An introduction to the Kalman Filter

Yves Piguet

November 2018

1 Introduction

This report gives a simple introduction to the Kalman filter. The purpose of the Kalman filter is to estimate the state of a linear system using the system input and (measured) output. It is based on the system model which includes the effect of noise. The Extended Kalman Filter (EKF) is a generalization to nonlinear systems. EKF plays an important role in making the best use of multiple sources of measurements, each with their specific precision, bandwidth, etc.; for instance to estimate the position, orientation and velocity of a drone using its accelerometers, gyroscopes and GPS. The estimation obtained with a Kalman filter can be used for analysis of a system behavior by a human, or in a feedback loop in automatic control.

The reader is assumed to have a basic understanding of dynamical systems, state-space models (non-linear and linear), linear algebra, and basic multivariate statistics.

2 A simple scalar example

Consider the following equation, which describes the position of a point moving at constant speed b (we reserve letter v for later):

$$x(t) = x_0 + bt$$

It can be represented by the following discrete-time state-space model:

$$x_{k+1} = x_k + d$$

where x_k is the position at sampled time k, x_{k+1} the position one sample later, and d the displacement. In a real motion, the speed is probably not perfectly constant, so a more realistic model takes imperfections into account with a term w_k, a stochastic value following a normal distribution with mean 0 and variance q:

$$x_{k+1} = x_k + d + w_k$$

If we try to measure the position, our measurement y_k is also corrupted by noise we can model with a term v_k, another stochastic value (independent of w_k) following a normal distribution with mean 0 and variance r:

$$y_k = x_k + v_k$$

What is the best approximation we can have for the real value x_k? Is it y_k? But then we would ignore the knowledge we have about the underlying system, a point moving at constant speed. A value obtained by linear regression using all our measurements? Two problems: first, we would have to wait until the end of the experiment before processing all of them; second, we still have w_k which needs to be handled in an optimal way.

Let's try another approach where we update recursively our knowledge for each new measurement. Assume that we know the position at the previous sample time $k-1$. We call it $\hat{x}_{k-1|k-1}$ to show it is an estimate of the position at time $k-1$ known at time $k-1$ (this will be important soon), not the real value x_{k-1}. We do not know it perfectly; we assume it has a normal distribution with mean $\hat{x}_{k-1|k-1}$ and variance $p_{k-1|k-1}$.

Our best guess for the current position, based on the previous position and the model, is

$$\hat{x}_{k|k-1} = \hat{x}_{k-1|k-1} + d$$

Of course, since w_k is stochastic, we do not know its actual value. But we can also guess the variance of $\hat{x}_{k|k-1}$: the variance of a sum of independent variables is the sum of the variances, hence

$$p_{k|k-1} = p_{k-1|k-1} + q$$

Let us compare the position guess (the technical name is *a priori estimate*, an estimated value based on prior knowledge) to the measure. The result is i_k, known as the *innovation*:

$$i_k = y_k - \hat{x}_{k|k-1} = x_k + v_k - \hat{x}_{k|k-1}$$

Its variance s_k is

$$s_k = p_{k|k-1} + r$$

($p_{k|k-1}$ and r are the variances of $\hat{x}_{k|k-1}$ and v_k, respectively, which are independent; and the variance of x_k is 0).

If the measurement differs from the a priori estimate, we should correct the estimate. The larger the difference, the larger the correction. Consider a correction factor K (the correction *gain*) which lets us improve the a priori estimate $\hat{x}_{k|k-1}$ and get an *a posteriori estimate* $\hat{x}_{k|k}$, an estimate which takes the measurement into account:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + Ki_k$$

How can we choose K? To have the best a posteriori estimate $\hat{x}_{k|k}$, we want to minimize its variance $p_{k|k}$ with respect to K, i.e. to have $\partial p_{k|k}/\partial K = 0$. We have

$$\begin{aligned}
p_{k|k} &= \operatorname{var}\left(x_k - \hat{x}_{k|k}\right) \\
&= \operatorname{var}\left(x_k - \hat{x}_{k|k-1} - Ki_k\right) \\
&= \operatorname{var}\left(x_k - \hat{x}_{k|k-1} - K(x_k + v_k - \hat{x}_{k|k-1})\right) \\
&= \operatorname{var}\left((1-K)(x_k - \hat{x}_{k|k-1}) - Kv_k\right)
\end{aligned}$$

The measurement noise v_k and the position noise w_k are uncorrelated, hence

$$p_{k|k} = \operatorname{var}(1-K)(x_k - \hat{x}_{k|k-1}) + \operatorname{var} Kv_k$$

Since $\operatorname{var}(x_k - \hat{x}_{k|k-1})$ is the a priori error variance $p_{k|k-1}$ and $\operatorname{var} v_k$ is the measurement variance r,

$$p_{k|k} = (K-1)^2 p_{k|k-1} + K^2 r$$

Its minimum with respect to K gives the K we want:

$$\frac{\partial}{\partial K}\left((K-1)^2 p_{k|k-1} + K^2 r\right) = 0$$

$$(K-1)p_{k|k-1} + Kr = 0$$

$$K = \frac{p_{k|k-1}}{p_{k|k-1} + r}$$

The a posteriori variance $p_{k|k}$ can be simplified:

$$\begin{aligned}
p_{k|k} &= (1-2K)p_{k|k-1} + K^2(r + p_{k|k-1}) \\
&= (1-K)p_{k|k-1}
\end{aligned}$$

To summarize, here is how to update the optimal estimate of the position $\hat{x}_{k|k}$ and its variance $p_{k|k}$ knowing these values at the previous step $k-1$, the model displacement d, the noise variances q and r, and the new position measurement y_k:

1. Compute the a priori state estimate

$$\hat{x}_{k|k-1} = \hat{x}_{k-1|k-1} + d$$

2. Compute the a priori error variance

$$p_{k|k-1} = p_{k-1|k-1} + q$$

3. Compute the innovation

$$i_k = y_k - \hat{x}_{k|k-1}$$

4. Compute the optimum correction gain

$$K = \frac{p_{k|k-1}}{p_{k|k-1} + r}$$

5. Compute the a posteriori state estimate

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + Ki_k$$

6. Compute the a posteriori error variance

$$p_{k|k} = (1 - K)p_{k|k-1}$$

This is a scalar version of the Kalman filter for the position estimation.

3 Kalman Filter

This section describes the Kalman filter for the general case of a multivariate linear discrete-time state-space model.

3.1 System model

Let a system described by the following discrete-time state-space model:

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{1}$$

$$y_k = Cx_k + v_k \tag{2}$$

where w_k is the process noise with covariance Q, and v_k the measurement noise with covariance R. The real state x_k is unknown.

3.2 Derivation

The purpose of the Kalman filter is to calculate an estimate $\hat{x}_k$, based on the system input u_k and the measurement system output y_k, as close as possible to x_k. The state estimate is assumed to have an uncertainty described by covariance matrix P_k.

Assuming we have an estimate $\hat{x}_{k-1|k-1}$ for time $k-1$ known at time $k-1$, (1) can be used to predict a value at time k, ignoring the unknown process noise w_{k-1}:

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \tag{3}$$

$\hat{x}_{k|k-1}$ is the *a priori state estimate*, i.e. an estimate based on past knowledge without taking into account the output measurement.

In a similar way, assuming we know the estimated error covariance $P_{k-1|k-1}$ at time $k-1$, (1) gives an estimate for the predicted error covariance (a priori error covariance) at time k:

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \tag{4}$$

The first term, $AP_{k-1|k-1}A^T$, comes from multiplying the state by A. The second term, Q, comes from the process noise w_k. The term Bu_k in (1) is known and does not add uncertainties, hence it has no effect on $P_{k|k-1}$.

The a priori estimates must be corrected with the measured output y_k to obtained the *a posteriori* state and error covariance estimates at time k, $\hat{x}_{k|k}$ and $P_{k|k}$. The reasons why the use of measurements is crucial is the same as for feedback control:

- Reject model uncertainties. If (or rather, since) the model used to compute the a priori estimates is different from the actual system, the estimate cannot converge to the real state.

- Take care of unstable systems. If the model is unstable (eigenvalues of A outside the unit circle), any error in the uncorrected a priori state estimate will diverge, even if u_k corresponds to a closed-loop control signal which stabilizes the real system.

- Reject the effect of initial conditions and process noise in an optimal way. Even if the system is stable, (3) has the same dynamics as the system, which can be slow or have resonance peaks.

To correct the a priori estimates, we consider the *innovation* i_k, which is the difference between the measurement and the estimated output. Using (2), the innovation is defined as

$$i_k = y_k - C\hat{x}_{k|k-1} \tag{5}$$

and its covariance S_k is

$$S_k = CP_{k|k-1}C^T + R \tag{6}$$

The innovation represents the estimate of the correction between the measurement and the output based on the a priori state estimate coming from the model. The idea of the Kalman filter is to use it in a linear way, by multiplying by matrix K (the *Kalman gain*), to correct the a priori state estimate and obtain the a posteriori estimate:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + Ki_k \tag{7}$$

The Kalman gain K is chosen to minimize the expected error between $\hat{x}_{k|k}$ and x_k, i.e. which minimizes the trace of the error covariance matrix $P_{k|k}$.

Developing $P_{k|k}$ (using (7), (5) and (2)) yields

$$\begin{aligned}
P_{k|k} &= \operatorname{cov}\left(x_k - \hat{x}_{k|k}\right) \\
&= \operatorname{cov}\left(x_k - \hat{x}_{k|k-1} - Ki_k\right) \\
&= \operatorname{cov}\left(x_k - \hat{x}_{k|k-1} - K(y_k - C\hat{x}_{k|k-1})\right) \\
&= \operatorname{cov}\left(x_k - \hat{x}_{k|k-1} - K(Cx_k + v_k - C\hat{x}_{k|k-1})\right) \\
&= \operatorname{cov}\left((I - KC)(x_k - \hat{x}_{k|k-1}) - Kv_k\right)
\end{aligned}$$

Assuming the output noise v_k and the system noise w_k are uncorrelated,

$$\begin{aligned}
&\operatorname{cov}\left((I - KC)(x_k - \hat{x}_{k|k-1}) - Kv_k\right) = \\
&\quad \operatorname{cov}\left((I - KC)(x_k - \hat{x}_{k|k-1})\right) + \operatorname{cov}\left(Kv_k\right)
\end{aligned}$$

$\operatorname{cov}(x_k - \hat{x}_{k|k-1})$ is the a priori error covariance $P_{k|k-1}$, and $\operatorname{cov} v_k$ is the output covariance R. Hence the a posteriori error covariance is

$$P_{k|k} = (I - KC)P_{k|k-1}(I - KC)^T + KRK^T \tag{8}$$

The optimum Kalman gain K minimizes the trace of $P_{k|k}$ (the sum of its diagonal elements), hence $\partial \operatorname{tr} P_{k|k}/\partial K = 0$. Expanding (8) and recognizing the innovation covariance (6) yields

$$\begin{aligned}
P_{k|k} &= P_{k|k-1} - KCP_{k|k-1} - P_{k|k-1}C^TK^T \\
&\quad + \underbrace{KCP_{k|k-1}C^TK^T + KRK^T}_{K(CP_{k|k-1}C^T+R)K^T=KSK^T} \\
&= P_{k|k-1} - KCP_{k|k-1} - P_{k|k-1}C^TK^T + KSK^T
\end{aligned} \tag{9}$$

Recalling that $\operatorname{tr} U^T = \operatorname{tr} U$, $\operatorname{tr}(U+V) = \operatorname{tr} U + \operatorname{tr} V$, $\partial \operatorname{tr} UV/\partial U = V^T$, and $\partial \operatorname{tr} UVU^T/\partial U = UV + UV^T$; and using the symmetry of P and S, we get

$$\begin{aligned}
\frac{\partial \operatorname{tr} P_{k|k}}{\partial K} &= \underbrace{\frac{\partial \operatorname{tr} P_{k|k-1}}{\partial K}}_{0} \\
&\quad - \underbrace{\frac{\partial \operatorname{tr} KCP_{k|k-1}}{\partial K}}_{(CP_{k|k-1})^T} - \underbrace{\frac{\partial \operatorname{tr} P_{k|k-1}C^TK^T}{\partial K}}_{(CP_{k|k-1})^T} \\
&\quad + \underbrace{\frac{\partial \operatorname{tr} KSK^T}{\partial K}}_{KS+KS^T} \\
&= -2P_{k|k-1}C^T + 2KS
\end{aligned}$$

Equating it with zero yields the optimum Kalman gain

$$K = P_{k|k-1}C^TS^{-1} \tag{10}$$

The a posteriori error covariance (9) can be simplified. From (10), $KS = P_{k|k-1}C^T$; substituting KS in (9) yields

$$P_{k|k} = (I - KC)P_{k|k-1} \tag{11}$$

3.3 Algorithm

To summarize, here is the algorithm to compute the best state estimate $\hat{x}_{k|k}$ (a posteriori estimate) and its associated covariance matrix $P_{k|k}$ with a Kalman filter, using the state estimate and covariance matrix at the previous step ($\hat{x}_{k-1|k-1}$ and $P_{k-1|k-1}$, respectively), and the model matrices A, B, C, Q, and R.

1. Compute the a priori state estimate

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1}$$

2. Compute the a priori error covariance

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q$$

3. Compute the innovation

$$i_k = y_k - C\hat{x}_{k|k-1}$$

4. Compute the innovation covariance

$$S_k = CP_{k|k-1}C^T + R$$

5. Compute the optimum Kalman gain

$$K_k = P_{k|k-1}C^T S_k^{-1}$$

6. Compute the a posteriori state estimate

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k i_k$$

7. Compute the a posteriori error covariance

$$P_{k|k} = (I - K_k C)P_{k|k-1}$$

3.4 Remarks

- If values for $x_{0|0}$ and $P_{0|0}$ are unknown, they can be initialized to zero (vector and square matrix of the correct size).

- In (7), the Kalman gain balances the part coming from the model and the part coming from the new output measurement. The larger the gain, the more important the effect of the measurement. To get a more rigorous understanding, subtituting (10), (6), and (5) into (7) yields

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}(y_k - C\hat{x}_{k|k-1})$$

For the covariance, substituting (10) and (6) into (8) yields

$$P_{k|k} = P_{k|k-1} - P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}CP_{k|k-1}$$

Assuming C is square and invertible, if $R \ll CP_{k|k-1}C^T$ (the output is much better known than the state estimate), $\hat{x}_{k|k} \approx C^{-1}y_k$ and $P_{k|k} \approx 0$: the state estimate comes only from the measured output y_k. If $R \gg CP_{k|k-1}C^T$ (the state estimate is much better known than the output), $\hat{x}_{k|k} \approx \hat{x}_{k|k-1}$ and $P_{k|k} \approx P_{k|k-1}$: the state estimate comes only from the model.

If C is not square and invertible, the system may not be observable, or observable but requiring past information to estimate the state with the a priori estimate; or with more outputs than states, the estimate results from the least-square solution of an overdetermined system.

- In the model, the term Bu_k in (1) suggests a "linear" input. But since u_k is not specified, Bu_k, as a whole, can represent any known input, be it a control signal in a feedback loop or a measured disturbance.
- The system has been assumed to be time-invariant, i.e. matrices A, B, C, Q, and R to be constant. The Kalman filter can also be used if they are changing slowly with respect to the system states and inputs.
- Since the contribution of the measurement using the innovation and the Kalman gain does not depend on previous values, it can be computed when it is available only, and even depend on which kind of measurement is available. At times k when no measurement is available, $x_{k|k} = x_{k|k-1}$ and $P_{k|k} = P_{k|k-1}$: the a posteriori estimate is the same as the a priori estimate. Depending on the measurement, C_k, i_k, and R_k can have a different meaning with different units and sizes.

3.5 Example

Consider the following second-order discrete-time system with a scalar measurement:

$$x_{k+1} = Ax_k + w_k$$
$$y_k = Cx_k + v_k$$

w_k is white noise with covariance Q, v_k is white noise with covariance R, and

$$A = \begin{bmatrix} 1 & -0.9 \\ 1 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad Q = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \quad R = \begin{bmatrix} 0.1 \end{bmatrix}$$

Measurements are as follows:

k	1	2	3	4	5	6	7	8
y_k	-0.1418	0.7094	0.8558	0.3455	-0.6060	-0.7966	-0.3689	0.2038

Starting with

$$\hat{x}_{0|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad P_{0|0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

we apply the Kalman filter steps of section 3.3. There is no system input term Bu_k.

1.1 Compute the a priori state estimate

$$\hat{x}_{1|0} = A\hat{x}_{0|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

1.2 Compute the a priori error covariance

$$P_{1|0} = AP_{0|0}A^T + Q = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$$

1.3 Compute the innovation

$$i_1 = y_1 - C\hat{x}_{1|0} = -0.1418$$

1.4 Compute the innovation covariance

$$S_1 = CP_{1|0}C^T + R = 0.2$$

1.5 Compute the optimum Kalman gain

$$K_1 = P_{1|0}C^T S_1^{-1} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$$

1.6 Compute the a posteriori state estimate

$$\hat{x}_{1|1} = \hat{x}_{1|0} + K_1 i_1 = \begin{bmatrix} -0.0709 \\ 0 \end{bmatrix}$$

1.7 Compute the a posteriori error covariance

$$P_{1|1} = (I - K_1 C) P_{1|0} = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.1 \end{bmatrix}$$

2.1 Compute the next a priori state estimate

$$\hat{x}_{2|1} = A\hat{x}_{1|1} = \begin{bmatrix} -0.709 \\ -0.709 \end{bmatrix}$$

...and so on.

3.6 State disturbance covariance Q

In (1), w_k stands for all the contributions from signals which do not come from the state feedback term Ax_k or the known exogenous inputs Bu_k: disturbances, model structure mismatch (including approximations coming from linearization), model parameter errors. In a model built from first principles, disturbances can be evaluated, and independent sources identified. For each disturbance source, we consider an independent white noise signal; all independent noise signals are gathered into a vector w_k^*, and they impact the model states through a matrix E:

$$x_{k+1} = Ax_k + Bu_k + Ew_k^* \tag{12}$$

Hence $w_k = Ew_k^*$, w_k^* has unit covariance (its covariance matrix is the identity matrix), and

$$Q = EE^T$$

E and A play the same role for disturbances as B and A for the known inputs. Depending on the way a source of white noise is filtered before impacting the system (i.e. a periodic disturbance coming from an imperfect wheel), the model order (number of states) should be extended. The matrix E is typically rectangular: the size of w_k^* can be smaller than w_k (for example vibrations in the frame of a machine which are transmitted differently to each moving part; or the wind speed which changes the lift and drag forces on each blade of the rotor of a helicopter). E should not be degenerate down to a matrix with 0 column ($Q = 0$), though; otherwise the a priori estimate would be exact (covariance $P_{k+1|k}$ shrinking to 0 for large k) and ultimately the measurements would have no effect on the a posteriori estimate.

3.7 Measurement noise covariance R

The measurement noise covariance matrix R describes the quality of measurements, and impact the weight measurements have on the Kalman state estimate. For a scalar sensor, the meaning of the variance is clear. Independent scalar sensors correspond to a diagonal R. A single sensor producing multiple values cannot always be considered as separate scalar sensors: for instance a 3D accelerometer can produce measurements which are not aligned correctly, with some cross-correlation. Combining multiple multi-value sensors results in a block-diagonal R.

3.8 Example

Let a mass m with position $x(t)$ and velocity $v(t) = \dot{x}(t)$, subject to a force $f(t)$. Its acceleration is $a(t) = \dot{v}(t) = f(t)/m$. $f(t)$ is an unknown perturbation whose frequency spectrum is flat between 0 and ω_0, with magnitude F_0, and decreases at $-20\,\mathrm{dB/dec}$ beyond. The sampling angular frequency is assumed to be much larger than ω_0. We measure the position with a sensor whose variance is r and want to estimate the position and the velocity.

Solution

The perturbation is modeled as white noise filtered by first-order transfer function $F_0/(1+s/\omega_0)$. In discrete time, the filter is approximated by substituting $s \to (z-1)/T_s$, where T_s is the sampling period; hence the discrete-time filter is $F(z)/W^*(z) = F_0 T_s \omega_0/(z + T_s\omega_0 - 1)$, where $F(z)$ is the z transform of the force and $W^*(z)$ is the z transform of the white noise source. The corresponding equation is $f_k = (1 - T_s\omega_0) f_{k-1} + F_0 T_s \omega_0 w^*_{k-1}$.

For the state-space discrete-time model, we choose the states x_k (position at sample k), v_k (velocity at sample k), and a_k (acceleration at sample k). The first-order filter of the perturbation model utilizes a_k as its state, since $f_k = m a_k$. Approximating integration with the Euler method ($y(t+T_s) \approx y(t) + T_s \mathrm{d}y(t)/\mathrm{d}t$), the system model is

$$\begin{bmatrix} x_{k+1} \\ v_{k+1} \\ a_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & T_s & 0 \\ 0 & 1 & T_s \\ 0 & 0 & 1 - T_s\omega_0 \end{bmatrix} \cdot \begin{bmatrix} x_k \\ v_k \\ a_k \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ F_0 T_s \omega_0/m \end{bmatrix} w^*_k$$

Hence

$$E = \begin{bmatrix} 0 \\ 0 \\ F_0 T_s \omega_0/m \end{bmatrix} \qquad Q = EE^T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & F_0^2 T_s^2 \omega_0^2/m^2 \end{bmatrix}$$

The measurement covariance matrix is a scalar:

$$R = r$$

4 Extended Kalman Filter

The Extended Kalman Filter is a Kalman filter applied to a non-linear system linearized at each discrete time k. The state-space discrete-time model is

$$x_{k+1} = f(x_k, u_k, k) + w_k$$

$$y_k = g(x_k, k) + v_k$$

where w_k and v_k are the process noise with covariance Q_k and the measurement noise with covariance R_k, respectively.

The EKF formulation is very similar to the plain Kalman filter where A and C are the jacobians of f and g, respectively:

$$A_k = \frac{\partial f(x, u_k, k)}{\partial x}$$

$$C_k = \frac{\partial g(x, k)}{\partial x}$$

The complete EKF algorithm follows.

1. Compute the jacobians of the state-space functions

$$A_k = \frac{\partial f(x, u_k, k)}{\partial x} \qquad C_k = \frac{\partial g(x, k)}{\partial x}$$

2. Compute the a priori state estimate

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_{k-1}, k)$$

3. Compute the a priori error covariance

$$P_{k|k-1} = A_k P_{k-1|k-1} A_k^T + Q$$

4. Compute the innovation

$$i_k = y_k - g(\hat{x}_{k|k-1}, k)$$

5. Compute the innovation covariance

$$S_k = C_k P_{k|k-1} C_k^T + R$$

6. Compute the optimum Kalman gain

$$K = P_{k|k-1} C_k^T S^{-1}$$

7. Compute the a posteriori state estimate

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K i_k$$

8. Compute the a posteriori error covariance

$$P_{k|k} = (I - K C_k) P_{k|k-1}$$