Kalman Filters - State Space models

Prof. Rozenn Dahyot
Trinity College Dublin, Ireland

2017

Given $p(\vec{s}_{j-1}|\vec{o}_{1:j-1}) = \mathcal{N}(\vec{s}_{j-1}; \mu_{j-1}, \Sigma_{j-1})$,

- **State equation** with $\vec{s}_j$ a random vector

$$\vec{s}_j = \mathrm{F}\,\vec{s}_{j-1} + \mathrm{G}\,\vec{\epsilon}_j \tag{1}$$

F is a square matrix, G is a matrix, and $\vec{\epsilon}_j \sim \mathcal{N}(0, \Sigma)$ is a random vector.

- **Observation equation**

$$\vec{o}_j = \mathrm{H}\,\vec{s}_j + \vec{e}_j \tag{2}$$

with H a matrix and $\vec{e}_j \sim \mathcal{N}(0, \Sigma_o)$.

1. Using the state equation (1), compute the predicted expectation $\mathbb{E}[\vec{s}_j]$ w.r.t. the expectation $\mathbb{E}[\vec{s}_{j-1}] = \mu_{j-1}$.
2. Using the state equation (1), compute the predicted covariance $\mathbb{C}[\vec{s}_j]$.
3. Show that answers to questions 1 and 2 lead to the prediction p.d.f. (3) - indicate which hypothesis is made when making this link.
4. Collecting the new observation o_j (eq. 2), compute the updates μ_j and Σ_j to recover results eq. (4).

State Space Models result:

- the prediction p.d.f. is

$$p(\vec{s}_j|\vec{o}_{1:j-1}) = \mathcal{N}(\vec{s}_j; \mathrm{F}\mu_{j-1}, \mathrm{R} = \mathrm{F}\Sigma_{j-1}\mathrm{F}^T + \mathrm{G}\Sigma\mathrm{G}^T) \tag{3}$$

- the update p.d.f. is

$$p(\vec{s}_j|\vec{o}_{1:j}) = \mathcal{N}(\vec{s}_j; \mu_j, \Sigma_j) \tag{4}$$

with

$$\Sigma_j = \left(\mathrm{H}^T\Sigma_o^{-1}\mathrm{H} + \mathrm{R}^{-1}\right)^{-1}$$

and

$$\mu_j = \Sigma_j\left(\mathrm{H}^T\Sigma_o^{-1}\vec{o}_j + \mathrm{R}^{-1}\mathrm{F}\mu_{j-1}\right)$$

Solutions:

1. linearity of $\mathbb{E}$ then
$$\mathbb{E}[\vec{s}_j] = \mathbb{E}[\mathrm{F}\,\vec{s}_{j-1} + \mathrm{G}\,\vec{\epsilon}_j] = \mathrm{F}\,\underbrace{\mathbb{E}[\vec{s}_{j-1}]}_{=\mu_{j-1}} + \mathrm{G}\,\underbrace{\mathbb{E}[\vec{\epsilon}_j]}_{=0} = \mathrm{F}\,\mu_{j-1}$$

2. $$\mathbb{C}[\vec{s}_j] = \mathbb{E}[(\vec{s}_j - \mathrm{F}\,\mu_{j-1})(\vec{s}_j - \mathrm{F}\,\mu_{j-1})^T]$$
replacing $\vec{s}_j = \mathrm{F}\,\vec{s}_{j-1} + \mathrm{G}\,\vec{\epsilon}_j$ then:
$$\mathbb{C}[\vec{s}_j] = \mathbb{E}[(\mathrm{F}\,(\vec{s}_{j-1} - \mu_{j-1}) + \mathrm{G}\,\vec{\epsilon}_j)(\mathrm{F}\,(\vec{s}_{j-1} - \mu_{j-1}) + \mathrm{G}\,\vec{\epsilon}_j)^T]$$
$$\mathbb{C}[\vec{s}_j] = \mathbb{E}[(\mathrm{F}\,(\vec{s}_{j-1}-\mu_{j-1})(\mathrm{F}\,(\vec{s}_{j-1}-\mu_{j-1}))^T + \mathrm{G}\,\vec{\epsilon}_j(\mathrm{F}\,(\vec{s}_{j-1}-\mu_{j-1}))^T + \mathrm{F}\,(\vec{s}_{j-1}-\mu_{j-1})(\mathrm{G}\,\vec{\epsilon}_j)^T + \mathrm{G}\,\vec{\epsilon}_j(\mathrm{G}\,\vec{\epsilon}_j)^T]$$
$$\mathbb{C}[\vec{s}_j] = \mathbb{E}[(\mathrm{F}\,(\vec{s}_{j-1} - \mu_{j-1})(\vec{s}_{j-1} - \mu_{j-1})^T\mathrm{F}^T + \mathrm{G}\,\vec{\epsilon}_j(\vec{s}_{j-1} - \mu_{j-1})^T\mathrm{F}^T + \mathrm{F}\,(\vec{s}_{j-1} - \mu_{j-1})\vec{\epsilon}_j^T\mathrm{G}^T + \mathrm{G}\,\vec{\epsilon}_j\vec{\epsilon}_j^T\mathrm{G}^T]$$
so using linearity of $\mathbb{E}$, independence ($\mathbb{E}[\vec{\epsilon}_j(\vec{s}_{j-1} - \mu_{j-1})^T] = 0$ noise at time j is not related to signal/stochastic process s_j at time j), and given $\mathbb{E}[(\vec{s}_{j-1} - \mu_{j-1})(\vec{s}_{j-1} - \mu_{j-1})^T] = \Sigma_{j-1}$ and $\mathbb{E}[\vec{\epsilon}_j\vec{\epsilon}_j^T] = \Sigma$:
$$\mathbb{C}[\vec{s}_j] = \mathrm{F}\Sigma_{j-1}\mathrm{F}^T + \mathrm{G}\Sigma\mathrm{G}^T$$

3. Assuming the prediction p.d.f. to be Normal, then we only need to compute its mean and covariance to get all information needed.

4. $p(\vec{s}_j|\vec{o}_{1:j}) \propto p(o_j|s_j) \times p(\vec{s}_j|\vec{o}_{1:j-1})$ with $p(o_j|s_j) = \mathcal{N}(o_j; \mathrm{H}s_j, \Sigma_o)$ so we need to rearrange:
$$(o_j - \mathrm{H}s_j)^T\Sigma_o^{-1}(o_j - \mathrm{H}s_j) + (s_j - F\mu_{j-1})^T\mathrm{R}^{-1}(s_j - F\mu_{j-1}) = (s_j - \mu_j)^T\Sigma_j^{-1}(s_j - \mu_j)$$
We extract the terms in $(s_j^T\ ?\ s_j)$ on the left side of the equality to identify Σ_j^{-1}:
$$s_j^T\mathrm{H}^T\Sigma_o^{-1}\mathrm{H}s_j + s_j^T\mathrm{R}^{-1}s_j = s_j^T\underbrace{(\mathrm{H}^T\Sigma_o^{-1}\mathrm{H} + \mathrm{R}^{-1})}_{=\Sigma_j^{-1}}s_j$$
Likewise the term $s_j^T\Sigma_j^{-1}\mu_j$ allows to identify μ_j:
$$s_j^T\mathrm{H}^T\Sigma_o^{-1}o_j + s_j^T\mathrm{R}^{-1}\mathrm{F}\mu_{j-1} = s_j^T\underbrace{(\mathrm{H}^T\Sigma_o^{-1}o_j + \mathrm{R}^{-1}\mathrm{F}\mu_{j-1})}_{\Sigma_j^{-1}\mu_j}$$