# Lecture 8

## Kalman Filter as a Data-driven Optimization Algorithm

Bao-Jun Cai, 4/22/2026

Introduction to Algorithms for Data Science and Physics IMP@Fudan, 2026

Topics of this lecture:

- Kalman filter, state and measurement
- linear Gaussian model $\mathbf{x}_k = \mathbf{A}_k\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{u}_k + \mathbf{w}_k,\ \mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{n}_k$
- variance reduction in Kalman Filter $\sigma_k^2 = (1 - K_kC_k)\overline{\sigma}_k^2$
- error propagation $\mathbf{K}_k = \overline{\mathbf{S}}_k\mathbf{C}_k^\top(\mathbf{Q}_k + \mathbf{C}_k\overline{\mathbf{S}}_k\mathbf{C}_k^\top)^{-1}$
- nonlinear extension $f(x) \approx \widehat{x} + F\delta x +$ noise
- bias estimate $\mathrm{E}[\widehat{w}] \neq w$

# Linear Gaussian model

state model

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

measurement model

$$\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{n}_k$$

*The Kalman filter is widely applied in physical problems to estimate the evolving state of a system, such as position, velocity, or field variables, by optimally combining imperfect measurements with dynamical models, for example in tracking particles in detectors, reconstructing trajectories in accelerator experiments, or filtering noisy signals in astrophysical observations.*

state: $\mathbf{x}_k \in \mathrm{R}^d$

measurement: $\mathbf{y}_k \in \mathrm{R}^m$

transition matrix: $\mathbf{A}_k \in \mathrm{R}^{d \times d}$

input: $\mathbf{u}_k \in R^u$

control matrix: $\mathbf{B}_k \in \mathrm{R}^{d \times u}$

noise on state: $\mathbf{w}_k \in \mathrm{R}^d \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$

observation matrix: $\mathbf{C}_k \in \mathrm{R}^{m \times d}$

noise on measurement: $\mathbf{n}_k \in \mathrm{R}^m \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$

# Equations for Kalman filter

prediction: $\overline{\mathbf{m}}_k = \mathbf{A}_k \mathbf{m}_{k-1} + \mathbf{B}_k \mathbf{u}_k$

$$\overline{\mathbf{S}}_k = \mathbf{A}_k \mathbf{S}_{k-1} \mathbf{A}_k^\top + \mathbf{R}_k$$

measurement: $\mathbf{m}_k = \overline{\mathbf{m}}_k + \mathbf{K}_k \left( \mathbf{y}_k - \mathbf{C}_k \overline{\mathbf{m}}_k \right)$

$$\mathbf{S}_k = \left( \mathsf{I} - \mathbf{K}_k \mathbf{C}_k \right) \overline{\mathbf{S}}_k$$

$$\mathbf{K}_k = \overline{\mathbf{S}}_k \mathbf{C}_k^\top \left( \mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{Q}_k \right)^{-1}$$

Kalman gain: $\mathbf{K}_k = \overline{\mathbf{S}}_k \mathbf{C}_k^\top \left( \mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{Q}_k \right)^{-1}$

Kalman innovation: $\mathbf{y}_k - \mathbf{C}_k \overline{\mathbf{m}}_k$

$$\mathbf{x}_k \sim \mathcal{N}(\mathbf{m}_k, \mathbf{S}_k)$$

Example: 1D KF

$$\overline{m}_k = A_k m_{k-1} + B_k u_k$$

$$\overline{\sigma}_k^2 = A_k^2 \sigma_{k-1}^2 + R_k$$

$$m_k = \overline{m}_k + K_k (y_k - C_k \overline{m}_k)$$

$$\sigma_k^2 = (1 - K_k C_k) \overline{\sigma}_k^2$$

$$K_k = \frac{\overline{\sigma}_k^2 C_k}{C_k^2 \overline{\sigma}_k^2 + Q_k}$$

$$\sigma_k^2 = \left( 1 - \frac{\overline{\sigma}_k^2 C_k^2}{C_k^2 \overline{\sigma}_k^2 + Q_k} \right) \overline{\sigma}_k^2 = \frac{\overline{\sigma}_k^2}{1 + (C_k^2 / Q_k) \overline{\sigma}_k^2}$$

if $C_k = 0 \rightarrow \sigma_k^2 = \overline{\sigma}_k^2$, $m_k = \overline{m}_k$

no correlation between state and measurement

Ex.: is there definite relation between $\sigma_k^2$ and $\sigma_{k-1}^2$?

# Sketch of KF derivation

$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_k)$: transition probability after obtaining the input $\mathbf{u}_k$ (Gaussian)

$$\bar{p}(\mathbf{x}_k) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{u}_k) p(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1}$$

$\bar{p}(\mathbf{x}_k)$: prediction; $p(\mathbf{x}_{k-1}) \sim \mathcal{N}(\mathbf{x}_{k-1}, \mathbf{S}_{k-1})$ (Gaussian)

$$=\Theta \int \exp\left[-\frac{1}{2}\left(\mathbf{x}_k - \mathbf{A}_k\mathbf{x}_{k-1} - \mathbf{B}_k\mathbf{u}_k\right)^\top \mathbf{R}_k^{-1}\left(\mathbf{x}_k - \mathbf{A}_k\mathbf{x}_{k-1} - \mathbf{B}_k\mathbf{u}_k\right)\right]$$

$$\times \exp\left[-\frac{1}{2}(\mathbf{x}_{k-1} - \mathbf{m}_{k-1})^\top \mathbf{S}_{k-1}^{-1}(\mathbf{x}_{k-1} - \mathbf{m}_{k-1})\right] d\mathbf{x}_{k-1}$$

normalization factor

$$=\Theta \int \exp\left[-\mathcal{L}_k(\mathbf{x}_{k-1}, \mathbf{x}_k)\right] d\mathbf{x}_{k-1}$$

quadratic in both $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$

$$\mathcal{L}_k(\mathbf{x}_{k-1}, \mathbf{x}_k) = \frac{1}{2}\left(\mathbf{x}_k - \mathbf{A}_k\mathbf{x}_{k-1} - \mathbf{B}_k\mathbf{u}_k\right)^\top \mathbf{R}_k^{-1}\left(\mathbf{x}_k - \mathbf{A}_k\mathbf{x}_{k-1} - \mathbf{B}_k\mathbf{u}_k\right)$$

$$+ \frac{1}{2}(\mathbf{x}_{k-1} - \mathbf{m}_{k-1})^\top \mathbf{S}_{k-1}^{-1}(\mathbf{x}_{k-1} - \mathbf{m}_{k-1})$$

$$\mathcal{L}_k(\mathbf{x}_{k-1}, \mathbf{x}_k) = \overbrace{\mathcal{U}_k(\mathbf{x}_{k-1}, \mathbf{x}_k)}^{\text{interaction term}} + \mathcal{L}_k(\mathbf{x}_k)$$

# Sketch of KF equations: continued

$$\mathcal{L}_k(\mathbf{x}_{k-1}, \mathbf{x}_k) = \overbrace{\mathcal{U}_k(\mathbf{x}_{k-1}, \mathbf{x}_k)}^{\text{interaction term}} + \mathcal{L}_k(\mathbf{x}_k)$$

$$\vec{\Phi}_k^{-1} \equiv \frac{\partial^2 \mathcal{L}_k(\mathbf{x}_k, \mathbf{x}_{k-1})}{\partial \mathbf{x}_{k-1}^2} = \mathbf{A}_k^\top \mathbf{R}_k^{-1} \mathbf{A}_k + \mathbf{S}_{k-1}^{-1}$$

$$\mathbf{x}_{k-1}^* = \vec{\Phi}_k \left[ \mathbf{A}_k^\top \mathbf{R}_k^{-1} (\mathbf{x}_k - \mathbf{B}_k \mathbf{u}_k) + \mathbf{S}_{k-1}^{-1} \mathbf{m}_{k-1} \right]$$

$$\mathcal{U}_k(\mathbf{x}_k, \mathbf{x}_{k-1}) = \frac{1}{2} (\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^*)^\top \vec{\Phi}_k^{-1} (\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^*)$$

$$\sim \mathcal{N}(\mathbf{x}_{k-1}^*, \vec{\Phi}_k^{-1}) \quad = \frac{1}{2} \left\{ \mathbf{x}_{k-1} - \Phi_k \left[ \mathbf{A}_k^\top \mathbf{R}_k^{-1} (\mathbf{x}_k - \mathbf{B}_k \mathbf{u}_k) + \mathbf{S}_{k-1}^{-1} \mathbf{m}_{k-1} \right] \right\}^\top$$

$$\times \Phi_k^{-1} \left\{ \mathbf{x}_{k-1} - \Phi_k \left[ \mathbf{A}_k^\top \mathbf{R}_k^{-1} (\mathbf{x}_k - \mathbf{B}_k \mathbf{u}_k) + \mathbf{S}_{k-1}^{-1} \mathbf{m}_{k-1} \right] \right\}$$

$$\bar{p}(\mathbf{x}_k) = \Theta' \exp[-\mathcal{L}_k(\mathbf{x}_k)], \;\; \Theta' = \Theta[(2\pi)^d \det \vec{\Phi}_k]^{1/2}$$

# Sketch of KF equations: continued

$$\mathcal{L}_k(\mathbf{x}_k) = \mathcal{L}_k(\mathbf{x}_{k-1}, \mathbf{x}_k) - \mathcal{U}_k(\mathbf{x}_{k-1}, \mathbf{x}_k)$$

$$= \frac{1}{2}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k)^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \frac{1}{2}\mathbf{m}_{k-1}^\top \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}$$

$$\vec{\Phi}_k^\top = \vec{\Phi}_k = \left(\mathbf{A}_k^\top \mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1}$$

$$- \frac{1}{2}\left[\mathbf{A}_k^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}\right]^\top \overbrace{\vec{\Phi}_k^\top \vec{\Phi}_k^{-1}}^{\vec{I}} \vec{\Phi}_k \left[\mathbf{A}_k^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}\right]$$

$$= \frac{1}{2}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k)^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \frac{1}{2}\mathbf{m}_{k-1}^\top \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}$$

$$- \frac{1}{2}\left[\mathbf{A}_k^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}\right]^\top \vec{\Phi}_k^\top \left[\mathbf{A}_k^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}\right]$$

$$= \frac{1}{2}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k)^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \frac{1}{2}\mathbf{m}_{k-1}^\top \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}$$

$$- \frac{1}{2}\left[\mathbf{A}_k^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}\right]^\top$$

$$\times \left(\mathbf{A}_k^\top \mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1} \left[\mathbf{A}_k^\top \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}\right]$$

next we need to calculate the first-order and second-order derivatives

# Sketch of KF equations: continued

$$\frac{\partial \mathcal{L}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k} = \mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) - \mathbf{R}_k^{-1}\mathbf{A}_k\left(\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1}\left[\mathbf{A}_k^\top\mathbf{R}_k^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) + \mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}\right]$$

$$= \left[\mathbf{R}_k^{-1} - \mathbf{R}_k^{-1}\mathbf{A}_k\left(\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1}\mathbf{A}_k^\top\mathbf{R}_k^{-1}\right](\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k)$$

$$- \mathbf{R}_k^{-1}\mathbf{A}_k\left(\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1}\mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1} \qquad \left(\mathbf{R} + \mathbf{P}\mathbf{Q}\mathbf{P}^\top\right)^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{P}\left(\mathbf{Q}^{-1} + \mathbf{P}^\top\mathbf{R}^{-1}\mathbf{P}\right)^{-1}\mathbf{P}^\top\mathbf{R}^{-1}$$

$$= \left(\mathbf{R}_k + \mathbf{A}_k\mathbf{S}_{k-1}\mathbf{A}_k^\top\right)^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) - \mathbf{R}_k^{-1}\mathbf{A}_k\left(\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1}\mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}$$

$$\frac{\partial \mathcal{L}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k} = 0 \rightarrow \boxed{\left(\mathbf{R}_k + \mathbf{A}_k\mathbf{S}_{k-1}\mathbf{A}_k^\top\right)^{-1}(\mathbf{x}_k - \mathbf{B}_k\mathbf{u}_k) = \mathbf{R}_k^{-1}\mathbf{A}_k\left(\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1}\mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1} = 0}$$

$$\mathbf{x}_k = \mathbf{B}_k\mathbf{u}_k + \left(\mathbf{R}_k + \mathbf{A}_k\mathbf{S}_{k-1}\mathbf{A}_k^\top\right)\left[\mathbf{R}_k^{-1}\mathbf{A}_k\left(\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1}\mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}\right]$$

$$= \mathbf{B}_k\mathbf{u}_k + \left(\mathbf{A}_k + \mathbf{A}_k\mathbf{S}_{k-1}\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k\right)\left(\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k + \mathbf{S}_{k-1}^{-1}\right)^{-1}\mathbf{S}_{k-1}^{-1}\mathbf{m}_{k-1}$$

$$= \mathbf{B}_k\mathbf{u}_k + \mathbf{A}_k\overbrace{\left(\mathbf{I} + \mathbf{S}_{k-1}\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k\right)\left(\mathbf{I} + \mathbf{S}_{k-1}\mathbf{A}_k^\top\mathbf{R}_k^{-1}\mathbf{A}_k\right)^{-1}}^{\vec{\mathbf{I}}}\mathbf{m}_{k-1}$$

prediction step: $\overline{\mathbf{m}}_k = \mathbf{B}_k\mathbf{u}_k + \mathbf{A}_k\mathbf{m}_{k-1}$

mean of $\overline{p}(\mathbf{x}_k)$

# Sketch of KF equations: continued

$$\frac{\partial^2 \mathcal{L}_k(\mathbf{x}_k)}{\partial \mathbf{x}_k^2} = \left(\mathbf{R}_k + \mathbf{A}_k \mathbf{S}_{k-1} \mathbf{A}_k^\top\right)^{-1} = \overline{\mathbf{S}}_k^{-1} \rightarrow \boxed{\bar{p}(\mathbf{x}_k) \sim \mathcal{N}(\overline{\mathbf{m}}_k, \overline{\mathbf{S}}_k^{-1})}$$

measurement update

$$p(\mathbf{x}_k) \sim p(\mathbf{y}_k|\mathbf{x}_k)\bar{p}(\mathbf{x}_k), \;\; p(\mathbf{y}_k|\mathbf{x}_k) \sim \mathcal{N}(\mathbf{y}_k|\mathbf{C}_k\mathbf{x}_k, \mathbf{Q}_k), \;\; \bar{p}(\mathbf{x}_k) \sim \mathcal{N}(\mathbf{x}_k|\overline{\mathbf{m}}_k, \overline{\mathbf{S}}_k)$$

$$\boxed{\text{loss function: } J_k(\mathbf{x}_k) = \frac{1}{2}\left(\mathbf{y}_k - \mathbf{C}_k\mathbf{x}_k\right)^\top \mathbf{Q}_k^{-1}\left(\mathbf{y}_k - \mathbf{C}_k\mathbf{x}_k\right) + \frac{1}{2}\left(\mathbf{x}_k - \overline{\mathbf{m}}_k\right)^\top \overline{\mathbf{S}}_k^{-1}\left(\mathbf{x}_k - \overline{\mathbf{m}}_k\right)}$$

$$\frac{\partial J_k(\mathbf{x}_k)}{\partial \mathbf{x}_k} = -\,\mathbf{C}_k^\top \mathbf{Q}_k^{-1}\left(\mathbf{y}_k - \mathbf{C}_k\mathbf{x}_k\right) + \overline{\mathbf{S}}_k^{-1}\left(\mathbf{x}_k - \overline{\mathbf{m}}_k\right), \quad \frac{\partial^2 J_k(\mathbf{x}_k)}{\partial \mathbf{x}_k^2} = \mathbf{C}_k^\top \mathbf{Q}_k^{-1}\mathbf{C}_k + \overline{\mathbf{S}}_k^{-1} \equiv \mathbf{S}_k^{-1}$$

measurement step: $\mathbf{m}_k = \overline{\mathbf{m}}_k + \mathbf{S}_k\mathbf{C}_k^\top \mathbf{Q}_k^{-1}\left(\mathbf{y}_k - \mathbf{C}_k\overline{\mathbf{m}}_k\right)$

# Sketch of KF equations: continued

Kalman gain:

$$\begin{aligned}
\mathbf{K}_k =& \mathbf{S}_k \mathbf{C}_k^\top \mathbf{Q}_k^{-1} = \mathbf{S}_k \mathbf{C}_k^\top \mathbf{Q}_k^{-1} \left(\mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{Q}_k\right)\left(\mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{Q}_k\right)^{-1} \\
=& \mathbf{S}_k \left(\mathbf{C}_k^\top \mathbf{Q}_k^{-1} \mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{C}_k^\top \mathbf{Q}_k^{-1} \mathbf{Q}_k\right)\left(\mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{Q}_k\right)^{-1} \\
=& \mathbf{S}_k \left(\mathbf{C}_k^\top \mathbf{Q}_k^{-1} \mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \overline{\mathbf{S}}_k^{-1} \overline{\mathbf{S}}_k \mathbf{C}_k^\top\right)\left(\mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{Q}_k\right)^{-1} \\
=& \mathbf{S}_k \left(\mathbf{C}_k^\top \mathbf{Q}_k^{-1} \mathbf{C}_k + \overline{\mathbf{S}}_k^{-1}\right) \overline{\mathbf{S}}_k \mathbf{C}_k^\top \left(\mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{Q}_k\right)^{-1} = \overline{\mathbf{S}}_k \mathbf{C}_k^\top \left(\mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top + \mathbf{Q}_k\right)^{-1}
\end{aligned}$$

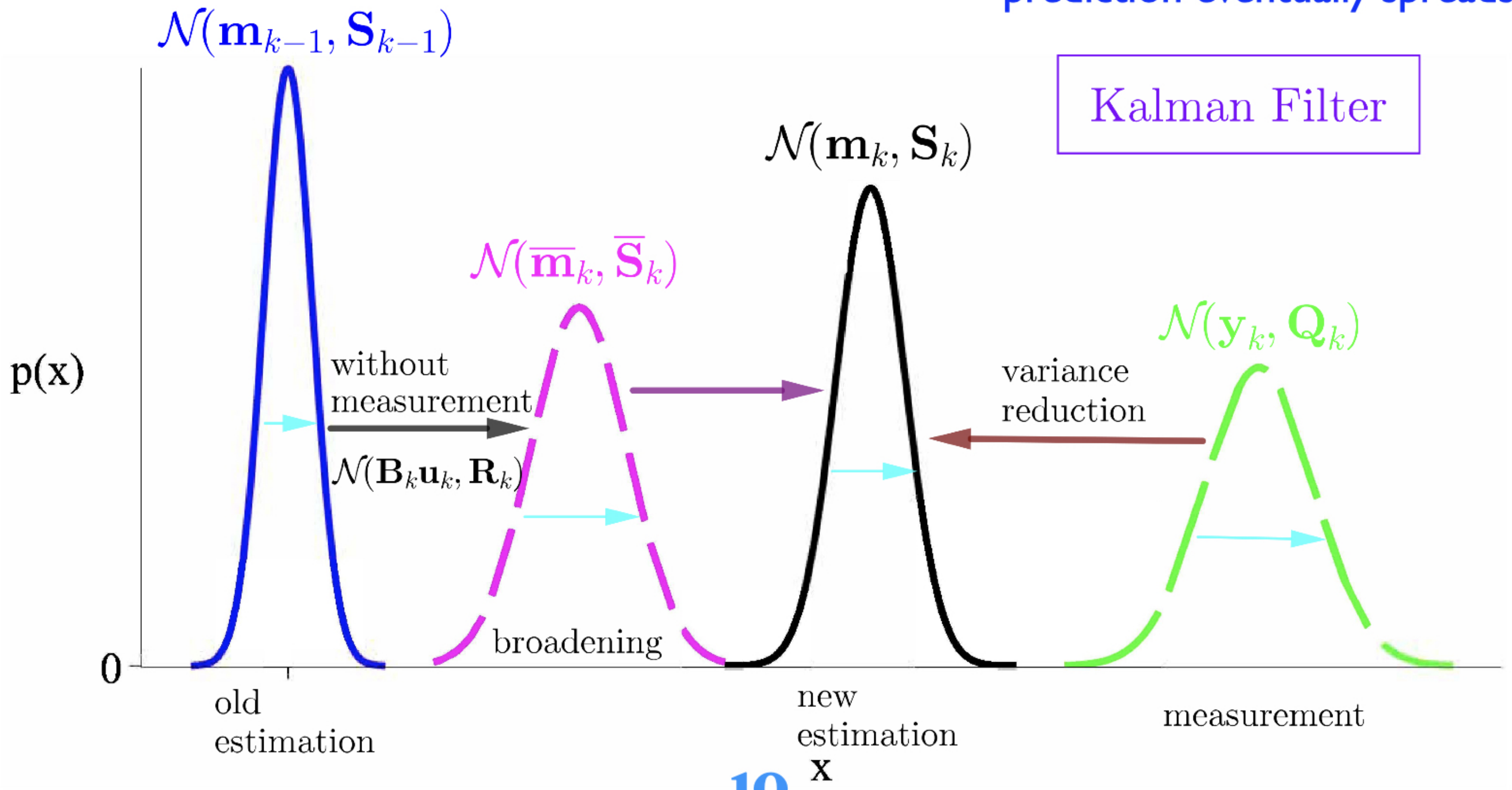
covariance update:

$$(\mathbf{D} + \mathbf{CAB})^{-1} = \mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{C}\left(\mathbf{A}^{-1} + \mathbf{BD}^{-1}\mathbf{C}\right)^{-1}\mathbf{BD}^{-1}$$

$$\begin{aligned}
\mathbf{S}_k =& \left(\mathbf{C}_k^\top \mathbf{Q}_k^{-1} \mathbf{C}_k + \overline{\mathbf{S}}_k^{-1}\right)^{-1} = \overline{\mathbf{S}}_k - \overline{\mathbf{S}}_k \mathbf{C}_k^\top \left(\mathbf{Q}_k + \mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top\right)^{-1} \mathbf{C}_k \overline{\mathbf{S}}_k \\
=& \left[\mathbf{I} - \overline{\mathbf{S}}_k \mathbf{C}_k^\top \left(\mathbf{Q}_k + \mathbf{C}_k \overline{\mathbf{S}}_k \mathbf{C}_k^\top\right)^{-1} \mathbf{C}_k\right] \overline{\mathbf{S}}_k = \left(\mathbf{I} - \mathbf{K}_k \mathbf{C}_k\right) \overline{\mathbf{S}}_k
\end{aligned}$$

# KF: prediction and measurement

measurements effectively reduce the variance, while the last prediction eventually spreads out



# Example 1: free falling (1D)

$Q = 10\,\mathrm{m}^2$, $dt = 0.001\,\mathrm{s}$, $g = 9.8\,\mathrm{m/s}^2$

$\mathbf{x}_0 = \begin{pmatrix} 30\,\mathrm{m} \\ 0\,\mathrm{m/s} \end{pmatrix}$ (GT), $\mathbf{x}_0 = \begin{pmatrix} 45\,\mathrm{m} \\ 0\,\mathrm{m/s} \end{pmatrix}$ (KF), $\mathbf{S}_0 = \begin{pmatrix} 10\,\mathrm{m}^2 & \\ & 5\,\mathrm{m}^2/\mathrm{s}^2 \end{pmatrix}$

positive direction

$$\mathbf{x}_k = \begin{pmatrix} h_k \\ \dot{h}_k \end{pmatrix}$$

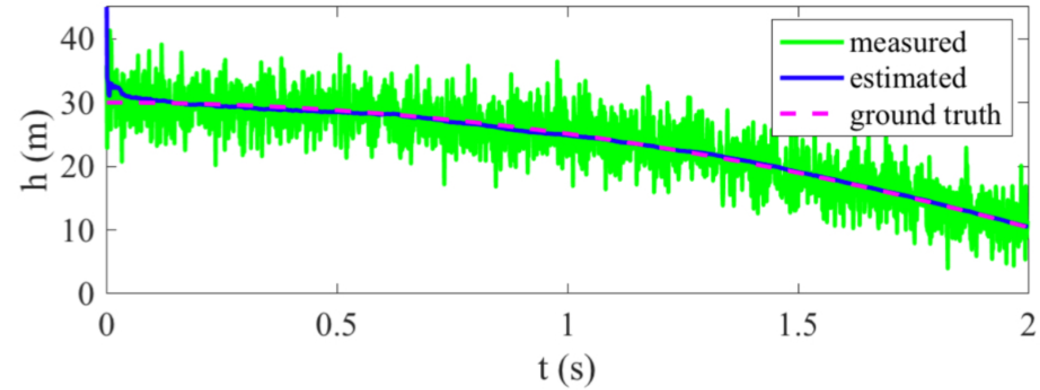
$$h_k = h_{k-1} + \dot{h}_{k-1} dt - \frac{1}{2} g dt^2$$

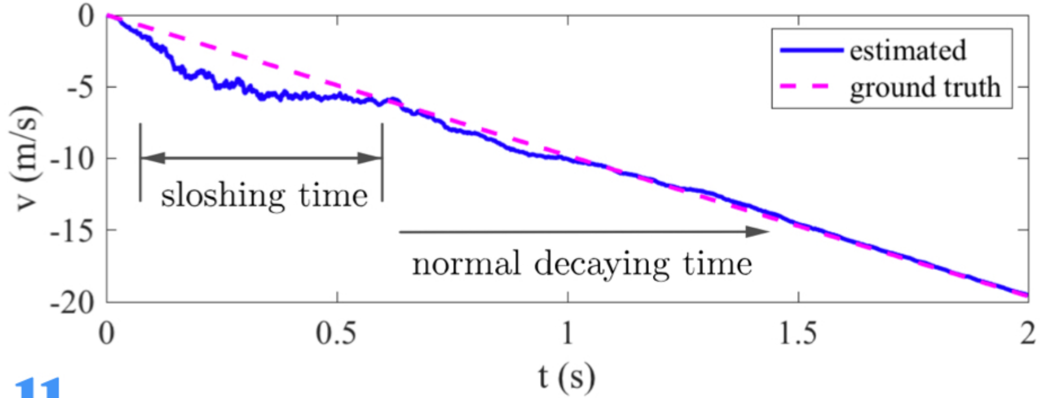
$$\dot{h}_k = \dot{h}_{k-1} - g dt$$

$$\mathbf{x}_k = \underbrace{\begin{pmatrix} 1 & dt \\ 0 & 1 \end{pmatrix}}_{\mathbf{A}} \mathbf{x}_{k-1} + \underbrace{\begin{pmatrix} -dt^2/2 \\ -dt \end{pmatrix}}_{\mathbf{B}} g$$

$h_k$: measurement

$y_k = \mathbf{C}\mathbf{x}_k + n_k$, $\mathbf{C} = (1, 0) \in \mathrm{R}^{1\times 2}$



# Example 2: 2D random motion

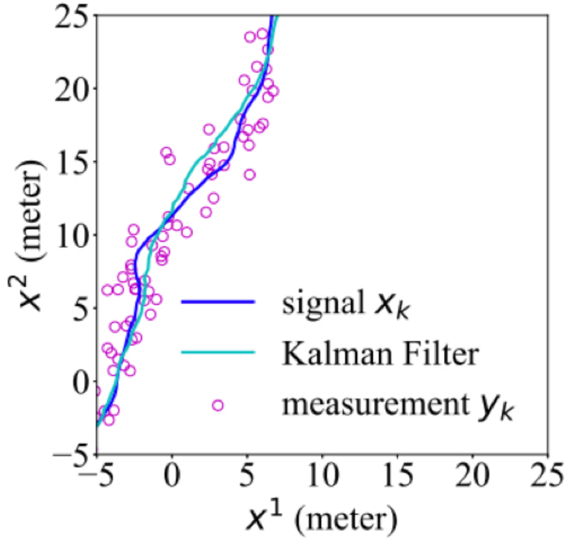
$$\frac{d\mathbf{x}_k}{dt} = \mathbf{D}\mathbf{x}_k + \mathbf{H}\mathbf{w}_k,\ \mathbf{D} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},\ \mathbf{H} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},\ \mathbf{x}_k = \begin{pmatrix} x_k^1 \\ x_k^2 \\ x_k^3 \\ x_k^4 \end{pmatrix},\ \mathbf{w}_k = \begin{pmatrix} w_k^1 \\ w_k^2 \end{pmatrix}$$

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \delta t \mathbf{H}\mathbf{w}_k,\ \mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{n}_k,\ \mathbf{A} = \delta t \mathbf{D} + \mathbf{I}$$

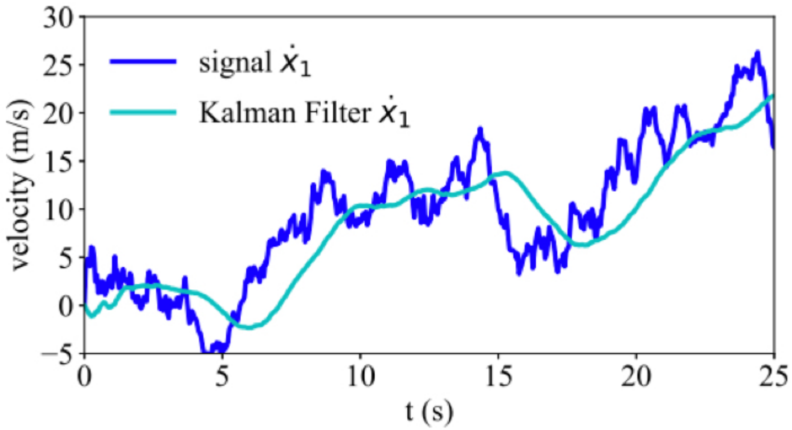
$$\overline{\mathbf{m}}_k = \mathbf{A}\mathbf{m}_{k-1},\ \overline{\mathbf{S}}_k = \mathbf{A}\mathbf{S}_{k-1}\mathbf{A}^\top + \delta t^2 \mathbf{R},\ \mathbf{K}_k = \overline{\mathbf{S}}_k \mathbf{C}^\top \left(\mathbf{C}\overline{\mathbf{S}}_k\mathbf{C}^\top + \mathbf{Q}\right)^{-1}$$

$$\mathbf{m}_k = \overline{\mathbf{m}}_k + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{C}\overline{\mathbf{m}}_k\right),\ \mathbf{S}_k = \left(\mathbf{I} - \mathbf{K}_k\mathbf{C}\right)\overline{\mathbf{S}}_k$$

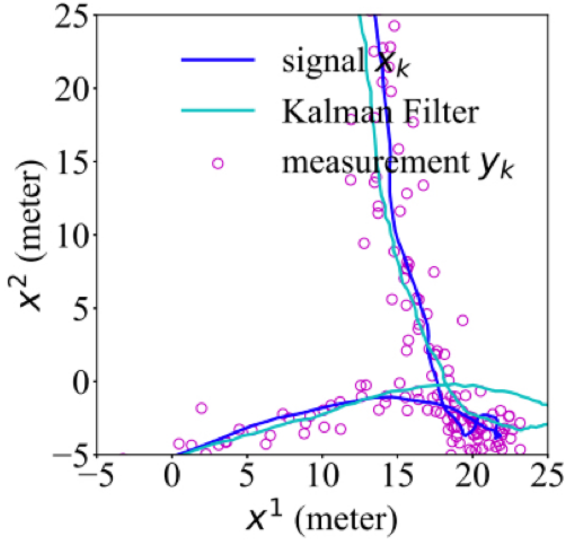
$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},\ \mathbf{R} = \mathrm{E}\left[(\mathbf{H}\mathbf{w}_k)(\mathbf{H}\mathbf{w}_k)^\top\right],\ \mathbf{Q} = \mathrm{E}\left[\mathbf{n}_k\mathbf{n}_k^\top\right]$$




$$d^2x^1/dt^2 = w^1(t)$$
$$d^2x^2/dt^2 = w^2(t)$$
$$x^3 = dx^1/dt$$
$$x^4 = dx^2/dt$$




**KF prediction is delayed**



# Error propagation

$$\mathrm{E}\left[\bar{\mathbf{e}}_k\bar{\mathbf{e}}_k^\top\right] = \bar{\mathbf{S}}_k,\ \mathrm{E}\left[\mathbf{e}_k\mathbf{e}_k^\top\right] = \mathbf{S}_k$$

two errors:

$$\bar{\mathbf{e}}_k = \bar{\mathbf{m}}_k - \mathbf{x}_k = \mathbf{A}_k\mathbf{e}_{k-1} - \mathbf{w}_k$$

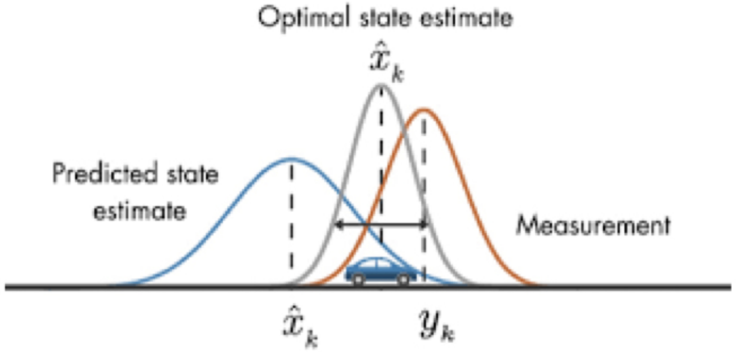
$$\mathbf{e}_k = \mathbf{m}_k - \mathbf{x}_k = (\vec{\mathbf{I}} - \mathbf{K}_k\mathbf{C}_k)\bar{\mathbf{e}}_k + \mathbf{K}_k\mathbf{n}_k$$

$$\boxed{\mathrm{E}[\mathbf{e}_0] = \vec{0} \rightarrow \mathrm{E}[\mathbf{e}_k] = \vec{0}}$$

proof: unbiased estimator

$$\mathrm{E}[\bar{\mathbf{e}}_k] = \mathbf{A}_k\mathrm{E}[\mathbf{e}_{k-1}] - \mathrm{E}[\mathbf{w}_k] = \vec{0}$$

$$\mathrm{E}[\mathbf{e}_k] = (\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\mathrm{E}[\bar{\mathbf{e}}_k] + \mathbf{K}_k\mathrm{E}[\mathbf{n}_k] = \vec{0}$$




proof:

$$\begin{aligned}
\mathrm{E}\left[\bar{\mathbf{e}}_k\bar{\mathbf{e}}_k^\top\right] =& \mathrm{E}\left[(\mathbf{A}_k\mathbf{e}_{k-1} - \mathbf{w}_k)(\mathbf{A}_k\mathbf{e}_{k-1} - \mathbf{w}_k)^\top\right] \\
=& \mathbf{A}_k\mathrm{E}\left[\mathbf{e}_{k-1}\mathbf{e}_{k-1}^\top\right]\mathbf{A}_k^\top - \mathbf{A}_k\mathrm{E}\left[\mathbf{e}_{k-1}\mathbf{w}_k^\top\right] \\
& - \mathrm{E}\left[\mathbf{w}_k\mathbf{e}_{k-1}^\top\right]\mathbf{A}_k^\top + \mathrm{E}\left[\mathbf{w}_k\mathbf{w}_k^\top\right] \\
=& \mathbf{S}_{k-1} + \mathbf{R}_k = \bar{\mathbf{S}}_k
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\left[\mathbf{e}_k\mathbf{e}_k^\top\right] =& \mathrm{E}\left[((\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\bar{\mathbf{e}}_k + \mathbf{K}_k\mathbf{n}_k)((\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\bar{\mathbf{e}}_k + \mathbf{K}_k\mathbf{n}_k)^\top\right] \\
=& (\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\mathrm{E}\left[\bar{\mathbf{e}}_k\bar{\mathbf{e}}_k^\top\right](\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)^\top \\
& + (\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\mathrm{E}\left[\bar{\mathbf{e}}_k\mathbf{n}_k^\top\right]\mathbf{K}_k^\top \\
& + \mathbf{K}_k\mathrm{E}\left[\mathbf{n}_k\bar{\mathbf{e}}_k^\top\right](\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)^\top + \mathbf{K}_k\mathrm{E}\left[\mathbf{n}_k\mathbf{n}_k^\top\right]\mathbf{K}_k^\top \\
=& (\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\bar{\mathbf{S}}_k(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)^\top + \mathbf{K}_k\mathbf{Q}_k\mathbf{K}_k^\top \\
=& \underbrace{(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\bar{\mathbf{S}}_k}_{\mathbf{S}_k} - \underbrace{(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k)\bar{\mathbf{S}}_k}_{\mathbf{S}_k}\mathbf{C}_k^\top\mathbf{K}_k^\top + \mathbf{K}_k\mathbf{Q}_k\mathbf{K}_k^\top = \mathbf{S}_k
\end{aligned}$$

# Kalman gain revisited: physical meaning encoded

magnitude of the covariance of the error $\mathbf{e}_k$

$\mathbf{K}_k$ is unknown here!

$$J(\mathbf{K}_k) = \frac{1}{2}\mathrm{tr}\left(\mathrm{E}\left[\mathbf{e}_k\mathbf{e}_k^\top\right]\right), \; \mathrm{E}\left[\mathbf{e}_k\mathbf{e}_k^\top\right] = (\mathsf{I} - \mathbf{K}_k\mathbf{C}_k)\overline{\mathbf{S}}_k(\mathsf{I} - \mathbf{K}_k\mathbf{C}_k)^\top + \mathbf{K}_k\mathbf{Q}_k\mathbf{K}_k^\top$$

error: $\mathbf{e}_k = \mathbf{m}_k - \mathbf{x}_k$

$\mathbf{m}_k = \overline{\mathbf{m}}_k + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}_k\overline{\mathbf{m}}_k)$

$$\vec{0} = \frac{\partial J(\mathbf{K}_k)}{\partial \mathbf{K}_k} = -(\mathsf{I} - \mathbf{K}_k\mathbf{C}_k)\overline{\mathbf{S}}_k\mathbf{C}_k^\top + \mathbf{K}_k\mathbf{Q}_k$$

$$\mathbf{K}_k = \overline{\mathbf{S}}_k\mathbf{C}_k^\top(\mathbf{Q}_k + \mathbf{C}_k\overline{\mathbf{S}}_k\mathbf{C}_k^\top)^{-1}$$

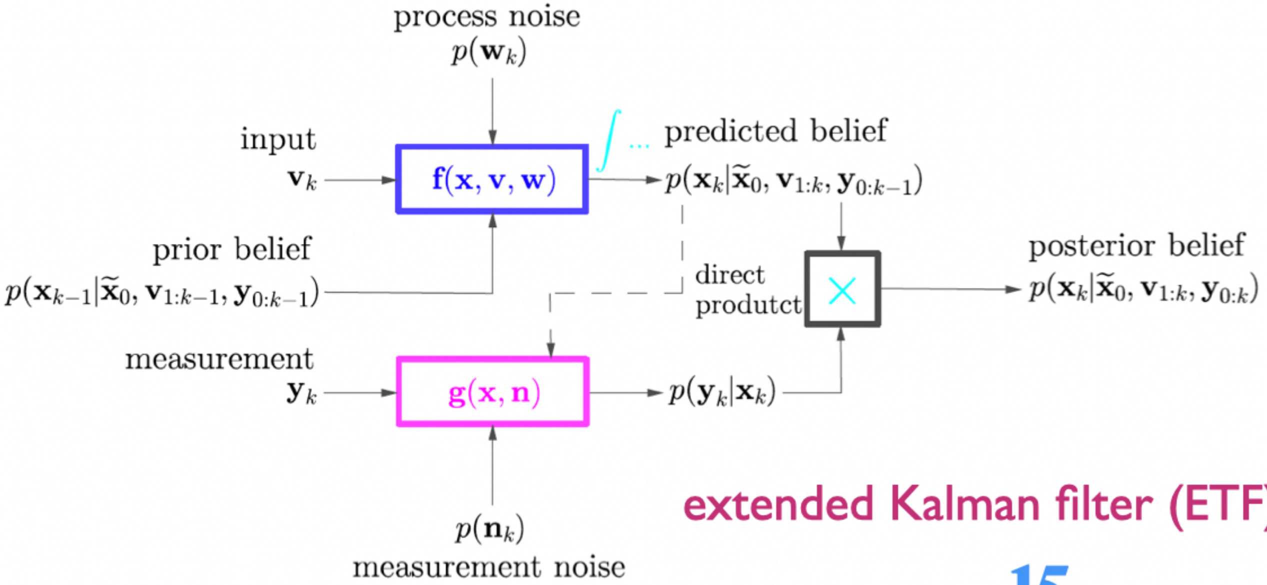
Kalman gain is the optimal "trust coefficient" that determines how much new data should correct your current belief, based on their relative uncertainties

Ex.: discuss when the measurement is very noisy or very accurate

# *Nonleianr extension (linearization)

motion model: $\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{v}_k, \mathbf{w}_k)$; observation model: $\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k, \mathbf{n}_k)$, $k = 0 \sim K$

$$\underbrace{\text{posterior belief}}_{p(\mathbf{x}_k|\tilde{\mathbf{x}}_0,\mathbf{v}_{1:k},\mathbf{y}_{0:k})} = \Theta \times \overbrace{\text{observation model g}}^{p(\mathbf{y}_k|\mathbf{x}_k)} \times \underbrace{\int \overbrace{\text{motion prediction f}}^{p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{v}_k)} \times \overbrace{\text{prior belief}}^{p(\mathbf{x}_{k-1}|\tilde{\mathbf{x}}_0,\mathbf{v}_{1:k-1},\mathbf{y}_{0:k-1})} \times d\mathbf{x}_{k-1}}_{p(\mathbf{x}_k|\tilde{\mathbf{x}}_0,\mathbf{v}_{1:k},\mathbf{y}_{0:k-1})}$$




extended Kalman filter (ETF)

$$\mathbf{f}(\mathbf{x}_{k-1}, \mathbf{v}_k, \mathbf{w}_k) \approx \tilde{\mathbf{x}}_k + \mathbf{F}_{k-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}) + \delta\mathbf{w}_k$$

$$\mathbf{g}(\mathbf{x}_k, \mathbf{n}_k) \approx \tilde{\mathbf{y}}_k + \mathbf{G}_k(\mathbf{x}_k - \tilde{\mathbf{x}}_k) + \delta\mathbf{n}_k$$

$$\tilde{\mathbf{x}}_k = \mathbf{f}(\hat{\mathbf{x}}_{k-1}, \mathbf{v}_k, \vec{0}), \ \tilde{\mathbf{y}}_k = \mathbf{g}(\tilde{\mathbf{x}}_k, \vec{0})$$

$$\mathbf{w}_k \sim \mathcal{N}(\vec{0}, \mathbf{R}_k), \ \mathbf{n}_k \sim \mathcal{N}(\vec{0}, \mathbf{Q}_k)$$

$\tilde{}$ : prior; $\hat{}$ : posterior

# Example using ETF: simple pendulum

$$\text{acc} = d^2(l\chi)/dt^2 = -g\sin\chi + w(t)$$

nosie

**prediction:**

$$\widetilde{\mathbf{x}}_k = \mathbf{f}(\widehat{\mathbf{x}}_{k-1}, \mathbf{v}_k, \vec{0})$$

$$\widetilde{\mathbf{S}}_k = \mathbf{F}_{k-1}\widehat{\mathbf{S}}_{k-1}\mathbf{F}_{k-1}^\top + \mathbf{R}_k$$

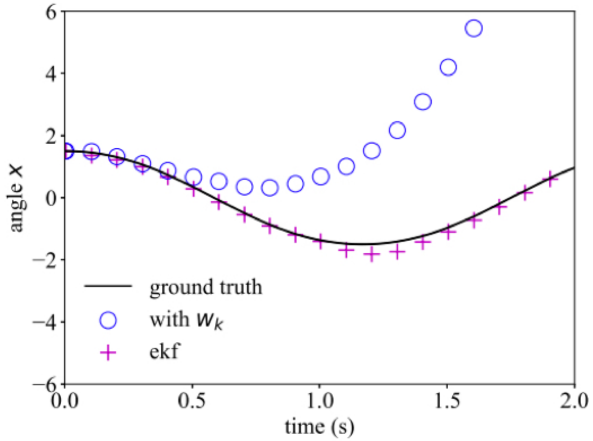
**Kalman gain:**

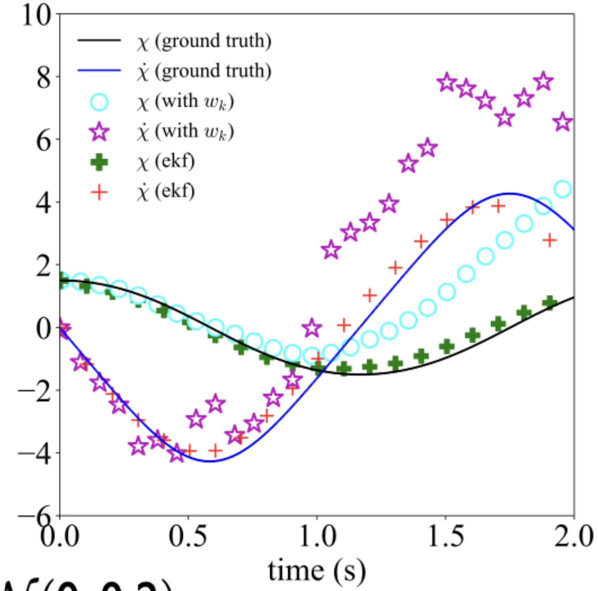
$$\mathbf{K}_k = \widetilde{\mathbf{S}}_k\mathbf{G}_k^\top\left(\mathbf{G}_k\widetilde{\mathbf{S}}_k\mathbf{G}_k^\top + \mathbf{Q}_k\right)^{-1}$$

**correction:**

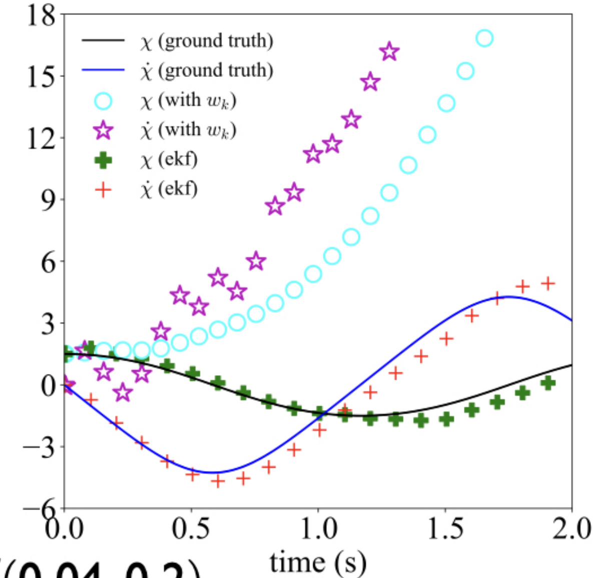
$$\mathbf{x}_k = \widetilde{\mathbf{x}}_k + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{g}(\widetilde{\mathbf{x}}_k, \vec{0})\right)$$

$$\widehat{\mathbf{S}}_k = (\vec{I} - \mathbf{K}_k\mathbf{G}_k)\widetilde{\mathbf{S}}_k$$




$w_k \sim \mathcal{N}(0, 0.2)$




$w_k \sim \mathcal{N}(0.04, 0.2)$

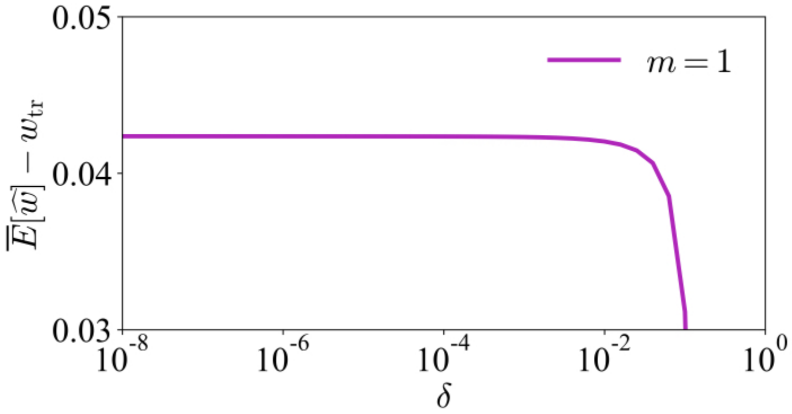
# Bias: nature of estimate problem (example)

problem: estimate $w$ of $f_w(x) = e^{-wx}$

$$x^* = x^{(1)} = \cdots = x^{(m)};\ y^{(i)} = f_w(x^*) + \epsilon^{(i)},\ \epsilon^{(i)} \sim \mathcal{N}(0, \sigma^2)$$

$$\bar{y} = \frac{1}{m}\sum_{i=1}^{m} y^{(i)} \sim \mathcal{N}(0, \sigma^2/m) \to \mathrm{E}[\bar{y}] = e^{-wx^*}$$

$w = w_{\mathrm{tr}} = 1$




**ML solution:** $\widehat{w} = -\frac{1}{x^*}\ln\bar{y}$

Jensen's inequality:

$$\mathrm{E}[\widehat{w}] = -\frac{1}{x^*}\mathrm{E}[\ln\bar{y}] \geq -\frac{1}{x^*}\ln\mathrm{E}[\bar{y}] = w$$

pseudo-expectation:

$$\overline{\mathrm{E}}[\widehat{w}] = \left(-\frac{1}{x^*}\sqrt{\frac{m}{2\pi\sigma^2}}\right)\int_{\delta}^{2e^{-wx^*}-\delta} d\bar{y}\,\ln\bar{y}\,\exp\left[-\frac{m}{2\sigma^2}\left(\bar{y} - e^{-wx^*}\right)^2\right]$$

$$(\bar{y} = e^{-wx^*} \pm \sigma\sqrt{x/m})$$

$$= w - \frac{1}{2x^*\sqrt{2\pi}}\int_0^{m[\exp(-wx^*)-\delta]^2/\sigma^2} \ln\left(1 - \frac{\sigma^2 x e^{2wx^*}}{m}\right) e^{-x/2}x^{-1/2}dx$$