THE KALMAN FILTER

RAUL ROJAS

ABSTRACT. This paper provides a gentle introduction to the Kalman filter, a numerical method that can be used for sensor fusion or for calculation of trajectories. First, we consider the Kalman filter for a one-dimensional system. The main idea is that the Kalman filter is simply a linear weighted average of two sensor values. Then, we show that the general case has a similar structure and that the mathematical formulation is quite similar.

1. AN EXAMPLE OF DATA FILTERING

The Kalman filter is widely used in aeronautics and engineering for two main purposes: for combining measurements of the same variables but from different sensors, and for combining an inexact forecast of a system's state with an inexact measurement of the state. The Kalman filter has also applications in statistics and function approximation.

When dealing with a time series of data points $x_1, x_2, \ldots, x_n$, a *forecaster* computes the best guess for the point x_{n+1}. A *smoother* looks back at the data, and computes the best possible x_i taking into account the points before and after x_i. A *filter* provides a correction for x_{n+1} taking into account all the points $x_1, x_2, \ldots, x_n$ and an inexact measurement of x_{n+1}.

An example of a filter is the following: Assume that we have a system whose one-dimensional state we can measure at successive steps. The readings of the measurements are $x_1, x_2, \ldots, x_n$. Our task is to compute the average μ_n of the time series given n points. The solution is

$$\mu_n = \frac{1}{n} \sum_1^n x_i.$$

If a new point x_{n+1} is measured, we can recompute μ_n, but it is more efficient to use the old value of μ_n and make a small correction using x_{n+1}. The correction is easy to derive, since

$$\mu_{n+1} = \frac{1}{n+1} \sum_1^{n+1} x_i = \frac{n}{n+1} \left(\frac{1}{n} \sum_1^n x_i + \frac{1}{n} x_{n+1} \right)$$

and so, μ_{n+1} can be written as

$$(1.1) \qquad \mu_{n+1} = \frac{n}{n+1} \mu + \frac{1}{n+1} x_{n+1} = \mu + K(x_{n+1} - \mu)$$

where $K = 1/(n+1)$. K is called the gain factor. The new average μ_{n+1} is a weighted average of the old estimate μ_n and the new value x_{n+1}. We trust μ_n more than the single value x_{n+1}; therefore, the weight of μ_n is larger than the weight of x_{n+1}. Equation 1.1 can be also read as stating that the average is corrected

using the difference of x_{n+1} and the old value μ_n. The gain K adjusts how big the correction will be.

We can also recalculate recursively the quadratic standard deviation of the time series (the variance). Given n points, the quadratic standard deviation σ_n is given by:

$$\sigma_n^2 = \frac{1}{n} \sum_1^n (x_i - \mu)^2.$$

If a new point x_{n+1} is measured, the new variance is

$$\sigma_{n+1}^2 = \frac{1}{n+1} \sum_1^{n+1} (x_i - \mu_{n+1})^2 = \frac{1}{n+1} \sum_1^{n+1} (x_i - \mu_n - K(x_{n+1} - \mu_n)^2)$$

where K is the gain factor defined above and we used Equation 1.1. The expression above can be expanded as follows:

$$\sigma_{n+1}^2 = \frac{1}{n+1} \left[\sum_1^n (x_i - \mu_n)^2 + 2K \sum_1^n (x_i - \mu_n)(x_{n+1} - \mu_n) + nK^2(x_{n+1} - \mu_n)^2 + (1-K)^2(x_{n+1} - \mu_n)^2 \right]$$

The second term inside the brackets is zero, because $\sum_1^n (x_i - \mu_n) = 0$. Therefore, the whole expression reduces to

$$\sigma_{n+1}^2 = \frac{1}{n+1} (n\sigma_n^2 + (x_{n+1} - \mu)^2 (nK^2 + (1-K)^2).$$

Since $nK^2 + (1-K)^2 = nK$ the last expression reduces to

$$\sigma_{n+1}^2 = \frac{n}{n+1} (\sigma_n^2 + K(x_{n+1} - \mu)^2) = (1-K)(\sigma_n^2 + K(x_{n+1} - \mu)^2).$$

The whole process can now be casted into as a series of steps to be followed iteratively. Given the first n points, and our calculation of μ_n and σ_n, then

- When a new point x_{n+1} is measured, we compute the gain factor $K = 1/(n+1)$.
- We compute the new estimation of the average

$$\mu_{n+1} = \mu_n + K(x_{n+1} - \mu)$$

- We compute also a provisional estimate of the new standard deviation

$$\sigma_n'^2 = \sigma_n^2 + K(x_{n+1} - \mu)^2$$

- Finally, we find the correct σ_{n+1} using the correction

$$\sigma_{n+1}^2 = (1-K)\sigma_n'^2$$

This kind of iterative computation is used in calculators for updating the average and standard deviation of numbers entered sequentially into the calculator without having to store all numbers. This kind of iterative procedure shows the general flavor of the Kalman filter, which is a kind of recursive least squares estimator for data points.

FIGURE 1. Gaussians

2. THE ONE-DIMENSIONAL KALMAN FILTER

The example above showed how to update a statistical quantity once more information becomes available. Assume now that we are dealing with two different instruments that provide a reading for some quantity of interest x. We call x_1 the reading from the first instrument and x_2 the reading from the second instrument. We know that the first instrument has an error modelled by a Gaussian with standard deviation σ_1. The error of the second instrument is also normally distributed around zero with standard deviation σ_2. We would like to combine both readings into a single estimation.

If both instruments are equally good ($\sigma_1 = \sigma_2$), we just take the average of both numbers. If the first instrument is absolutely superior ($\sigma_1 << \sigma_2$), we will keep x_1 as our estimate, and vice versa if the second instrument is clearly superior to the first. In any other case we would like to form a weighted average of both readings to generate an estimate of x which we call $\hat{x}$.

The question now is which is the best weighted average. One possibility is weighting each reading inversely proportional to its precision, that is,

$$\hat{x} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

or simplifying

$$\hat{x} = \frac{x_1 \sigma_2^2 + x_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

This estimation of $\hat{x}$ fulfills the boundary conditions mentioned above.

Note that the above estimate can be also rewritten as

$$\hat{x} = x_1 + K(x_2 - x_1)$$

where now the gain $K = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. The update equation has the same general form as in the example in Section 1.

The expression used above is optimal given our state of knowledge about the measurements. Since the error curve from the instruments is a Gaussian, we can write the probability of x being the right measurement as

$$p_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}(x-x_1)^2/\sigma_1^2}$$

for instrument 1, and as

$$p_2(x) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2}(x-x_2)^2/\sigma_2^2}$$

for instrument 2. In the first case, x_1 is the most probable measurement, in the second, x_2, but all points x have a non-vanishing probability of being the right measurement due to the instruments' errors.

Since the two measurements are independent we can combine them best, by multiplying their probability distributions and normalizing. Multiplying we obtain:

$$p(x) = p_1(x)p_2(x) = Ce^{-\frac{1}{2}(x-x_1)^2/\sigma_1^2 - \frac{1}{2}(x-x_2)^2/\sigma_2^2}$$

where C is a constant obtained after the multiplication (including the normalization factor needed for the new Gaussian).

The expression for $p(x)$ can be expanded into

$$p(x) = Ce^{-\frac{1}{2}((\frac{x^2}{\sigma_1^2} - \frac{2xx_1}{\sigma_1^2} + \frac{x_1^2}{\sigma_1^2}) + (\frac{x^2}{\sigma_2^2} - \frac{2xx_2}{\sigma_2^2} + \frac{x_2^2}{\sigma_2^2}))}$$

which grouping some terms reduces to

$$p(x) = Ce^{-\frac{1}{2}(x^2(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_1^2}) - 2x(\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2})) + D}$$

where D is a constant. The expression can be rewritten as

$$p(x) = Ce^{-\frac{1}{2}(\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2}(x^2 - 2x(\frac{x_1\sigma_2^2 + x_2\sigma_1^2}{\sigma_1^2 + \sigma_2^2}))) + D}.$$

Completing the square in the exponent we obtain:

$$p(x) = Fe^{-\frac{1}{2}\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \sigma_2^2}\left[x - \frac{(x_1\sigma_2^2 + x_2\sigma_1^2)}{\sigma_1^2 + \sigma_2^2}\right]^2}$$

where all constants in the exponent (also those arising from completing the square) and in front of the exponential function have been absorbed into the constant F.

From this result, we see that the most probable result $\hat{x}$ obtained from combining the two measurements (that is, the center of the distribution) is

$$\hat{x} = \frac{(x_1\sigma_2^2 + x_2\sigma_1^2)}{\sigma_1^2 + \sigma_2^2}$$

and the variance of the combined result is

$$\hat{\sigma}^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

If we introduce a gain factor $K = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$ we can rewrite the best estimate $\hat{x}$ of the state as

$$\hat{x} = x_1 + K(x_2 - x_1)$$

and the change to the variance σ_1^2 as

$$\hat{\sigma} = (1 - K)\sigma_1^2$$

This is the general form of the classical Kalman filter.

Note that x_1 does not need to be a measurement. It can be a forecast of the system state, with a variance σ_1^2, and x_2 can be a measurement with the error variance σ_2^2. The Kalman filter would in that case combine the forecast with the measurement in order to provide the best possible linear combination of both as the final estimate.

3. Multidimensional Kalman Filter - uncorrelated dimensions

We can generalize the result obtained above to the case of multiple dimensions, when every dimension is independent of each other, that is, when there is no correlation between the measurements obtained for each dimension. In that case the measurements x_1 and x_2 are vectors, and we can combine their coordinates independently.

In the case of uncorrelated dimensions, given measurements x_1 and x_2 (vectors of dimension n), the variances for each dimension can be arranged in the diagonal of two matrices Σ_1 and Σ_2, as follows:

$$\Sigma_1 = \begin{bmatrix} \sigma_{11}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{12}^2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_{1n}^2 \end{bmatrix}$$

and

$$\Sigma_2 = \begin{bmatrix} \sigma_{21}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{22}^2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_{2n}^2 \end{bmatrix}$$

where σ_{1i}^2 represents the variance of the i-th coordinate for the first instrument, and σ_{2i}^2 the variance of the i-th coordinate for the second.

The best estimate for the system state is given by

$$\hat{x} = (\Sigma_2 x_1 + \Sigma_1 x_2)(\Sigma_1 + \Sigma_2)^{-1}$$

Introducing the gain $K = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$, we can rewrite the expression above as

$$\hat{x} = x_1 + K(x_2 - x_1)$$

and the new estimate of all variances can be put in the diagonal of a matrix $\hat{\Sigma}$ computed as

$$\hat{\Sigma} = (I - K)\Sigma_1$$

where I is the $n \times n$ identity matrix. The reader can verify all these expressions.

The matrices Σ_1 and Σ_2 are the covariance matrices for the two instruments. In the special case of uncorrelated dimensions, the covariance matrix is diagonal. Remember that the covariance matrix Σ for m n-dimensional data points $x_1, x_2, \ldots, x_m$ is defined as

$$\Sigma = \frac{1}{n}(x_1 x_1^T + x_2 x_2^T + \cdots + x_m x_m^T).$$

4. The general case

The general case is very similar to the one-dimensional case. The mathematics is almost identical, but now we have to operate with vectors and matrices.

Given two n-dimensional measurements x_1 and x_2, we assume that the two instruments that produced them have a normal error distribution. The first measurement is centered at x_1 with covariance matrix Σ_1. The second is centered at x_2 with covariance matrix Σ_2. The first distribution is denoted by $N(x_1, \Sigma_1)$, the second by $N(x_2, \Sigma_2)$

If the two instruments are independent, the combined distribution is the product of the two distributions. The product of two normalized Gaussians, after normalization, is another normalized Gaussian. We only have to add the exponents of the two Gaussians in order to find the exponent of the new Gaussian. Any constants in the exponent can be moved out of the exponent: they can be converted to constants in front of the exponential function, which will be later normalized. The new distribution is $N(\hat{x}, \hat{\Sigma})$.

Remember that in a Gaussian with center μ and covariance matrix Σ, the exponent of the Gaussian is $-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)$.

The sum of the exponents of the two Gaussians is:

$$-\frac{1}{2}\left[(x - x_1)^T \Sigma_1^{-1}(x - x_1) + (x - x_2)^T \Sigma_2^{-1}(x - x_2)\right]$$

We will drop the factor $-1/2$ in the intermediate steps and will recover it at the end. The expression above (without considering the factor $-1/2$) can be expanded as:

$$x^T(\Sigma_1^{-1} + \Sigma_2^{-1})x - 2(x_1^T \Sigma_1^{-1} + x_2^T \Sigma_2^{-1})x + C$$

where C is a constant. Since the product of symmetric matrices is commutative and the two covariance matrices Σ_1 and Σ_2 are commutative, we can transform the expression above into:

$$x^T(\Sigma_1 + \Sigma_2)(\Sigma_1\Sigma_2)^{-1}x - 2(x_1^T\Sigma_2 + x_2^T\Sigma_1)(\Sigma_1\Sigma_2)^{-1}(\Sigma_1 + \Sigma_2)^{-1}(\Sigma_1 + \Sigma_2)x + C$$

Note that we used the fact that $(\Sigma_1^{-1} + \Sigma_2^{-1})\Sigma_1\Sigma_2 = \Sigma_1 + \Sigma_2$. Now we can complete the square in the exponent and rewrite this expression as:

$$(x - (x_1\Sigma_2 + x_2\Sigma_1)(\Sigma_1 + \Sigma_2)^{-1})^T(\Sigma_1 + \Sigma_2)(\Sigma_1\Sigma_2)^{-1}(x - (x_1\Sigma_2 + x_2\Sigma_1)(\Sigma_1 + \Sigma_2)^{-1}) + C'$$

where C' is a constant absorbing C and any other constants needed for the square completion. Multiplying the above expression by $-1/2$, the exponent has the required form for a Gaussian, and by direct inspection we see that the expected average of the combination of measurements is

$$\hat{x} = (x_1\Sigma_2 + x_2\Sigma_1)(\Sigma_1 + \Sigma_2)^{-1}$$

and the new covariance matrix is

$$\hat{\Sigma} = (\Sigma_1 + \Sigma_2)^{-1}(\Sigma_1\Sigma_2)$$

These expressions correspond to the Kalman filter for the combination of two measurements. To see this more clearly we can define the Kalman gain K as

$$K = \Sigma_1(\Sigma_1 + \Sigma_2)^{-1}$$

Then the expressions above can be rewritten as

$$\hat{x} = x_1 + K(x_2 - x_1)$$

and

$$\hat{\Sigma} = (I - K)\Sigma_1$$

where I is the identity matrix. Notice that in all the algebraic manipulations above we used the fact that the product and the sum of symmetric matrices is symmetric, and that the inverse of a symmetric matrix is symmetric. Symmetric matrices commute under matrix multiplication.

5. Forecast and measurement of different dimension

The equations printed in books for the Kalman filter are more general than the expressions we have derived, because they handle the more general case in which the measurement can have a different dimension than the system's state. But it is easy to derive the more general form, all is needed is to "move" all calculations to measurement space.

In the general case we have a process going through state transitions $x_1, x_2, \ldots$. The propagator matrix A allows us to forecast the state at time $k + 1$ given our best past estimate $\hat{x}_k$ of the state at time k:

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

The forecast is error prone. The error is given by ν, a variable with Gaussian distribution centered at 0. If the covariance matrix of the system state estimate at time k is called P, the covariance matrix of $\hat{x}_{k+1}^-$ is given by

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

where Q is the covariance matrix of the noise ν.

The measurement z_{k+1} is also error prone. The measurement is linearly related to the state, namely as follows:

$$z_k = Hx_k + R$$

where H is a matrix and R the covariance matrix of the measurement error.

Kalman gain:

$$K_{k+1} = P_{k+1}^-H^T[HP_{k+1}^-H^T + R]^{-1}$$

State update:

$$\hat{x_{k+1}} = \hat{x}_{k+1}^- + K_{k+1}[z_{k+1} - H\hat{x}_{k+1}^-]$$

Covariance update:

$$P_{k+1} = (I - K_{k+1}H)P_{k+1}^-$$

References

1. Peter Maybeck, *Stochastic Models, Estimationd and Control*, Academic Press. Inc, 1979.
2. N.D. Singpurwalla and R.J. Meinhold, "Understanding the Kalman Filter", *The American Statistician*, Vol. 37, N. 2, pp. 123-7.
3. Greg Welch and Gary Bishop, *An Introduction to the Kalman Filter*, TR-95-041, Department of Computer Science, University of North Carolina at Chapell Hill, updated in 2002.

Freie Universität Berlin, Institut für Informatik