The Kalman Filter

A brute force derivation

Weijie Chen*

Department of Political and Economic Studies

University of Helsinki

6 June, 2012

*Email: weijie.chen@helsinki.fi

1 Introduction

The Kalman filter algorithm has very high shreshhold for comprehension, and the logic is twisted with sophisticated mathematical notations. In this note, we derive the standard Kalman filter following the logic of Harvey (1990)[1]. All essential steps are presented, readers are required to think through each step carefully. The preliminary knowledge requires some advanced knowledge of matrix algebra and basic knowledge of multivariate normal distribution. The specific advanced matrix algebra techniques used in this notes will be fully explained in the first section. Then the second section will be derivation of the Kalman filter based on previous results.

2 Matrix Partitioning

There are three partitioning techniques employed in Kalman filter derivation, they are *inverse of partitioned symmetric matrix* and *determinant calculation of partitioned matrix* and *quadratic decomposition* respectively.

2.1 Inverse of Partitioned Symmetric Matrices

We need some basic facts from matrix algebra before we move onwards. The first one is

$$(A + CBD)^{-1} = A^{-1} + A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1} \tag{1}$$

It is easy to show it holds,

$$\begin{aligned}
&(A + CBD)[A^{-1} - A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1}] \\
=&(A + CBD)A^{-1} - (A + CBD)A^{-1}C(B^{-1} + DA^{-1}C)^{-1}DA^{-1} \\
=&I + CBDA^{-1} - (C + CBDA^{-1}C)(B^{-1} + DA^{-1}C)^{-1}DA^{-1} \\
=&I + CBDA^{-1} - CB(B^{-1} + DA^{-1}C)(B^{-1} + DA^{-1}C)^{-1}DA^{-1} \\
=&I + CBDA^{-1} - CBDA^{-1} = I
\end{aligned}$$

With this fact in mind, we move forwards. We have symmetric matrix A, partitioned into four blocks

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{bmatrix}$$

A_{11} has the dimension of $p \times p$, A_{22} is $q \times q$. With the same partitioned dimension, the inverse matrix B, which is also a symmetric matrix, can be partitioned accordingly

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B'_{12} & B_{22} \end{bmatrix} = A^{-1}$$

The identity matrix I_n can be written as

$$I_n = AB = \begin{bmatrix} A_{11} & A_{12} \\ A'_{12} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B'_{12} & B_{22} \end{bmatrix}$$

$$= \begin{bmatrix} A_{11}B_{11} + A_{12}B'_{12} & A_{11}B_{12} + A_{12}B_{22} \\ A'_{12}B_{11} + A_{22}B'_{12} & A'_{12}B_{12} + A_{22}B_{22} \end{bmatrix} = \begin{bmatrix} I_p & 0 \\ 0 & I_q \end{bmatrix}$$

Now we have four identities

$$A_{11}B_{11} + A_{12}B'_{12} = I_p$$

$$A_{11}B_{12} + A_{12}B_{22} = 0$$

$$A'_{12}B_{11} + A_{22}B'_{12} = 0$$

$$A'_{12}B_{12} + A_{22}B_{22} = I_q$$

Isolate B_{11}, B_{12}, B'_{12} and B_{22} on one side, four identities above become

$$B_{11} = A_{11}^{-1} - A_{11}^{-1}A_{12}B'_{12}$$

$$B_{12} = -A_{11}^{-1}A_{12}B_{22}$$

$$B'_{12} = -A_{22}^{-1}A'_{12}B_{11}$$

$$B_{22} = A_{22}^{-1} - A_{22}^{-1}A'_{12}B_{12}$$

The idea is to represent B_{11}, B_{12}, B'_{12} and B_{22} completely by blocks of A, we plug B'_{12} into B_{11}

$$\begin{aligned} B_{11} &= A_{11}^{-1} - A_{11}^{-1} A_{12}(-A_{22}^{-1} A'_{12} B_{11}) \\ &= A_{11}^{-1} + A_{11}^{-1} A_{12} A_{22}^{-1} A'_{12} B_{11} \end{aligned}$$

then collect B_{11}

$$\begin{aligned} B_{11} - A_{11}^{-1} A_{12} A_{22}^{-1} A'_{12} B_{11} &= A_{11}^{-1} \\ (I_p - A_{11}^{-1} A_{12} A_{22}^{-1} A'_{12}) B_{11} &= A_{11}^{-1} \\ (A_{11} - A_{12} A_{22}^{-1} A'_{12}) 11 &= I_p \\ (A_{11} - A_{12} A_{22}^{-1} A'_{12})^{-1} &= B_{11} \end{aligned}$$

Use the formula (1), we can get another form of B_{11}

$$B_{11} = A_{11}^{-1} - A_{11}^{-1} A_{12} (A_{22} + A'_{12} A_{11}^{-1} A_{12})^{-1} A'_{12} A_{11}^{-1}$$

Similarly for others, substitute B_{11} into B'_{12}

$$\begin{aligned} B'_{12} &= -A_{22}^{-1} A'_{12} B_{11} \\ A_{22} B'_{12} &= -A'_{12} B_{11} = -A'_{12}(A_{11}^{-1} - A_{11}^{-1} A_{12} B'_{12}) \\ A_{22} B'_{12} &= -A'_{12} A_{11}^{-1} + A_{12} A_{11}^{-1} A_{12} B'_{12} \\ B_{12} A_{22} &= -A_{11}^{-1} A_{12} + B_{12} A'_{12} A_{11}^{-1} A_{12} \\ B_{12} A_{22} - B_{12} A'_{12} A_{11}^{-1} A_{12} &= -A_{11}^{-1} A_{12} \\ B_{12}(A_{22} - A'_{12} A_{11}^{-1} A_{12}) &= -A_{11}^{-1} A_{12} \\ B_{12} &= -A_{11}^{-1} A_{12} (A_{22} - A'_{12} A_{11}^{-1} A_{12})^{-1} \end{aligned}$$

Next one, substitute B_{22} into B_{12}

$$\begin{aligned}
B_{12} &= -A_{11}^{-1}A_{12}B_{22} \\
A_{11}B_{12} &= -A_{12}B_{22} = -A_{12}(A_{22}^{-1} - A_{22}^{-1}A'_{12}B_{12}) \\
B'_{12}A_{11} &= -A_{22}^{-1}A'_{12} + B'_{12}A_{12}A_{22}A'_{12} \\
B'_{12}(A_{11} - A_{12}A_{22}^{-1}A'_{12}) &= -A_{22}^{-1}A'_{12} \\
B'_{12} &= -A_{22}^{-1}A'_{12}(A_{11} - A_{12}A_{22}^{-1}A'_{12})^{-1}
\end{aligned}$$

And last one B_{22},

$$\begin{aligned}
B_{22} &= A_{22} - A_{22}^{-1}A'_{12}B_{12} = A_{22} - A_{22}^{-1}A'_{12}(-A_{11}^{-1}A_{12}B_{22}) \\
B_{22} &= A_{22}^{-1} + A_{22}^{-1}A'_{12}A_{11}^{-1}A_{12}B_{22} \\
A_{22}^{-1} &= B_{22} - A_{22}^{-1}A'_{12}A_{11}^{-1}A_{12}B_{22} \\
A_{22}^{-1} &= (I_q - A_{22}^{-1}A'_{12}A_{11}^{-1}A_{12})B_{22} \\
B_{22} &= (A_{22} - A'_{12}A_{11}^{-1}A_{12})^{-1}
\end{aligned}$$

In summary, we can represent partitioned matrix B by the blocks of A

$$B_{11} = (A_{11} - A_{12}A_{22}^{-1}A'_{12})^{-1} \tag{2}$$

$$B_{12} = -A_{11}^{-1}A_{12}(A_{22} - A'_{12}A_{11}^{-1}A_{12})^{-1} \tag{3}$$

$$B'_{12} = -A_{22}^{-1}A'_{12}(A_{11} - A_{12}A_{22}^{-1}A'_{12})^{-1} \tag{4}$$

$$B_{22} = (A_{22} - A'_{12}A_{11}^{-1}A_{12})^{-1} \tag{5}$$

2.2 Determinant of Partitioned Matrices

Decompose the partitioned matrix A into two matrices as below

$$\begin{aligned}
A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} &= \begin{bmatrix} A_{11} & 0 \\ A'_{12} & I \end{bmatrix} \begin{bmatrix} I & A_{11}^{-1}A_{12} \\ 0 & A_{22} - A'_{12}A_{11}^{-1}A_{12} \end{bmatrix} \\
&= \begin{bmatrix} I & A_{12} \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A'_{12} & 0 \\ A_{22}^{-1}A_{21} & I \end{bmatrix}
\end{aligned}$$

Then we could calculate the determinant

$$|A| = \left| \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \right| = |A_{11}||A_{22} - A'_{12}A_{11}^{-1}A_{12}| \tag{6}$$

$$= |A_{22}||A_{11} - A_{12}A_{22}^{-1}A'_{12}| \tag{7}$$

2.3 Quadratic decomposition

For any vector u, v and a symmetric matrix $A = A'$, we have following theorem

$$u'Au - 2u'Av + v'Av = u'Au - u'Av - u'Av + v'Av$$

$$= u'A(u - v) - (u' - v')Av$$

$$= (u - v)'A(u - v) \tag{8}$$

$$= (v - u)'A(v - u) \tag{9}$$

To understand the step from the second to the third equation, you shall be awared of the fact that $u'Av = v'Au$, which is a scalar. Transpose only change the notation, but not the scalar itself.

3 Conditional Multivariate Guassian Distribution

We have a n dimensional random vector

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

x_1 and x_2 have dimension p and q respectively. x follows a multivariate normal distribution

$$x \sim N(\mu, \Sigma)$$

where $\mu = E(x)$ and $\Sigma = E(x - \mu)^2$, partitioned accordingly

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \qquad \text{and} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{11} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

The joint density of x is

$$f(x) = f(x_1, x_2) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\Sigma|}} \exp\Big(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\Big)$$
$$= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}Q(x_1, x_2)\Big)$$

where we define the quadratic form

$$Q(x_1, x_2) = (x-\mu)'\Sigma^{-1}(x-\mu)$$

Partition the equation above

$$Q(x_1, x_2) = [(x_1-\mu_1)' \quad (x_2-\mu_2)'] \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

Multiply the matices out then collect terms,

$$Q(x_1, x_2) = \Big[\; (x_1-\mu_1)'\Sigma^{11} + (x_2-\mu_2)'\Sigma^{21} \quad (x_1-\mu_1)'\Sigma^{12} + (x_2-\mu_2)'\Sigma^{22} \; \Big] \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$
$$= (x_1-\mu_1)'\Sigma^{11}(x_1-\mu_1) + (x_2-\mu_2)'\Sigma^{21}(x_1-\mu_1)$$
$$+ (x_1-\mu_1)'\Sigma^{12}(x_2-\mu_2) + (x_2-\mu_2)'\Sigma^{22}(x_2-\mu_2)$$

Because $\Sigma^{21} = \Sigma^{12}$ and $(x_1-\mu_1)'\Sigma^{12}(x_2-\mu_2)$ is a scalar, thus

$$Q(x_1, x_2) = (x_1-\mu_1)'\Sigma^{11}(x_1-\mu_1) + 2(x_1-\mu_1)'\Sigma^{12}(x_2-\mu_2) + (x_2-\mu_2)'\Sigma^{22}(x_2-\mu_2) \tag{10}$$

We need to figure out what Σ^{11}, Σ^{12} and Σ^{22} are by using facts presented in Section 2.2,

$$\Sigma^{11} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}')^{-1} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - A_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}'\Sigma_{11}^{-1} \tag{11}$$
$$\Sigma^{22} = (\Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1} = \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{12}'(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}')^{-1}\Sigma_{12}\Sigma_{22}^{-1} \tag{12}$$
$$\Sigma^{12} = -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1} = (\Sigma^{21})' \tag{13}$$

Substitute (11),(12) and (13) into (10)

$$
\begin{aligned}
Q(x_1,x_2) &= (x_1-\mu_1)'(\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}')^{-1}(x_1-\mu_1)\\
&\quad -2(x_1-\mu_1)'[\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2-\mu_2)\\
&\quad +(x_2-\mu_2)'[(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2-\mu_2)\\
&= (x_1-\mu_1)'[\Sigma_{11}^{-1}+\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}'\Sigma_{11}^{-1}](x_1-\mu_1)\\
&\quad -2(x_1-\mu_1)'[\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2-\mu_2)\\
&\quad +(x_2-\mu_2)'[(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}](x_2-\mu_2)\\
&= (x_1-\mu_1)'\Sigma_{11}^{-1}(x_1-\mu_1)\\
&\quad +(x_1-\mu_1)'\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{12}'\Sigma_{11}^{-1}(x_1-\mu_1)\\
&\quad -2(x_1-\mu_1)'\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}(x_2-\mu_2)\\
&\quad +(x_2-\mu_2)'(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}(x_2-\mu_2)\\
&= (x_1-\mu_1)'\Sigma_{11}^{-1}(x_1-\mu_1)\\
&\quad +\underbrace{[\Sigma_{12}\Sigma_{11}^{-1}(x_1-\mu_1)]'}_{u'}\underbrace{(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}}_{A}\underbrace{\Sigma_{12}'\Sigma_{11}^{-1}(x_1-\mu_1)}_{u}\\
&\quad \underbrace{-2[\Sigma_{12}\Sigma_{11}^{-1}(x_1-\mu_1)]'}_{-2u'}\underbrace{(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}}_{A}\underbrace{(x_2-\mu_2)}_{v}\\
&\quad +\underbrace{(x_2-\mu_2)'}_{v'}\underbrace{(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}}_{A}\underbrace{(x_2-\mu_2)}_{v}\\
&= (x_1-\mu_1)'\Sigma_{11}^{-1}(x_1-\mu_1)\\
&\quad +\underbrace{[(x_2-\mu_2)-\Sigma_{12}'\Sigma_{11}^{-1}(x_1-\mu_1)]'}_{(v-u)'}\underbrace{(\Sigma_{22}-\Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}}_{A}\underbrace{[(x_2-\mu_2)-\Sigma_{12}'\Sigma_{11}^{-1}(x_1-\mu_1)]}_{(v-u)}
\end{aligned}
$$

The first term of the second equation uses the fact (1), the last equation is using the fact (9).

Make more definitions to assist the derivation

$$
\begin{aligned}
b &= \mu_2 + \Sigma'_{12}\Sigma_{11}^{-1}(x_1 - \mu_1) \\
A &= \Sigma_{22} - \Sigma'_{12}\Sigma_{11}^{-1}\Sigma_{12} \\
Q_1(x_1) &= (x_1 - \mu_1)'\Sigma_{11}^{-1}(x_1 - \mu_1) \\
Q_2(x_1, x_2) &= [(x_2 - \mu_2) - \Sigma'_{12}\Sigma_{11}^{-1}(x_1 - \mu_1)]'(\Sigma_{22} - \Sigma'_{12}\Sigma_{11}^{-1}\Sigma_{12})^{-1}[(x_2 - \mu_2) - \Sigma'_{12}\Sigma_{11}^{-1}(x_1 - \mu_1)] \\
&= \{x_2 - \underbrace{[\mu_2 + \Sigma'_{12}\Sigma_{11}^{-1}(x_1 - \mu_1)]}_{b}\}' \underbrace{(\Sigma_{22} - \Sigma'_{12}\Sigma_{11}^{-1}\Sigma_{12})^{-1}}_{A^{-1}}\{(x_2 - \underbrace{[\mu_2 + \Sigma'_{12}\Sigma_{11}^{-1}(x_1 - \mu_1)]}_{b}\} \\
&= (x_2 - b)'A^{-1}(x_2 - b)
\end{aligned}
$$

The manipulation above is for change of variables in the density function, it will be sooner clear below. And notice that

$$
Q(x_1, x_2) = Q_1(x_1) + Q_2(x_1, x_2) = (x_1 - \mu_1)'\Sigma_{11}^{-1}(x_1 - \mu_1) + (x_2 - b)'A^{-1}(x_2 - b)
$$

Rewrite the joint distribution $f(x_1, x_2)$

$$
\begin{aligned}
f(x_1, x_2) &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\Big(- \frac{1}{2} Q(x_1, x_2)\Big) \\
&= \frac{1}{(2\pi)^{n/2}|\Sigma_{11}|^{1/2}|\Sigma_{22} - \Sigma'_{12}\Sigma_{11}^{-1}\Sigma_{12}|^{1/2}} \exp\Big(- \frac{1}{2} Q(x_1, x_2)\Big) \\
&= \frac{1}{(2\pi)^{p/2}|\Sigma_{11}|^{1/2}} \exp\Big(- \frac{1}{2}(x_1 - \mu_1)'\Sigma_{11}^{-1}(x_1 - \mu_1)\Big) \times \\
&\quad \frac{1}{(2\pi)^{q/2}|A|^{1/2}} \exp\Big(- \frac{1}{2}(x_2 - b)'A^{-1}(x_2 - b)\Big) \\
&= P(x_1, \mu_1, \Sigma_{11}) \; P(x_2, b, A)
\end{aligned}
$$

The second equation uses fact (7) and rest of steps are clear. The conditional distribution of x_2 given realisation of x_1 is

$$
f_{2|1}(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)} = \frac{f(x_1, x_2)}{\int_{-\infty}^{\infty} f(x_1, x_2)\mathrm{d}x_2}
$$

where we could see that

$$
\int_{-\infty}^{\infty} f(x_1, x_2)\mathrm{d}x_2 = P(x_1, \mu_1, \Sigma_{11}) \int_{-\infty}^{\infty} P(x_2, b, A)\mathrm{d}x_2 = P(x_1, \mu_1, \Sigma_{11})
$$

because $P(x_2, b, A)$ is a probability density function. Then it is crystal clear that

$$f_{2|1}(x_2|x_1) = \frac{P(x_1, \mu_1, \Sigma_{11})\ P(x_2, b, A)}{P(x_1, \mu_1, \Sigma_{11})} = P(x_2, b, A)$$

And we end the derivation of condition properties

$$f_{2|1}(x_2|x_1) = \frac{1}{(2\pi)^{q/2}|A|^{1/2}} \exp\Big(-\frac{1}{2}(x_2 - b)' A^{-1}(x_2 - b)\Big)$$

where mean vector is

$$b = \mu_2 + \Sigma_{12}' \Sigma_{11}^{-1}(x_1 - \mu_1) \tag{14}$$

and covariance matrix is

$$A = \Sigma_{22} - \Sigma_{12}' \Sigma_{11}^{-1} \Sigma_{12} \tag{15}$$

4 Time-Variant State Space Model

The Kalman filter is based the recursive state space model. This section will present you the general setting of state space model we need in derivation of the Kalman filter. First, we define the transition equation

$$x_t = c_t + F_t x_{t-1} + B_t v_t \qquad E(v_t) = 0 \text{ and } \operatorname{Var}(v_t) = Q_t \tag{16}$$

the coefficient matrix F_t and B_t are labelled with time t, which indicates a time-variant state space model[1]. The x_t is the state vector which describes the state of the system, and it is an unobservable vector. And c_t is a deterministic term.

The measurement equation is given by

$$y_t = H_t x_t + d_t + \omega_t \qquad E(\omega_t) = 0 \text{ and } \operatorname{Var}(\omega_t) = R_t \tag{17}$$

y_t is observable, namely input data. d_t is a deterministic term. Assume that

$$E(\omega_t v_k') = \mathbf{0}$$

[1] If all variables are stationary, the state-space model is time-invariant.

and also

$$E(v_t x_0') = \mathbf{0} \text{ and } E(\omega_t x_0') = \mathbf{0}$$

5 Derivation of The Kalman Filter

With the Gaussian assumption of v_t, x_0 has multivariate Gaussian distribution

$$x_0 \sim N(E(x_0), P_0)$$

From transition equation

$$x_1 = c_1 + F_1 x_0 + B_1 v_1$$

Take expectation

$$E(c_1 + F_1 x_0 + B_1 v_1) = c_1 + F_1 E(x_0) = E(x_{1|0}) \tag{18}$$

which is a conditional expectation based on information known at time 1. Also take variance

$$\begin{aligned}
&\mathrm{Var}\,(c_1 + F_1 x_0 + B_1 v_1) \\
&= F_1 \,\mathrm{Var}\,(x_0) F_1' + B_1 \,\mathrm{Var}\,(v_1) B_1' = \mathrm{Var}\,(x_{1|0}) \\
&= F_1 P_0 F_1' + B_1 Q_1 B_1' = P_{1|0}
\end{aligned} \tag{19}$$

which is the conditional variance based on information know at time 1. Rewrite the state space model as

$$\begin{aligned}
x_1 &= E(x_{1|0}) + [x_1 - E(x_{1|0})] \\
y_1 &= H_1 E(x_{1|0}) + d_1 + H_1[x_1 - E(x_{1|0})] + \omega_1
\end{aligned}$$

the purpose of rewriting is clear by noticing that $[x_1 - E(x_{1|0})]$ is one-step ahead forecast error. Then they follow a multivariate normal distribution

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} E(x_{1|0}) \\ H_1 E(x_{1|0} + d_1) \end{bmatrix}, \begin{bmatrix} P_{1|0} & P_{1|0} H_1' \\ H_1 P_{1|0} & H_1 P_{1|0} H_1' + R_1 \end{bmatrix} \right)$$

We have assumed that the expectation of forecast error $E([x_1 - E(x_{1|0})]) = 0$. Use the results from Section 3, (14) and (15), we have[2]

$$E(x_1) = \underbrace{E(x_{1|0})}_{\mu_1} + \underbrace{P_{1|0}H_1'}_{\Sigma_{21}'}\underbrace{(H_1P_{1|0}H_1' + R_1)^{-1}}_{\Sigma_{22}^{-1}}\{\underbrace{y_1}_{x_2} - \underbrace{[H_1E(x_{1|0}) + d_1]}_{\mu_2}\} \tag{20}$$

$$P_1 = \underbrace{P_{1|0}}_{\Sigma_{11}} - \underbrace{P_{1|0}H_1'}_{\Sigma_{21}'}\underbrace{(H_1P_{1|0}H_1' + R_1)^{-1}}_{\Sigma_{22}^{-1}}\underbrace{H_1P_{1|0}}_{\Sigma_{21}} \tag{21}$$

All above derivation is just the first iteration of the Kalman filter from period 0 to 1.

In general we have prediction equations and updating equations in the Kalman filtering algorithm, from (18)

$$E(x_{t|t-1}) = c_t + F_tE(x_{t-1})$$

and also from (19)

$$P_{t|t-1} = F_tP_{t-1}F_t' + B_tQ_tB_t'$$

The two equations are *prediction equations*, the philosophy behind the name indicates that we are using the information at time $t-1$, and $t-1$ only, to predict the variables at t.

After prediction, we are waiting for the new observations coming into the system, once a new observation y_t comes, we update the system

$$\begin{aligned} E(x_t) &= E(x_{t|t-1}) + P_{t|t-1}H_1'(H_tP_{t|t-1}H_t' + R_t)^{-1}[y_t - H_tE(x_{t|t-1}) - x_t] \\ P_t &= P_{t|t-1}H_t'(H_tP_{t|t-1}H_t' + R_t)^{-1}H_tP_{t|t-1} \end{aligned}$$

these two equations are *update equations*, they are from (20) and (21).

In conclusion, the philosophy behind the Kalman filter is to recursively update the information of the system, this is the most original version of the Kalman filter. The

[2] The notation for x_2 in the underbrace might be confused with state vector, however this is in line with the notation used in Section 3, specifically equation (14).

Kalman filter has very restrict assumptions about the stochastic components of the system, if we relax those assumptions, the Kalman filter will be modified accordingly to handle more general situations.

References

[1] Harvey, A. (1990): *The Econometric Analysis of Time Series*, the MIT Press, 2nd Edition

[2] Kalman, R. (1960): 'A New Approach to Linear Filtering and Prediction Problem,' *Journal of Basic Engineering*,pp. 35-45 (March 1960)