

Kalman Filter: A Simple Derivation

Ali M. Mosammam

Department of Statistics, University of Zanjan, Zanjan, Iran

*Corresponding Author: a.m.mosammam@znu.ac.ir

Copyright ©2015 Horizon Research Publishing All rights reserved.

Abstract The Kalman filter is a recursive estimator and plays a fundamental role in statistics for filtering, prediction and smoothing. The key element in any recursive estimator is the estimate of the current state, x_k , at time k , based on observations up to and including observation k and the Kalman filter enables the estimate of the state to be updated as new observations become available. In this paper we have tried to derive the Kalman filter as simple as possible.

Keywords Kalman Filter, State-space Model, Dynamic System, Gaussian Process

1 Introduction

The analysis of a time series [5] and spatial analysis [2] can be done in one of two ways. Firstly, it could be analysed in a batch. In other words, the whole time series to date is analysed. Secondly, analysis of the time series sequentially. With this approach the current estimate of the state of the system is updated over time as new measurements are made. Whilst both methods will give the same answer, the sequential method has two advantages:

1. reduction in the computational cost
2. reduction in the storage capacity.

The disadvantage of sequential methods is appeared when the posterior can not be calculated exactly. In this paper we will concentrate on the state-space approach to modelling dynamic systems. In order to analyze a dynamic system, at least two models are required:

- First, a model describing the evolution of the state with time (the state model)
- Second, a model relating the noisy measurements to the state (the measurement model).

In the Bayesian approach to dynamic state estimation, one attempts to construct the posterior probability density function of the state based on all available information, including the set of received measurements. For many problems, an estimate is required every time that a measurement is received. In this case, a recursive filter

is a convenient solution. A recursive filtering approach means that received data can be processed sequentially rather than as a batch so that it is not necessary to store the complete data set nor to reprocess existing data if a new measurement becomes available. There are classes of problems for which the recursive Bayesian solutions are tractable. The most important of these classes is the set of problems where the state and observation equations are linear, and the distributions of the prior, and observation and state noise are Gaussian. In this case no algorithm can ever do better than a Kalman filter [4].

In this paper we have tried to derive the Kalman filter properly. If the state space model is linear with uncorrelated Gaussian noise and a Gaussian prior, the prior and posterior distributions at each time step are themselves Gaussian random variables. A Gaussian random variable's distribution is uniquely defined by the specification of its mean and variance. The Kalman filter is a recursive estimator and plays a fundamental role in statistics for filtering, prediction and smoothing. The key element in any recursive estimator is the estimate of the current state, x_k , at time k , based on observations up to and including observation k and the Kalman filter enables the estimate of the state to be updated as new observations become available.

2 Basic model

The basic model in order to derive the Kalman filter is described by linear, discrete-time, finite-dimensional state-space equations

$$\begin{aligned}x_{k+1} &= ax_k + w_k, \quad k \geq 0 \quad (\text{state equation}) \quad (1) \\y_k &= cx_k + v_k \quad (\text{measurement model}) \quad (2)\end{aligned}$$

Here we shall make the following assumptions:

1. $\{v_k\}$ and $\{w_k\}$ are individually independent, zero-mean, Gaussian processes with known variances.

If $w_k \sim N(0, \sigma_w^2)$, then we have $E(w_k) = 0$, $\text{var}(w_k) = \sigma_w^2$ and $E(w_k w_\ell) = E(w_k)E(w_\ell) = 0$ for all $k \neq \ell$.

If $v_k \sim N(0, \sigma_v^2)$, we have $E(v_k) = 0$, $\text{var}(v_k) = \sigma_v^2$ and $E(v_k v_\ell) = 0$ for all $k \neq \ell$.

2. $\{v_k\}$ and $\{w_k\}$ are independent processes. Hence $E(w_k v_\ell) = 0$ for all k, ℓ .

3. At the initial time, $k = 0$, the initial state, $x_0 \sim N(\tilde{x}_0, P_0)$, further we shall assume that x_0 is independent of v_k and w_k for any k . Hence

- (a) $E(x_k v_\ell) = 0$ for all k, ℓ .
 (b) $E(x_k w_\ell) = 0$ for all k, ℓ .

This state-space model has some properties:

1. Since x_k is a linear combination of the jointly Gaussian random variables $x_0, w_0, w_1, \dots, w_{k-1}$, it is a Gaussian variable.
2. $\{x_k\}$ is a Markov process.
3. $\{y_k\}$ is a Gaussian process.
4. $\{x_k\}$ and $\{y_k\}$ are jointly Gaussian.

We have made these assumptions in order to derive the Kalman filter easily. Some of these assumptions can be relaxed in practice.

3 Scalar Kalman filter

We have already introduced the basic models (1, 2) to derive the Kalman filter. We want to show that under some assumptions conditional expectation yields a set of recursive equations which is known the Kalman filter recursions. Here we shall denote the following quantities:

- $\hat{x}_{k|k} = E(x_k | y_0, \dots, y_k)$ is the estimate of the current state, x_k , at time k , based on observations up to and including observation k .
- $\hat{x}_{k|k-1} = E(x_k | y_0, \dots, y_{k-1})$ is the predictor of the current state, x_k , at time k , based on $k-1$ observations.
- $\Sigma_{k|k} = \text{var}(x_k | y_0, \dots, y_k)$ is the variance of the current state, x_k , at time k , based on observations up to and including observation k .
- $\Sigma_{k|k-1} = \text{var}(x_k | y_0, \dots, y_{k-1})$ is the variance of the current state, x_k , at time k , based on first $k-1$ observations.

We might interested in prediction of $E(y_k | y_0, \dots, y_{k-1})$. Similarly notations can be established for observation equation. From (2) and the third assumption it can be written

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \sim N_2 \left(\begin{bmatrix} \tilde{x}_0 \\ c\tilde{x}_0 \end{bmatrix}, \begin{bmatrix} P_0 & cP_0 \\ cP_0 & c^2P_0 + \sigma_v^2 \end{bmatrix} \right).$$

Hence the conditional distribution takes the form $x_0 | y_0 \sim N(\hat{x}_{0|0}, \Sigma_{0|0})$, which is completely specified by following conditional mean and variance:

$$\begin{aligned} \hat{x}_{0|0} &= E(x_0 | y_0) = \tilde{x}_0 + \frac{cP_0}{c^2P_0 + \sigma_v^2} (y_0 - c\tilde{x}_0), \\ \Sigma_{0|0} &= \text{var}(x_0 | y_0) = P_0 - \frac{(cP_0)^2}{c^2P_0 + \sigma_v^2}. \end{aligned}$$

Using state-space equation (1, 2) for $k = 0$, we get

$$\begin{aligned} x_1 &= ax_0 + w_0 \\ y_1 &= cx_1 + v_1. \end{aligned}$$

Using,

$$\begin{aligned} \begin{bmatrix} x_1 \\ y_0 \end{bmatrix} &\sim N_2 \left(\begin{bmatrix} a\tilde{x}_0 \\ c\tilde{x}_0 \end{bmatrix}, \begin{bmatrix} a^2P_0 + \sigma_w^2 & acP_0 \\ acP_0 & c^2P_0 + \sigma_v^2 \end{bmatrix} \right) \\ \begin{bmatrix} y_1 \\ y_0 \end{bmatrix} &\sim N_2 \left(\begin{bmatrix} ac\tilde{x}_0 \\ c\tilde{x}_0 \end{bmatrix}, \begin{bmatrix} c^2(a^2P_0 + \sigma_w^2) + \sigma_v^2 & ac^2P_0 \\ ac^2P_0 & c^2P_0 + \sigma_v^2 \end{bmatrix} \right), \end{aligned}$$

the conditional distributions are themselves normal and completely specified by conditional means and variances

$$\begin{aligned} x_1 | y_0 &\sim N(\hat{x}_{1|0}, \Sigma_{1|0}) \\ \hat{x}_{1|0} &= a\tilde{x}_0 + \frac{acP_0}{c^2P_0 + \sigma_v^2} (y_0 - c\tilde{x}_0) = a\hat{x}_{0|0} \\ \Sigma_{1|0} &= a^2P_0 + \sigma_w^2 - \frac{(acP_0)^2}{c^2P_0 + \sigma_v^2} = a^2\Sigma_{0|0} + \sigma_w^2 \\ y_1 | y_0 &\sim N(c\hat{x}_{1|0}, c^2\Sigma_{1|0} + \sigma_v^2) \\ E(y_1 | y_0) &= ac\tilde{x}_0 + \frac{ac^2P_0}{c^2P_0 + \sigma_v^2} (y_0 - c\tilde{x}_0) = c\hat{x}_{1|0} \\ \text{var}(y_1 | y_0) &= c^2(a^2P_0 + \sigma_w^2) + \sigma_v^2 - \frac{(ac^2P_0)^2}{c^2P_0 + \sigma_v^2} = c^2\Sigma_{1|0} + \sigma_w^2. \end{aligned}$$

Furthermore,

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} | y_0 \sim N_2 \left(\begin{bmatrix} \hat{x}_{1|0} \\ c\hat{x}_{1|0} \end{bmatrix}, \begin{bmatrix} \Sigma_{1|0} & c\Sigma_{1|0} \\ c\Sigma_{1|0} & c^2\Sigma_{1|0} + \sigma_v^2 \end{bmatrix} \right).$$

Hence $x_1 | y_0, y_1 \sim N(\hat{x}_{1|1}, \Sigma_{1|1})$, where its mean and variance are

$$\begin{aligned} \hat{x}_{1|1} &= E(x_1 | y_0, y_1) = \hat{x}_{1|0} + \frac{c\Sigma_{1|0}}{c^2\Sigma_{1|0} + \sigma_v^2} (y_1 - c\hat{x}_{1|0}) \\ \Sigma_{1|1} &= \text{var}(x_1 | y_0, y_1) = \Sigma_{1|0} - \frac{(c\Sigma_{1|0})^2}{c^2\Sigma_{1|0} + \sigma_v^2}. \end{aligned}$$

Let we have repeated this method and found $x_{k-1} | y_0, \dots, y_{k-1} \sim N(\hat{x}_{k-1|k-1}, \Sigma_{k-1|k-1})$, using state (1, 2) we can update equations by

$$\begin{aligned} x_k &= ax_{k-1} + w_{k-1} && \text{(state equation)} \\ y_{k-1} &= cx_{k-1} + v_{k-1} && \text{(measurement model)}. \end{aligned}$$

Hence the conditional distributions are

$$\begin{aligned} x_k | y_0, \dots, y_{k-1} &\sim N(\hat{x}_{k|k-1}, \Sigma_{k|k-1}) \\ y_k | y_0, \dots, y_{k-1} &\sim N(c\hat{x}_{k|k-1}, c^2\Sigma_{k|k-1} + \sigma_v^2) \end{aligned}$$

where $\hat{x}_{k|k-1} = a\hat{x}_{k-1|k-1}$ and $\Sigma_{k|k-1} = a^2\Sigma_{k-1|k-1} + \sigma_w^2$ and

$$\begin{aligned} \begin{bmatrix} x_k \\ y_k \end{bmatrix} | y_0, \dots, y_{k-1} \\ \sim N_2 \left(\begin{bmatrix} \hat{x}_{k|k-1} \\ c\hat{x}_{k|k-1} \end{bmatrix}, \begin{bmatrix} \Sigma_{k|k-1} & c\Sigma_{k|k-1} \\ c\Sigma_{k|k-1} & c^2\Sigma_{k|k-1} + \sigma_v^2 \end{bmatrix} \right) \end{aligned}$$

Hence $x_k | y_0, \dots, y_k \sim N(\hat{x}_{k|k}, \Sigma_{k|k})$, where

$$\begin{aligned} \hat{x}_{k|k} &= E(x_k | y_0, \dots, y_k) = \hat{x}_{k|k-1} + \frac{c\Sigma_{k|k-1}}{c^2\Sigma_{k|k-1} + \sigma_v^2} (y_k - c\hat{x}_{k|k-1}) \\ \Sigma_{k|k} &= \text{var}(x_k | y_0, \dots, y_k) = \Sigma_{k|k-1} - \frac{(c\Sigma_{k|k-1})^2}{c^2\Sigma_{k|k-1} + \sigma_v^2}. \end{aligned} \quad (3)$$

So the Kalman filter equations can be written as

$$\begin{aligned} \hat{x}_{k|k} &= E(x_k | y_0, \dots, y_k) = \hat{x}_{k|k-1} + b_k (y_k - c\hat{x}_{k|k-1}) \\ \Sigma_{k|k} &= \text{var}(x_k | y_0, \dots, y_k) = \Sigma_{k|k-1} - cb_k \Sigma_{k|k-1}. \end{aligned} \quad (4)$$

where

$$b_k = \frac{c\Sigma_{k|k-1}}{c^2\Sigma_{k|k-1} + \sigma_v^2}. \tag{5}$$

Notice that the conditional variance is independent of y_0, \dots, y_k . Alternatively, using $\hat{x}_{k|k-1} = a\hat{x}_{k-1|k-1}$ the Kalman filter can be written by

$$\hat{x}_{k|k} = E(x_k|y_0, \dots, y_k) = a\hat{x}_{k-1|k-1} + b_k(y_k - ac\hat{x}_{k-1|k-1}).$$

Note that $\hat{x}_{k|k}$ is the estimate of state x_k at time k based on the previous estimate and only one data at time k . No algorithm can ever do better than a Kalman filter in the linear Gaussian state space. It should be noted that it is possible to derive the same results by minimizing MSE.

4 Kalman filter predictor

We might intrested in $E(x_{k+\ell}|y_0, \dots, y_k)$, which is known as the Kalman filter predictor.

1. One-step predictor:

The same as the last section we again update x_{k+1} applying $x_{k+1} = ax_k + w_k$. Therefore $x_{k+1}|y_0, \dots, y_k \sim N(a\hat{x}_{k|k}, a^2\Sigma_{k|k} + \sigma_w^2)$ and applying equation(4), we obtain

$$\begin{aligned} \hat{x}_{k+1|k} &= a\hat{x}_{k|k} = a\hat{x}_{k|k-1} + ab_k(y_k - c\hat{x}_{k|k-1}) \\ \Sigma_{k+1|k} &= a^2\Sigma_{k|k} + \sigma_w^2 = a^2(1 - cb_k)\Sigma_{k|k-1} + \sigma_w^2. \end{aligned} \tag{6}$$

Note that if we define the innovation process $\{\tilde{y}_k\}$ by $y_k - c\hat{x}_{k|k-1}$ and also assume that $\tilde{y}_0 = y_0 - E(y_0)$, then we have $\hat{x}_{k+1|k} = a\hat{x}_{k|k-1} + ab_k\tilde{y}_k$ in terms of \tilde{y}_k .

2. Multi-step predictor:

Using (1), and substituting these sequences we can write

$$x_{k+\ell} = a^\ell x_k + \sum_{i=1}^{\ell} a^{i-1} w_{k+\ell-i}.$$

Then the ℓ -step prediction is given by

$$\hat{x}_{k+\ell|k} = E(x_{k+\ell}|y_0, \dots, y_k) = a^\ell \hat{x}_{k|k},$$

and $\Sigma_{k+\ell|k} = \text{var}(x_{k+\ell}|k) = a^{2\ell}\Sigma_{k|k} + \sigma_w^2 \frac{1-a^{2\ell}}{1-a^2}$. Since $y_{k+\ell} = cx_{k+\ell} + v_{k+\ell}$ we have, then

$$\begin{aligned} E(y_{k+\ell}|y_0, \dots, y_k) &= cE(x_{k+\ell}|y_0, \dots, y_k) = c\hat{x}_{k+\ell|k} \\ \text{var}(y_{k+\ell}|y_0, \dots, y_k) &= c^2\Sigma_{k+\ell|k} + \sigma_v^2. \end{aligned}$$

5 Vector Kalman filter

We have dealt so far with the scalar state equation by a first order autoregressive process. We can write directly the vector Kalman filter which each one is the first-order autoregressive process:

$$x_{jk+1} = a_j x_{jk} + w_{jk} \quad j = 1, 2, \dots, q. \tag{7}$$

Let $X_k = [x_{1k}, x_{2k}, \dots, x_{qk}]^T$ and $W_k = [w_{1k}, w_{2k}, \dots, w_{qk}]^T$.

The q equations (7) can be written as the first-order vector equation, $X_{k+1} = AX_k + W_k$, where A is a $q \times q$ matrix, in this case given by $A = \text{diag}(a_1, a_2, \dots, a_q)$ and

$W_k \sim N(0, Q_k)$. Assume that observation equations at time k are

$$y_{ik} = c_i x_{ik} + v_{ik} \quad i = 1, \dots, r \quad r < q.$$

We define observation vector equation by $Y_k = CX_k + V_k$, where

$$C = \begin{pmatrix} c_1 & 0 & \dots & 0 & \dots & 0 \\ 0 & c_2 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & c_r & \dots & 0 \end{pmatrix}$$

$Y_k = [y_{1k}, y_{2k}, \dots, y_{rk}]^T$, $V_k = [v_{1k}, v_{2k}, \dots, v_{rk}]^T$ and $V_k \sim N(0, R_k)$.

Hence we can directly transfer the scalar Kalman filter, given by equations (3) to (5) in to the corresponding vector Kalman filter form

$$\begin{aligned} \hat{X}_{k|k} &= \hat{X}_{k|k-1} + \beta_k(Y_k - C\hat{X}_{k|k-1}) \\ \Sigma_{k|k} &= \Sigma_{k|k-1} - \beta_k C \Sigma_{k|k-1}, \end{aligned}$$

where $\beta_k = \Sigma_{k|k-1} C^T [C \Sigma_{k|k-1} C^T + R_k]^{-1}$ and $\Sigma_{k|k-1} = A \Sigma_{k-1|k-1} A^T + Q_{k-1}$.

- We can obtain the following vector and matrix set of equation directly from scalar Kalman predictor(6).

$$\begin{aligned} \hat{X}_{k+1|k} &= A\hat{X}_{k|k-1} + A\beta_k(Y_k - C\hat{X}_{k|k-1}) \\ \Sigma_{k+1|k} &= A(I - \beta_k C)\Sigma_{k|k-1}A^T + Q_k, \end{aligned}$$

which β_k is already defined.

6 Some properties of Kalman filter

1. The Kalman filter defined in the last sections was conditional expectation of $E(x_k|y_0, \dots, y_k)$ which minimizes the mean square error. It is also maximum likelihood estimator and the best linear unbiased estimator whose variance is less than any other linear unbiased estimators. Even we drop normality assumption, $E(x_k|y_0, \dots, y_k)$ also minimizes the variance. Here $E(x_k|y_0, \dots, y_k)$ is a function of y_0, \dots, y_k and without the normality assumption will not necessarily be linear.
2. Let $AY + b$ be a linear estimator of X given Y , where A is a fixed matrix and b is a fixed vector. We define a linear minimum variance estimator, $\hat{X}(Y) = A_0Y + b_0$, contrast with $E(X|Y)$ the calculation of $\hat{X}(Y)$ does not require the joint probability density. If X and Y are jointly Gaussian, the minimum variance and linear minimum variance estimators coincide.
3. If $\hat{X}(Y)$ is a linear minimum variance estimate of X , then $C\hat{X}(Y) + d$ is a linear minimum variance estimator of $CX + d$, where C and d are fixed matrix and vector respectively. It is clearly property of linearity of the expectation operator.
4. $\hat{X}(Y)$ is unbiased, because $E(\hat{X}(Y)) = E\{\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - \mu_y)\} = \mu_x$. Recall that $E(X|Y)$ is a conditional minimum variance estimator. Infact $\hat{X}(y)$ evaluated at $Y = y$ is a conditional minimum variance estimate for all y if and only if $E(X|Y) = cy + d$ for some c, d and this implies $\hat{X}(y) = E(X|Y)$.

5. Kalman filter can be extended to many nonlinear problems [3]. As most real world models are non linear and non Gaussian and many analysis tasks involve estimating the state of a dynamic model when only partial or inaccurate observations are available, it is of great interest to develop efficient computational methods to solve this so called Bayesian filtering problem numerically. Many approximation methods such as a Extended Kalman filter (EKF) have been developed to cope with this problem. The EKF linearises the state and observation equations, and then uses the Kalman Filter to obtain estimates for the state. For more details see e.g. [1].

7 Smoothing Kalman filter

We already have considered $\hat{x}_{k|k} = E(x_k|y_0, \dots, y_k)$ and $\hat{x}_{k+\ell|k} = E(x_{k+\ell}|y_0, \dots, y_k)$ as classic Kalman filter and predictor. One might be interested in $\hat{x}_{j|k} = E(x_j|y_0, \dots, y_k), j < k$, which is known as smoothing Kalman filter. We expect this estimator to be more accurate than $E(x_j|y_0, \dots, y_k)$, because more measurements are used in producing $\hat{x}_{j|k}$. The simplest smoothing is fixed-point smoothing where, we want to determine $\hat{x}_{j|k} = E(x_j|y_0, \dots, y_k)$ and the associated variance, $\Sigma_{j|k} = E\{(x_j - \hat{x}_{j|k})(x_j - \hat{x}_{j|k})^T | y_0, \dots, y_k\}$ for some fixed j and all $k > j$. Let an augmenting state x_k^* is defined by $x_{k+1}^* = x_k^*, k \geq j$ with $x_j^* = x_j$ at the initial time, then $x_{k+1}^* = x_j, \forall k \geq j$. Hence $\hat{x}_{k+1|k}^* = \hat{x}_{j|k}$ and $\Sigma_{k+1|k}^{**} = \Sigma_{j|k}$, where $\Sigma_{k+1|k}^{**}$ denotes the variance of error ($\hat{x}_{k+1|k}^* - x_{k+1}^*$). The augment state space model is therefore

$$\begin{aligned} X_{k+1} &= \begin{bmatrix} x_{k+1} \\ x_{k+1}^* \end{bmatrix} \\ &= \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ x_k^* \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_k = AX_k + W_k \\ Y_k = y_k &= \begin{bmatrix} c & 0 \end{bmatrix} \begin{bmatrix} x_k \\ x_k^* \end{bmatrix} + V_k = CX_k + V_k, \end{aligned}$$

with the state vector at $k = j$ satisfying

$$\begin{bmatrix} x_j \\ x_j^* \end{bmatrix} = \begin{bmatrix} x_j \\ x_j \end{bmatrix}.$$

Using the Kalman filter predictor

$$\begin{aligned} \hat{X}_{k+1|k} &= \begin{bmatrix} \hat{x}_{k+1|k} \\ \hat{x}_{k+1|k}^* \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{x}_{k|k-1}^* \end{bmatrix} \\ &+ \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_k \\ b_k^* \end{bmatrix} \left(Y_k - \begin{bmatrix} c & 0 \end{bmatrix} \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{x}_{k|k-1}^* \end{bmatrix} \right), \end{aligned}$$

where

$$\begin{aligned} \beta_k &= \begin{bmatrix} b_k \\ b_k^* \end{bmatrix} = \begin{bmatrix} \Sigma_{k|k-1} & \Sigma_{k|k-1}^* \\ \Sigma_{k|k-1}^* & \Sigma_{k|k-1}^{**} \end{bmatrix} \begin{bmatrix} c \\ 0 \end{bmatrix} \\ &\times \left(\begin{bmatrix} c & 0 \end{bmatrix} \begin{bmatrix} \Sigma_{k|k-1} & \Sigma_{k|k-1}^* \\ \Sigma_{k|k-1}^* & \Sigma_{k|k-1}^{**} \end{bmatrix} \begin{bmatrix} c \\ 0 \end{bmatrix} + \sigma_w^2 \right)^{-1}, \end{aligned}$$

and

$$\begin{aligned} \begin{bmatrix} \Sigma_{k+1|k} & \Sigma_{k+1|k}^* \\ \Sigma_{k+1|k}^* & \Sigma_{k+1|k}^{**} \end{bmatrix} &= \\ \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} b_k \\ b_k^* \end{bmatrix} \begin{bmatrix} c & 0 \end{bmatrix} \right) \times \\ \begin{bmatrix} \Sigma_{k|k-1} & \Sigma_{k|k-1}^* \\ \Sigma_{k|k-1}^* & \Sigma_{k|k-1}^{**} \end{bmatrix} &+ \begin{bmatrix} \sigma_w^2 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

It can be written as

$$\begin{bmatrix} \Sigma_{k+1|k} & \Sigma_{k+1|k}^* \\ \Sigma_{k+1|k}^* & \Sigma_{k+1|k}^{**} \end{bmatrix} = \begin{bmatrix} a^2(1 - cb_k)\Sigma_{k|k-1} + \sigma_w^2 & a(1 - cb_k)\Sigma_{k|k-1}^* \\ a(1 - cb_k^*)\Sigma_{k|k-1} & \Sigma_{k|k-1}^{**} - cb_k^*\Sigma_{k|k-1}^* \end{bmatrix}.$$

Note that $\Sigma_{k+1|k}$ appearing here is precisely that associated with the scalar Kalman filter equation, because the first element of the augmented Kalman filter is $\hat{x}_{k|k-1}$. Hence for $k \geq j$,

$$\begin{aligned} \hat{x}_{j|k} &= E(x_j|y_0, \dots, y_k) = \hat{x}_{j|k-1} + b_k^*(y_k - c\hat{x}_{k|k-1}), \\ b_k^* &= \frac{c\Sigma_{k|k-1}^*}{c^2\Sigma_{k|k-1} + \sigma_w^2}, \\ \Sigma_{k+1|k}^* &= a(1 - cb_k)\Sigma_{k|k-1}^*, \end{aligned}$$

with initial value $\Sigma_{j|j-1}^* = \Sigma_{j|j-1}$.

The variance of the error term, $(x_j - \hat{x}_{j|k})$, is given by

$$\Sigma_{j|k} = \Sigma_{j|k-1} - cb_k^*\Sigma_{k|k-1}^*.$$

The fixed-point smoother has some properties:

1. The fixed point smoother is driven from the innovations process, $\tilde{y}_k = y_k - c\hat{x}_{k|k-1}$ of the Kalman filter for the nonaugmented model.
2. The smoother is a linear discrete time model of dimension equal to that of the filter.
3. As in the case of filter, $x_j|y_0, \dots, y_k$ is Gaussian and therefore defined by the conditional mean $\hat{x}_{j|k}$ and $\Sigma_{j|k}$.

8 General form of the vector Kalman filter

In general we can write the state-space equation by

$$\begin{aligned} X_{k+1} &= A_k X_k + G_k W_k, \quad k \geq 0 \quad (\text{state equation}) \\ Y_k &= C_k X_k + V_k \quad (\text{measurement model}), \end{aligned}$$

where we have all the assumptions which we made in the scalar one. Furthermore we assume that the covariance of the $\{V_k\}$ is given by $E[V_k V_k^T] = R_k \delta_{k\ell}$ where

$$\delta_{k\ell} = \begin{cases} 1 & k = \ell \\ 0 & o.w \end{cases}$$

and also covariance of the $\{W_k\}$ process is $E[W_k W_k^T] = Q_k \delta_{k\ell}$. Hence $\{V_k\}$ and $\{W_k\}$ are zero mean, independent Gaussian processes. Note that in the general form A_k, G_k and C_k are dependent on time. Let

$X_0 \sim N(\tilde{X}_0, P_0)$. Similar to scalar one by using conditional expectation we have

$$\begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} \sim N \left(\begin{bmatrix} \tilde{X}_0 \\ C_0 \tilde{X}_0 \end{bmatrix}, \begin{bmatrix} P_0 & P_0 C_0^T \\ C_0 P_0 & C_0 P_0 C_0^T + R_0 \end{bmatrix} \right).$$

Hence $\hat{X}_{0|0} = \tilde{X}_0 + P_0 C_0^T (C_0 P_0 C_0^T + R_0)^{-1} (Y_0 - C_0 \tilde{X}_0)$ and covariance is defined by $\Sigma_{0|0} = P_0 - P_0 C_0^T (C_0 P_0 C_0^T + R_0)^{-1} C_0 P_0$. We can update state equation by $X_1 = A_0 X_0 + G_0 W_0$. So $\hat{X}_{1|0} = A_0 \hat{X}_{0|0}$ and $\Sigma_{1|0} = A_0 \Sigma_{0|0} A^T + G_0 Q_0 G_0^T$. Also from $Y_1 = C_1 X_1 + V_1$ it follows that Y_1 condition on Y_0 is Gaussian with mean, $\hat{Y}_{1|0} = C_1 \hat{X}_{1|0}$, and covariance, $C_1 \Sigma_{1|0} C_1^T + R_1$. Hence

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} | Y_0 \sim N \left(\begin{bmatrix} \hat{X}_{1|0} \\ C_1 \hat{X}_{1|0} \end{bmatrix}, \begin{bmatrix} \Sigma_{1|0} & \Sigma_{1|0} C_1^T \\ C_1 \Sigma_{1|0} & C_1 \Sigma_{1|0} C_1^T + R_1 \end{bmatrix} \right).$$

Then X_1 given Y_0, Y_1 has normal distribution by conditional mean and covariance,

$$\begin{aligned} \hat{X}_{1|1} &= \hat{X}_{1|0} + \Sigma_{1|0} C_1^T (C_1 \Sigma_{1|0} C_1^T + R_1)^{-1} (Y_1 - C_1 \hat{X}_{1|0}) \\ \Sigma_{1|1} &= \Sigma_{1|0} - \Sigma_{1|0} C_1^T (C_1 \Sigma_{1|0} C_1^T + R_1)^{-1} C_1 \Sigma_{1|0}. \end{aligned}$$

Repeating this method yields

$$\begin{aligned} \hat{X}_{k|k} &= \hat{X}_{k|k-1} + B_k (Y_k - C_k \hat{X}_{k|k-1}) \\ \Sigma_{k|k} &= \Sigma_{k|k-1} - B_k C_k \Sigma_{k|k-1}, \end{aligned}$$

where $B_k = \Sigma_{k|k-1} C_k^T (C_k \Sigma_{k|k-1} C_k^T + R_k)^{-1}$ and $\Sigma_{k|k-1} = A_{k-1} \Sigma_{k-1|k-1} A_{k-1}^T + G_{k-1} Q_{k-1} G_{k-1}^T$.

REFERENCES

- [1] B. D. Anderson and J. B. Moore. Optimal filtering, Courier Corporation, 2012.
- [2] N. Cressie. Statistics for Spatial Data, Revised edition, John Wiley, New York, 1993.
- [3] D. Crisan and B. Rozovskii. The Oxford handbook of nonlinear filtering, Oxford University Press , 2011.
- [4] R. E. Kalman. A new approach to linear filtering and prediction problems, Journal of Fluids Engineering, Vol.82, No.1, 35-45, 1960.
- [5] R. H. Shumway and D. S. Stoffer, Time series analysis and its applications: with R examples, Springer., 2011