STAT 576 Bayesian Analysis

Lecture 10: State-space Models and Sequential Monte Carlo I

Chenceng Cai

Washington State University

State-space Models

The **state-space model** is a general framework for modeling time series data. It consists of two components:

- The **state equation**: describes the evolution of the latent state variables over time.
- The **observation equation**: describes the relationship between the latent state variables and the observed data.
- The state-space model is also known as the **hidden Markov model (HMM)** when the state space is finite and the process is Markovian.

State-space Models

$$\begin{tikzcd}
 & Y_1 & Y_2 & & Y_t & & Y_{T-1} & Y_T \\
X_0 \arrow[r, "f_1"] & X_1 \arrow[u, "g_1"] \arrow[r, "f_2"] & X_2 \arrow[u, "g_2"] \arrow[r, "f_3"] & \cdots \arrow[r, "f_t"] & X_t \arrow[u, "g_t"] \arrow[r, "f_{t+1}"] & \cdots \arrow[r, "f_{T-1}"] & X_{T-1} \arrow[u, "g_{T-1}"] \arrow[r, "f_T"] & X_T \arrow[u, "g_T"]
\end{tikzcd}$$

- Observed data: $\boldsymbol{Y} = (Y_1, \ldots, Y_T)$
- Latent states: $\boldsymbol{X} = (X_0, X_1, \ldots, X_T)$
- The state equation:

$$p(X_0) = f_0(X_0), \quad p(X_t \mid \boldsymbol{X}_{t-1}) = f_t(X_t \mid \boldsymbol{X}_{t-1})$$

- The observation equation:

$$p(Y_t \mid \boldsymbol{X}_t) = g_t(Y_t \mid X_t)$$

State-space Model

- If the state equation satisfies

$$p(X_t \mid \boldsymbol{X}_{t-1}) = p(X_t \mid X_{t-1})$$

then the state-space model is **Markovian**.

- The (Markovian) state-space model is **linear** if

$$\mathbb{E}[X_t \mid X_{t-1}] = \boldsymbol{A}_t X_{t-1}$$

and

$$\mathbb{E}[Y_t \mid X_t] = \boldsymbol{B}_t X_t,$$

for some matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$.

- The (Markovian) state-space model is **linear Gaussian** if

$$X_t \mid X_{t-1} \sim \mathcal{N}(\boldsymbol{A}_t X_{t-1}, \boldsymbol{\Sigma}_t) \text{ and } Y_t \mid X_t \sim \mathcal{N}(\boldsymbol{A}_t X_t, \boldsymbol{R}_t)$$

Example: Object Tracking

- Consider the problem that tracks the position of an object moving in a 2D plane.
- The data contains the observed positions (with noise) of the object at different time points. $Y_t = (a_t, b_t)^T$.
- We can assume the latent states $X_t = (x_t, y_t)$, the true positions of the object.
- The observation equation is

$$Y_t = X_t + \epsilon_t$$

  where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R})$. $\boldsymbol{R}$ is the accuracy of the sensor.
- For the latent states X_t, we can assume a linear Gaussian model (random walk):

$$X_t = X_{t-1} + \eta_t,$$

  where $\eta_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{\Sigma}$ is the process noise.

Example: Object Tracking

The previous model has a continuous path, but quite stochastic velocities. We can add a velocity component to the model to stablize the dynamics.

- The latent states $X_t = (x_t, y_t, v_t, u_t)$, where (x_t, y_t) is the position and (v_t, u_t) is the velocity.
- The observation equation is

$$Y_t = (x_t, y_t)^T + \epsilon_t$$

  where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R})$.
- The state equation is

$$\begin{aligned} x_t &= x_{t-1} + v_{t-1} \\ y_t &= y_{t-1} + u_{t-1} \\ v_t &= v_{t-1} + \eta_t \\ u_t &= u_{t-1} + \xi_t, \end{aligned}$$

  where $\eta_t, \xi_t \sim \mathcal{N}(0, \sigma^2)$.

Example: Object Tracking

The previous model is a linear Gaussian model. We can write it in the matrix form:

$$
\begin{aligned}
\boldsymbol{X}_t &= \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{\eta}_t \\
\boldsymbol{Y}_t &= \boldsymbol{B}\boldsymbol{X}_t + \boldsymbol{\epsilon}_t,
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{A} &= \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
\boldsymbol{B} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \\
\boldsymbol{\epsilon}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}) \\
\boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).
\end{aligned}
$$

The Probabilities

The state-space model is a full probabilistic model.

- The joint distribution of the latent states and the observed data is

$$p(\boldsymbol{X}, \boldsymbol{Y}) = p(X_0) \prod_{t=1}^{T} p(X_t \mid \boldsymbol{X}_{t-1}) p(Y_t \mid X_t) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t)$$

- The joint distribution of the latent states is

$$p(\boldsymbol{X}) = p(X_0) \prod_{t=1}^{T} p(X_t \mid \boldsymbol{X}_{t-1}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1})$$

- The joint distribution of the observed data is

$$p(\boldsymbol{Y}) = \int p(\boldsymbol{X}, \boldsymbol{Y})\, d\boldsymbol{X} = \int f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t) d\boldsymbol{X}$$

Bayesian Framework

- The prior:

$$p(\boldsymbol{X}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1})$$

- The likelihood:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \prod_{t=1}^{T} g_t(Y_t \mid X_t)$$

- The posterior:

$$p(\boldsymbol{X} \mid \boldsymbol{Y}) = \frac{p(\boldsymbol{X}, \boldsymbol{Y})}{p(\boldsymbol{Y})} \propto p(\boldsymbol{X}, \boldsymbol{Y}) = f_0(X_0) \prod_{t=1}^{T} f_t(X_t \mid \boldsymbol{X}_{t-1}) g_t(Y_t \mid X_t)$$

Direct sampling from this posterior distribution can be difficult. We need to utilize the **sequential** structure of the model.

The Sequential Structure

Suppose we are at time t.

- We have observed data $Y_1, \ldots, Y_t$.
- We have the latent states $X_0, \ldots, X_t$.
- The **sequential** joint prior for the latent states is

$$p(\boldsymbol{X}_t) = f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1})$$

- The **sequential** likelihood for the observed data (so far) is

$$p(\boldsymbol{Y}_t \mid \boldsymbol{X}_t) = \prod_{s=1}^{t} g_t(Y_t \mid X_t)$$

- The **sequential** posterior for the latent states up to time t is (also called the **filtering** distribution)

$$p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t) \propto f_t(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1}) g_t(Y_t \mid X_t)$$

The Sequential Structure

At time t,

- The **predictive** distribution for the latent state at time $t+1$ is
$$p(X_{t+1} \mid \boldsymbol{Y}_t) = \int p(X_{t+1} \mid X_t) p(X_t \mid \boldsymbol{Y}_t) \, dX_t$$
- The joint distribution of the latent states up to time $t+1$ is
$$\begin{aligned} p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t) &= p(X_{t+1} \mid \boldsymbol{Y}_t) p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t) \\ &\propto f_{t+1}(X_{t+1} \mid \boldsymbol{X}_t) f_0(X_0) \prod_{s=1}^{t} f_s(X_s \mid \boldsymbol{X}_{s-1}) g_t(Y_t \mid X_t) \end{aligned}$$
- The incremental likelihood for the observed data at time $t+1$ is
$$p(Y_{t+1} \mid \boldsymbol{X}_{t+1}) = g_{t+1}(Y_{t+1} \mid X_{t+1})$$
- The filtering distribution for the latent states up to time $t+1$ is
$$p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_{t+1}) \propto p(Y_{t+1} \mid \boldsymbol{X}_t) p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t) \propto f_0(X_0) \prod_{s=1}^{t+1} f_s(X_s \mid \boldsymbol{X}_{s-1}) g_t(Y_t \mid X_t)$$

The Sequential Structure

The sequential structure of the state-space model allows us to update the latent states one by one.

- $p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t)$ is the prior
- $p(Y_{t+1} \mid X_{t+1})$ is the likelihood
- $p(\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_{t+1})$ is the posterior

A rudiment of sequential Monte Carlo:

- If we have a sample from $\boldsymbol{X}_t \mid \boldsymbol{Y}_t$.
- We can draw a sample from $\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_t$ by drawing X_{t+1} from $p(X_{t+1} \mid \boldsymbol{X}_t)$.
- We can update the sample to $\boldsymbol{X}_{t+1} \mid \boldsymbol{Y}_{t+1}$ by adjusting its weight according $p(Y_{t+1} \mid X_{t+1})$.

Remark:

- The distribution $p(\boldsymbol{X}_t \mid \boldsymbol{Y}_t)$ is called the **filtering** distribution.
- The distribution $p(\boldsymbol{X}_t \mid \boldsymbol{Y})$ is called the **smoothing** distribution.

Multivariate Normal Distribution

- The vector $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if its density is

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

- The vector $\boldsymbol{X}$ is multivariate normal if and only if every linear combination of its components is normally distributed.
- If $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{A}\boldsymbol{X} + \boldsymbol{b} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)$.
- Marginally normal does not imply jointly normal:

$$X_1 \sim \mathcal{N}(0,1),\ X_2 = sX_1$$

  where s is a Rademacher random variable.

Multivariate Normal Distribution

Suppose

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

Then the conditional distribution of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is

$$\boldsymbol{X}_1 \mid \boldsymbol{X}_2 \sim \mathcal{N}\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \right)$$

Multivariate Normal Distribution

Proof 1:

The joint density of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is

$$
\begin{aligned}
& p(\boldsymbol{x}_1, \boldsymbol{x}_2) \\
& \propto \exp\left(-\frac{1}{2}\begin{pmatrix}\boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2\end{pmatrix}^T \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}^{-1} \begin{pmatrix}\boldsymbol{x}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{x}_2 - \boldsymbol{\mu}_2\end{pmatrix}\right) \\
& \propto_{\boldsymbol{x}_1} \exp\Bigg(-\frac{1}{2}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T \left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}\boldsymbol{\Sigma}_{21}\right)^{-1} (\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \\
& \qquad + (\boldsymbol{x}_1 - \boldsymbol{\mu}_1)^T \left(\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)^{-1} \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)\Bigg)
\end{aligned}
$$

The conditional distribution of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$ is

$$
\boldsymbol{X}_1 \mid \boldsymbol{X}_2 \sim \mathcal{N}\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)
$$

Multivariate Normal Distribution

Proof 2:
Construct

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{X}_2 \\ \boldsymbol{X}_2 \end{pmatrix}$$

Since Y_1 and Y_2 are linear combinations of X_1 and X_2, they are jointly normal:

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

Therefore, both $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are normal and they are independent. And

$$\boldsymbol{X}_1 \mid \boldsymbol{X}_2 = (\boldsymbol{Y}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{Y}_2) \mid \boldsymbol{Y}_2 \sim \mathcal{N}\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$$

Sequential Structure Under Linear Gaussian Models

Consider the following linear Gaussian state-space model:

$$X_t \mid X_{t-1} \sim \mathcal{N}(\boldsymbol{A}_t X_{t-1}, \boldsymbol{\Sigma}_t)$$
$$Y_t \mid X_t \sim \mathcal{N}(\boldsymbol{B}_t X_t, \boldsymbol{R}_t)$$

Or, in a constructive way,

$$X_t = \boldsymbol{A}_t X_{t-1} + \boldsymbol{\epsilon}_t$$
$$Y_t = \boldsymbol{B}_t X_t + \boldsymbol{\eta}_t.$$

with $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t)$ and $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}_t)$.

Sequential Structure Under Linear Gaussian Models

Notice that

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_t & \boldsymbol{I}_x & 0 \\ \boldsymbol{B}_t\boldsymbol{A}_t & \boldsymbol{B}_t & \boldsymbol{I}_y \end{pmatrix} \begin{pmatrix} X_{t-1} \\ \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix}$$

If $X_{t-1} \sim \mathcal{N}(\boldsymbol{\mu}_{t-1}, \boldsymbol{V}_{t-1})$, then

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{A}_t\boldsymbol{\mu}_{t-1} \\ \boldsymbol{B}_t\boldsymbol{A}_t\boldsymbol{\mu}_{t-1} \end{pmatrix}, \begin{pmatrix} \boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T + \boldsymbol{\Sigma}_t & \boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T\boldsymbol{B}_t^T + \boldsymbol{\Sigma}_t\boldsymbol{B}_t^T \\ \boldsymbol{B}_t\boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T + \boldsymbol{B}_t\boldsymbol{\Sigma}_t & \boldsymbol{B}_t\boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T\boldsymbol{B}_t^T + \boldsymbol{B}_t\boldsymbol{\Sigma}_t\boldsymbol{B}_t^T + \boldsymbol{R}_t \end{pmatrix}\right)$$

Using the conditional probability of multivariate normal distribution, we have

$$X_t \mid Y_t \sim \mathcal{N}\left(\boldsymbol{\mu}_t, \boldsymbol{V}_t\right)$$

with

$$\begin{aligned}
&\boldsymbol{\mu}_t = \\
&\boldsymbol{A}\boldsymbol{\mu}_{t-1} + (\boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T + \boldsymbol{\Sigma}_t)\boldsymbol{B}_t^T(\boldsymbol{B}_t\boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T\boldsymbol{B}_t^T + \boldsymbol{B}_t\boldsymbol{\Sigma}_t\boldsymbol{B}_t^T + \boldsymbol{R}_t)^{-1}(Y_t - \boldsymbol{B}_t\boldsymbol{A}_t\boldsymbol{\mu}_{t-1}) \\
&\boldsymbol{V}_t = \boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T + \boldsymbol{\Sigma}_t \\
&- (\boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T + \boldsymbol{\Sigma}_t)\boldsymbol{B}_t^T(\boldsymbol{B}_t\boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T\boldsymbol{B}_t^T + \boldsymbol{B}_t\boldsymbol{\Sigma}_t\boldsymbol{B}_t^T + \boldsymbol{R}_t)^{-1}\boldsymbol{B}_t(\boldsymbol{A}_t\boldsymbol{V}_{t-1}\boldsymbol{A}_t^T + \boldsymbol{\Sigma}_t)
\end{aligned}$$

Sequential Structure Under Linear Gaussian Models

A simplified version of the previous formula:

- If $X_{t-1} \sim \mathcal{N}(\boldsymbol{\mu}_{t-1}, \boldsymbol{V}_{t-1})$, then
- $X_t \mid Y_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{V}_t)$
- with

$$
\begin{aligned}
\boldsymbol{Q}_t &= \boldsymbol{A}_t \boldsymbol{V}_{t-1} \boldsymbol{A}_t^T + \boldsymbol{\Sigma}_t \\
\boldsymbol{K}_t &= \boldsymbol{B}_t \boldsymbol{Q}_t \boldsymbol{B}_t^T + \boldsymbol{R}_t \\
\boldsymbol{\mu}_t &= \boldsymbol{A}_t \boldsymbol{\mu}_{t-1} + \boldsymbol{Q}_t \boldsymbol{B}_t^T \boldsymbol{K}_t^{-1} (Y_t - \boldsymbol{B}_t \boldsymbol{A}_t \boldsymbol{\mu}_{t-1}) \\
\boldsymbol{V}_t &= \boldsymbol{Q}_t - \boldsymbol{Q}_t \boldsymbol{B}_t^T \boldsymbol{K}_t^{-1} \boldsymbol{B}_t \boldsymbol{Q}_t
\end{aligned}
$$

Kalman Filter

Consider the following linear Gaussian state-space model:

$$X_t \mid X_{t-1} \sim \mathcal{N}(\boldsymbol{A}_t X_{t-1}, \boldsymbol{\Sigma}_t)$$
$$Y_t \mid X_t \sim \mathcal{N}(\boldsymbol{B}_t X_t, \boldsymbol{R}_t)$$

with $X_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{V}_0)$.
The **Kalman filter** is a recursive algorithm to compute the filtering distribution $X_t \mid \boldsymbol{Y}_t \sim \mathcal{N}(\boldsymbol{\mu}_t, \boldsymbol{V}_t)$:

1. for $t = 1, 2, \ldots, T$:
2. Compute

$$\begin{aligned}
\boldsymbol{Q}_t &= \boldsymbol{A}_t \boldsymbol{V}_{t-1} \boldsymbol{A}_t^T + \boldsymbol{\Sigma}_t \\
\boldsymbol{K}_t &= \boldsymbol{B}_t \boldsymbol{Q}_t \boldsymbol{B}_t^T + \boldsymbol{R}_t \\
\boldsymbol{\mu}_t &= \boldsymbol{A}_t \boldsymbol{\mu}_{t-1} + \boldsymbol{Q}_t \boldsymbol{B}_t^T \boldsymbol{K}_t^{-1} (Y_t - \boldsymbol{B}_t \boldsymbol{A}_t \boldsymbol{\mu}_{t-1}) \\
\boldsymbol{V}_t &= \boldsymbol{Q}_t - \boldsymbol{Q}_t \boldsymbol{B}_t^T \boldsymbol{K}_t^{-1} \boldsymbol{B}_t \boldsymbol{Q}_t
\end{aligned}$$

The Smoothing Problem

Now we consider the smoothing problem, that is, to find the smoothing distribution $X_t \mid \boldsymbol{Y}_T$.
From the previous calculation, we have

$$\begin{pmatrix} X_t \\ X_{t+1} \end{pmatrix} \Bigg| \, \boldsymbol{Y}_t \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_t \\ \boldsymbol{A}_{t+1}\boldsymbol{\mu}_t \end{pmatrix}, \begin{pmatrix} \boldsymbol{V}_t & \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \\ \boldsymbol{A}_{t+1}\boldsymbol{V}_t & \boldsymbol{Q}_{t+1} \end{pmatrix} \right)$$

Using the conditional probability of multivariate normal distribution, we have

$$X_t \mid X_{t+1}, \boldsymbol{Y}_t \sim \mathcal{N}\left(\boldsymbol{\mu}_t + \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} (X_{t+1} - \boldsymbol{A}_{t+1}\boldsymbol{\mu}_t), \boldsymbol{V}_t - \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} \boldsymbol{A}_{t+1} \boldsymbol{V}_t \right)$$

In the smoothing case, we assume $X_t \mid \boldsymbol{Y}_T \sim \mathcal{N}(\boldsymbol{\nu}_t, \boldsymbol{U}_t)$.

The Smoothing Problem

Using the law of total expectation, we have

$$\begin{aligned}
\boldsymbol{\nu}_t &= \mathbb{E}[X_t \mid \boldsymbol{Y}_T] \\
&= \mathbb{E}[\mathbb{E}[X_t \mid X_{t+1}, \boldsymbol{Y}_T] \mid \boldsymbol{Y}_T] \\
&= \mathbb{E}[\mathbb{E}[X_t \mid X_{t+1}, \boldsymbol{Y}_t] \mid \boldsymbol{Y}_T] \\
&= \mathbb{E}[\boldsymbol{\mu}_t + \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} (X_{t+1} - \boldsymbol{A}_{t+1} \boldsymbol{\mu}_t) \mid \boldsymbol{Y}_T] \\
&= \boldsymbol{\mu}_t + \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} (\boldsymbol{\nu}_{t+1} - \boldsymbol{A}_{t+1} \boldsymbol{\mu}_t)
\end{aligned}$$

The Smoothing Problem

Using the law of total variance, we have

$$
\begin{aligned}
\boldsymbol{U}_t &= \mathrm{Var}[X_t \mid \boldsymbol{Y}_T] \\
&= \mathbb{E}[\mathrm{Var}[X_t \mid X_{t+1}, \boldsymbol{Y}_T] \mid \boldsymbol{Y}_T] + \mathrm{Var}[\mathbb{E}[X_t \mid X_{t+1}, \boldsymbol{Y}_T] \mid \boldsymbol{Y}_T] \\
&= \mathbb{E}[\mathrm{Var}[X_t \mid X_{t+1}, \boldsymbol{Y}_t] \mid \boldsymbol{Y}_T] + \mathrm{Var}[\mathbb{E}[X_t \mid X_{t+1}, \boldsymbol{Y}_t] \mid \boldsymbol{Y}_T] \\
&= \mathbb{E}[\boldsymbol{V}_t - \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} \boldsymbol{A}_{t+1} \boldsymbol{V}_t \mid \boldsymbol{Y}_T] + \mathrm{Var}[\boldsymbol{\mu}_t + \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} (X_{t+1} - \boldsymbol{A}_{t+1} \boldsymbol{\mu}_t) \mid \boldsymbol{Y}_T] \\
&= \boldsymbol{V}_t - \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} \boldsymbol{A}_{t+1} \boldsymbol{V}_t + \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} \boldsymbol{U}_{t+1} \boldsymbol{Q}_{t+1}^{-1} \boldsymbol{A}_{t+1} \boldsymbol{V}_t \\
&= \boldsymbol{V}_t + \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} (\boldsymbol{U}_{t+1} - \boldsymbol{Q}_{t+1}) \boldsymbol{Q}_{t+1}^{-1} \boldsymbol{A}_{t+1} \boldsymbol{V}_t
\end{aligned}
$$

The Smoothing Problem

In summary, if we know $X_{t+1} \mid \boldsymbol{Y}_T \sim \mathcal{N}(\boldsymbol{\nu}_{t+1}, \boldsymbol{U}_{t+1})$, then

$$X_t \mid \boldsymbol{Y}_T \sim \mathcal{N}(\boldsymbol{\nu}_t, \boldsymbol{U}_t)$$

with

$$\begin{aligned}
\boldsymbol{\nu}_t &= \boldsymbol{\mu}_t + \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} (\boldsymbol{\nu}_{t+1} - \boldsymbol{A}_{t+1} \boldsymbol{\mu}_t) \\
\boldsymbol{U}_t &= \boldsymbol{V}_t + \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} (\boldsymbol{U}_{t+1} - \boldsymbol{Q}_{t+1}) \boldsymbol{Q}_{t+1}^{-1} \boldsymbol{A}_{t+1} \boldsymbol{V}_t
\end{aligned}$$

Kalman Smoother

The **Kalman Smoother** is a recursive algorithm to compute the smoothing distribution $X_t \mid \boldsymbol{Y}_T \sim \mathcal{N}(\boldsymbol{\nu}_t, \boldsymbol{U}_t)$:

1. Run the Kalman filter.
2. Initialize $\boldsymbol{\nu}_T = \boldsymbol{\mu}_T$ and $\boldsymbol{U}_T = \boldsymbol{V}_T$.
3. for $t = T-1, T-2, \ldots, 1$:
4. Compute

$$\begin{aligned}
\boldsymbol{C}_t &= \boldsymbol{V}_t \boldsymbol{A}_{t+1}^T \boldsymbol{Q}_{t+1}^{-1} \\
\boldsymbol{\nu}_t &= \boldsymbol{\mu}_t + \boldsymbol{C}_t(\boldsymbol{\nu}_{t+1} - \boldsymbol{A}_{t+1}\boldsymbol{\mu}_t) \\
\boldsymbol{U}_t &= \boldsymbol{V}_t + \boldsymbol{C}_t(\boldsymbol{U}_{t+1} - \boldsymbol{Q}_{t+1})\boldsymbol{C}_t^T
\end{aligned}$$

Example: Object Tracking

$$
\begin{aligned}
\boldsymbol{X}_t &= \boldsymbol{A}\boldsymbol{X}_{t-1} + \boldsymbol{\eta}_t \\
\boldsymbol{Y}_t &= \boldsymbol{B}\boldsymbol{X}_t + \boldsymbol{\epsilon}_t,
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{A} &= \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
\boldsymbol{B} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \\
\boldsymbol{\epsilon}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{R}) \\
\boldsymbol{\eta}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).
\end{aligned}
$$

with $\boldsymbol{\Sigma} = \mathrm{diag}(0.3, 0.3, 0.5, 0.5)$ and $\boldsymbol{R} = \mathrm{diag}(10, 10)$.

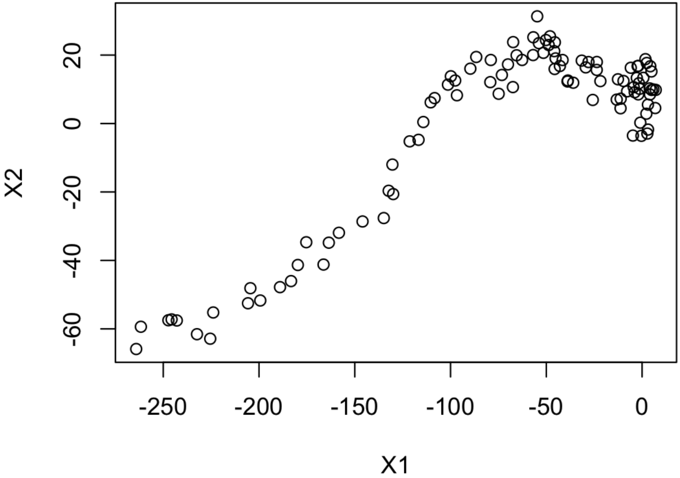
Example: Object Tracking

```
A = diag(4)
A[1, 3] = 1
A[2, 4] = 1
B = matrix(0, nrow=2, ncol=4)
B[1, 1] = 1
B[2, 2] = 1

Sigma = diag(c(0,3, 0.3, 0.5, 0.5))
R = diag(c(10, 10))

T = 100
Y = array(0, dim=c(2, T))
X = c(0, 0, 0, 0)
for(t in 1:T){
    X = A%*%X + sqrt(Sigma) %*% rnorm(4)
    Y[,t] = B%*%X + sqrt(R) %*% rnorm(2)
}
```

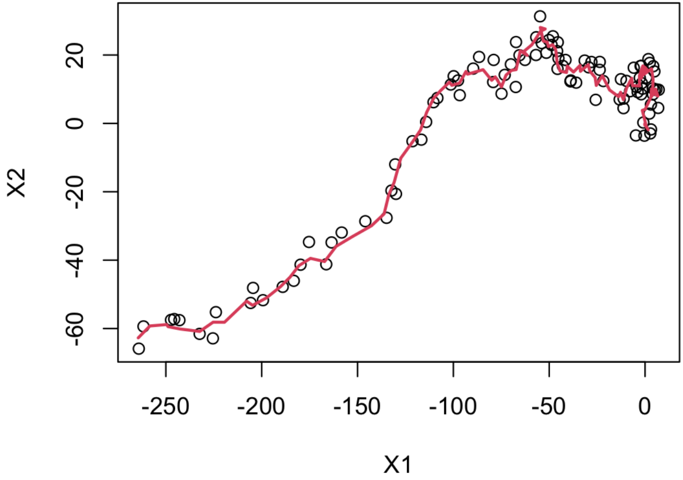
Example: Object Tracking



Example: Object Tracking

```
mu = array(0, dim=c(4, T+1))
V = array(0, dim=c(4, 4, T+1))
Q = array(0, dim=c(4, 4, T+1))
for(t in 1:T){
    Q[,,t+1] = A%*%V[,,t]%*%t(A) + Sigma
    K = B%*%Q[,,t+1]%*%t(B) + R
    mu[,t+1] = A%*%mu[,t] + Q[,,t+1]%*%t(B)%*%solve(K)%*%(Y[,t] - B%*%A%*
        %mu[,t])
    V[,,t+1] = Q[,,t+1] - Q[,,t+1]%*%t(B)%*%solve(K)%*%B%*%Q[,,t+1]
}
```

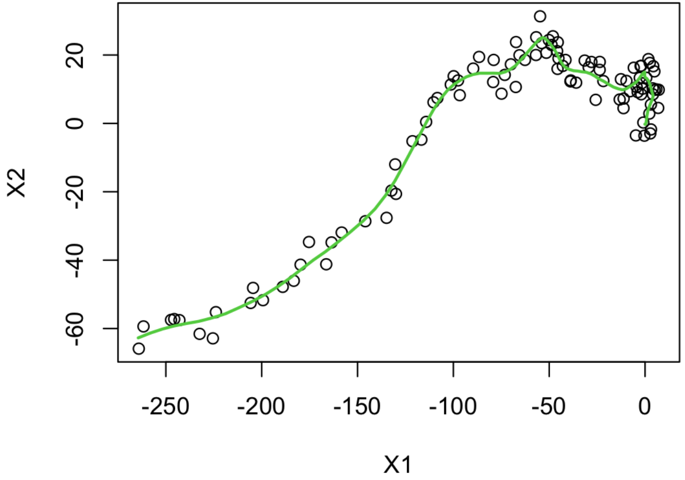
Example: Object Tracking



Example: Object Tracking

```
nu = array(0, dim=c(4, T))
U = array(0, dim=c(4, 4, T))
nu[,T] = mu[,T+1]
U[,,T] = V[,,T+1]
for(t in (T-1):1){
    C = V[,,t+1]%*%t(A)%*%solve(Q[,,t+2])
    nu[,t] = mu[, t+1] + C%*%(nu[,t+1] - A%*%mu[,t+1])
    U[,,t] = V[,,t+1] + C%*%(U[,,t+1] - Q[,,t+2])%*%t(C)
}
```

Example: Object Tracking



Example: Object Tracking