AN INTUITIVE EXPLANATION OF THE KALMAN UPDATE

ADAM WILLIAMS

1. Formulation

Suppose we have a normally-distributed random variable x with known *prior* mean and covariance,

$$x \sim \mathcal{N}(\tilde{x}, \tilde{P}).$$

Suppose also that we have a linear-Gaussian measurement model producing a measurement z of x,

$$z = Hx + r,$$

where $r \sim \mathcal{N}(0, R)$.

The **Kalman update** of $(\tilde{x}, \tilde{P})$ by the measurement z is the normal distribution $(\hat{x}, P)$ that maximizes the log-likelihood of the measurement z under the prior distribution of x. In other words,

$$\hat{x} = \arg\min_{x} \left(\|z - Hx\|_R^2 + \|\tilde{x} - x\|_{\tilde{P}}^2 \right), \tag{1}$$

where $\|u\|_A^2 = u^\top A^{-1} u$ is the Mahalanobis norm of u under A.

2. Solution

The solution can be computed by subsituting $x \leftarrow \tilde{x} + \delta$ into (1), differentiating with respect to δ, setting equal to zero, and solving for δ (note that this is a typical linear-least-squares problem and solution). The normal equation is thus

$$\left(H^\top R^{-1} H + \tilde{P}^{-1} \right) \delta = H^\top R^{-1} (z - Hx) + \tilde{P}^{-1} (\tilde{x} - x). \tag{2}$$

Rearranging and setting $\Lambda = H^\top R^{-1} H + \tilde{P}^{-1}$, we have:

$$\hat{x} = \Lambda^{-1} \left(H^\top R^{-1} z + \tilde{P}^{-1} \tilde{x} \right) \tag{3}$$

$$\hat{P} = \Lambda^{-1}. \tag{4}$$

Note that the *Woodbury matrix identity* can be used to convert this expression into the standard Kalman update form; see [1] for details.

3. Interpretation

3.1. **As a Posterior Distribution.**

As suggested above, the Kalman update can be viewed as yielding the posterior distribution maximizing the likelihood of the measurement and the prior state simultaneously, according to their respective uncertainties.

3.2. **As a Gauss-Newton Step.** If the measurement model is nonlinear and is being approximated by its derivative, $h(x+\delta) \approx h(x) + H\delta$, then the Kalman update is equivalent to taking a single Gauss-Newton optimization step towards minimizing the objective function (1). Under this interpretation, the extracted covariance P is the inverse of the Hessian approximation $J^\top \Sigma^{-1} J$ of the objective function $\jmath(x) = \frac{1}{2} r(x)^T \Sigma^{-1} r(x)$, with $r(x+\delta) \approx r(x) + J\delta$.

3.3. **As an Information-Weighted Mean.** Note that the **information** (i.e. inverse covariance) of the posterior distribution is simply the sum of the prior information and the measurement information:

$$\Lambda = \tilde{P}^{-1} + H^\top R^{-1} H,$$

and that the prior mean is simply the *information-weighted mean* of the prior and measurement distributions (each mean is scaled by the appropriate inverse covariance; the two are then added and normalized). Explicitly (where matrix division means left-multiplication by the inverse):

$$\hat{x} = \frac{\tilde{P}^{-1}\tilde{x} + H^\top R^{-1} z}{\tilde{P}^{-1} + H^\top R^{-1} H}.$$

Note that, in each case, the measurement information has to be "unprojected" into state space according to the measurement Jacobian.

References

[1] Ethan Eade, *Monocular Simultaneous Localization and Mapping*, 2008.