Equations for the Prediction Stage of the Information Filter

Hugh Durrant-Whyte
hugh@acfr.usyd.edu.au

August 24, 2000

Conventional Kalman filters deal with the estimation of states $\mathbf{x}(i)$, and yield estimates $\hat{\mathbf{x}}(i \mid j)$ together with a corresponding estimate variance $\mathbf{P}(i \mid j)$. The information filter deals instead with the information state vector $\hat{\mathbf{y}}(i \mid j)$ and information matrix $\mathbf{Y}(i \mid j)$ defined as

$$\hat{\mathbf{y}}(i \mid j) = \mathbf{P}^{-1}(i \mid j)\hat{\mathbf{x}}(i \mid j), \qquad \mathbf{Y}(i \mid j) = \mathbf{P}^{-1}(i \mid j). \tag{1}$$

A set of recursion equations for the information state and information matrix can be derived directly from the equations for the Kalman filter. The resulting information filter is mathematically identical to the conventional Kalman filter.

We have

$$\mathbf{1} - \mathbf{W}(k)\mathbf{H}(k) = \mathbf{P}(k \mid k)\mathbf{P}^{-1}(k \mid k-1), \tag{2}$$

and

$$\mathbf{W}(k) = \mathbf{P}(k \mid k)\mathbf{H}^T(k)\mathbf{R}^{-1}(k). \tag{3}$$

Substituting Equations 2 and 3 into the state update equations for the Kalman Filter and premultiplying through by $\mathbf{P}^{-1}(k \mid k)$ gives the update equation for the information-state vector as

$$\mathbf{P}^{-1}(k \mid k)\hat{\mathbf{x}}(k \mid k) = \mathbf{P}^{-1}(k \mid k-1)\hat{\mathbf{x}}(k \mid k-1) + \mathbf{H}^T(k)\mathbf{R}^{-1}(k)\mathbf{z}(k). \tag{4}$$

A similar expression can be found for the information matrix. Substituting Equations 2 and 3 into the covariance update equation for the Kalman filter and rearranging gives

$$\mathbf{P}^{-1}(k \mid k) = \mathbf{P}^{-1}(k \mid k-1) + \mathbf{H}^T(k)\mathbf{R}^{-1}(k)\mathbf{H}(k). \tag{5}$$

Defining

$$\mathbf{i}(k) \stackrel{\triangle}{=} \mathbf{H}^T(k)\mathbf{R}^{-1}(k)\mathbf{z}(k) \tag{6}$$

as the information-state contribution from an observation $\mathbf{z}(k)$, and

$$\mathbf{I}(k) \stackrel{\triangle}{=} \mathbf{H}^T(k)\mathbf{R}^{-1}(k)\mathbf{H}(k) \tag{7}$$

as its associated information matrix, and with the definitions defined in Equation 1, Equations 4 and 5 become

$$\hat{\mathbf{y}}(k \mid k) = \hat{\mathbf{y}}(k \mid k-1) + \mathbf{i}(k) \tag{8}$$

and

$$\mathbf{Y}(k \mid k) = \mathbf{Y}(k \mid k-1) + \mathbf{I}(k). \tag{9}$$

Recall the covariance prediction equation

$$\mathbf{P}(k \mid k-1) = \mathbf{F}(k)\mathbf{P}(k-1 \mid k-1)\mathbf{F}^T(k) + \mathbf{G}(k)\mathbf{Q}(k)\mathbf{G}^T(k) \tag{10}$$

To derive the prediction stage for the information filter, the following version of the matrix inversion lemma is noted [2]

$$\left(\mathbf{A} + \mathbf{B}^T\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}^T\left(\mathbf{1} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B}^T\right)^{-1}\mathbf{C}\mathbf{A}^{-1}.$$

Identifying

$$A = \mathbf{F}(k)\mathbf{P}(k-1 \mid k-1)\mathbf{F}^T(k), \qquad \mathbf{B}^T = \mathbf{G}(k)\mathbf{Q}(k), \qquad \mathbf{C} = \mathbf{G}^T(k),$$

the inverse of Equation 10 becomes

$$\mathbf{P}^{-1}(k \mid k-1) = \mathbf{M}(k) - \mathbf{M}(k)\mathbf{G}(k)\left[\mathbf{G}^T(k)\mathbf{M}(k)\mathbf{G}(k) + \mathbf{Q}^{-1}(k)\right]^{-1}\mathbf{G}^T(k)\mathbf{M}(k) \tag{11}$$

where

$$\mathbf{M}(k) = \mathbf{F}^{-T}(k)\mathbf{P}^{-1}(k-1 \mid k-1)\mathbf{F}^{-1}(k) \tag{12}$$

when $\mathbf{Q}(k)$ is non singular. Note the definition

$$\mathbf{F}(k) \triangleq \mathbf{\Phi}(t_k, t_{k-1})$$

implies $\mathbf{F}^{-1}(k)$ always exists and indeed

$$\mathbf{F}^{-1}(k) = \mathbf{\Phi}(t_{k-1}, t_k)$$

is simply the state transition matrix defined backwards from a time t_k to t_{k-1}. Now, defining

$$\mathbf{\Sigma}(k) \triangleq \left[\mathbf{G}^T(k)\mathbf{M}(k)\mathbf{G}(k) + \mathbf{Q}^{-1}(k)\right], \tag{13}$$

and

$$\mathbf{\Omega}(k) \triangleq \mathbf{M}(k)\mathbf{G}(k)\left[\mathbf{G}^T(k)\mathbf{M}(k)\mathbf{G}(k) + \mathbf{Q}^{-1}(k)\right]^{-1} = \mathbf{M}(k)\mathbf{G}(k)\mathbf{\Sigma}^{-1}(k), \tag{14}$$

the information matrix prediction equation becomes

$$\mathbf{P}^{-1}(k \mid k-1) = \mathbf{Y}(k \mid k-1) = \mathbf{M}(k) - \mathbf{\Omega}(k)\mathbf{\Sigma}(k)\mathbf{\Omega}^T(k). \tag{15}$$

Alternate expressions are

$$\mathbf{Y}(k \mid k-1) = \left[\mathbf{1} - \mathbf{\Omega}(k)\mathbf{G}^T(k)\right]\mathbf{M}(k) \tag{16}$$

and

$$\mathbf{Y}(k \mid k-1) = \left[\mathbf{1} - \mathbf{\Omega}(k)\mathbf{G}^T\right]\mathbf{M}(k)\left[\mathbf{1} - \mathbf{\Omega}(k)\mathbf{G}^T\right]^T + \mathbf{\Omega}(k)\mathbf{Q}^{-1}(k)\mathbf{\Omega}^T(k). \tag{17}$$

The information-state prediction equations may also be obtained as

$$\begin{aligned}
\hat{\mathbf{y}}(k \mid k-1) &= \left[\mathbf{1} - \mathbf{\Omega}(k)\mathbf{G}^T(k)\right]\mathbf{F}^{-T}(k) \\
&\quad \times \left[\hat{\mathbf{y}}(k-1 \mid k-1) + \mathbf{Y}(k-1 \mid k-1)\mathbf{F}^{-1}(k)\mathbf{B}(k)\mathbf{u}(k)\right] \\
&= \left[\mathbf{1} - \mathbf{\Omega}(k)\mathbf{G}^T(k)\right]\left[\mathbf{F}^{-T}(k)\hat{\mathbf{y}}(k-1 \mid k-1) + \mathbf{M}(k)\mathbf{B}(k)\mathbf{u}(k)\right] \\
&= \left[\mathbf{1} - \mathbf{\Omega}(k)\mathbf{G}^T(k)\right]\mathbf{F}^{-T}(k)\hat{\mathbf{y}}(k-1 \mid k-1) \\
&\quad + \mathbf{Y}(k \mid k-1)\mathbf{B}(k)\mathbf{u}(k).
\end{aligned} \tag{18}$$

It should be noted that the complexity of the inversion of $\mathbf{\Sigma}(k)$ is only of order the dimension of the driving noise (often scalar). Further $\mathbf{Q}(k)$ is almost never singular, and indeed if it were, singularity can be eliminated by appropriate definition of $\mathbf{G}(k)$. The special case in which $\mathbf{Q}(k) = \mathbf{0}$ yields prediction equations in the form:

$$\mathbf{Y}(k \mid k-1) = \mathbf{M}(k), \qquad \hat{\mathbf{y}}(k \mid k-1) = \mathbf{F}^{-T}(k)\hat{\mathbf{y}}(k-1 \mid k-1) + \mathbf{M}(k)\mathbf{B}(k)\mathbf{u}(k) \tag{19}$$

There is a clear duality between the information filter and the conventional Kalman filter, in which the prediction stage of the information filter is related to the update stage of the Kalman filter, and the update stage of the information filter to the prediction stage of the Kalman filter [1]. In particular, identifying the correspondences

$$\mathbf{\Omega}(k) \to \mathbf{W}(k), \quad \mathbf{\Sigma}(k) \to \mathbf{S}(k), \quad \mathbf{G}^T(k) \to \mathbf{H}(k)$$

$\mathbf{\Omega}(k)$ is seen to take on the role of an information prediction gain matrix, $\mathbf{\Sigma}(k)$ an 'information innovation matrix', and $\mathbf{G}^T(k)$ an 'information observation'. This duality is instructive in understanding the relationship of information to state as equivalent representations and also in terms of implementation of filtering equations.

The information filter is now summarised

Prediction:

$$\hat{\mathbf{y}}(k \mid k-1) = \left[\mathbf{1} - \mathbf{\Omega}(k)\mathbf{G}^T(k)\right]\mathbf{F}^{-T}(k)\hat{\mathbf{y}}(k-1 \mid k-1) + \mathbf{Y}(k \mid k-1)\mathbf{B}(k)\mathbf{u}(k) \tag{20}$$

$$\mathbf{Y}(k \mid k-1) = \mathbf{M}(k) - \mathbf{\Omega}(k)\mathbf{\Sigma}(k)\mathbf{\Omega}^T(k) \tag{21}$$

where

$$\mathbf{M}(k) = \mathbf{F}^{-T}(k)\mathbf{P}^{-1}(k-1 \mid k-1)\mathbf{F}^{-1}(k), \tag{22}$$

$$\mathbf{\Omega}(k) = \mathbf{M}(k)\mathbf{G}(k)\mathbf{\Sigma}^{-1}(k), \tag{23}$$

and

$$\mathbf{\Sigma}(k) = \left[\mathbf{G}^T(k)\mathbf{M}(k)\mathbf{G}(k) + \mathbf{Q}^{-1}(k)\right]. \tag{24}$$

Estimate:

$$\hat{\mathbf{y}}(k \mid k) = \hat{\mathbf{y}}(k \mid k-1) + \mathbf{i}(k) \tag{25}$$

$$\mathbf{Y}(k \mid k) = \mathbf{Y}(k \mid k-1) + \mathbf{I}(k). \tag{26}$$

The information-filter form has the advantage that the update Equations 25 and 26 for the estimator are computationally simpler than the equations for the Kalman Filter, at the cost of increased complexity in prediction. The value of this in decentralized sensing is that estimation occurs locally at each node, requiring partition of the estimation equations which are simpler in their information form. Prediction, which is more complex in this form, relies on a propagation coefficient which is independent of the observations made and so is again simpler to decouple and decentralize amongst a network of sensor nodes. This property is exploited in subsequent sections.

The information matrix $\mathbf{Y}(i \mid j)$ has a straight-forward interpretation as the Fisher information; the compactness of the distribution surface or certainty in state estimates. The information state vector $\hat{\mathbf{y}}(i \mid j)$ is technically known as the *score function*. It has no obvious metric properties; the difference between two information-state vectors does not relate at all to the difference in state vectors because of scaling by the information matrix.

References

[1] B.D.O. Anderson and J.B. Moore. *Optimal Filtering.* Prentice Hall, 1979.

[2] P.S. Maybeck. *Stochastic Models, Estimaton and Control, Vol. I.* Academic Press, 1979.