

## 8. Kalman Filtering

In this section we will discuss the technique called *Kalman filtering*. Kalman filtering is a technique for estimating the state of a dynamical system from a sequence of incomplete or noisy measurements. The measurements need not be of the state variables themselves, but must be related to the state variables through a linearizable functional relationship. It is a solution to the linear-quadratic-Gaussian problem, which is the problem of estimating the instantaneous state of a linear dynamical system which is perturbed by Gaussian white noise-by using measurements of observables that are linearly related to the state, but are corrupted by white noise. It is optimal in the usual quadratic mean-square sense. It is one of the greatest innovations in statistical estimation theory and is widely used in a variety of applications.

Rudolf Emil Kalman born in Budapest in 1930, and emigrated with his family to the US in 1943. He went to MIT and completed a Ph.D. at Columbia in 1957. He studied Wiener's work on filtering and came up with the idea of applying it in state space and equating the expectation operator with a projection in a finite dimensional state space, and the Kalman filter resulted. The Wiener filter is used with analog electronics, whereas the Kalman filter is ideally suited to dealing with digital data. The Kalman filter was used as part of the onboard guidance system on the Apollo Project, one of its first applications. A large fraction of guidance and process control systems include Kalman filters of one stripe or another-there are many variations and extensions now.

In numerical weather prediction the functional relationship used to connect the observables and the state vector is the numerical weather prediction model, augmented by other models that may relate observables to state variables. Kalman filtering allows an initial analysis to be made in an optimal way from observations taken at random times (the model is used to do an optimal temporal interpolation that includes the atmospheric dynamics and physics as represented in the NWP model), at random locations (the spatial interpolation is done with the model equations), and from a diverse set of observations (rawinsondes, aircraft, buoys, satellites, etc.). In recent years NWP models have begun to incorporate satellite radiance measurements, rather than derived variables such as temperature or humidity. In this way the high resolution details that the model may have constructed (correctly), but which can not be resolved by remote sensing, are not corrupted by the satellite retrieval. Rather the model adjusts its (e.g.) temperature in a way that is consistent with the observed satellite radiance, but which disturbs the model's useful information in the minimum way. For this we need to include the forward model that predicts the radiances from the model temperatures as part of the Kalman filter. A full-blown Kalman filtering data assimilation scheme has yet to be implemented, but so-called "four-dimensional variational" assimilation schemes now coming on line can be considered to be practical approximations to a Kalman filter approach. The first application of Kalman filters in meteorology that I am aware of was to the optimal estimation of zonal Fourier coefficients of temperature in the stratosphere from observations taken from a polar-orbiting satellite (Rodgers, 1977).

We will start with very simple concepts and work up to the idea of a Kalman filter. We will not go into the actual implementation of a Kalman filter in an NWP assimilation scheme, but will concentrate on the basic concepts, and a concrete example or two.

### Outline:

#### 1. *Combination of estimates*

Suppose we have two estimates of the state of a system. How can we “best” combine the two estimates to obtain a combined estimate that is “better” than both individually? How do we estimate the improvement and the quality of the new estimate?

#### 2. *Sequential estimation*

How can we “best” improve our estimate of a “state vector” by adding new observations? How do we keep track of how good the new estimates are?

#### 2. *Kalman Filtering*

How can we do sequential estimation and also take into account a linearized version of the known mathematical relationship between the observations and the state variables or among the state variables themselves?

### Problems: General Case

In general objective analysis problems we have several complexities and imperfections to deal with and need to make the most of what we have to work with.

- a. Observations are noisy and not complete, *e.g.*, 1 point on a map of geopotential, 1 velocity component, 1 point on a latitude circle- all with different error characteristics.
- b. Observations are not all taken at synoptic times, but are distributed inconveniently in time and space, *e.g.* satellite measurements, aircraft measurements, etc.
- c. Observables are not part of the chosen state vector, but can be related to it functionally.
  1. State variables are parameters of an orbit; observable is the position of the satellite at a particular time.
  2. State vector is a functional map of temperature (*e.g.* Fourier coefficients, spherical spectral coefficients); Observables are temperature at points irregularly distributed in time and space.

3. State vector includes regions where we seldom or never get observations (over oceans, in boundary layer, in Southern Hemisphere), but we have a global model that relates variables at all locations.
4. State vector includes variables that are not measured, but we have a model that relates all variables. For example, we have lots of radiance measurements over the oceans from satellites that are related to the temperature profile, but few direct wind observations.

### 8.1 Combining Estimates

In order to more easily understand how the principles of Kalman filtering work, we can first consider the simple problem of combining two or more estimates of which the accuracy can be estimated. An old, but readable source on this is Deutsch (1965).

Suppose we have a measurement estimate,  $\hat{\mathbf{x}}_1$  of a vector  $\mathbf{x}$  with which we associate some error  $\mathbf{q}_1$

$$\mathbf{q}_1 = \hat{\mathbf{x}}_1 - \mathbf{x}$$

Suppose  $\mathbf{q}$  is unbiased,

$$E\{\mathbf{q}_1\} = 0$$

where  $E\{\mathbf{x}\}$  is the expected value of  $\mathbf{x}$ , for example the average value.

The covariance matrix of the error is

$$\mathbf{S}_1 = E\{\mathbf{q}_1\mathbf{q}_1^T\} \quad (8.2)$$

Suppose we have a second unbiased measurement estimate of  $\mathbf{x}$ ,  $\hat{\mathbf{x}}_2$  with error  $\mathbf{q}_2$  and error covariance matrix  $\mathbf{S}_2$ .

$$\mathbf{q}_2 = \hat{\mathbf{x}}_2 - \mathbf{x}, E\{\mathbf{q}_2\} = 0, \mathbf{S}_2 = E\{\mathbf{q}_2\mathbf{q}_2^T\} \quad (8.3)$$

#### Optimality and Risk:

In optimization theory we normally choose a measure of a bad (good) consequence and try to minimize (maximize) it. The classical example is the mean squared error  $E\{\mathbf{q}_1^2\}$

which is minimized and leads to least squares regression, which was first used by Gauss(1777-1855). This is only one possible definition of a risk function, but it is by far the most common and useful. In general we want to choose a risk function that:

1. Is a good measure of a bad consequence whose minimization will lead to a desirable result.
2. Has a form that is convenient to manipulate and the minimization of which can be easily obtained. This is where quadratic error shines, since the solution for the optimization problem is linear.

### 8.1.1 Scalar case

Here we consider the example of combining two estimates of a scalar quantity. Once we have combined two estimates, we can combine an infinite number sequentially if we can estimate the error of the combined estimate at every step. This forms the basis for sequential estimation, which is the basic process of which Kalman filtering is a sterling example.

$$E(\mathbf{q}_1^2) = \sigma_1^2, E(\mathbf{q}_2^2) = \sigma_2^2 \quad (8.4)$$

$$R(x) = \underbrace{(x_1 - x)^2 \sigma_1^{-2}}_{\downarrow} + \underbrace{(x_2 - x)^2 \sigma_2^{-2}}_{\downarrow} \quad (8.5)$$

$$\frac{\text{Squared error of 1}}{\text{Expected error of 1}} \quad \frac{\text{Squared error of 2}}{\text{Expected error of 2}}$$

If  $x_1$  has a much higher than expected rms error than  $x_2$ , then we should not weight the deviation of  $x_1$  from  $x$  as heavily as that of  $x_2$ . That is,  $x_1$  should not dominate the minimization since we know it is very noisy anyway. We don't want to draw too closely to noisy data.

Next we minimize  $R$  with respect to the unknown  $x$ .

$$\frac{\partial R}{\partial x} = 0 = -2(x_1 - x)\sigma_1^{-2} - 2(x_2 - x)\sigma_2^{-2} \quad (8.6)$$

The choice of  $x$  which minimizes the risk, i.e., the optimal estimate of  $x$ ,  $\hat{x}$  is the  $x$  for which

$$\frac{(\hat{x}_1 - \hat{x})}{\sigma_1^2} + \frac{(\hat{x}_2 - \hat{x})}{\sigma_2^2} = 0 \quad (8.7)$$

or

$$\hat{x} = \frac{\left( \frac{\hat{x}_1}{\sigma_1^2} + \frac{\hat{x}_2}{\sigma_2^2} \right)}{\left( \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right)} = \hat{\sigma}^2 \left( \frac{\hat{x}_1}{\sigma_1^2} + \frac{\hat{x}_2}{\sigma_2^2} \right) \quad (8.8)$$

This is the combination of  $x_1$  and  $x_2$  that minimizes the risk function.

Exercise: Find the expected error associated with the combined estimate  $\hat{x}$ .

### 8.1.2 Vector Operations Form

Suppose that rather than a single  $x$  we have a collection of observables that we choose to represent with a vector  $\mathbf{x}_1$ . A risk function for the vector problem can be formulated as follows:

$$R(\mathbf{x}) = (\hat{\mathbf{x}}_1 - \mathbf{x})^T \mathbf{S}_1^{-1} (\hat{\mathbf{x}}_1 - \mathbf{x}) + (\hat{\mathbf{x}}_2 - \mathbf{x})^T \mathbf{S}_2^{-1} (\hat{\mathbf{x}}_2 - \mathbf{x}) \quad (8.9)$$

We have taken the projection of the error on itself with a weighting matrix that is the inverse of the error covariance matrix. This gives us the desired weighted measure of quadratic error as a discrete projection in state space. We differentiate with respect to the unknown  $\mathbf{x}$  to minimize this risk function.

$$\frac{\partial}{\partial \mathbf{x}} R(\mathbf{x}) = \left( -\mathbf{S}_1^{-1} (\hat{\mathbf{x}}_1 - \mathbf{x}) \right)^T - (\hat{\mathbf{x}}_1 - \mathbf{x})^T \mathbf{S}_1^{-1} - \left( -\mathbf{S}_2^{-1} (\hat{\mathbf{x}}_2 - \mathbf{x}) \right)^T - (\hat{\mathbf{x}}_2 - \mathbf{x})^T \mathbf{S}_2^{-1} = 0$$

or, noting that  $\mathbf{S}$  is symmetric,

$$\mathbf{S}_1^{-1} (\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}) + \mathbf{S}_2^{-1} (\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}) = 0$$

The  $\mathbf{x}$  that satisfies this,  $\hat{\mathbf{x}}$ , is our best estimate.

$$\mathbf{S}_1^{-1} \hat{\mathbf{x}}_1 + \mathbf{S}_2^{-1} \hat{\mathbf{x}}_2 = (\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}) \hat{\mathbf{x}}$$

our best estimate is thus:

$$\hat{\mathbf{x}} = (\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1})^{-1} (\mathbf{S}_1^{-1} \hat{\mathbf{x}}_1 + \mathbf{S}_2^{-1} \hat{\mathbf{x}}_2) \quad (8.10)$$

It is useless to construct a new estimate, unless we can estimate the error of the new estimate. So we next need to calculate the error covariance matrix for the combined estimate, which is.

$$\begin{aligned}\hat{\mathbf{S}} &= E\{\hat{\mathbf{q}}\hat{\mathbf{q}}^T\} = E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T\} \\ &= E\left\{\left[\left(\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}\right)^{-1}\left(\mathbf{S}_1^{-1}\mathbf{x}_1 + \mathbf{S}_2^{-1}\mathbf{x}_2\right) - \mathbf{x}\right]\left[\left(\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}\right)^{-1}\left(\mathbf{S}_1^{-1}\mathbf{x}_1 + \mathbf{S}_2^{-1}\mathbf{x}_2\right) - \mathbf{x}\right]^T\right\}\end{aligned}$$

Where  $E$  represents an expected value or ensemble average operator. Now we can rearrange this productively by noting that

$$\left(\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}\right)\mathbf{x} = \mathbf{S}_1^{-1}\mathbf{x} + \mathbf{S}_2^{-1}\mathbf{x}$$

or

$$\mathbf{x} = \left(\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}\right)^{-1}\left(\mathbf{S}_1^{-1}\mathbf{x} + \mathbf{S}_2^{-1}\mathbf{x}\right)$$

Substituting this into the expression for  $\mathbf{S}$  we get:

$$\begin{aligned}\hat{\mathbf{S}} &= E\left\{\left[\left(\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}\right)^{-1}\left(\mathbf{S}_1^{-1}(\mathbf{x}_1 - \mathbf{x}) + \mathbf{S}_2^{-1}(\mathbf{x}_2 - \mathbf{x})\right)\right] \cdot \right. \\ &\quad \left. \left[\left(\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}\right)^{-1}\left(\mathbf{S}_1^{-1}(\mathbf{x}_1 - \mathbf{x}) + \mathbf{S}_2^{-1}(\mathbf{x}_2 - \mathbf{x})\right)\right]^T\right\}\end{aligned}$$

Now use

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$$

to rearrange the second half of the product within the curly brackets in the above equation

$$\begin{aligned}\hat{\mathbf{S}} &= E\left\{\left[\left(\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}\right)^{-1}\left(\mathbf{S}_1^{-1}(\mathbf{x}_1 - \mathbf{x}) + \mathbf{S}_2^{-1}(\mathbf{x}_2 - \mathbf{x})\right)\right] \cdot \right. \\ &\quad \left. \left[\left(\mathbf{S}_1^{-1}(\mathbf{x}_1 - \mathbf{x}) + \mathbf{S}_2^{-1}(\mathbf{x}_2 - \mathbf{x})\right)^T \cdot \left(\mathbf{S}_1^{-1} + \mathbf{S}_2^{-1}\right)^{-1T}\right]\right\}\end{aligned}$$

or doing some more transposing of the second part,

$$\hat{\mathbf{S}} = \mathbb{E} \left\{ \left[ \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1} \left( \mathbf{S}_1^{-1} (\mathbf{x}_1 - \mathbf{x}) + \mathbf{S}_2^{-1} (\mathbf{x}_2 - \mathbf{x}) \right) \right] \cdot \left[ \left( (\mathbf{x}_1 - \mathbf{x})^T \mathbf{S}_1^{-1T} + (\mathbf{x}_2 - \mathbf{x})^T \mathbf{S}_2^{-1T} \right) \cdot \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1T} \right] \right\}$$

We can now expand this out, to get,

$$= \mathbb{E} \left\{ \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1} \left( \begin{array}{l} \mathbf{S}_1^{-1} (\mathbf{x}_1 - \mathbf{x}) (\mathbf{x}_1 - \mathbf{x})^T \mathbf{S}_1^{-1T} + \mathbf{S}_1^{-1} (\mathbf{x}_1 - \mathbf{x}) (\mathbf{x}_2 - \mathbf{x})^T \mathbf{S}_2^{-1T} \\ + \mathbf{S}_2^{-1} (\mathbf{x}_2 - \mathbf{x}) (\mathbf{x}_1 - \mathbf{x})^T \mathbf{S}_1^{-1T} + \mathbf{S}_2^{-1} (\mathbf{x}_2 - \mathbf{x}) (\mathbf{x}_2 - \mathbf{x})^T \mathbf{S}_2^{-1T} \end{array} \right) \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1T} \right\}$$

If we assume that the errors of the two measurements are uncorrelated, then the second and third terms inside the parentheses are zero and we get:

$$\hat{\mathbf{S}} = \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1} \left( \mathbf{S}_1^{-1T} + \mathbf{S}_2^{-1T} \right) \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1T}$$

$$\hat{\mathbf{S}} = \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1} \quad (8.11)$$

This is a simple result achieved after much effort. It is entirely analogous to the scalar result expressed in (8.8) above.

Since we have determined a method to combine two estimates of a vector for which we know the covariance matrix of the measurement error, and we can derive a formula for the covariance matrix of the error of the combined estimate, we have set the stage for sequential estimation. Suppose we have started with two observations,  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , from which we get a combined best estimate,

$$\hat{\mathbf{x}} = \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1} \left( \mathbf{S}_1^{-1} \mathbf{x}_1 + \mathbf{S}_2^{-1} \mathbf{x}_2 \right)$$

with a new, improved error covariance matrix,

$$\hat{\mathbf{S}} = \left( \mathbf{S}_1^{-1} + \mathbf{S}_2^{-1} \right)^{-1}$$

Now suppose we have a new observation  $\mathbf{x}_3$ , which we can combine with the best estimate from the previous two to get a new best estimate.

$$\hat{\mathbf{x}}_{\text{new}} = (\hat{\mathbf{S}}_{\text{old}}^{-1} + \mathbf{S}_3^{-1})^{-1} (\hat{\mathbf{S}}_{\text{old}}^{-1} \hat{\mathbf{x}}_{\text{old}} + \mathbf{S}_3^{-1} \mathbf{x}_3)$$

with a new error covariance matrix,

$$\hat{\mathbf{S}}_{\text{new}} = (\hat{\mathbf{S}}_{\text{old}}^{-1} + \mathbf{S}_3^{-1})^{-1}$$

Apparently we can continue this process forever, getting better and better estimates of  $\mathbf{x}$  and reducing the measure of the error covariance matrix. If each new observation has a similar error covariance matrix, the estimates of  $\hat{\mathbf{x}}_{\text{new}}$  and  $\hat{\mathbf{S}}_{\text{new}}$  will converge and become very insensitive to the addition of new data, provided that the true state is not changing in the mean time. In most applications of interest, the true value of  $\mathbf{x}$  will change on a time scale that is not large compared to the rate at which we get new observations, and we will have to find some method of taking account of this in the estimation process. This usually involves some assumption about the growth with time of the error covariance of the combined estimate. In this way the most recent data will always have an important impact on the latest estimate.



## 8.2 Sequential Estimation when the state vector is not directly observable

### 8.2.1 Estimating the state vector and its error covariance matrix from the observables.

In this section we present a derivation of a discrete Kalman filter, in which we use a linearized version of a function that relates the observables to the state variables. Sequential estimation is the process of successively adding estimates of a state vector and at each step estimating the confidence attributed to the combined estimate, so that the expected error of the estimate decreases. Suppose we have a system that is described completely by a state vector  $\mathbf{b}$ . Suppose further that we cannot observe  $\mathbf{b}$  directly but have a set of observables  $\mathbf{W}$  and a relation between the state vector and the observables.

$$\mathbf{W} = \omega(\mathbf{b}) \quad (8.12a)$$

In order to begin we need at least as many observables,  $p$ , as there are elements in the state vector so that we can invert the relation.

$$\omega_0(\hat{\mathbf{b}}_0) = \begin{bmatrix} \omega_{01}(\hat{\mathbf{b}}_0) \\ \omega_{02}(\hat{\mathbf{b}}_0) \\ \dots\dots\dots \end{bmatrix} = \mathbf{W}_0 \quad (8.12b)$$

to obtain the first estimate  $\hat{\mathbf{b}}_0$  of  $\mathbf{b}$  from  $\mathbf{W}_0$ . Otherwise we must simply guess  $\hat{\mathbf{b}}_0$  to start the estimation process.

The initial error will be  $\mathbf{q}_0 = \hat{\mathbf{b}}_0 - \mathbf{b}$

If we have more than  $p$  observables then we must define a risk function to enable us to choose the optimal  $\omega_0(\mathbf{b})$  given  $\mathbf{W}_0$  that contains more than  $p$  data.

$$\mathbf{R}(\mathbf{b}) = [\mathbf{W}_0 - \omega_0(\mathbf{b})]^T \mathbf{S}_0^{-1} [\mathbf{W}_0 - \omega_0(\mathbf{b})] \quad (8.13)$$

here  $\mathbf{S}_0$  is the covariance matrix of the measurement error. Here the subscript zero on the  $\omega_0(\mathbf{b})$  function indicates that we may only be able to make a linear approximation to the transfer function between the state vector and the observables, and that this linear approximation depends on our estimate of where in state space we are. We next define the measurement error,  $\mathbf{a}$  and its error covariance matrix  $\mathbf{S}$ .

$$\mathbf{a} = \mathbf{W} - \omega(\mathbf{b})$$

$$\mathbf{S} = E\left\{(\mathbf{a} - E\{\mathbf{a}\})(\mathbf{a} - E\{\mathbf{a}\})^T\right\}$$

which are assumed known.  $E$  is the expected value operator-you can think of it as the ensemble mean.

How do we construct the error covariance matrix for  $\mathbf{b}$ ? We do not know it, since the state vector cannot be directly observed. Try a Taylor Series: in hopes of expressing a small deviation of the state vector  $\mathbf{b}$  in terms of small deviations in the observables,  $\mathbf{W} = \omega_0(\mathbf{b})$ .

$$\omega_0(\hat{\mathbf{b}}_0) = \omega_0(\mathbf{b}) + \mathbf{F}_0(\mathbf{b}) \cdot (\hat{\mathbf{b}}_0 - \mathbf{b}) + \dots$$

$$\mathbf{F}_0(\mathbf{b}) = \nabla_{\mathbf{b}} \omega_0(\mathbf{b})$$

$$p \times p \quad p \times 1 \cdot 1 \times p$$

Here we evaluate the gradient at a specific value of  $\mathbf{b}$ . So now we can write,

$$\mathbf{F}_0(\mathbf{b}) \cdot \underset{\substack{\uparrow \\ \text{error} \\ \text{in } \mathbf{b}}}{\mathbf{q}_0} = \underbrace{\omega_0(\hat{\mathbf{b}}_0) - \omega_0(\mathbf{b})}_{\text{error in obs}} \quad (8.14)$$

Then we can write the error covariance matrix of the observables in terms of the error in the state vector.

$$E\{\mathbf{F}_0 \mathbf{q}_0 \mathbf{q}_0^T \mathbf{F}_0^T\} = E\{\mathbf{a}_0 \mathbf{a}_0^T\} = \mathbf{S}_0 \quad (8.15)$$

Suppose the parameter ( $\mathbf{b}$ ) error covariance matrix is given by

$$\mathbf{\Gamma}_0 = E\{\mathbf{q}_0 \mathbf{q}_0^T\}$$

We can now go back and forth between observational error covariance  $\mathbf{S}$ , and the state vector error covariance  $\mathbf{\Gamma}$ , by using the linearized conversion function,  $\mathbf{F}$ . First we can relate the observational error covariance matrix to the state vector error covariance matrix,

$$\mathbf{S}_0 = E\{\mathbf{W}_0 \mathbf{W}_0^T\} = E\{\mathbf{F}_0 \mathbf{q}_0 \mathbf{q}_0^T \mathbf{F}_0^T\} = \mathbf{F}_0 \mathbf{\Gamma}_0 \mathbf{F}_0^T$$

and then, more usefully, the state vector error covariance to the observation error covariance,

$$\Gamma_0 = E\{\mathbf{q}_0\mathbf{q}_0^T\} = E\{\mathbf{F}_0^{-1}\mathbf{W}_0\mathbf{W}_0^T\mathbf{F}_0^{-1T}\} = \mathbf{F}_0^{-1}\mathbf{S}_0\mathbf{F}_0^{-1T}$$

and, if we want the inverse,

$$\begin{matrix} \Gamma_0^{-1} & = & \mathbf{F}_0^{+1T} & \mathbf{S}_0^{-1} & \mathbf{F}_0^{+1} \\ \uparrow & & & \uparrow & \\ \text{inverse} & & & \text{inverse} & \\ \text{of error} & & & \text{error} & \\ \text{covariance of} & & & \text{covariance} & \\ \text{parameters} & & & \text{of obs.} & \end{matrix}$$

Now we can obtain  $\Gamma_0^{-1}$ , the error in inferred state vector, from the error covariance matrix of the observables.

Remember that,

$$\text{Sensitivity matrix} \rightarrow \mathbf{F}_0 \underbrace{(\mathbf{b}_0 - \mathbf{b})}_{\substack{\uparrow \\ \text{error in} \\ \text{parameters}}} = \underbrace{\omega_0(\mathbf{b}) - \omega(\mathbf{b})}_{\substack{\uparrow \\ \text{error} \\ \text{in obs.}}}$$

Now we have:

- (1) An initial estimate  $\hat{\mathbf{b}}_0$  of the state vector variables based upon an initial set of measurements.  $\mathbf{W}_0 = \omega_0(\mathbf{b}_0)$ .
- (2) An estimate of the error covariance  $\Gamma_0$  of  $\hat{\mathbf{b}}_0$  given the error covariance matrix of measurements  $\mathbf{S}_0$ ,

Let's stop for a minute and consider some examples of where this might be applied:

- 1) Determining the orbital parameters of an artificial satellite from observations of its position as a function of time.

- $\mathbf{b}$  = orbital parameters
- $\mathbf{W}$  = angles or positions that can be observed
- $\mathbf{W} = \omega(\mathbf{b})$
- $\omega$  = celestial mechanics equations

- 2) Determining a zonal Fourier representation of the temperature field from observations taken at nadir under a polar-orbiting satellite or from limb scanning measurements

- $\mathbf{b}$  = Coefficients of zonal Fourier expansion of temperature
- $\mathbf{W}$  = Temperatures taken at particular points along a latitude circle
- $\omega$  = formulas of zonal Fourier expansion

$$\mathbf{W} = \omega(\mathbf{b}) = \sum \mathbf{b}_i \cos kx + \sum \mathbf{b}_{i+1} \sin kx$$

- 3) Determining the state of the atmosphere, as represented in a numerical weather prediction model, given observations of some incomplete subset of the full state vector at an odd assortment of places and times, plus things like radiances that are not part of the state vector.

$\mathbf{b}$  = full state vector of the numerical model at some particular time

$\mathbf{W}$  = some observations at some points in space-time not necessarily where we want them. They may be variables that are part of the state vector, but are certainly not all of them. Some of them are weird-like radiances from sounding satellites and imagers or backscatter coefficients from scatterometers.

$$\mathbf{W} = \omega(\mathbf{b})$$

$\omega$  = The equations of the numerical model of the atmosphere, plus equations that may relate observables to the state vector.

### 8.2.1 Improving the Estimate of the state vector with more observations

Suppose we now wish to improve this initial estimate with a new set of observations  $\mathbf{W}_1$ . The number of observations in  $\mathbf{W}_1$ ,  $s$ , need not equal the number of parameters,  $p$ , contained in the vector of state parameters  $\mathbf{b}$ . Let's construct a risk function for the combination of estimates.

$$\begin{aligned} R(\mathbf{b}|\mathbf{q}_0, \mathbf{a}_1) &= \mathbf{q}_0^T \Gamma_0^{-1} \mathbf{q}_0 + \mathbf{a}_1^T \mathbf{S}^{-1} \mathbf{a}_1 \\ &= (\hat{\mathbf{b}}_0 - \mathbf{b})^T \Gamma_0^{-1} (\hat{\mathbf{b}}_0 - \mathbf{b}) + (\mathbf{W}_1 - \omega_1(\mathbf{b}))^T \mathbf{S}^{-1} (\mathbf{W}_1 - \omega_1(\mathbf{b})) \end{aligned}$$

Now operate on this risk function with operator  $\nabla_{\mathbf{b}}$  (matrix differential operator) and try to minimize the risk. To find the new best estimate of  $\mathbf{b}$ ,  $\hat{\mathbf{b}}_1$ . We get,

$$\Gamma_0^{-1} \cdot (\hat{\mathbf{b}}_0 - \hat{\mathbf{b}}_1) + [\nabla_{\mathbf{b}} \omega_1^T(\hat{\mathbf{b}}_1)] \mathbf{S}^{-1} [\mathbf{W}_1 - \omega_1(\hat{\mathbf{b}}_1)] = 0 \quad (8.16)$$

We can now expand  $\omega_1(\hat{\mathbf{b}}_1)$  about  $\mathbf{b}$  (true value) in a Taylor series and retain only the first term.

$$\omega_1(\hat{\mathbf{b}}_1) = \omega_1(\mathbf{b}) + \mathbf{T}_{\mathbf{b}} (\hat{\mathbf{b}}_1 - \mathbf{b}) \quad (8.17)$$

where  $\nabla_{\mathbf{b}} \omega = \mathbf{T}_{\mathbf{b}}$  (8.18)

$\mathbf{T}_b$  thus represents the linear tangent to the model evaluated at the state  $\mathbf{b}$ . If  $\omega(\mathbf{b})$  is a linear function then this procedure can be carried through without a Taylor Series expansion. In numerical weather prediction,  $\mathbf{T}_b$  is the linear tangent operator of the numerical model. It depends on the state vector, since it is a linear perturbation about it.

Now apply(8.17) to (8.16 ) to get

$$\Gamma_0^{-1}(\hat{\mathbf{b}}_0 - \hat{\mathbf{b}}_1) + \mathbf{T}_b^T \mathbf{S}^{-1} [\mathbf{W}_1 - \omega_1(\mathbf{b}) - \mathbf{T}_b(\hat{\mathbf{b}}_1 - \mathbf{b})] = 0 \quad (8.19)$$

or

$$\begin{aligned} \Gamma_0^{-1}(\hat{\mathbf{b}}_0 - \hat{\mathbf{b}}_1) &= \mathbf{T}_b^T \mathbf{S}^{-1} \mathbf{T}_b (\hat{\mathbf{b}}_1 - \mathbf{b}) - \mathbf{T}_b^T \mathbf{S}^{-1} (\mathbf{W}_1 - \omega_1(\mathbf{b})) \\ &= \mathbf{N}(\hat{\mathbf{b}}_1 - \mathbf{b}) - \mathbf{T}_b^T \mathbf{S}^{-1} \mathbf{a}_1 \end{aligned} \quad (8.20)$$

where

$$\mathbf{T}_b^T = \nabla_{\mathbf{b}} \omega^T(\mathbf{b}) \quad (8.21)$$

$$\mathbf{N} = \mathbf{T}_b^T \mathbf{S}^{-1} \mathbf{T}_b$$

We can manipulate the observation error in the following way.

$$\begin{aligned} \mathbf{a}_1 &= \mathbf{W}_1 - \omega_1(\mathbf{b}) \\ &= \mathbf{W}_1 - \omega_1(\hat{\mathbf{b}}_0) - \mathbf{T}_b \cdot (\mathbf{b} - \hat{\mathbf{b}}_0) \\ &= \mathbf{d}_1 - \mathbf{T}_b \cdot (\mathbf{b} - \hat{\mathbf{b}}_0) \end{aligned} \quad (8.22)$$

Here  $\mathbf{d}_1$  is the measurement error estimated using the estimated state vector  $\hat{\mathbf{b}}_0$ .

$$\mathbf{d}_1 = \mathbf{W}_1 - \omega_1(\hat{\mathbf{b}}_0) \quad (8.23)$$

Combine (8.20) and (8.22)

$$\begin{aligned} \Gamma_0^{-1}(\hat{\mathbf{b}}_0 - \hat{\mathbf{b}}_1) &= \mathbf{N}(\hat{\mathbf{b}}_1 - \mathbf{b}) - \mathbf{T}_b^T \mathbf{S}^{-1} (\mathbf{d}_1 - \mathbf{T}_b \cdot (\mathbf{b} - \hat{\mathbf{b}}_0)) \\ &= \mathbf{N}(\hat{\mathbf{b}}_1 - \mathbf{b}) - \mathbf{T}_b^T \mathbf{S}^{-1} \mathbf{d}_1 + \mathbf{N} \cdot (\mathbf{b} - \hat{\mathbf{b}}_0) \end{aligned} \quad (8.24)$$

=

Solve for  $\mathbf{b}_1$ : (Note:  $\mathbf{N} \mathbf{b}$  drops out, fortunately.)

$$\begin{aligned} \hat{\mathbf{b}}_1 &= \hat{\mathbf{b}}_0 + [\mathbf{N} + \Gamma_0^{-1}]^{-1} \mathbf{T}_b^T \mathbf{S}^{-1} \mathbf{d}_1 \\ &\approx \hat{\mathbf{b}}_0 + [\mathbf{N}_0 + \Gamma_0^{-1}]^{-1} \mathbf{T}_0^T \mathbf{S}^{-1} \mathbf{d}_1 \end{aligned} \tag{8.25}$$

where we have made the approximation of evaluating  $\mathbf{T}_b$  using our guess of  $\mathbf{b} = \mathbf{b}_0$ .

$$\mathbf{T}_0^T = \mathbf{T}_{\mathbf{b}_0}^T$$

so that 
$$\mathbf{N}_0 = \mathbf{T}_0^T \mathbf{S}^{-1} \mathbf{T}_0 \tag{8.26}$$

and  $\hat{\mathbf{b}}_0$  is the first guess.

Let's simplify (8.25) a little by writing it as,

$$\hat{\mathbf{b}}_1 = \hat{\mathbf{b}}_0 + \underbrace{\mathbf{M}_1(\hat{\mathbf{b}}_0)}_{\substack{\uparrow \\ \text{Smoothing or} \\ \text{weighting} \\ \text{matrix}}} \cdot \underbrace{[\mathbf{W}_1 - \omega_1(\hat{\mathbf{b}}_0)]}_{\substack{\uparrow \\ \text{Observation} \\ \text{residual}}} \tag{8.27}$$

$$\mathbf{M}_1(\hat{\mathbf{b}}_0) = [\mathbf{N}_0 + \Gamma_0^{-1}]^{-1} \mathbf{T}_0^T \mathbf{S}^{-1} \tag{8.28}$$

The expression (8.27) indicates the optimum manner for modifying the initial estimate  $\hat{\mathbf{b}}_0$ , by an appropriate weighting of new observation  $\mathbf{W}_1$  given the measurement error matrix  $\mathbf{S}$ , prediction function  $\omega$ , and sensitivity matrix  $\mathbf{T}$ .

Note:  $M_1(\hat{\mathbf{b}}_0)$  depends on the last (best) estimate of  $\mathbf{b}$ .

What is the dimension of  $M_1(\hat{\mathbf{b}}_0)$ ?

$$\hat{\mathbf{b}}_1^{[p \times 1]} = \hat{\mathbf{b}}_0^{[p \times 1]} + \underbrace{[\mathbf{N}_0 + \Gamma_0^{-1}]^{-1} \mathbf{T}_0^T \mathbf{S}^{-1}}_{\substack{\uparrow \\ \mathbf{M}_1(\hat{\mathbf{b}}_0)^{[p \times s]}}} \underbrace{[\mathbf{W}_1 - \omega_1(\hat{\mathbf{b}}_0)]}_{\substack{\uparrow \\ \mathbf{d}_1^{[1 \times s]}}} \tag{8.28}$$

[p×1] ok if s < p

In Rodgers applic. s=1, p=13.

We can construct a recursive scheme to improve our estimate of  $\mathbf{b}$  sequentially using observations  $\mathbf{W}_i$

$$\begin{aligned}\hat{\mathbf{b}}_{i+1} &= \hat{\mathbf{b}}_i + \mathbf{M}_{i+1}(\hat{\mathbf{b}}_i, \hat{\mathbf{b}}_{i+1}, \hat{\mathbf{b}}_0) \cdot [\mathbf{W}_{i+1} - \omega_{i+1}(\hat{\mathbf{b}}_i)] \\ \mathbf{M}_{i+1}(\hat{\mathbf{b}}_i, \hat{\mathbf{b}}_{i+1}, \hat{\mathbf{b}}_0) &= [\mathbf{N}_i + \Gamma_{i-1}^{-1}]^{-1} \mathbf{T}_i^T \mathbf{S}^{-1} \\ \mathbf{N}_{i+1} &= \mathbf{T}_{i+1}^T \mathbf{S}^{-1} \mathbf{T}_{i+1} \\ \Gamma_i &= [\mathbf{N}_i(\mathbf{b}_i) + \Gamma_{i-1}^{-1}]^{-1}\end{aligned}\tag{8.29}$$

By considerable manipulation one can reduce this to a form requiring only one inverse.

Kalman filtering is theoretically very appealing and is correspondingly very powerful and effective. Its main problem is that it takes a great deal of computational resources to apply it to realistic problems in meteorology, and it is currently impractical to apply it in its full form to today's numerical prediction models with millions of degrees of freedom. In the next section we will do a simple practical example in mapping satellite data into Fourier coefficients, without the use of any other model equations.

References:

Deutsch, R., 1965, Estimation Theory: Englewood Cliffs, N.J., Prentice-Hall, 269 p.

Grewal, M. S., 1993, Kalman Filtering: Theory and Practice: Prentice-Hall Information and System Science Series: Englewood Cliffs, N.J., Prentice-Hall, 381 p.

Rodgers, C. R., 1977, Statistical principles of inversion theory, Inversion Methods in Atmospheric Remote Sounding: New York, Academic Press, p. 117-138.

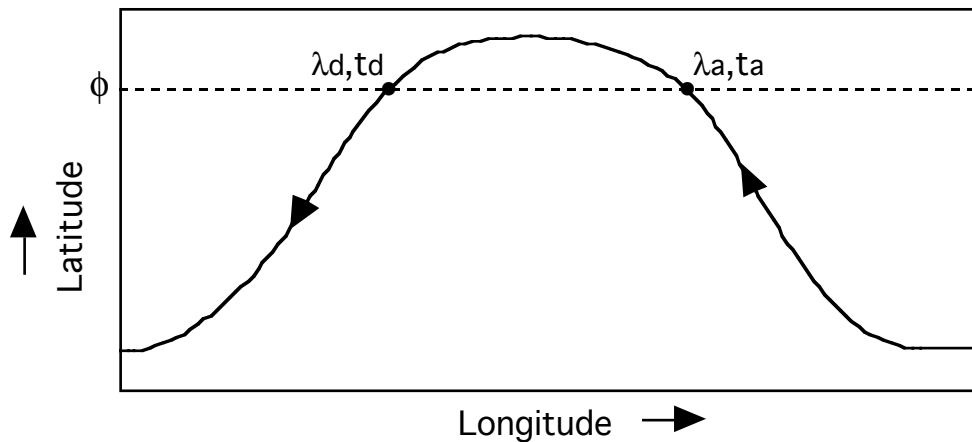
### 8.3 Kalman Filter Application to Mapping Satellite Data

In this section we will consider the use of the Kalman filter for making global synoptic maps from polar orbiting satellites. Sun-synchronous satellites are in about 98° inclination orbits whose equator crossing time remains fixed relative to the sun. For every latitude less than about 82°, they produce two observations each orbit at each latitude. An orbit takes about 90 to 120 minutes, depending on the orbital altitude, so 13-15 orbits are executed each day, and the ground tracks of successive orbits are separated by about 15 degrees of longitude. For a given latitude  $\phi_0$ , a polar orbiting satellite produces a sequence of observations of  $T$ , say temperature for instance.

$$T(\lambda_{a_1}, t_{a_1}), T(\lambda_{d_1}, t_{d_1}), T(\lambda_{a_2}, t_{a_2}), T(\lambda_{d_2}, t_{d_2}), \dots$$

Polar orbiting satellites:

Track of Polar-Orbiting Satellite  
Sampling at latitude  $\phi$



A non sun-synchronous, or medium inclination orbiter, would sample in much the same way, except it would not reach as high a latitude, and the local solar time of the equator crossings would vary with time.

An observation can be taken at a particular latitude at most once each time the satellite crosses the latitude circle. A typical polar orbiting satellite crosses the equator twice during each 90 minute orbit, once on the ascending node (heading toward the north pole) and once on the descending node. If the satellite is sun-synchronous, then it takes 24 hours for the ascending node crossings to precess through all longitudes. How can we use a Kalman Filter Algorithm to make synoptic maps out of this mess? We can use a sequential estimator or Kalman filter.

$$\mathbf{b}_{i+1} = \mathbf{b}_i + [\mathbf{N}_i + \Gamma_{i-1}^{-1}]^{-1} \mathbf{T}_i^T \mathbf{S}^{-1} [\mathbf{W}_{i+1} - \omega_{i+1}(\hat{\mathbf{b}}_i)] \tag{8.30}$$



take  $\mathbf{b}$ , the state vector, to be the coefficients of a zonal harmonic analysis of the temperature  $T$

$$T(\lambda) = b_1 + b_2 \cos \lambda + b_3 \sin \lambda + b_4 \cos 2\lambda + b_5 \sin 2\lambda + \dots + b_{13} \sin 6\lambda$$

$$\mathbf{b} = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ b_{13} \end{bmatrix}$$

where we have arbitrarily truncated the zonal harmonic expansion at wavenumber 6, 13 coefficients. This is about the maximum number of functions you can expect to define with only 13 orbits per day. The temperature at a particular set of longitudes is given by,

$$W(\lambda_i) = \begin{bmatrix} 1, \cos \lambda_0, \sin \lambda_0, \cos 2\lambda_0, \sin 2\lambda_0 \dots \\ 1, \cos \lambda_1, \sin \lambda_1, \cos 2\lambda_1, \sin 2\lambda_1 \dots \\ 1, \cos \lambda_2, \sin \lambda_2, \cos 2\lambda_2, \sin 2\lambda_2 \dots \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \cdot \\ b_{13} \end{bmatrix} = \mathbf{T} \mathbf{b}$$

In this case, the transformation from observable to state vector (observation at a longitude along a latitude circle to the Fourier coefficients of an expansion along that latitude) is linear and constant, so we don't need to do a Taylor Series expansion.

$$W = \omega(\mathbf{b}) = \mathbf{T} \mathbf{b}$$

$$\mathbf{N}_i = \mathbf{T}_i^T \mathbf{S}^{-1} \mathbf{T}_i = \text{constant}$$

then

$$\mathbf{b}_{i+1} = \mathbf{b}_i + \left[ \mathbf{T}^T \mathbf{S}^{-1} \mathbf{T} + \Gamma_{i-1}^{-1} \right]^{-1} \mathbf{T}^T \mathbf{S}^{-1} [\mathbf{W}_{i+1} - \mathbf{T} \mathbf{b}_i] \quad (8.31)$$

$$\Gamma_i = \left[ \mathbf{T}^T \mathbf{S}^{-1} \mathbf{T} + \Gamma_{i-1}^{-1} \right]^{-1} \quad (8.32)$$

Let's manipulate this so that only one matrix inverse is required for each time step. Start by inverting (8.32),

$$\Gamma_i^{-1} = \mathbf{T}^T \mathbf{S}^{-1} \mathbf{T} + \Gamma_{i-1}^{-1} \quad (8.33)$$

$$\Gamma_i \times \rightarrow \Gamma_i \Gamma_i^{-1} = \Gamma_i \mathbf{T}^T \mathbf{S}^{-1} \mathbf{T} + \Gamma_i \Gamma_{i-1}^{-1} \quad (8.34)$$

$$\times \Gamma_{i-1} \rightarrow \Gamma_{i-1} = \Gamma_i \mathbf{T}^T \mathbf{S}^{-1} \mathbf{T} \Gamma_{i-1} + \Gamma_i \quad (8.35)$$

$$\begin{aligned} \times \mathbf{T}^T \rightarrow \Gamma_{i-1} \mathbf{T}^T &= \Gamma_i \mathbf{T}^T + \Gamma_i \mathbf{T}^T \mathbf{S}^{-1} \mathbf{T} \Gamma_{i-1} \mathbf{T}^T \\ &= \Gamma_i \mathbf{T}^T \mathbf{S}^{-1} (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T) \end{aligned} \quad (8.36)$$

$$\Gamma_{i-1} \mathbf{T}^T (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} = \Gamma_i \mathbf{T}^T \mathbf{S}^{-1} \quad (8.37)$$

post multiply by  $\mathbf{T} \Gamma_{i-1}$

$$\Gamma_{i-1} \mathbf{T}^T (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} \mathbf{T} \Gamma_{i-1} = \Gamma_i \mathbf{T}^T \mathbf{S}^{-1} \mathbf{T} \Gamma_{i-1} \quad (8.38)$$

Inserting (8.38) into (8.35)

$$\begin{aligned} \Gamma_{i-1} &= \Gamma_{i-1} \mathbf{T}^T (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} \mathbf{T} \Gamma_{i-1} + \Gamma_i \\ &\quad \text{or} \\ \Gamma_i &= \Gamma_{i-1} - \Gamma_{i-1} \mathbf{T}^T (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} \mathbf{T} \Gamma_{i-1} \end{aligned} \quad (8.39)$$

i.e., only 1 matrix inversion is now required to get  $\Gamma_i$  from  $\Gamma_{i-1}$ .

Note that (8.31) can be written, subs (8.32) into (8.31),

$$\mathbf{b}_{i+1} = \mathbf{b}_i + \Gamma_i \mathbf{T}^T \mathbf{S}^{-1} [\mathbf{W}_{i+1} - \mathbf{T} \mathbf{b}_i] \quad (8.40)$$

substitute (8.37) into (8.40) to get

$$\mathbf{b}_{i+1} = \mathbf{b}_i + \Gamma_{i-1} \mathbf{T}^T (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} [\mathbf{W}_{i+1} - \mathbf{T} \mathbf{b}_i] \quad (8.41)$$

(8.39) and (8.41) form a new system where only one matrix inverse  $(\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1}$  is required.

We march messily ahead:

$$\mathbf{b}_{i+1} = \mathbf{b}_i + \Gamma_{i-1} \mathbf{T}^T (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} [\mathbf{W}_{i+1} - \mathbf{T} \mathbf{b}_i] \quad (8.42)$$

$$\Gamma_i = \Gamma_{i-1} - \Gamma_{i-1} \mathbf{T}^T (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} \mathbf{T} \Gamma_{i-1} \quad (8.43)$$

In practice, for the satellite mapping problem we are only inserting one temperature at a time so that  $\mathbf{W}_i = W_i$ , a scalar i.e.,  $s = 1$ .

$$\begin{aligned} \mathbf{b}_{i+1} &= \mathbf{b}_i + \Gamma_{i-1} \mathbf{T}^T \overbrace{(\sigma^2 + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1}}^{[1 \times 1]} [W_{i+1} - \mathbf{T} \mathbf{b}_i] \\ &= [13 \times 1] \end{aligned}$$

$$\Gamma_i = \Gamma_{i-1} - \Gamma_{i-1} \mathbf{T}^T (\sigma^2 + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} \mathbf{T} \Gamma_{i-1}$$

In this simple case no matrix inversion is required at all. The state vector is 13x1 (a column vector), the T matrix operator is 1x13 (a row vector), the error covariance matrix of the state vector is 13x13, and the error matrix of the observations is a scalar. The matrix we are required to invert is also a scalar.

Note that we can write (8.43) as:

$$\Gamma_i = \Gamma_{i-1} \underbrace{\left[ \mathbf{I} - \mathbf{T}^T (\mathbf{S} + \mathbf{T} \Gamma_{i-1} \mathbf{T}^T)^{-1} \mathbf{T} \Gamma_{i-1} \right]}_{\mathbf{A}}$$

It can be shown that if the eigenvalues of  $\mathbf{A}$  are not all less than one, the iteration will not converge. However, if it does converge  $\Gamma_i$  will become smaller with time. This is O.K. if  $\mathbf{b}$  is not changing in time. Since we know that as we go along adding our data, time is passing and the atmosphere is changing, we need to augment the error covariance matrix of the state vector  $\Gamma_i$  so that it will reflect this additional uncertainty.

Rather than use  $\Gamma_{i-1}$  in (8.42) and (8.43) use

$$\mathbf{S}_i^0 = \Gamma_{i-1} + \Delta \Gamma \Delta t \quad (8.44)$$

$$\mathbf{b}_{i+1} = \mathbf{b}_i + \mathbf{S}_i^0 \mathbf{T}^T (\sigma^2 + \mathbf{T} \mathbf{S}_i^0 \mathbf{T}^T)^{-1} [W_{i+1} - \mathbf{T} \mathbf{b}_i] \quad (8.45)$$

$$\Gamma_i = \mathbf{S}_i^0 - \mathbf{S}_i^0 (\sigma^2 + \mathbf{T} \mathbf{S}_i^0 \mathbf{T}^T)^{-1} \mathbf{T} \mathbf{S}_i^0 \quad (8.46)$$

Here  $\sigma^2$  is the covariance of the error of the measurements (now a scalar).

$$\Delta \Gamma \text{ can be chosen to be } = \frac{\Gamma_{\text{clim}}}{\tau} \quad (8.47)$$

Where  $\Gamma_{\text{clim}}$  is the climatological covariance of the coefficients and  $\tau$  is the time it takes the state vector to change so that

$$E\{[\mathbf{b}_t - \mathbf{b}_{t+\Delta t}][\mathbf{b}_t - \mathbf{b}_{t+\Delta t}]^T\} \cong \Gamma_{\text{clim}} \tag{8.48}$$

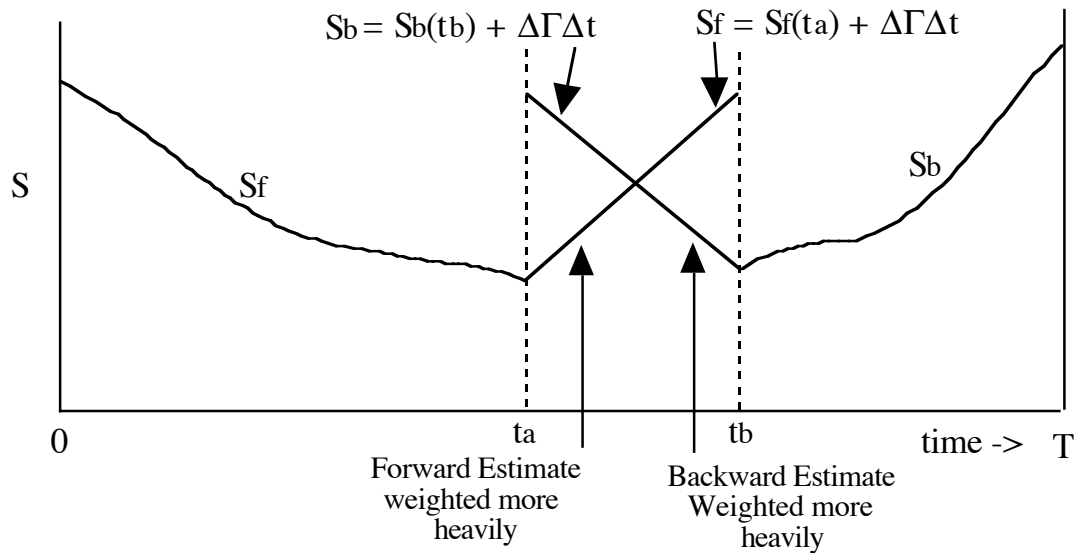
If we assume that the mapping process is done retrospectively, so that we have a long data series to work with, then we can pass over the data twice, once forward in time and once backward in time and obtain two estimates at each map time.

$\hat{\mathbf{b}}_{\text{forward}}(t_1), \hat{\mathbf{b}}_{\text{backward}}(t_1)$  and save their associated error covariance matrices.

These can then be combined in the usual manner that takes into account the objective measures of confidence we have calculated.

$$\hat{\mathbf{b}}_{\text{comb}}(t_i) = (\mathbf{S}_f^{-1} + \mathbf{S}_b^{-1})^{-1} (\mathbf{S}_f^{-1} \hat{\mathbf{b}}_{\text{forward}} + \mathbf{S}_b^{-1} \hat{\mathbf{b}}_{\text{backward}}) \tag{8.49}$$

If you have a data gap as shown below, then the linear growth with time of the error assumed in (8.44) will mean that the forward solution will be weighted more heavily at the beginning of the gap, both will be weighted equally at the center, and the backward solution will be weighted more heavily at the end of the gap.



Kalman filtering works very well for this purpose and many others. For this type of regular satellite data ingestion, there is also a transformed Fourier Method, which is much faster and very accurate, but which doesn't do as well on missing data (Salby, 1982a,b).

References:

Salby, M. L., 1982a, Sampling theory for asynoptic satellite observations. Part I: Space-time spectra, resolution and aliasing: *J. Atmos. Sci.*, v. 39, p. 2577-2600.

Salby, M. L., 1982b, Sampling theory for asynoptic satellite observations. Part II: Fast Fourier synoptic mapping: *J. Atmos. Sci.*, v. 39, p. 2601-2614.

### 8.4 A Review of the Kalman Filter

In this section we will briefly review the equations for a discrete Kalman filter developed in the previous sections. The approach used here is a little more general than those shown previously, but is very similar in final structure to the derivation for the one-inversion-only formulation given in (8.42) and (8.43). In particular, the formulation below explicitly separates the model or system dynamic operator from the operator that relates the observables to the state vector. In the previous two examples, the observation problem and the dynamical evolution problem were mixed together in a somewhat unsatisfactory way.

*A Discrete Dynamical System with Noise:*

Consider the following discrete linear process with state variable  $\mathbf{b}_k$  and observation  $\mathbf{w}_k$  where the subscript  $k$  indicates an approximation step or time.

$$\mathbf{b}_k = \mathbf{G}_{k-1} \mathbf{b}_{k-1} + \mathbf{H}_{k-1} \boldsymbol{\varepsilon}_k \quad (8.50)$$

$$\mathbf{w}_k = \mathbf{T}_k \mathbf{b}_k + v_k \quad (8.51)$$

The operator  $\mathbf{G}_k$  describes the system dynamics of how the state evolves. It would be the numerical weather prediction model, the celestial mechanics equations, the factory process, or whatever system you are describing. This linear operator is presumed known. There is a separate set of mathematical and statistical tools to use if you know the state vector and want to estimate the system operator. The system dynamics are subjected to the introduction of the noise process  $\boldsymbol{\varepsilon}_k$  (it will be assumed to be Gaussian white noise here). This noise is modified by the system dynamics by the operator  $\mathbf{H}_k$ . The observables are related to the state vector through the operator  $\mathbf{T}_k$ , and the observation process is also subjected to noise,  $v_k$ . If the governing equations are nonlinear, then we must use a linear tangent approximation to construct both the process operator  $\mathbf{G}_k$  and the observation operator  $\mathbf{T}_k$ . Conditions that we will apply to the noise processes are that they are unbiased,

$$E\langle \boldsymbol{\varepsilon}_k \rangle = 0; \quad E\langle v_k \rangle = 0 \quad (8.52)$$

and they have known error covariances matrices, which are diagonal,

$$E\langle \boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_i^T \rangle = \delta(k-i) \mathbf{E}_k; \quad E\langle v_k v_i^T \rangle = \delta(k-i) \mathbf{V}_k \quad (8.53)$$

We further assume that the noise processes are uncorrelated with each other and with the state and observation variables. We will not show the derivation of the Kalman

filter equations here, since we have shown a derivation above. Let's just look at the result.

*The Kalman Filter Equations:*

The Kalman filter to produce the best estimate of this system at each step is given by the following set of equations. The index  $k$  is the iteration or time step, and the superscript - and + indicate the state of the estimation process before and after the new data have been incorporated, respectively.

$$\hat{\mathbf{b}}_k^- = \mathbf{G}_{k-1} \hat{\mathbf{b}}_{k-1}^+ \quad (8.54)$$

The *a priori* estimate (-) at iteration  $k$ ,  $\hat{\mathbf{b}}_k^-$ , is the last best estimate at iteration  $k-1$ ,  $\hat{\mathbf{b}}_{k-1}^+$ , projected forward without noise using the system dynamics operator evaluated at the previous step  $\mathbf{G}_{k-1}$ . The error covariance matrix of this estimate is the sum of the state vector error projected forward with the system operator and the process noise projected forward with its system operator.

$$\Gamma_k^- = \mathbf{G}_{k-1} \Gamma_{k-1}^+ \mathbf{G}_{k-1}^T + \mathbf{H}_{k-1} \mathbf{E}_k \mathbf{H}_{k-1}^T \quad (8.55)$$

The new estimate of the state vector is the first guess  $\hat{\mathbf{b}}_k^-$  plus a correction that is the deviation of the observation from the first guess, weighted by the Kalman smoothing matrix .

$$\hat{\mathbf{b}}_k^+ = \hat{\mathbf{b}}_k^- + \bar{\mathbf{K}}_k (\mathbf{w}_k - \mathbf{T}_{k-1} \hat{\mathbf{b}}_k^-) \quad (8.56)$$

The Kalman smoothing matrix for the iteration step  $k$  is given by,

$$\bar{\mathbf{K}}_k = \Gamma_k^- \mathbf{T}_k^T (\mathbf{T}_k \Gamma_k^- \mathbf{T}_k^T + \mathbf{V}_k)^{-1} \quad (8.57)$$

The matrix  $\mathbf{V}_k$  is the error covariance matrix of the observations, which was called  $\mathbf{S}$  in the derivations of the previous sections. The new estimate of the error covariance matrix of the state vector to be used at the next step is,

$$\Gamma_k^+ = \Gamma_k^- - \bar{\mathbf{K}}_k \mathbf{T}_k \Gamma_k^- \quad (8.58)$$

The equations (8.54) through (8.58) constitute the Kalman filter estimator for the linear dynamical system (8.50) observed by the process (8.51). They are similar to (8.42-3), but we have included the intermediate steps more explicitly, particularly the first one (8.54) is always used in numerical weather prediction, and was not explicitly called out before. The problem with applying this in a numerical prediction model is that the matrix

to be inverted is too big to be inverted, so further approximations to this step must be made.

The Kalman filter estimator takes into account: the known system dynamics, the effect of the system dynamics on Gaussian white noise (e.g. it might redden it), what is known about the error characteristics of the observations, and the relationships between the observables and the state variables. It makes an optimum mean square estimate of the system state based on all of this input, and provides an estimate of the error of this estimate.