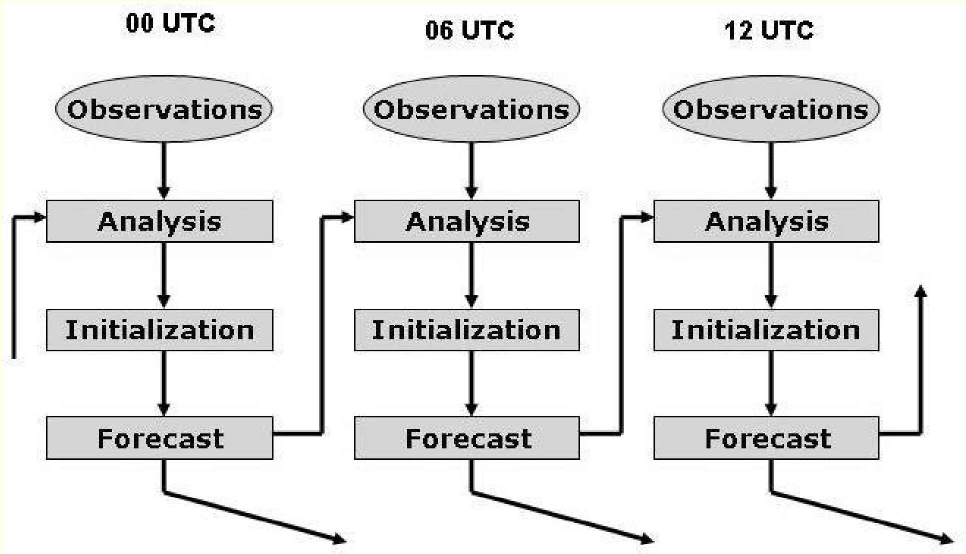
Application to Analysis

If the background is a forecast, we can use the four equations to create a simple sequential analysis cycle.

The observation is used once at the time it appears and then discarded.

Assume that we have completed the analysis at time t_i (e.g., at 06 UTC), and we want to proceed to the next cycle (time t_{i+1}, or 12 UTC).

The analysis cycle has two phases, a forecast phase to update the background T_b and its error variance σ_b^2, and an analysis phase, to update the analysis T_a and its error variance σ_a^2.



Typical 6-hour analysis cycle.

Forecast Phase

In the forecast phase of the analysis cycle, the background is first obtained through a forecast:

$$T_b(t_{i+1}) = M\left[T_a(t_i)\right]$$

where M represents the forecast model.

We also need the error variance of the background.

In OI, this is obtained by making a suitable simple assumption, such as that the model integration increases the initial error variance by a fixed amount, a factor a somewhat greater than 1:

$$\sigma_b^2(t_{i+1}) = a\sigma_a^2(t_i)$$

This allows the new weight $W(t_{i+1})$ to be estimated using

$$W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

Analysis Phase

In the analysis phase of the cycle we get the new observation $T_o(t_{i+1})$, and we derive the new analysis $T_a(t_{i+1})$ using

$$T_a = T_b + W(T_o - T_b)$$

The estimates of σ_b^2 is from

$$\sigma_b^2(t_{i+1}) = a\sigma_a^2(t_i)$$

The new analysis error variance $\sigma_a^2(t_{i+1})$ comes from

$$\sigma_a^2 = (1 - W)\sigma_b^2$$

It is smaller than the background error.

After the analysis, the cycle for time t_{i+1} is completed, and we can proceed to the next cycle.

Reading Assignment

Study the Remarks in Kalnay, §5.3.1