Benjamin Born
Johannes Pfeifer

Spring Term 2018
University of Cologne

Kalman Filter and Smoother[1]

1 The Setup

Consider a linear Gaussian state space system of the form

$$x_{t+1} = Fx_t + w_{t+1},\ w_{t+1} \overset{iid}{\sim} N(0, Q) \tag{1}$$

$$y_t = Hx_t + \nu_t,\ \nu_t \overset{iid}{\sim} N(0, R)\ , \tag{2}$$

where x_t is an $n_x \times 1$ vector of states and y_t is an $n_y \times 1$ vector of observables. w_t is a $p \times 1$ vector of structural errors and ν_t a vector of measurement error. We assume that w_t and ν_t are orthogonal:

$$E_t(w_{t+1}\nu_s) = 0 \tag{3}$$

for all $t+1$ and all $s \geq 0$.

We assume the initial value x_0 is distributed as

$$x_0 \sim N\left(\hat{x}_{0|-1}, \Sigma_{0|-1}\right)\ , \tag{4}$$

where the subscript -1 denotes the information we have at the beginning of times (often this is simply denoted as x_0 and Σ_0). In the following, we will denote a particular observation with a time subscript, e.g. y_t and the complete history of observations up to a certain point in time with a time superscript, i.e. $y^t = \{y_t, \ldots, y_0\}$.

This described setup implies for $y_0 = Hx_0 + \nu_0$ that

$$\hat{y}_{0|-1} \equiv E(y_0|y^{-1}) = E(Hx_0 + \nu_0|y^{-1}) = H\hat{x}_{0|-1} \tag{5}$$

$$\begin{aligned} E\left[\left(y_0 - \hat{y}_{0|-1}\right)\left(y_0 - \hat{y}_{0|-1}\right)'|y^{-1}\right] &= E\left[\left(Hx_0 + \nu_0 - H\hat{x}_{0|-1}\right)\left(Hx_0 + \nu_0 - H\hat{x}_{0|-1}\right)'|y^{-1}\right] \\ &= E\left[H(x_0 - \hat{x}_{0|-1})(x_0 - \hat{x}_{0|-1})'H' + \nu_0\nu_0'|y^{-1}\right] \\ &= H\Sigma_{0|-1}H' + R\,, \end{aligned} \tag{6}$$

where the cross-terms in ν have been dropped due to them being uncorrelated with everything else. Hence:

$$y_0 \sim N\left(H\hat{x}_{0|-1}, H\Sigma_{0|-1}H' + R\right) \tag{7}$$

Note that (6) denotes the Mean Squared Error of y_0 given information up to $t = -1$.

1.1 The Goal

The economist only observes the observables $y^t = \{y_t, \ldots, y_0\}$, but wants to infer the $x_t, \ldots, x_0$. We assume that the economist knows the structure (1)-(2) and the first and second moments implied by this structure. The aim is to find *recursive* formulas for the state forecast

$$\hat{x}_{t|t-1} \equiv E[x_t|y^{t-1}] \tag{8}$$

and the Mean Squared Error/covariance matrices of the forecast error

$$\Sigma_{t|t-1} \equiv E\left[\left(x_t - \hat{x}_{t|t-1}\right)\left(x_t - \hat{x}_{t|t-1}\right)'|y^{t-1}\right] \tag{9}$$

Recursiveness allows for online tracking: at time t when a new observation becomes available the old forecast and the new observation can be combined to build the new forecast.

[1] Jointly developed with Benjamin Born, Frankfurt School of Finance and Management.

1.2 The Idea

Use a regression of the unknown difference between the true state x_t and our forecast of it made yesterday $\hat{x}_{t|t-1}$, i.e. $x_t - \hat{x}_{t|t-1}$, on the new information contained in the observation y_t compared to what we had predicted it to be $y_t - \hat{y}_{t|t-1}$:[2]

$$x_t - \hat{x}_{t|t-1} = L_t \left(y_t - \hat{y}_{t|t-1}\right) + \eta_t \tag{10}$$

where L_t is the regression coefficient. Our forecast $\hat{y}_{t|t-1}$ of y_t is given from (2) by

$$\hat{y}_{t|t-1} = H\hat{x}_{t|t-1} \tag{11}$$

Thus, the forecast error can be defined as

$$a_t \equiv y_t - H\hat{x}_{t|t-1} \tag{12}$$

and the regression equation be written as

$$x_t - \hat{x}_{t|t-1} = L_t \left(y_t - H\hat{x}_{t|t-1}\right) + \eta_t\ , \tag{13}$$

where η_t is orthogonal to the variables contained in the information set at time t by being a regression residual. Of course, we don't know the left-hand side to actually run the regression and compute L_t. But if we somehow know L_t, we could form a forecast of our forecast error for the state.

2 The Kalman filter recursion

Let's try to compute L_t at time t.

Remember the formula for a population regression

$$\beta = E\left(YX'\right)\left[E\left(XX'\right)\right]^{-1} \tag{14}$$

Plugging in

$$\begin{aligned}
L_t =& E\left(\left(x_t - \hat{x}_{t|t-1}\right)\left(y_t - H\hat{x}_{t|t-1}\right)'\right)\left[E\left(\left(y_t - H\hat{x}_{t|t-1}\right)\left(y_t - H\hat{x}_{t|t-1}\right)'\right)\right]^{-1} \\
\overset{(6)}{=}& E\left(\left(x_t - \hat{x}_{t|t-1}\right)\left(Hx_t - H\hat{x}_{t|t-1}\right)'\right)\left(H\Sigma_{t|t-1}H' + R\right)^{-1} \\
=& E\left(\left(x_t - \hat{x}_{t|t-1}\right)\left(x_t - \hat{x}_{t|t-1}\right)'\right)H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1} \\
=& \Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}\ ,
\end{aligned} \tag{15}$$

where the second equality follows from the fact that ν_t is orthogonal to all other terms and can be dropped.

From equation (1) we know that given our best forecast for the state at time 0, our best forecast for tomorrow's state is given by

$$\hat{x}_{1|0} = E\left[x_1|y^0\right] = E\left[Fx_0 + w_1|y^0\right] = FE\left[x_0|y^0\right] + E\left[w_1|y^0\right] = F\hat{x}_{0|0} \tag{16}$$

At the same time, equation (1) can be rewritten as

$$\begin{aligned}
x_1 &= Fx_0 + w_1 \\
&= F\hat{x}_{0|-1} + F\left(x_0 - \hat{x}_{0|-1}\right) + w_1 && (17)\\
&\overset{(13)}{=} F\hat{x}_{0|-1} + F\left(L_0\left(y_0 - H\hat{x}_{0|-1}\right) + \eta_0\right) + w_1 && (18)
\end{aligned}$$

[2] This derivation is based on Ljungqvist and Sargent (2012, Chapter 2.5), for a different derivation making use of updating of linear projections, see Hamilton (1994, Chapter 13) or Durbin and Koopman (2012).

Now try to forecast tomorrow's state given information until time 0:[3]

$$\hat{x}_{1|0} = E\left[x_1|y_0\right] = F\hat{x}_{0|-1} + \underbrace{FL_0}_{K_0}\left(y_0 - H\hat{x}_{0|-1}\right) = F\hat{x}_{0|-1} + K_0\left(y_0 - H\hat{x}_{0|-1}\right) \tag{19}$$

This allows to forecast tomorrow's state just based on yesterday's forecast and today's observation. The matrix

$$K_0 = FL_0 = F\Sigma_{0|-1}H'\left(H\Sigma_{0|-1}H' + R\right)^{-1} \tag{20}$$

is called the *Kalman Gain*. It determines by how much your state estimate is updated based on your previous forecast error.

But our goal is not just to forecast x_1, but also x_2 and later values. For this, we need the population regression (15), which makes use of information on the full distribution of x_t (which is conditionally normal). We already computed the expected value, but still need the covariance/mean squared error matrix $\Sigma_{t|t-1}$, i.e. $\Sigma_{1|0}$. To compute this, note that the forecast error is given by

$$\begin{aligned} x_1 - \hat{x}_{1|0} &\overset{(1)}{=} Fx_0 + w_1 - \hat{x}_{1|0} \overset{(19)}{=} Fx_0 + w_1 - \left(F\hat{x}_{0|-1} + K_0\left(y_0 - H\hat{x}_{0|-1}\right)\right) \\ &= F\left(x_0 - \hat{x}_{0|-1}\right) + w_1 - K_0\left(y_0 - H\hat{x}_{0|-1}\right) \end{aligned} \tag{21}$$

Thus

$$\begin{aligned} \Sigma_{1|0} =& E\left[\left(x_1 - \hat{x}_{1|0}\right)\left(x_1 - \hat{x}_{1|0}\right)'|y^0\right] \\ \overset{(21)}{=}& E\left[\left(F\left(x_0 - \hat{x}_{0|-1}\right) + w_1 - K_0\left(y_0 - H\hat{x}_{0|-1}\right)\right)\left(F\left(x_0 - \hat{x}_{0|-1}\right) + w_1 - K_0\left(y_0 - H\hat{x}_{0|-1}\right)\right)'|y^0\right] \\ =& E[F\left(x_0 - \hat{x}_{0|-1}\right)\left(x_0 - \hat{x}_{0|-1}\right)'F' + F\left(x_0 - \hat{x}_{0|-1}\right)w_1' - F\left(x_0 - \hat{x}_{0|-1}\right)\left(y_0 - H\hat{x}_{0|-1}\right)'K_0' \\ &+ w_1\left(x_0 - \hat{x}_{0|-1}\right)'F' + w_1w_1' - w_1\left(y_0 - H\hat{x}_{0|-1}\right)'K_0' \\ &- K_0\left(y_0 - H\hat{x}_{0|-1}\right)\left(x_0 - \hat{x}_{0|-1}\right)'F' - K_0\left(y_0 - H\hat{x}_{0|-1}\right)w_1' \\ &+ K_0\left(y_0 - H\hat{x}_{0|-1}\right)\left(y_0 - H\hat{x}_{0|-1}\right)'K_0'|y^0] \end{aligned}$$

[3]The same formula can be derived by noting that equation (32) implies

$$x_{t|t} = \hat{x}_{t|t-1} + \Sigma_t H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}\left(y_t - H\hat{x}_{t|t-1}\right)$$

Combining this with (16) yields the equation in the text.

Now noting that w_1 is orthogonal to everything else on the right hand side;

$$\begin{aligned}
\Sigma_{1|0} \overset{(2)}{=} & \underbrace{E_t[F\left(x_0 - \hat{x}_{0|-1}\right)\left(x_0 - \hat{x}_{0|-1}\right)' F']}_{F\Sigma_{0|-1}F'} \\
& - \underbrace{E\left[F\left(x_0 - \hat{x}_{0|-1}\right)\left(Hx_0 + \nu_0 - H\hat{x}_{0|-1}\right)' K_0'\right]}_{F\Sigma_{0|-1}H'K_0'} \\
& + \underbrace{E\left[w_1 w_1'\right]}_{Q} \\
& - \underbrace{E\left[K_0\left(Hx_0 + \nu_0 - H\hat{x}_{0|-1}\right)\left(x_0 - \hat{x}_{0|-1}\right)' F'\right]}_{K_0 H\Sigma_{0|-1}F'} \\
& + \underbrace{E\left[K_0\left(y_0 - H\hat{x}_{0|-1}\right)\left(y_0 - H\hat{x}_{0|-1}\right)' K_0'\right]}_{\overset{(6),(11)}{=} K_0\left(H\Sigma_{0|-1}H' + R\right)K_0'} \\
= & F\Sigma_{0|-1}F' - F\Sigma_{0|-1}H'K_0' + Q - K_0 H\Sigma_{0|-1}F' + K_0\left(H\Sigma_{0|-1}H' + R\right)K_0'
\end{aligned}$$

Factoring out results in:

$$\Sigma_{1|0} = \left(F - K_0 H\right)\Sigma_{0|-1}\left(F - K_0 H\right)' + Q + K_0 R K_0' \tag{22}$$

Note that the update of the covariance matrix/MSE matrix, $\Sigma_{1|0}$, only requires the knowledge of the previous period's covariance matrix $\Sigma_{0|-1}$ and the given matrices of the state space system.

This completes the recursion as we now know that

$$x_1 \sim N\left(\hat{x}_{1|0}, \Sigma_{1|0}\right) \tag{23}$$

where $\hat{x}_{1|0}$ is given by (19) and $\Sigma_{1|0}$ by (22)

3 Summary

At time t, given $\hat{x}_{t|t-1}, \Sigma_{t|t-1}$ and observing y_t

1. Compute the forecast error in the observations using

$$a_t = y_t - H\hat{x}_{t|t-1} \tag{12}$$

2. Compute the *Kalman Gain* K_t using

$$K_t = F\Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1} \tag{20'}$$

3. Compute the state forecast for next period given today's information, $\hat{x}_{t+1|t}$ using (19)

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_t\left(y_t - H\hat{x}_{t|t-1}\right) = F\hat{x}_{t|t-1} + K_t a_t \tag{24}$$

4. Update the covariance matrix using (22)

$$\Sigma_{t+1|t} = \left(F - K_t H\right)\Sigma_{t|t-1}\left(F - K_t H\right)' + Q + K_t R K_t' \tag{25}$$

4 Initialization

How does one initialize the Kalman filter at time $t = 0$ before any observations are available? One typically starts with the unconditional mean $E(x)$ and variance Σ. Given covariance stationarity, i.e. all eigenvalues of F inside the unit circle, the unconditional mean is given from equation (1) by

$$x = Ex_{t+1} = E(Fx_t + w_{t+1}) = Fx \Rightarrow (I - F)x = 0$$

Given that the eigenvalues of F are all inside of the unit circle $I - F$ is non-singular. Thus $x = 0$. For the covariance matrix we have

$$\begin{aligned} \Sigma &= E\left[(Fx_t + w_t)(Fx_t + w_t)'\right] \\ &= E\left[Fx_t x_t{}' F' + w_t w_t{}'\right] \\ &= F\Sigma F' + Q \end{aligned} \tag{26}$$

This is a so-called *Lyapunov-equation*.

Proposition 10.4 of Hamilton (1994, p. 265) states that if A, B, C are conformable matrices, then

$$vec(ABC) = (C' \otimes A)vec(B)$$

Using this on (26) yields

$$vec(\Sigma) = (F \otimes F)vec(\Sigma) + vec(Q)$$

This has solution

$$vec(\Sigma) = \left(I_{n_x^2} - (F \otimes F)\right)^{-1} vec(Q) \tag{27}$$

While this is a neat formula that can be easily implemented in Matlab, it involves the inversion of an n_x^2 matrix, which can be really slow. Thus, it is often advisable to use a different technique, the so-called *doubling algorithm*.

5 The Importance of the Kalman Filter for Estimation

Denote with $y^t = \{y_t, \ldots, y_0\}$ the complete history of observables up to time t. The likelihood function $f(y_T, \ldots, y_0)$ can be factored as

$$f(y_T, \ldots, y_0) = f\left(y_T | y^{T-1}\right) \times f\left(y_{T-1} | y^{T-2}\right) \times \ldots \times f\left(y_1 | y^0\right) \times f(y_0) \ ,$$

where $y_0 = y_{0|-1}$ denotes the information at the beginning of times. We saw in the derivation of the Kalman filter from (7)

$$y_t \sim N\left(H\hat{x}_{t|t-1}, H\Sigma_{t|t-1}H' + R\right) \tag{7}$$

where $\hat{x}_{t|t-1} = E[x_t | y^{t-1}]$ is a function of y^{t-1} and

$$\Omega_t \equiv H\Sigma_{t|t-1}H' + R \tag{28}$$

only depends on the population moments and not on the data. Thus, $H\hat{x}_{t|t-1}$ and Ω_t are sufficient statistics to compute the likelihood as they describe the first two moments of the conditional normal distribution, which the observables follow.

Hence, the probability density of observing y_t given y^{t-1} is given by

$$f\left(y_t | y^{t-1}\right) = \frac{1}{\sqrt{(2\pi)^{n_y} \det(\Omega_t)}} e^{-\frac{1}{2}\left(y_t - H\hat{x}_{t|t-1}\right)' \Omega_t^{-1} \left(y_t - H\hat{x}_{t|t-1}\right)} \ , \tag{29}$$

where *det* denotes the matrix determinant. Taking logs leads to the log-likelihood function:

$$\log\left(f\left(y_t|y^{t-1}\right)\right) = -\frac{n_y}{2}\log(2\pi) - \frac{1}{2}\log(\det(\Omega_t)) - \frac{1}{2}a_t{}'\Omega_t^{-1}a_t \tag{30}$$

Hence, the output of the Kalman filter, as a byproduct, delivers everything needed to compute the likelihood of the data.

6 Filtered estimates

Sometimes we are not interested in getting a forecast $x_{t|t-1}$, but rather obtaining "filtered estimates" $x_{t|t}$, i.e. a best estimate of the unobserved state at time t given information up to the current time period t. From equation (13) we know that

$$x_t = \hat{x}_{t|t-1} + L_t\left(y_t - H\hat{x}_{t|t-1}\right) + \eta \tag{31}$$

and can use this to compute the forecast of our state at time t given the observation of y_t:

$$x_{t|t} \equiv E_t\left(x_t|\, y^t\right) = \hat{x}_{t|t-1} + L_t\left(y_t - H\hat{x}_{t|t-1}\right)\ , \tag{32}$$

where the equality follows from i) the regression residual being unpredictable given the information set and ii) the law of iterated expectations: our best forecast today of yesterday's forecast is yesterday's forecast.

Given this, we can compute the MSE of the filtered estimates

$$\begin{aligned}
\Sigma_{t|t} =& E\left[\left(x_t - \hat{x}_{t|t}\right)\left(x_t - \hat{x}_{t|t}\right)'\right] \\
\overset{(32)}{=}& E\left[\left(x_t - \left(\hat{x}_{t|t-1} + L_t\left(y_t - H\hat{x}_{t|t-1}\right)\right)\right)\left(x_t - \hat{x}_{t|t-1} + L_t\left(y_t - H\hat{x}_{t|t-1}\right)\right)'\middle| y^t\right] \\
=& E\left[\left(x_t - \left(\hat{x}_{t|t-1} + L_t\left(Hx_t + \nu_t - H\hat{x}_{t|t-1}\right)\right)\right)\left(x_t - \hat{x}_{t|t-1} + L_t\left(Hx_t + \nu_t - H\hat{x}_{t|t-1}\right)\right)'\middle| y^t\right] \\
=& E\left[\left(x_t - \hat{x}_{t|t-1} - LH\left(x_t - \hat{x}_{t|t-1}\right) + L\nu_t\right)\left(x_t - \hat{x}_{t|t-1} - LH\left(x_t - \hat{x}_{t|t-1}\right) + L\nu_t\right)'\middle| y^t\right] \\
=& E[\left(x_t - \hat{x}_{t|t-1}\right)\left(x_t - \hat{x}_{t|t-1}\right)' \\
& - \left(x_t - \hat{x}_{t|t-1}\right)\left(x_t - \hat{x}_{t|t-1}\right)' H'L' - LH\left(x_t - \hat{x}_{t|t-1}\right)\left(x_t - \hat{x}_{t|t-1}\right)' \\
& + LH\left(x_t - \hat{x}_{t|t-1}\right)\left(x_t - \hat{x}_{t|t-1}\right)' H'L' \\
& + LH\left(x_t - \hat{x}_{t|t-1}\right)\nu_t{}'L' \\
& + L\nu_t\left(x_t - \hat{x}_{t|t-1}\right)' H'L' \\
& + L\nu_t\nu_t{}'L'|y^t] \\
=& \Sigma_{t|t-1} - \Sigma_{t|t-1}H'L' - LH\Sigma_{t|t-1} + LH\Sigma_{t|t-1}H'L' + LRL' \\
=& \Sigma_{t|t-1} + L\left(H\Sigma_{t|t-1}H' + R\right)L' - \Sigma_{t|t-1}H'L' - LH\Sigma_{t|t-1} \\
\overset{(15)}{=}& \Sigma_{t|t-1} + \Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}\left(H\Sigma_{t|t-1}H' + R\right)\left(\Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}\right)' \\
& - \Sigma_{t|t-1}H'\left(\Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}\right)' - \Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}H\Sigma_{t|t-1} \\
=& \Sigma_{t|t-1} + \Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}H\Sigma_{t|t-1} \\
& - 2\Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}H\Sigma_{t|t-1} \\
=& \Sigma_{t|t-1} - \Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}H\Sigma_{t|t-1}
\end{aligned} \tag{33}$$

where we have used the fact that the observational error ν_t is uncorrelated with the forecast error of the state and that both R and the MSE matrix $\Sigma_{t|t-1}$ are symmetric matrices.

7 Kalman Smoothing

The Kalman filter naturally delivers forecasts of the state tomorrow given today's observation $x_{t|t-1}$ and current estimates of the state $x_{t|t}$. However, if the state has a structural interpretation, e.g. a news shock about future fiscal policy that cannot be observed by the econometrician, we are often interested in our best estimate of the historical values of the states given all observations, i.e.

$$x_{t|T} \equiv E(x_t|y^T) \, \forall \, t \in \{0, \ldots, T\} \tag{34}$$

Moreover, we are interested in the distribution of those states and would like to know the MSE

$$\Sigma_{t|T} = E\left[\left(x_t - x_{t|T}\right)\left(x_t - x_{t|T}\right)'|y^T\right] \tag{35}$$

We start again with our idea to use a population regression for updating. Assume that at time T, we are told the true value of tomorrow's state x_{t+1}. Given this observation, we want to use our forecast error for time $t+1$, $x_{t+1} - x_{t+1|t}$ to predict the prediction error we made at time t, i.e. $x_t - x_{t|t}$:

$$x_t - \hat{x}_{t|t} = J_t\left(x_{t+1} - \hat{x}_{t+1|t}\right) + \eta_t \, , \tag{36}$$

where J_t is the regression coefficient. As before, we cannot actually run the regression, but we can compute the population regression coefficient J_t

$$\begin{aligned} J_t =& E\left(\left(x_t - \hat{x}_{t|t}\right)\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right)\left[E\left(\left(x_{t+1} - \hat{x}_{t+1|t}\right)\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right)\right]^{-1} \\ =& E\left(\left(x_t - \hat{x}_{t|t}\right)\left(Fx_t + w_t - F\hat{x}_{t|t}\right)'\right)\left[E\left(\left(x_{t+1} - \hat{x}_{t+1|t}\right)\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right)\right]^{-1} \\ =& E\left(\left(x_t - \hat{x}_{t|t}\right)\left(x_t - \hat{x}_{t|t}\right)'F'\right)\left[E\left(\left(x_{t+1} - \hat{x}_{t+1|t}\right)\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right)\right]^{-1} \\ =& \Sigma_{t|t}F'\Sigma_{t+1|t}^{-1} \end{aligned} \tag{37}$$

All required objects to compute J_t can be computed while performing the Kalman filter recursion:

- $\Sigma_{t+1|t}$ follows from equation (25),
- $x_{t|t}$ from (32),
- and $\Sigma_{t|t}$ from (33).

We now proceed as before. Rewrite the actual state today as

$$\begin{aligned} x_t =& \hat{x}_{t|t} + \left(x_t - \hat{x}_{t|t}\right) \\ \overset{(36)}{=}& \hat{x}_{t|t} + J_t\left(x_{t+1} - \hat{x}_{t+1|t}\right) + \eta_t \end{aligned} \tag{38}$$

Now take expectations given our assumed information set at time T, which includes x_{t+1}:

$$E\left(x_t|x_{t+1}, y^T\right) = E\left(\hat{x}_{t|t} + J_t\left(x_{t+1} - \hat{x}_{t+1|t}\right) + \eta_t|x_{t+1}, y^T\right) = \hat{x}_{t|t} + J_t\left(x_{t+1} - \hat{x}_{t+1|t}\right) \, ,$$

where the second equality follows from the law of iterated expectations and the fact that η_t is a regression residual that is orthogal to everything contained in y_{t+j} beyond x_{t+1}.[4]

Unfortunately, the previous equation was based on the fiction that we actually know x_{t+1}. What happens if we don't know it, but only y^T. The answer comes from the law of iterated expectations:

$$
\begin{aligned}
\hat{x}_{t|T} \equiv & E\left(x_t|y^T\right) \\
= & E\left[E\left(x_t|x_{t+1}, y^T\right)\Big| y^T\right] \\
= & E\left(\hat{x}_{t|t} + J_t\left(x_{t+1} - \hat{x}_{t+1|t}\right) + \eta_t|y^T\right) \\
= & \hat{x}_{t|t} + J_t\left(\hat{x}_{t+1|T} - \hat{x}_{t+1|t}\right)
\end{aligned} \tag{39}
$$

Thus, to compute $x_{t|T}$, we need $J_t, \hat{x}_{t+1|t}$, and $\hat{x}_{t|t}$. All these objects are or can be computed during the Kalman filter recursion. One only needs to store them. However, (39) is a backwards recursion, making use of the the smoothed estimate of next period's state $x_{t+1|T}$, which is unknown yet. How can one proceed and solve this dependence on future values?

The answer lies in the last period of the Kalman filter recursion, where the filtered estimate $x_{T|T}$ coincides with the smoothed estimate. Thus, working backwards from period T, the recursion can be solved.

Finally, we are also interested in the distribution of the smoothed estimates, i.e. we would like to know the MSE $\Sigma_{t|T}$. Rewrite (39) as

$$
x_t - \hat{x}_{t|T} + J_t\hat{x}_{t+1|T} = x_t - \hat{x}_{t|t} + J_t\hat{x}_{t+1|t} \tag{40}
$$

Now, multiply each side of (40) with its transposed:

$$
\left[(x_t - \hat{x}_{t|T}) + J_t\hat{x}_{t+1|T}\right]\left[(x_t - \hat{x}_{t|T}) + J_t\hat{x}_{t+1|T}\right]' = \left[(x_t - \hat{x}_{t|t}) + J_t\hat{x}_{t+1|t}\right]\left[(x_t - \hat{x}_{t|t}) + J_t\hat{x}_{t+1|t}\right]' \tag{41}
$$

Taking expectations and recognizing that the forecast for time $t+1$ given information up to T is orthogonal to the forecast error at time t, the left-hand side of (41) evaluates to

$$
\begin{aligned}
& E\left[(x_t - \hat{x}_{t|T}) + J_t\hat{x}_{t+1|T}\right]\left[(x_t - \hat{x}_{t|T}) + J_t\hat{x}_{t+1|T}\right]' \\
& = \Sigma_{t|T} + J_t E\left(\hat{x}_{t+1|T}(\hat{x}_{t+1|T})'\right) J_t'
\end{aligned} \tag{42}
$$

Similarly, as the forecast for time $t+1$ made at time t is orthogonal to the forecast error made at time t, the right-hand side of (41) evaluates to:

$$
\begin{aligned}
& E\left[(x_t - \hat{x}_{t|t}) + J_t\hat{x}_{t+1|t}\right]\left[(x_t - \hat{x}_{t|t}) + J_t\hat{x}_{t+1|t}\right]' \\
& = \Sigma_{t|t} + J_t E\left(\hat{x}_{t+1|t}(\hat{x}_{t+1|t})'\right) J_t'
\end{aligned}
$$

[4] To be precise, we have that

$$
y_{t+j} = Hx_{t+j} = H\left(Fx_{t+j-1} + v_{t+j}\right) = H\left(F^{j-1}x_{t+1} + \sum_{i=1}^{j-1} F^{j-1-i}v_{t+1+i}\right) + w_{t+j}
$$

Thus, apart from x_{t+j}, the observations only contain information about the future error terms $\nu_{t+j}, w_{t+j}, j > 0$, which under our assumptions are of no use in predicting η_t.

Equation (40) therefore is equal to

$$\Sigma_{t|T} = \Sigma_{t|t} + J_t \left(-E\left(\hat{x}_{t+1|T}(\hat{x}_{t+1|T})'\right) + E\left(\hat{x}_{t+1|t}(\hat{x}_{t+1|t})'\right)\right) J'_t \tag{43}$$
$$= \Sigma_{t|t} + J_t \left(E\left(x_{t+1}(x_{t+1})'\right) - E\left(\hat{x}_{t+1|T}(\hat{x}_{t+1|T})'\right) + E\left(\hat{x}_{t+1|t}(\hat{x}_{t+1|t})'\right) - E\left(x_{t+1}(x_{t+1})'\right)\right) J'_t$$

where the second line expands by $0 = E\left(x_{t+1}(x_{t+1})'\right) - E\left(x_{t+1}(x_{t+1})'\right)$.

We can now use the fact that that forecast errors are uncorrelated with the forecasts to twice add a 0 to the following terms and rewrite them:

$$\begin{aligned}
E\left(\hat{x}_{t+1|T}(\hat{x}_{t+1|T})'\right) &= E\left(\hat{x}_{t+1|T}(\hat{x}_{t+1|T})'\right) + E\left[\left(x_{t+1} - \hat{x}_{t+1|T}\right)(\hat{x}_{t+1|T})'\right] + E\left[\hat{x}_{t+1|T}\left(x_{t+1} - \hat{x}_{t+1|T}\right)'\right] \\
&= E\left[\left(x_{t+1} - \hat{x}_{t+1|T} + \hat{x}_{t+1|T}\right)(\hat{x}_{t+1|T})'\right] + E\left[\hat{x}_{t+1|T}\left(x_{t+1} - \hat{x}_{t+1|T}\right)'\right] \\
&= E\left[x_{t+1}(\hat{x}_{t+1|T})'\right] + E\left[\hat{x}_{t+1|T}\left(x_{t+1} - \hat{x}_{t+1|T}\right)'\right] \qquad (44) \\
E\left(\hat{x}_{t+1|t}(\hat{x}_{t+1|t})'\right) &= E\left(\hat{x}_{t+1|t}(\hat{x}_{t+1|t})'\right) + E\left[\left(x_{t+1} - \hat{x}_{t+1|t}\right)(\hat{x}_{t+1|t})'\right] + E\left[\hat{x}_{t+1|t}\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right] \\
&= E\left[\left(x_{t+1} - \hat{x}_{t+1|t} + \hat{x}_{t+1|T}\right)(\hat{x}_{t+1|t})'\right] + E\left[\hat{x}_{t+1|t}\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right] \\
&= E\left[x_{t+1}(\hat{x}_{t+1|t})'\right] + E\left[\hat{x}_{t+1|t}\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right] \qquad (45)
\end{aligned}$$

Therefore:

$$\begin{aligned}
\Sigma_{t|T} &= \Sigma_{t|t} + J_t \left(E\left(x_{t+1}(x_{t+1})'\right) - E\left(\hat{x}_{t+1|T}(\hat{x}_{t+1|T})'\right) + E\left(\hat{x}_{t+1|t}(\hat{x}_{t+1|t})'\right) - E\left(x_{t+1}(x_{t+1})'\right)\right) J'_t \\
&= \Sigma_{t|t} + J_t \begin{pmatrix} E\left(x_{t+1}(x_{t+1})'\right) - \left(E\left[x_{t+1}(\hat{x}_{t+1|T})'\right] + E\left[\hat{x}_{t+1|T}\left(x_{t+1} - \hat{x}_{t+1|T}\right)'\right]\right) \\ + \left(E\left[x_{t+1}(\hat{x}_{t+1|t})'\right] + E\left[\hat{x}_{t+1|t}\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right]\right) - E\left(x_{t+1}(x_{t+1})'\right) \end{pmatrix} J'_t \\
&= \Sigma_{t|t} + J_t \left(E\left(\left(x_{t+1} - \hat{x}_{t+1|T}\right)\left(x_{t+1} - \hat{x}_{t+1|T}\right)'\right) - E\left[\left(x_{t+1} - \hat{x}_{t+1|t}\right)\left(x_{t+1} - \hat{x}_{t+1|t}\right)'\right]\right) J'_t \\
&= \Sigma_{t|t} + J_t \left(\Sigma_{t+1|T} - \Sigma_{t+1|t}\right) J'_t
\end{aligned} \tag{46}$$

Again the required objects can be stored during the Kalman filter recursion. At time T, we automatically get $\Sigma_{T|T}$ and can then work backwards.

References

Durbin, J. and Siem J. Koopman (2012). *Time series analysis by state space methods.* Second Revised Edition. Oxford: Oxford University Press.

Hamilton, James D. (1994). *Time series analysis.* Princeton, NJ: Princeton University Press.

Ljungqvist, Lars and Thomas J. Sargent (2012). *Recursive macroeconomic theory.* 3rd ed. Cambridge, MA: MIT Press.