Filtering

Andrew Stuart

21 October 2013

The Filtering Problem

Suppose that a system of interest evolves according to deterministic discrete time dynamics

$$v_{j+1} = \Psi(v_j) \tag{1}$$

where v_j in some Banach space $\mathbb{V}$ denotes the state of the system at the j^{th} time step, and the initial condition is a Gaussian random variable $v_0 \sim \mathcal{N}(m_0, C_0)$. Suppose also that we have noisy observations in some other Banach space $\mathbb{Y}$ (usually of much lower dimension than $\mathbb{V}$):

$$y_{j+1} = Hv_{j+1} + \eta_{j+1} \tag{2}$$

where $\eta_j \sim \mathcal{N}(0, \Gamma)$, and $H \colon \mathbb{V} \to \mathbb{Y}$ is some (linear) operator. Suppose that v_0 and all the η_j are pairwise independent, and let $Y_j := (y_\ell)_{\ell=0}^{j}$. The *filtering problem* is to determine the *filtering distribution* $\mathbb{P}(v_j | Y_j)$. This can be split into two steps, that of *prediction*

$$\mathbb{P}(v_j | Y_j) \mapsto \mathbb{P}(v_{j+1} | Y_j)$$

and *analysis/correction*

$$\mathbb{P}(v_{j+1} | Y_j) \mapsto \mathbb{P}(v_{j+1} | Y_{j+1}).$$

This is a Bayesian approach in which the prediction performs the role of the prior.

The Kálmán Filter

The Kálmán Filter (KF) is the case of linear dynamics, i.e. $\Psi(v) = Lv$ for some linear map $L \colon \mathbb{V} \to \mathbb{V}$. In this case, the filtering distribution is always Gaussian (since a linear image of a Gaussian measure is again Gaussian). If $\mathbb{P}(v_j | Y_j) = \mathcal{N}(m_j, C_j)$, then the prediction is that $\mathbb{P}(v_{j+1} | Y_j) = \mathcal{N}(\hat{m}_{j+1}, \hat{C}_{j+1})$ where

$$\begin{aligned} \hat{m}_{j+1} &= Lm_j, \\ \hat{C}_{j+1} &= LC_jL^\top. \end{aligned}$$

For the analysis step, if we introduce a cost function

$$J(m) := \frac{1}{2} \left\| \hat{C}_{j+1}^{-1/2} (m - \hat{m}_{j+1}) \right\|^2 + \frac{1}{2} \left\| \Gamma^{-1/2} (y_{j+1} - Hm) \right\|^2$$

then the posterior mean and covariance are given by completing the square:

$$\begin{aligned} m_{j+1} &= \operatorname*{arg\,min}_m J(m) \\ C_{j+1}^{-1} &= \hat{C}_{j+1}^{-1} + H^\top \Gamma^{-1} H. \end{aligned}$$

Gaussian Approximate Filters

When the forward dynamics Ψ are non-linear, the filtering distribution is at best *approximately* Gaussian. Inspired by the KF, we take

$$m_{j+1} = \operatorname*{arg\,min}_m J(m)$$
$$J(m) = \frac{1}{2}\left\|\hat{C}_{j+1}^{-1/2}(m - \Psi(m_j))\right\|^2 + \frac{1}{2}\left\|\Gamma^{-1/2}(y_{j+1} - Hm)\right\|^2.$$

Once the $\{\hat{C}_{j+1}\}_{j=0}^{\infty}$ are specified, this minimization determines a map $(m_j, y_{j+1}) \mapsto m_{j+1}$. We have the following equations for the evolution of the state v_j and the mean state estimate m_j:

$$C_{j+1}^{-1} m_{j+1} = \hat{C}_{j+1}^{-1}\Psi(m_j) + H^\top \Gamma^{-1} y_{j+1} \tag{3}$$
$$C_{j+1}^{-1} v_{j+1} = \hat{C}_{j+1}^{-1}\Psi(v_j) + H^\top \Gamma^{-1} H\Psi(v_j) \tag{4}$$
$$y_{j+1} = H\Psi(v_j) + \eta_{j+1}. \tag{5}$$

Equations (3–5) imply that

$$C_{j+1}^{-1}(m_{j+1} - v_{j+1}) = \hat{C}_{j+1}^{-1}(\Psi(m_{j+1}) - \Psi(v_j)) + H^\top \Gamma^{-1}\eta_{j+1}. \tag{6}$$

Equation (6) underwrites the long-time asymptotic behaviour of the filter, and in particular the recovery from a (large) initial error. If we let $\delta_j := m_j - v_j$ denote the error, then δ_j evolves according to

$$\delta_{j+1} = C_{j+1}\hat{C}_{j+1}^{-1}(\Psi(v_j + \delta_j) - \Psi(v_j)) + \xi_{j+1} \tag{7}$$
$$\xi_{j+1} = C_{j+1} H^\top \Gamma^{-1} \eta_{j+1}. \tag{8}$$

In the good cases, the operator on the right-hand side of (7) will be a contraction overall, and so δ_j will shrink to 0 in long time.

3DVar

This is an approximate filter. The "3D" comes from physical application in which minimization (the **var**iational approach) is performed over the spatial dimensions at fixed time, as opposed to the 4DVar filter in which minimization is performed over time as well. In 3DVar, $\hat{C}_j$ is a fixed matrix/operator $\hat{C}$, inspired by the fact that in the KF the covariance matrix C_j tends to a limit as $j \to \infty$. The use of a 'large' $\hat{C}$ is referred to as 'variance inflation', and is used to compensate for model error (a wrong Ψ).

ExKF

The Extended Kálmán Filter uses $\mathrm{D}\Psi$ as L, and $\hat{C}_{j+1} = \mathrm{D}\Psi(v_j) C_j \mathrm{D}\Psi(v_j)^\top$. This can be a prohibitively large matrix to calculate and store, e.g. in weather applications with $10^9 \times 10^9$ matrices.

EnKF

Ensemble Kálmán Filters use an ensemble $v^{(1)}, \dots, v^{(K)}$ of state estimates with

$$v_{j+1}^{(k)} = \operatorname*{arg\,min}_m J^{(k)}(m)$$
$$J^{(k)}(m) = \frac{1}{2}\left\|\hat{C}_{j+1}^{-1/2}(m - \Psi(v_j^{(k)}))\right\|^2 + \frac{1}{2}\left\|\Gamma^{-1/2}(y_{j+1}^{(k)} - Hm)\right\|^2.$$

The empirical covariance of $\{\Psi(v_j^{(k)})\}_{k=1}^K$ is used as $\hat{C}_{j+1}$. One implementation of the EnKF uses data perturbation, so that the k^{th} member of the ensemble uses the perturbed observation

$$y_{j+1}^{(k)} = y_{j+1} + \eta_j^{(k)}$$

where $\eta_j^{(k)} \sim \mathcal{N}(0, \Gamma)$ are independent and identically distributed for each $k = 1, \dots, K$.