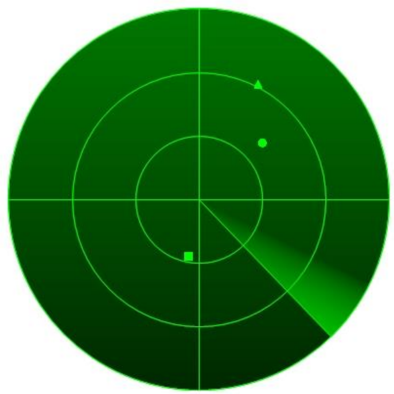
Kalman Filtering

Jur van den Berg

Kalman Filtering

- (Optimal) estimation of the (hidden) **state** of a linear dynamic process of which we obtain noisy (partial) **measurements**
- Example: radar tracking of an airplane.
  What is the state of an airplane given noisy radar measurements of the airplane's position?



Model

- Discrete time steps, continuous state-space
- (Hidden) state: $\mathbf{x}_t$, measurement: $\mathbf{y}_t$
- Airplane example: $$\mathbf{x}_t = \begin{pmatrix} x_t \\ \dot{x}_t \\ \ddot{x}_t \end{pmatrix}, \quad \mathbf{y}_t = \left(\tilde{x}_t \right)$$
- Position, speed and acceleration

Dynamics and Observation model

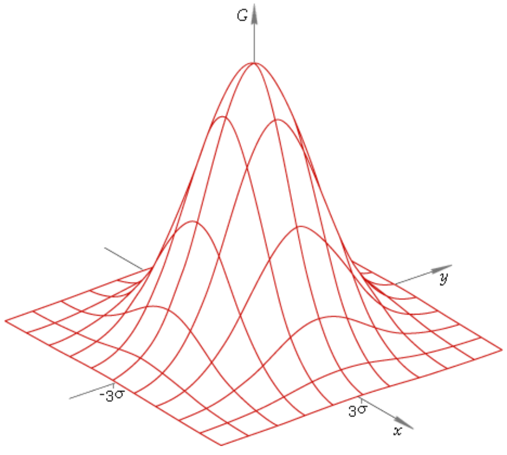
- Linear dynamics **model** describes relation between the state and the next state, and the observation:

$$
\begin{aligned}
\mathbf{x}_{t+1} &= A\mathbf{x}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim W_t = N(\mathbf{0}, Q) \\
\mathbf{y}_t &= C\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim V_t = N(\mathbf{0}, R)
\end{aligned}
$$

- Airplane example (if process has time-step δ):

$$
A = \begin{pmatrix} 1 & \delta & \frac{1}{2}\delta^2 \\ 0 & 1 & \delta \\ 0 & 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}
$$

Normal distributions



- Let X_0 be a normal distribution of the initial state $\mathbf{x}_0$
- Then, every X_t is a normal distribution of hidden state $\mathbf{x}_t$. Recursive definition:

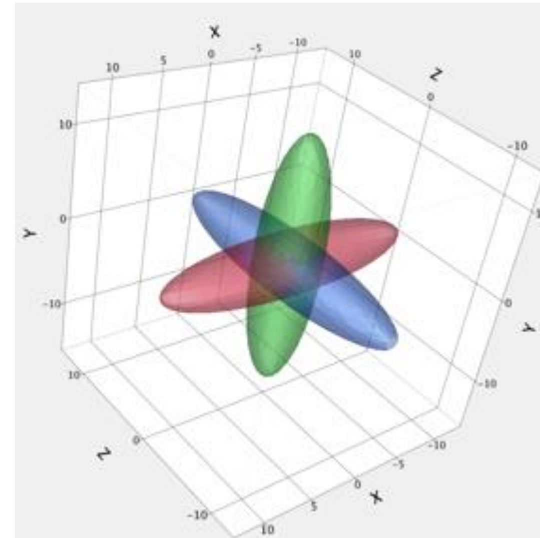
$$X_{t+1} = AX_t + W_t$$

- And every Y_t is a normal distribution of observation $\mathbf{y}_t$. Definition:

$$Y_t = CX_t + V_t$$

- **Goal of filtering**: compute conditional distribution $\left(X_t \mid Y_0 = \mathbf{y}_0, \ldots, Y_t = \mathbf{y}_t\right)$

Normal distribution



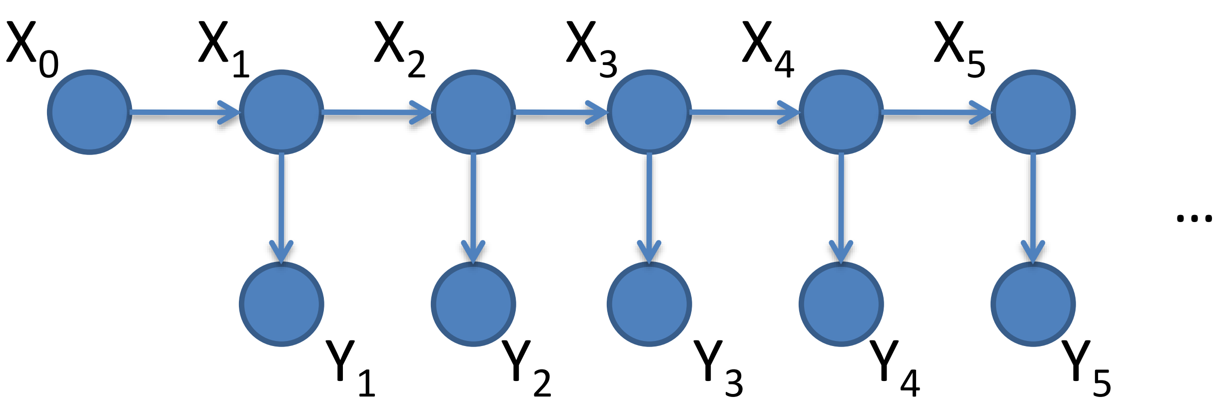
- Because X_t's and Y_t's are normal distributions, $\left(X_t \mid Y_0 = \mathbf{y}_0, \ldots, Y_t = \mathbf{y}_t\right)$ is also a normal distribution
- Normal distribution is fully specified by mean and covariance
- We denote:

$$
\begin{aligned}
X_{t|s} &= \left(X_t \mid Y_0 = \mathbf{y}_0, \ldots, Y_s = \mathbf{y}_s\right) \\
&= N\left(\mathrm{E}\left(X_t \mid Y_0 = \mathbf{y}_0, \ldots, Y_s = \mathbf{y}_s\right), \mathrm{Var}\left(X_t \mid Y_0 = \mathbf{y}_0, \ldots, Y_s = \mathbf{y}_s\right)\right) \\
&= N\left(\hat{\mathbf{x}}_{t|s}, P_{t|s}\right)
\end{aligned}
$$

Problem reduces to computing $\underline{\mathbf{x}}_{t|t}$ and $P_{t|t}$

Recursive update of state

- Kalman filtering algorithm: repeat…
  - Time update:
    from $X_{t|t}$, compute **a priori** distrubution $X_{t+1|t}$
  - Measurement update:
    from $X_{t+1|t}$ (and given $\mathbf{y}_{t+1}$), compute **a posteriori** distribution $X_{t+1|t+1}$



Time update

- From $X_{t|t}$, compute **a priori** distribution $X_{t+1|t}$:

$$
\begin{aligned}
X_{t+1|t} &= AX_{t|t} + W_t \\
&= N\left(\mathrm{E}\left(AX_{t|t} + W_t\right), \mathrm{Var}\left(AX_{t|t} + W_t\right)\right) \\
&= N\left(A\,\mathrm{E}\left(X_{t|t}\right) + \mathrm{E}\left(W_t\right), A\,\mathrm{Var}\left(X_{t|t}\right)A^T + \mathrm{Var}\left(W_t\right)\right) \\
&= N\left(A\hat{\mathbf{x}}_{t|t}, AP_{t|t}A^T + Q\right)
\end{aligned}
$$

- So,

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

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

Measurement update

- From $X_{t+1|t}$ (and given $\mathbf{y}_{t+1}$), compute $X_{t+1|t+1}$.

- 1. Compute **a priori** distribution of the observation $Y_{t+1|t}$ from $X_{t+1|t}$:

$$
\begin{aligned}
Y_{t+1|t} &= CX_{t+1|t} + V_{t+1} \\
&= N\left(\mathrm{E}\left(CX_{t+1|t} + V_{t+1}\right), \mathrm{Var}\left(CX_{t+1|t} + V_{t+1}\right)\right) \\
&= N\left(C\,\mathrm{E}\left(X_{t+1|t}\right) + \mathrm{E}\left(V_{t+1}\right), C\,\mathrm{Var}\left(X_{t+1|t}\right)C^T + \mathrm{Var}\left(V_{t+1}\right)\right) \\
&= N\left(C\hat{\mathbf{x}}_{t+1|t}, CP_{t+1|t}C^T + R\right)
\end{aligned}
$$

Measurement update (cont'd)

- 2. Look at joint distribution of $X_{t+1|t}$ and $Y_{t+1|t}$:

$$
\begin{aligned}
\left(X_{t+1|t}, Y_{t+1|t}\right) &= N\left(\begin{pmatrix} \mathrm{E}\left(X_{t+1|t}\right) \\ \mathrm{E}\left(Y_{t+1|t}\right) \end{pmatrix}, \begin{pmatrix} \mathrm{Var}\left(X_{t+1|t}\right) & \mathrm{Cov}\left(X_{t+1|t}, Y_{t+1|t}\right) \\ \mathrm{Cov}\left(Y_{t+1|t}, X_{t+1|t}\right) & \mathrm{Var}\left(Y_{t+1|t}\right) \end{pmatrix}\right) \\
&= N\left(\begin{pmatrix} \hat{\mathbf{x}}_{t+1|t} \\ C\hat{\mathbf{x}}_{t+1|t} \end{pmatrix}, \begin{pmatrix} P_{t+1|t} & P_{t+1|t}C^T \\ CP_{t+1|t} & CP_{t+1|t}C^T + R \end{pmatrix}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{Cov}\left(Y_{t+1}, X_{t+1|t}\right) &= \mathrm{Cov}\left(CX_{t+1|t} + V_{t+1}, X_{t+1|t}\right) \\
&= C\,\mathrm{Cov}\left(X_{t+1|t}, X_{t+1|t}\right) + \mathrm{Cov}\left(V_{t+1}, X_{t+1|t}\right) \\
&= C\,\mathrm{Var}\left(X_{t+1|t}\right) \\
&= CP_{t+1|t}
\end{aligned}
$$

Measurement update (cont'd)

- Recall from undergrad that if

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

then

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

- 3. Compute $X_{t+1|t+1} = (X_{t+1|t} | Y_{t+1|t} = \mathbf{y}_{t+1})$:

$$\begin{aligned} X_{t+1|t+1} &= \left(X_{t+1|t} \mid Y_{t+1|t} = \mathbf{y}_{t+1}\right) \\ &= N\left(\hat{\mathbf{x}}_{t+1|t} + P_{t+1|t}C^T\left(CP_{t+1|t}C^T + R\right)^{-1}\left(\mathbf{y}_{t+1} - C\hat{\mathbf{x}}_{t+1|t}\right),\right. \\ &\qquad \left. P_{t+1|t} - P_{t+1|t}C^T\left(CP_{t+1|t}C^T + R\right)^{-1}CP_{t+1|t}\right) \end{aligned}$$

Measurement update (cont'd):

- Often written in terms of **Kalman gain** matrix:

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

Kalman filter summary

- Model: $$\begin{aligned} \mathbf{x}_{t+1} &= A\mathbf{x}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim W_t = N(\mathbf{0}, Q) \\ \mathbf{y}_t &= C\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim V_t = N(\mathbf{0}, R) \end{aligned}$$
- 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:

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

Initialization

- Choose distribution of initial state by picking $\underline{\mathbf{x}}_0$ and P_0
- Start with measurement update given measurement $\mathbf{y}_0$
- Choice for Q and R (identity)
  - small Q: dynamics “trusted” more
  - small R: measurements “trusted” more

Conclusion

- Kalman filter can be used in real time
- 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

- Dynamic process with known **control input**
- **Non-linear** dynamic process
- **Kalman smoothing**: compute optimal estimate of state $\mathbf{x}_t$ given all data $\mathbf{y}_1, \ldots, \mathbf{y}_T$, with $T > t$ (not real-time).
- Automatic parameter (Q and R) fitting using EM-algorithm