Chapter 6

The Kalman filter

In the last chapter, we saw that in Data Assimilation, we ultimately desire knowledge of the full *a posteriori* p.d.f., that is the conditional p.d.f. of the state given the observations. In practice, it is difficult to determine the complete p.d.f., so we can instead try to estimate some parameter of the p.d.f. that can serve as a "best" estimate of the state. Various principles lead to various estimators. The Minimum Variance (MV) estimator is the conditional mean (of the state given the observations) while the Maximum of the *A Posteriori* (MAP) p.d.f. leads to the mode of conditional p.d.f. For our measurement equation, assuming Gaussian background and observation errors allows us to determine the complete *a posteriori* p.d.f. Also, the MV and MAP estimators are identical. The MV estimator leads to the OI equations. The MAP estimator leads to 3DVAR. For linear observation operators and Gaussian error statistics, 3DVAR and OI are therefore equivalent.

Thus far, we have only considered the spatial estimation problem, that is, when observations are distributed in space, but at a single point in time. For stationary processes, the mean and covariance are constant in time so that the data assimilation scheme could be applied at different times using the same statistics. 3DVAR and OI are both examples of 3D data assimilation schemes. For the global NWP problem, the errors are assumed stationary over the time scale of 1 month. However, ultimately, for environmental applications, we must consider nonstationary processes, because the governing equations are time-varying.

In this chapter, we will extend our analysis of the previous chapter to include the time dimension. This will enable us to derive the linear Kalman Filter (KF) for discrete time processes. If we are interested in data assimilation, why do we talk of a "filter"? Historically, in signal processing theory, the idea was to separate the signal from the noise, based on the the assumption of a frequency gap or separation between the signal and noise. (In fact, if small overlaps occurred in the frequency domain, it didn't matter so long as the noise frequencies are primarily filtered.) For stationary processes, Wiener developed his filter to extract the signal. The Kalman filter, developed in 1960, solved the problem in state-space and for time-varying systems.

Before we can define our stochastic-dynamic system describing the signal and measurement processes, we must first briefly review linear dynamical systems.

6.1 Linear Dynamical Systems

Consider the linear dynamical system:

$$\dot{\mathbf{x}}(t) = \frac{d\mathbf{x}}{dt} = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{L}(t)\mathbf{u}(t) + \mathbf{G}(t)\mathbf{w}(t). \tag{6.1}$$

Let us consider each of the components of this equation. The state vector, $\mathbf{x}(t)$, is composed of a set of variables which completely describe the unforced motion of the system. In other words, given the state vector at a point in time, and the forcing and the controlling inputs from that time forward, the state at any other time can be computed. For a given physical system, the state vector is *not* unique. i.e. If $\mathbf{x}(t)$ is a state vector, so too is $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$. $\mathbf{F}$ represents the dynamics of the model. $\mathbf{w}(t)$ is the random system or model error and $\mathbf{G}(t)$ is the operator which maps this error into state space. The existence of this term is a recognition of the fact that our models are only approximations to the true (atmospheric) dynamics. Thus, even if we knew the truth (on our model basis), a forecast from this truth using our imperfect model would ultimately diverge from reality. Finally, $\mathbf{u}(t)$ are the inputs to the system (sources) and $\mathbf{L}$ is the operator mapping the inputs to state space. For example, if our system was a chemical-transport process, $\mathbf{u}(t)$ would be the emissions of pollutants into the atmosphere. If emissions are occurring at only some spatial locations, then $\mathbf{L}$ is an interpolation operator from emission space to state space.

6.1.1 Stability of the model

Consider the unforced system

$$\dot{\mathbf{x}} = \mathbf{F}\mathbf{x} \tag{6.2}$$

This system can be written in difference form:

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k \mathbf{x}_k. \tag{6.3}$$

Recursive substitution for time-invariant $\boldsymbol{\phi}$ yields:

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}^{k+1} \mathbf{x}_0. \tag{6.4}$$

Thus the state remains bounded only if $|\boldsymbol{\phi}| \leq 1$. The model dynamics are stable if eigenvalues of $\boldsymbol{\phi}$ are less than or equal to 1.

6.1.2 Transition matrix

Let us ignore model error for the moment. The solution to the system, (6.1), can be written in terms of its trajectory in state space as

$$\boldsymbol{\phi}(t; \mathbf{u}(\tau), \mathbf{x}_0, t_0). \tag{6.5}$$

If there are no inputs, the solution is

$$\boldsymbol{\phi}(t; \mathbf{x}_0, t_0). \tag{6.6}$$

Definition: The *transition matrix* $\boldsymbol{\phi}(t, t_0)$ is the $n \times n$ matrix such that

$$\mathbf{x}(t) = \boldsymbol{\phi}(t; \mathbf{x}_0, t_0) = \boldsymbol{\phi}(t, t_0)\mathbf{x}_0. \tag{6.7}$$

Substituting (6.7) into

$$\dot{\mathbf{x}} = \mathbf{F}(t)\mathbf{x} \tag{6.8}$$

yields

$$\dot{\boldsymbol{\phi}}(t, t_0) = \mathbf{F}(t)\boldsymbol{\phi}(t, t_0) \tag{6.9}$$

with initial condition $\boldsymbol{\phi}(t_0, t_0) = \mathbf{I}$. For a time-invariant system,

$$\boldsymbol{\phi}(t, t_0) = \boldsymbol{\phi}(t - t_0)$$

and (6.9) becomes

$$\dot{\boldsymbol{\phi}}(t - t_0) = \mathbf{F}\boldsymbol{\phi}(t - t_0) \tag{6.10}$$

or

$$\boldsymbol{\phi}(t - t_0) = \mathrm{e}^{\mathbf{F}(t-t_0)}.$$

There is NO general solution for the transition matrix of time-varying linear systems.

For a discrete-time system with no stochastic error, we can write our system as

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k\mathbf{x}_k + \boldsymbol{\Lambda}_k\mathbf{u}_k \tag{6.11}$$

where $\mathbf{x}_k = \mathbf{x}(t_k)$, $\boldsymbol{\Lambda}_k = \mathbf{L}(t_k)$ and the transition matrix takes the state from time t_k to time t_{k+1}:

$$\boldsymbol{\phi}_k = \boldsymbol{\phi}(k, k + 1).$$

Properties of the transition matrix

Property	Continuous time	Discrete time
(1) Transition:	$\boldsymbol{\phi}(t_2, t_0) = \boldsymbol{\phi}(t_2, t_1)\boldsymbol{\phi}(t_1, t_0)$	$\boldsymbol{\phi}(k, m) = \boldsymbol{\phi}(k, l)\boldsymbol{\phi}(l, m)$
(2) Inversion:	$\boldsymbol{\phi}(t_0, t_1) = \boldsymbol{\phi}^{-1}(t_1, t_0)$	$\boldsymbol{\phi}(m, k) = \boldsymbol{\phi}^{-1}(k, m)$
(3) Identity:	$\boldsymbol{\phi}(t_0, t_0) = \mathbf{I}$	$\boldsymbol{\phi}(m, m) = \mathbf{I}$

Now that we have the homogeneous solution, what is the solution to the system equation with inputs? i.e.

$$\dot{\mathbf{x}}(t) = \frac{d\mathbf{x}}{dt} = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{L}(t)\mathbf{u}(t) \tag{6.12}$$

with $\mathbf{x}(t_0) = \mathbf{x}_0$. The solution is

$$\mathbf{x}(t) = \boldsymbol{\phi}(t, t_0)\mathbf{x}_0 + \int_{t_0}^{t} \boldsymbol{\phi}(t, \tau)\mathbf{L}(\tau)\mathbf{u}(\tau)d\tau. \tag{6.13}$$

Before checking that this solution satisfies (6.12), recall Leibnitz' rule:

$$\frac{d}{dt}\int_{A(t)}^{B(t)} f(t, \tau)d\tau = \int_{A(t)}^{B(t)} \frac{d}{dt}f(t, \tau)d\tau + f(t, B(t))\frac{dB}{dt} - f(t, A(t))\frac{dA}{dt} \tag{6.14}$$

Now substitute (6.13) into the left side of (6.12):

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \dot{\boldsymbol{\phi}}(t,t_0)\mathbf{x}_0 + \int_{t_0}^{t} \dot{\boldsymbol{\phi}}(t,\tau)\mathbf{L}(\tau)\mathbf{u}(\tau)d\tau + \boldsymbol{\phi}(t,t)\mathbf{L}(t)\mathbf{u}(t) \\ &= \mathbf{F}(t)\boldsymbol{\phi}(t,t_0)\mathbf{x}_0 + \int_{t_0}^{t} \mathbf{F}(t)\boldsymbol{\phi}(t,\tau)\mathbf{L}(\tau)\mathbf{u}(\tau)d\tau + \mathbf{L}(t)\mathbf{u}(t) \\ &= \mathbf{F}(t)\left[\boldsymbol{\phi}(t,t_0)\mathbf{x}_0 + \int_{t_0}^{t} \boldsymbol{\phi}(t,\tau)\mathbf{L}(\tau)\mathbf{u}(\tau)d\tau\right] + \mathbf{L}(t)\mathbf{u}(t) \\ &= \mathbf{F}(t)\mathbf{x}(t) + \mathbf{L}(t)\mathbf{u}(t) \end{aligned} \tag{6.15}$$

which equals the right side of (6.12). Thus (6.12) is satisfied by (6.13). Similarly, the solution of (6.1) is

$$\mathbf{x}(t) = \boldsymbol{\phi}(t,t_0)\mathbf{x}_0 + \int_{t_0}^{t} \boldsymbol{\phi}(t,\tau)\mathbf{L}(\tau)\mathbf{u}(\tau)d\tau \int_{t_0}^{t} \boldsymbol{\phi}(t,\tau)\mathbf{G}(\tau)\mathbf{w}(\tau)d\tau. \tag{6.16}$$

For the corresponding discrete-time system,

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k\mathbf{x}_k + \boldsymbol{\Lambda}_k\mathbf{u}_k + \boldsymbol{\Gamma}_k\mathbf{w}_k \tag{6.17}$$

$$\begin{aligned}\boldsymbol{\Gamma}_k\mathbf{w}_k &= \int_{t_k}^{t_{k+1}} \boldsymbol{\phi}(t_{k+1},\tau)\mathbf{G}(\tau)\mathbf{w}(\tau)d\tau \\ \boldsymbol{\Lambda}_k\mathbf{u}_k &= \int_{t_k}^{t_{k+1}} \boldsymbol{\phi}(t_{k+1},\tau)\mathbf{L}(\tau)\mathbf{u}(\tau)d\tau \end{aligned} \tag{6.18}$$

6.2 Observability and controllability

In this section, we introduce the concepts of observability and controllability. These concepts can be established for linear models and observation operators. The following discussion will consider only time-invariant linear models and observation operators for which these concepts are most easily established and demonstrated. The extension for general linear models is noted in Todling (1999). Because we cannot determine these principles for general nonlinear models and observation operators, they are of little relevance for us, in practice. Nevertheless, it is useful to know that these concepts exist, for simple systems. Furthermore, the concepts of observability and controllability are needed to establish the conditions for stability of the Kalman filter, in later sections.

6.2.1 Observability

A system is observable if we can determine the sequence of states, $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k$ from the sequence of measurements, $\mathbf{z}_o, \mathbf{z}_1, \ldots, \mathbf{z}_k$. Consider the case of a linear, time-independent system. For a perfect model (i.e. no model error):

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}\mathbf{x}_k \tag{6.19}$$

when there are no inputs, $\mathbf{u}_k = 0$, and $\mathbf{x}_k$ is an n-vector. Assume there are r scalar noise-free observations:

$$z_k = \mathbf{H}\mathbf{x}_k, \qquad k = 0, 1, 2, \ldots, r-1 \tag{6.20}$$

where $\mathbf{H}$ is a time-invariant $1 \times n$ matrix. Then,

$$\begin{aligned} z_0 &= \mathbf{H}\mathbf{x}_0 \\ z_1 &= \mathbf{H}\mathbf{x}_1 = \mathbf{H}\boldsymbol{\phi}\mathbf{x}_0 \\ z_2 &= \mathbf{H}\mathbf{x}_2 = \mathbf{H}\boldsymbol{\phi}^2\mathbf{x}_0 \\ &\vdots \\ z_r &= \mathbf{H}\mathbf{x}_r = \mathbf{H}\boldsymbol{\phi}^{r-1}\mathbf{x}_0. \end{aligned} \tag{6.21}$$

If we define

$$\mathbf{y} = \begin{bmatrix} z_o \\ z_1 \\ z_2 \\ \vdots \\ z_r \end{bmatrix} \quad \text{and} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{H}^{\mathrm{T}} & \boldsymbol{\phi}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} & (\boldsymbol{\phi}^{\mathrm{T}})^2\mathbf{H}^{\mathrm{T}} & \cdots & (\boldsymbol{\phi}^{\mathrm{T}})^{r-1}\mathbf{H}^{\mathrm{T}} \end{bmatrix} \tag{6.22}$$

then

$$\mathbf{y} = \mathbf{Z}^{\mathrm{T}}\mathbf{x}_0$$

and $\mathbf{x}_0$ is uniquely determined if $\mathbf{Z}^{\mathrm{T}}$ is non-singular, or has a rank of r. If we can determine $\mathbf{x}_0$ uniquely, then from (6.19), we can determine $\mathbf{x}_k$ for $k = 1, 2, \ldots, r$. A system is *observable* at $t_1 > t_0$, if $\mathbf{x}(t_0)$ can be determined by observing $\mathbf{z}(t)$, $t \in [t_0, t_1]$. If all states $\mathbf{x}(t)$ corresponding to all $\mathbf{z}(t)$ are observable, the system is *completely observable*.

Example: (Gelb, Ex. 3.5-1, p. 69)

Is the system

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ 0 \end{bmatrix}, z = x_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{6.23}$$

observable? To answer this question, note that

$$\mathbf{H}^{\mathrm{T}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{\phi}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{6.24}$$

Therefore

$$\mathbf{Z} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \tag{6.25}$$

and $\det(\mathbf{Z}) = 0$. The system is NOT observable. If we measure only the sum of x_1 and x_2, we cannot determine both x_1, and x_2 since they are spectrally identical. This is clearer if we write the system as

$$\begin{aligned} \dot{x}_1 &= w_1 \\ \dot{x}_2 &= w_2 \\ \dot{x}_3 &= x_1 + x_2 \end{aligned}$$

If instead

$$\begin{aligned}\dot{x}_1 &= x_1 + w_1 \\ \dot{x}_2 &= w_2 \\ \dot{x}_3 &= x_1 + x_2\end{aligned}$$

the system would be observable. Check this yourself.

6.3 Controllability

Controllability is concerned with the effect of inputs upon model states. A discrete time system model is completely controllable if for any vectors, $\mathbf{x}_0$,$\mathbf{x}_N \in \mathcal{R}^n$, there exists a sequence $\mathbf{u}(0), \mathbf{u}(1), \ldots, \mathbf{u}(N-1)$, which can drive any $\mathbf{x}_0$ to state $\mathbf{x}_N$. Consider a linear time-invariant system model:

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}\mathbf{x}_k + \boldsymbol{\Lambda}\mathbf{u}_k, \tag{6.26}$$

where $\mathbf{x}_k$ is an n-vector, $\boldsymbol{\phi}$ is $n \times n$, $\mathbf{u}_k$ is an m-vector and $\boldsymbol{\Lambda}$ is $n \times m$. We can write $\mathbf{x}_N$ in terms of $\mathbf{x}_0$.

$$\begin{aligned}\mathbf{x}_N &= \boldsymbol{\phi}\mathbf{x}_{N-1} + \boldsymbol{\Lambda}\mathbf{u}_{N-1} \\ &= \boldsymbol{\phi}[\boldsymbol{\phi}\mathbf{x}_{N-2} + \boldsymbol{\Lambda}\mathbf{u}_{N-2}] + \boldsymbol{\Lambda}\mathbf{u}_{N-1} \\ &= \boldsymbol{\phi}^2\mathbf{x}_{N-2} + \boldsymbol{\phi}\boldsymbol{\Lambda}\mathbf{u}_{N-2} + \boldsymbol{\Lambda}\mathbf{u}_{N-1} \\ &= \boldsymbol{\phi}^3\mathbf{x}_{N-3} + \boldsymbol{\phi}^2\boldsymbol{\Lambda}\mathbf{u}_{N-3} + \boldsymbol{\phi}\boldsymbol{\Lambda}\mathbf{u}_{N-2} + \boldsymbol{\Lambda}\mathbf{u}_{N-1} \\ &= \boldsymbol{\phi}^N\mathbf{x}_0 + \boldsymbol{\phi}^{N-1}\boldsymbol{\Lambda}\mathbf{u}_0 + \boldsymbol{\phi}^{N-2}\boldsymbol{\Lambda}\mathbf{u}_1 + \ldots + \boldsymbol{\Lambda}\mathbf{u}_{N-1}\end{aligned} \tag{6.27}$$

If we define

$$\begin{aligned}\boldsymbol{\Theta} &= [\ \boldsymbol{\Lambda} \quad \boldsymbol{\phi}\boldsymbol{\Lambda} \quad \boldsymbol{\phi}^2\boldsymbol{\Lambda} \quad \ldots \quad \boldsymbol{\phi}^{N-1}\boldsymbol{\Lambda}\ ] \\ \mathbf{U} &= [\ \mathbf{u}_{N-1}^{\mathrm{T}} \quad \mathbf{u}_{N-2}^{\mathrm{T}} \quad \ldots \quad \mathbf{u}_1^{\mathrm{T}} \quad \mathbf{u}_0^{\mathrm{T}}\ ]^{\mathrm{T}}\end{aligned} \tag{6.28}$$

then we can write

$$\mathbf{x}_N = \boldsymbol{\phi}^N\mathbf{x}_0 + \boldsymbol{\Theta}\mathbf{U} \tag{6.29}$$

where $\boldsymbol{\Theta}$ is $n \times Nm$ and $\mathbf{U}$ is $Nm \times 1$. In order for $\mathbf{x}_N$ to be completely controllable, all $\mathbf{u}_i$, $0 \le i \le N-1$ must have an impact on $\mathbf{x}_0$. Therefore, the $n \times Nm$ matrix $\boldsymbol{\Theta}$ must have rank N.

Example: Consider the system:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -\alpha & 0 \\ 0 & -\beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u \tag{6.30}$$

Find α and β such that the system is controllable. Answer:

$$\boldsymbol{\Theta} = [\ \boldsymbol{\Lambda} \quad \boldsymbol{\phi}\boldsymbol{\Lambda}\ ] = \begin{bmatrix} 1 & -\alpha \\ 1 & -\beta \end{bmatrix} \tag{6.31}$$

so that $\det(\mathbf{\Theta})=-\beta+\alpha \neq 0$. Thus, the system is controllable if $\alpha \neq \beta$. If $\alpha = \beta$, the two first order systems are identical and there is no way a single input u could by itself produce different values for x_1 and x_2. If we consider instead,

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -\alpha & 1 \\ 0 & -\beta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ c \end{bmatrix} u \tag{6.32}$$

then

$$\mathbf{\Theta} = [\ \mathbf{\Lambda} \quad \boldsymbol{\phi}\mathbf{\Lambda}\ ] = \begin{bmatrix} 1 & -\alpha + c \\ c & -\beta c \end{bmatrix} \tag{6.33}$$

and $\det(\mathbf{\Theta})=c(-\beta+\alpha-c) \neq 0$. Thus, this system is controllable when $\alpha \neq \beta + c$.

More generally, for linear, discrete-time systems,

$$\begin{aligned} \mathbf{x}_{k+1} &= \boldsymbol{\phi}_k \mathbf{x}_k + \mathbf{w}_k, \\ \mathbf{z}_k &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{aligned} \tag{6.34}$$

is observable if and only if $\mathcal{I}(k,0) \geq 0$ where

$$\mathcal{I}(k, k-N) = \sum_{i=k-N}^{k} \boldsymbol{\phi}_{i,k}^{\mathrm{T}} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1} \mathbf{H}_i \boldsymbol{\phi}_{i,k}. \tag{6.35}$$

This system is completely controllable if and only if $\mathcal{C}(k,0) \geq 0$ where

$$\mathcal{C}(k, k-N) = \sum_{i=k-N}^{k} \boldsymbol{\phi}_{i,k}^{\mathrm{T}} \mathbf{Q}_i^{-1} \boldsymbol{\phi}_{i,k}. \tag{6.36}$$

Derivation of (6.35) and (6.36) are beyond the scope of this course. Note that observability does NOT depend on the observations, just on their location and accuracy and the system dynamics.

Suppose we have a linearized, global shallow water model with variables being the normal modes of the system, i.e. rotational and gravitational modes. If the gravity waves are unobserved, or poorly observed, $\sigma^2_{\mathrm{grav}} \to \infty$ so $\sigma^{-2}_{\mathrm{grav}} \to 0$. Therefore, $\mathbf{R}_i^{-1}$ has some 0 eigenvalues and is not positive definite. Thus the system is not observable because the gravity waves are not observable. This is the usual case for a global forecast model. The high frequency waves are not observable because the observation network is too widely spaced.

Similarly, controllability depends only on the model dynamics and model error covariance matrix. If some modes are very inaccurately predicted by the model, these modes cannot be controlled by specifying inputs.

For nonlinear system models, there is no condition to establish observability or controllability. Even for linear models, because of model truncation errors and round-off errors, it is hard to establish controllability and observability numerically.

6.4 Derivation of the discrete linear Kalman Filter

Consider the following stochastic-dynamic system:

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k \mathbf{x}_k + \mathbf{w}_k, \tag{6.37}$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k. \tag{6.38}$$

$\mathbf{w}_k$ is the model error and $\mathbf{v}_k$ is the observation error. The model state at time t_k is $\mathbf{x}_k$, an n-vector, while the observed state is $\mathbf{z}_k$, an m-vector. The transition matrix, $\boldsymbol{\phi}_k$ is then $n \times n$ and the observation operator, $\mathbf{H}_k$, is $m \times n$. Let us make the following assumptions about the stochastic inputs: they are unbiased and independent of each other. They are also white in time. This means that for all k and l,

$$\begin{aligned} < \mathbf{w}_k >= 0, \qquad & < \mathbf{w}_k(\mathbf{w}_l)^{\mathrm{T}} >= \mathbf{Q}_k\delta_l^k \\ < \mathbf{v}_k >= 0, \qquad & < \mathbf{v}_k(\mathbf{v}_l)^{\mathrm{T}} >= \mathbf{R}_k\delta_l^k \\ < \mathbf{w}_k(\mathbf{v}_l)^{\mathrm{T}} >= 0. & \end{aligned} \tag{6.39}$$

The equations (6.37) and (6.38) define our stochastic- dynamic model for our system. That is, if we had the true state, $\mathbf{x}$, and true measurements, $\mathbf{z}$, and our imperfect discrete forecast model, $\boldsymbol{\phi}_k$, and observation operator, $\mathbf{H}_k$, the state would evolve according to (6.37) and (6.38). Thus, we are assuming that we know the actual error statistics of the model error, $\mathbf{w}_k$, and the observation error, $\mathbf{v}_k$. The assumptions about unbiased errors and uncorrelated errors (in time or with each other) are not critical. Extensions of the standard KF can be derived should any of these assumptions not hold.

The KF problem is this: given a prior (background) estimate, $\hat{\mathbf{x}}_k^f$, of the system state at time t_k, what is the update or analysis, $\hat{\mathbf{x}}_k^a$, based on the measurements, $\mathbf{z}_k$? The background, $\hat{\mathbf{x}}_k^f$, bears a superscript f, referring to the fact that it is derived from a model forecast. The superscript a refers to the analysis, or estimate. We shall seek this estimate (or analysis) in the linear, recursive form:

$$\hat{\mathbf{x}}_k^a = \tilde{\mathbf{L}}_k\hat{\mathbf{x}}_k^f + \tilde{\mathbf{K}}_k\mathbf{z}_k. \tag{6.40}$$

To start this recursive process, we must have an initial estimate. The mean of the initial state is assumed to be given by a forecast,

$$< \mathbf{x}_0 >= \hat{\mathbf{x}}_0^f$$

and its error covariance matrix is given by

$$\mathbf{P}_0^f =< (\hat{\mathbf{x}}_0^f - \mathbf{x}_0)(\hat{\mathbf{x}}_0^f - \mathbf{x}_0)^{\mathrm{T}} > .$$

This initial estimate is assumed uncorrelated with the model and observation errors for all time. Let us define our errors. The analysis and forecast errors are:

$$\begin{aligned} \mathbf{e}_k^a &= \hat{\mathbf{x}}_k^a - \mathbf{x}_k \\ \mathbf{e}_k^f &= \hat{\mathbf{x}}_k^f - \mathbf{x}_k \end{aligned} \tag{6.41}$$

where $\mathbf{x}_k$ is the true state at time t_k.

The analysis equation can be rewritten in terms of errors by subtracting the truth from both sides of (6.40).

$$\begin{aligned} \mathbf{e}_k^a &= \tilde{\mathbf{L}}_k(\hat{\mathbf{x}}_k^f + \mathbf{x}_k - \mathbf{x}_k) - \mathbf{x}_k + \tilde{\mathbf{K}}_k\mathbf{z}_k \\ &= \tilde{\mathbf{L}}_k\mathbf{e}_k^f + \tilde{\mathbf{L}}_k\mathbf{x}_k - \mathbf{x}_k + \tilde{\mathbf{K}}_k(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k) \\ &= (\tilde{\mathbf{L}}_k + \tilde{\mathbf{K}}_k\mathbf{H}_k - \mathbf{I})\mathbf{x}_k + \tilde{\mathbf{L}}_k\mathbf{e}_k^f + \tilde{\mathbf{K}}_k\mathbf{v}_k. \end{aligned} \tag{6.42}$$

The advantage of writing the analysis equation in terms of errors is that we know something about these errors, namely, their means and covariances. Thus, we can use this information to help us determine an optimal estimate. First consider the bias of the analysis error. By definition, $< \mathbf{v}_k >= 0$. Now if the forecast error is unbiased, $< \mathbf{e}_k^f >= 0$, then

$$< \mathbf{e}_k^a > \;=\; (\tilde{\mathbf{L}}_k + \tilde{\mathbf{K}}_k\mathbf{H}_k - \mathbf{I}) < \mathbf{x}_k > . \tag{6.43}$$

If we want an unbiased estimate, then we must require that

$$\tilde{\mathbf{L}}_k = \mathbf{I} - \tilde{\mathbf{K}}_k\mathbf{H}_k \tag{6.44}$$

so that the estimator (6.40) becomes

$$\hat{\mathbf{x}}_k^a \;=\; \hat{\mathbf{x}}_k^f + \tilde{\mathbf{K}}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^f). \tag{6.45}$$

The estimation error can then be written as

$$\mathbf{e}_k^a \;=\; (\mathbf{I} - \tilde{\mathbf{K}}_k\mathbf{H}_k)\mathbf{e}_k^f + \tilde{\mathbf{K}}_k\mathbf{v}_k. \tag{6.46}$$

$\tilde{\mathbf{K}}_k$ is the weight matrix or gain matrix. $\tilde{\mathbf{K}}_k$ is the weight given to the observations. In the scalar case, $0 \leq \tilde{\mathbf{K}}_k \leq 1$. Different estimators will result in different choices for $\tilde{\mathbf{K}}_k$. It is interesting to note that the form of (6.45) is that of a linear combination of our two data sources: the background and the observations. What remains is to determine the Kalman gain, $\tilde{\mathbf{K}}_k$. To do this let us form the analysis error covariance matrix (i.e. the estimation error covariance matrix). By definition,

$$\begin{aligned}
\mathbf{P}_k^a &= < (\mathbf{e}_k^a)(\mathbf{e}_k^a)^{\mathrm{T}} > \\
&= (\mathbf{I} - \tilde{\mathbf{K}}_k\mathbf{H}_k) < (\mathbf{e}_k^f)(\mathbf{e}_k^f)^{\mathrm{T}} > (\mathbf{I} - \tilde{\mathbf{K}}_k\mathbf{H}_k)^{\mathrm{T}} + \tilde{\mathbf{K}}_k < (\mathbf{v}_k)(\mathbf{v}_k)^{\mathrm{T}} > \tilde{\mathbf{K}}_k^{\mathrm{T}} \\
&= (\mathbf{I} - \tilde{\mathbf{K}}_k\mathbf{H}_k)\mathbf{P}_k^f(\mathbf{I} - \tilde{\mathbf{K}}_k\mathbf{H}_k)^{\mathrm{T}} + \tilde{\mathbf{K}}_k\mathbf{R}_k\tilde{\mathbf{K}}_k^{\mathrm{T}}
\end{aligned} \tag{6.47}$$

Note that we used the fact that the observation error and background (forecast) errors are uncorrelated for all times. This fact can be proven by induction using the fact that the initial error is uncorrelated with the observation error. The last line of (6.47) is called Joseph's formula. This equation says that the analysis error is due to background and observation errors. If the observations are accurate and plentiful, $\tilde{\mathbf{K}}_k\mathbf{H}_k \approx \mathbf{I}$ so that the analysis error is given by the observation error projected onto state space. If there are no observations or they are very inaccurate, $\tilde{\mathbf{K}}_k = 0$ and the analysis error is given by the background error.

To update the estimate, we use our forecast model:

$$\hat{\mathbf{x}}_{k+1}^f \;=\; \boldsymbol{\phi}_k\hat{\mathbf{x}}_k^a. \tag{6.48}$$

The truth actually evolves according to (6.37) since our model is imperfect. Thus we can subtract (6.48) - (6.37) to define our forecast error. i.e.

$$\mathbf{e}_{k+1}^f \;=\; \boldsymbol{\phi}_k\mathbf{e}_k^a - \mathbf{w}_k. \tag{6.49}$$

Now if our analysis is unbiased, then the forecast is unbiased since our model error was assumed to be unbiased. Our forecast error covariance is by definition,

$$\begin{aligned}
\mathbf{P}_{k+1}^f &= < (\mathbf{e}_{k+1}^f)(\mathbf{e}_{k+1}^f)^{\mathrm{T}} > \\
&= < (\boldsymbol{\phi}_k\mathbf{e}_k^a - \mathbf{w}_k)(\boldsymbol{\phi}_k\mathbf{e}_k^a - \mathbf{w}_k)^{\mathrm{T}} > \\
&= \boldsymbol{\phi}_k < (\mathbf{e}_k^a)(\mathbf{e}_k^a)^{\mathrm{T}} > \boldsymbol{\phi}_k^{\mathrm{T}} + < (\mathbf{w}_k)(\mathbf{w}_k)^{\mathrm{T}} > \\
&= \boldsymbol{\phi}_k\mathbf{P}_k^a\boldsymbol{\phi}_k^{\mathrm{T}} + \mathbf{Q}_k.
\end{aligned} \tag{6.50}$$

Note that we used the fact that the analysis error and observation errors are uncorrelated at all times. This can be proven by induction using the fact that the initial state error and model errors are uncorrelated at all times. (6.50) describes the evolution of forecast errors. Forecast error is seen to be due to two terms: the amplification of analysis error at time step k through the dynamics of the model, and model error. The first term can also be viewed as predictability error. Thus, if our model dynamics contains unstable growing modes, analysis errors can grow. This error growth can be unbounded if we don't have observations to damp it (see (6.47)). If our model contains only decaying modes at time step k, the analysis error can actually be damped, if this term is larger than the model error term. If a perfect model assumption is made, the second term, $\mathbf{Q}_k = 0$.

Now, (6.47) and (6.50) together with the initial conditions, completely describe the evolution of forecast error covariances. Note that the evolution of forecast errors does not depend on the observations, the background or analyses, themselves. It depends only on the observation and model error covariances and on the observation distribution, $\mathbf{H}_k$.

The derivation here is simpler than that in Todling (1999) because observations are assumed to be available at every time step. Todling writes more generally that

$$\hat{\mathbf{x}}_k^f = \boldsymbol{\phi}_{k,k-l}\hat{\mathbf{x}}_{k-l}^f.$$

Thus, observations are available at every l time steps rather than at every time step. This notation uses the fact that when observations come every l steps we can use the transition matrix from time step $k-l$ to l. (Recall the transition property of transition matrices.) Thus, in (6.48), the update step is not the model time step but the interval between observations.

The Kalman filter is obtained from a particular choice of $\tilde{\mathbf{K}}_k$, that obtained by minimizing the analysis error variance. Thus we want to minimize:

$$\begin{aligned} \mathcal{J}_k^a &= E(|\mathbf{e}_k^a|_{\mathbf{S}_k}^2) \\ &= <(\mathbf{e}_k^a)^{\mathrm{T}}\mathbf{S}_k(\mathbf{e}_k^a)> \\ &= <\mathrm{Tr}(\mathbf{S}_k(\mathbf{e}_k^a)(\mathbf{e}_k^a)^{\mathrm{T}})> \\ &= \mathrm{Tr}(\mathbf{S}_k\mathbf{P}_k^a) \end{aligned} \tag{6.51}$$

where $\mathbf{S}_k$ is a positive definite scaling matrix. Setting the derivative with respect to $\tilde{\mathbf{K}}_k$ to 0 and substituting for $\mathbf{P}_k^a$ using (6.47) yields

$$\frac{d}{d\tilde{\mathbf{K}}_k}\mathcal{J}_k^a = \mathbf{S}_k\left[-2\mathbf{H}_k\mathbf{P}_k^f(\mathbf{I}-\tilde{\mathbf{K}}_k\mathbf{H}_k)^{\mathrm{T}} + 2\mathbf{R}_k\tilde{\mathbf{K}}_k^{\mathrm{T}}\right] = 0 \tag{6.52}$$

or

$$\tilde{\mathbf{K}}_k = \mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}. \tag{6.53}$$

$\mathbf{K}_k$ is the optimal weight matrix called the Kalman gain. For this choice of $\mathbf{K}_k$, (6.47) becomes

$$\mathbf{P}_k^a = (\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f \tag{6.54}$$

This is shown in problem 5.1.

Summary of discrete Kalman filter equations

system model:	$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k \mathbf{x}_k + \mathbf{w}_k, < \mathbf{w}_k >= 0, < \mathbf{w}_k (\mathbf{w}_l)^{\mathrm{T}} >= \mathbf{Q}_k \delta_l^k$
measurement model:	$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, < \mathbf{v}_k >= 0, < \mathbf{v}_k (\mathbf{v}_l)^{\mathrm{T}} >= \mathbf{R}_k \delta_l^k$
other assumptions:	$< \mathbf{w}_k (\mathbf{v}_l)^{\mathrm{T}} >= 0$
initial conditions:	$< \mathbf{x}_0 >= \hat{\mathbf{x}}_0^f, \mathbf{P}_0^f =< (\hat{\mathbf{x}}_0^f - \mathbf{x}_0)(\hat{\mathbf{x}}_0^f - \mathbf{x}_0)^{\mathrm{T}} >$
forecast step:	$\hat{\mathbf{x}}_{k+1}^f = \boldsymbol{\phi}_k \hat{\mathbf{x}}_k^a$
	$\mathbf{P}_{k+1}^f = \boldsymbol{\phi}_k \mathbf{P}_k^a \boldsymbol{\phi}_k^{\mathrm{T}} + \mathbf{Q}_k$
analysis step:	$\hat{\mathbf{x}}_k^a = \hat{\mathbf{x}}_k^f + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^f)$
	$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f$
Kalman gain:	$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k + \mathbf{R}_k)^{-1}$

Theorem: For the system (6.37) and (6.38) with $\mathbf{w}_k$, $\mathbf{v}_k$ uncorrelated zero mean processes and $< \mathbf{w}_k (\mathbf{w}_l)^{\mathrm{T}} >= \mathbf{Q}_k \delta_l^k$, $< \mathbf{v}_k (\mathbf{v}_l)^{\mathrm{T}} >= \mathbf{R}_k \delta_l^k$, and where $\mathbf{x}_0$ has mean $\hat{\mathbf{x}}_0^f$ and covariance $\mathbf{P}_0^f$ and is uncorrelated with $\mathbf{w}_k$ and $\mathbf{v}_k$, the Kalman filter is the best estimator of a certain type of linear estimator (6.40) in that it produces the smallest estimation error covariance.

If the errors are additionally assumed to be Gaussian, then the KF is the best estimator of any kind (linear or nonlinear). In this case, we can see that the Kalman filter is a recursive algorithm for estimating the *a posteriori* p.d.f. of the state given the observations. After an estimate is made at time step k, the analysis error covariance matrix can be computed, thus defining the complete *a posteriori* p.d.f. for time step k. The state estimate and error covariance are then propagated to the next time step according to the model dynamics. Using this information and the new observations at step $k+1$, a new estimate is made for this time step and the cycle is repeated. The KF will be examined from the conditional density viewpoint in section 6.7.

6.5 Simple KF examples

Example 1: No observations. With no observations, $\mathbf{K}_k = 0$ so that

$$\hat{\mathbf{x}}_k^a = \hat{\mathbf{x}}_k^f \tag{6.55}$$

$$\mathbf{P}_k^a = \mathbf{P}_k^f \tag{6.56}$$

$$\hat{\mathbf{x}}_{k+1}^f = \boldsymbol{\phi}_k \hat{\mathbf{x}}_k^a \tag{6.57}$$

$$\mathbf{P}_{k+1}^f = \boldsymbol{\phi}_k \mathbf{P}_k^a \boldsymbol{\phi}_k^{\mathrm{T}} + \mathbf{Q}_k \tag{6.58}$$

Thus, we can drop the analysis stage of the algorithm to get

$$\hat{\mathbf{x}}_{k+1}^f = \boldsymbol{\phi}_k \hat{\mathbf{x}}_k^f \tag{6.59}$$

$$\mathbf{P}_{k+1}^f = \boldsymbol{\phi}_k \mathbf{P}_k^f \boldsymbol{\phi}_k^{\mathrm{T}} + \mathbf{Q}_k \tag{6.60}$$

with initial conditions $\hat{\mathbf{x}}_0^f = \hat{\mathbf{x}}_0$, $\mathbf{P}_0^f = \mathbf{P}_0$. Therefore the model runs with no data injection. For neutral or unstable dynamics, the forecast error grows without bounds. During an NWP assimilation cycle, observations come every 6 hours, so the model runs for 6 hours (i.e. 6 time steps for the GEM model) without data. Thus during this time, the forecast error grows and is damped only when data arrives. This results in the typical "sawtooth" pattern of error variance evolution.

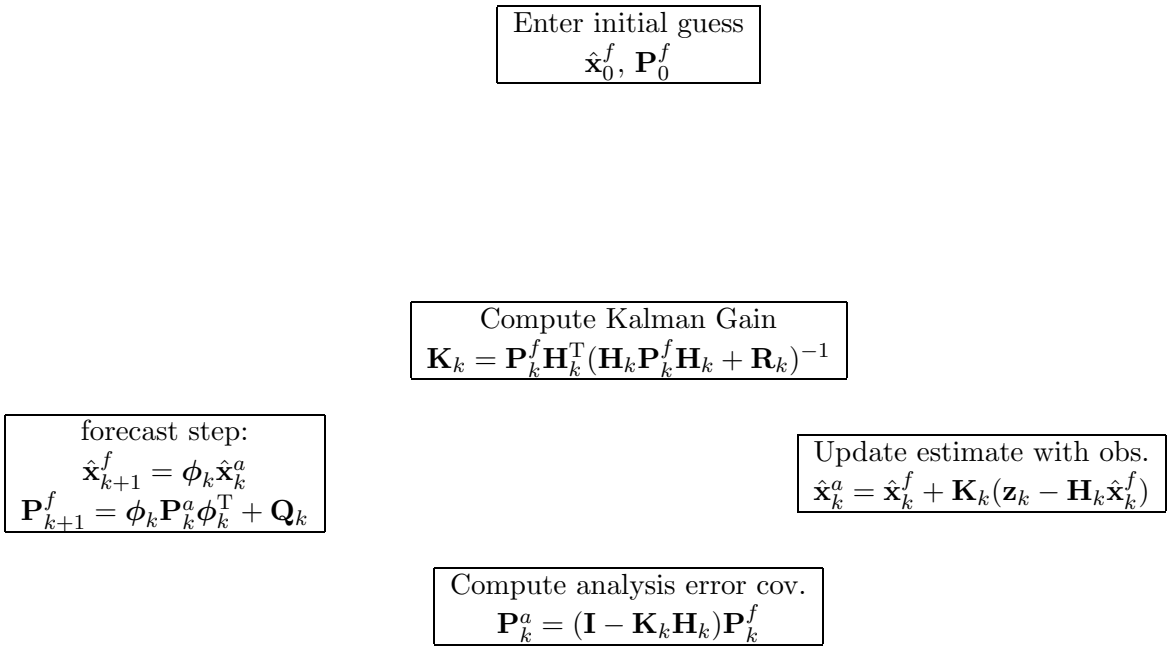

Figure 6.1: The Kalman Filter algorithm

Example 2: Perfect observations of model variables at every gridpoint. If observations are perfect then $\mathbf{R}_k = 0$. Observations of model variables at every gridpoint means that $\mathbf{H}_k = \mathbf{I}$. Thus the KF equations reduce to:

$$\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k)^{-1} = \mathbf{I} && (6.61)\\
\hat{\mathbf{x}}_k^a &= \hat{\mathbf{x}}_k^f + (\mathbf{z}_k - \hat{\mathbf{x}}_k^f) = \mathbf{z}_k && (6.62)\\
\mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}_k^f = 0 && (6.63)\\
\hat{\mathbf{x}}_{k+1}^f &= \boldsymbol{\phi}_k \hat{\mathbf{x}}_k^a = \boldsymbol{\phi}_k \mathbf{z}_k && (6.64)\\
\mathbf{P}_{k+1}^f &= \boldsymbol{\phi}_k \mathbf{P}_k^a \boldsymbol{\phi}_k^{\mathrm{T}} + \mathbf{Q}_k = \mathbf{Q}_k && (6.65)
\end{aligned}$$

This is the ideal: perfect observations at every gridpoint. In this case, we can again skip the analysis step since the observations are perfect. The KF reduces to

$$\begin{aligned}
\hat{\mathbf{x}}_0^f &= \mathbf{z}_0 && (6.66)\\
\mathbf{P}_0^f &= 0 && (6.67)\\
\hat{\mathbf{x}}_{k+1}^f &= \boldsymbol{\phi}_k \mathbf{z}_k && (6.68)\\
\mathbf{P}_{k+1}^f &= \mathbf{Q}_k. && (6.69)
\end{aligned}$$

Since $\mathbf{R}$ is the sum of instrument and representativeness error, $\mathbf{R} = 0$ means that only scales resolved by the model are observed. The forecast is then obtained by integrating the observed state. The forecast error is limited to the model error because the observations are perfect.

Example 3: Brownian motion Consider the scalar system:

$$x_{k+1} = x_k + w_k \quad (6.70)$$

$$z_k = x_k + v_k \tag{6.71}$$

where $\mathbf{Q}_k = 1$, $\mathbf{R}_k = 1/4$, $x_0^f = 0$ and $P_0^f = 0$. Also, in the KF equations, the transition matrix $\boldsymbol{\phi} = 1$ and the observation operation, $\mathbf{H}_k = 1$. Thus the KF equations are:

$$K_k = P_k^f (P_k^f + 1/4)^{-1} \tag{6.72}$$
$$P_k^a = (1 - K_k) P_k^f \tag{6.73}$$
$$P_{k+1}^f = P_k^a + 1 \tag{6.74}$$
$$\hat{x}_{k+1}^f = \hat{x}_k^a \tag{6.75}$$
$$\hat{x}_k^a = \hat{x}_k^f + K_k(z_k - \hat{x}_k^f) \tag{6.76}$$

The Kalman gain can be written in terms of P_k^a:

$$K_k = \frac{P_{k-1}^a + 1}{P_{k-1}^a + 5/4}$$

so that the update for the error variance can be written:

$$P_k^a = \frac{P_{k-1}^a + 1}{4P_{k-1}^a + 5}$$

To summarize:

$$\hat{x}_k^a = \hat{x}_{k-1}^a + \frac{P_{k-1}^a + 1}{P_{k-1}^a + 5/4}(z_k - \hat{x}_{k-1}^a) \tag{6.77}$$
$$P_k^a = \frac{P_{k-1}^a + 1}{4P_{k-1}^a + 5}. \tag{6.78}$$

To start, note that $K_0 = 0$ so that $\hat{x}_0^a = 0$ and $P_0^a = 0$. Then for $k = 1$,

$$K_1 = \frac{1}{5/4} = \frac{4}{5} = 0.8$$
$$\hat{x}_1^a = 0 + K_1(z_1 - 0) = \frac{4}{5} z_1 \tag{6.79}$$
$$P_1^a = \frac{1}{5} = 0.2. \tag{6.80}$$

For $k = 2$,

$$K_2 = \frac{\frac{1}{5} + 1}{\frac{1}{5} + \frac{5}{4}} = \frac{24}{29} \approx 0.827$$
$$\hat{x}_2^a = \frac{4}{5} z_1 + K_2(z_2 - \frac{4}{5} z_1) = \frac{4}{29} z_1 + \frac{24}{29} z_2 \tag{6.81}$$
$$P_2^a = \frac{\frac{1}{5} + 1}{\frac{4}{5} + 5} = \frac{6}{29} \approx 0.207 \tag{6.82}$$

What happens in the limit that $k \to \infty$? When $k \to \infty$, $P_k \approx P_{k-1}$. Thus

$$P_\infty^a = \frac{P_\infty^a + 1}{4P_\infty^a + 5}.$$

The solutions are

$$P_\infty^a = \frac{1}{2}(-1 \pm \sqrt{2}).$$

The positive definite solution is

$$P_\infty^a = \frac{1}{2}(-1 + \sqrt{2}) \approx 0.2071$$

and

$$K_\infty = \frac{P_\infty^a + 1}{P_\infty^a + 5/4} = \frac{2 + 2\sqrt{2}}{3 + 2\sqrt{2}} \approx 0.828$$

In this case, the KF approaches the steady state filter in just two steps! This is because the dynamics are neutral and the observation error covariance $R = 1/4 << Q = 1$. The observations are relatively accurate compared to the model error. The state, being scalar, is completely observed when an observation is available. Thus dense, accurate observations combined with steady, linear dynamics lead to a stable filter.

6.6 The information filter

In this section we derive a different form of the discrete KF, which is very useful when we have no background information.

First recall the equation for the analysis error covariance, (6.54) and substitute for the optimal gain using (6.53):

$$\begin{aligned} \mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f \\ &= \mathbf{P}_k^f - \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}\mathbf{H}_k\mathbf{P}_k^f. \end{aligned} \tag{6.83}$$

Another form for the analysis error covariance equation is

$$(\mathbf{P}_k^a)^{-1} = (\mathbf{P}_k^f)^{-1} + \mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k. \tag{6.84}$$

To verify that (6.84) is correct, multiply (6.83) by it:

$$\begin{aligned} & [\mathbf{P}_k^f - \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}\mathbf{H}_k\mathbf{P}_k^f][(\mathbf{P}_k^f)^{-1} + \mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k] \\ = \;& \mathbf{I} - \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}\mathbf{H}_k - \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k + \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k \\ = \;& \mathbf{I} + \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}[(\mathbf{R}_k)^{-1} - (\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}(\mathbf{I} + \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1})]\mathbf{H}_k \\ = \;& \mathbf{I} + \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}[(\mathbf{R}_k)^{-1} - (\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}(\mathbf{R} + \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}})(\mathbf{R}_k)^{-1}]\mathbf{H}_k \\ = \;& \mathbf{I} + \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}[(\mathbf{R}_k)^{-1} - (\mathbf{R}_k)^{-1}]\mathbf{H}_k \\ = \;& \mathbf{I}. \end{aligned} \tag{6.85}$$

Now that we have verifed that (6.84) is correct, let us substitute this into (6.53) to get

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k + \mathbf{R}_k)^{-1} \\ &= [(\mathbf{P}_k^f)^{-1} + \mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k]^{-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1} \\ &= [(\mathbf{P}_k^a)^{-1}]^{-1}\mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1} \\ &= \mathbf{P}_k^a\mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1}. \end{aligned} \tag{6.86}$$

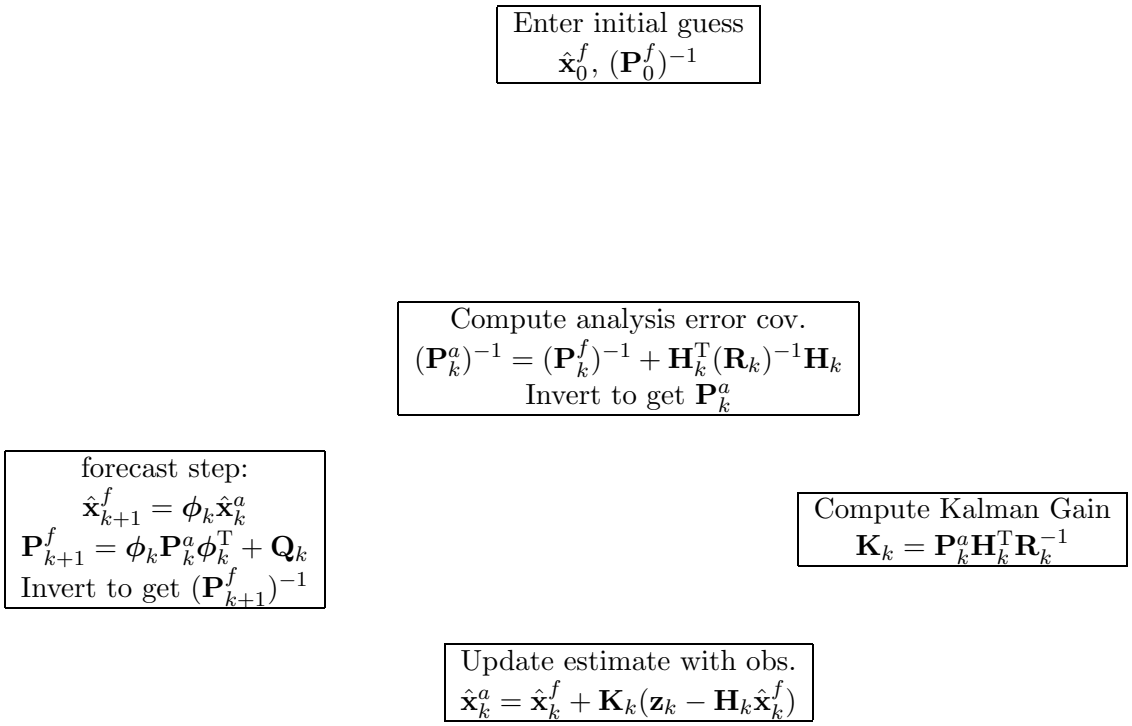

Figure 6.2: The alternative Kalman Filter algorithm

To get the second line, we used the Sherman-Morrison-Woodbury formula defined in Chapter 4, equation (5.27). Now we can summarize this form of the KF:

Summary of the information filter equations

system model:	$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k\mathbf{x}_k + \mathbf{w}_k, < \mathbf{w}_k >= 0, < \mathbf{w}_k(\mathbf{w}_l)^{\mathrm{T}} >= \mathbf{Q}_k\delta_l^k$
measurement model:	$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k, < \mathbf{v}_k >= 0, < \mathbf{v}_k(\mathbf{v}_l)^{\mathrm{T}} >= \mathbf{R}_k\delta_l^k$
other assumptions:	$< \mathbf{w}_k(\mathbf{v}_l)^{\mathrm{T}} >= 0$
initial conditions:	$< \mathbf{x}_0 >= \hat{\mathbf{x}}_0^f, \mathbf{P}_0^f =< (\hat{\mathbf{x}}_0^f - \mathbf{x}_0)(\hat{\mathbf{x}}_0^f - \mathbf{x}_0)^{\mathrm{T}} >$
forecast step:	$\hat{\mathbf{x}}_{k+1}^f = \boldsymbol{\phi}_k\hat{\mathbf{x}}_k^a$
	$\mathbf{P}_{k+1}^f = \boldsymbol{\phi}_k\mathbf{P}_k^a\boldsymbol{\phi}_k^{\mathrm{T}} + \mathbf{Q}_k$
analysis step:	$\hat{\mathbf{x}}_k^a = \hat{\mathbf{x}}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^f)$
	$(\mathbf{P}_k^a)^{-1} = (\mathbf{P}_k^f)^{-1} + \mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1}\mathbf{H}_k$
Kalman gain:	$\mathbf{K}_k = \mathbf{P}_k^a\mathbf{H}_k^{\mathrm{T}}(\mathbf{R}_k)^{-1}$

This process is depicted in Fig. 6.2.

Example 6.1 *Estimation of a random constant.*

Suppose we have a sequence of noisy measurements and we wish to estimate a random constant. The deterministic model is then

$$\mathbf{x}_{k+1} = \mathbf{x}_k.$$

The transition matrix, $\boldsymbol{\phi}_k = \mathbf{I}$, and the model error covariance matrix is $\mathbf{Q}$=$\mathbf{0}$. The measurements are $\mathbf{z} = [z_1, z_2, \ldots, z_N]^{\mathrm{T}}$ and the measurement error variance is the same for all of them. Thus,

the measurement equation can be written as

$$\mathbf{z} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \mathrm{x} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}$$

and the observation error covariance is $\mathbf{R} = \sigma^2\mathbf{I}$. *There is no initial background state. To represent this lack of knowledge, we set*

$$\mathbf{P}_0^f = \infty, \qquad \hat{\mathbf{x}}_0^f = 0.$$

We can't deal with infinite variances with the usual KF algorithm, but the alternative algorithm can be used.

Proceeding with the first step of the alternative form of the KF:

$$\begin{aligned} (\mathbf{P}_0^a)^{-1} &= (\mathbf{P}_0^f)^{-1} + \mathbf{H}_0^{\mathrm{T}}(\mathbf{R}_0)^{-1}\mathbf{H}_0 \\ &= [\ 1 \quad 1 \quad \cdots \quad 1\ ]\sigma^{-2}\mathbf{I} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \\ &= \frac{N}{\sigma^2}. \end{aligned} \tag{6.87}$$

Thus

$$\mathbf{P}_0^a = \frac{\sigma^2}{N}.$$

Next, the Kalman gain is computed as

$$\begin{aligned} \mathbf{K}_0 &= \mathbf{P}_0^a\mathbf{H}_0^{\mathrm{T}}(\mathbf{R}_0)^{-1} \\ &= \frac{\sigma^2}{N}[\ 1 \quad 1 \quad \cdots \quad 1\ ]\sigma^{-2}\mathbf{I} \\ &= \frac{1}{N}[\ 1 \quad 1 \quad \cdots \quad 1\ ] \end{aligned} \tag{6.88}$$

Finally, the estimate is given by

$$\begin{aligned} \hat{\mathbf{x}}_0^a &= \hat{\mathbf{x}}_0^f + \mathbf{K}_0(\mathbf{z} - \mathbf{H}_0\hat{\mathbf{x}}_0^f) \\ &= \mathbf{K}_0\mathbf{z} \\ &= \frac{1}{N}[\ 1 \quad 1 \quad \cdots \quad 1\ ] \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{bmatrix} \\ &= \frac{z_1}{N} + \frac{z_2}{N} + \ldots + \frac{z_N}{N}. \end{aligned} \tag{6.89}$$

Thus the Kalman filter says that the best estimate is given by the average of all the observations in the absence of a background.

Example 6.2 *A recursive algorithm.*

In the previous example, we took all observations at the same time, since the dynamics were steady. We could equally well have done the problem, one observation at a time, using N time steps. In this case, the measurement equation is simply the scalar equation:

$$z_k = x_k + v_k.$$

Here, $\mathrm{H}_k = 1$ and the observation error covariance matrix is a scalar: $\mathbf{R}_k = \sigma^2$. Now let's start again. Since there is no observation at time t_0, $\mathbf{K}_0 = 0$ and $\hat{\mathrm{x}}_1^f = \hat{\mathrm{x}}_0^a = \hat{\mathrm{x}}_0^f = 0$ and $\mathbf{P}_1^f = \mathbf{P}_0^a = \infty$. The first step is

$$(\mathrm{P}_1^a)^{-1} = (\mathrm{P}_1^f)^{-1} + \mathrm{H}_1^{\mathrm{T}}\mathrm{R}_1^{-1}\mathrm{H}_1 = \sigma^{-2}. \tag{6.90}$$

Thus, $\mathrm{P}_1^a = \sigma^2$. Next, the gain is

$$\mathrm{K}_1 = \mathrm{P}_1^a\mathrm{H}_1^{\mathrm{T}}\mathrm{R}_1^{-1} = \sigma^2\sigma^{-2} = 1.$$

Then the updated estimate is

$$\hat{\mathrm{x}}_1^a = \hat{\mathrm{x}}_1^f + \mathrm{K}_1(z_1 - \mathrm{H}_1\hat{\mathrm{x}}_1^f) = z_1.$$

In the forecast step, we obtain

$$\begin{aligned} \hat{\mathrm{x}}_2^f &= \hat{\mathrm{x}}_1^a = z_1 \\ \mathrm{P}_2^f &= \mathrm{P}_1^a = \sigma^2. \end{aligned}$$

In the next iteration, we start with the analysis equations again:

$$\begin{aligned} (\mathrm{P}_2^a)^{-1} &= (\mathrm{P}_2^f)^{-1} + \mathrm{H}_2^{\mathrm{T}}\mathrm{R}_2^{-1}\mathrm{H}_2 = 2\sigma^{-2} \\ \mathrm{K}_2 &= \mathrm{P}_2^a\mathrm{H}_2^{\mathrm{T}}\mathrm{R}_2^{-1} = 1/2 \\ \hat{\mathrm{x}}_2^a &= \hat{\mathrm{x}}_2^f + \mathrm{K}_2(z_2 - \mathrm{H}_2\hat{\mathrm{x}}_2^f) = z_1 + 0.5z_2 - 0.5z_1 = 0.5(z_1 + z_2) \end{aligned} \tag{6.91}$$

then follow with the forecast step:

$$\begin{aligned} \hat{\mathrm{x}}_3^f &= \hat{\mathrm{x}}_2^a = 0.5(z_1 + z_2) \\ \mathrm{P}_3^f &= \mathrm{P}_2^a = 0.5\sigma^2. \end{aligned}$$

The next analysis step yields:

$$\begin{aligned} (\mathrm{P}_3^a)^{-1} &= (\mathrm{P}_3^f)^{-1} + \mathrm{H}_3^{\mathrm{T}}\mathrm{R}_3^{-1}\mathrm{H}_3 = 3\sigma^{-2} \\ \mathrm{K}_3 &= \mathrm{P}_3^a\mathrm{H}_3^{\mathrm{T}}\mathrm{R}_3^{-1} = 1/3 \\ \hat{\mathrm{x}}_3^a &= (1 - \mathrm{K}_3\mathrm{H}_3)\hat{\mathrm{x}}_3^f + \mathrm{K}_3 z_3 = \frac{2}{3}\frac{1}{2}(z_1 + z_2) + \frac{1}{3}z_3 = \frac{1}{3}(z_1 + z_2 + z_3). \end{aligned} \tag{6.92}$$

It should now be apparent that with each step of the Kalman filter, an observation is ingested and the estimate is a running average. This procedure can be continued until step N. The estimate would be as found in the previous example, except, instead of taking all observations together, we used them one by one in a recursive manner.

6.7 The conditional density viewpoint

We have derived the KF by minimizing the analysis error variance. However, in the previous chapter, we showed that the minimum variance estimator is the conditional mean. How does this relate to the KF derivation? In this section, we re-interpret the KF problem from the conditional density viewpoint.

Suppose we have $\mathbf{Z}_k = (z_0, z_1, \ldots, z_k)^{\mathrm{T}}$ the set of all observations up to and including t_k. What does knowledge of $\mathbf{Z}_k$ tell us about the state $\mathbf{x}_k$? We know from chapter 4 that we ultimately want to know $p_{\mathbf{x}_k|\mathbf{Z}_k}(\mathbf{x}_k|\mathbf{Z}_k)$. We also know that the minimum variance estimator is the conditional mean, i.e.

$$\mathbf{x}_k^a = E(\mathbf{x}_k|\mathbf{Z}_k).$$

Now using the definition of conditional density, we can rewrite this p.d.f. in terms of those we know. i.e.

$$\begin{aligned} p_{\mathbf{x}_k|\mathbf{Z}_k}(\mathbf{x}_k|\mathbf{Z}_k) &= p_{\mathbf{x}_k|\mathbf{Z}_{k-1},\mathbf{z}_k}(\mathbf{x}_k|\mathbf{Z}_{k-1},\mathbf{z}_k) \\ &= \frac{p(\mathbf{x}_k,\mathbf{Z}_{k-1},\mathbf{z}_k)}{p(\mathbf{Z}_{k-1},\mathbf{z}_k)} \\ &= \frac{p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{Z}_{k-1})p(\mathbf{x}_k,\mathbf{Z}_{k-1})}{p(\mathbf{Z}_{k-1},\mathbf{z}_k)} \\ &= \frac{p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{Z}_{k-1})p(\mathbf{x}_k|\mathbf{Z}_{k-1})p(\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})p(\mathbf{Z}_{k-1})} \\ &= \frac{p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{Z}_{k-1})p(\mathbf{x}_k|\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})} \end{aligned} \tag{6.93}$$

The above was derived by repeatedly applying Bayes' theorem. After the second equality, the subscripts on the probabilities were dropped for convenience. Now, if we assume that all quantities are Gaussian, then we can completely specify the *a posteriori* p.d.f. $p(\mathbf{x}_k|\mathbf{Z}_k)$. Therefore, assume that $\mathbf{w}_k$ is $\mathrm{N}(\mathbf{0}, \mathbf{Q}_k)$, $\mathbf{v}_k$ is $\mathrm{N}(\mathbf{0}, \mathbf{R}_k)$, $\mathbf{x}_0$ is $\mathrm{N}(\hat{\mathbf{x}}_0^f, \mathbf{P}_0)$, and $\mathbf{w}_k$ and $\mathbf{x}_k$ are white in time and uncorrelated with each other. Also the initial state error is assumed uncorrelated with the model and observations errors at all times. Because the observation error is white, we can write

$$p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{Z}_{k-1}) = p(\mathbf{z}_k|\mathbf{x}_k)$$

so the conditional p.d.f. we are looking for (6.93) can be simplified to

$$p_{\mathbf{x}_k|\mathbf{Z}_k}(\mathbf{x}_k|\mathbf{Z}_k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Z}_{k-1})}{p(\mathbf{z}_k|\mathbf{Z}_{k-1})}. \tag{6.94}$$

What are the terms in (6.94)? In ch. 4, we already determined $p(\mathbf{z}_k|\mathbf{x}_k)$ (except without subscripts). For Gaussian errors (as assumed here),

$$p(\mathbf{z}_k|\mathbf{x}_k) \sim \mathrm{N}(\mathbf{H}_k\boldsymbol{\mu}_k, \mathbf{R}_k). \tag{6.95}$$

To determine the next term, we need to have at our disposal, a prior estimate (its mean and error covariance), $\hat{\mathbf{x}}_k^f$, $\mathbf{P}_k^f$. We shall stretch our notation to let $\mathbf{x}_k$ be the true state conditioned on the data stream, $\mathbf{Z}_{k-1}$. Then,

$$p(\mathbf{x}_k|\mathbf{Z}_{k-1}) \sim \mathrm{N}(\hat{\mathbf{x}}_k^f, \mathbf{P}_k^f). \tag{6.96}$$

We can determine what the background estimate's mean is. It is the conditional mean of $\mathbf{x}_k$ given $\mathbf{Z}_{k-1}$. Thus

$$\begin{aligned}\hat{\mathbf{x}}_k^f &= E(\mathbf{x}_k|\mathbf{Z}_{k-1}) \\ &= E(\boldsymbol{\phi}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}|\mathbf{Z}_{k-1}) \\ &= \boldsymbol{\phi}_{k-1}E(\mathbf{x}_{k-1}|\mathbf{Z}_{k-1}) + E(\mathbf{w}_{k-1}|\mathbf{Z}_{k-1}) \\ &= \boldsymbol{\phi}_{k-1}\hat{\mathbf{x}}_{k-1}^a. \end{aligned} \tag{6.97}$$

This equation says that to update the mean of the analysis estimate, propagate it in time using the model dynamics to get the background estimate for the next analysis step. When we derived the KF earlier in this note, we assumed exactly this update equation on an *ad hoc* basis. Here, however, we have produced exactly the same update/forecast step, justifying our previous choice.

We can similarly develop the covariance matrix for the background. It is given by the conditional covariance of the state at time t_k given the sequence of observations, $\mathbf{Z}_{k-1}$.

$$\begin{aligned}\mathbf{P}_k^f &= E(\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}|\mathbf{Z}_{k-1}) \\ &= E((\boldsymbol{\phi}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1})(\boldsymbol{\phi}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1})^{\mathrm{T}}|\mathbf{Z}_{k-1}) \\ &= \boldsymbol{\phi}_{k-1}E(\mathbf{x}_{k-1}\mathbf{x}_{k-1}^{\mathrm{T}}|\mathbf{Z}_{k-1})\boldsymbol{\phi}_{k-1}^{\mathrm{T}} + E(\mathbf{w}_{k-1}\mathbf{w}_{k-1}^{\mathrm{T}}|\mathbf{Z}_{k-1}) \\ &= \boldsymbol{\phi}_{k-1}\mathbf{P}_k^a\boldsymbol{\phi}_{k-1}^{\mathrm{T}} + \mathbf{Q}_k. \end{aligned} \tag{6.98}$$

The cross terms have disappeared from the third line because of the whiteness assumption of model errors. Because our prior (background) estimate is coming from the evolution of a previous analysis, the error of the state includes the error of using the conditional mean update. Thus, the forecast error includes not only the evolution of the analysis error but also the model error.

What remains now is to determine the denominator of (6.94). Since $\mathbf{z}_k = \mathbf{x}_k + \mathbf{H}_k\mathbf{v}_k$ we can determine the mean and covariance of $p(\mathbf{z}_k|\mathbf{Z}_{k-1})$.

$$\begin{aligned}E(\mathbf{z}_k|\mathbf{Z}_{k-1}) &= E((\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k)|\mathbf{Z}_{k-1}) \\ &= \mathbf{H}_kE(\mathbf{x}_k|\mathbf{Z}_{k-1}) + E(\mathbf{v}_k|\mathbf{Z}_{k-1}) \\ &= \mathbf{H}_k\hat{\mathbf{x}}_k^f. \end{aligned} \tag{6.99}$$

The covariance is

$$\begin{aligned}E(\mathbf{z}_k\mathbf{z}_k^{\mathrm{T}}|\mathbf{Z}_{k-1}) &= E((\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k)(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k)^{\mathrm{T}}|\mathbf{Z}_{k-1}) \\ &= \mathbf{H}_kE(\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}|\mathbf{Z}_{k-1})\mathbf{H}_k^{\mathrm{T}} + E(\mathbf{v}_k\mathbf{v}_k^{\mathrm{T}}|\mathbf{Z}_{k-1}) \\ &= \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k \end{aligned} \tag{6.100}$$

As a result we can say that

$$p(\mathbf{z}_k|\mathbf{Z}_{k-1}) \sim \mathrm{N}(\mathbf{H}_k\hat{\mathbf{x}}_k^f, \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k). \tag{6.101}$$

Finally, we can substitute (6.95),(6.96) and (6.101) into (6.94) to get

$$p_{\mathbf{x}_k|\mathbf{Z}_k}(\mathbf{x}_k|\mathbf{Z}_k) = \frac{\mathrm{N}(\mathbf{H}_k\boldsymbol{\mu}_k, \mathbf{R}_k)\mathrm{N}(\hat{\mathbf{x}}_k^f, \mathbf{P}_k^f)}{\mathrm{N}(\mathbf{H}_k\hat{\mathbf{x}}_k^f, \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)} = \mathrm{N}(\hat{\mathbf{x}}_k^a, \mathbf{P}_k^a) \tag{6.102}$$

where

$$\begin{aligned} \hat{\mathbf{x}}_k^a &= \hat{\mathbf{x}}_k^f + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^f) \\ \mathbf{K}_k &= \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k + \mathbf{R}_k)^{-1} \\ (\mathbf{P}_k^a)^{-1} &= (\mathbf{P}_k^f)^{-1} + \mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{H}_k \end{aligned} \tag{6.103}$$

and the update step is given by (6.97) and (6.98). Thus, we can now see that in chapter 4, we derived the analysis step of the Kalman filter (with no update equations for the background). Thus 3D schemes such as OI and 3DVAR can be viewed as degraded KF's where the covariance update is omitted since it is assumed constant in time. Also, the schemes lose the optimality property because the error covariance matrices are only approximated and not known exactly.

For Gaussian errors, the conditional mean is also the most likely value (the mode or MAP estimator). Therefore the KF is the best estimator by almost any measure. If the errors are not assumed Gaussian, then the estimator $\hat{\mathbf{x}}(\mathbf{z})$ will be a (not necessarily linear) function of $\mathbf{z}$. If we seek an estimator which is a linear function of $\mathbf{z}$, then the KF is the estimator which corresponds to the minimum variance. Thus it is said to have the Best Linear Unbiased Estimator (BLUE) property. However the KF would not give a conditional mean estimate which would be the minimum variance estimate since a nonlinear estimator may have lower variance than the linear one. Still the KF would provide the minimum variance of all linear estimators.

We have seen that the KF provides a procedure for updating the conditional p.d.f., in the Gaussian case since on the first 2 moments are needed to define the p.d.f. The conditional means and covariances are:

$$\begin{aligned} \hat{\mathbf{x}}_k^f &= E(\mathbf{x}_k|\mathbf{Z}_{k-1}) & \mathbf{P}_k^f &= E(\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}|\mathbf{Z}_{k-1}) \\ \hat{\mathbf{x}}_k^a &= E(\mathbf{x}_k|\mathbf{Z}_k) & \mathbf{P}_k^a &= E(\mathbf{x}_k\mathbf{x}_k^{\mathrm{T}}|\mathbf{Z}_k) \end{aligned} \tag{6.104}$$

6.8 Properties of the discrete KF

1. The KF is a linear discrete-time finite dimensional system.
2. The input to the filter is $\{\mathbf{z}_k\}$ and the output is $\{\hat{\mathbf{x}}_k\}$.
3. The conditional covariances

$$\begin{aligned} \mathbf{P}_k^f &= E[(\mathbf{x}_k - E(\mathbf{x}_k))(\mathbf{x}_k - E(\mathbf{x}_k))^{\mathrm{T}}|\mathbf{Z}_{k-1}] \\ &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_k^f)(\mathbf{x}_k - \hat{\mathbf{x}}_k^f)^{\mathrm{T}}|\mathbf{Z}_{k-1}] \\ &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_k^f)(\mathbf{x}_k - \hat{\mathbf{x}}_k^f)^{\mathrm{T}}] \end{aligned}$$

$$\begin{aligned} \mathbf{P}_k^a &= E[(\mathbf{x}_k - E(\mathbf{x}_k))(\mathbf{x}_k - E(\mathbf{x}_k))^{\mathrm{T}}|\mathbf{Z}_k] \\ &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_k^a)(\mathbf{x}_k - \hat{\mathbf{x}}_k^a)^{\mathrm{T}}|\mathbf{Z}_k] \\ &= E[(\mathbf{x}_k - \hat{\mathbf{x}}_k^a)(\mathbf{x}_k - \hat{\mathbf{x}}_k^a)^{\mathrm{T}}] \end{aligned}$$

   are INDEPENDENT of the observations and are therefore also the unconditional covariances.
4. When $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are jointly Gaussian, then $\{\mathbf{x}_k|\mathbf{Z}_{k-1}\}$ can be shown to be Gaussian. The conditional p.d.f. is completely defined by the conditional mean and covariance. Therefore, the KF describes a procedure for updating the entire conditional p.d.f. of $\mathbf{x}_k$.

5. In the Kalman gain definition, the inverse of $\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$ is required. Since $\mathbf{P}_k^f$ and $\mathbf{R}_k$ are covariance matrices, they are positive semi-definite. One way to force positive definiteness of $\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$ is to require that $\mathbf{R}_k$ be positive definite. i.e. Thus no measurement is allowed to be perfect. Since $\mathbf{R}_k$ actually includes representativeness error as well, this is not unrealistic.

6. If the forecast model is time invariant and the input and output noises are stationary, $\boldsymbol{\phi}_k$, $\mathbf{H}_k$, $\mathbf{Q}_k$, $\mathbf{R}_k$ are all constant. However the KF will in general be time-varying.

7. If $\{\mathbf{x}_k\}$ and $\{\mathbf{z}_k\}$ are not Gaussian, the KF describes the Best Linear minimum variance estimator (BLUE property). However, it does not give a conditional mean analysis which would be the minimum variance estimator because a nonlinear estimator may have lower variance.

8. The innovation sequence is white. Let

$$\begin{aligned} \mathbf{d}_k &= \mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^f \\ &= \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^f \qquad (6.105) \\ &= \mathbf{H}_k(\boldsymbol{\phi}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}) + \mathbf{v}_k - \mathbf{H}_k\boldsymbol{\phi}_{k-1}\hat{\mathbf{x}}_{k-1}^a \\ &= -\mathbf{H}_k\boldsymbol{\phi}_{k-1}\mathbf{e}_{k-1}^a + \mathbf{H}_k\mathbf{w}_{k-1} + \mathbf{v}_k \qquad (6.106) \end{aligned}$$

   then $< \mathbf{d}_k >= 0$ (from (6.106), and

$$< \mathbf{d}_k\mathbf{d}_k^{\mathrm{T}} >= \mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k$$

   from (6.105). Because $\mathbf{e}_{k-1}^a$, $\mathbf{w}_{k-1}$ and $\mathbf{v}_k$ are all independent of $\mathbf{Z}_{k-1}$, then $\mathbf{d}_k$ is also independent of $\mathbf{Z}_{k-1}$. However, by definition, $\mathbf{d}_j$ for $j < k$ are all linear functions of $\mathbf{Z}_{k-1}$. Therefore $\mathbf{d}_k$ is independent of $\mathbf{d}_j$ for $j < k$. Thus $\mathbf{d}_k$ is a WHITE sequence. This fact is important because it means that we have a way to monitor the filter's performance while it is operating: check the whiteness of $\mathbf{d}_k$ (it is already computed as part of the KF).

6.9 Filter divergence

Filter divergence is the name given to the phenomenon where the filter seems to behave well, with low predicted analysis error variance, while the analysis is actually drifting away from reality.

Divergence occurs when our modelling assumptions are not true. For example, the model error is higher than we said it was, or the system model has the wrong form, or the system is unstable or has bias errors when none was expected. When you have assumed large model errors, and have a very stable model and no biases, divergence problems are avoided.

How do you know if divergence is occuring if you don't know the truth? You can monitor the innovation sequence. If it is not white, then some of your assumptions must be inappropriate. Another indication is that the Kalman gain will tend to 0 as t increases. This is not a necessary nor a sufficient condition for filter divergence. However, it indicates that less and less weight is being given to the data and the filter may be asymptoting to an erroneous value.

How do you cope with divergence?

1. Don't under-estimate model errors. Over-estimate them.

2. Try to adaptively estimate model errors using innovations.

3. Overweight most recent data, thereby reducing filter memory of old data. This forces data into the KF.

4. Put an *ad hoc* lower bound on the Kalman gain.

6.10 Stability of the KF

Asymptotic stability of the KF means that its solution will gradually become insensitive to its initial conditions. One can see that observability will play a role because if there are sufficient observations, the true state will be well approximated. Also, controllability will play a role because if the system is not controllable in some modes, then any number of observations cannot help damp the analysis errors.

Theorem (Jaswinski Theorem 7.4)
If the system

$$\begin{aligned} \mathbf{x}_{k+1} &= \boldsymbol{\phi}_k \mathbf{x}_k + \mathbf{w}_k \\ \mathbf{z}_k &= \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \end{aligned}$$

with $\mathbf{x}_0$, $\{\mathbf{w}_k\}$, $\{\mathbf{v}_k\}$, independent, is uniformly completely observable and uniformly completely controllable and if $\mathbf{P}_0 \geq 0$ then the discrete time KF is uniformly asymptotically stable.

What does this mean? The KF analysis may be written as

$$\hat{\mathbf{x}}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\boldsymbol{\phi}_{k-1} \hat{\mathbf{x}}_{k-1}^a + \mathbf{K}_k \mathbf{z}_k.$$

For the unforced equation, it is easy to see that stability of the KF solution ($\hat{\mathbf{x}}_k^a$ remains bounded) depends on

$$|(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\boldsymbol{\phi}_{k-1}|.$$

In fact

$$\hat{\mathbf{x}}_k^a = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\boldsymbol{\phi}_{k-1}(\mathbf{I} - \mathbf{K}_{k-1} \mathbf{H}_{k-1})\boldsymbol{\phi}_{k-2} \ldots (\mathbf{I} - \mathbf{K}_1 \mathbf{H}_1)\boldsymbol{\phi}_0 \hat{\mathbf{x}}_0^a.$$

Let us examine the operator:

$$(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\boldsymbol{\phi}_{k-1}.$$

$\mathbf{K}_k$ represents the observation accuracy, while $\mathbf{H}_k$ reflects the observation distribution in space. $\boldsymbol{\phi}_{k-1}$ represents the propagation of the model from t_{k-1} to t_k.

- For accurate, abundant observations, $\mathbf{K}_k \mathbf{H}_k \approx \mathbf{I}$. Then even for unstable dynamics, the KF is stable.

- If the model is very stable, even with no observations ($\mathbf{K}_k = 0$), the KF eventually loses memory of the initial condition.

- If the model is unstable, sufficient observations (in terms of accuracy and coverage) are needed to keep errors from growing.

6.11 Generalizations of the Kalman Filter

When we derived the KF, we made a number of assumptions such as the system and measurement noise being white in time, and uncorrelated with each other. It was noted that these assumptions were not critical to the KF derivation. In this section, we describe modifications to the standard KF when these assumptions are not true.

First recall the true system equation, (6.37), and the measurement equation, (6.38):

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k.$$

Here we have allowed for the system noise to be a different dimension than the state. Thus we need a mapping from the system noise space to the model space, $\mathbf{G}_k$.

6.11.1 Serially correlated measurement noise

If the measurement noise is correlated (not white) in time, then we can write

$$\mathbf{v}_{k+1} = \mathbf{E}_k \mathbf{v}_k + \mathbf{u}_k \tag{6.107}$$

where $\mathbf{u}_k$ is white in time and Gaussian with mean $\mathbf{0}$ and covariance $\mathbf{S}_k$. Because the measurement noise is not white, we will try to define a new measurement equation which is white. First note that

$$\begin{aligned} \mathbf{z}_{k+1} &= \mathbf{H}_{k+1}\mathbf{x}_{k+1} + \mathbf{v}_{k+1} \\ &= \mathbf{H}_{k+1}\boldsymbol{\phi}_k \mathbf{x}_k + \mathbf{H}_{k+1}\mathbf{G}_k \mathbf{w}_k + \mathbf{E}_k \mathbf{v}_k + \mathbf{u}_k. \end{aligned} \tag{6.108}$$

Now it is clear that

$$\mathbf{z}_{k+1} - \mathbf{E}_k \mathbf{z}_k = (\mathbf{H}_{k+1}\boldsymbol{\phi}_k - \mathbf{E}_k \mathbf{H}_k)\mathbf{x}_k + \mathbf{H}_{k+1}\mathbf{G}_k \mathbf{w}_k + \mathbf{u}_k. \tag{6.109}$$

Thus we can define a new measurement equation:

$$\mathbf{z}_k^* = \mathbf{H}_k^* \mathbf{x}_k + \mathbf{v}_k^*. \tag{6.110}$$

where

$$\mathbf{z}_k^* = \mathbf{z}_{k+1} - \mathbf{E}_k \mathbf{z}_k,$$

$$\mathbf{H}_k^* = \mathbf{H}_{k+1}\boldsymbol{\phi}_k - \mathbf{E}_k \mathbf{H}_k,$$

and

$$\mathbf{v}_k^* = \mathbf{H}_{k+1}\mathbf{G}_k \mathbf{w}_k + \mathbf{u}_k.$$

Note that $\mathbf{v}_k^*$ is also white and Gaussian with mean $\mathbf{0}$ and covariance matrix, $\mathbf{H}_{k+1}\mathbf{G}_k \mathbf{Q}_k \mathbf{G}_k^{\mathrm{T}} \mathbf{H}_{k+1}^{\mathrm{T}} + \mathbf{S}_k$. With this new measurement equation, and the system equation, a KF can be defined with the usual assumptions.

6.11.2 Correlated system and measurement noise

What if the system and measurement noise were correlated, i.e.

$$< \mathbf{w}_k(\mathbf{v}_k)^{\mathrm{T}} >= \mathbf{C}_k\delta_l^k.$$

Here $\mathbf{w}_k$ and $\mathbf{v}_k$ must have the same dimension. (If not, new mapping operators must be introduced.) Consider the following state equation:

$$\begin{aligned} \mathbf{x}_{k+1} &= \boldsymbol{\phi}_k\mathbf{x}_k + \mathbf{G}_k\mathbf{w}_k + \mathbf{D}_k(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \mathbf{v}_k) \\ &= (\boldsymbol{\phi}_k - \mathbf{D}_k\mathbf{H}_k)\mathbf{x}_k + \mathbf{D}_k\mathbf{z}_k + \mathbf{G}_k\mathbf{w}_k - \mathbf{D}_k\mathbf{v}_k. \end{aligned} \tag{6.111}$$

Let us define

$$\mathbf{s}_k = \mathbf{G}_k\mathbf{w}_k - \mathbf{D}_k\mathbf{v}_k.$$

Now, choose $\mathbf{D}_k$ such that $< \mathbf{s}_k(\mathbf{v}_k)^{\mathrm{T}} >=0$. Thus,

$$\mathbf{D}_k = \mathbf{G}_k\mathbf{C}_k\mathbf{R}_k^{-1}.$$

Thus we have defined a new system equation for which the noise $\mathbf{s}_k$ is uncorrelated with the measurement noise, $\mathbf{v}_k$. Note that the system equation now has inputs $\mathbf{z}_k$.

6.11.3 Serially correlated system noise

In this case, the system noise $\mathbf{w}_k$ is not white but is Markov:

$$\mathbf{w}_{k+1} = \mathbf{A}_k\mathbf{w}_k + \mathbf{u}_k \tag{6.112}$$

where $\mathbf{u}_k$ is white in time and Gaussian with mean $\mathbf{0}$ and covariance matrix $\mathbf{S}_k$. We can combine the system equation and the model for system noise to get an augmented state, $\mathbf{y}_k^{\mathrm{T}} = [\mathbf{x}_l^{\mathrm{T}}\mathbf{w}_k^{\mathrm{T}}]$. Thus we can write

$$\begin{pmatrix} \mathbf{x}_{k+1} \\ \mathbf{w}_{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\phi}_k & \mathbf{G}_k \\ \mathbf{0} & \mathbf{A}_k \end{pmatrix} \begin{pmatrix} \mathbf{x}_k \\ \mathbf{w}_k \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix} \mathbf{u}_k \tag{6.113}$$

or

$$\mathbf{y}_{k+1} = \mathbf{M}_k\mathbf{y}_k + \tilde{\mathbf{G}}_k\mathbf{u}_k. \tag{6.114}$$

Thus we have a new system equation for which the system noise is white. However, the new state has dimension equal to the sum of the dimensions of vectors $\mathbf{x}$ and $\mathbf{w}$. The measurement equation can also be written in terms of this new augmented state as:

$$\mathbf{z}_k = \tilde{\mathbf{H}}_k\mathbf{y}_k + \mathbf{v}_k, \tag{6.115}$$

where $\tilde{\mathbf{H}} = [\mathbf{H}_k\,\mathbf{0}]$.

6.12 Nonlinear filtering

The Kalman filter is applicable to linear systems. In Earth systems science, we usually deal with nonlinear systems so in this section we introduce a standard extension to the Kalman Filter for nonlinear models.

6.12.1 The Extended Kalman Filter

Consider the following stochastic-dynamic system:

$$\mathbf{x}_{k+1} = M_k(\mathbf{x}_k) + \mathbf{w}_k \tag{6.116}$$
$$\mathbf{z}_k = H_k(\mathbf{x}_k) + \mathbf{v}_k \tag{6.117}$$

where M_k represents the model which is a nonlinear function of the state at time step k. H_k is the nonlinear observation operator. If the model and observation operators are only weakly nonlinear, then we can approximate them using the first two terms of a Taylor expansion:

$$\begin{aligned} M_k(\bar{\mathbf{x}} + \delta\mathbf{x}) &\approx M_k(\bar{\mathbf{x}}) + \frac{\partial M_k(\bar{\mathbf{x}})}{\partial \mathbf{x}}\delta\mathbf{x} \\ H_k(\bar{\mathbf{x}} + \delta\mathbf{x}) &\approx H_k(\bar{\mathbf{x}}) + \frac{\partial H_k(\bar{\mathbf{x}})}{\partial \mathbf{x}}\delta\mathbf{x}. \end{aligned}$$

Here we have introduced the Tangent Linear Model (TLM),

$$\frac{\partial M_k(\bar{\mathbf{x}})}{\partial \mathbf{x}} = \mathbf{M}_k,$$

and the Tangent Linear observation operator

$$\frac{\partial H_k(\bar{\mathbf{x}})}{\partial \mathbf{x}} = \mathbf{H}_k.$$

Since the forecast step is given by:

$$\mathbf{x}^f_{k+1} = M_k(\mathbf{x}^a_k), \tag{6.118}$$

the forecast error evolves according to

$$\begin{aligned} \mathbf{e}^f_{k+1} &= M_k(\mathbf{x}^a_k) - M_k(\mathbf{x}_k) - \mathbf{w}_k \\ &= M_k(\mathbf{x}^a_k + \mathbf{x}_k - \mathbf{x}_k) - M_k(\mathbf{x}_k) - \mathbf{w}_k \\ &\approx \mathbf{M}_k\mathbf{e}^a_k - \mathbf{w}_k. \end{aligned} \tag{6.119}$$

Now, if the analysis error at step k is unbiased, then the forecast error at step $k+1$ is unbiased and the error covariance at $k+1$ is

$$\mathbf{P}^f_{k+1} = \mathbf{M}_k\mathbf{P}^a_k\mathbf{M}^{\mathrm{T}}_k + \mathbf{Q}_k. \tag{6.120}$$

Because the forecast error evolution is described by the Tangent Linear model, the forecast error covariance also evolves according to the TLM dynamics. Note that this equation is only valid when the model is weakly nonlinear.

As in OI or 3DVAR with a nonlinear observation operator, the analysis step for the Extended Kalman Filter is given by:

$$\boxed{\begin{aligned} \mathbf{x}^a_k &= \mathbf{x}^f_k + \mathbf{K}_k(\mathbf{z}_k - H_k(\mathbf{x}^f_k)) \\ \mathbf{K}^{\mathrm{T}}_k &= \mathbf{P}^f_k\mathbf{H}^{\mathrm{T}}_k(\mathbf{H}_k\mathbf{P}^f_k\mathbf{H}^{\mathrm{T}}_k + \mathbf{R}_k)^{-1} \\ \mathbf{P}^a_k &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}^f_k \end{aligned}}$$

Thus the full nonlinear model is used to obtain the innovations for the analysis equation. The gain matrix and the analysis error covariance matrix involve the tangent linear forward model operator.

The forecast step of the Extended Kalman Filter (EKF) is summarized as:

$$\boxed{\begin{array}{lcl} \mathbf{x}^f_{k+1} & = & M_k(\mathbf{x}^a_k) \\ \mathbf{P}^f_{k+1} & = & \mathbf{M}_k\mathbf{P}^a_k\mathbf{M}^{\mathrm{T}}_k + \mathbf{Q}_k \end{array}}$$

with initial conditions: $\mathbf{x}^a_0 = \mathbf{x}_0$ and $\mathbf{P}^a_0 = \mathbf{P}_0$.

An important assumption made in deriving the EKF is that model is weakly nonlinear. When this assumption is valid, the EKF should work, but in general, the EKF is not a consistent method for propagating the mean and covariance of the *a posteriori* p.d.f. To see why, we must first define some higher order statistical moments. For components x_i of vector $\mathbf{x}$, the first statistical moment is:

$$\bar{\mathrm{x}}_i =< \mathrm{x}_i >$$

the second statistical moment is:

$$\mathbf{P}_{ij} =< (\mathrm{x}_i - \bar{\mathrm{x}}_i)(\mathrm{x}_j - \bar{\mathrm{x}}_j) >$$

the third statistical moment is:

$$\mathbf{\Theta}_{ijk} =< (\mathrm{x}_i - \bar{\mathrm{x}}_i)(\mathrm{x}_j - \bar{\mathrm{x}}_j)(\mathrm{x}_k - \bar{\mathrm{x}}_k) >$$

and the fourth statistical moment is:

$$\mathbf{\Gamma}_{ijkl} =< (\mathrm{x}_i - \bar{\mathrm{x}}_i)(\mathrm{x}_j - \bar{\mathrm{x}}_j)(\mathrm{x}_k - \bar{\mathrm{x}}_k)(\mathrm{x}_l - \bar{\mathrm{x}}_l) >$$

where the subscript indicating time step was dropped.

Now consider the nonlinear model (6.116). If we had many model integrations, the mean state at time step $k+1$ would be given by

$$\bar{\mathbf{x}}_{k+1} = \overline{M_k(\mathbf{x}_k)}.$$

Note that the right hand side is not the forecast of the mean state at time step k but rather the mean of the forecasts from time step k. For a linear model, these two would be the same. Now define the TLM, as well as higher order derivatives of the model:

$$\mathbf{M}_k = \frac{\partial M_k(\bar{\mathbf{x}})}{\partial \mathbf{x}} \quad \mathbf{S}_k = \frac{\partial^2 M_k(\bar{\mathbf{x}})}{\partial \mathbf{x}^2} \quad \mathbf{T}_k = \frac{\partial^3 M_k(\bar{\mathbf{x}})}{\partial \mathbf{x}^3}.$$

Then the Taylor expansion for the model is:

$$M(\bar{\mathbf{x}} + \delta\mathbf{x}) = M(\bar{\mathbf{x}}) + \mathbf{M}\delta\mathbf{x} + \frac{1}{2}\mathbf{S}(\delta\mathbf{x}\delta\mathbf{x}^{\mathrm{T}}) + \frac{1}{6}\mathbf{T}(\delta\mathbf{x}\delta\mathbf{x}\delta\mathbf{x}) + \ldots$$

Substitute this into the nonlinear model to get:

$$\mathbf{x}_{k+1} = M_k(\bar{\mathbf{x}}_k) + \mathbf{w}_k + \mathbf{M}_k\delta\mathbf{x} + \frac{1}{2}\mathbf{S}_k(\delta\mathbf{x}_k\delta\mathbf{x}^{\mathrm{T}}_k) + \frac{1}{6}\mathbf{T}_k(\delta\mathbf{x}_k\delta\mathbf{x}_k\delta\mathbf{x}_k) + \ldots$$

Thus the mean state at time step $k+1$ is given by:

$$\bar{\mathbf{x}}_{k+1} = M_k(\bar{\mathbf{x}}_k) + \frac{1}{2}\mathbf{S}_k\mathbf{P}^a_k + \frac{1}{6}\mathbf{T}_k\mathbf{\Theta}_k + \ldots \tag{6.121}$$

Similarly, we can derive the evolution equation for the forecast error covariance matrix by taking the outer product of the forecast error (minus its mean) with itself. The result is:

$$\begin{aligned}\mathbf{P}^f_{k+1} &= \mathbf{M}_k\mathbf{P}^a_k\mathbf{M}^{\mathrm{T}}_k + \mathbf{Q}_k + \mathbf{M}_k\mathbf{\Theta}_k\mathbf{S}^{\mathrm{T}}_k \\ &+ \frac{1}{4}\mathbf{S}_k\mathbf{\Gamma}_k\mathbf{S}^{\mathrm{T}}_k + \frac{1}{3}\mathbf{M}_k\mathbf{\Gamma}_k\mathbf{T}^{\mathrm{T}}_k + \frac{1}{4}\mathbf{S}_k\mathbf{P}^a_k(\mathbf{P}^a_k)^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}_k \\ &- \frac{1}{6}\mathbf{S}_k\mathbf{P}^a_k\mathbf{\Theta}^{\mathrm{T}}_k\mathbf{T}^{\mathrm{T}}_k - \frac{1}{36}\mathbf{T}_k\mathbf{\Theta}_k{\mathbf{\Theta}_k}^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}_k + \ldots \end{aligned} \tag{6.122}$$

Note that the evolution of the mean state and the covariance matrix involves all higher order statistical moments. Obviously, we can't afford to calculate all of these, so to close the system of equations, we might choose to keep only up to the second order statistics. Thus we would drop all terms involving $\mathbf{\Theta}_k$ and $\mathbf{\Gamma}_k$ and higher order terms. Then the mean state and covariance evolution would be given by

$$\begin{aligned}\bar{\mathbf{x}}_{k+1} &= M_k(\bar{\mathbf{x}}_k) + \frac{1}{2}\mathbf{S}_k\mathbf{P}^a_k \\ \mathbf{P}^f_{k+1} &= \mathbf{M}_k\mathbf{P}^a_k\mathbf{M}^{\mathrm{T}}_k + \mathbf{Q}_k + \frac{1}{4}\mathbf{S}_k\mathbf{P}^a_k(\mathbf{P}^a_k)^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}_k. \end{aligned} \tag{6.123}$$

Compared to the EKF forecast equations, (6.118) and (6.120), there are extra terms not only for the covariance propagation but also for the mean state propagation equation. Thus the EKF is not a consistent approximation even to second order statistics. However, if the model is weakly nonlinear so that $\mathbf{S}_k$ can be ignored, then the EKF forecast equations result.

The problem with the EKF forecast equations can be evident when dealing with geophysical forecast models. The TLM is supposed to describe the evolution of forecast errors. With a nonlinear model, error growth would saturate eventually. However, with a linear model, error growth can continue indefinitely. When do errors grow? Error growth occurs when the model is unstable. For atmospheric models, in the midlatitudes, this corresponds to growing baroclinic disturbances. Thus precisely when and where we are interested in an accurate forecast, the EKF equations are incorrect. The forecast error covariances can be overestimated and if these regions are not sufficiently well observed, the error growth can lead to filter divergence. Gauthier et al. (1993) showed such unrealistic error growth using the barotropic vorticity equation. Using a perfect model, the true initial conditions and observations of u,v and ϕ at 500 hPa, the EKF still failed. Unbounded error growth was found where the flow was unstable, and where there were no observations to damp the error growth. The the EKF can diverge simply due to the presence of dynamical instabilities since a linearized forecast error evolution was assumed.

In summary, a standard extension of the KF for weakly nonlinear models (the so-called Extended Kalman Filter or EKF) was derived. The EKF was shown to be an inconsistent approximation to the full nonlinear evolution equations, even to second order. However, if the nonlinear model is weakly nonlinear, the omitted terms may be small and the EKF may work. When the model is not weakly nonlinear, the forecast for the state may develop a bias, and the covariance evolution may be incorrect.

6.13 Sub-Optimal Kalman Filters

The KF is designed for linear systems. In reality, we are dealing not only with nonlinear systems, but very large and complex ones. Thus the KF is not practical. The main problem is the evolution

of the covariance matrix. If the state is of dimension 10^7, then the covariance matrix has 10^{14} elements and propagation of this matrix in time would be impossible. Thus several people have come up with schemes to reduce the cost of the covariance propagation equation. Because the KF equations are altered, and therefore no longer optimal (this is not guaranteed in the nonlinear case anyway), such scheme are called "Sub-Optimal Schemes" or SOS.

As noted above, the expensive part of the EKF is the covariance forecast:

$$\mathbf{P}^f_{k+1} = \mathbf{M}_k \mathbf{P}^a_k \mathbf{M}^{\mathrm{T}}_k + \mathbf{Q}_k = \mathbf{P}^p_k + \mathbf{Q}_k.$$

The expensive part of this calculation is the first term on the right side which is called the predictability term, $\mathbf{P}^p_k$. To avoid matrix multiplication, this term can be rewritten as

$$\mathbf{M}_k (\mathbf{M}_k \mathbf{P}^a_k)^{\mathrm{T}}.$$

The term in round brackets describes the evolution of columns of the analysis error covariance matrix using the TLM. If $\mathbf{M}_k$ is $n \times n$, then this involves n model integrations. Then the final multipliation involves an addition n integrations. In total, this term involves $2n$ model integrations. If $n = 10^7$, then clearly this is impractical.

In addition to the fact that the covariance evolution step is expensive, the EKF approximation of this step is not even correct so why both evaluating this term exactly according to the EKF equation? Finally, and most importantly, we do not have sufficient knowledge of observation and especially model errors, so what is the point in evolving erroneous covariances exactly? All of these arguments point to the necessity of simplifying the EKF forecast step for covariances.

One method of reducing the cost of (6.120) is to use a simpler model to propagate the forecast errors. So, if a full NWP model is used for the analysis, use a reduced resolution model to propagate the errors. Since we know the analysis step involves a filtering, the analysis errors may involve larger scales anyway so using a reduced resolution model might be sufficient. This method was tried by Cohn and Todling (1996).

Another method also due to Cohn and Tolding (1996) is the Partial singular value decomposition (or PSKF). The idea here is to decompose the model dynamics into singular vectors and keep only some of the most important directions. Thus,

$$\mathbf{M}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{V}_k$$

where

$$\mathbf{U}_k = [\mathbf{U}^L_k \mathbf{U}^T_k],\ \mathbf{V}_k = [\mathbf{V}^L_k \mathbf{V}^T_k],\ \mathbf{D}_k = \mathrm{diag}[\mathbf{D}^L_k \mathbf{D}^T_k].$$

The left and right singular vectors are $\mathbf{U}_k$ and $\mathbf{V}_k$. The superscript L refers to the leading singular values and corresponding vectors while T refers to the trailing values and vectors. Thus the approximation involves dropping the trailing singular values and vectors so that (6.120) becomes:

$$\mathbf{P}^p_k = \mathbf{S}_k \mathbf{P}^a_k \mathbf{S}^{\mathrm{T}}_k + \mathbf{P}^T_k$$

where

$$\mathbf{S}_k = \mathbf{U}^L_k \mathbf{D}^L_k (\mathbf{V}^L_k)^{\mathrm{T}}.$$

Another method (Partial Eigenvalue decomposition PEKF) is to use an eigenvalue decomposition of the whole prediction error covariance matrix:

$$\mathbf{M}_k \mathbf{P}^a_k \mathbf{M}^{\mathrm{T}}_k = \mathbf{W}_k \mathbf{T}_k \mathbf{W}^{\mathrm{T}}_k$$

where the leading and trailing eigenvalues and vectors are given by:

$$\mathbf{W}_k = [\mathbf{W}_k^L \mathbf{W}_k^T],\ \mathbf{T}_k = \mathrm{diag}[\mathbf{T}_k^L \mathbf{T}_k^T].$$

Thus (6.120) becomes:

$$\mathbf{P}_k^p = \mathbf{W}_k^L \mathbf{T}_k^L (\mathbf{W}_k^L)^{\mathrm{T}} + \mathbf{T}_k^T.$$

This method is also due to Cohn and Todling (1996).

Is it better to approximate the model dynamics by keeping only a few singular vectors, or to approximate the whole prediction error covariance matrix using a partial eigenvalue decomposition? It depends on the model in question. If the spectrum of singular values has a steep slope so that some values are clearly unimportant, it is easy to define the leading and trailing values. Similarly, for the eigenvalue decomposition, if there is a steep spectrum of eigenvalues, then it is clear how to define what the leading and trailing values are. From Cohn and Todling (1995)'s Fig. 1, the eigenvalues vary over 10 orders of magnitude while the singular vectors vary over only 1 order of magnitude. However they were using a simple barotropic model and these results will depend on the model.

Another variation of the PEKF is the RRSQRT-EKF. The idea is the same as the PEKF in that the first term of (6.120) is approximated using only the leading eigenvalues and eigenvectors. However, instead of the standard EKF algorithm, the square root (SQRT-EKF) algorithm is used. The latter is more robust because one doesn't have to worry about the development of non symmetric positive definite covariance matrices (due to numerical errors). This can happen with the standard EKF forecast equation for covariances. This algorithm is described in Verlaan and Heemink (1995).

An obvious approximation to (6.120) is to simply use a different (simpler) model to propagate the forecast errors than was used to propagate the state. For example Dee (1991) tried propagating the mass variable only and deriving the winds from the geostrophic relationship.

One can also try approximating the forecast error covariance by a banded matrix. However, negative values can still result unless a square root formulation is used (Boggs et al. 1995).

In summary there are many ways to approximate the covariance forecast equation of the EKF. All will have advantages and disadvantanges, and before one chooses a method, it is necessary to determine the assumptions involved and whether they are valid. It should also be noted that there are many more sub optimal schemes than have been mentioned here.

6.14 The Ensemble Kalman Filter

One of the most intriguing of the KF approximations is the so-called "Ensemble Kalman Filter" or EnsKF. The idea behind the EnsKF is to dispense with the troublesome covariance propagation equation of the EKF and directly evaluate the forecast error covariance matrix using an ensemble of forecasts. Since a nonlinear model is used for the forecasts, the covariance so calculated does not involve any linearization and hence is better than the EKF. Thus, the ensemble mean forecast is

$$< \mathbf{x}_k^f >= \frac{1}{s} \sum_{i=1}^{s} \mathbf{x}_k^{f,i}$$

and the ensemble-based forecast error covariance matrix is

$$\mathbf{P}_k^f \approx \frac{1}{s-1} \sum_{i=1}^{s} (\mathbf{x}_k^{f,i} - < \mathbf{x}_k^f >)(\mathbf{x}_k^{f,i} - < \mathbf{x}_k^f >)^{\mathrm{T}}. \tag{6.124}$$

Thus, you run not one but s analyses, using the observation set $\mathbf{z}$ but perturbed by a different realization of the observation error, for each analysis. This is to ensure that the analyses are reasonably different from each other so as to reflect an analysis error. Then you can propagate the s ensemble members forward in time using the complete nonlinear model. No TLM or adjoint models are required! This gives an ensemble of s forecasts which are then used to calculate $\mathbf{P}_k^f$ using (6.124). This $\mathbf{P}_k^f$ is then used in the calculation of the Kalman gain matrix for the analysis step. To start the EnsKF you need to generate an ensemble of s perturbations of the initial state which are used to calculate $\mathbf{P}_0^f$. With chaotic model dynamics (as is the case for atmospheric motion), ensemble members will diverge in time but saturate at the climatological values of forecast error. Thus the ensemble members will grow apart with time. During the analysis step, the ensemble spread is reduced because all members are constrained by the same observations (but with different random errors added).

The calculation (6.124) will underestimate forecast error covariances because each member is actually used in the calculation of its own error covariance. Thus, it is better to exclude the ith member when calculating the ith covariance matrix:

$$\mathbf{P}_k^{f,j} \approx \frac{1}{s-2} \sum_{i \neq j}^{s} (\mathbf{x}_k^{f,i} - < \mathbf{x}_k^{f,j} >)(\mathbf{x}_k^{f,i} - < \mathbf{x}_k^{f,j} >)^{\mathrm{T}}. \tag{6.125}$$

Thus the gain matrix for the ith member does not involve the ith member.

The appeal of the EnsKF is that it is conceptually simple but valid, requires no TLM or adjoint models (which takes months to develop) and can be easily parallelized for MPI machines. However, there are a few tricky points. First of all, the definition of a covariance matrix involves realizations of forecast error while (6.124) uses the spread of forecasts about their mean to represent forecast error. Thus there is an underlying assumption that the ensemble mean is close to the truth and that the distribution of ensemble members about their mean truly represents the forecast error.

Because there is no assumption of weakly nonlinear dyanmics, the EnsKF can be applied to highly nonlinear models. The EnsKF was introduced to the meteorological literature by Evensen (1994) to avoid the EKF divergence due to dynamic instabilities. It was applied by Evensen (1997) to demonstrate its use for low order chaotic systems.

REFERENCES

1. Anderson, B. D. O., and J. B. Moore, 1979: *Optimal filtering.* Prentice-Hall, 357 pp.

2. Boggs, D., M. Ghil, and C. Keppenne, 1995: A stabilized spare-matrix U-D square-root implementation of a large-state extended Kalman filter. *2nd WMO symposium on Assimilation of observations in Meteorology and Oceanography*, 219-224.

3. Brown, R. G. and P. Y. C. Hwang, 1997: *Introduction to random signals and applied Kalman filtering.* John Wiley and sons, 484 pp.

4. Cohn, S. E. and R. Todling, 1995: Approximate Kalman Filters for unstable dynamics. *2nd WMO symposium on Assimilation of observations in Meteorology and Oceanography*, 241-246.

5. Cohn, S. E. and R. Todling, 1996: Approximate Data assimilation schemes for stable and unstable dynamics *J. Meteor. Soc. Japan*, **74**, 63-75.