Kalman Filtering Notes

Portions of these notes are adapted from [3], [5], [4], [2], and [1].

What is the Kalman Filter?

Optimal recursive data processing algorithm for processing series of measurements generated from a linear dynamic system.

Define $x_t \in \mathbb{R}^n$ to be (unobserved) state of dynamic system at time t, $t \in \{0, 1, \ldots, T\}$. Define $z_t \in \mathbb{R}^m$ to be (observed) measurement at time t.

Kalman filter is an algorithm for determining $P(x_t|z_{0:t})$, given some particular assumptions about these random variables.

What is it used for?

Applications of the Kalman filter:

- Radar tracking of planes/missles (classical application)
- Tracking heads/hands/people from video data
- Economics (stock market data)
- Navigation

Simple example

Let's say that Greg and Huyen are hungry and lost in downtown Vancouver, and are trying to get to Fatburger, located at x_f. Greg thinks that Fatburger is located at z_g, and his estimates of any location x come from a Gaussian distribution $N(x, \sigma_g^2)$ (for a small σ_g). Huyen thinks that Fatburger is located at z_h, and her estimates come from a Gaussian distribution $N(x, \sigma_h^2)$ (no comment on the relative size of σ_h).

Given these two measurements (let's say z_h and z_g are scalars to make things easier), how do we combine them to get an estimate of the location of Fatburger?

$$
\begin{aligned}
P(x_f|z_g, z_h) &= P(z_g, z_h|x_f)\frac{P(x_f)}{P(z_g, z_h)} \text{ (by Bayes' rule)} && (1)\\
&= \alpha P(z_g, z_h|x_f)P(x_f) && (2)\\
&= \alpha P(z_g|x_f)P(z_h|x_f)P(x_f) \text{ (assuming conditional ind.)} && (3)\\
& && (4)
\end{aligned}
$$

As another simplification, let's assume that the prior $P(x_f)$ is uniform (a frequentist sort of assumption).

Then,

$$\begin{aligned}
P(x_f|z_g, z_h) &\propto P(z_g|x_f)P(z_h|x_f) && (5)\\
&\propto \exp\left(-\frac{1}{2}(z_g - x_f)^2/\sigma_g^2\right)\exp\left(-\frac{1}{2}(z_h - x_f)^2/\sigma_h^2\right) && (6)\\
&= \exp\left(-\frac{1}{2}\frac{(z_g - x_f)^2\sigma_h^2 + (z_h - x_f)^2\sigma_g^2}{\sigma_h^2\sigma_g^2}\right) && (7)\\
&= \exp\left(-\frac{1}{2}\frac{(\sigma_h^2 + \sigma_g^2)x_f^2 - 2(z_g\sigma_h^2 + z_h\sigma_g^2)x_f + (z_g^2\sigma_h^2 + z_h^2\sigma_g^2)}{\sigma_h^2\sigma_g^2}\right) && (8)\\
&\propto \exp\left(-\frac{1}{2}\frac{(\sigma_h^2 + \sigma_g^2)x_f^2 - 2(z_g\sigma_h^2 + z_h\sigma_g^2)x_f}{\sigma_h^2\sigma_g^2}\right) && (9)
\end{aligned}$$

This looks messy, but can be converted into something familiar, using the trick of "completing the square."

$$ax^2 + bx + c = a(x - \frac{-b}{2a})^2 + (c - \frac{b^2}{4a}) \qquad (10)$$

Applying this trick to Equation 9 gives us:

$$\begin{aligned}
P(x_f|z_g, z_h) &\propto \exp\left[-\frac{1}{2}\left(\frac{\sigma_g^2 + \sigma_h^2}{\sigma_g^2\sigma_h^2}\right)\left(x_f - \frac{2z_g\sigma_h^2 + 2z_h\sigma_g^2}{2(\sigma_h^2 + \sigma_g^2)}\right)^2 + R\right] && (11)\\
&\propto \exp\left[-\frac{1}{2}\frac{\left(x_f - \frac{z_g\sigma_h^2 + z_h\sigma_g^2}{(\sigma_h^2+\sigma_g^2)}\right)^2}{\left(\frac{\sigma_h\sigma_g}{\sqrt{\sigma_g^2+\sigma_h^2}}\right)^2}\right] && (12)
\end{aligned}$$

i.e. a Gaussian distribution with mean $\frac{z_g\sigma_h^2 + z_h\sigma_g^2}{(\sigma_h^2+\sigma_g^2)}$ and variance $\frac{\sigma_h^2\sigma_g^2}{\sigma_g^2+\sigma_h^2}$

More General Assumptions

We will make the following assumptions:

- Markovian conditional independence for states, $P(x_t|x_{0:t-1}) = P(x_t|x_{t-1})$, and measurements $P(z_t|x_{0:t}, z_{0:t-1}) = P(z_t|x_t)$
- x_0 is drawn from a Gaussian distribution $N(\mu_0, \Sigma_0)$
- $P(x_t|x_{t-1})$ is a linear Gaussian distribution, $P(x_t|x_{t-1}) = N(Ax_{t-1}, \Sigma_x)$. I.e. $x_t = Ax_{t-1} + w_t$, linear transformation of x_{t-1} plus (white) Gaussian noise.
- $P(z_t|x_t)$ is a linear Gaussian distribution, $P(z_t|x_t) = N(Cx_t, \Sigma_z)$. I.e. $z_t = Cx_t + v_t$

Note that A, Σ_x, C, Σ_z could also vary over time t in general.

A general important fact about Bayesian networks of this sort, in which all conditional distributions are linear Gaussians, is that the joint probability distribution is a multivariate Gaussian distribution. Further, all conditional distributions are also multivariate Gaussians.

The general case

$$
\begin{aligned}
P(x_t|z_{0:t}) &= P(x_t|z_{0:t-1}, z_t) && (13)\\
&= \alpha P(z_t|x_t, z_{0:t-1})P(x_t|z_{0:t-1}) && (14)\\
&= \alpha P(z_t|x_t) \int P(x_t, x_{t-1}|z_{0:t-1})dx_{t-1} && (15)\\
&= \alpha P(z_t|x_t) \int P(x_{t-1}|z_{0:t-1})P(x_t|x_{t-1}, z_{0:t-1})dx_{t-1} && (16)\\
&= \alpha P(z_t|x_t) \int P(x_{t-1}|z_{0:t-1})P(x_t|x_{t-1})dx_{t-1} && (17)
\end{aligned}
$$

$P(x_t|z_{0:t})$ is a multivariate Gaussian for all t, denote the mean μ_t and the variance Σ_t. Equation 17 defines the recurrence relation between these parameters for t and $t-1$.

A slightly less simple 1-D example

We will derive the Kalman filter updates for a 1-D state vector. Let $P(x_t|x_{t-1}) = N(ax_{t-1}, \sigma_x^2)$, $P(z_t|x_t) = N(cx_t, \sigma_z^2)$.

Following the derivation in [1], use the notation

$$
g(x; \mu, \nu) = exp\left(-\frac{(x-\mu)^2}{2\nu}\right) \quad (18)
$$

The following identities then hold:

$$
\begin{aligned}
g(x; \mu, \nu) &= g(x-\mu; 0, \nu); && (19)\\
g(m; n, \nu) &= g(n; m, \nu); && (20)\\
g(ax; \mu, \nu) &= g(x; \mu/a, \nu/a^2); && (21)
\end{aligned}
$$

Also,

$$
\int_{-\infty}^{\infty} g(x-u, \mu, \nu_a)g(u; 0, \nu_b)du \propto g(x; \mu, \nu_a^2 + \nu_b^2) \quad (22)
$$

This fact can be obtained by thinking about the distribution of $Z = X + Y$ where X and Y are normally distributed random variables.

Finally,

$$
g(x; a, b)g(x; c, d) = g\left(x; \frac{ad+cb}{b+d}, \frac{bd}{b+d}\right) f(a, b, c, d) \quad (23)
$$

Note that the $f(\cdot)$ does not depend on x. The derivation of this fact is the same as that in the first simple example.

Using these facts, we can evaluate the integral in Equation 17.

$$
\begin{aligned}
P(x_t|z_{0:t-1}) &= \int P(x_{t-1}|z_{0:t-1})P(x_t|x_{t-1})dx_{t-1} && (24)\\
&= \int P(x_t|x_{t-1})P(x_{t-1}|z_{0:t-1})dx_{t-1} && (25)\\
&\propto \int g(x_t; ax_{t-1}, \sigma_x^2)g(x_{t-1}; \mu_{t-1}, \sigma_{t-1}^2)dx_{t-1} && (26)\\
&\propto \int g((x_t - ax_{t-1}); 0, \sigma_x^2)g((x_{t-1} - \mu_{t-1}); 0, \sigma_{t-1}^2)dx_{t-1} && (27)\\
&\propto \int g((x_t - a(u + \mu_{t-1})); 0, \sigma_x^2)g(u; 0, \sigma_{t-1}^2)du && (28)\\
&\propto \int g((x_t - au); a\mu_{t-1}, \sigma_x^2)g(u; 0, \sigma_{t-1}^2)du && (29)\\
&\propto \int g((x_t - v); a\mu_{t-1}, \sigma_x^2)g(v; 0, (a\sigma_{t-1})^2)du && (30)\\
&\propto g(x_t; a\mu_{t-1}, \sigma_x^2 + (a\sigma_{t-1})^2) && (31)
\end{aligned}
$$

Denote the above mean and variances by $\mu_t^- = a\mu_{t-1}$ and $\sigma_t^- = \sigma_x^2 + (a\sigma_{t-1})^2$. The final update, multiplying Equation 31 into Equation 17 gives:

$$
\begin{aligned}
P(x_t|z_{0:t}) &= \alpha P(z_t|x_t, z_{0:t-1})P(x_t|z_{0:t-1}) && (32)\\
&\propto g(z_t; cx_t, \sigma_z^2)g(x_t; \mu_t^-, \sigma_t^-) && (33)\\
&= g(cx_t; z_t, \sigma_z^2)g(x_t; \mu_t^-, \sigma_t^-) && (34)\\
&= g(x_t; z_t/c, (\sigma_z/c)^2)g(x_t; \mu_t^-, \sigma_t^-) && (35)
\end{aligned}
$$

Applying our identities, we obtain:

$$
\begin{aligned}
\mu_t &= \left(\frac{\mu_t^- \sigma_z^2 + cz_t(\sigma_t^-)^2}{\sigma_z^2 + c^2(\sigma_t^-)^2}\right) && (36)\\
\sigma_t &= \sqrt{\left(\frac{\sigma_z^2(\sigma_t^-)^2}{\sigma_z^2 + c^2(\sigma_t^-)^2}\right)} && (37)
\end{aligned}
$$

The multivariate case

Similar identities for multivariate Gaussian distributions can be derived. In the full multivariate case, the final update equations for μ_t and Σ_t are:

$$
\begin{aligned}
\mu_t &= A\mu_{t-1} + K_t(z_t - CA\mu_{t-1}) && (38)\\
\Sigma_t &= (I - K_t)(A\Sigma_{t-1}A^T + \Sigma_x) && (39)
\end{aligned}
$$

where

$$
K_t = (A\Sigma_{t-1}A^T + \Sigma_x)C^T(C(A\Sigma_{t-1}A^T + \Sigma_x)C^T + \Sigma_z)^{-1} \qquad (40)
$$

Other issues

The *data association* problem arises when trying to track multiple (possibly interacting) objects. The basic problem is which measurement goes with which state variable.

A simple approach is to use nearest-neighbour data association, where measurements are assigned to closest forward projected state variables. Distance can be measured using the Mahalanobis distance, reweighting coordinates based on measurement covariance matrix $||x-y||^2_\Sigma = (x-y)^T \Sigma^{-1} (x-y)^T$.

Probabilistic techniques that average over possible assignments (there are $m(m-1)(m-2)\ldots(m-n+1)$ assignments with n objects and m measurements) are also used.

References

[1] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach.* Prentice Hall, 2003.

[2] M. Jordan and C. Bishop. An introduction to graphical models.

[3] P. S. Maybeck. *Stochastic models, estimation, and control*, volume 141 of *Mathematics in Science and Engineering.* 1979.

[4] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach.* Prentice-Hall, Englewood Cliffs, NJ, second edition, 2003.

[5] G. Welch and G. Bishop. Kalman filter webpage. `http://www.cs.unc.edu/~welch/kalman/`.