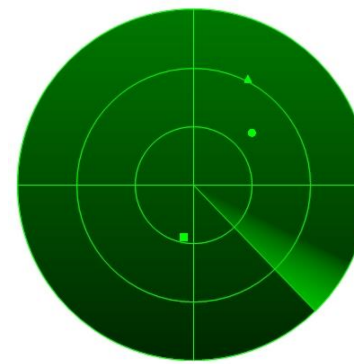
# Kalman Smoothing

Jur van den Berg

# Kalman Filtering vs. Smoothing

- Dynamics and Observation model

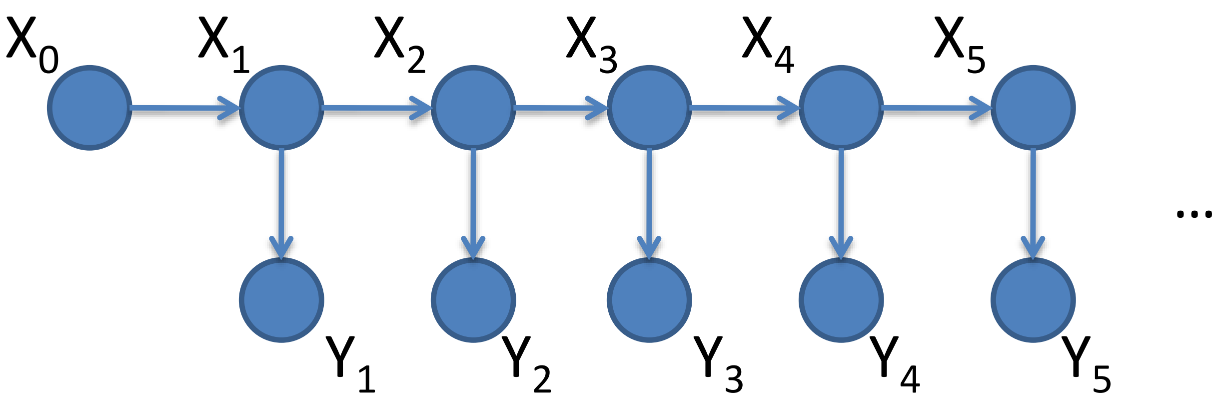
$$X_{t+1} = AX_t + W_t, \quad W_t = N(\mathbf{0}, Q)$$
$$Y_t = CX_t + V_t, \quad V_t = N(\mathbf{0}, R)$$



- Kalman Filter:
  - Compute $\left(X_t \mid Y_0 = \mathbf{y}_0, \ldots, Y_t = \mathbf{y}_t\right)$
  - Real-time, given data so far
- Kalman Smoother:
  - Compute $\left(X_t \mid Y_0 = \mathbf{y}_0, \ldots, Y_T = \mathbf{y}_T\right), \quad t < T$
  - Post-processing, given all data

# Kalman Filtering Recap

- Time update
  - $X_{t+1|t} = AX_{t|t} + W_t$
- Measurement update:
  - $Y_{t+1|t} = CX_{t+1|t} + V_{t+1}$
  - Compute joint distribution $(X_{t+1|t}, Y_{t+1|t})$
  - Compute conditional $X_{t+1|t+1} = (X_{t+1|t} \mid Y_{t+1|t} = \mathbf{y}_{t+1})$



# Kalman filter summary

- Model: $$X_{t+1} = AX_t + W_t, \quad W_t = N(\mathbf{0}, Q)$$
  $$Y_t = CX_t + V_t, \quad V_t = N(\mathbf{0}, R)$$
- Algorithm: repeat…
  - Time update: $$\hat{\mathbf{x}}_{t+1|t} = A\hat{\mathbf{x}}_{t|t}$$
    $$P_{t+1|t} = AP_{t|t}A^T + Q$$
  - Measurement update:
    $$\begin{aligned} K_{t+1} &= P_{t+1|t}C^T\left(CP_{t+1|t}C^T + R\right)^{-1} \\ \hat{\mathbf{x}}_{t+1|t+1} &= \hat{\mathbf{x}}_{t+1|t} + K_{t+1}\left(\mathbf{y}_{t+1} - C\hat{\mathbf{x}}_{t+1|t}\right) \\ P_{t+1|t+1} &= P_{t+1|t} - K_{t+1}CP_{t+1|t} \end{aligned}$$

# Kalman Smoothing

- Input: initial distribution $X_0$ and data $\mathbf{y}_1, \ldots, \mathbf{y}_T$
- Algorithm: forward-backward pass (Rauch-Tung-Striebel algorithm)
- Forward pass:
  - Kalman filter: compute $X_{t+1|t}$ and $X_{t+1|t+1}$ for $0 \le t < T$
- Backward pass:
  - Compute $X_{t|T}$ for $0 \le t < T$
  - Reverse "horizontal" arrow in graph

# Backward Pass

- Compute $X_{t|T}$ given $X_{t+1|T} = N(\hat{\mathbf{x}}_{t+1|T}, P_{t+1|T})$
- Reverse arrow: $X_{t|t} \rightarrow X_{t+1|t}$
- Same as incorporating measurement in filter
  - 1. Compute joint ($X_{t|t}$, $X_{t+1|t}$)
  - 2. Compute conditional ($X_{t|t}$ | $X_{t+1|t}$ = $\mathbf{x}_{t+1}$)
- New: $\mathbf{x}_{t+1}$ is not “known”, we only know its distribution: $\mathbf{x}_{t+1} \sim X_{t+1|T}$
  - 3. “Uncondition” on $\mathbf{x}_{t+1}$ to compute $X_{t|T}$ using laws of total expectation and variance

# Backward pass. Step 1

- Compute joint distribution of $X_{t|t}$ and $X_{t+1|t}$:

$$\left(X_{t|t}, X_{t+1|t}\right) = N\left(\begin{pmatrix} \mathrm{E}\left(X_{t|t}\right) \\ \mathrm{E}\left(X_{t+1|t}\right) \end{pmatrix}, \begin{pmatrix} \mathrm{Var}\left(X_{t|t}\right) & \mathrm{Cov}\left(X_{t|t}, X_{t+1|t}\right) \\ \mathrm{Cov}\left(X_{t+1|t}, X_{t|t}\right) & \mathrm{Var}\left(X_{t+1|t}\right) \end{pmatrix}\right)$$

$$= N\left(\begin{pmatrix} \hat{\mathbf{x}}_{t|t} \\ \hat{\mathbf{x}}_{t+1|t} \end{pmatrix}, \begin{pmatrix} P_{t|t} & P_{t|t}A^T \\ AP_{t|t} & P_{t+1|t} \end{pmatrix}\right)$$

where

$$\begin{aligned} \mathrm{Cov}\left(X_{t+1|t}, X_{t|t}\right) &= \mathrm{Cov}\left(AX_{t|t} + W_t, X_{t|t}\right) \\ &= A\,\mathrm{Cov}\left(X_{t|t}, X_{t|t}\right) + \mathrm{Cov}\left(W_t, X_{t|t}\right) \\ &= A\,\mathrm{Var}\left(X_{t|t}\right) \\ &= AP_{t|t} \end{aligned}$$

# Backward pass. Step 2

- Recall that if

$$(Z_1, Z_2) = N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right)$$

then

$$(Z_1 \mid Z_2 = \mathbf{z}_2) = N\left(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{z}_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- Compute $(X_{t|t} | X_{t+1|t} = \mathbf{x}_{t+1})$:

$$\begin{aligned}(X_{t|t} \mid X_{t+1|t} = \mathbf{x}_{t+1}) \quad = \quad & N\Big(\hat{\mathbf{x}}_{t|t} + P_{t|t}A^T P_{t+1|t}{}^{-1}(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}),\\ & P_{t|t} - P_{t|t}A^T P_{t+1|t}{}^{-1} A P_{t|t}\Big)\end{aligned}$$

# Backward pass Step 3

- Conditional only valid for *given* $\mathbf{x}_{t+1}$.

$$\left(X_{t|t} \mid X_{t+1|t} = \mathbf{x}_{t+1}\right) = N\left(\hat{\mathbf{x}}_{t|t} + P_{t|t} A^T P_{t+1|t}^{-1}\left(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}\right), P_{t|t} - P_{t|t} A^T P_{t+1|t}^{-1} A P_{t|t}\right)$$

$$= N\left(\hat{\mathbf{x}}_{t|t} + L_t\left(\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1|t}\right), P_{t|t} - L_t P_{t+1|t} L_t^T\right)$$

  – Where $L_t = P_{t|t} A^T P_{t+1|t}^{-1}$

- But we don't know its value, but only its distribution: $\mathbf{x}_{t+1} \sim X_{t+1|T}$
- Uncondition on $\mathbf{x}_{t+1}$ to compute $X_{t|T}$ using law of total expectation and law of total variance

# Law of total expectation/variance

- Law of total expectation:
  - $E(X) = E_Z( E(X|Y = Z) )$
- Law of total variance:
  - $Var(X) = E_Z( Var(X|Y = Z) ) + Var_Z( E(X|Y = Z) )$
- Compute $X_{t|T} = N(E(X_{t|T}), \mathrm{Var}(X_{t|T}))$
  - where

$$E(X_{t|T}) = E_{X_{t+1|T}}\left(E(X_{t|t} \mid X_{t+1|t} = X_{t+1|T})\right)$$

$$\begin{aligned}\mathrm{Var}(X_{t|T}) &= E_{X_{t+1|T}}\left(\mathrm{Var}(X_{t|t} \mid X_{t+1|t} = X_{t+1|T})\right) + \\ &\quad \mathrm{Var}_{X_{t+1|T}}\left(E(X_{t|t} \mid X_{t+1|t} = X_{t+1|T})\right)\end{aligned}$$

# Unconditioning

- Recall from step 2 that

$$E(X_{t|t} \mid X_{t+1|t} = X_{t+1|T}) = \hat{\mathbf{x}}_{t|t} + L_t\left(X_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}\right)$$

$$\mathrm{Var}(X_{t|t} \mid X_{t+1|t} = X_{t+1|T}) = P_{t|t} - L_t P_{t+1|t} L_t^T$$

- So,

$$\begin{aligned} E(X_{t|T}) &= E_{X_{t+1|T}}\left(E(X_{t|t} \mid X_{t+1|t} = X_{t+1|T})\right) \\ &= \hat{\mathbf{x}}_{t|t} + L_t\left(\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}\right) \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X_{t|T}) &= E_{X_{t+1|T}}\left(\mathrm{Var}(X_{t|t} \mid X_{t+1|t} = X_{t+1|T})\right) + \\ &\quad \mathrm{Var}_{X_{t+1|T}}\left(E(X_{t|t} \mid X_{t+1|t} = X_{t+1|T})\right) \\ &= P_{t|t} - L_t P_{t+1|t} L_t^T + L_t P_{t+1|T} L_t^T \\ &= P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t}) L_t^T \end{aligned}$$

# Backward pass

- Summary:

$$
\begin{aligned}
L_t &= P_{t|t} A^T P_{t+1|t}{}^{-1} \\
\hat{\mathbf{x}}_{t|T} &= \hat{\mathbf{x}}_{t|t} + L_t \left( \hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t} \right) \\
P_{t|T} &= P_{t|t} + L_t (P_{t+1|T} - P_{t+1|t}) L_t^T
\end{aligned}
$$

# Kalman smoother algorithm

- for (t = 0; t < T; ++t)        // Kalman filter

$$\hat{\mathbf{x}}_{t+1|t} = A\hat{\mathbf{x}}_{t|t}$$

$$P_{t+1|t} = AP_{t|t}A^T + Q$$

$$\begin{aligned}
K_{t+1} &= P_{t+1|t}C^T\left(CP_{t+1|t}C^T + R\right)^{-1} \\
\hat{\mathbf{x}}_{t+1|t+1} &= \hat{\mathbf{x}}_{t+1|t} + K_{t+1}\left(\mathbf{y}_{t+1} - C\hat{\mathbf{x}}_{t+1|t}\right) \\
P_{t+1|t+1} &= P_{t+1|t} - K_{t+1}CP_{t+1|t}
\end{aligned}$$

- for (t = T – 1; t ≥ 0; --t)     // Backward pass

$$\begin{aligned}
L_t &= P_{t|t}A^T P_{t+1|t}{}^{-1} \\
\hat{\mathbf{x}}_{t|T} &= \hat{\mathbf{x}}_{t|t} + L_t\left(\hat{\mathbf{x}}_{t+1|T} - \hat{\mathbf{x}}_{t+1|t}\right) \\
P_{t|T} &= P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t^T
\end{aligned}$$

# Conclusion

- Kalman smoother can in used a post-processing
- Use $\underline{\mathbf{x}}_{t|T}$'s as optimal estimate of state at time t, and use $P_{t|T}$ as a measure of uncertainty.

# Extensions

- Automatic parameter (Q and R) fitting using EM-algorithm
  - Use Kalman Smoother on “training data” to learn Q and R (and A and C)