Josue Casco-Rodriguez

Discriminative Kalman Filter Derivation

I. Traditional Kalman System Model

Consider a state-space model $Z_{1:T} = \{Z_1, Z_2, \ldots, Z_T\}$ (latent states) and $X_{1:T} = \{X_1, X_2, \ldots, X_T\}$ (observations) represented as a Bayesian network:

$$\begin{array}{ccccccccccc} Z_1 & \longrightarrow & \cdots & \longrightarrow & Z_{t-1} & \longrightarrow & Z_t & \longrightarrow & \cdots & \longrightarrow & Z_T \\ \downarrow & & & & \downarrow & & \downarrow & & & & \downarrow \\ X_1 & & & & X_{t-1} & & X_t & & & & X_T \end{array}$$

The conditional density of Z_T can be expressed recursively using the Chapman-Kolmogorov equation and Bayes' rule:

1. $p(z_t \mid x_{1:t-1}) = \int dz_{t-1}\, p(z_t \mid z_{t-1})\, p(z_{t-1} \mid x_{1:t-1})$
2. $p(z_t \mid x_{1:t}) = p(x_t \mid z_t)\, p(z_t \mid x_{1:t-1}) / p(x_t \mid x_{1:t-1})$,

where $p(z_0 \mid x_{1:0}) = p(z_0)$.

As with the traditional Kalman filter, the network is modeled as follows:

3. $p(z_0 \mid x_{1:0}) = p(z_0) = N(0, S)$
4. $p(z_t \mid z_{t-1}) = N(Az_{t-1}, \Gamma)$,

where N() denotes a multivariate Gaussian distribution, S is the covariance of z_t when not conditioned on any z_{t-1}, A is the state transition matrix, and Γ is the process noise covariance. Furthermore, the network is modeled such that the observation model, $p(x_t \mid z_t)$, and the process transformation model, $p(z_t \mid z_{t-1})$ are stationary.

II. Discriminative Kalman Filter Model

Using the aforementioned details from the traditional Kalman system model, the Discriminative Kalman Filter (DKF) begins by making the following approximations:

5. $p(z_t \mid x_t) \sim N(f(x_t), Q(x_t))$
6. $f(x) = E(z_t \mid x_t = x)$
7. $Q(x) = V(z_t \mid x_t = x)$

Since, in neural recordings, the dimensionality of observations x is usually much larger than the dimensionality of latent states z, the Bernstein-von Mises Theorem states that there exist such functions f() and Q() such that the approximation of $p(z_t \mid x_t)$ is accurate, requiring only mild regularity conditions on the aforementioned stationary observation model $p(x_t \mid z_t)$.

To make use of Equation 5 for approximating Equation 2, $p(x_t \mid z_t)$ can be rewritten using Bayes' rule ($p(x_t \mid z_t) = p(z_t \mid x_t)\, p(x_t) / p(z_t)$):

8. $p(z_t \mid x_{1:t}) = (p(x_t) / p(x_t \mid x_{1:t-1})) * (p(z_t \mid x_t) / p(z_t)) * p(z_t \mid x_{1:t-1})$

Next, the first set of terms (i.e., all terms that do not depend on z_t) are condensed into a normalizing constant K, and $p(z_t \mid x_{1:t-1})$ is replaced by its integral form expressed in Equation 1:

9. $p(z_t \mid x_{1:t}) = K * (p(z_t \mid x_t) / p(z_t)) * \int dz_{t-1}\, p(z_t \mid z_{t-1})\, p(z_{t-1} \mid x_{1:t-1})$

Finally, the DKF model (specifically, Equations 3, 4, 5) are substituted into Equation 9:

10. $p(z_t \mid x_{1:t}) = K * (N(f(x_t), Q(x_t)) / N(0, S)) * \int dz_{t-1}\, N(Az_{t-1}, \Gamma)\, p(z_{t-1} \mid x_{1:t-1})$,

where the $p(z_{t-1} \mid x_{1:t-1})$ is approximated as Gaussian and is defined as Gaussian for the case when z = 1 (see Equation 3).

III. Discriminative Kalman Filter Algorithm Derivation (Lemma 1)

To derive the full DKF algorithm, Equation 10 must be simplified into a recursive series of linear algebra operations.

From traditional Kalman literature, the integral in Equation 10 simplifies to:

11. $\int dz_{t-1}\, N(Az_{t-1}, \Gamma)\, p(z_{t-1} \mid x_{1:t-1}) = N(Az_{t-1}, A\sum_{t-1}A + \Gamma) = N(v_t, M_t)$,

where $\sum_t$ is the uncertainty of the state estimate z_t, $v_t = Az_{t-1}$, and $M_t = A\sum_{t-1}A + \Gamma$.

As for the quotient of Gaussians N(f(x), Q(x)) / N(0, S) in Equation 10, its simplification follows as such:

- $N(f(x), Q(x)) \propto \exp((-1/2)(z - f)^TQ^{-1}(z - f))$, where f = f(x) and Q = Q(x)
- $N(0, S) \propto \exp((-1/2)z^TS^{-1}z)$
- $N(f, Q) / N(0, S) \propto \exp((-1/2)[z^TQ^{-1}z - 2z^TQ^{-1}f - z^TS^{-1}z])$
- $N(f, Q) / N(0, S) \propto \exp((-1/2)[z^T(Q^{-1} - S^{-1})z] - 2z^TQ^{-1}f])$
- $N(f, Q) / N(0, S) \propto \exp((-1/2)(z-b)^TU^{-1}(z-b))$

12. $N(f, Q) / N(0, S) \propto N(b, U)$,

where $U = Q(x)^{-1} - S^{-1}$ and $b = U*Q(x)^{-1}*f(x)$.

Thus, ignoring the constant of proportionality K, Equation 10 simplifies to:

- $p(z_t \mid x_{1:t}) \propto N(b, U) * N(v_t, G_t)$
- $p(z_t \mid x_{1:t}) \propto N(\sum_t[U^{-1}b + M_t^{-1}v_t], \sum_t)$

13. $p(z_t \mid x_{1:t}) \propto N(\mu_t, \sum_t)$,

where the uncertainty of the state estimate μ_t is $\sum_t = (M_t^{-1} + Q(x_t)^{-1} - S^{-1})^{-1}$, and the state estimate itself is $\mu_t = \sum_t(M_t^{-1}v_t + Q(x_t)^{-1}f(x_t))$.

Therefore, the resulting DKF estimate, using regression to estimate f(x) and Q(x), of μ_t and $\sum_t$ follows as:

$$\boxed{v_t = A\mu_{t-1}, M_t = A\,\Sigma_{t-1}\,A^{\top} + \Gamma\,,\ \Sigma_t = \left(M_t^{-1} + Q(x_t)^{-1} - S^{-1}\right)^{-1}, \mu_t = \Sigma_t\left(M_t^{-1}v_t + Q(x_t)^{-1}f(x_t)\right).}$$