2.160 Identification, Estimation, and Learning

Part 2 Kalman and Bayes Filters

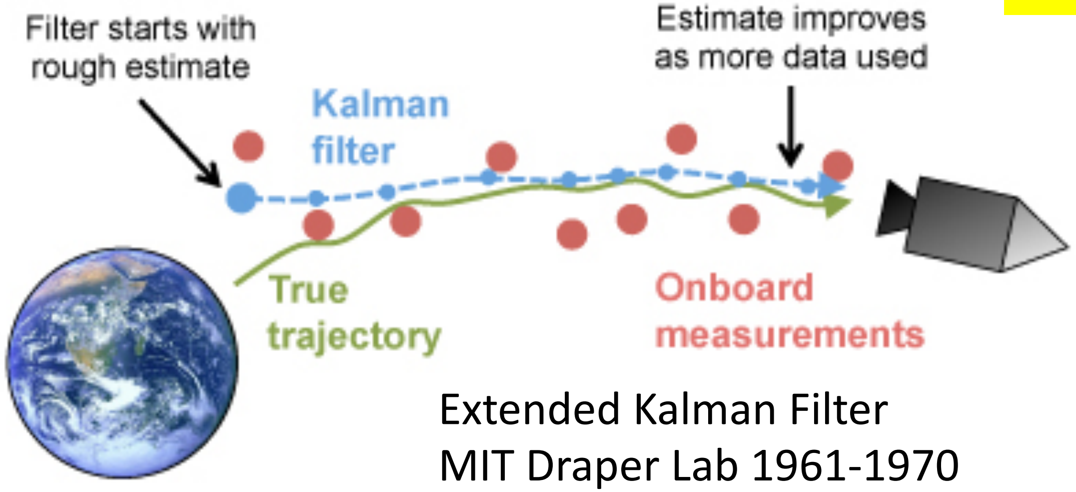
Lecture 7

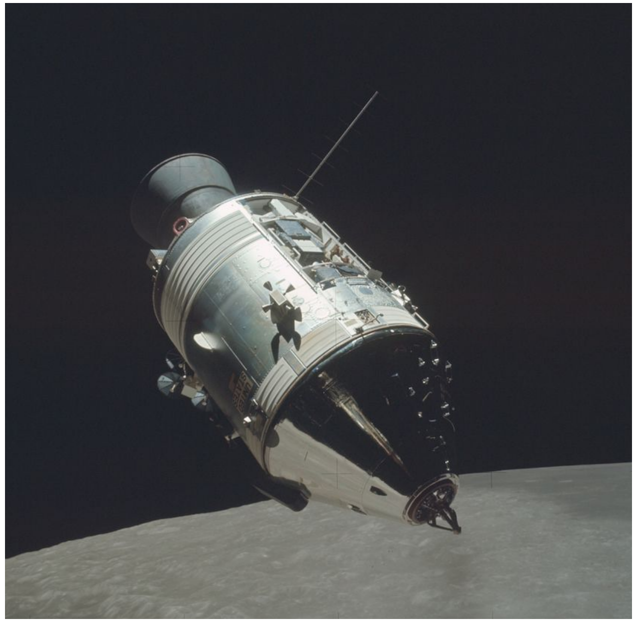
Discrete-Time Kalman Filter

H. Harry Asada
Department of Mechanical Engineering
MIT

Kalman Filter applied to the Apollo Moon Mission

Moon Shot



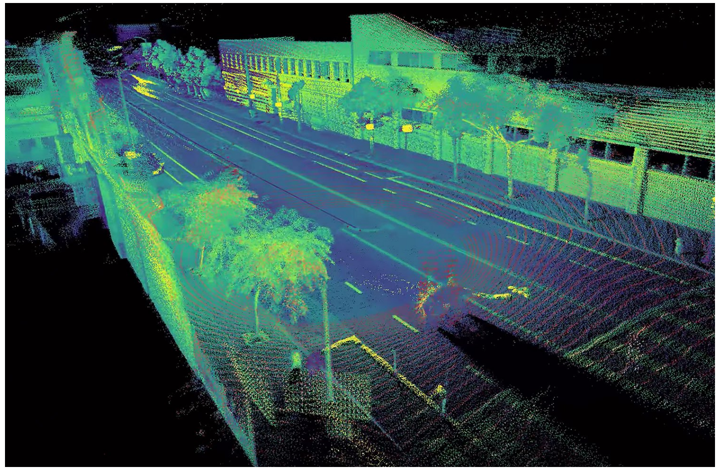


Rudolf E. Kalman

Extended Kalman Filter
MIT Draper Lab 1961-1970

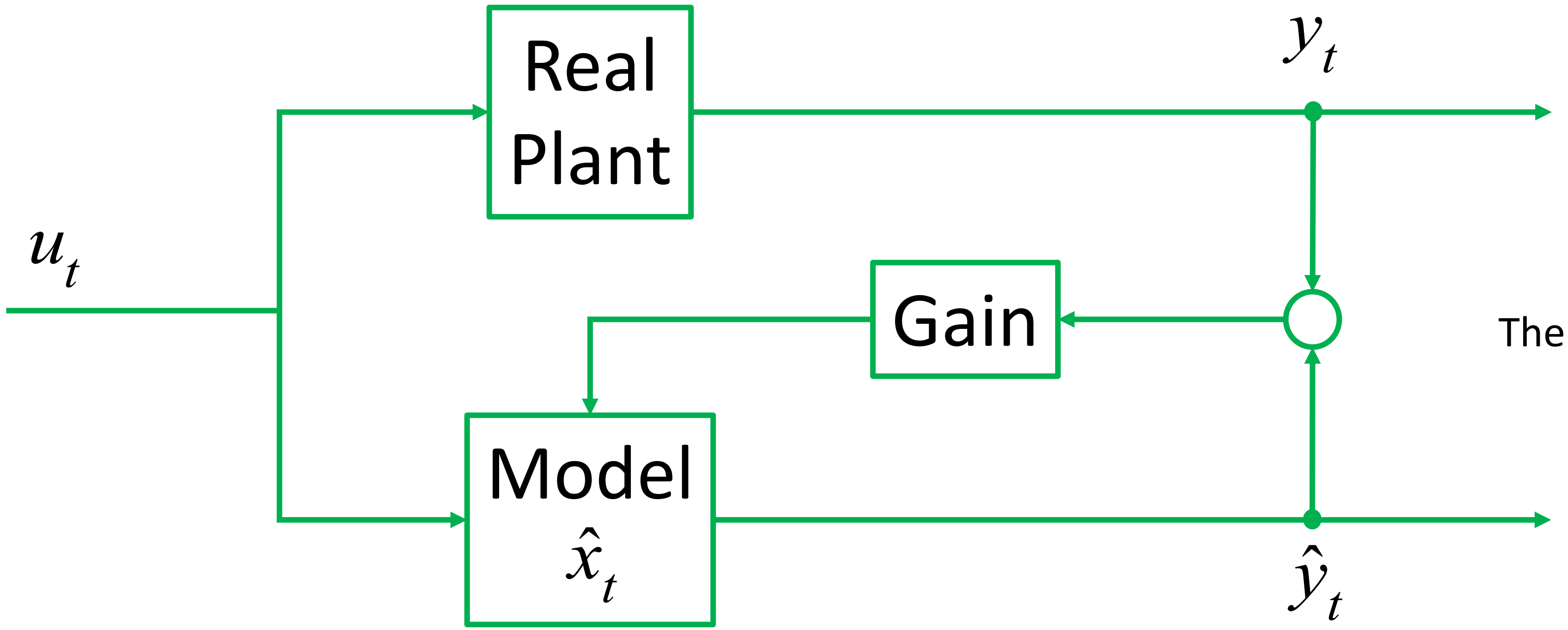
Over 50 years later

Self-Driving Car



Simultaneous Localization And Mapping (SLAM)

Prediction Error Correction Formalism



The Luenberger Observer

- ❑ Estimating state variables, as opposed to plant parameters: State Observer

$$\hat{\theta} \rightarrow \hat{x}_t$$

- ❑ Real-time recursive computation
  - Prediction error fed back to state estimation

Kalman Filter

- Quantifying Uncertainty
  - Process noise
  - Measurement noise

+

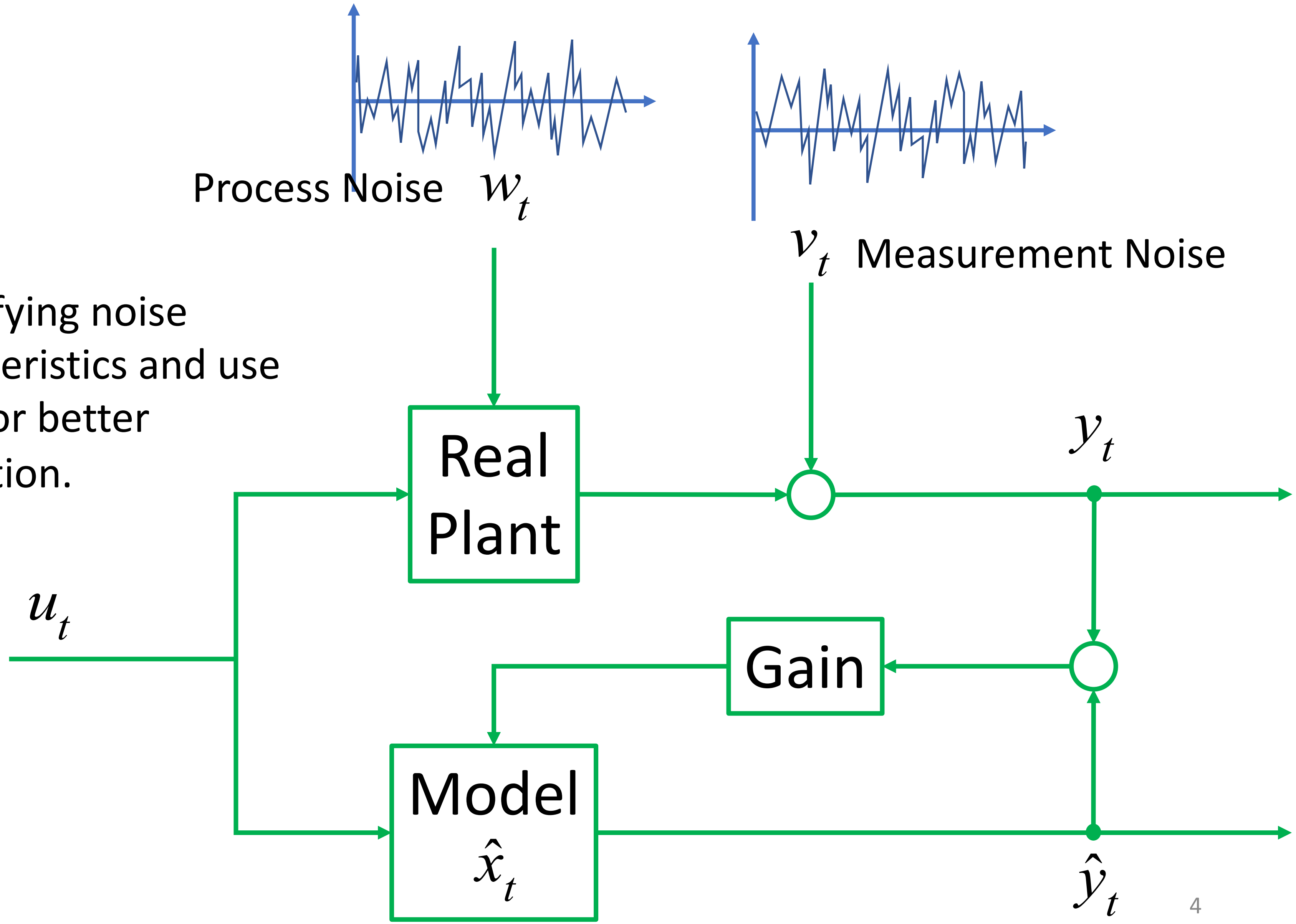
- State Observer

=

Kalman Filter

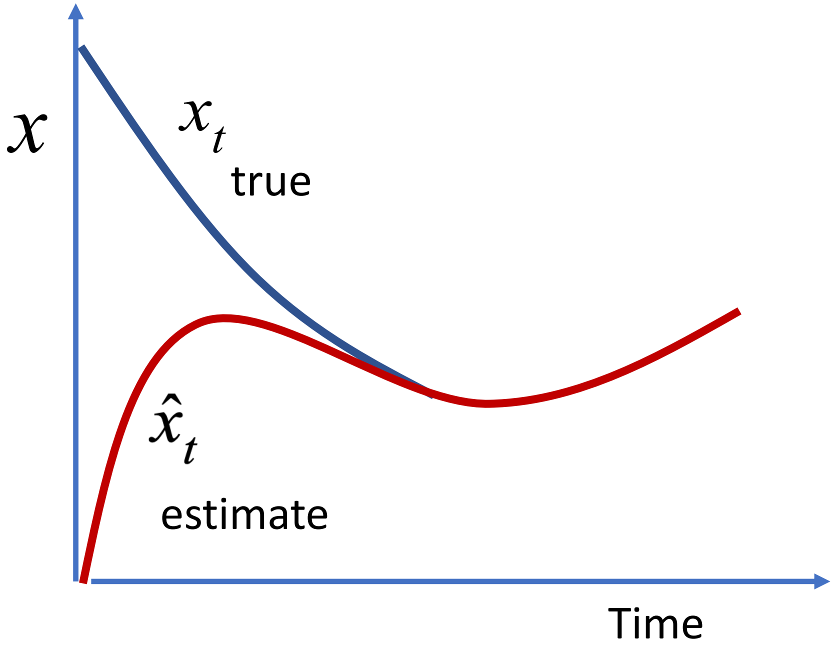
Recursive Least Squares — Bayes Filter

Quantifying noise characteristics and use them for better estimation.



Recap

Discrete-Time State Observer Formulation



- Plant Model: Linear Time-Varying System
  - State (transition) equation

$$x_{t+1} = A_t x_t + B_t u_t$$

where $x_t \in \Re^{n\times 1}$ State vector $u_t \in \Re^{r\times 1}$ input

  - Output equation (measurement equation)

$$y_t = H_t x_t \qquad y_t \in \Re^{\ell\times 1}$$

- Luenberger Observer

$$\hat{x}_{t+1} = A_t \hat{x}_t + B_t u_t + L(\underline{y_t - \hat{y}_t})$$

Prediction Error: negative feedback

  - If the system is observable, the estimated state exponentially converges to the true state.

$$\hat{x}_t \xrightarrow[t\to\infty]{} x_t$$

Convergence speed : Pole placement

- Kalman filter uses an optimal gain based on statistical properties of noise.

Kalman Filter v.s. Recursive Least Squares

There is no fundamental difference between parameter estimation and state estimation.

$$\hat{x}_{t+1} = A_t \hat{x}_t + B_t u_t + K_t (y_t - \hat{y}_t)$$

Treat parameters to estimate as state variables that are constant but unknown. Replacing A_t by the identity matrix and setting B_t to 0 yield

$$\hat{\theta}_{t+1} = I \hat{\theta}_t + 0 u_t + K_t (y_t - \hat{y}_t)$$

Kalman Filter has been extended to many filters.

Kalman Filter

Discrete time

Continuous time

Kalman-Bucy Filter

Riccati Equation

Convergence analysis

Linear plant dynamics

Nonlinear plant dynamics

Extended Kalman Filter

Unscented Kalman Filter

Gaussian noise

Non-Gaussian noise

Bayes Filter

Particle Filter

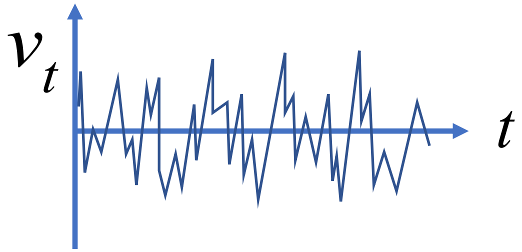
Formulation of Discrete Kalman Filter: Quantification of Uncertainty

- First we quantify both measurement noise and process noise with respect to mean and correlation/covariance. We assume that noise is wide-sense stationary.
- The mean of noise is assumed zero, and constant. If the mean is not zero, then we can shift the origin of coordinate axes.

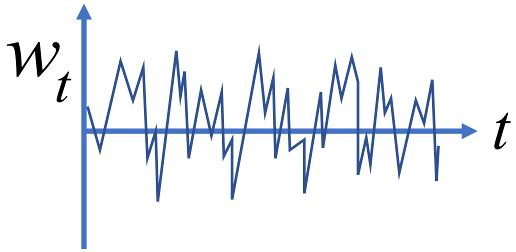
Measurement Noise

$$E[v_t] = 0$$

Zero Mean



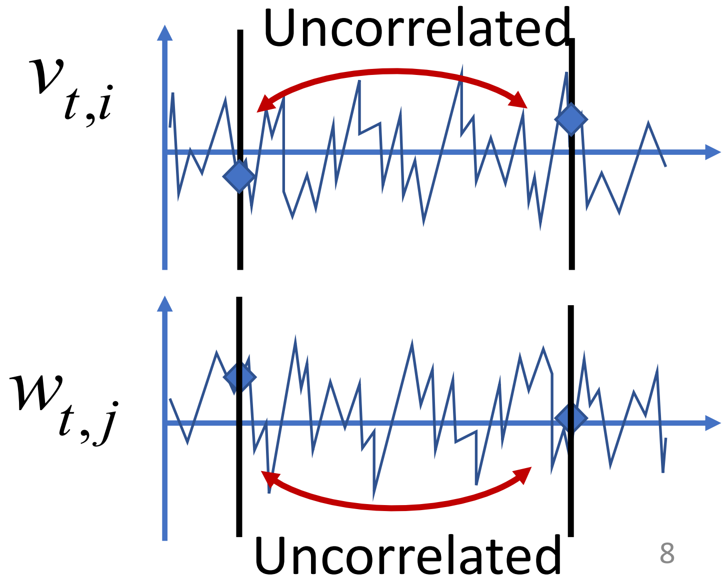
Process Noise



$$E[w_t] = 0$$

- We assume Uncorrelated (White) noise.
- Note that each of measurement and process noise is a vectorial quantity. Auto-correlation is the correlation between two time slices of the same random process (the same component of a noise vector).

Uncorrelated (White) Noise



Auto-Correlation

$$E[v_{t,i} v_{s,i}] = 0$$

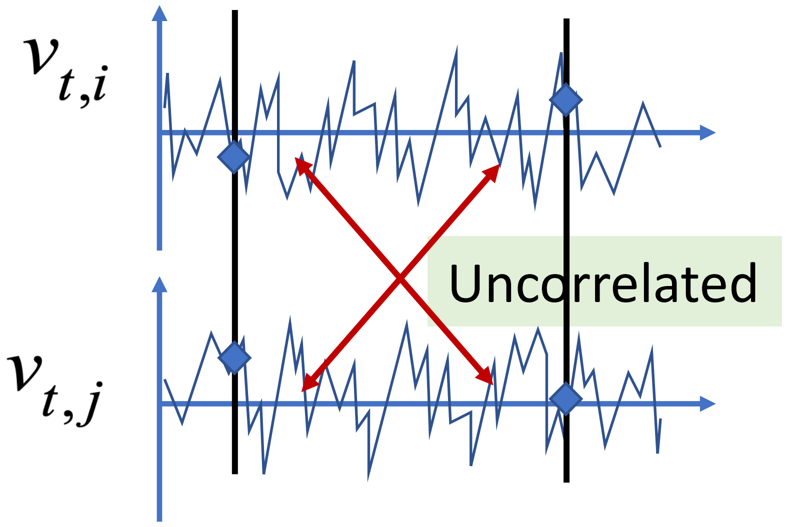
For all t and s, $t \neq s$

$$E[w_{t,j} w_{s,j}] = 0$$

Quantification of Uncertainty: Cross-Correlation

- We also characterize noise properties with respect to Cross-Correlation: the correlation between two different components of the same noise vector or the one between measurement and process noise.

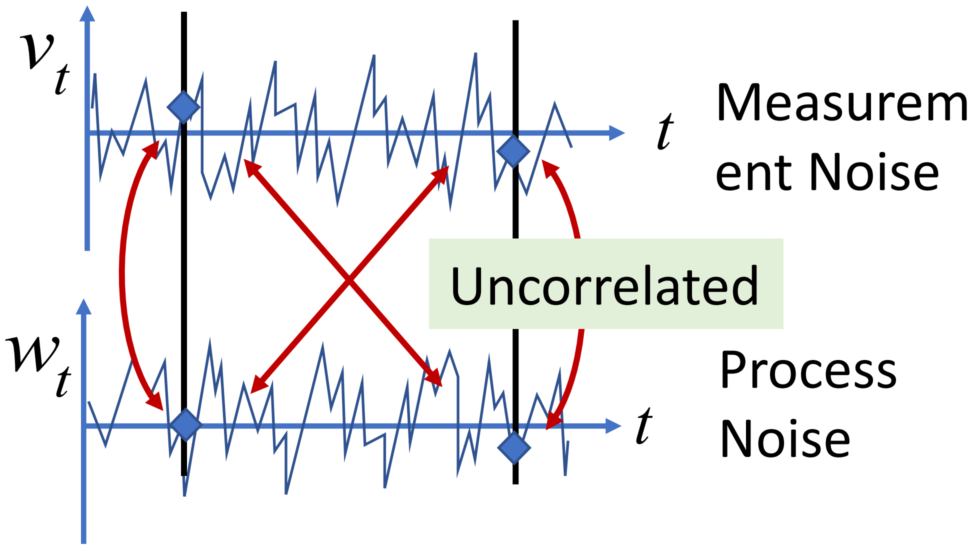
Measurement or Process Noise



$$v_t = \begin{pmatrix} v_{t,1} \\ \vdots \\ v_{t,\ell} \end{pmatrix}$$

$$w_t = \begin{pmatrix} w_{t,1} \\ \vdots \\ w_{t,\ell} \end{pmatrix}$$

Correlation between measurement and process noise



$E[v_{t,i}v_{s,j}] = 0$ For all t and s, $t \neq s$

$E[w_{t,i}w_{s,j}] = 0$ For all i and j, $i \neq j$

Uncorrelated

$E[v_{t,i}w_{s,j}] = 0$ For all t and s, For all i and j,

Quantification of Uncertainty: Covariance Matrices

- In vector and matrix form, auto-correlation and cross-correlation can be collectively expressed as

$$E[v_t v_s^T] = \begin{pmatrix} E[v_{t,1}v_{s,1}] & \cdots & E[v_{t,1}v_{s,\ell}] \\ \vdots & \ddots & \vdots \\ E[v_{t,\ell}v_{s,1}] & \cdots & E[v_{t,\ell}v_{s,\ell}] \end{pmatrix}$$

- When $t = s$, the diagonal terms represent variance of the individual noise component and off-diagonal terms co-variance. $E[v_{t,i}^2] = \sigma_i^2$
- In summary,
  - Measurement noise covariance

$$E[v_t v_s^T] = \begin{cases} R_t; & t = s \\ 0; & t \neq s \end{cases}$$

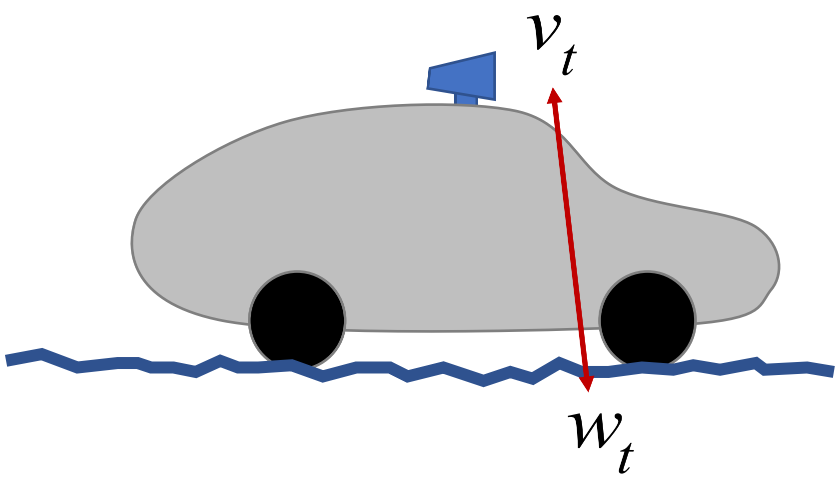
  - Covariance R_t is assumed to be positive definite.
  - There is no perfect sensor.

  - Process noise covariance

$$E[w_t w_s^T] = \begin{cases} Q_t; & t = s \\ 0; & t \neq s \end{cases}$$

  - Covariance Q_t is assumed to be positive semi-definite.

Measurement and Process Noise Cross-Correlation



PS#2 Problem 2

$$E[w_t v_s^T] = 0$$ For all t and s,

- However, in some application, process noise also influences measurement, as in the case of a self-driving car.
- For the sake of simplicity, we assume that there is no correlation between them, but this assumption can be removed.

- Finally, we assume that these noise terms additively disturb the process. Namely the **state and measurement equations** are given by

$$x_{t+1} = A_t x_t + B_t u_t + \boxed{G_t w_t}$$
$$y_t = H_t x_t + \boxed{v_t}$$

Additive noise

- G_t represents how the process noise disturb the state variables.

Optimal Filtering Problem

- Find a state estimate, $\hat{x}_t$, that minimizes the mean squared prediction error:

$$\bar{J}_t = E[|\hat{x}_t - x_t|^2]$$

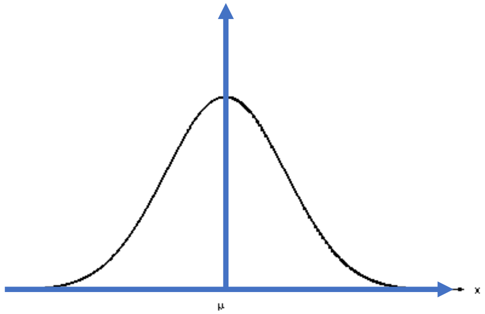
Subject to state and measurement equations

$$x_{t+1} = A_t x_t + B_t u_t + G_t w_t$$

$$y_t = H_t x_t + v_t$$

Linear Time Varying System

where process noise, w_t, and measurement noise, v_t , are uncorrelated (White) noise as characterized above.

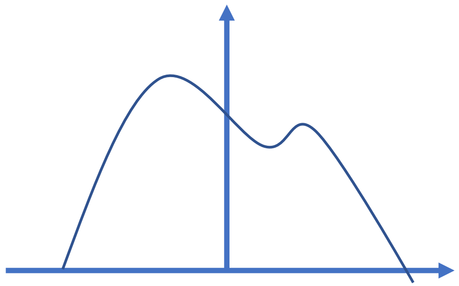


Gaussian

- Assuming that the noise distribution is Gaussian, Kalman Filter is the optimal among linear and nonlinear filters.

$$\hat{x}_{t+1} = A_t \hat{x}_t + B_t u_t + K_t [y_t - \hat{y}_t]$$

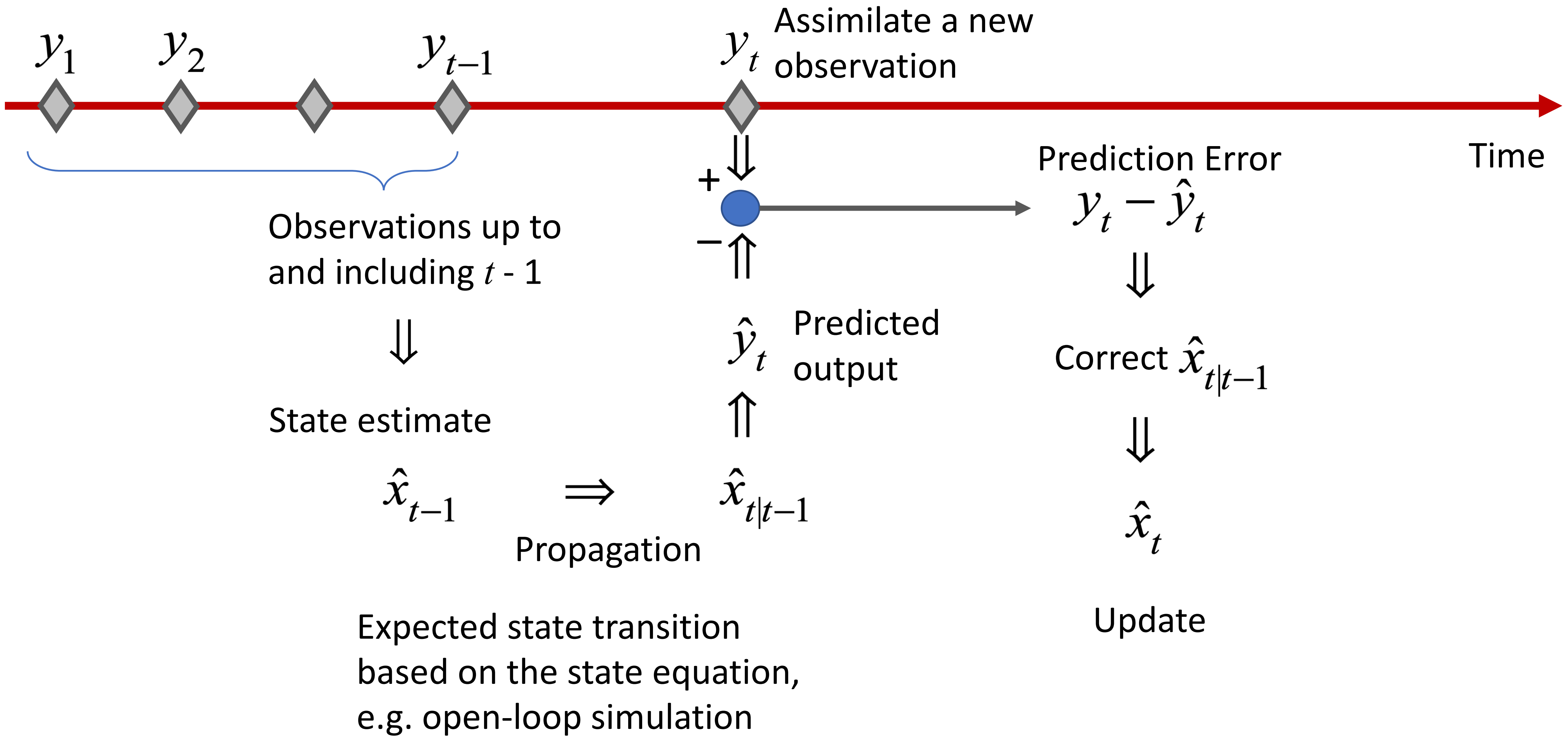
- Assuming that the filter structure is linear, $K_t[y_t - \hat{y}_t]$, Kalman Filter is the optimal linear filter, regardless of noise distribution.



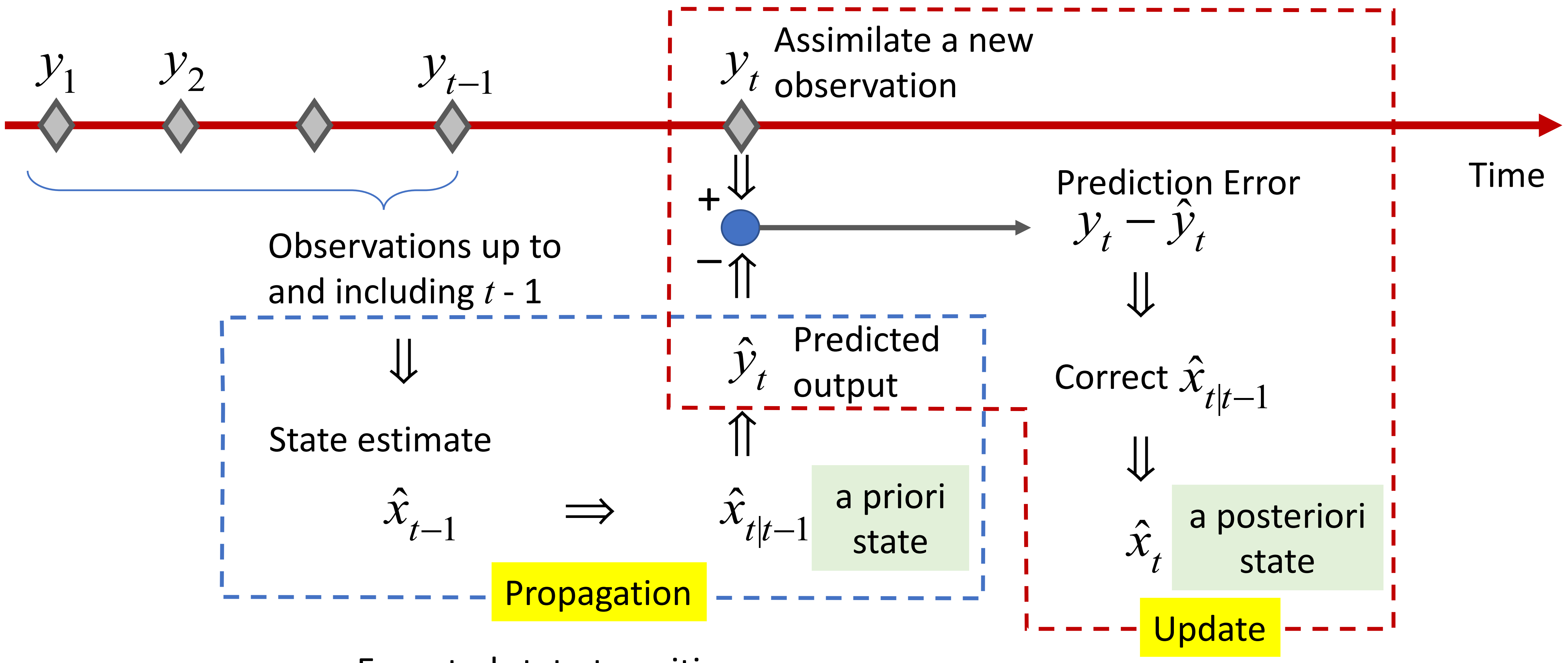
Non-Gaussian

We first prove the second problem, and show the proof for the first problem at the Bayes Filter lecture.

The Flow of the Discrete Kalman Filter Algorithm

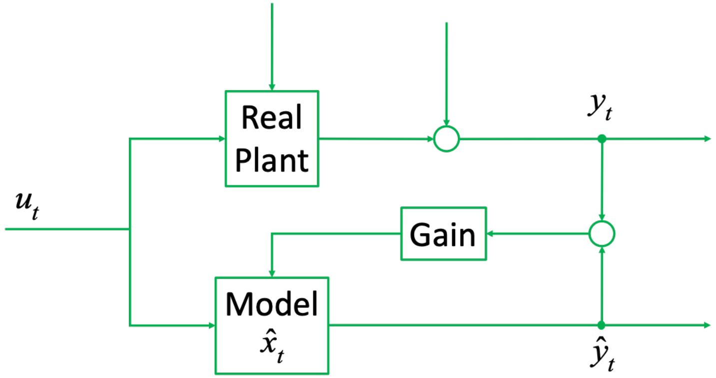


The Flow of the Discrete Kalman Filter Algorithm



Expected state transition based on the state equation, e.g. open-loop simulation

Propagation of State



- Using the state equation, we want to predict the transition of state:

$$x_t = A_{t-1}x_{t-1} + \cancel{B_{t-1}u_{t-1}} + G_{t-1}w_{t-1}$$

- The deterministic term, $B_{t-1}u_{t-1}$, can be omitted by setting, $u_{t-1} = 0$, without loss of generality.
- Taking expectation yields

This is the state estimate at time t - 1 Zero mean

$$E[x_t] = E[A_{t-1}x_{t-1} + G_{t-1}w_{t-1}] = A_{t-1}\overset{\hat{x}_{t-1}}{E[x_{t-1}]} + G_{t-1}\cancel{E[w_{t-1}]}$$

- This expected value of state, $E[x_t]$, is the predicted state at time t based on the estimated state at $t-1$. This is ***a priori*** *state* estimate denoted by

$$\hat{x}_{t|t-1} = A_{t-1}\hat{x}_{t-1}$$

Before measurement of output $y(t)$ is available.

- Using the output equation, the expected output is constructed as

$$\hat{y}_t = E[y_t] = E[H_t x_t + v_t] = H_t E[x_t] + \cancel{E[v_t]}$$

Zero mean

$$\therefore \hat{y}_t = H_t \hat{x}_{t|t-1} = H_t A_{t-1}\hat{x}_{t-1}$$

- This is a predicted output to be compared to an actual measured output.

State Update

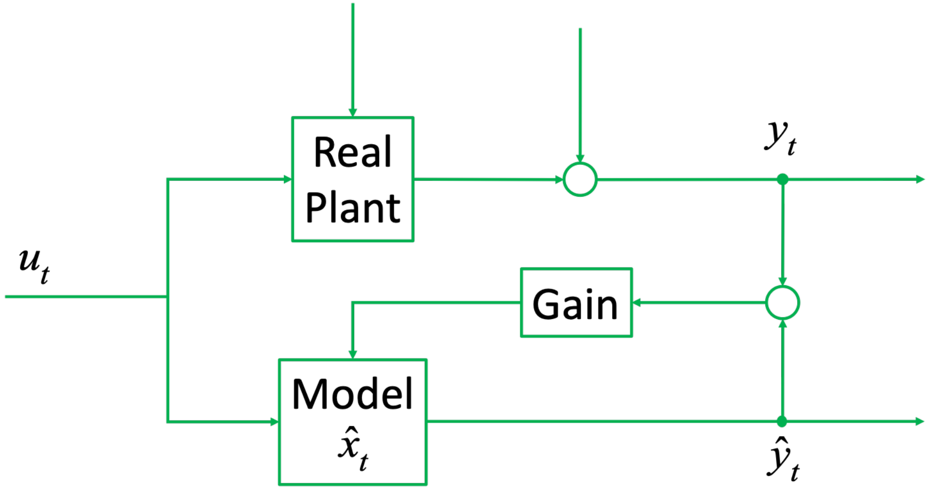
- Let y_t be a newly observed output. The predicted output $\hat{y}_t$ based on the previous state estimate is then compared to the actual measured output and the error is used for correcting, or updating, the a priori state estimate:

$$\hat{x}_t = \hat{x}_{t|t-1} + K_t[y_t - \hat{y}_t] \qquad (23)$$

- Note that we assume this linear update law in this proof. Gain $K_t \in \Re^{n\times\ell}$ is called the Kalman gain. Our goal is to find an optimal gain that minimizes the mean squared prediction error: $\bar{J}_t = E[|\hat{x}_t - x_t|^2]$

- Optimal gain: .

$$K_t = \arg\min_{K_t} \bar{J}_t = \arg\min_{K_t} E[|\hat{x}_t - x_t|^2]$$



Effect of Kalman Gain K_t

- Prediction Law: $$\hat{x}_t = \hat{x}_{t|t-1} + K_t[y_t - \hat{y}_t] \qquad (23)$$

- Note that there are two types of prediction error.
  - a posteriori error: $e_t \triangleq \hat{x}_t - x_t$
  - a priori error: $\varepsilon_t \triangleq \hat{x}_{t|t-1} - x_t$

- These two prediction errors are related to each other. Using the state update law (23),

$$\begin{aligned} e_t &= \hat{x}_t - x_t = \hat{x}_{t|t-1} + K_t[y_t - \hat{y}_t] - x_t \\ &= \hat{x}_{t|t-1} + K_t[H_t x_t + v_t - H_t \hat{x}_{t|t-1}] - x_t \\ &= \underbrace{\hat{x}_{t|t-1} - x_t}_{\varepsilon_t} + K_t H_t \underbrace{(x_t - \hat{x}_{t|t-1})}_{-\varepsilon_t} + K_t v_t \end{aligned}$$

$$\therefore \quad e_t = (I - K_t H_t)\varepsilon_t + K_t v_t$$

a priori error ε_t, is attenuated by $(I - K_t H_t)$

But, the measurement noise is amplified by K_t .

Trade-off

- As the gain K_t becomes higher, the a priori error is more reduced but the measurement noise is amplified.
- An optimal gain may exist by making the trade-off between the two.

Computation of an optimal gain

- Computation of the squared a posteriori error: $\bar{J}_t = E[|\hat{x}_t - x_t|^2]$
- Recall $e_t = \hat{x}_t - x_t = (I - K_t H_t)\varepsilon_t + K_t v_t$
- Omitting t for brevity,

$$|e|^2 = [(I-KH)\varepsilon + Kv]^T [(I-KH)\varepsilon + Kv]$$
$$= \varepsilon^T \varepsilon + \varepsilon^T H^T K^T KH\varepsilon + v^T K^T Kv - 2\varepsilon^T KH\varepsilon - 2\varepsilon^T H^T K^T Kv + 2\varepsilon^T Kv$$

- Necessary conditions for min $\bar{J}_t(K)$

$$\frac{d\bar{J}_t}{dK} = 0$$ But, $K = \{K_{ij}\}$ is a matrix

$$\frac{d\bar{J}_t}{dK_{ij}} = 0$$ For all i and j, $1 \le i \le n,\ 1 \le j \le \ell$

Lemma Matrix differentiation Rules

- Consider a scalar function: $$f = a^T K\, b = \sum_i \sum_j a_i b_j K_{ij}$$

where $a \in \Re^{n\times 1}, b \in \Re^{\ell\times 1}, K \in \Re^{n\times \ell}$

$$\frac{\partial f}{\partial K_{pq}} = a_p\, b_q$$

Because all others $i \neq p,\ j \neq q$ are zeros, when differentiating by K_{pq} .

$$\therefore \frac{df}{dK} = \left\{ \frac{\partial f}{\partial K_{pq}} \right\} = a\, b^T$$

Note: $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} b_1 & \cdots & b_\ell \end{pmatrix} = \left\{ a_p\, b_q \right\}$

- Consider another scalar function:

$$g = c^T K^T K\, b, \quad c \in \Re^{\ell\times 1}$$

Similarly, we can show that $$\frac{d}{dK}(c^T K^T K b) = Kbc^T + Kcb^T$$

DIY

Computation of Optimal Gain K_t

$$|e|^2 = \varepsilon^T\varepsilon + \varepsilon^T H^T K^T K H\varepsilon + v^T K^T K v - 2\varepsilon^T K H\varepsilon - 2\varepsilon^T H^T K^T K v + 2\varepsilon^T K v$$

- Applying these rules of differentiation by matrix K to the derivative of squared error,

$$\frac{d|e|^2}{dK} = \cancel{\frac{d}{dK}\varepsilon^T\varepsilon} + \frac{d}{dK}(2\varepsilon^T K v - 2\varepsilon^T K H\varepsilon) \quad \text{-----------------} \quad \text{Rule 1}$$

$$+\frac{d}{dK}(\varepsilon^T H^T K^T K H\varepsilon + v^T K^T K v - 2\varepsilon^T H^T K^T K v) \quad \text{----} \quad \text{Rule 2}$$

$$= 2\varepsilon v^T - 2\varepsilon\varepsilon^T H^T + KH\varepsilon\varepsilon^T H^T + KH\varepsilon\varepsilon^T H^T + 2Kvv^T - 2Kv\varepsilon^T H^T - 2KH\varepsilon v^T$$

Rule 1 $\qquad \dfrac{d}{dK}(a^T K\, b) = a\, b^T$

Rule 2 $\qquad \dfrac{d}{dK}(c^T K^T K b) = Kbc^T + Kcb^T$

Computation of Optimal Gain K_t

$$\frac{d|e|^2}{dK} = 2\varepsilon v^T - 2\varepsilon\varepsilon^T H^T + KH\varepsilon\varepsilon^T H^T + KH\varepsilon\varepsilon^T H^T + 2Kvv^T - 2Kv\varepsilon^T H^T - 2KH\varepsilon v^T$$

- Taking expectation and setting it to zero,

$$E[\varepsilon_t v_t^T] - \underbrace{E[\varepsilon_t \varepsilon_t^T]}_{P_{t|t-1}} H_t^T + K_t H_t E[\varepsilon_t \varepsilon_t^T] H_t^T + K_t \underbrace{E[v_t v_t^T]}_{R_t} - K_t E[v_t \varepsilon_t^T] H_t^T = 0$$

Measurement noise covariance

- Define a priori error covariance: $P_{t|t-1} \triangleq E[\varepsilon_t \varepsilon_t^T]$

- Examine $E[\varepsilon_t v_t^T]$

Examine $E[\varepsilon_t v_t^T]$

$$E[\varepsilon_t v_t^T] = E[(\hat{x}_{t|t-1} - x_t)v_t^T] \quad \leftarrow \varepsilon_t = \hat{x}_{t|t-1} - x_t$$

$$= A_{t-1}\boxed{E[\hat{x}_{t-1}v_t^T]} - \boxed{E[x_t v_t^T]} \quad \leftarrow \hat{x}_{t|t-1} = A_{t-1}\hat{x}_{t-1}$$

Examine

$$x_t = A_{t-1}x_{t-1} + G_{t-1}w_{t-1}$$

$$\boxed{E[x_t v_t^T]} = E[(A_{t-1}x_{t-1} + G_{t-1}w_{t-1})v_t^T]$$

$$= A_{t-1}E[\boxed{x_{t-1}}v_t^T] + G_{t-1}\cancel{E[w_{t-1}v_t^T]}$$

From x_{t-1}, only process noise terms $w_{t-2}, w_{t-3}, \cdots$ come out, which do not correlate with the measurement noise v_t. $E[w_t v_s^T] = 0$

$$\therefore E[x_t v_t^T] = 0$$

For all t and s,

Process noise and measurement noise are not correlated.

Examine

$$\boxed{E[\hat{x}_{t-1}v_t^T]} = E[(\hat{x}_{t-1|t-2} + K_{t-1}(y_{t-1} - \hat{y}_{t-1}))v_t^T] \quad \leftarrow \hat{x}_{t-1} = \hat{x}_{t-1|t-2} + K_{t-1}(y_{t-1} - \hat{y}_{t-1})$$

$$= E[(A_{t-2}\hat{x}_{t-2} + K_{t-1}(H_{t-1}x_{t-1} + v_{t-1} - \hat{y}_{t-1})v_t^T] \quad \leftarrow y_{t-1} = H_{t-1}x_{t-1} + v_{t-1}$$

$$= A_{t-2}\cancel{E[\hat{x}_{t-2}v_t^T]} + K_{t-1}H_{t-1}\cancel{E[x_{t-1}v_t^T]} + \cancel{E[v_{t-1}v_t^T]} - \cancel{E[\hat{y}_{t-1}v_t^T]} = 0$$

v_t does not correlate with past $\hat{x}_{t-2}, x_{t-1}, \hat{y}_{t-1}$

Uncorrelated $E[v_t v_s^T] = 0; \quad t \neq s$

Optimal Gain K_t

- From the above examination: $E[\varepsilon_t v_t^T] = 0$

- Back to the optimality conditions:

$$\cancel{E[\varepsilon_t v_t^T]} - \underbrace{E[\varepsilon_t \varepsilon_t^T]}_{P_{t|t-1}} H_t^T + K_t H_t E[\varepsilon_t \varepsilon_t^T] H_t^T + K_t \underbrace{E[v_t v_t^T]}_{R_t} - K_t \cancel{E[v_t \varepsilon_t^T]} H_t^T = 0$$

$$-P_{t|t-1} H_t^T + K_t H_t P_{t|t-1} H_t^T + K_t R_t = 0$$

- The optimal gain can be obtained from: $K_t(\underbrace{H_t P_{t|t-1} H_t^T}_{\text{Positive semi-definite}} + \overbrace{R_t}^{\text{Positive definite}}) = P_{t|t-1} H_t^T$

- Note that matrix $H_t P_{t|t-1} H_t^T + R_t$ is positive-definite and invertible:

$$K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1}$$

This is the Kalman Gain

So, how can we find a priori error covariance ?

Recursive Formula for Obtaining $P_{t|t-1} \triangleq E[\varepsilon_t \varepsilon_t^T]$

- a priori error covariance can be computed recursively together with another error covariance.
- Define a posteriori error covariance: $P_t \triangleq E[e_t e_t^T]$
- Recall that a priori error and a posteriori error are related: $e_t = (I - K_t H_t)\varepsilon_t + K_t v_t$
- With this, the two error covariances are related as:

$$
\begin{aligned}
E[e_t e_t^T] &= E[((I - K_t H_t)\varepsilon_t + K_t v_t)((I - K_t H_t)\varepsilon_t + K_t v_t)^T] \\
&= (I - K_t H_t) E[\varepsilon_t \varepsilon_t^T](I - K_t H_t)^T \\
&\quad + K_t E[v_t v_t^T] K_t^T + (I - K_t H_t)\cancel{E[\varepsilon_t v_t^T]} K_t^T + K_t \cancel{E[v_t \varepsilon_t^T]}(I - K_t H_t)^T
\end{aligned}
$$

$$
\therefore P_t = (I - K_t H_t) P_{t|t-1} (I - K_t H_t)^T + K_t R_t K_t^T
$$

Recursive Formula of Covariances, $P_{t|t-1}$ and P_t

- The previous expression can be further simplified by using the optimal (Kalman) gain solution.

$$P_t = (I - K_t H_t) P_{t|t-1} (I - K_t H_t)^T + K_t R_t K_t^T \Rightarrow P_t = (I - K_t H_t) P_{t|t-1}$$

$$K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1}$$

- This implies that a priori covariance is reduced by updating the a priori state estimate $\hat{x}_{t|t-1}$ with a newly assimilated measurement, y_t.

This is called Covariance Update: $P_{t|t-1} \mapsto P_t$.

- In turn, a priori covariance $P_{t+1|t}$ can be derived from a posteriori covariance P_t
- Recall: $\varepsilon_{t+1} = \hat{x}_{t+1|t} - x_{t+1} = A_t \hat{x}_t - (A_t x_t + G_t w_t) = A_t e_t - G_t w_t$
- Compute: $P_{t+1|t} = E[\varepsilon_{t+1} \varepsilon_{t+1}^T] = E[(A_t e_t - G_t w_t)(A_t e_t - G_t w_t)^T]$

$$= A_t E[e_t e_t^T] A_t^T + G_t E[w_t w_t^T] G_t^T - A_t E[e_t w_t^T] G_t^T - G_t E[w_t e_t^T] A_t^T$$

- We can show $E[e_t w_t^T] = 0$. Therefore, $P_{t+1|t} = A_t P_t A_t^T + G_t Q_t G_t^T$

DIY

Recursive Formula of Covariances, $P_{t|t-1}$ and P_t

- The last formula is called Covariance Propagation : $P_t \mapsto P_{t+1|t}$

Covariance Propagation

$$P_{t+1|t} = A_t P_t A_t^T + G_t Q_t G_t^T$$

$$t = t + 1$$

$$P_t = (I - K_t H_t) P_{t|t-1}$$

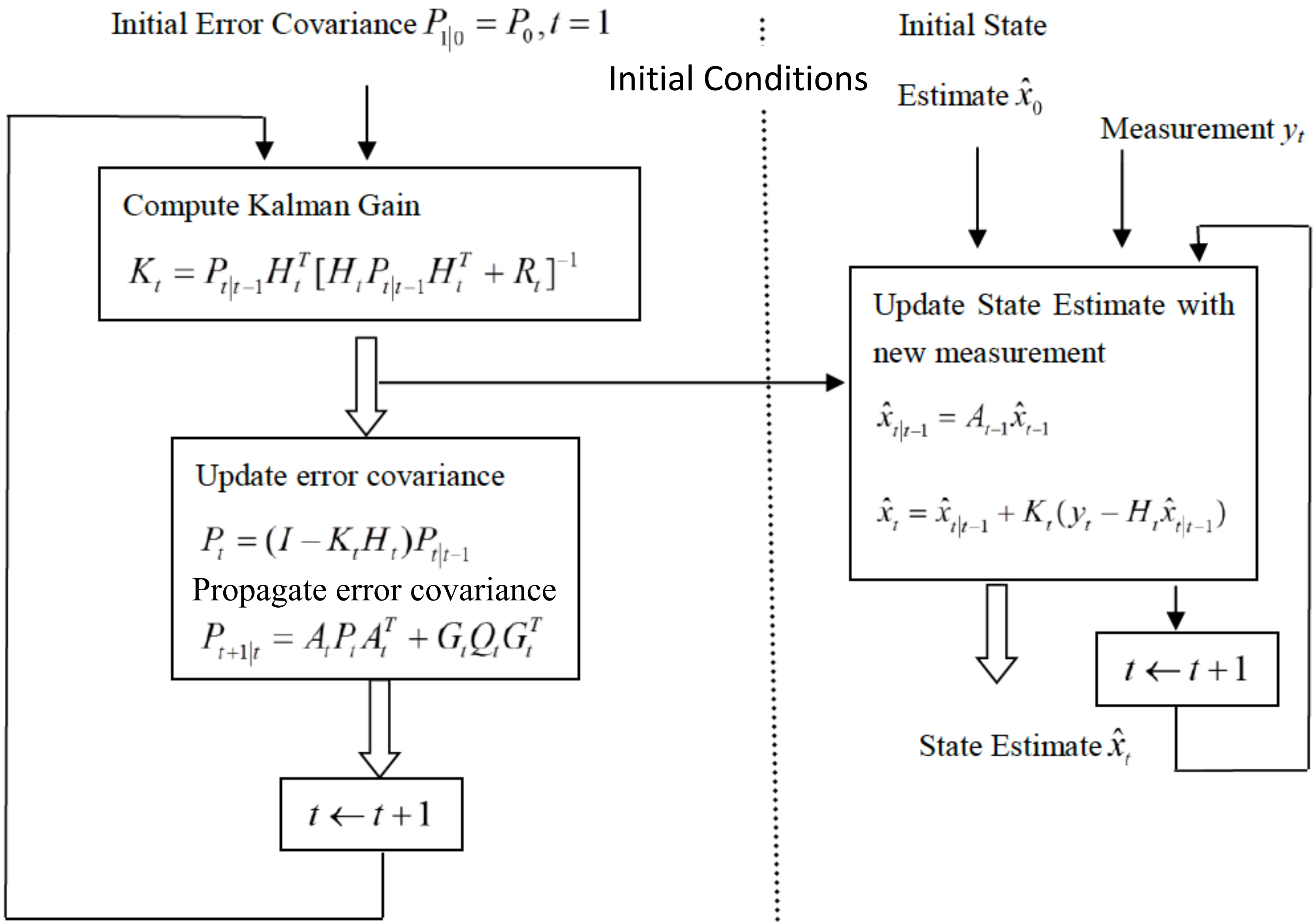
Covariance Update

- Given initial conditions, $P_{1|0}$, covariance matrices can be computed recursively along with the Kalman gain

$$K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1}$$

Kalman Filter

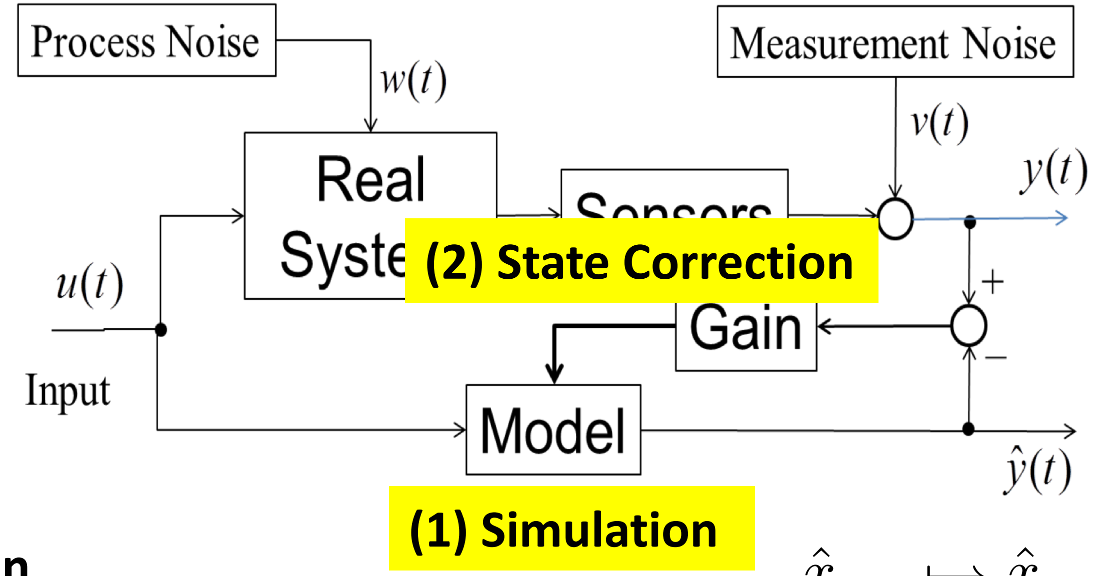
Recursive Computation Algorithm



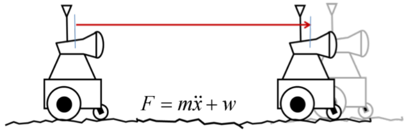
The covariance and Kalman gain computation does not depend on measurements. Therefore it can be computed off-line.

Real-Time
Online Computation

$$P_{t+1|t} = A_t P_t A_t^T + G_t Q_t G_t^T$$



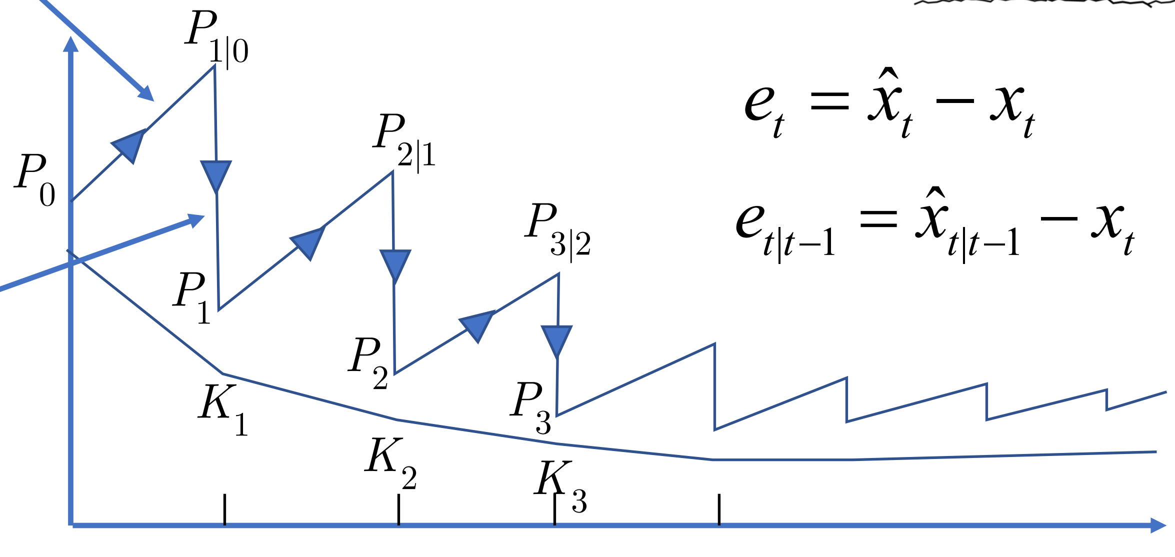
$$\hat{x}_{t-1} \mapsto \hat{x}_{t|t-1}$$



(1) Simulation: Propagation

$$e_t = \hat{x}_t - x_t \qquad P_t = E[e_t e_t^T]$$

$$e_{t|t-1} = \hat{x}_{t|t-1} - x_t \qquad P_{t|t-1} = E[e_{t|t-1} e_{t|t-1}^T]$$



(2) State Correction: Update

$$K_t = P_t H^T R^{-1}$$

$$P_t = (I - K_t H_t) P_{t|t-1}$$