# Paper Summary
# Computer Integrated Surgery II

Alessandro Luca Asoni

April 1, 2014

This is a summary of the paper **An Introduction to the Kalman Filter** by Welch and Bishop.

## The Kalman Filter

The Kalman Filter is a recursive algorithm that provides an efficient computational means to estimate the state of a process governed by the following two equations:

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \tag{1}$$
$$z_k = Hx_k + v_k \tag{2}$$

Where $(1)$ is a stochastic difference equation that encodes the system's dynamics. The dynamics are assumed to be linearly dependent on the state of the system at the previous time step $x_{k-1} \in \mathfrak{R}^n$ and on some driving input on the system $u_{k-1}$. The dynamics are also subject to *process* noise that is assumed to be distributed normally:

$$p(w) \sim N(0, Q) \tag{3}$$

And $(2)$ is the measurement equation that relates some sensor reading $z \in \mathfrak{R}^n$ to the state vector $x_k$ through a linear transformation $H$. One of the requirements of the Kalman Filter is that this transformation be invertible, in other words $H^{-1}$ must exist.

### Mathematical Derivation

The Kalman Filter is a two step process:

1) First an *a priori* estimate of the state $\hat{x}_k^-$ and its covariance $P_k^-$ is computed by using equation $(1)$ to propagate the previous state and covariance.

2) Then new *a posteriori* estimates $\hat{x}_k, P_k$ are computed using sensor information.

Notice that from this two estimates we have the following two errors:

1) *a priori* error $e_k^- = x_k - \hat{x}_k^-$

2) *a posteriori* error $e_k = x_k - \hat{x}_k$

Our goal is then to minimize the *a posteriori* error covariance

$$\begin{aligned} P_k &= E\left[e_k e_k^\top\right] \\ &= E\left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^\top\right] \end{aligned}$$

More precisely we want to minimize the trace of $P_k$ which corresponds to the sum of the variances of each entry in the state vector.

To determine what the optimal value of $\hat{x}_k$ should be, we start from a linear blending of the noisy measurement and the *a priori* estimate:

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \tag{4}$$

This equation is setting the *a posteriori* estimate to be the *a priori* estimate plus the difference between the actaual measurement $z_k$ and a measurement prediction $H\hat{x}_k$ weighted by a gain $K$. This difference is called the $residual$.

For a justification of why we pick this equation for $\hat{x}_k$ see Appendix II. We can plug this value into the expression for $P_k$ and then to minimize the value of trace$[P_k]$ we take the derivative with respect to $K$ and set to zero. Solving for $K$ then gives the following expressions for $K$ and $P_k$:

$$K = P_k^- H^\top (H P_k^- H^\top + R_k)^{-1} \tag{5}$$

$$P_k = (I - KH)P_k^- \tag{6}$$

For a derivation of this result see Appendix I.

Now to see that this behaves in an intuitively correct way, we check what happens in the limit when the measurement noise covariance approaches zero and when the *a priori* state covariance approaches zero.

$$\begin{aligned} \lim_{R_k \to 0} K &= H^{-1} \\ \hat{x}_k &= H^{-1} z_k \end{aligned}$$

Hence if the measurement noise covariance goes to zero we trust the measurements more and more. This is the intuitive behaviour that we expect.

$$\begin{aligned} \lim_{P_k^- \to 0} K &= 0 \\ \hat{x}_k &= \hat{x}_k^- \end{aligned}$$

So the *a priori* estimate is trusted more and more when the state *a priori* covariance approaches zero. This also is in line with our intuition of how the system should behave.

### The Filter in action

#### Time Update

1) Project the state ahead

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

2) Project the error covariance ahead

$$P_k^- = AP_{k-1}A^\top + Q$$

#### Measurement Update

1. Compute the Kalman gain

$$K_k = P_k^- H^\top (HP_k^- H^\top + R)^{-1}$$

2. Update estimate with measurement $z_k$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$$

3. Update error covariance

$$P_k = (I - K_k)P_k^-$$

These two update steps are performed iteratively for each time step.
It is important to note that for this discussion the matrices $H, A, R, Q$ were considered constant. This need not be the case, and in practice it is common for them to vary with time. The only change in the above equations would be the addition of a subscript $_k$ to these matrices.

## The Extended Kalman Filter

The Extended Kalman Filter is designed to deal with a similar system as presented earlier, but without the linearity constraint. In other words the system dynamics and measurment equations are:

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1})$$
$$z_k = h(x_k, v, k)$$

There is conceptually very little difference between the Extended Kalman Filter and the standard Kalman Filter. All that is required is to linearize the functions $f$ and $h$. This can be done by computing the appropriate Jacobian matrices:

- $A$: Jacobian of $f$ with respect of $x$

- $W$: Jacobian of $f$ with respect of $w$

- $H$: Jacobian of $h$ with respect of $x$

- $V$: Jacobian of $h$ with respect of $v$

This gives the following equations for the *prediction* and *update* steps:

**Time Update**

1) Project the state ahead

$$\hat{x}_k^- = f(\hat{x}_{k-1}, u_{k-1}, 0)$$

2) Project the error covariance ahead

$$P_k^- = AP_{k-1}A^\top + Q + WQW^\top$$

**Measurement Update**

1. Compute the Kalman gain

$$K_k = P_k^- H^\top (HP_k^- H^\top + R)^{-1}$$

2. Update estimate with measurement $z_k$

$$\hat{x}_k = \hat{x}_k^- + K_k \left(z_k - h(\hat{x}_k^-, 0)\right)$$

3. Update error covariance

$$P_k = (I - K_k)P_k^-$$

## Appendix I - Derivation of the Kalman Gain Expression

This derivation isn't done in the original paper. It is included in this summary for completeness. Recall that we want to minimize the trace of $P_k$:

$$\begin{aligned} P_k &= E\left[e_k e_k^\top\right] \\ &= E\left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^\top\right] \end{aligned}$$

And the following expression was given for $\hat{x}_k$

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-)$$

We can plug this back into the expression for $P_k$ to obtain:

$$= E\left[\left(x_k - \hat{x}_k^- + K(z_k - H\hat{x}_k^-)\right)\left(x_k - \hat{x}_k^- + K(z_k - H\hat{x}_k^-)\right)^\top\right]$$

We notice that $x_k - \hat{x}_k^- = e_k^-$. We also substitute the measurement equation for $z_k$:

$$\begin{aligned} &= E\left[\left(e_k^- + K(Hx_k + v_k - H\hat{x}_k^-)\right)\left(e_k^- + K(Hx_k + v_k - H\hat{x}_k^-)\right)^\top\right] \\ &= E\left[\left(e_k^- + K(He_k^- + v_k)\right)\left(e_k^- + K(He_k^- + v_k)\right)^\top\right] \\ &= E\left[\left((I - KH)e_k^- + Kv_k\right)\left((I - KH)e_k^- + Kv_k\right)^\top\right] \end{aligned}$$

Notice now that some cross terms cancel because $E[e_k v_k^\top] = 0$ since they are assumed to be independent.

$$\begin{aligned} &= E\left[(I - KH)e_k^- e_k^{-\top}(I - KH)^\top + Kv_k v_k^\top K^\top\right] \\ &= (I - KH)P_k^-(I - KH)^\top + KRK^\top \end{aligned}$$

this can be rewritten as:

$$P_k = P_k^- - KHP_k^- - P_K^- H^\top K^\top + K(HP_k^- H^\top + R)K^\top$$

We want to take the derivative of the trace of this expression with respect to $K$. In order to do this, the following matrix differentiation formulas are needed:

$$\begin{aligned} \frac{d[\text{trace}(\mathbf{AB})]}{d\mathbf{A}} &= \mathbf{B}^\top \\ \frac{d[\text{trace}(\mathbf{ACA}^\top)]}{d\mathbf{A}} &= 2\mathbf{AC} \end{aligned}$$

Using this formulas (and noticing that the trace of $KHP_k^-$ is equal to the trace of $P_K^- H^\top K^\top$) we obtain:

$$\frac{d[\text{trace}P_k]}{dK} = -2(HP_k^-)^\top + 2K(HP_k^- H^\top + R)$$

Setting this expression equal to zero and solving for $K$ we obtain:

$$K = P_k^- H^\top (H P_k^- H^\top + R)^{-1}$$

Plugging this value back into the expression for $P_k$ we obtain:

$$P_k = (I - KH) P_k^-$$

## Appendix II - The Conditional Density Viewpoint

This section aims to justify the initial choice for $\hat{x}_k$:

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \tag{7}$$

This is also not present in the paper, but is included here for completeness. We will look at the filter from a conditional density viewpoint. The discussion will not always be rigorous, since the goal is to give a general idea of this alternate approach and how this leads to the above expression for $\hat{x}_k$.

Assume that at some time step $k$ we have by some means an optimal estimate $\hat{x}_k^-$ and its associated covariance $P_k^-$. The probability density of $x_k$ is:

$$f_{x_k} \sim N(\hat{x}_k^-, P_k^-)$$

Recall the measurement equation:

$$z_k = H\hat{x}_k^- + v_k$$

Since $z_k$ is the sum of two Gaussian distributed random variables its distribution can easily be seen to be:

$$f_{z_k} \sim N(H\hat{x}_k^-,\ HP_k^- H^\top + R)$$

Now say that we know $x_k$, meaning that $x_k$ is constant. The density of $z_k$ conditioned on $x_k$ is:

$$f_{z_k|x_k} \sim N(Hx_k, R)$$

In order to obtain the density of the probability of $x_k$ given the measurment $z_k$ we can apply Bayes formula:

$$\begin{aligned} f_{x_k|z_k} &= \frac{f_{z_k|x_k} f_{x_k}}{f_{z_k}} \\ &\sim \frac{[N(Hx_k, R)][N(\hat{x}_k^-, P_k^-)]}{[N(H\hat{x}_k^-,\ HP_k^- H^\top + R)]} \end{aligned}$$

Notice that this is the distribution that we are after for the optimal update of $x_k$ and $P_k$. Since it is the optimal value of $x_k$ given that we have seen the last measurement $z_k$. Expanding the product of Gaussians is laborious and so it is not done here. The resulting distribution has the following parameters:

$$\begin{aligned} \text{Mean} &= \hat{x}_k^- + P_k^- H^\top (HP_k^- H^\top + R)^{-1}(z_k - H\hat{x}_k^-) \\ \text{Covariance} &= [(P_k^-)^{-1} + H^\top R H]^{-1} \end{aligned}$$

Notice that the mean corresponds in fact to the earlier result that was found. And it turns out that this expression for the covariance $P_k$ is identical to the one presented in the previous section.
The key here is to see that the mean does in fact have the form:

$$\hat{x}_k^- + K(z_k - H\hat{x}_k^-)$$

which hopefully justifies this initial choice. Note that this derivation of the Kalman filter is less intuitive then the one that was presented in this summary, which is why the derivation was started by directly assuming the linear blending as the correct value for $\hat{x}_k$.

## References

[1] Welch, G., and G. Bishop (1995), An introduction to the Kalman Filter. Technical Report TR 95-041, University of North Carolina, Department of Computer Science

[2] Brown R.G. and P. Y. C. Hwang (1922). *Introduction to Random Signals and Applied Kalman Filtering, Second Edition*, John Wiley & Sons, Inc.