I225E Statistical Signal Processing

14. Signal Processing II

MAWALIM and UNOKI

candylim@jaist.ac.jp and unoki@jaist.ac.jp

School of Information Science

JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY
HOKURIKU 1990

Wiener filter and Kalman filter

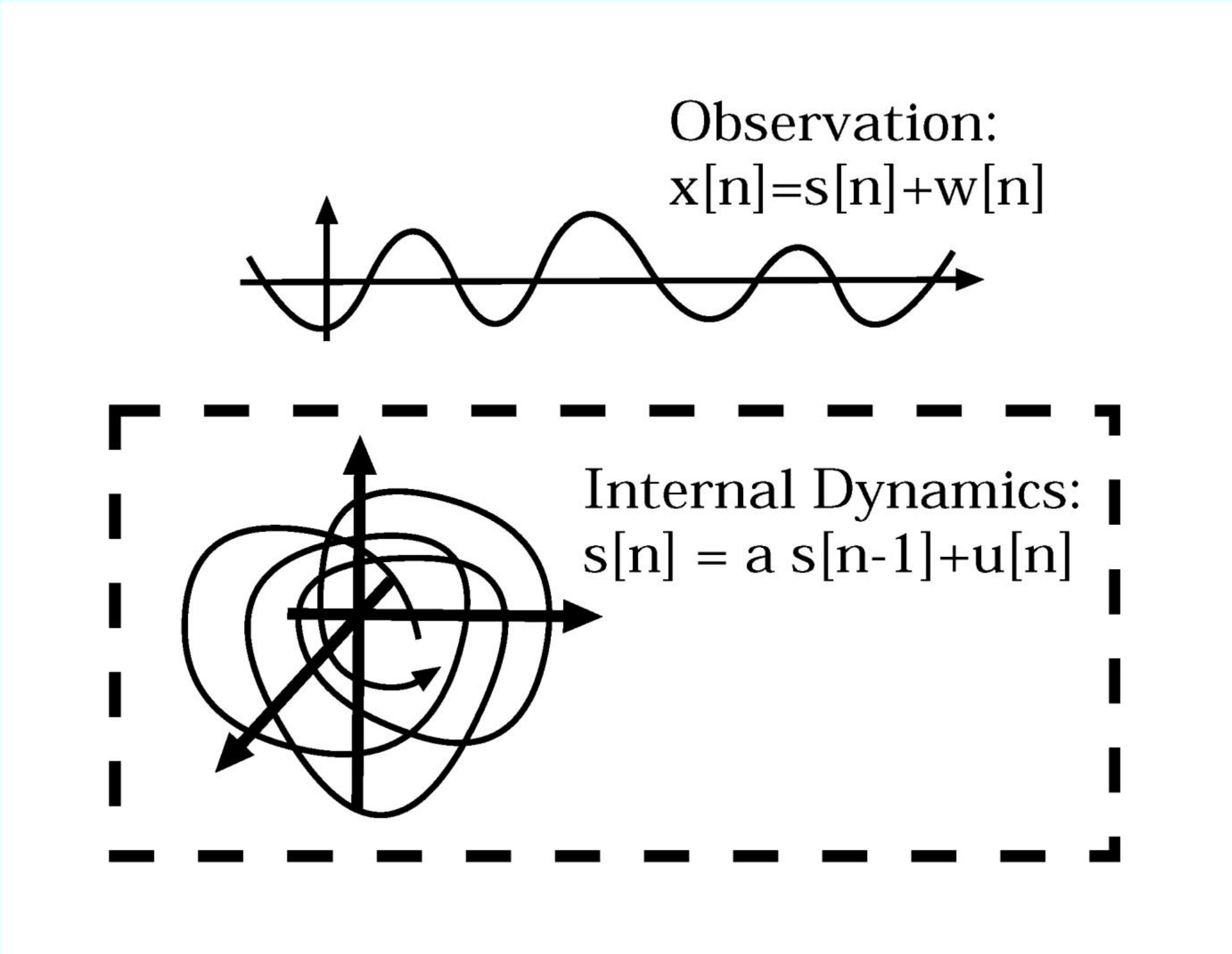
Kalman filter is an important generalization of Wiener filter.

- ***Wiener Filter***
  - WSS (Wide-sense-stationary) Process
  - Data from infinite past
  - Scalar signals
  - Non-adaptive
- ***Kalman Filter***
  - Gauss-Markov Process
  - Data from a specific point in time
  - Vector signals
  - Adaptive (model may evolve over time)

1. Introduction Kalman Filter

- The filter is named after **Rudolf E. Kálmán** (May 19, 1930 – July 2, 2016).
- In 1960, Kálmán published his famous paper on the Kalman Filter, an optimal estimation algorithm.
- Kalman Filter is a filter to find the best estimate from noisy input data by filtering out the noise.
- Kalman filter projects the data measurements onto the state estimate.

Estimate internal system state from observed data



Applications

- **Computer vision applications** for feature tracking or cluster tracking
- **Track objects** like missiles or people based on their current measured position to more accurately estimate their position and velocity in the future.
- **Navigation systems** utilize sensor output from an inertial measurement unit (IMU) and a global navigation satellite system (GNSS) receiver as input to estimate the vehicle state, position, and velocity.
- **Economics and Finance**
- **Robotics**

Key Concepts

1. **State-Space Model:**
   - **State Equation (System Model):** Describes how the system's state evolves over time.

     $$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \mathbf{w}_k$$

     - $\mathbf{x}_k$ is the state vector at time step k.
     - $\mathbf{F}_k$ is the state transition matrix, relating the previous state to the current state.
     - $\mathbf{B}_k$ is the input matrix, relating the control input $\mathbf{u}_k$ to the current state.
     - $\mathbf{u}_k$ is the control input vector at time step k.
     - $\mathbf{w}_k$ is the process noise, assumed to be zero-mean Gaussian with covariance $\mathbf{Q}_k$.
   - **Measurement Equation (Observation Model):** Describes how the measurements are related to the system's state.

     $$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$$

     - $\mathbf{z}_k$ is the measurement vector at time step k.
     - $\mathbf{H}_k$ is the observation matrix, relating the state to the measurement.
     - $\mathbf{v}_k$ is the measurement noise, assumed to be zero-mean Gaussian with covariance $\mathbf{R}_k$.

Key Concepts

2. **Recursive Nature:** The Kalman filter is a recursive algorithm, meaning it processes measurements sequentially. It doesn't need to store the entire history of measurements.
3. **Optimal Estimation:** The Kalman filter provides the optimal estimate of the system's state in the linear minimum mean squared error (LMMSE) sense, given the system model, the measurement model, and the statistics of the process and measurement noise.
4. **Two-Step Process:** The Kalman filter works in a two-step predict-update cycle:
   - Predict Step (Time update)
     - Predicts the current state based on the previous state estimate
     - Predicts the current error covariance
   - Update Step (Measurement update)
     - Computes the Kalman gain
     - Updates the state estimate using the measurement
     - Update the error covariance

2. Scalar Kalman Filter

First-order Gauss-Markov model

$$x[n] = s[n] + w[n],$$
$$s[n] = as[n-1] + u[n].$$

From observed data $\boldsymbol{X}[n] = [x[0], x[1], \cdots, x[n]]^T$, estimate $s[n]$ $(n \geq 0)$

- Constant a is known ($|a| < 1$).
- $u[n] \sim N(0, \sigma_u^2)$, $w[n] \sim N(0, \sigma_w^2)$, $s[-1] \sim N(0, \sigma_s^2)$.
- $s[-1]$, $u[n]$, $w[n]$ are all independent from each other.
- Denote estimate of $s[n]$ based on $\boldsymbol{X}[m] = [x[0], x[1], \cdots, x[m]]^T$ by $\hat{s}[n|m]$.

Find estimator $\hat{s}[n|n]$ that minimizes mean square error

$$E[(s[n] - \hat{s}[n|n])^2].$$

Computational Procedure of Kalman Filter

- **Prediction**

$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1]$$

- **Minimum prediction error**

$$M[n|n-1] = a^2 M[n-1|n-1] + \sigma_u^2$$

- **Kalman gain**

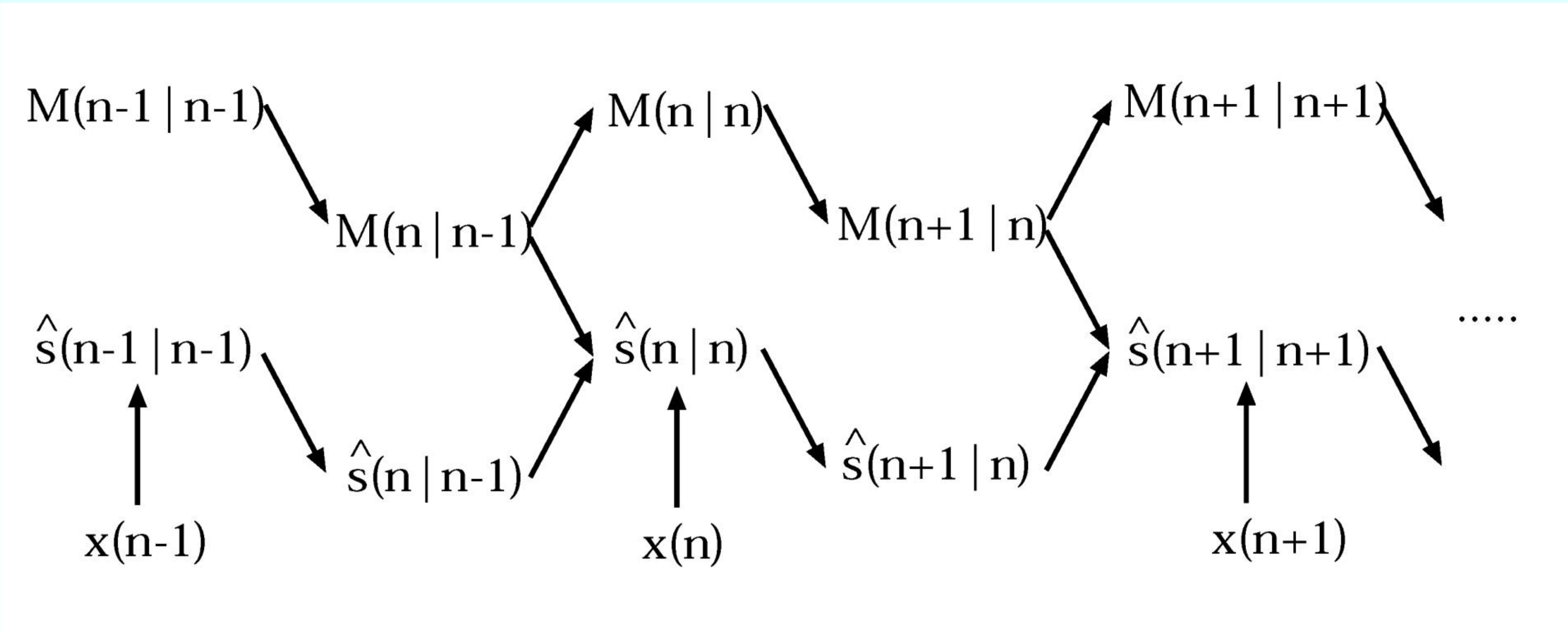
$$K[n] = \frac{M[n|n-1]}{\sigma_w^2 + M[n|n-1]}$$

- **Correction**

$$\hat{s}[n|n] = \hat{s}[n|n-1] + K[n](x[n] - \hat{s}[n|n-1])$$

- **Minimum mean square error**

$$M[n|n] = (1 - K[n])M[n|n-1]$$


M(n-1 | n-1)
M(n | n)
M(n+1 | n+1)
M(n | n-1)
M(n+1 | n)
ŝ(n-1 | n-1)
ŝ(n | n)
ŝ(n+1 | n+1)
.....
ŝ(n | n-1)
ŝ(n+1 | n)
x(n-1)
x(n)
x(n+1)

3. Dynamical Model

Theorem: Gauss-Markov model

$$\boldsymbol{s}[n] = \boldsymbol{A}\boldsymbol{s}[n-1] + \boldsymbol{B}\boldsymbol{u}[n] \qquad n \geq 0.$$

- $\boldsymbol{A} \in \Re^{p\times p}$, $\boldsymbol{B} \in \Re^{p\times r}$
  Eigenvalues of $\boldsymbol{A}$ have amplitude smaller than 1.
- $\boldsymbol{s}[n] \in \Re^{p\times 1}$, $\boldsymbol{s}[-1] \sim N(\mu_s, \boldsymbol{C}_s)$
- $\boldsymbol{u}[n] \in \Re^{r\times 1}$, $\boldsymbol{u}[n] \sim N(0, \boldsymbol{Q})$
- $\boldsymbol{s}[-1]$, $\boldsymbol{u}[-1]$ are all independent from each other.

Then, mean and covariance of signal $\boldsymbol{s}[n]$ are

- **Mean**: $E\{\boldsymbol{s}[n]\} = \boldsymbol{A}^{n+1}\mu_s$
- **Covariance**: $\boldsymbol{C}_s[m,n] =$

$$E\left\{(\boldsymbol{s}[m] - E\{\boldsymbol{s}[m]\})\boldsymbol{s}[n] - E\{\boldsymbol{s}[n]\}\right\}$$

$(m \geq n)$ $$= \boldsymbol{A}^{m+1}\boldsymbol{C}_s(\boldsymbol{A}^{n+1})^T + \sum_{k=m-n}^{m} \boldsymbol{A}^k\boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}^T\left(\boldsymbol{A}^{n-m+k}\right)^T$$

$(m < n)$ $$\boldsymbol{C}_s[m,n] = \boldsymbol{C}_s^T[n,m]$$

$(m = n)$ $$\boldsymbol{C}[n] = \boldsymbol{C}_s[n,n]$$

$$= \boldsymbol{A}^{n+1}\boldsymbol{C}_s(\boldsymbol{A}^{n+1})^T + \sum_{k=0n} \boldsymbol{A}^k\boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}^T\left(\boldsymbol{A}^k\right)^T$$

Time evolution of mean and covariance is

$$E\{\boldsymbol{s}[n]\} = \boldsymbol{A}E\{\boldsymbol{s}[n-1]\},$$
$$\boldsymbol{C}[n] = \boldsymbol{A}\boldsymbol{C}[n-1]\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}^T.$$

4. Derivation of scalar Kalman filter

First-order Gauss-Markov model

$$x[n] = s[n] + w[n]$$
$$s[n] = as[n-1] + u[n]$$

From observed data $\boldsymbol{X}[n] = [x[0], x[1], \cdots, x[n]]^T$, estimate $s[n] (n \geq 0)$.

- Constant a is known ($|a| < 1$).
- $u[n] \sim N(0, \sigma_u^2)$, $w[n] \sim N(0, \sigma_w^2)$, $s[-1] \sim N(0, \sigma_s^2)$.
- $s[-1], u[n], w[n]$ are all independent from each other.
- Denote estimate of $s[n]$ based on $\boldsymbol{X}[m] = [x[0], x[1], \cdots, x[m]]^T$ by $\hat{s}[n|m]$.
- Denote error by $\tilde{x}[n] = x[n] - \hat{x}[n|n-1]$.

With respect to minimum mean square error (MMSE)

$$E\{(s[n]-\hat{s}[n|n])^2\},$$

Corresponding minimum mean square estimation is

$$\hat{s}[n|n]=E\{s[n]|\boldsymbol{X}[n]\}.$$

Basic properties of minimum mean square error estimator used for derivation:

- With respect to uncorrelated data x_1, x_2, minimum mean square error estimator $\hat{\theta}$ is (in case of $E\{\theta\}=0$),

$$\hat{\theta}=E\{\theta|x_1,x_2\}=E\{\theta|x_1\}+E\{\theta|x_2\}.$$

- For $\theta = \theta_1 + \theta_2$, minimum mean square error estimator $\hat{\theta}$ is,

$$\hat{\theta} = E\{\theta | x\}$$
$$= E\{\theta_1 + \theta_2 | x\}$$
$$= E\{\theta_1 | x\} + E\{\theta_2 | x\}.$$

$$\tilde{s}[n|n] = E\{s[n] | \boldsymbol{X}[n]\} = E\{s[n] | \boldsymbol{X}[n-1], x[n]\}$$
$$= E\{s[n] | \boldsymbol{X}[n-1], \tilde{x}[n] + \hat{x}[n]\}$$

because $\hat{x}[n]$ is represented by linear summation of $\{x[0], x[1], \cdots, x[n-1]\}$

$$= E\{s[n] | \boldsymbol{X}[n-1], \tilde{x}[n]\}$$
$$= E\{s[n] | \boldsymbol{X}[n-1]\} + E\{s[n] | \tilde{x}[n]\}.$$

$$
\begin{aligned}
\tilde{s}[n|n-1] &= E\{s[n]|\boldsymbol{X}[n-1]\} \\
&= E\{as[n-1]+u[n]|\boldsymbol{X}[n-1]\} \\
&= aE\{s[n-1]|\boldsymbol{X}[n-1]\ \} \\
&\qquad \text{because } \ E\{u[n]|X[n-1]\} = E\{u[n]\} = 0 \\
&= a\hat{s}[n-1|n-1]
\end{aligned}
$$

$$
\hat{s}[n|n] = \hat{s}[n|n-1] + E\{s[n]|\tilde{x}[n]\},
$$

$$
\hat{s}[n|n-1] = a\hat{s}[n-1|n-1]
$$

Since $E\{s[n]|\tilde{x}[n]\}$ is MMSE estimator of $s[n]$ based on $\tilde{x}[n]$,

$$\begin{aligned} E\{s[n]|\tilde{x}[n]\} &= K[n]\tilde{x}[n] \\ &= K[n](x[n] - \hat{x}[n|n-1]) \end{aligned}$$

$$K[n] = \frac{E\{s[n]\tilde{x}[n]\}}{E\{\tilde{x}^2[n]\}}$$

$$\begin{aligned} \hat{x}[n|n-1] &= E\{x[n]|\boldsymbol{X}[n-1]\} \\ &= E\{s[n] + w[n]|\boldsymbol{X}[n-1]\} \\ &= E\{s[n]|\boldsymbol{X}[n-1]\} + E\{w[n]|\boldsymbol{X}[n-1]\} \\ &= \hat{s}[n|n-1] + \widehat{w}[n|n-1] \\ &= \hat{s}[n|n-1]. \ \text{(from } \widehat{w}[n|n-1] = 0) \end{aligned}$$

Summarizing above

$$\hat{s}[n|n] = \hat{s}[n|n-1] + K[n](x[n] - \hat{x}[n|n-1])$$
$$= \hat{s}[n|n-1] + K[n](x[n] - \hat{s}[n|n-1]),$$
$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1].$$

Denominator and numerator of $K[n] = \frac{E\{s[n]\tilde{x}[n]\}}{E\{\tilde{x}^2[n]\}}$ are

$$E\{s[n]\tilde{x}[n]\} = E\{(s[n] - \hat{s}[n|n-1])\tilde{x}[n]\}$$

from orthogonality principle, $E\{\hat{s}[n|n-1]\tilde{x}[n]\} = 0$

using $\tilde{x}[n] = x[n] - \hat{x}[n|n-1] = x[n] - \hat{s}[n|n-1]$

$$= E\{(s[n] - \hat{s}[n|n-1])(x[n] - \hat{s}[n|n-1])\}$$

$$= E\{(s[n] - \hat{s}[n|n-1])(s[n] + w[n] - \hat{x}[n|n-1])\}$$

using $E\{(s[n] - \hat{s}[n|n-1])w[n]\} = 0$

$$= E\{(s[n] - \hat{s}[n|n-1])(s[n] - \hat{s}[n|n-1])\}$$

$$= E\{(s[n] - \hat{s}[n|n-1])^2\}.$$

$$E\{\tilde{x}^2[n]\} = E\{(x[n] - \hat{x}[n|n-1])^2\}$$

$$= E\{(s[n] - \hat{s}[n|n-1] + w[n])^2\}$$

$$= \sigma_w^2 + E\{(s[n] - \hat{s}[n|n-1])^2\}$$

Hence $$K[n] = \frac{E\{(s[n]-\hat{s}[n|n-1])^2\}}{\sigma_w^2 + E\{(s[n]-\hat{s}[n|n-1])^2\}}$$

$$= \frac{M[n|n-1]}{\sigma_w^2 + M[n|n-1]},$$

where

$$M[n|n-1] = E\{(s[n] - \hat{s}[n|n-1])^2\}.$$

$$M[n|n-1] = E\{(s[n] - \hat{s}[n|n-1])^2\}$$

$$= E\{(as[n-1] + u[n] - \hat{s}[n|n-1])^2\}$$

$$= E\{(a(s[n-1] - \hat{s}[n|n-1]) + u[n])^2\}$$

because $E\{(s[n-1] - \hat{s}[n-1|n-1])u[n]\} = 0$

$$= a^2 M[n-1|n-1] + \sigma_u^2.$$

$$
\begin{aligned}
M[n|n] &= E\{(s[n]-\hat{s}[n|n])^2\} \\
&= E\{(s[n]-\hat{s}[n|n-1]-K[n](x[n]-\hat{s}[n|n-1]))^2\} \\
&= E\{(s[n]-\hat{s}[n|n-1])^2\} \\
&\qquad -2K[n]E\{(s[n]-\hat{s}[n|n-1])(x[n]-\hat{s}[n|n-1])\} \\
&\qquad\quad +K^2[n]E\{(x[n]-\hat{s}[n|n-1])^2\} \\
&= M[n|n-1]-2K[n]M[n|n-1] \\
&\qquad\qquad +K[n]M[n|n-1] \\
&= (1-K[n])M[n|n-1].
\end{aligned}
$$

In summary, scalar Kalman filter is obtained as:

$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1],$$

$$M[n|n-1] = a^2 M[n-1|n-1] + \sigma_u^2,$$

$$K[n] = \frac{M[n|n-1]}{\sigma_w^2 + M[n|n-1]},$$

$$\hat{s}[n|n] = \hat{s}[n|n-1] + K[n](x[n] - \hat{s}[n|n-1]),$$

$$M[n|n] = (1 - K[n])M[n|n-1].$$

4. Extension to vector form

Gauss-Markov model

$$x[n] = \boldsymbol{h}^T s[n-1] + w[n]$$
$$\boldsymbol{s}[n] = \boldsymbol{A}\boldsymbol{s}[n-1] + \boldsymbol{B}\boldsymbol{u}[n], \quad n \geq 0$$

Estimate $s[n]$ from observed data $[x[0], x[1], \cdots, x[n]]^T$

- $\boldsymbol{A} \in \Re^{p\times p}$, $\boldsymbol{B} \in \Re^{p\times r}$, $\boldsymbol{h}[n] \in \Re^{p\times 1}$
- $\boldsymbol{s}[n] \in \Re^{p\times 1}$, $\boldsymbol{s}[-1] \sim N(\mu_s, \boldsymbol{C}_s)$.
- $\boldsymbol{u}[n] \in \Re^{r\times 1}$, $\boldsymbol{u}[n] \sim N(0, \boldsymbol{Q})$
- $w[n] \in \Re^{1\times 1}$, $w[n] \sim N(0, \sigma_w^2)$

Vector Kalman filter:

- ***Prediction*** $\hat{\boldsymbol{s}}[n|n-1] = \boldsymbol{A}\hat{\boldsymbol{s}}[n-1|n-1]$,
- ***Minimum prediction error***

$$\boldsymbol{M}[n|n-1] = \boldsymbol{A}\boldsymbol{M}[n-1|n-1]\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{Q}\boldsymbol{B}^T,$$

- ***Kalman gain*** $\boldsymbol{K}[n] = \frac{\boldsymbol{M}[n|n-1]\boldsymbol{h}[n]}{\sigma_w^2+\boldsymbol{h}^T[n]\boldsymbol{M}[n|n-1]\boldsymbol{h}[n]}$,
- ***Correction***

$$\hat{\boldsymbol{s}}[n|n] = \hat{\boldsymbol{s}}[n|n-1] + \boldsymbol{K}[n](x[n] - \boldsymbol{h}^T[n]\hat{s}[n|n-1]),$$

- ***Minimum mean square error***

$$\boldsymbol{M}[n|n] = (\boldsymbol{I} - \boldsymbol{K}[n]\boldsymbol{h}^T[n])\boldsymbol{M}[n|n-1].$$

Kalman filter

- Process and measurement equations:

$$\begin{cases} \mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{w}_n \\ \mathbf{z}_n = \mathbf{C}\mathbf{x}_n + \mathbf{v}_n \end{cases}$$

where **w** and **v** are Gaussian noises with mean **0** and covariance **Q** and **R**.

$$\mathbf{w}_n \sim N(\mathbf{0}, \mathbf{Q}), \quad \mathbf{v}_n \sim N(\mathbf{0}, \mathbf{R})$$

Problem: Given the posterior probability at step *n*

$$p(\mathbf{x}_n | \mathbf{Z}_{1:n}) = p(\mathbf{x}_n | \mathbf{z}_1, \cdots, \mathbf{z}_n),$$

Compute the posterior probability at step *n*+1

$$p(\mathbf{x}_{n+1} | \mathbf{Z}_{1:n+1}) = p(\mathbf{x}_{n+1} | \mathbf{z}_1, \cdots, \mathbf{z}_n, \mathbf{z}_{n+1}).$$

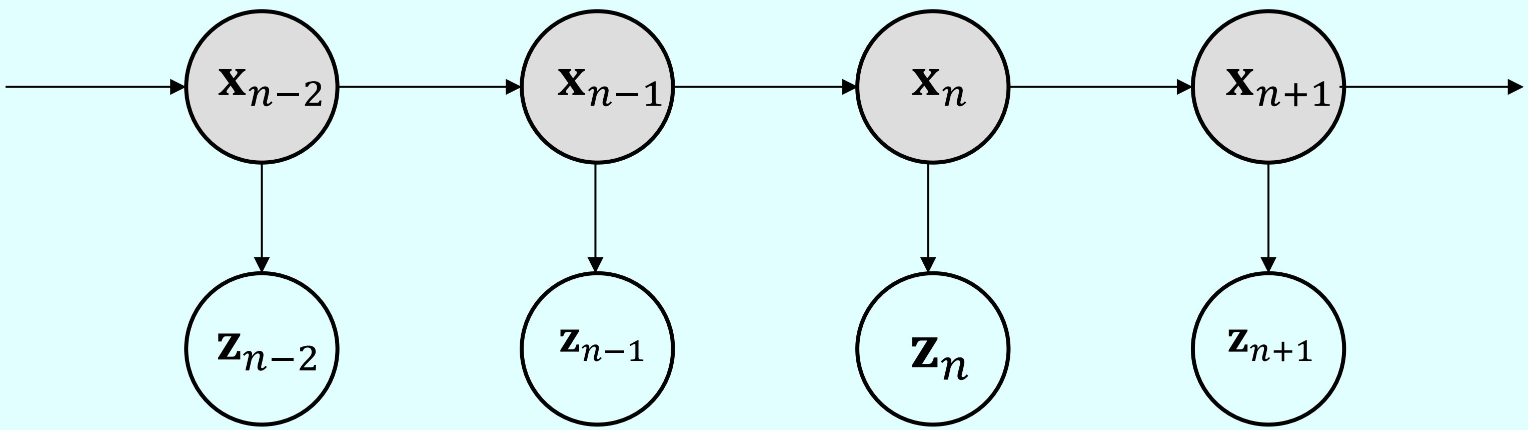
Kalman filter

- Process and measurement equations:

$$\begin{cases} \mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{w}_n \\ \mathbf{z}_n = \mathbf{C}\mathbf{x}_n + \mathbf{v}_n \end{cases}$$

where **w** and **v** are Gaussian noises with mean **0** and covariance **Q** and **R**.

$$\mathbf{w}_n \sim N(\mathbf{0}, \mathbf{Q}), \quad \mathbf{v}_n \sim N(\mathbf{0}, \mathbf{R})$$



Kalman filter

- Problem: Given the posterior probability at step *n*

$$p(\mathbf{x}_n|\mathbf{Z}_{1:n}) = p(\mathbf{x}_n|\mathbf{z}_1, \cdots, \mathbf{z}_n),$$

Compute the posterior probability at step *n*+1

$$p(\mathbf{x}_{n+1}|\mathbf{Z}_{1:n+1}) = p(\mathbf{x}_{n+1}|\mathbf{z}_1, \cdots, \mathbf{z}_n, \mathbf{z}_{n+1}).$$

$$p(\mathbf{x}_n|\mathbf{Z}_{1:n}) \xrightarrow[\text{prediction}]{} p(\mathbf{x}_{n+1}|\mathbf{Z}_{1:n}) \xrightarrow[\text{filtering}]{} p(\mathbf{x}_{n+1}|\mathbf{Z}_{1:n+1})$$

$$\begin{matrix} \hat{\mathbf{x}}_{n|n} \\ \Sigma_{n|n} \end{matrix} \xrightarrow[\text{prediction}]{} \begin{matrix} \hat{\mathbf{x}}_{n+1|n} \\ \Sigma_{n+1|n} \end{matrix} \xrightarrow[\text{filtering}]{} \begin{matrix} \hat{\mathbf{x}}_{n+1|n+1} \\ \Sigma_{n+1|n+1} \end{matrix}$$

Kalman filter

- Kalman filter: result

$$\begin{array}{ccccc} \hat{\mathbf{x}}_{n|n} & & \hat{\mathbf{x}}_{n+1|n} & & \hat{\mathbf{x}}_{n+1|n+1} \\ & \xrightarrow{\hspace{3em}} & & \xrightarrow{\hspace{3em}} & \\ \Sigma_{n|n} & \text{prediction} & \Sigma_{n+1|n} & \text{filtering} & \Sigma_{n+1|n+1} \end{array}$$

prediction step:

$$\boxed{\begin{aligned} \hat{\mathbf{x}}_{n+1|n} &= \mathbf{A}\hat{\mathbf{x}}_{n|n} \\ \Sigma_{n+1|n} &= \mathbf{A}\Sigma_{n|n}\mathbf{A}^{\mathrm{T}} + \mathbf{Q} \end{aligned}}$$

filtering step:

$$\boxed{\begin{aligned} &\hat{\mathbf{x}}_{n+1|n+1} \\ &= \hat{\mathbf{x}}_{n+1|n} + \mathbf{K}_{n+1}\left(\mathbf{z}_{n+1} - \mathbf{C}\hat{\mathbf{x}}_{n+1|n}\right) \\ &\Sigma_{n+1|n+1} = (\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})\Sigma_{n+1|n} \end{aligned}}$$

$$\mathbf{K}_{n+1} = \Sigma_{n+1|n}\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\Sigma_{n+1|n}\mathbf{C}^{\mathrm{T}} + \mathbf{R}\right)^{-1}$$

(Simplified) derivation based on mean and variance (1/3)

- Prediction step:

$$\hat{\mathbf{x}}_{n+1|n} = \mathrm{E}[\mathbf{x}_{n+1}|\mathbf{Z}_{1:n}] = \mathrm{E}[\mathbf{A}\mathbf{x}_n + \mathbf{w}_n|\mathbf{Z}_{1:n}] = \mathbf{A}\hat{\mathbf{x}}_{n|n}$$

$$\Sigma_{n+1|n} = \mathrm{Cov}[\mathbf{x}_{n+1}|\mathbf{Z}_{1:n}]$$
$$= \mathrm{E}\left[\left(\mathbf{x}_{n+1} - \hat{\mathbf{x}}_{n+1|n}\right)\left(\mathbf{x}_{n+1} - \hat{\mathbf{x}}_{n+1|n}\right)^{\mathrm{T}}|\mathbf{Z}_{1:n}\right]$$

$$= \mathrm{E}\left[\left\{\mathbf{A}\left(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n}\right) + \mathbf{w}_n\right\}\left\{\mathbf{A}\left(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n}\right) + \mathbf{w}_n\right\}^{\mathrm{T}}|\mathbf{Z}_{1:n}\right]$$
$$= \mathbf{A}\Sigma_{n|n}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}$$

$$\boxed{\begin{aligned}\hat{\mathbf{x}}_{n+1|n} &= \mathbf{A}\hat{\mathbf{x}}_{n|n}\\ \Sigma_{n+1|n} &= \mathbf{A}\Sigma_{n|n}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}\end{aligned}}$$

(Simplified) derivation based on mean and variance (2/3)

- Filtering step:
  Assume that the posterior mean has a linear form:

$$\hat{\mathbf{x}}_{n+1|n+1} = \mathrm{E}[\mathbf{x}_{n+1}|\mathbf{Z}_{1:n+1}] = \hat{\mathbf{x}}_{n+1|n} + \mathbf{K}_{n+1}\left(\mathbf{z}_{n+1} - \mathbf{C}\hat{\mathbf{x}}_{n+1|n}\right).$$

  Then, the Kalman gain K is determined so as to minimize the trace of covariance matrix:

$$\mathrm{Cov}[\mathbf{x}_{n+1}|\mathbf{Z}_{1:n+1}].$$

  By noting

$$\mathbf{x}_{n+1} - \hat{\mathbf{x}}_{n+1|n+1} = (\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})\left(\mathbf{x}_{n+1} - \hat{\mathbf{x}}_{n+1|n}\right) + \mathbf{K}_{n+1}\mathbf{v}_{n+1},$$

  we obtain

$$\mathrm{Cov}[\mathbf{x}_{n+1}|\mathbf{Z}_{1:n+1}] = (\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})\Sigma_{n+1|n}(\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})^{\mathrm{T}} + \mathbf{K}_{n+1}\mathbf{R}\mathbf{K}_{n+1}^{\mathrm{T}}$$

(Simplified) derivation based on mean and variance (3/3)

- Finding the Kalman gain:

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \mathbf{K}_{n+1}} \operatorname{tr} \operatorname{Cov}[\mathbf{x}_{n+1} | \mathbf{Z}_{1:n+1}] \\
&= \frac{\partial}{\partial \mathbf{K}_{n+1}} \operatorname{tr}\left\{(\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})\Sigma_{n+1|n}(\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})^{\mathrm{T}} + \mathbf{K}_{n+1}\mathbf{R}\mathbf{K}_{n+1}^{\mathrm{T}}\right\} \\
&= -2(\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})\Sigma_{n+1|n}\mathbf{C}^{\mathrm{T}} + 2\mathbf{K}_{n+1}\mathbf{R}
\end{aligned}
$$

$$
\therefore\ \mathbf{K}_{n+1} = \Sigma_{n+1|n}\mathbf{C}^{\mathrm{T}}\left(\mathbf{C}\Sigma_{n+1|n}\mathbf{C}^{\mathrm{T}} + \mathbf{R}\right)^{-1}
$$

$$
\boxed{
\begin{aligned}
&\hat{\mathbf{x}}_{n+1|n+1} \\
&= \hat{\mathbf{x}}_{n+1|n} + \mathbf{K}_{n+1}\left(\mathbf{z}_{n+1} - \mathbf{C}\hat{\mathbf{x}}_{n+1|n}\right) \\
&\Sigma_{n+1|n+1} = (\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})\Sigma_{n+1|n}
\end{aligned}
}
$$

(More rigorous) derivation based on probability densities (1/3)

- Prediction step:

$$
\begin{aligned}
p(\mathbf{x}_{n+1}|\mathbf{Z}_{1:n}) &= \int d\mathbf{x}_n p(\mathbf{x}_{n+1}, \mathbf{x}_n|\mathbf{Z}_{1:n+1}) \\
&= \int d\mathbf{x}_n p(\mathbf{x}_{n+1}|\mathbf{x}_n) p(\mathbf{x}_n|\mathbf{Z}_{1:n+1}) \\
&= \int d\mathbf{x}_n \frac{1}{(2\pi)^{n/2}|\mathbf{Q}|} \exp\left(-\frac{1}{2}(\mathbf{x}_{n+1} - \mathbf{A}\mathbf{x}_n)^{\mathrm{T}} \mathbf{Q}^{-1} (\mathbf{x}_{n+1} - \mathbf{A}\mathbf{x}_n)\right) \\
&\qquad \times \frac{1}{(2\pi)^{n/2}\left|\Sigma_{n|n}\right|} \exp\left(-\frac{1}{2}\left(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n}\right)^{\mathrm{T}} \Sigma_{n|n}^{-1} \left(\mathbf{x}_n - \hat{\mathbf{x}}_{n|n}\right)\right) \\
&= N\left(\mathbf{x}_{n+1}; \hat{\mathbf{x}}_{n+1|n}, \Sigma_{n+1|n}\right)
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
\hat{\mathbf{x}}_{n+1|n} &= \mathbf{A}\hat{\mathbf{x}}_{n|n} \\
\Sigma_{n+1|n} &= \mathbf{A}\Sigma_{n|n}\mathbf{A}^{\mathrm{T}} + \mathbf{Q}
\end{aligned}
}
$$

(More rigorous) derivation based on probability densities (2/3)

- Filtering step:

$$
\begin{aligned}
p(\mathbf{x}_{n+1}|\mathbf{Z}_{1:n+1}) &= p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1}, \mathbf{Z}_{1:n}) \\
&= \frac{p(\mathbf{z}_{n+1}|\mathbf{x}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{Z}_{1:n})}{p(\mathbf{z}_{n+1}|\mathbf{Z}_{1:n})} \\
&\propto p(\mathbf{z}_{n+1}|\mathbf{x}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{Z}_{1:n}) \\
&\propto \exp\left(-\frac{1}{2}\left(\mathbf{z}_{n+1} - \mathbf{C}\mathbf{x}_{n+1|n}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{z}_{n+1} - \mathbf{C}\mathbf{x}_{n+1|n}\right)\right)\exp\left(-\frac{1}{2}\left(\mathbf{x}_{n+1} - \hat{\mathbf{x}}_{n+1|n}\right)^{\mathrm{T}}\Sigma_{n+1|n}^{-1}\left(\mathbf{x}_{n+1} - \hat{\mathbf{x}}_{n+1|n}\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\mathbf{x}_{n+1} - \hat{\mathbf{x}}_{n+1|n+1}\right)^{\mathrm{T}}\Sigma_{n+1|n+1}^{-1}\left(\mathbf{x}_{n+1} - \hat{\mathbf{x}}_{n+1|n+1}\right)\right)
\end{aligned}
$$

$$
\boxed{
\begin{aligned}
&\hat{\mathbf{x}}_{n+1|n+1} \\
&= \left(\Sigma_{n+1|n}^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C}\right)^{-1}\left(\Sigma_{n+1|n}^{-1}\hat{\mathbf{x}}_{n+1|n} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{z}_{n+1}\right) \\
&\Sigma_{n+1|n+1} = \left(\Sigma_{n+1|n}^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C}\right)^{-1}
\end{aligned}
}
$$

(More rigorous) derivation based on probability densities (3/3)

- Filtering step:

$$\Sigma_{n+1|n+1} = \left(\Sigma_{n+1|n}^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C}\right)^{-1}$$

$$= \Sigma_{n+1|n} - \Sigma_{n+1|n}\mathbf{C}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{C}\Sigma_{n+1|n}\mathbf{C}^{\mathrm{T}}\right)^{-1}\mathbf{C}\Sigma_{n+1|n}$$
$$= (\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})\Sigma_{n+1|n}$$

$$\hat{\mathbf{x}}_{n+1|n+1} = \left(\Sigma_{n+1|n}^{-1} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{C}\right)^{-1}\left(\Sigma_{n+1|n}^{-1}\hat{\mathbf{x}}_{n+1|n} + \mathbf{C}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{z}_{n+1}\right)$$
$$= \hat{\mathbf{x}}_{n+1|n} + \mathbf{K}_{n+1}\left(\mathbf{z}_n - \mathbf{C}\hat{\mathbf{x}}_{n+1|n}\right)$$

$$\boxed{\begin{aligned}&\hat{\mathbf{x}}_{n+1|n+1}\\&= \hat{\mathbf{x}}_{n+1|n} + \mathbf{K}_{n+1}\left(\mathbf{z}_{n+1} - \mathbf{C}\hat{\mathbf{x}}_{n+1|n}\right)\\&\Sigma_{n+1|n+1} = (\mathbf{I} - \mathbf{K}_{n+1}\mathbf{C})\Sigma_{n+1|n}\end{aligned}}$$