Let us assume that

$$\mathbf{Y}_i - \boldsymbol{h}_i(\hat{\boldsymbol{X}}_i^-) \approx \mathbf{H}_i \mathbf{e}_i^- + \mathbf{v}_i \tag{1.38}$$

Then (1.32) implies that

$$\mathbf{e}_i^+ = \mathbf{e}_i^- - \mathbf{K}_i \mathbf{H}_i \mathbf{e}_i^- - \mathbf{K}_i \mathbf{v}_i \tag{1.39}$$

and by our prior assumption that $\mathrm{E}\big[\mathbf{e}_i^- \mathbf{v}_i^\mathsf{T}\big] = \mathbf{0}$,

$$\mathbf{P}_i^+ \approx \mathrm{E}\big[\mathbf{e}_i^+ (\mathbf{e}_i^+)^\mathsf{T} | \mathbb{Y}_i\big] \tag{1.40}$$

$$= (\mathbf{I} - \mathbf{K}_i \mathbf{H}_i)\, \mathbf{P}_i^- \,(\mathbf{I} - \mathbf{K}_i \mathbf{H}_i)^\mathsf{T} + \mathbf{K}_i \mathbf{R}_i \mathbf{K}_i^\mathsf{T} \tag{1.41}$$

Equation (1.41) is *Joseph's Formula*, and it holds for any gain $\mathbf{K}_i$. Only for the optimal gain and true covariance does (1.41) reduce to (1.33). Since (1.31) was computed with only an approximate covariance, and due to the various other approximations listed above as well, $\mathbf{K}_i$ cannot be the optimal gain, so at best, (1.33) will only hold approximately. At worst, such approximations may lead to $\mathbf{P}_i^+$ becoming non-positive definite, which is a significant issue. Because of its symmetric and additive form, (1.41) is much less likely (but not impossible!) to produce a non-positive definite $\mathbf{P}_i^+$.

- By our assumption that $\mathrm{E}[\mathbf{w}(t)\mathbf{w}^\mathsf{T}(\tau)] = \mathbf{Q}(t)\delta(t-\tau)$, one of the integrals in (1.29) should be annihilated by the Dirac function, resulting in

$$\mathbf{S}_i = \int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i,\tau)\mathbf{B}(\tau)\mathbf{Q}(\tau)\mathbf{B}^\mathsf{T}(\tau)\boldsymbol{\Phi}^\mathsf{T}(t_i,\tau)\,\mathrm{d}\tau \tag{1.42}$$

In any case, unlike for a discrete time dynamics model,

$$\mathrm{E}\left[\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i,\tau)\mathbf{B}(\tau)\mathbf{w}(\tau)\mathbf{w}^\mathsf{T}(\sigma)\mathbf{B}^\mathsf{T}(\sigma)\boldsymbol{\Phi}^\mathsf{T}(t_i,\sigma)\,\mathrm{d}\tau\,\mathrm{d}\sigma\right]$$
$$\neq \mathrm{E}\left[\int_{t_{i-1}}^{t_i} \boldsymbol{\Phi}(t_i,\tau)\mathbf{B}(\tau)\mathbf{w}(\tau)\,\mathrm{d}\tau \int_{t_{i-1}}^{t_i} \mathbf{w}^\mathsf{T}(\tau)\mathbf{B}^\mathsf{T}(\tau)\boldsymbol{\Phi}^\mathsf{T}(t_i,\tau)\,\mathrm{d}\tau\right] \tag{1.43}$$

- In general, one would need to simultaneously integrate (1.26) and (1.28) due to their interdependence via the Jacobian $\mathbf{A}$. If the time between measurements is small enough, then if one were to employ a suitable approximation for (1.28), perhaps as simple as

$$\boldsymbol{\Phi}(t_i, t_{i-1}) \approx \mathbf{I} + \mathbf{A}(t_i)\,(t_i - t_{i-1}) \tag{1.44}$$

then one may reasonably expect that a carefully chosen approximation would be no worse than the many other approximations inherent in the EKF. One may also consider the same or simpler approximations when considering approximations to (1.42).

- Because there no way to prove global observability for a nonlinear system, the EKF may fail to converge from some initial conditions, even if the system is locally observable in particular neighborhoods.

In light of the above observations, we conclude this section by presenting a slightly improved version of the EKF as Algorithm 1.3. In subsequent chapters, we shall describe additional improvements to the EKF.