Recursive Optimization

Andrius Buinovskij - 18-940-270

1 Kalman Filter Continued

Recall We have

$$x_m(0) = x(0) \tag{1}$$
$$x_p(k) = A(k-1)x_m(k-1) + v(k-1) + u(k-1) \tag{2}$$
$$z_m(k) = H(k-1)x_p(k) + w(k) \tag{3}$$
$$p_{x_m(k)} = p_{x(k)|z(1:k)}(\xi|\bar{z}(k)) \tag{4}$$

And of course We had

$$p_{x_p(k)}(\xi) = p_{x(k)|z(1:k-1)}(\xi|\bar{z}(1:k-1) \tag{5}$$
$$p_{x_m(k)}(\xi) == p_{x(k)|z(1:k)}(\xi|\bar{z}(1:k) \tag{6}$$

Wee.

Next We show that $x_p(k)$ and $x_m(k)$ are Gaussian random variables (GRVs), and find ways to compute their variance and mean. If We can do that, We have everything We need.

1.1 Auxiliary variables are GRVs

The proofs are by induction.

Immediately, $x_p(k)$ is a function of independent Gaussians. If You assume that the previous state $x_m(k-1)$ is GRV, You're good to go, and of course that's fine since our starting point by definition is a GRV.

The second one's trickier - $x_m(k)$. We drop the (k) for convenience.

So, by the definition and then Bayes' rule We have

$$p_{x_m} = p_{x_p|z_m}(\xi|\bar{z}) = \frac{p_{z_m|x_p}(\bar{z}|\xi) \cdot p_{x_p}(\xi)}{p_{z_m}(\bar{z})} \tag{7}$$

So nothing crazy there - We have $p_{x_m}(\xi)$, then We use the definition of x_m which is just conditioning x_p on z_m, and then use Bayes' rule to condition.

Now then, We can immediately ignore the normalizing constant - does not depend on ξ.

The numerator is a product of two GRVs:

$$p_{x_p}(\xi) \propto \exp\Bigg(-\frac{1}{2}(\xi - \hat{x}_p)^\top P_p^{-1}(\xi - \hat{x}_p)\Bigg) \tag{8}$$

Where of course $\hat{x}_p$ is the mean of the random variable x_pn and P_p is the variance matrix of x_p.

Similarly:

$$p_{z_m|x_p}(\bar{z}|\xi) \propto \exp\Bigg(-\frac{1}{2}(\bar{z} - H\xi)^\top R^{-1}(\bar{z} - H\xi)\Bigg) \tag{9}$$

Where $\bar{z}$ is the mean of $z_m|x_p$ which is just the mean of the noise term with an offset since x_p becomes a constant.

And then of course when You multiply these two exponential terms together You add the exponents yielding:

$$p_{x_m} \propto \exp\Bigg(-\frac{1}{2}\Bigg((\bar{z} - H\xi)^\top R^{-1}(\bar{z} - H\xi) + (\xi - \hat{x}_p)^\top P_p^{-1}(\xi - \hat{x}_p)\Bigg)\Bigg) \tag{10}$$

The R there is the variance matrix of the measurement noise w.

And since it's just some quadratic mess up in that exponent, there must exist a μ and Σ such that

$$p_{x_m} \propto \exp\Bigg(-\frac{1}{2}(\xi - \mu)^\top \Sigma^{-1}(\xi - \mu)\Bigg) \tag{11}$$

And so We're done.

Note that I think in general We could have just said: GRVs are closed under conditioning, but I suppose this is more rigorous.

1.2 Mean and variance of auxiliary variables

First the mean:

$$\begin{aligned} \hat{x}_p(k) = E(x_p(k)) &= A(k-1)E(x_m(k-1) + u(k-1) + E(v(k-1)) && (12)\\ &= A(k-1)\hat{x}_m(k-1) + u(k-1) && (13) \end{aligned}$$

So no worries there, and We can assume that We have the previous mean.

$$\begin{aligned} P_p(k) = Var(x_p(k)) &= E((x_p(k) - \hat{x}_p(k))(x_p(k) - \hat{x}_p(k))^\top) && (14)\\ &= E(x_p(k)(x_p(k) - \hat{x}_p(k))^\top - \hat{x}_p(k)(x_p(k) - \hat{x}_p(k))^\top =) && (15) \end{aligned}$$

Alright I am just going to skip the expansion here and stick with the notes. The main takeaway is to keep an eye out for independent things that cancel, since for independent A, B We have $E(AB) = E(A)E(B)$, and recall that $Var(X) = E(X^2) - E(X)^2$.

After simplification You get

$$P_p(k) = A(k-1)P_m(k-1)A^\top(k-1) + Q(k-1) \tag{16}$$

Where $P_m(k-1)$ is of course the variance of $x_m(k-1)$, and Q is the variance of the process noise.

Honestly just write these down, it's just linear algebra.

The formulas for x_m are:

$$\Sigma^{-1} = P_p^{-1} + H^\top R^{-1} H \tag{17}$$

So P_p is the variance of the prior, H is the linear matrix for transforming state to observation, and R is the measurement noise variance matrix.

The mean is given by:

$$\mu = \hat{x}_p + \Sigma H^\top R^{-1}(\bar{z} - H\hat{x}_p) \tag{18}$$

Where $\bar{z}$ is just the mean of the wonky conditional variable as mentioned before.

1.3 Summary

Alright! So

First We initialize $\hat{x}_m(0) = x_0$ and $P_m(0) = P_0$.

Prior update/Prediction Step:

$$\hat{x}_p(k) = A(k-1)\hat{x}_m(k-1) + u(k) \tag{19}$$

$$P_p(k) = A(k-1)P_m(k-1)A^\top(k-1) + Q(k-1) \tag{20}$$

So We're just calculating relevant parameters here. Strictly speaking We don't need to calculate any actual values, there's no propagation happening in that We're going to receive new information.

A posteriori update/Measurement update step:

$$P_m(k) = (P_p^{-1}(k) + H^\top(k)R^{-1}(k)H(k))^{-1} \tag{21}$$

$$\hat{x}_m(k) = \hat{x}_p(k) + P_m(k)H^\top(k)R^{-1}(k)(\bar{z}(k) - H(k)\hat{x}_p(k)) \tag{22}$$