1 Filtering

1.1 Fully observed

Suppose we have noisy measurements $D = \{y_0, \dots y_M\}$ of the process x_t at discrete times t_i with Gaussian noise $\mathcal{N}(y_i|x_i, S)$, and let θ be the parameters of the system. The likelihood can be written as

$$L(D \mid \theta) = p(y_0, \dots, y_M|\theta) = p(y_0) \prod_{i=1}^{M} p(y_i|y_{i-1}, \dots y_0) = \prod_{i=0}^{M} L_i. \tag{1.1}$$

The factors can be written as

$$L_i = p(y_i|y_{i-1}, \dots y_0) = \int dx_i dx_{i-1} p(y_i|x_i, x_{i-1}) p(x_i, x_{i-1}|y_{i-1}, \dots y_0) \tag{1.2}$$

$$= \int dx_i dx_{i-1} p(y_i|x_i) p(x_i|x_{i-1}) p(x_{i-1}|y_{i-1}, \dots y_0) \tag{1.3}$$

$$= \int dx_i p(y_i|x_i) p(x_i|y_{i-1}, \dots y_0). \tag{1.4}$$

Here, $p(x_i|x_{i-1})$ fulfills the forward chemical master equation (CME). The *predictive distribution* $p(x_i|y_{i-1}, \dots y_0)$ is obtained from $p(x_i|y_{i-1}, \dots y_0) = \int dx_{i-1} p(x_i|x_{i-1}) p(x_{i-1}|y_{i-1}, \dots y_0)$, where $p(x_{i-1}|y_{i-1}, \dots y_0)$ is the *posterior* of the previous step. The posterior of the current step is obtained using Bayes rule as

$$p(x_i|y_i, \dots y_0) = \frac{p(y_i|x_i, y_{i-1}, \dots y_0) p(x_i|y_{i-1}, \dots y_0)}{p(y_i|y_{i-1}, \dots y_0)} \tag{1.5}$$

$$= \frac{p(y_i|x_i) p(x_i|y_{i-1}, \dots y_0)}{\int dx_i p(y_i|x_i) p(x_i|y_{i-1}, \dots y_0)} \tag{1.6}$$

$$= \frac{p(y_i|x_i) p(x_i|y_{i-1}, \dots y_0)}{L_i}, \tag{1.7}$$

and serves as the initial condition for the next step. The likelihood contribution L_i is just the normalization of the posterior.

In summary, the ith inference step comprises

i) Solve CME from $t_{i-1} \to t_i$ to obtain $p(x_i|x_{i-1})$ and compute $p(x_i|y_{i-1}, \dots y_0) = \int dx_{i-1} p(x_i|x_{i-1}) p(x_{i-1}|y_{i-1}, \dots y_0)$, where $p(x_{i-1}|y_{i-1}, \dots y_0)$ is the posterior of the previous step.

ii) Perform measurement (bayesian) update $p(y_i|x_i) p(x_i|y_{i-1}, \dots y_0)$.

iii) Compute

$$L_i = \int dx_i p(y_i|x_i)p(x_i|y_{i-1}, \dots y_0). \tag{1.8}$$

iv) Compute the posterior $p(x_i|y_i, \dots y_0) = p(y_i|x_i)p(x_i|y_{i-1}, \dots y_0)/L_i$ which is needed in i) of the next step.

1.2 LNA

For the linear noise approximation (LNA) this becomes (the subscripts "$-$" and denote "$+$" denote values before and after measurement updates, respectively)

i) Integrate LNA from $t_{i-1} \to t_i$ with initial conditions given by posterior $p(x_{i-1}|y_{i-1}, \dots y_0)$ of previous step (denote mean and variance of the latter by $\mu_{(i-1)+}$ and $\Sigma_{(i-1)+}$, respectively) to obtain μ_{i-} and Σ_{i-}. Equations:

$$\partial_t \mu = A, \tag{1.9}$$

$$\partial_t \Sigma_n = J \cdot \Sigma_n + \Sigma_n \cdot J^T + \tilde{D}, \tag{1.10}$$

$$g(\mu) = \Omega f(\phi = \mu/\Omega), \tag{1.11}$$

$$A = Sg(\mu), \tag{1.12}$$

$$J_{ij} = \partial_{\mu_j} A_i, \tag{1.13}$$

$$\tilde{D} = S\text{diag}(g)S^T. \tag{1.14}$$

ii) Perform measurement update $p(y_i|x_i)p(x_i|y_{i-1}, \dots y_0) = c_i \mathcal{N}_{x_i}(\mu_{i+}, \Sigma_{i+})$

$$c_i = \mathcal{N}_{\mu_{i-}}(y_i, \Sigma_{i-} + S), \tag{1.15}$$

$$\mu_{i+} = (\Sigma_{i-}{}^{-1} + S^{-1})^{-1}(\Sigma_-^{i\ -1}\mu_{i-} + S^{-1}y_i), \tag{1.16}$$

$$\Sigma_{i+} = (\Sigma_{i-}{}^{-1} + S^{-1})^{-1}. \tag{1.17}$$

iii) Compute

$$L_i = \int dx_i p(y_i|x_i)p(x_i|y_{i-1}, \dots y_0) = c_i \int dx_i \mathcal{N}_{x_i}(\mu_{i+}, \Sigma_{i+}) = c_i \tag{1.18}$$

iv) The posterior $p(x_i|y_i, \dots y_0) = p(y_i|x_i)p(x_i|y_{i-1}, \dots y_0)/L_i$ is needed for the next step. Since it is normalized, its simply a Gaussian with mean μ_{i+} and variance Σ_{i+}.

The full likelihood is then given by

$$L(D \,|\, \theta) = p(y_0) \prod_{i=1}^{M} p(y_i|y_{i-1}, \dots y_0) = \prod_{i=0}^{M} c_i. \tag{1.19}$$

1.2.1 Initial conditions with measurement noise

The initial conditions for a flat prior $p_0(n) = const.$ are

$$c_0 = \int dx_0 p(y_0|x_0)p(x_0) \tag{1.20}$$

$$= \int dx_0 p(y_0|x_0) = 1, \tag{1.21}$$

$$\mu_{0+} = y_0, \tag{1.22}$$

$$\Sigma_{0+} = S. \tag{1.23}$$

If we assume instead $p_0(x) = \mathcal{N}_x(y_0, S)$,

$$c_0 = \int dx_0 \mathcal{N}_{x_0}(y_0, S)\mathcal{N}_{x_0}(y_0, S) \tag{1.24}$$

$$= \mathcal{N}_{y_0}(y_0, 2S) \int dx_0 \mathcal{N}_{x_0}(\ldots) \tag{1.25}$$

$$= ((2\pi)^N \det(2S))^{-1/2} \tag{1.26}$$

$$= ((4\pi)^N \det(S))^{-1/2}, \tag{1.27}$$

$$\mu_{0+} = y_0, \tag{1.28}$$

$$\Sigma_{0+} = 2S. \tag{1.29}$$

1.2.2 Initial conditions without measurement noise

For a flat prior $p_0(n) = const.$ we get the inital conditions

$$c_0 = 1, \tag{1.30}$$

$$\mu_{0+} = y_0, \tag{1.31}$$

$$\Sigma_{0+} = 0, \tag{1.32}$$

and jump conditions

$$c_i = \lim_{S\to 0} \mathcal{N}_{\mu_{i-}}(y_i, \Sigma_{i-} + S) = \mathcal{N}_{\mu_{i-}}(y_i, \Sigma_{i-}), \tag{1.33}$$

$$\mu_{i+} = \lim_{S\to 0} ({\Sigma_{i-}}^{-1} + S^{-1})^{-1}({\Sigma_{i-}}^{-1}\mu_{i-} + S^{-1}y_i) = y_i, \tag{1.34}$$

$$\Sigma_{i+} = \lim_{S\to 0} ({\Sigma_{i-}}^{-1} + S^{-1})^{-1} = 0. \tag{1.35}$$

The initial conditions for a flat prior $p_0(n) = const.$ are

$$c_0 = \int dx_0 p(y_0|x_0)p(x_0) \tag{1.36}$$

$$= \int dx_0 p(y_0|x_0) = 1, \tag{1.37}$$

$$\mu_{0+} = y_0, \tag{1.38}$$

$$\Sigma_{0+} = 0. \tag{1.39}$$

1.3 Observation of projection of system

Suppose we have observations $y \in \mathbb{R}^m$of projections of the system $x \in \mathbb{R}^m$, given by $p(y_t|x_t) = \mathcal{N}(y_t|Px_t, S)$ with projection $P \in \mathbb{R}^{m\times n}$. For example, if only the first species is measured we have $P_{11} = 1$ and zero otherwise. Write the likelihood as

$$L(D\,|\,\theta) = p(y_0)\prod_{i=1}^{M} p(y_i|y_{i-1},\dots y_0) = \prod_{i=0}^{M} L_i. \tag{1.40}$$

The factors can be written as

$$L_i = p(y_i|y_{i-1},\dots y_0) = \int dx_i dx_{i-1} p(y_i|x_i)p(x_i|x_{i-1})p(x_{i-1}|y_{i-1},\dots y_0) \tag{1.41}$$

$$= \int dx_i dx_{i-1} p(y_i|x_i)p(x_i|x_{i-1})p(x_{i-1}|y_{i-1},\dots y_0) \tag{1.42}$$

$$= \int dx_i p(y_i|x_i)p(x_i|y_{i-1},\dots y_0). \tag{1.43}$$

The posterior is obtained from

$$p(x_i|y_i,\dots y_0) = \frac{p(y_i|x_i, y_{i-1},\dots y_0)p(x_i|y_{i-1},\dots y_0)}{p(y_i|y_{i-1},\dots y_0)} \tag{1.44}$$

$$= \frac{p(y_i|x_i)p(x_i|y_{i-1},\dots y_0)}{p(y_i|y_{i-1},\dots y_0)} \tag{1.45}$$

$$= \frac{\mathcal{N}(y_i|Px_i, S)p(x_i|y_{i-1},\dots y_0)}{p(y_i|y_{i-1},\dots y_0)}, \tag{1.46}$$

$$L_i = p(y_i|y_{i-1},\dots y_0) \tag{1.47}$$

$$= \int dx_i p(y_i|x_i)p(x_i|y_{i-1},\dots y_0). \tag{1.48}$$

1.4 Gaussian approximation

Suppose $p(x_i|y_{i-1},\dots y_0)$ is Gaussian with mean μ_{i-} and variance Σ_{i-}. The bayesian update thus becomes

$$p(x_i, y_i|y_{i-1},\dots y_0) = p(y_i|x_i, y_{i-1},\dots y_0)p(x_i|y_{i-1},\dots y_0) \tag{1.49}$$

$$= \mathcal{N}(y_i|Px_i, S)\mathcal{N}(x_i|\mu_{i-}, \Sigma_{i-}) \tag{1.50}$$

$$\sim \begin{bmatrix} \mu_{i-} \\ P\mu_{i-} \end{bmatrix} \begin{bmatrix} \Sigma_{i-} & \Sigma_{i-}P^T \\ P\Sigma_{i-} & P\Sigma_{i-}P^T + S \end{bmatrix}, \tag{1.51}$$

where the last step is obtained from explicit integrating. Marginalizing gives the LH contribution

$$
\begin{aligned}
L_i &= p(y_i|y_{i-1}, \dots y_0) && (1.52)\\
&= \int dx_i p(y_i|x_i) p(x_i|y_{i-1}, \dots y_0) && (1.53)\\
&= \int dx_i \mathcal{N}\Big(x_i, y_i| \begin{bmatrix} \mu_{i-} \\ P\mu_{i-} \end{bmatrix} \begin{bmatrix} \Sigma_{i-} & \Sigma_{i-}P^T \\ P\Sigma_{i-} & P\Sigma_{i-}P^T + S \end{bmatrix}\Big) && (1.54)\\
&= \mathcal{N}(y_i|P\mu_{i-}\ ,\ P\Sigma_{i-}P^T + S) && (1.55)
\end{aligned}
$$

Conditioning gives the posterior (see App. C)

$$
\begin{aligned}
p(x_i|y_i, \dots y_0) &= \mathcal{N}(x_i|\mu_{i+}, \Sigma_{i+}), && (1.56)\\
\mu_{i+} &= \mu_{i-} + \Sigma_{i-}P^T(P\Sigma_{i-}P^T + S)^{-1}(y_i - P\mu_{i-}), && (1.57)\\
\Sigma_{i+} &= \Sigma_{i-} - \Sigma_{i-}P^T(P\Sigma_{i-}P^T + S)^{-1}P\Sigma_{i-}. && (1.58)
\end{aligned}
$$

2 Smoothing

2.1 Fully observed

Suppose we have noisy measurements $D = \{y^0, \dots y^M\}$ of the process x_t at discrete times t^i with noise $\mathcal{N}(y_{t_i}|x_{t_i}, S)$. The system has the parameters θ. Suppose we want to compute the likelihood

$$L(D \,|\, \theta) = p(y_M) \prod_{i=0}^{M-1} p(y_{M-i-1}|y_{M-i}, \dots y_M) = \prod_{i=0}^{M} L_i. \tag{2.1}$$

Using $j = M - i$, the factors can be written as

$$L_i = L_{M-j} = p(y_{j-1}|y_j, \dots y_M) \tag{2.2}$$

$$= \int dx_j dx_{j-1} p(y_{j-1}|x_{j-1}) p(x_{j-1}|x_j) p(x_j|y_j, \dots y_M) \tag{2.3}$$

$$= \int dx_{j-1} p(y_{j-1}|x_{j-1}) p(x_{j-1}|y_j, \dots y_M). \tag{2.4}$$

Here, $p(x_{j-1}|x_j)$ fulfills the *backward* CME. Then $p(x_{j-1}|y_j, \dots y_M) = \int dx_j p(x_{j-1}|x_j) p(x_j|y_j, \dots y_M)$ is the solution with initial condition given by the posterior $p(x_j|y_j, \dots y_M)$ of the previous step. The posterior of the current step is obtained from

$$p(x_{j-1}|y_{j-1}, \dots y_M) = \frac{p(y_{j-1}|x_{j-1}, y_j, \dots y_M) p(x_{j-1}|y_j, \dots y_M)}{p(y_{j-1}|y_j, \dots y_M)} \tag{2.5}$$

$$= \frac{p(y_{j-1}|x_{j-1}) p(x_{j-1}|y_j, \dots y_M)}{L_{M-j}}. \tag{2.6}$$

2.2 LNA

The ith inference step comprises

i) Integrate backward LNA from $t_j \rightarrow t_{j-1}$ with initial conditions given by posterior $p(x_{j-1}|y_{j-1}, \dots y_M)$ of previous step (denote mean, variance and normalization of the latter by $\mu_{(j)-}$, $\Sigma_{(j)-}$ and $c_{(j)-}$, respectively) to obtain $\mu_{(j-1)+}$, $\Sigma_{(j-1)+}$ and $c_{(j-1)-}$, where we assume the solution to be of the form

$$p(x) = \frac{c}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2}((x-\mu)^T \Sigma^{-1} (x-\mu))}. \tag{2.7}$$

Equations:

$$\partial_t c = \text{tr}[J]c, \tag{2.8}$$
$$\partial_t \mu = A, \tag{2.9}$$
$$\partial_t \Sigma_n = J \cdot \Sigma_n + \Sigma_n \cdot J^T + \tilde{D}, \tag{2.10}$$
$$g(\mu) = \Omega f(\phi = \mu/\Omega), \tag{2.11}$$
$$A = Sg(\mu), \tag{2.12}$$
$$J_{ij} = \partial_{\mu_j} A_i, \tag{2.13}$$
$$\tilde{D} = S\text{diag}(g)S^T. \tag{2.14}$$

ii) measurement update $p(y_{j-1}|x_{j-1})p(x_{j-1}|y_j, \ldots y_M) = c_{(j-1)-}\mathcal{N}_{x_{j-1}}(\mu_{(j-1)-}, \Sigma_{(j-1)-})$

$$c_{(j-1)-} = c_{(j-1)+}\mathcal{N}_{\mu_{(j-1)+}}(y_{j-1}, \Sigma_{(j-1)+} + S), \tag{2.15}$$
$$\mu_{(j-1)-} = ({\Sigma_{(j-1)+}}^{-1} + S^{-1})^{-1}({\Sigma_{(j-1)+}}^{-1}\mu_{(j-1)+} + S^{-1}y_{j-1}), \tag{2.16}$$
$$\Sigma_{(j-1)-} = ({\Sigma_{(j-1)+}}^{-1} + S^{-1})^{-1}. \tag{2.17}$$

iii) Compute

$$L_i = \int dx_{j-1} p(y_{j-1}|x_{j-1})p(x_{j-1}|y_j, \ldots y_M) = c_{(j-1)-}\int dx_i \mathcal{N}_{x_{j-1}}(\mu_{(j-1)-}, \Sigma_{(j-1)-}) = c_{(j-1)-} \tag{2.18}$$

iv) The posterior $p(x_{j-1}|y_{j-1}, \ldots y_M) = p(y_{j-1}|x_{j-1})p(x_{j-1}|y_j, \ldots y_M)/L_{M-j}$ is needed for the next step. It is a Gaussian with mean and variance $\mu_{(j-1)-}$ and $\Sigma_{(j-1)-}$, respectivey.

2.3 Fully observed 2

Suppose we have noisy measurements $y = \{y^0, \ldots y^M\}$ of the process x_t at discete times t^i with noise $\mathcal{N}(y_{t_i}|x_{t_i}, S)$. The system has the parameters θ. Suppose we want to compute the likelihood

$$L(y \,|\, \theta) = \sum_{x_0} p(y_0, \ldots y_M|x_0, \theta)p(x_0). \tag{2.19}$$

The conditioned likelihood $p(y_0, \ldots y_M|x_0, \theta)$ can be computed as follows. Consider (we omit the conditioning on the parameters here)

$$p(y_t, \ldots y_M|x_t) = p(y_t|y_{t+1} \ldots y_M, x_t)p(y_{t+1}, \ldots y_M|x_t) \tag{2.20}$$
$$= p(y_t|x_t)\int dx_{t+1} p(x_{t+1}, y_{t+1}, \ldots y_M|x_t) \tag{2.21}$$
$$= p(y_t|x_t)\int dx_{t+1} p(y_{t+1}, \ldots y_M|x_{t+1}, x_t)p(x_{t+1}|x_t) \tag{2.22}$$
$$= p(y_t|x_t)\int dx_{t+1} p(y_{t+1}, \ldots y_M|x_{t+1})p(x_{t+1}|x_t). \tag{2.23}$$

Here, $p(x_{t+1}|x_t)$ and thus $p(y_{t+1}, \dots y_M|x_t)$ fulfills the backward master equation in x_t. If we define

$$r_t(x) = p(y_{\text{ceiling}(t)}, \dots y_M|x_t = x), \tag{2.24}$$

$r_t(x)$ fulfills the backward equation inbetween two measurements

$$\partial_t r_t(x) = \sum_{r=1}^{R} f_r(n)(r_t(x) - r_t(x + S_r)), \tag{2.25}$$

and the jump condition at measurement i

$$\lim_{t \to t_i -} r_t(x) = p(y_t|x_t) \lim_{t \to t_i +} r_t(x), \tag{2.26}$$

which we can write as

$$p(y_i, \dots y_M|x_i) = p(y_i|x_i)p(y_{i+1}, \dots y_M|x_i) \tag{2.27}$$

Starting with the end condition which is just given by the noise model $p(y_M|x_M)$ we can thus recursively compute the conditioned likelihood. The full likelihood is then given by

$$L(y\,|\,\theta) = \sum_{x_0} p(y_0, \dots y_M|x_0, \theta)p_0(x_0) \tag{2.28}$$

$$= \sum_{x} r_0(x)p_0(x). \tag{2.29}$$

2.4 LNA

The LNA solution of the backward equation is

$$r_t(x) = \frac{z}{\sqrt{(2\pi)^N \text{det}(\Sigma)}} e^{-\frac{1}{2}((x-\mu)^T \Sigma^{-1}(x-\mu))}, \tag{2.30}$$

$$z = e^T, \tag{2.31}$$

$$\mu = \Omega\phi, \tag{2.32}$$

where ϕ is the solution of the macroscopic rate equations and T is a function satisfying

$$\partial_t T = \text{tr}[J], \quad T = \int_{t_0}^{t} \text{tr}[J(t')]dt', \tag{2.33}$$

which means

$$\partial_t z = \text{tr}[J]z. \tag{2.34}$$

Σ_n satisfies

$$\partial_t \Sigma_n = J \cdot \Sigma_n + \Sigma_n \cdot J^T - \Omega D, \tag{2.35}$$

$$J_{ij} = S_{ir}\partial_{\phi_j} f_r(\phi), \tag{2.36}$$

$$D_{ij} = S_{ir}S_{jr}f_r(\phi). \tag{2.37}$$

2.4.1 Jump and end conditions

In Appendix A.1 we show that the jump and end conditions imply

$$\mu_M = y_M, \tag{2.38}$$
$$\Sigma_M = S, \tag{2.39}$$
$$z_M = 1, \tag{2.40}$$

and

$$z_{i-} = z_{i+}\mathcal{N}_{\mu_{i+}}(y_i, \Sigma_{i+} + S), \tag{2.41}$$
$$\mu_{i-} = ({\Sigma_{i+}}^{-1} + S^{-1})^{-1}({\Sigma_{i+}}^{-1}\mu_{i+} + S^{-1}y_i), \tag{2.42}$$
$$\Sigma_{i-} = ({\Sigma_{i+}}^{-1} + S^{-1})^{-1}. \tag{2.43}$$

Using (App. C)

$$(A^{-1} + B^{-1})^{-1} = A(B + A)^{-1}B, \tag{2.44}$$
$$(A^{-1} + B^{-1})^{-1}(A^{-1}a + B^{-1}b) = B(B + A)^{-1}a + A(B + A)^{-1}b, \tag{2.45}$$

the update equations can be simplified as

$$z_{i-} = z_{i+}\mathcal{N}_{\mu_{i+}}(y_i, \Sigma_{i+} + S), \tag{2.46}$$
$$\mu_{i-} = S(\Sigma_{i+} + S)^{-1}\mu_{i+} + \Sigma_{i+}(\Sigma_{i+} + S)^{-1}y_i, \tag{2.47}$$
$$\Sigma_{i-} = \Sigma_{i+}(\Sigma_{i+} + S)^{-1}S. \tag{2.48}$$

2.4.2 Final full likelihood

For a flat prior $p_0(x) = const.$ the likelihood is given by

$$L(D\,|\,\theta) = \sum_{x_0} p(y_0, \dots y_M|x_0, \theta)p(x_0) \tag{2.49}$$
$$= \int dx z_{0+}\mathcal{N}(x|...) \tag{2.50}$$
$$= z_{0+}\int dx\mathcal{N}(x|...) \tag{2.51}$$
$$= z_{0+}. \tag{2.52}$$

If we assume instead $p_0(x) = \mathcal{N}_x(y_0, S)$,

$$L(D\,|\,\theta) = \sum_{x_0} p(y_0, \dots y_M|x_0, \theta)p(x_0) \tag{2.53}$$
$$= \int dx z_{0+}\mathcal{N}(x|y_0, S)\mathcal{N}(x|\mu_{0-}, \Sigma_{0-}) \tag{2.54}$$
$$= \mathcal{N}(\mu_{0-}|y_0, S + \Sigma_{0-})\int dx\mathcal{N}(x|...) \tag{2.55}$$
$$= \mathcal{N}(\mu_{0-}|y_0, S + \Sigma_{0-}). \tag{2.56}$$

2.5 Without measurement noise

Without measurement noise the final and jump conditions are

$$\mu_M = y_M, \tag{2.57}$$
$$\Sigma_M = 0, \tag{2.58}$$
$$z_M = 1, \tag{2.59}$$

and

$$z_{i-} = z_{i+}\mathcal{N}_{\mu_{i+}}(y_i, \Sigma_{i+}), \tag{2.60}$$
$$\mu_{i-} = y_i, \tag{2.61}$$
$$\Sigma_{i-} = 0. \tag{2.62}$$

A flat prior $p_0(x) = const.$ gives the likelihood

$$L(D \,|\, \theta) = \int dx z_{0+}\delta(x - \mu_{0-}) \tag{2.63}$$
$$= z_{0+}. \tag{2.64}$$

2.6 Projection observed

Suppose we have observations $y \in \mathbb{R}^m$of projections of the system $x \in \mathbb{R}^m$, given by $p(y_t|x_t) = \mathcal{N}(y_t|Px_t, S)$ with projection $P \in \mathbb{R}^{m\times n}$ Only the update (and final) conditions change. The bayesian update becomes

$$p(y_i, \ldots y_M|x_i) = p(y_i|x_i)p(y_{i+1}, \ldots y_M|x_i) \tag{2.65}$$
$$= \mathcal{N}(y_i|Px_i, S)z_{i+}\mathcal{N}(x_i|\mu_{i+}, \Sigma_{i+}) \tag{2.66}$$
$$\sim z_{i+}\mathcal{N}\begin{bmatrix}\mu_{i-}\\ P\mu_{i-}\end{bmatrix}\begin{bmatrix}\Sigma_{i+} & \Sigma_{i+}P^T\\ P\Sigma_{i+} & P\Sigma_{i+}P^T + S\end{bmatrix} \tag{2.67}$$

The jump conditions are

$$z_{i-} = z_{i+}\mathcal{N}(y_i|P\mu_{i-}\ ,\ P\Sigma_{i-}P^T + S), \tag{2.68}$$
$$\mu_{i-} = \mu_{i+} + \Sigma_{i+}P^T(P\Sigma_{i+}P^T + S)^{-1}(y_i - P\mu_{i+}), \tag{2.69}$$
$$\Sigma_{i-} = \Sigma_{i+} - \Sigma_{i+}P^T(P\Sigma_{i+}P^T + S)^{-1}P\Sigma_{i+}. \tag{2.70}$$