Reinforcement and Online Learning

Mahesan Niranjan

Week Two and Three: State-Space Models and Kalman Filter

9, 15, 16 February 2021

Plan: Tuesday 9 Feb 2021

Approximate

- Quick review RLS, Coursework (15 Minutes)
- Discussion with UG cohort [Cameras ON, recording OFF] (15 minutes)
- Work on coursework on your own (45 minutes)
- Any clarifications on coursework (15 minutes)
- Start on State-Space models and what is coming next (30 minutes)

State Space Models

- We will look at systmes that are dynamic in time
  - Flight of an aircraft governed by laws of motion
  - Chemical plant governed by laws of reactions
  - Parameters of a model being estimated by stochastic gradient descent
  - A robot following a path
- In these we can usually identify
  - A dynamics determined by knowledge of the system
  - Noise in the dynamics that
  - A control input we exert
- The above are defined on the *state* of a systm
- We can make obervations from the system
  - Observations are functions of the state
  - There will be observation noise / instrument noise

Some Preliminaries

Linear Algebra and Multi-variate Gaussian Densities

- $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x} \,|\, \boldsymbol{m},\ C)$
- $\boldsymbol{x}$ is vector, dimension p
- $\boldsymbol{m}$ is mean, vector $\boldsymbol{m} \in \mathcal{R}^p$, estimated as:

$$\widehat{\boldsymbol{m}} = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}_n$$

- C is covariance matrix $C \in \mathcal{R}^{p \times p}$, estimated as:

$$\widehat{C} = \frac{1}{N} \sum_{n=1}^{N} (\boldsymbol{x}_n - \boldsymbol{m})\,(\boldsymbol{x}_n - \boldsymbol{m})^t$$

- $\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b}$ is a linear transform of $\boldsymbol{x}$
- $\boldsymbol{y}$ has probability distribution $\mathcal{N}(\boldsymbol{y} \,|\, A\boldsymbol{m} + \boldsymbol{b},\ ACA^t)$
- Suppose we add Gaussian noise: $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0},\ Q)$.... $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{w}$ has distribution $\mathcal{N}(\boldsymbol{z} \,|\, \boldsymbol{m},\ C + Q)$

Preliminaries (cont'd)

- Partitioned matrices – just like numbers, we can have matrices, but the dimensions must agree. Example

$$X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

will work if... if A is $m \times n$, B is $m \times l$, C is $p \times n$ and D is $p \times l$

- We can do matrix calculations with such partitioned matrices (as long as the dimensions match up!)

$$X = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \text{ and } Y = \begin{bmatrix} P & Q \\ R & S \end{bmatrix} \quad XY = \begin{bmatrix} AP + BR & AQ + BS \\ CP + DR & CQ + DS \end{bmatrix}$$

- For a covariance matrix:

$$C = \begin{bmatrix} P & Q \\ Q^t & R \end{bmatrix}$$

- We will see this in joint Gaussian distributions:

$$p\left(\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}, \begin{bmatrix} A & C \\ C^t & B \end{bmatrix}\right)$$

- Example: five variables $\{x_1, x_2, y_1, y_2, y_3\}$, jointly Gaussian in five dimensions. There may be relationships between them, inducing conditional probabilities.
- Conditionals:

$$\begin{aligned}
\boldsymbol{x} &\sim \mathcal{N}(\boldsymbol{a}, A) \\
\boldsymbol{y} &\sim \mathcal{N}(\boldsymbol{b}, B) \\
\boldsymbol{x}|\boldsymbol{y} &\sim \mathcal{N}\left(\boldsymbol{a} + CB^{-1}(\boldsymbol{y} - \boldsymbol{b}),\ A - CB^{-1}C^t\right) \\
\boldsymbol{y}|\boldsymbol{x} &\sim \mathcal{N}\left(\boldsymbol{b} + C^tA^{-1}(\boldsymbol{x} - \boldsymbol{a}),\ B - C^tA^{-1}C\right)
\end{aligned}$$

- Compare and check if the above is same as in Bishop: PRML.

Plan: Monday 15 Feb 2021

- Quick review of State Space Models
- Walk through Derivation of Kalman Filter

State Space Models: State Dynamics

- State dynamics (Control, System Identification literature):

$$\boldsymbol{x}_n = F_n \, \boldsymbol{x}_{n-1} + B_n \, \boldsymbol{u}_n + \boldsymbol{w}_n$$

- The state $\boldsymbol{x}$ makes a transition from $\boldsymbol{x}_{n-1}$ to $\boldsymbol{x}_n$, under control of command $\boldsymbol{u}_n$ and disturbed by noise $\boldsymbol{w}_n$
- This is a linear system because the transition happens by a matrix multiplication and additive control.
- The transitions (F_n) and gain on the control B_n can be time-varying or constant: F and B
- The noise is zero mean Gaussian with covariance matrix Q: $\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, Q)$
- Sometimes it is convenient to write (statistics literature):

$$p\left(\boldsymbol{x}_n \,|\, \boldsymbol{x}_{n-1}\right) = \mathcal{N}\left(F_n \boldsymbol{x}_{n-1} + B_n \boldsymbol{u}_n, \, Q\right)$$

State Space Models: Observations

- System output (seen by an observer) is a function of state

$$\boldsymbol{z}_n = H_n \boldsymbol{x}_n + \boldsymbol{v}_n$$

- Again, H_n could be constant H
- $\boldsymbol{v}_n$ is measurement noise / observation noise
- $\boldsymbol{v}_n \sim \mathcal{N}(\mathbf{0}, R)$
- One could also write (statistics literature):

$$p(\boldsymbol{z}_n \mid \boldsymbol{x}_n) = \mathcal{N}(H_n \boldsymbol{x}_n, R)$$

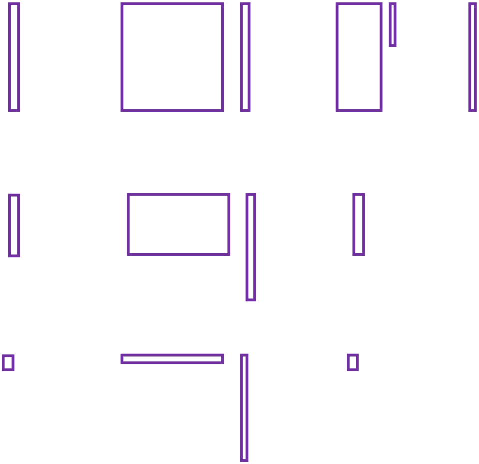
State Space Models

Note the matrix dimensions

Dynamical System

Observations

One dimensional observations



State Space Models

- We have

$$\begin{aligned} \boldsymbol{x}_n &= F\,\boldsymbol{x}_{n-1} + B\,\boldsymbol{u}_n + \boldsymbol{w}_n \\ \boldsymbol{z}_n &= H\,\boldsymbol{x}_n + \boldsymbol{v}_n \end{aligned}$$

  With the two noise processes, $\boldsymbol{w}_n \sim \mathcal{N}(\mathbf{0}, Q)$ and $\boldsymbol{v}_n \sim \mathcal{N}(\mathbf{0}, R)$, zero mean Gaussian.
- Here $\boldsymbol{z}_n$ are observations.
- Our job is to estimate the state $\boldsymbol{x}$, and an uncertainty P, on it based on the observations and knowledge of the dynamics
- We do linear transform of state and the noise is all Gaussian $\longrightarrow$ the distribution over state$\boldsymbol{x}$ is going to be Gaussian.
- We will propagate

$$\left\{ \begin{array}{c} \widehat{\boldsymbol{x}}_{n-1|n-1} \\ P_{n-1|n-1} \end{array} \right\} \longrightarrow \left\{ \begin{array}{c} \widehat{\boldsymbol{x}}_{n|n-1} \\ P_{n|n-1} \end{array} \right\} \longrightarrow \left\{ \begin{array}{c} \widehat{\boldsymbol{x}}_{n|n} \\ P_{n|n} \end{array} \right\}$$

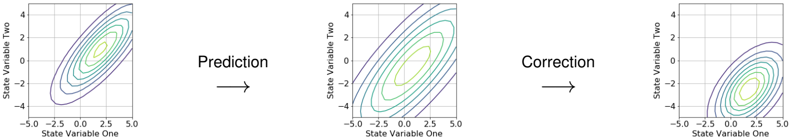
State Space Models (cont'd)

- We had:

$$\left\{ \begin{array}{c} \widehat{\boldsymbol{x}}_{n-1|n-1} \\ P_{n-1|n-1} \end{array} \right\} \longrightarrow \left\{ \begin{array}{c} \widehat{\boldsymbol{x}}_{n|n-1} \\ P_{n|n-1} \end{array} \right\} \longrightarrow \left\{ \begin{array}{c} \widehat{\boldsymbol{x}}_{n|n} \\ P_{n|n} \end{array} \right\}$$

- It is important to understand this as updating Gaussian density in two stages between time $n-1$ and n:
  - A prediction step from knowledge of dynamics
  - A correction step from the new data
- It is also important to understand that with linear models and Gaussian noise, the uncertainty over the estimated state is Gaussian:
  - Recall Lab One in COMP6245

$$\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{m}, \boldsymbol{\Sigma}\right), \quad \boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{b} \sim \mathcal{N}\left(A\,\boldsymbol{m} + \boldsymbol{b},\; A\Sigma A^t\right)$$



- When we do not have linear functions and Gaussian noise:
  - Approximation: Extended Kalman Filter (EKF)
  - Non-Parametric: Sequential Monte Carlo aka Particle Filter

Kalman Filter: Prediction Step

$$\left\{ \begin{array}{l} \widehat{\boldsymbol{x}}_{n-1|n-1} \\ P_{n-1|n-1} \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \widehat{\boldsymbol{x}}_{n|n-1} \\ P_{n|n-1} \end{array} \right\}$$

- Knowledge of dynamics helps us predict state and uncertainty on state:

$$\boldsymbol{x}_n = F\,\boldsymbol{x}_{n-1} + \boldsymbol{w}_n$$

- We are multiplying the state by a matrix (linear) and adding Gausian noise:

$$\mathcal{N}\left(\widehat{\boldsymbol{x}}_{n-1|n-1},\, P_{n-1|n-1}\right) \longrightarrow \text{becomes} \longrightarrow \mathcal{N}\left(F\,\widehat{\boldsymbol{x}}_{n-1|n-1},\quad F\,P_{n-1|n-1}\,F^t + Q\right)$$

- Alternate way of writing the prediction step:

$$\begin{aligned} \widehat{\boldsymbol{x}}_{n|n-1} &= F\,\widehat{\boldsymbol{x}}_{n-1|n-1} \\ P_{n|n-1} &= F\,P_{n-1|n-1}\,F^t + Q \end{aligned}$$

- We can predict the observation at next step:

$$\widehat{\boldsymbol{z}}_n = H\,\widehat{\boldsymbol{x}}_{n|n-1}$$

- We make a prediction error, aka *innovation*:

$$\widetilde{\boldsymbol{y}}_n = \boldsymbol{z}_n - H\,\widehat{\boldsymbol{x}}_{n|n-1}$$

Kalman Filter: Posterior Update

$$\left\{ \begin{array}{c} \widehat{\boldsymbol{x}}_{n|n-1} \\ P_{n|n-1} \end{array} \right\} \longrightarrow \left\{ \begin{array}{c} \widehat{\boldsymbol{x}}_{n|n} \\ P_{n|n} \end{array} \right\}$$

- First the results (important), then the formal derivation (for completeness)
- Innovation/ error $\widetilde{\boldsymbol{y}}_n$ is a random vector – same dimension as observation $\boldsymbol{z}_n$.
- Covariance of innovation:

$$S_n = H \, P_{n|n-1} \, H^t + R$$

- We have a term *Kalman gain*

$$K_n = P_{n|n-1} \, H \, S^{-1}$$

- Updates for posterior state:

$$\widehat{\boldsymbol{x}}_{n|n} = \widehat{\boldsymbol{x}}_{n|n-1} + K_n \, \widetilde{\boldsymbol{y}}_n$$

- Note the structure: New = Old $+$ Gain $\times$ Error
- Updates of uncertainty:

$$P_{n|n} = (I - K_n \, H) \; P_{n|n-1}$$

Derivation: $P_{n|n}$

Minimising the Mean Squared Error

- $P_{n|n}$ is uncertainty on the state estimate
- That is: $\texttt{Cov}(\boldsymbol{x}_n - \widehat{\boldsymbol{x}}_{n|n})$
- In the space of $\boldsymbol{x}$, we want to have a Gaussian distribution with mean at $\widehat{\boldsymbol{x}}_{n|n}$ and covariance matrix $P_{n|n}$.
- $P_{n|n-1} = \texttt{Cov}(\boldsymbol{x}_n - \widehat{\boldsymbol{x}}_{n|n-1})$
- Similarly, Innovation covariance: $S_n = \texttt{Cov}(\widetilde{\boldsymbol{y}}_n)$
- A few steps of substitution:

$$
\begin{aligned}
P_{n|n} &= \texttt{Cov}[\boldsymbol{x}_n - \widehat{\boldsymbol{x}}_{n|n}] \\
&= \texttt{Cov}[\boldsymbol{x}_n - (\widehat{\boldsymbol{x}}_{n|n-1} + K_n\,\widetilde{\boldsymbol{y}}_n)] \\
&= \texttt{Cov}[\boldsymbol{x}_n - (\widehat{\boldsymbol{x}}_{n|n-1} + K_n\,(\boldsymbol{z}_n - H\,\widehat{\boldsymbol{x}}_{n|n-1}))] \\
&= \texttt{Cov}[\boldsymbol{x}_n - (\widehat{\boldsymbol{x}}_{n|n-1} + K_n\,(H\,\boldsymbol{x}_n + \boldsymbol{v}_n - H\,\widehat{\boldsymbol{x}}_{n|n-1}))] \\
&= \texttt{Cov}[(I - K_n\,H)(\boldsymbol{x}_n - \widehat{\boldsymbol{x}}_{n|n-1}) - K_n\,\boldsymbol{v}_n] \\
&= \texttt{Cov}[(I - K_n\,H)(\boldsymbol{x}_n - \widehat{\boldsymbol{x}}_{n|n-1})] + \texttt{Cov}[K_n\,\boldsymbol{v}_n] \\
&= (I - K_n\,H)\,\texttt{Cov}[(\boldsymbol{x}_n - \widehat{\boldsymbol{x}}_{n|n-1})]\,(I - K_n\,H)^t + K_n\,\texttt{Cov}[\boldsymbol{v}_n]\,K_n^t \\
&= (I - K_n\,H)\,P_{n|n-1}\,(I - K_n\,H)^t + K_n\,R\,K_n^t
\end{aligned}
$$

Derivation: K_n

- Expected error in the posterior state estimate:

$$E\left[||\boldsymbol{x} - \boldsymbol{x}_{n|n}||^2\right]$$

- To minimize this, we minimize the trace of the error covariance matrix: $P_{n|n}$
- We already have an expresion for $P_{n|n}$:

$$\begin{aligned}
P_{n|n} &= (I - K_n H)\, P_{n|n-1}\, (I - K_n H)^t + K_n\, R\, K_n^t \\
&= P_{n|n-1} - K_n\, H\, P_{n|n-1} - P_{n|n-1}\, H^t\, K_n^t + K_n\, (H\, P_{n|n-1}\, H^t + R)\, K_n^t \\
&= P_{n|n-1} - K_n\, H\, P_{n|n-1} - P_{n|n-1}\, H^t\, K_n^t + K_n\, S_n\, K_n^t
\end{aligned}$$

- We have an expresion for $P_{n|n}$ with the Kalman gain K_n as unknown.
- Our task is to maximize the trace of $P_{n|n}$ with respect to K_n.

Derivation: K_n (cont'd)

- $P_{n|n} \; = \; P_{n|n-1} \; - \; K_n \, H \, P_{n|n-1} \; - \; P_{n|n-1} \, H^t \, K_n^t \; + \; K_n \, S_n \, K_n^t$
- Differentiate trace of $P_{n|n}$ with respect to K_n, set to zero and solve:

$$\begin{aligned} \frac{\partial \, \mathrm{tr}(P_{n|n-1})}{\partial \, K_n} \quad &= \quad -2 \, (H \, P_{n|n-1})^t \; + \; 2 \, K_n \, S_n \\ &= \quad \mathbf{0} \end{aligned}$$

- This gives

$$K_n \; = \; P_{n|n-1} \, H^t \, S_n^{-1}$$

- We can simplify the expression for $P_{n|n}$.
- Substituting for K_n cancels the last two terms:

$$\begin{aligned} P_{n|n} \quad &= \quad P_{n|n-1} \; - \; K_n \, H \, P_{n|n-1} \; - \; P_{n|n-1} \, H^t \, K_n^t \; + \; K_n \, S_n \, K_n^t \\ &= \quad P_{n|n-1} \; - \; K_n \, H \, P_{n|n-1} \\ &= \quad (I \; - \; K_n \, H) \; P_{n|n-1} \end{aligned}$$

Plan: Tuesday 16 Feb 2021

- Bayesian Derivation of Kalman Filter
- Estimating parameters of an Autoregressive Time Series
- Work on Assignment Two: Kalman Filtering

Derivation: Bayesian

Important: Sequential Monte Carlo builds on this!

- We start with a Markov Process: $p\left(\boldsymbol{x}_n \,|\, \boldsymbol{x}_0,\, \boldsymbol{x}_1 \dots \boldsymbol{x}_{n-1}\right) \;= p\left(\boldsymbol{x}_n \,|\, \boldsymbol{x}_{n-1}\right)$
- Observation depends on current state only (not on history):

$$p\left(\boldsymbol{z}_n \,|\, \boldsymbol{x}_0,\, \boldsymbol{x}_1 \dots \boldsymbol{x}_{n-1}\right) \;= p\left(\boldsymbol{z}_n \,|\, \boldsymbol{x}_{n-1}\right)$$

- Joint distribution (over states and observations)

$$p\left(\boldsymbol{x}_0,\, \boldsymbol{x}_1 \dots \boldsymbol{x}_n, \boldsymbol{z}_0,\, \boldsymbol{z}_1 \dots \boldsymbol{z}_n\right) \;=\; p(\boldsymbol{x}_0) \prod_{i=1}^{n} p(\boldsymbol{z}_i|\boldsymbol{x}_i)\, p(\boldsymbol{x}_i|\boldsymbol{x}_{i-1})$$

- Define a symbol for past observations $\boldsymbol{Z}_t \;=, \{\boldsymbol{z}_1,\, \boldsymbol{z}_1, \dots \boldsymbol{z}_t\}$
- Distribution over state at n is an integral over the distribution at $n-1$, multiplied by the transition probability:

$$p(\boldsymbol{x}_n|\boldsymbol{Z}_n) \;=\; \int p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})\, p(\boldsymbol{x}_{n-1}|\boldsymbol{Z}_{n-1})\, d\boldsymbol{x}_{n-1}$$

- At time n, we apply Bayes' rule:

$$p(\boldsymbol{x}_n|\boldsymbol{Z}_n) \;=\; \frac{p(\boldsymbol{z}_n|\boldsymbol{x}_n)\;\, p(\boldsymbol{x}_n|\boldsymbol{Z}_{n-1})}{p(\boldsymbol{z}_n|\boldsymbol{Z}_{n-1})}$$

- Denominator is integral over numerator:

Kalman Filter: Bayesian Derivation

- Bayes rule at time n

$$p(\boldsymbol{x}_n|\boldsymbol{Z}_n) = \frac{p(\boldsymbol{z}_n|\boldsymbol{x}_n)\ p(\boldsymbol{x}_n|\boldsymbol{Z}_{n-1})}{p(\boldsymbol{z}_n|\boldsymbol{Z}_{n-1})}$$

- Do we know these probabilities? All of them Gaussian!

Transition	$p(\boldsymbol{x}_n|\boldsymbol{x}_{n-1})$	$\mathcal{N}\left(F\,\boldsymbol{x}_{n-1},\ Q\right)$
Observation	$p(\boldsymbol{z}_n|\boldsymbol{x}_n)$	$\mathcal{N}\left(H\,\boldsymbol{x}_n,\ R\right)$
Posterior at $(n-1)$	$p(\boldsymbol{x}_{n-1}|\boldsymbol{Z}_{n-1})$	$\mathcal{N}\left(\widehat{\boldsymbol{x}}_{n-1|n-1},\ P_{n-1|n-1}\right)$

- Gives us a view as a model that makes transitions (stochastic) and generates observations (also stochastic), as we move in time.
- Sample transitions from probability density $\mathcal{N}\left(F\,\boldsymbol{x}_{n-1},\ Q\right)$
- Sampe observations from probability density $\mathcal{N}\left(H\,\boldsymbol{x}_n,\ R\right)$
- At time $n-1$:

$$p\left(\boldsymbol{x}_{n-1}\,|\,Z_{n-1}\right) = \mathcal{N}\left(\boldsymbol{x}_{n-1}\,|\,\widehat{\boldsymbol{x}}_{n-1|n-1},\ P_{n-1|n-1}\right)$$

- Let's predict and correct from here....

Bayesian Derivation (cont'd)

- Predicting from time $n-1$ to time n:

$$\begin{aligned} p\left(\boldsymbol{x}_n \,|\, \boldsymbol{Z}_{n-1}\right) &= \mathcal{N}\left(\boldsymbol{x}_n \,|\, \widehat{\boldsymbol{x}}_{n|n-1},\, P_{n|n-1}\right) \\ &= \mathcal{N}\left(A\,\widehat{\boldsymbol{x}}_{n-1|n-1},\ \ AP_{n-1|n-1}A^t + Q\right) \end{aligned}$$

- Now we are at time n. Joint distribution:

$$\begin{aligned} p\left(\boldsymbol{x}_n,\, \boldsymbol{z}_n \,|\, \boldsymbol{Z}_{n-1}\right) &= p\left(\boldsymbol{z}_n \,|\, \boldsymbol{x}_n\right)\ p\left(\boldsymbol{x}_n \,|\, Z_{n-1}\right) \\ &= \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x}_n \\ \boldsymbol{z}_n \end{bmatrix} \,\middle|\, \begin{bmatrix} \widehat{\boldsymbol{x}}_{n|n-1} \\ H\,\widehat{\boldsymbol{x}}_{n|n-1} \end{bmatrix},\ \begin{bmatrix} P_{n|n-1} & P_{n|n-1}H^t \\ HP_{n|n-1} & HP_{n|n-1}H^t + R \end{bmatrix}\right) \end{aligned}$$

- The conditional we need is

$$p\left(\boldsymbol{x}_n \,|\, \boldsymbol{z}_n\right) = \mathcal{N}\left(\boldsymbol{x}_n \,|\, \widehat{\boldsymbol{x}}_{n|n},\, P_{n|n}\right)$$

- We can write this out immediately (next page)!

Bayesian Derivation (cont'd)

- What we have

$$p\left(\boldsymbol{x}_n,\, \boldsymbol{z}_n \,|\, Z_{n-1}\right) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x}_n \\ \boldsymbol{z}_n \end{bmatrix} \,\middle|\, \begin{bmatrix} \widehat{\boldsymbol{x}}_{n|n-1} \\ H\,\widehat{\boldsymbol{x}}_{n|n-1} \end{bmatrix},\ \begin{bmatrix} P_{n|n-1} & P_{n|n-1}H^t \\ HP_{n|n-1} & HP_{n|n-1}H^t + R \end{bmatrix} \right)$$

- What we know:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} &\sim \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}, \begin{bmatrix} A & C \\ C^t & B \end{bmatrix} \\ \boldsymbol{x}|\boldsymbol{y} &\sim \mathcal{N}\left(\boldsymbol{a} + CB^{-1}(\boldsymbol{y}-\boldsymbol{b}),\ A - CB^{-1}C^t\right) \end{aligned}$$

- Apply the relationship to find: $p\left(\boldsymbol{x}_n \,|\, \boldsymbol{z}_n\right) = \mathcal{N}\left(\boldsymbol{x}_n \,|\, \widehat{\boldsymbol{x}}_{n|n},\, P_{n|n}\right)$
- Hence we have:

$$\begin{aligned} \widehat{\boldsymbol{x}}_{n|n} &= \widehat{\boldsymbol{x}}_{n|n-1} + P_{n|n-1}\,H^t\left(H\,P_{n|n-1}\,H^t + R\right)^{-1}\left(\boldsymbol{z}_n - H\,\widehat{\boldsymbol{x}}_{n|n-1}\right) \\ &= \widehat{\boldsymbol{x}}_{n|n-1} + K_n\left(\boldsymbol{z}_n - H\,\widehat{\boldsymbol{x}}_{n|n-1}\right) \\ P_{n|n} &= P_{n|n-1} - P_{n|n-1}H^t\left(H\,P_{n|n-1}\,H^t + R\right)^{-1} H\,P_{n|n-1} \\ &= \left(I - K_n\,H\right)P_{n|n-1} \end{aligned}$$