Parallel Batch Kalman Filtering

Pete Bunch

July 6, 2012

1 Introduction

Sometimes we want to run a Kalman filter offline, for example when using Rao-Blackwellisation in an MCMC algorithm, such as the MCMC-DA system for target tracking. In this case, the sequential nature of the Kalman filter prevents it from being easily parallelisable. Here we try to fix that.

2 System Equations

Consider a standard discrete-time linear-Gaussian state-space model,

$$x_n = Ax_{n-1} + w_n \tag{1}$$
$$y_n = Cx_n + v_n \tag{2}$$

$$w_n \sim \mathcal{N}(.|0, Q) \tag{3}$$
$$v_n \sim \mathcal{N}(.|0, R). \tag{4}$$

For now, we assume that the starting state, x_0 is fixed and known. We'll want to relax this condition later.

We can stack up all N of the state and observation equations into matrix equations,

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} A \\ A^2 \\ A^3 \\ \vdots \\ A^N \end{bmatrix}}_{F} x_0 + \underbrace{\begin{bmatrix} I & 0 & 0 & \dots & 0 \\ A & I & 0 & \dots & 0 \\ A^2 & A & I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A^N & A^{N-1} & A^{N-2} & \dots & I \end{bmatrix}}_{G} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_N \end{bmatrix}}_{W} \tag{5}$$

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} C & & \\ & \ddots & \\ & & C \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}}_{X} + \underbrace{\begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}}_{V}. \tag{6}$$

$$W \sim \mathcal{N}\left(.|0, \underbrace{\begin{bmatrix} Q & & \\ & \ddots & \\ & & Q \end{bmatrix}}_{S}\right) \tag{7}$$

$$C \sim \mathcal{N}\left(.|0, \underbrace{\begin{bmatrix} R & & \\ & \ddots & \\ & & R \end{bmatrix}}_{T}\right). \tag{8}$$

This gives us the following distributions,

$$X \sim \mathcal{N}(.|Fx_0, GSG^T) \tag{9}$$

$$Y \sim \mathcal{N}(.|HX, T). \tag{10}$$

Now its time for Bayes,

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \tag{11}$$

$$= \mathcal{N}(X|\mu, \Sigma). \tag{12}$$

For completeness, here's the derivation of μ and Σ in full. Its pretty standard completing the square, throughout which we though away constant terms knowing that they'll come out in the wash when we normalise.

$$\begin{aligned}
P(X|Y) &\propto P(Y|X)P(X) \\
&= \mathcal{N}(X|Fx_0, GSG^T)\mathcal{N}(Y|HX, T) \\
&\propto \exp\left\{-\frac{1}{2}\left[(X - Fx_0)^T(GSG^T)^{-1}(X - Fx_0) + (Y - HX)^T T^{-1}(Y - HX)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[X^T(GSG^T)^{-1}X - 2x_0^T F^T(GSG^T)^{-1}X + X^T H^T T^{-1}HX - 2Y^T T^{-1}HX\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[X^T\left((GSG^T)^{-1} + H^T T^{-1}H\right)X - 2\left(x_0^T F^T(GSG^T)^{-1} + Y^T T^{-1}H\right)X\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[X^T\Sigma^{-1}X - 2\mu^T\Sigma^{-1}X\right]\right\}.
\end{aligned}$$

Comparing terms gives us,

$$\Sigma = \left[(GSG^T)^{-1} + H^T T^{-1}H\right]^{-1} \tag{13}$$

$$\mu = \Sigma\left[(GSG^T)^{-1}Fx_0 + H^T T^{-1}Y\right]. \tag{14}$$

Good. μ is the vector of posterior means for each state. Σ is the complete covariance matrix for all states over time. To replicate a Kalman smoother, we only need the blocks on the diagonal of this matrix. The other blocks are the covariances between states at different times, and are less interesting.

Σ^{-1} is highly structured, and in fact we can right is out explicitly. G is square and clearly full rank (because A has to be full rank for a valid HMM), and it turns out that its inverse has the following wizard form,

$$G^{-1} = \begin{bmatrix} I & 0 & 0 & \dots & 0 & 0 \\ -A & I & 0 & \dots & 0 & 0 \\ 0 & -A & I & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & I & 0 \\ 0 & 0 & 0 & \dots & -A & I \end{bmatrix}. \tag{15}$$

Thus we can write,

$$\begin{aligned} \Sigma^{-1} &= G^{-T} S^{-1} G^{-1} + H^T T^{-1} H \\ &= \begin{bmatrix} Q^{-1} + A^T Q^{-1} A & -A^T Q^{-1} & 0 & \dots & 0 & 0 \\ -Q^{-1} A & Q^{-1} + A^T Q^{-1} A & -A^T Q^{-1} & \dots & 0 & 0 \\ 0 & -Q^{-1} A & Q^{-1} + A^T Q^{-1} A & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & Q^{-1} + A^T Q^{-1} A & -A^T Q^{-1} \\ 0 & 0 & 0 & \dots & -Q^{-1} A & Q^{-1} \end{bmatrix} \\ &+ \begin{bmatrix} C^T R^{-1} C & 0 & 0 & \dots & 0 & 0 \\ 0 & C^T R^{-1} C & 0 & \dots & 0 & 0 \\ 0 & 0 & C^T R^{-1} C & \vdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & C^T R^{-1} C & 0 \\ 0 & 0 & 0 & \dots & 0 & C^T R^{-1} C \end{bmatrix} \end{aligned} \tag{16}$$

The tasks that remain then are as follows:

- Calculate the diagonal blocks of Σ.
- Evaluate the vector $\xi = \left[(GSG^T)^{-1} F x_0 + H^T T^{-1} Y \right]$.
- Solve the system of equations $\Sigma^{-1} \mu = \xi$ to give us μ.