5.1 Monte Carlo/Ensemble Methods

5.1.1 Ensemble Methods and Particle Filters

When a model is non-linear, one of the fundamental computational steps of the Kalman filter (and any related calculation such as a smoother) is no longer possible. Consider a Kalman filter problem in which the initial conditions, $\tilde{\mathbf{x}}(0)$, contain errors characterized by the covariance matrix, $\mathbf{P}(0)$, and the zero-mean disturbance or control, $\mathbf{u}(0)$, is unknown with covariance, $\mathbf{Q}(0)$. The state forecast step, Eq. (4.49) proceeds as before. Computation of the forecast error covariance, Eq. (4.50), however, which sums the error owing to the initial conditions and that of the unknown controls, $\mathbf{A}(0)\mathbf{P}(0)\mathbf{A}(0)^T + \mathbf{\Gamma Q \Gamma}^T$, depends directly upon the linearity assumption, and can no longer be carried out rigorously. For weak nonlinearities, the extended or linearized Kalman filters and associated smoothers may be adequate. But when linearizing assumptions fail, some other method must be used. A commonly discussed example of a non-linear (but scalar) model is,

$$\begin{aligned} x_t &= \frac{1}{2}x_{t-1} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos 1.2t + \varepsilon_t, \\ y_t &= \frac{x_t^2}{20} + \eta_t, \end{aligned}$$

Continued on next page...

which is nonlinear in both the evolution equation and in the measurement; ε_t, η_t are Gaussian white noise processes.[152] Extended Kalman filters work badly for this low-dimensional example.

The basic idea behind so-called ensemble or Monte Carlo methods is in some ways even simpler than the use of Eq. (4.50). (In the signal processing literature, closely related approaches are usually called "sequential Monte Carlo methods," or "particle filtering".[153]) One directly simulates a sufficiently large number of forecasts, $\tilde{\mathbf{x}}^{(i)}(t,-)$, all having the same statistical properties, that $\mathbf{P}(t,-)$ can be estimated by brute force computation from the many simulations. If a sufficiently large ensemble can be generated, one can contemplate estimating not just the second moments, but calculating the empirical frequency function for the forecast step.

To see how this approach might work, generate an ensemble of initial conditions, $\tilde{\mathbf{X}}(0)$, where each column of the $N \times L$, $L < N$ matrix corresponds to a possible initial condition consistent with both $\tilde{\mathbf{x}}(0)$ and $\mathbf{P}(0)$. Form a similar ensemble for $\tilde{\mathbf{u}}(0)$ based upon $\langle \mathbf{u}(0) \rangle = 0$, and $\mathbf{Q}(0)$. (We discuss generation of such ensembles below.) Then one can run the model on each column of $\tilde{\mathbf{X}}(0)$, with a disturbance from the corresponding column of $\tilde{\mathbf{U}}(0)$, and compute the ensemble of forecasts $\tilde{\mathbf{X}}(1,-)$. Computation of $\mathbf{A}(0)\mathbf{P}(0)\mathbf{A}(0)^T$ requires running the model $2N$ times, where computation of $\tilde{\mathbf{X}}(1,-)$ requires only L such model runs. Assuming that the true mean of $\mathbf{x}(1)$ is zero, estimate,

{ensemble1}
$$\tilde{\mathbf{P}}(1,-) = \frac{1}{L}\tilde{\mathbf{X}}(1,-)\,\tilde{\mathbf{X}}(1,-)^T = \frac{1}{L}\sum_{j=1}^{L}\tilde{\mathbf{x}}_j(1,-)\,\tilde{\mathbf{x}}_j(1,-)^T \tag{5.1}$$

where $\tilde{\mathbf{x}}_j(1,-)$ is column j of $\tilde{\mathbf{X}}(t,-)$, as an estimate of $\mathbf{P}(1,-)$. Note that if the mean is computed from the columns of $\tilde{\mathbf{X}}$, and subtracted from the estimate, the factor in front becomes $1/(L-1)$. With $\tilde{\mathbf{P}}(1,-)$ known, the filter average step (4.51) can be carried out, although if the probability densities of the model and data errors are different, the average may have little meaning. Because the average is a linear one, the conventional filter error covariance calculation, Eq. (4.53) is still appropriate, and one can continue in this fashion through the filter loop. In essence, this approach characterizes the so-called ensemble Kalman filter method. The main issue here concerns the reliability of the estimates for small ensemble sizes.[154] Because of its structure, if L is less than N, the maximum rank of $\tilde{\mathbf{P}}(1,-)$ is $k = L < N$, and the matrix will be singular. Singularity implies that some structures (those in the nullspace of $\tilde{\mathbf{P}}(1,-)$) are impossible in the initial conditions—a potentially troublesome outcome.

In principle, one can use the ensemble members to produce estimates of the complete probability densities of $\tilde{\mathbf{x}}, \tilde{\mathbf{u}}$, no matter how non-linear the model—leading to the use of maximum likelihood methods. These are computationally more demanding, however. Even small ensembles provide at least a qualitative indication of where maximum uncertainty is likely to lie, but their use should not stretch beyond their actual limited information content.

How does one generate ensemble members with zero mean and given spatial covariance, $\mathbf{P}(0)$? Let

$$\mathbf{P}(0) = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T,$$

and suppose $\alpha_i^{(p)}$ is white noise from a pseudo-random number generator such that, $\left\langle \alpha_i^{(p)} \right\rangle = 0$, $\left\langle \alpha_i^{(p)} a_j^{(p)} \right\rangle = \delta_{ij}$ where p is the ensemble member label. Form

$$\tilde{\mathbf{x}}_p(0) = \sum_{j=1}^{N} \sqrt{\lambda}_j \alpha_j^{(p)} \mathbf{v}_j.$$

Then it follows that,

$$\begin{aligned}
\left\langle \tilde{\mathbf{x}}^{(p)}(0) \right\rangle &= 0, \\
\left\langle (\tilde{\mathbf{x}}_p(0) - \tilde{\mathbf{x}}(0))(\tilde{\mathbf{x}}_p(0) - \tilde{\mathbf{x}}(0))^T \right\rangle &= \left\langle \left(\sum_{j=1}^{N} \sqrt{\lambda}_j \alpha_j^{(p)} \mathbf{v}_j \right) \left(\sum_{n=1}^{N} \sqrt{\lambda_n} \alpha_n^{(p)} \mathbf{v}_n \right)^T \right\rangle \\
&= \sum_{j=1}^{N} \lambda_j \left\langle \alpha_j^{(p)\ 2} \right\rangle \mathbf{v}_j \mathbf{v}_j^T = \mathbf{P}(0),
\end{aligned}$$

as required. It is readily confirmed too, that the ensemble members are uncorrelated with each other.

The members of the ensemble of initial conditions can have highly non-Gaussian probability densities. One would then select the $\alpha_j^{(p)}$ from populations with whatever is the appropriate probability density.[155] More generally, the initial condition disturbances may have specific structures related to the dynamics. Some of those structures may give rise to particularly rapidly growing disturbances, and which if excited can give an ensemble spread much larger than that obtained from purely random components. A very large effort in weather forecasting in particular, has gone into generating small ensembles that have a useful breadth.[156] Unknown model parameters extend far beyond initial conditions, including for example, mixing parameterizations, boundary conditions, source/sink components, etc. Ensembles can be generated by calculating solutions from random perturbations to any and all of these problem elements simultaneously.

Example.

Let the initial estimate for the mass-spring oscillator of the Example on P. 194, be $\tilde{\mathbf{x}}(0) = [1, 1]^T$, and *have error covariance*

$$\mathbf{P}(0) = \left\{ \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right\}.$$

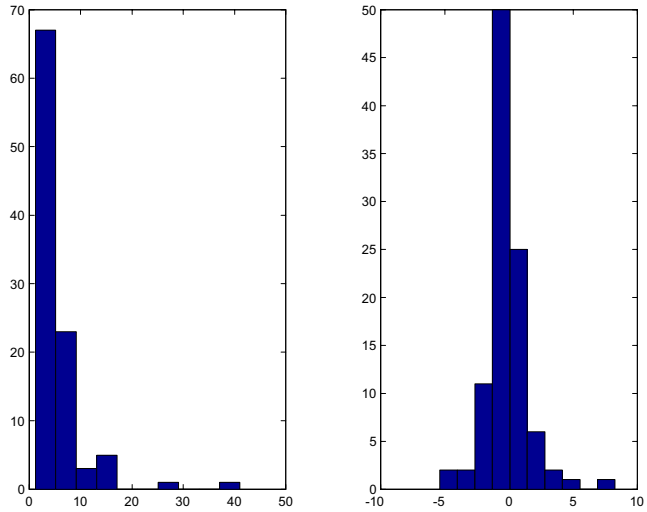

Figure 5.1: (Left) Empirical histogram of an ensemble of 100 values of $\tilde{x}_1(0)$ obtained from a log-normal distribution with parameters (1,1). A similar ensemble of values was generated for $x_2(0)$, with no correlation between the two. The random variables are regarded as different realizations of the noise occurring in the true value $\mathbf{x}(0) = [1,1]^T$.(Right) Historgram of 100 realizations of $x_1(t = 500)$ from the mass spring oscillator ($k = 0.1, \Delta t = 1, r = 0.01$). Note the tendency for the frequency function to tend towards Gaussian.

ognorm1.eps}

It is thought that the errors follow a log-normal distribution,

$$p(\xi) = \frac{1}{\sqrt{2\pi}\xi} \exp\left(-(\ln \xi - 1)^2 / 2\right). \tag{5.2}$$ {lognorm1}

An ensemble of initial conditions of the form $\tilde{\mathbf{x}}(0) = [1,1]^T + \boldsymbol{\xi}$ *was generated, producing a frequency function (histogram) of values for* $\tilde{x}_1(0)$ *as shown in Fig. 5.1. After 500 time steps (with* $k = 0.1, r = 0.01, \Delta t = 1$) $\tilde{x}_1(500)$ *tends toward Gaussian, and in this linear system, it would be reasonable to calculate the mean position as* $\overline{\tilde{x}_1(500)}$ *where the overbar indicates an average of the ensemble, and the error of the mean position is computed simply from its standard deviation in the ensemble. The latter can be used in the Kalman averaging step. But in a highly non-Gaussian distribution, as seen in the left panel of Fig. 5.1, the sample mean and variance may differ greatly from the true mean and variance owing to the presence of a few strong outliers. (One would be strongly advised to work with the logarithm of* $\tilde{x}_1(0)$; *see Aitchison and Brown, 1957; such a transformation would not be possible with* $\tilde{x}_1(500)$ *because it can be negative.)*

Calculation of probability densities, or the defining low moments of the probability densities for filter/smoother solutions can also be approached by solving equations for the evolution of the

densities or moments. Dynamical evolution equations can be used to write an explicit equation (the so-called Fokker-Planck, or Kolmogorov equation in the physics and mathematics literature, respectively) for the evolution of the state probability density.[157] Because the solution of such equations in high dimensions is forbidding, one can alternatively seek approximate equations for the evolution of the low moments.

This subject is a large and sophisticated one; a growing literature describes applications at various levels of approximation in oceanography and meteorology. But we must leave it to the references to deal with it further.[158]