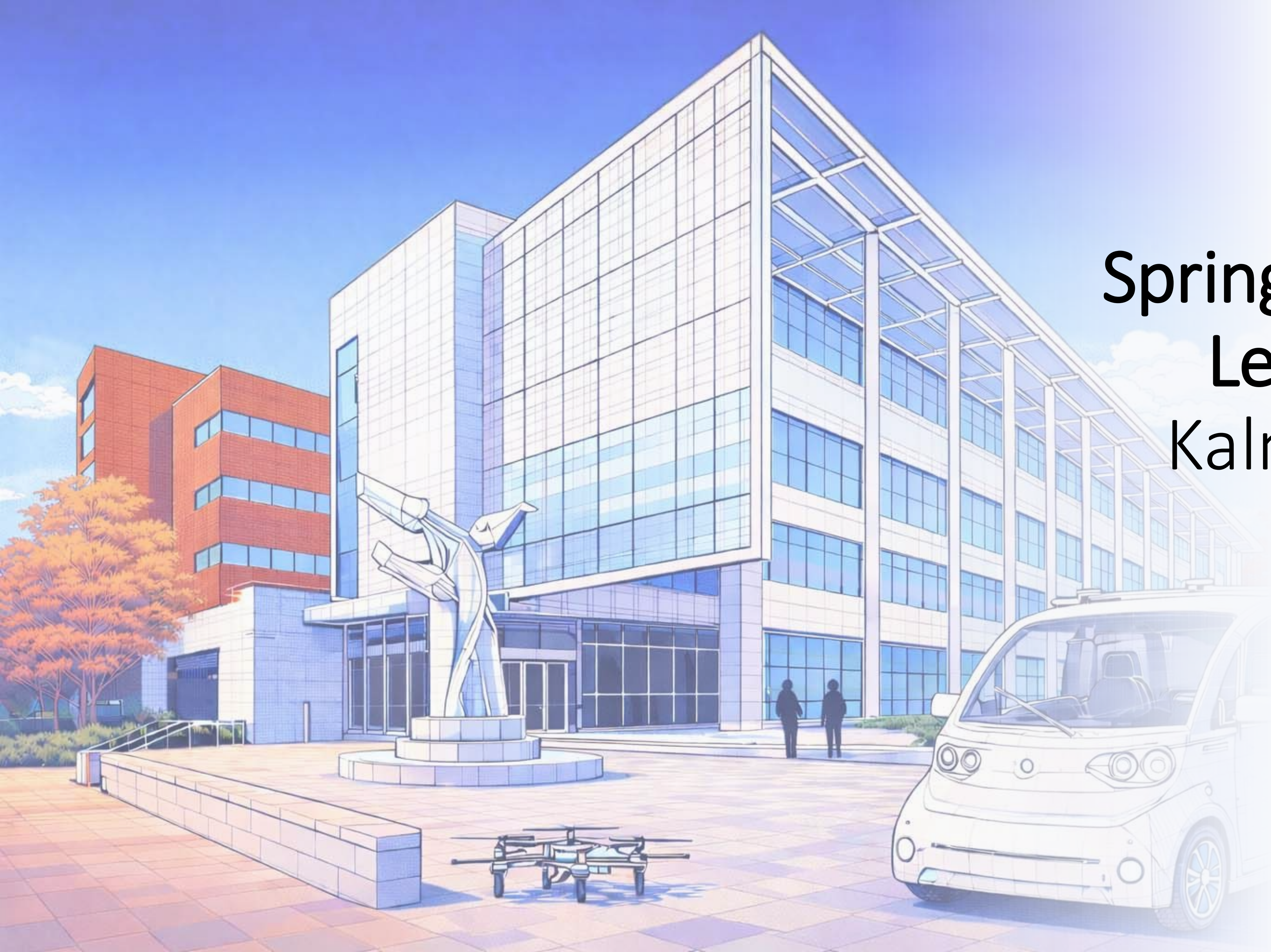


Spring 26 ECE484
Lecture 13
Kalman Filter

Sayan Mitra

The Bayes Filter Algorithm

BAYES_FILTER($bel(x_{t-1}), u_t, z_t$ **)**

PREDICT

for all x_t do:

$$\overline{bel}(x_t) = \sum_{x \in Q} p_D(x_t | x_{t-1}, u_t)\, bel(x_{t-1})$$

UPDATE

$$bel(x_t) = \eta \cdot p_M(z_t \mid x_t)\ \overline{bel}(x_t)$$

endfor

return $bel(x_t)$

Key Properties

- Recursive filter for computing bel(x_{t-1})
- Only needs bel(x_{t-1}), not full history
- General: works for any state-space model
- Problem: the sum/integral is often intractable in closed form!

Bayes Filter: Continuous Distributions

Bayes_filter($bel(x_{t-1}), u_t, z_t$**)**

for all x_t do:

$$\overline{bel}(x_t) = \int p(x_t | u_{t,} x_{t-1}) bel(x_{t-1}) dx_{t-1}$$

$$bel(x_t) = \eta \, p(z_t | x_t) \, \overline{bel}(x_t)$$

end for

return $bel(x_t)$

Discrete Kalman Filter

The Kalman filter estimates state of a Discrete Linear System with Gaussian noise

$$x_t = A_t x_{t-1} + B_t u_t + \epsilon_t$$
$$z_t = C_t x_t + \delta_t$$

$x_t \in \mathbb{R}^n$: State vector
$\mathrm{u_t} \in \mathbb{R}^m$: : Input vector
$z_t \in \mathbb{R}^\ell$: : Output vector
$\varepsilon_t \sim N(0, \mathrm{Q_t})$: Process noise with covariance $\mathrm{R_t}$
$\delta_t \sim N(0, R_t)$: Measurement noise with covariance $\mathrm{Q_t}$

Note that we no longer have discrete states or measurements!

The Gaussian (Normal) Distribution

1D Gaussian

$$p(x) = N(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

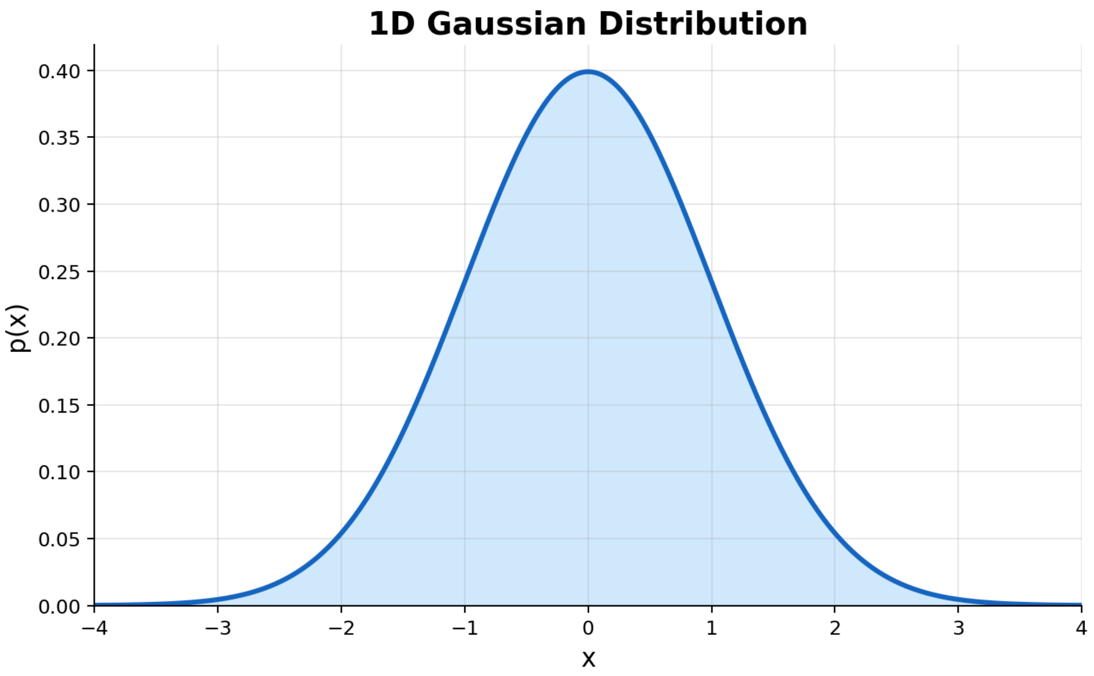
μ = mean (center of distribution)

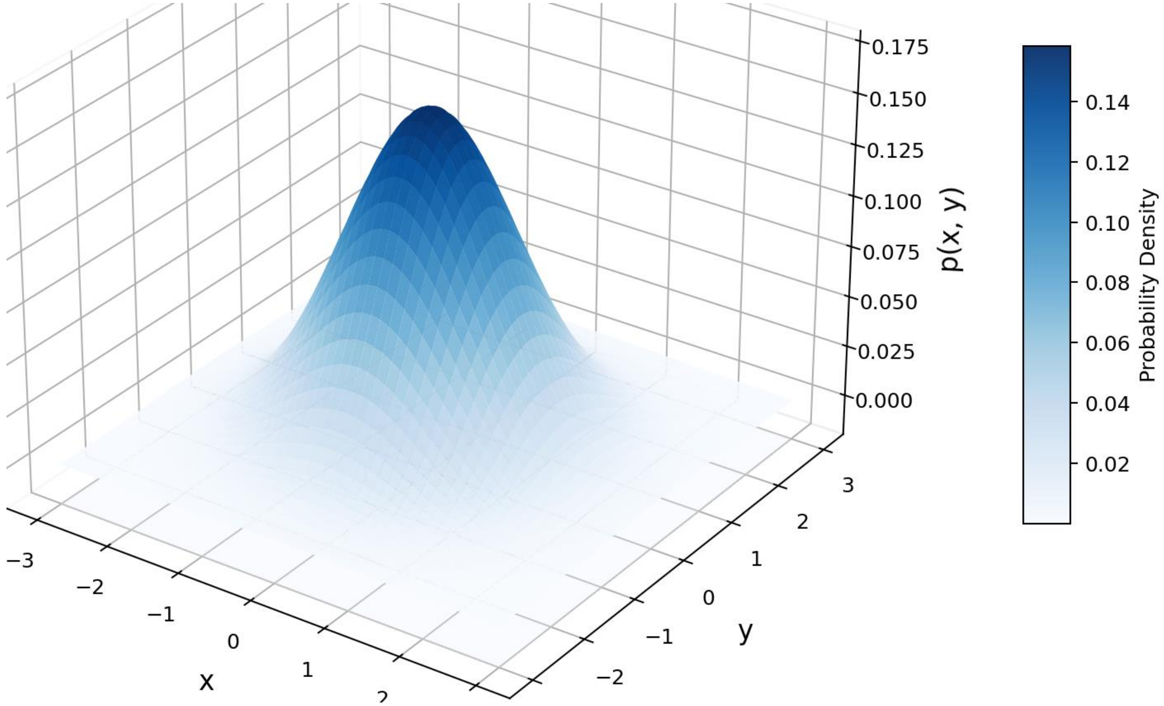
σ^2 = variance

Multivariate nD Gaussian

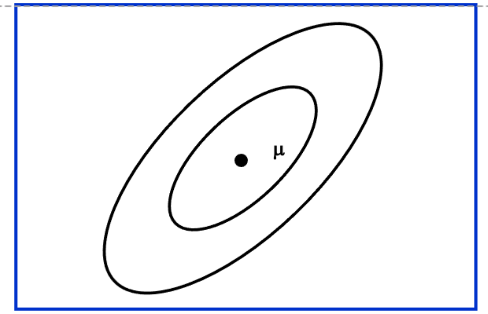
$$p(x) = (2\pi)^{-\frac{n}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp(-\frac{1}{2}\,(x-\mu)^{\mathrm{T}}\,\Sigma^{-1}\,(x-\mu))$$

μ = mean vector ($\mathbb{R}^n$) Σ = covariance matrix ($\mathbb{R}^{nxn}$)





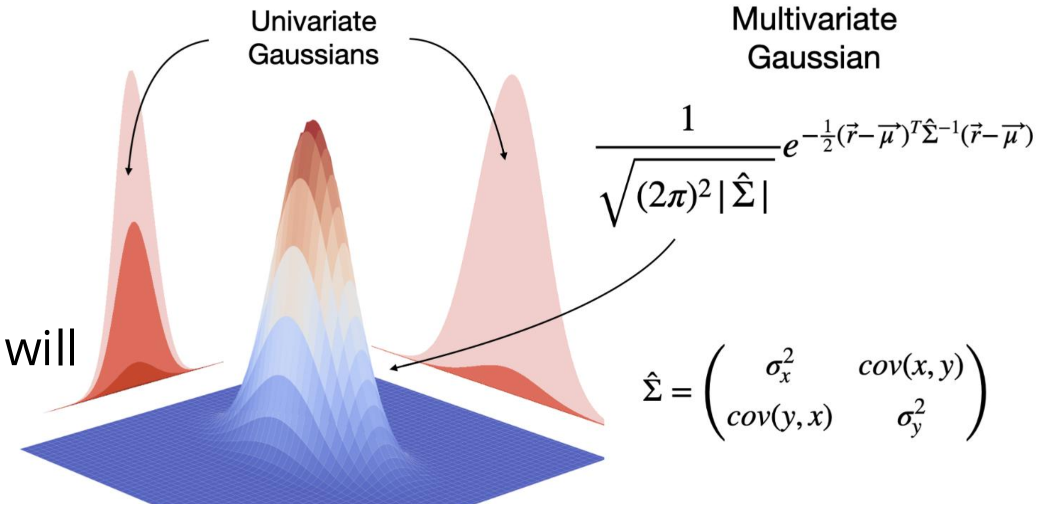
Multivariate Gaussians

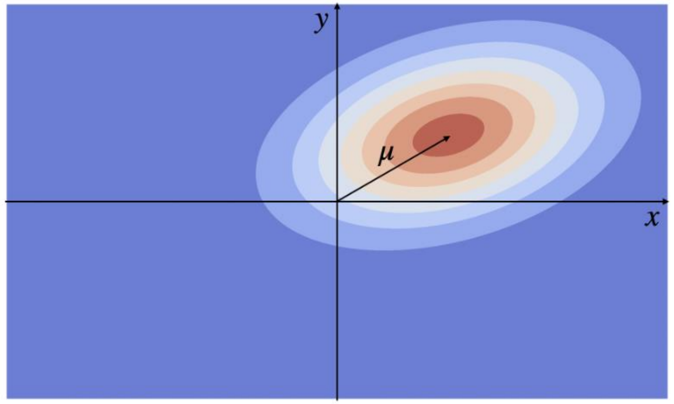


$$p(x) = (2\pi)^{-\frac{n}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp(-\frac{1}{2}\,(x-\mu)^{\mathrm{T}}\,\Sigma^{-1}\,(x-\mu))$$

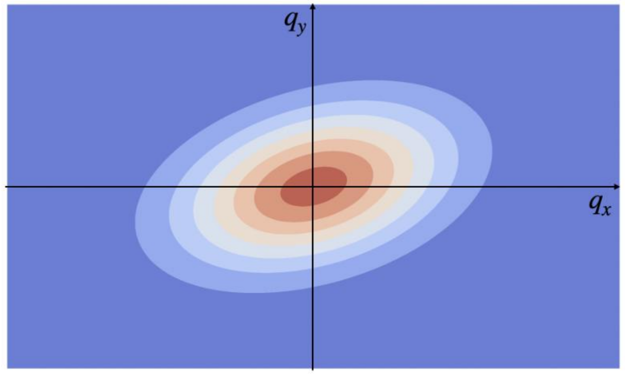
Every single variable x_i in x has a normal distribution $N(\mu_i, \sigma_i)$

If the variables are uncorrelated then the covariance matrix Σ will be a diagonal matrix with the diagonal terms $\{\sigma_i^2\}$

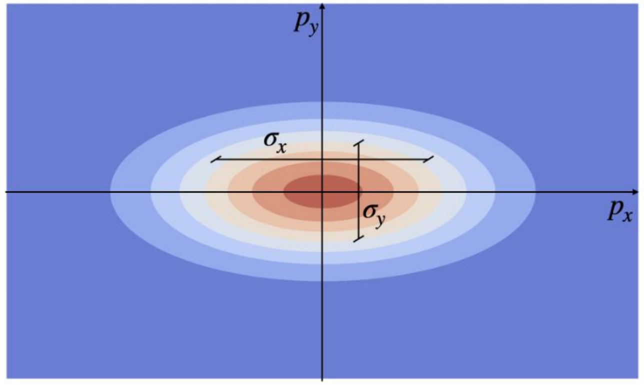




Random bivariate Gaussian distribution



Bivariate Gaussian distribution centered in zero



Bivariate Gaussian distribution diagonalised

Properties of Gaussians

Linear transformations of Gaussians are Gaussians
Gaussian are closed under linear transformations

$$\left.\begin{array}{l} X \sim N(\mu, \sigma^2) \\ Y = aX + b \end{array}\right\} \Rightarrow \quad Y \sim N(a\mu + b, a^2\sigma^2)$$

Products of Gaussian densities is a Gaussian

$$\left.\begin{array}{l} X_1 \sim N(\mu_1, \sigma_1^{\,2}) \\ X_2 \sim N(\mu_2, \sigma_2^{\,2}) \end{array}\right\} \Rightarrow p(X_1) \cdot p(X_2) \sim N\left(\frac{\sigma_2^{\,2}}{\sigma_1^{\,2} + \sigma_2^{\,2}} \mu_1 + \frac{\sigma_1^{\,2}}{\sigma_1^{\,2} + \sigma_2^{\,2}} \mu_2, \quad \frac{1}{\sigma_1^{\,-2} + \sigma_2^{\,-2}} \right)$$

Multivariate Gaussians

$$\left.\begin{array}{r} X \sim N(\mu, \Sigma) \\ Y = AX + B \end{array}\right\} \Rightarrow \quad Y \sim N(A\mu + B, A\Sigma A^T)$$

$$\left.\begin{array}{r} X_1 \sim N(\mu_1, \Sigma_1) \\ X_2 \sim N(\mu_2, \Sigma_2) \end{array}\right\} \Rightarrow p(X_1) \cdot p(X_2) \sim N\left(\frac{\Sigma_2}{\Sigma_1 + \Sigma_2}\mu_1 + \frac{\Sigma_1}{\Sigma_1 + \Sigma_2}\mu_2, \quad \frac{1}{\Sigma_1^{-1} + \Sigma_2^{-1}} \right)$$

We stay in the “Gaussian world” as long as we start with Gaussians and perform only linear transformations.

Kalman Filter

$$x_t = A_t x_{t-1} + B_t u_t + \epsilon_t$$
$$z_t = C_t x_t + \delta_t$$

$x_t \in \mathbb{R}^n$: State vector
$\mathrm{u_t} \in \mathbb{R}^m$: : Input vector
$z_t \in \mathbb{R}^\ell$: : Output vector
$\varepsilon_t \sim N(0, Q_t)$: Process noise with covariance $\mathrm{R_t}$
$\delta_t \sim N(0, R_t)$: Measurement noise with covariance $\mathrm{Q_t}$

$$\mathrm{p_D(x_t | x_{t-1}, u_t)}$$
$$= \mathrm{N(x_t;\ A_t\, x_{t-1} + B_t u_t, Q_t)}$$
$$= \mathrm{(2\pi)^{-\frac{n}{2}} \det(Q_t)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x_t - A_t x_{t-1} - B_t u_t)^T Q_t^{-1} (x_t - A_t x_{t-1} - B_t u_t)\right)}$$

Mean $A_t\, x_{t-1} + B_t\, u_t$

Covariance Q_t

Kalman Filter — Soundness

Theorem (Gaussian Closure) If the initial belief $bel(x_0)$ is Gaussian, the state transition is linear-Gaussian, and the measurement model is linear-Gaussian, then $bel(x_t)$ is Gaussian for all $t \geq 0$. i.e., belief is fully parameterized by μ_t and Σ_t

Proof sketch:

1. Linear transformation and marginalization: If $x \sim N(\mu, \Sigma)$, then $Ax + b \sim N(A\mu + b, A\Sigma A^T)$ → Used prediction

2. Product of Gaussians: If p(x) and p(z|x) are both Gaussian, then p(x|z) is Gaussian → Used correction (Bayes' rule)

By induction: predict and correct keeps distributions Gaussian, so $bel(x_t)$ remains Gaussian for all t.

Kalman Filter Algorithm

Kalman_Filter($\mu_{t-1}, \Sigma_{t-1}, u_t, z_t$):

Prediction

1. $\bar{\mu}_t = A_t \mu_{t-1} + B_t u_t$
2. $\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^\top + Q_t$

Correction:

1. $K_t = \bar{\Sigma}_t C_t^\top (C_t \bar{\Sigma}_t C_t^\top + R_t)^{-1}$
2. $\mu_t = \bar{\mu}_t + K_t (z_t - C_t \bar{\mu}_t)$
3. $\Sigma_t = (I - K_t C_t) \bar{\Sigma}_t$

Return μ_t,Σ_t

Kalman Filter represents the belief $bel(x_t)$ by mean μ_t and covariance Σ_t

Prediction uses the property that Gaussians are closed under linear transformations

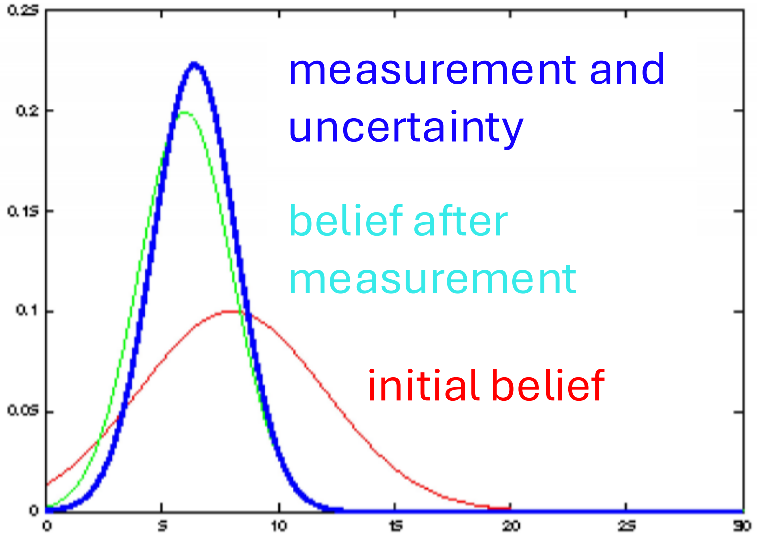
Correction computes the ***Kalman gain*** K_t which is used to weight the impact of new measurements against the predicted value

The key concept here is ***innovation*** = $z_t - C_t \bar{\mu}_t$

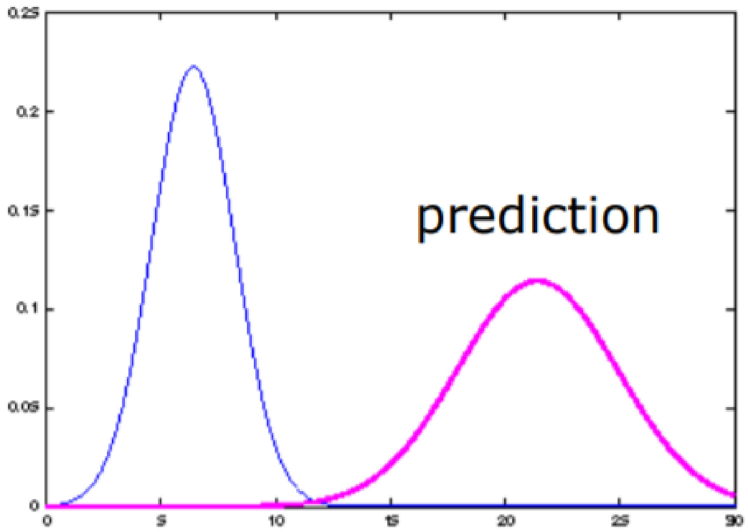
Image Credit: Wikipedia

Rudolf Kalman

- One of the most influential people on control theory and is most known for his co-invention of the Kalman filter in 1950 (or Kalman-Bucy Filter)
- *Today Kalman filters are inside many robots, commercial airplanes, nuclear power plant instrumentation, and demographic models, as well as applications in econometrics*
- Initially met with vast skepticism, so much so that he was forced to do the first publication of his results in mechanical engineering, rather than in electrical engineering
- This worked out fine as some of the first use cases was with NASA on the Apollo spacecraft to and from Moon



Apply control action

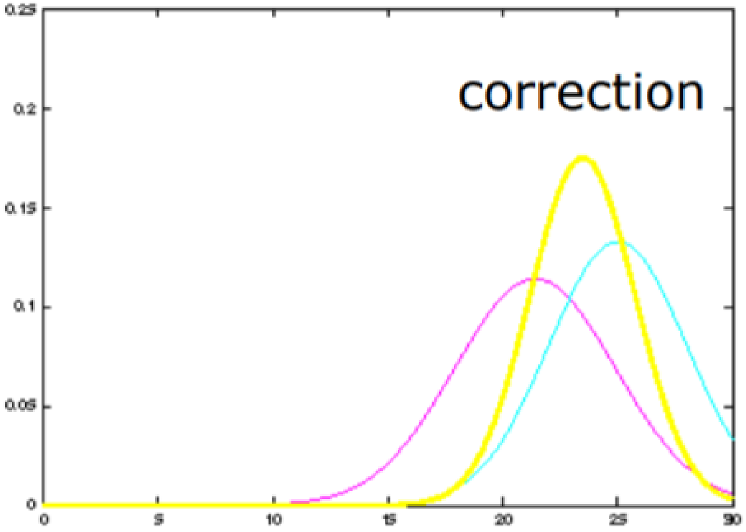


Correction:

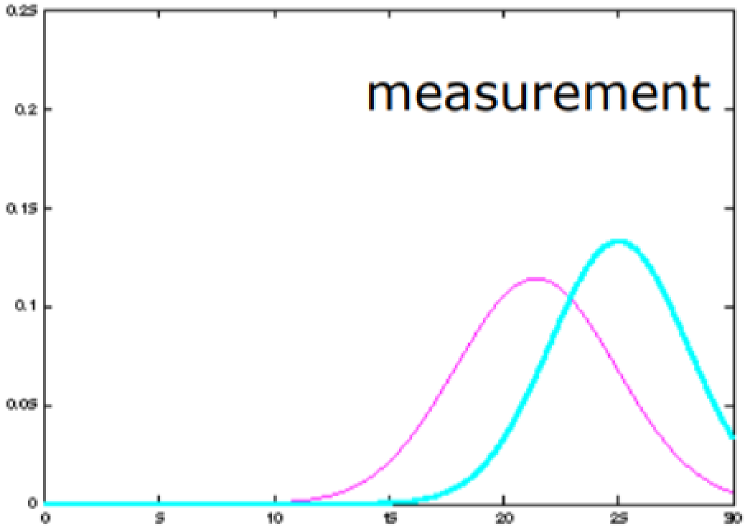
1. $K_t = \bar{\Sigma}_t C_t^\top (C_t \bar{\Sigma}_t C_t^\top + R_t)^{-1}$
2. $\mu_t = \bar{\mu}_t + K_t (z_t - C_t \bar{\mu}_t)$
3. $\Sigma_t = (I - K_t C_t) \bar{\Sigma}_t$

Prediction:

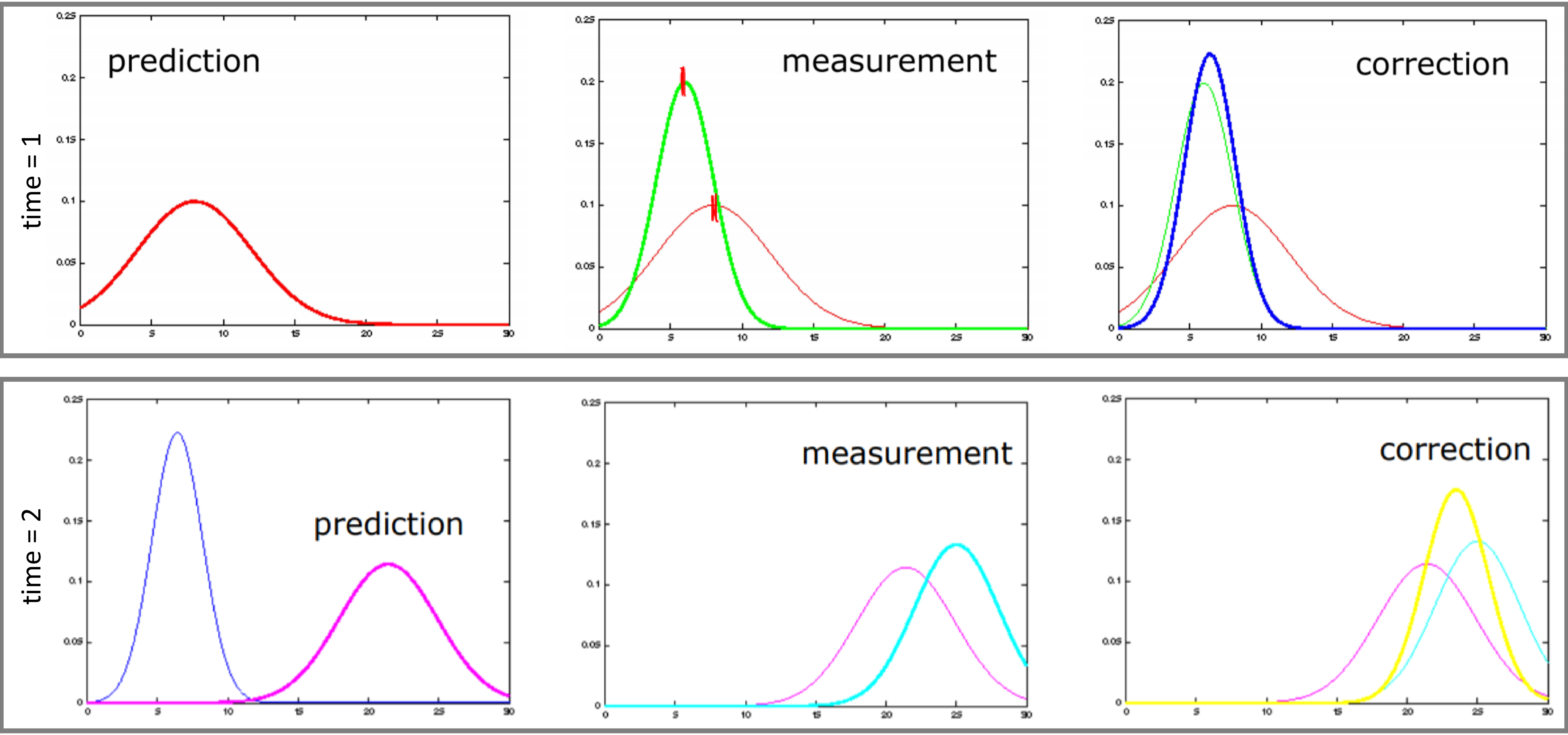
1. $\bar{\mu}_t = A_t \mu_{t-1} + B_t u_t$
2. $\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^\top + Q_t$



Get sensor measurement



Kalman Filter Example



KF Derivation — Prediction Step

Starting point: $\overline{\text{bel}}(x_t) = \int p(x_t \mid x_{t-1}, u_t) \cdot \text{bel}(x_{t-1})\, dx_{t-1}$

Step 1: The prior belief is Gaussian

$\text{bel}(x_{t-1}) = N(x_{t-1}\,;\ \mu_{t-1}, \Sigma_{t-1})$

Step 2: The motion model is linear-Gaussian

$p(x_t \mid x_{t-1}, u_t) = N(x_t\,;\ A_t x_{t-1} + B_t u_t,\ Q_t)$

Step 3: Apply linear transformation property

The convolution of two Gaussians yields a Gaussian

Result: Predicted belief $\overline{\text{bel}}(x_t) = N(x_t\,;\ \bar{\mu}_t,\ \bar{\Sigma}_t)$

Predicted mean: $\bar{\mu}_t = A_t \mu_{t-1} + B_t u_t$

Predicted covariance: $\bar{\Sigma}_t = A_t \Sigma_{t-1} A_t^T + Q_t$

Intuition: Mean propagates through dynamics. Covariance grows by $A\Sigma A^T$ (transform) + R (process noise).

Deriving the Kalman Gain K_t

Goal. Find the gain K_t that minimizes the posterior estimation error — optimally fusing prediction and measurement.

Setup

Predicted state: $\bar{\mu}_t$ with covariance $\bar{\Sigma}_t$

Measurement model: $z_t = C_t \mu_t^- + \delta_t$, $\delta_t \sim \mathcal{N}(0, R_t)$

General Update Rule

$$\mu_t = \bar{\mu}_t + K_t (z_t - C_t \bar{\mu}_t)$$

Key Question: What is the optimal K_t?

1. The posterior error $x_t - \mu_t$

$\bar{e}_t = x_t - \bar{\mu}_t$ (predicted error: true state minus predicted mean)

$e_t = x_t - \mu_t$ (posterior error: true state minus updated estimate)

Start from the update equation

$$\mu_t = \bar{\mu}_t + K_t(z_t - C_t\,\bar{\mu}_t)$$

$$= \bar{\mu}_t + K_t(C_t x_t + \delta_t - C_t\,\bar{\mu}_t)$$

[Substitute measurement model $z_t = C_t x_t + \delta_t$]

$$= \bar{\mu}_t + K_t C_t(x_t - \bar{\mu}_t) + K_t\delta_t$$

$$= \bar{\mu}_t + K_t C_t\,\bar{e}_t + K_t\delta_t$$

Compute posterior error:

$$e_t = x_t - \mu_t = x_t - \bar{\mu}_t - K_t C_t\,\bar{e}_t - K_t\delta_t$$

$$= \bar{e}_t - K_t C_t\,\bar{e}_t - K_t\delta_t$$

$$= (I - K_t C_t)\,\bar{e}_t - K_t\delta_t$$

2. The posterior covariance Σ_t

$$\Sigma_t = E[e_t e_t^\top]$$

$$= E\left[\left((I - K_t C_t)\bar{e}_t - K_t \delta_t\right)\left((I - K_t C_t)\bar{e}_t - K_t \delta_t\right)^\top\right]$$

$= E[(I - K_t C_t)\ \bar{e}_t\ \bar{e}_t^\top (I - K_t C_t)^\top] + E[K_t \delta_t\ \delta_t^\top K_t^\top]$ [Expanding, using linearity of expectation and independence of $\bar{e}_t$ and δ_t with 0 mean]

$$= (I - K_t C_t)\ E[\bar{e}_t\ \bar{e}_t^\top](I - K_t C_t)^\top + K_t E[\delta_t\ \delta_t^\top] K_t^\top$$

$= (I - K_t C_t)\bar{\Sigma}_t (I - K_t C_t)^\top + K_t \mathrm{R}_t K_t^\top$ [$\bar{\Sigma}_t = \bar{e}_t\ \bar{e}_t^\top$ and $\mathrm{R}_t = \delta_t\ \delta_t^\top$]

This is called the **Joseph Form**: Always symmetric and positive definite-> Numerically stable

3. Expanding Σ_t and the Cost Function

Why minimize $tr(\Sigma_t)$?

$tr(\Sigma_t)$ = sum of variances $\Longrightarrow$ scalar measure of total uncertainty

Expand and collect terms

$$\Sigma_t = (I - K_t C_t)\bar{\Sigma}_t (I - K_t C_t)^\top + K_t R_t K_t^\top$$

$$= \bar{\Sigma}_t - K_t C_t \bar{\Sigma}_t - \bar{\Sigma}_t C_t^T K_t^T + K_t (C_t \bar{\Sigma}_t C_t^T + R_t) K_t^T$$

Take the trace

$$tr(\Sigma_t) = tr(\bar{\Sigma}_t) - 2 \cdot tr(K_t C_t \bar{\Sigma}_t) + tr(K_t (C_t \bar{\Sigma}_t C_t^T + R_t) K_t^T)$$

Using trace identities: $tr(A) = tr(A^\top)$ and $tr(AB) = tr(BA)$

4. Minimizing $tr(\Sigma_t)$ w.r.t. K_t

Set the matrix derivative to zero

$$\frac{\partial}{\partial K_t} tr(\Sigma_t) = -2(C_t \bar{\Sigma}_t)^T + 2K_t(C_t \bar{\Sigma}_t C_t^T + R_t) = 0$$

Using: $\frac{\partial}{\partial K_t} tr(KAK^T) = 2KA$ and $\frac{\partial}{\partial K_t} tr(KB) = B^T$

Rearranging $K_t (C_t \bar{\Sigma}_t C_t^T + R_t) = \bar{\Sigma}_t C_t^T$

Solve for K_t: $K_t = \bar{\Sigma}_t C_t^T (C_t \bar{\Sigma}_t C_t^T + R_t)^{-1}$

The Kalman Gain Formula

$K_t \;= \frac{\bar{\Sigma}_t C_t^T}{C_t \bar{\Sigma}_t C_t^T + R_t} \approx$ (prediction uncertainty) / (pred. uncertainty + measurement noise)

$\bar{\Sigma}_t C_t^T$: maps prediction uncertainty into measurement space

$C_t \bar{\Sigma}_t C_t^T + R_t$: total innovation covariance

Case 1: $R_t \to \infty$ (sensor very noisy)

$K_t \to 0 \Longrightarrow \mu_t \to \bar{\mu}_t \;\; and \;\; \Sigma_t \to \bar{\Sigma}_t$ (measurement ignored, prediction unchanged)

Case 2: $R_t \to 0$ (perfect sensor)

$K_t \to C_t^{-1} \quad \Longrightarrow \quad \mu_t \to C_t^{-1} z_t \;\; and \;\; \Sigma_t \to 0$ (pred. discarded, sensor trusted fully)

Takeaway: K_t automatically weights prediction vs. measurement by their covariances — optimal in the MMSE sense.

Kalman Filter — Optimality

Theorem (Gaussian Noise — MVUE)
When the noise is Gaussian, the KF is the Minimum Variance Unbiased Estimator. No estimator (linear or nonlinear) can achieve lower MSE.

Key insight

The Kalman gain K_t is chosen to minimize the posterior error covariance:

$$K_t = \text{argmin } \operatorname{tr}(\Sigma_t) = \text{argmin } E[\|x_t - \mu_t\|^2]$$

This is what makes the KF optimal — it finds the exact gain that minimizes estimation error at every time step.

Summary and Limitations of KF

KF implements Bayesian filter for linear systems with Gaussian noise representing the posterior by a mean μ_t and covariance Σ_t

Under the assumptions the conditional distribution $p(\, x_t \mid z_{0:t}\,) = \mathrm{N}(\mu_t, \Sigma_t)$ as computed by KF

KF computes **Minimum Mean Square Error (MMSE) Estimator**: $\mu_t = \mathrm{argmin}_{\tilde{x}}\ \mathrm{E} \parallel x_t - \tilde{x} \parallel^2$

Computational advantage: update requires time that is **polynomial in the dimension of the system**

Limitations from Linearity

Extensions via linearization (Taylor expansion, Jacobian) EFK, stochastic regression: UKF