Observers and Kalman Filters

CS 393R: Autonomous Robots

Slides Courtesy of
Benjamin Kuipers

Good Afternoon Colleagues

- Are there any questions?

Stochastic Models of an Uncertain World

$$\begin{aligned} \dot{\mathbf{x}} &= F(\mathbf{x},\mathbf{u}) \\ \mathbf{y} &= G(\mathbf{x}) \end{aligned} \quad \Rightarrow \quad \begin{aligned} \dot{\mathbf{x}} &= F(\mathbf{x},\mathbf{u},\varepsilon_1) \\ \mathbf{y} &= G(\mathbf{x},\varepsilon_2) \end{aligned}$$

- Actions are uncertain.
- Observations are uncertain.
- $\varepsilon_i \sim N(0,\sigma_i)$ are random variables

Observers

$$\begin{aligned} \dot{\mathbf{x}} &= F(\mathbf{x},\mathbf{u},\varepsilon_1) \\ \mathbf{y} &= G(\mathbf{x},\varepsilon_2) \end{aligned}$$

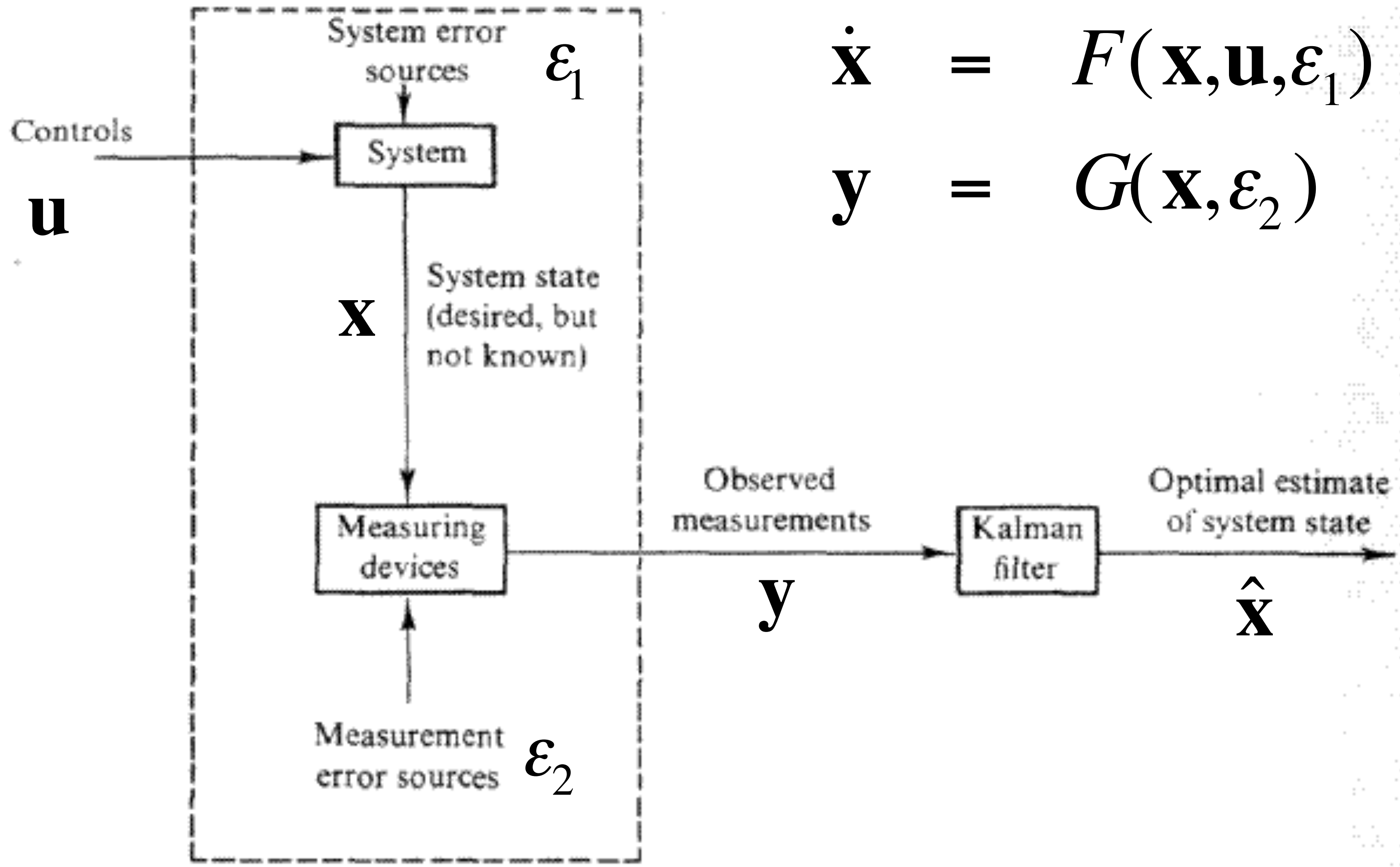
- The state **x** is unobservable.
- The sense vector **y** provides noisy information about **x**.
- An *observer* $\hat{\mathbf{x}} = Obs(\mathbf{y})$ is a process that uses sensory history to estimate **x**.
- Then a control law can be written

$$\mathbf{u} = H_i(\hat{\mathbf{x}})$$

Kalman Filter: Optimal Observer

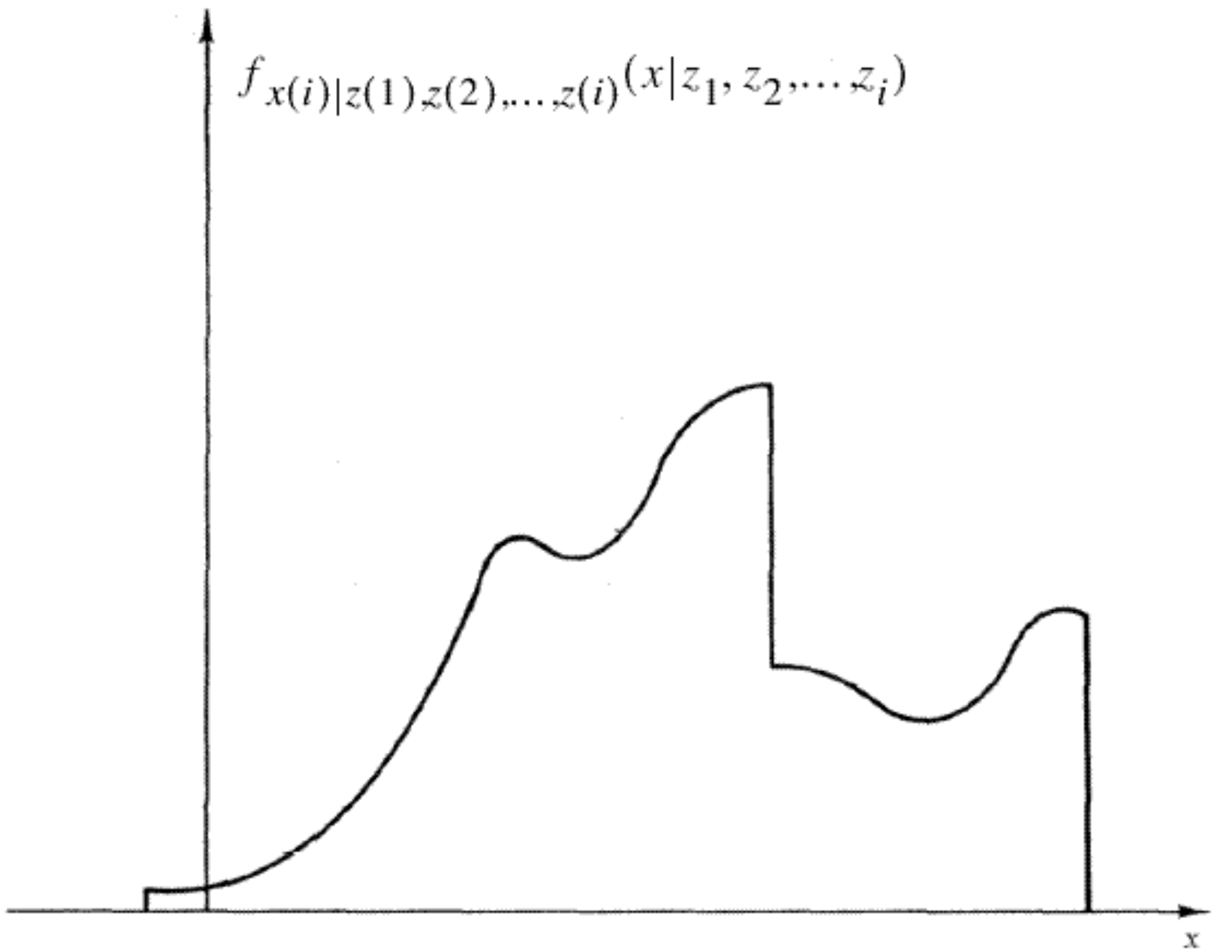
$$\dot{\mathbf{x}} = F(\mathbf{x}, \mathbf{u}, \varepsilon_1)$$

$$\mathbf{y} = G(\mathbf{x}, \varepsilon_2)$$



Estimates and Uncertainty

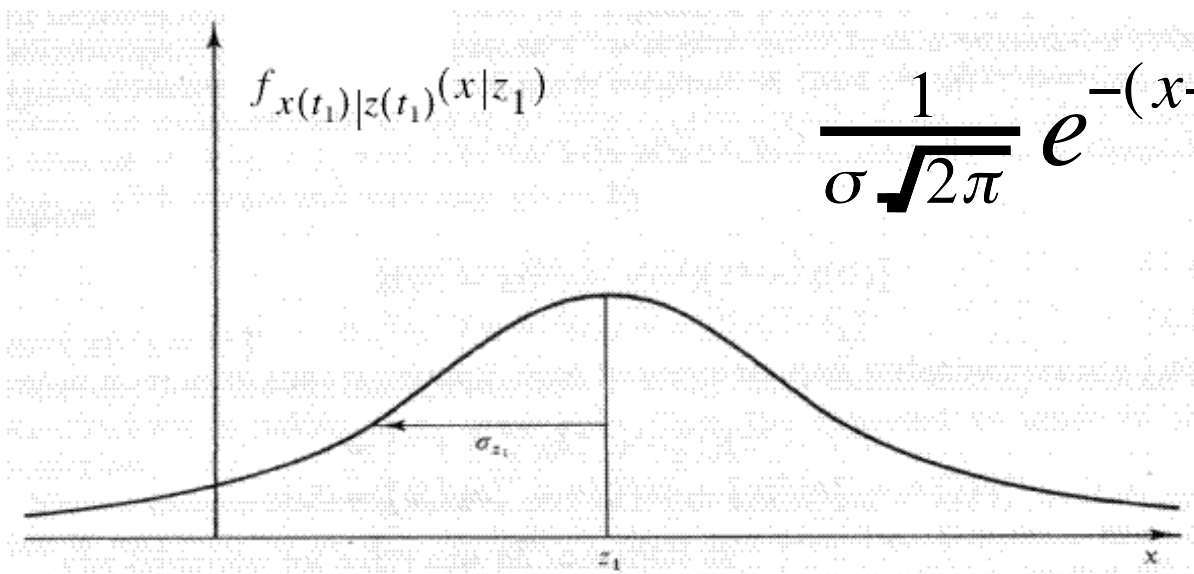
- Conditional probability density function



Gaussian (Normal) Distribution

- Completely described by $N(\mu,\sigma^2)$
  - Mean μ
  - Standard deviation σ, variance σ^2

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}$$



The Central Limit Theorem

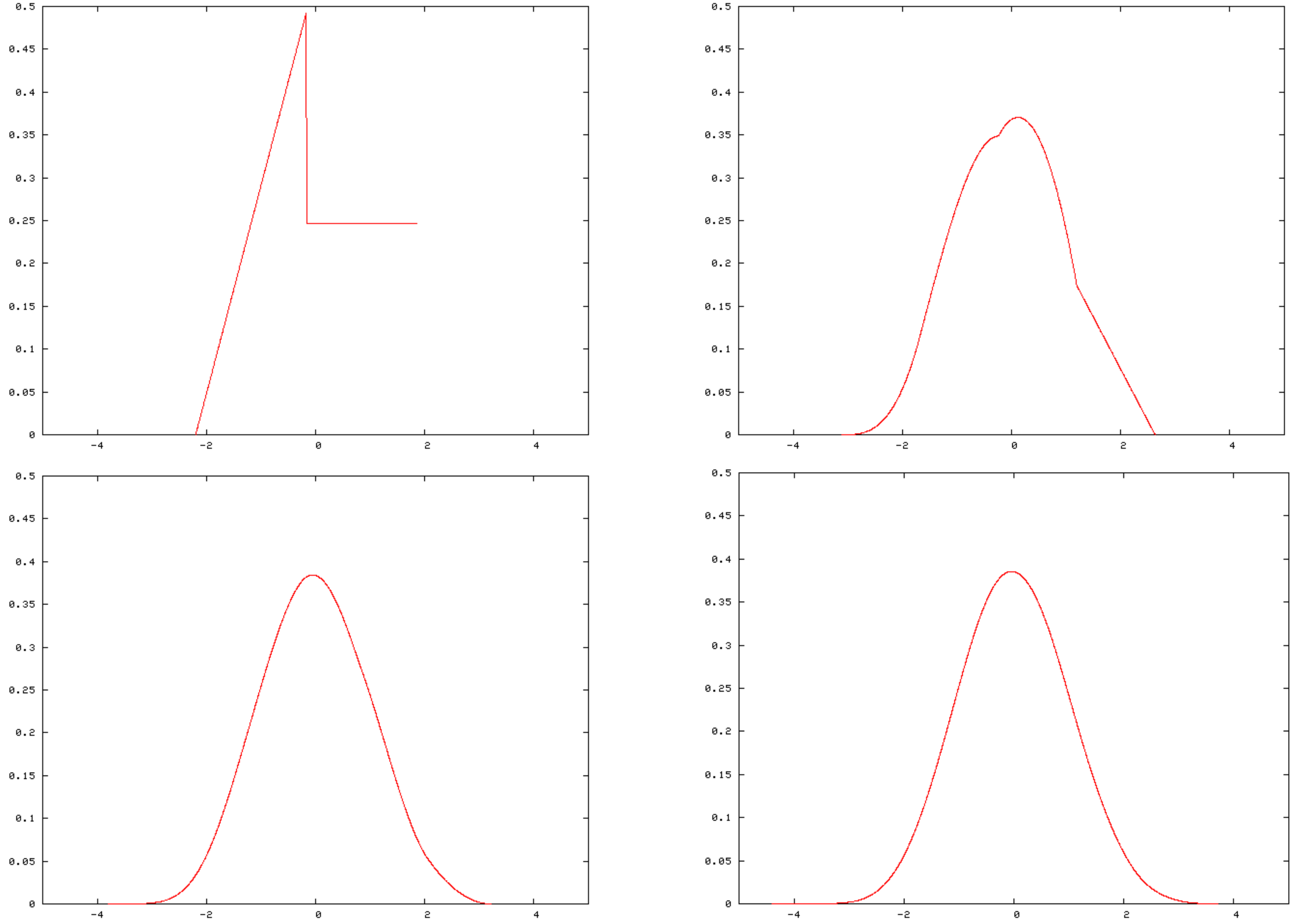
- The sum of many random variables
  - with the same mean, but
  - with arbitrary conditional density functions,

  converges to a Gaussian density function.

- If a model omits many small unmodeled effects, then the resulting error should converge to a Gaussian density function.

Illustrating the Central Limit Thm

– Add 1, 2, 3, 4 variables from the same distribution.



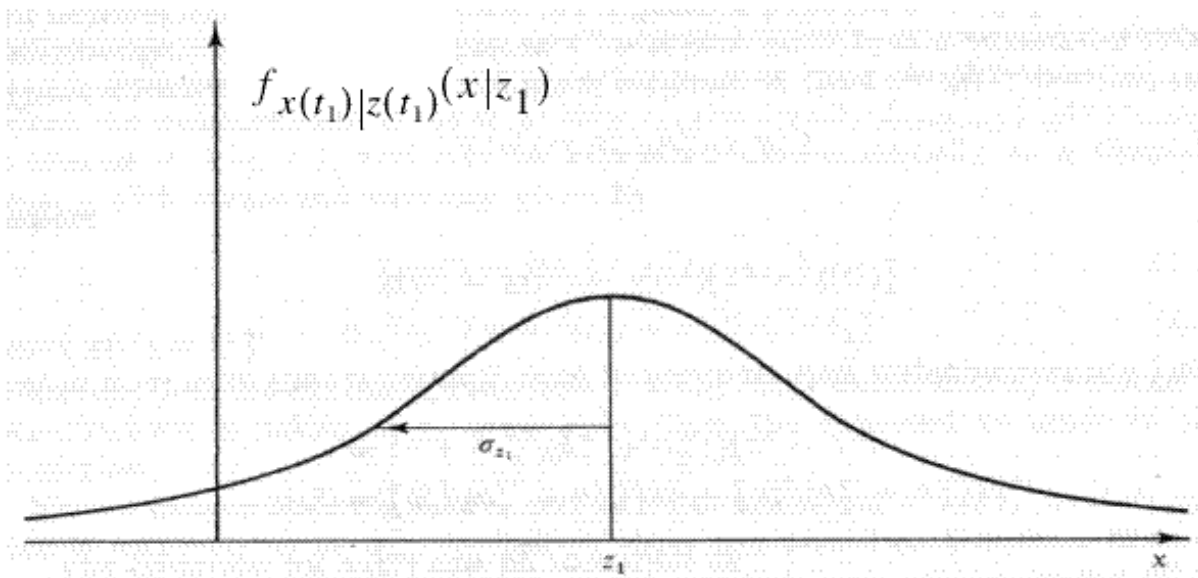
Detecting Modeling Error

- Every model is incomplete.
  - If the omitted factors are *all* small, the resulting errors should add up to a Gaussian.

- If the error between a model and the data is not Gaussian,
  - Then some omitted factor is *not* small.
  - One should find the dominant source of error and add it to the model.

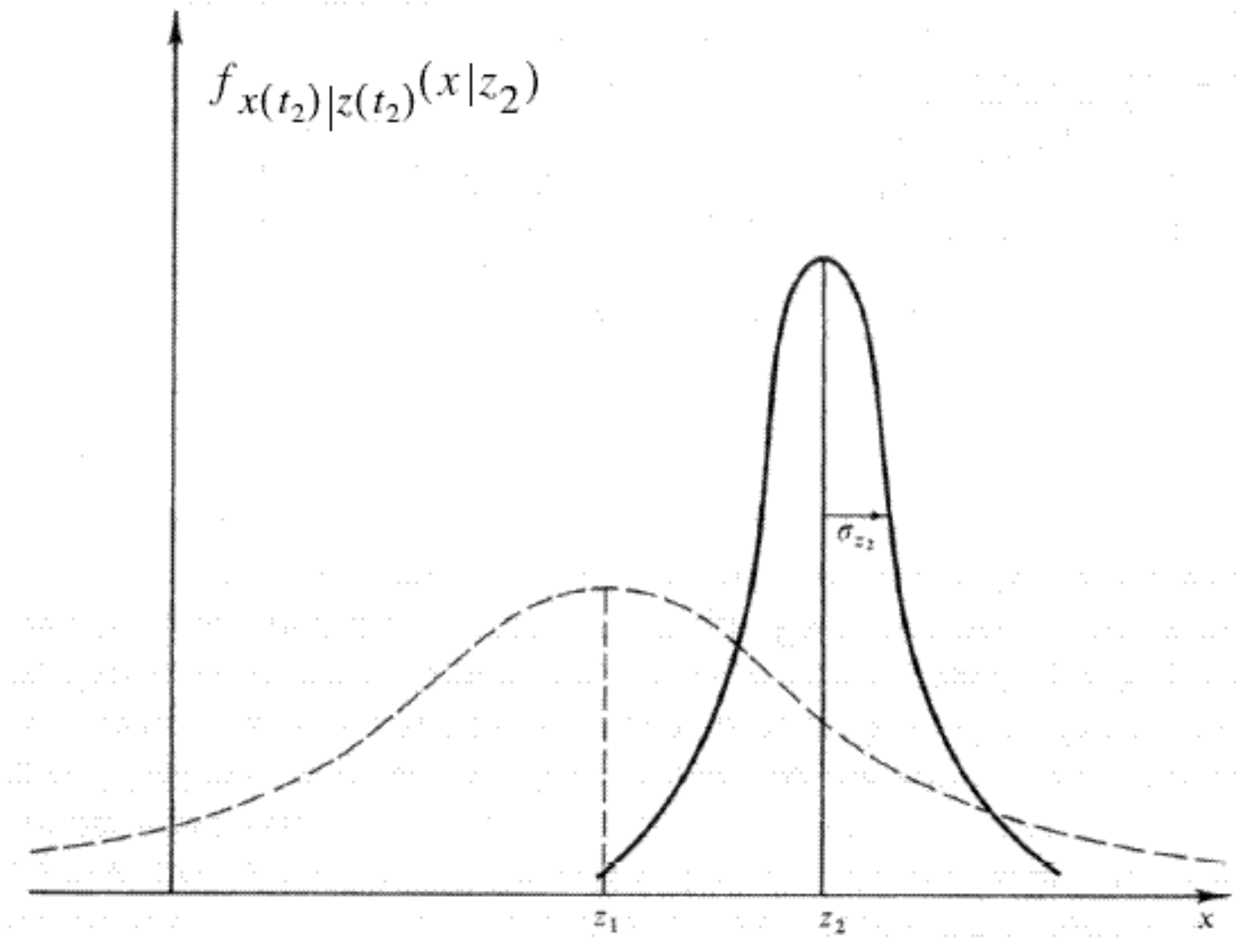
Estimating a Value

- Suppose there is a *constant* value x.
  - Distance to wall; angle to wall; etc.
- At time t_1, observe value z_1 with variance σ_1^2
- The optimal estimate is $\hat{x}(t_1) = z_1$ with variance σ_1^2

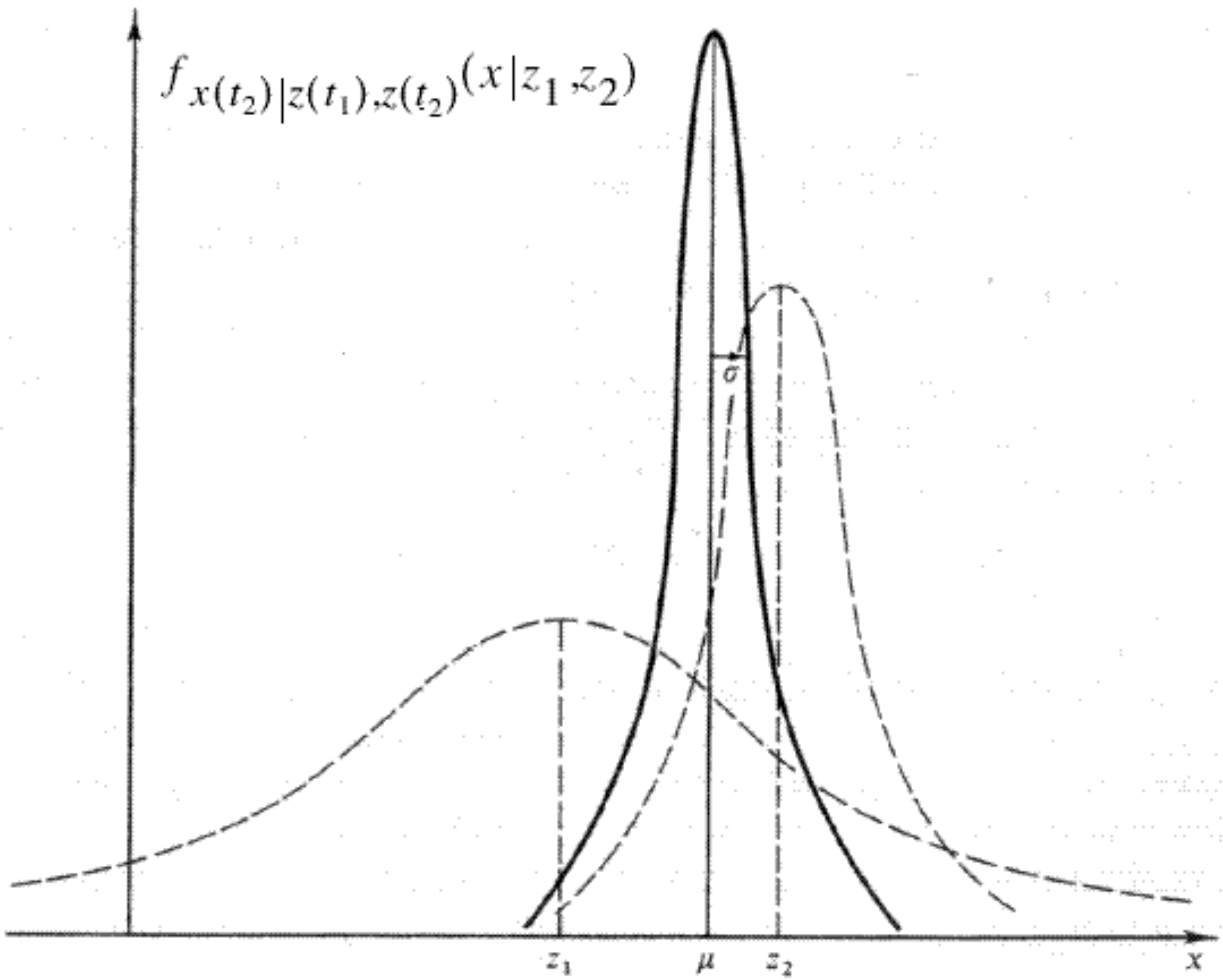


A Second Observation

- At time t_2, observe value z_2 with variance σ_2^2



Merged Evidence



Update Mean and Variance

- Weighted average of estimates.

$$\hat{x}(t_2) = Az_1 + Bz_2 \qquad\qquad A + B = 1$$

- The weights come from the variances.
  - Smaller variance = more certainty

$$\hat{x}(t_2) = \left[\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right] z_1 + \left[\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right] z_2$$

$$\frac{1}{\sigma^2(t_2)} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

From Weighted Average to Predictor-Corrector

- Weighted average:

$$\hat{x}(t_2) = Az_1 + Bz_2 = (1 - K)z_1 + Kz_2$$

- Predictor-corrector:

$$\begin{aligned}\hat{x}(t_2) &= z_1 + K(z_2 - z_1) \\ &= \hat{x}(t_1) + K(z_2 - \hat{x}(t_1))\end{aligned}$$

  – This version can be applied “recursively”.

Predictor-Corrector

- Update best estimate given new data

$$\hat{x}(t_2) = \hat{x}(t_1) + K(t_2)(z_2 - \hat{x}(t_1))$$

$$K(t_2) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

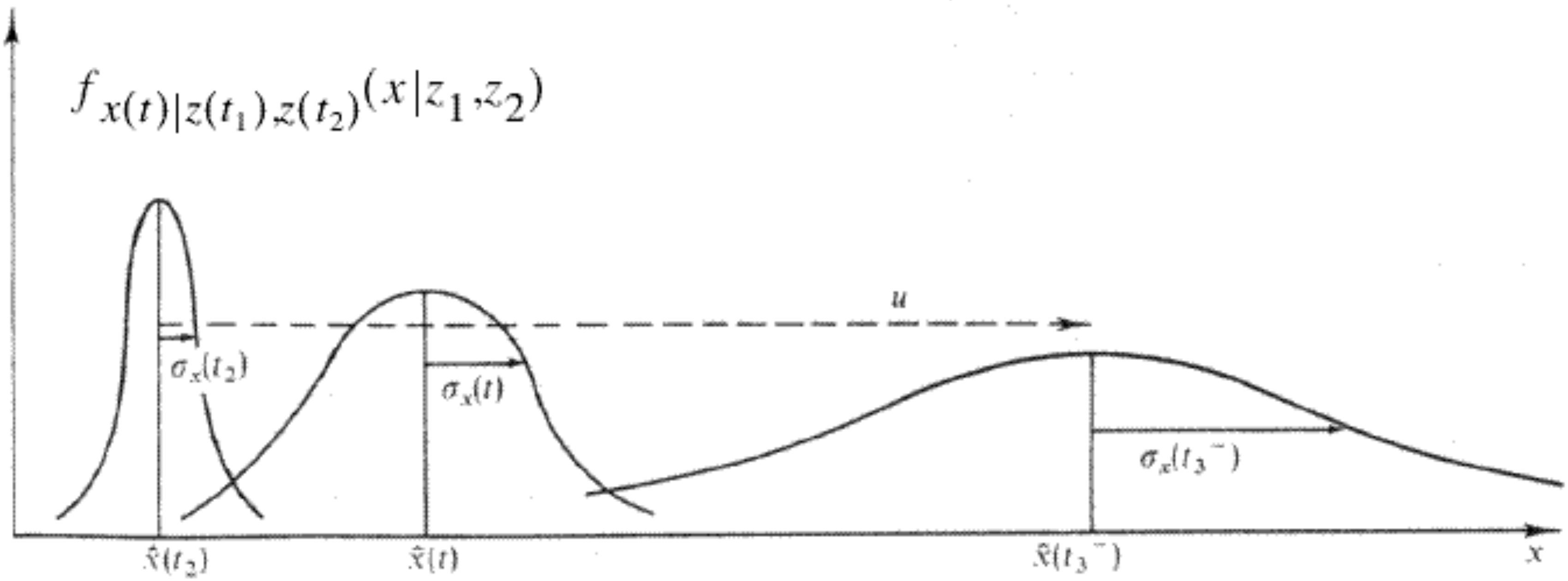
- Update variance:

$$\sigma^2(t_2) = \sigma^2(t_1) - K(t_2)\sigma^2(t_1)$$

$$= (1 - K(t_2))\,\sigma^2(t_1)$$

Static to Dynamic

- Now suppose *x* changes according to
  $\dot{x} = F(x, u, \varepsilon) = u + \varepsilon \qquad (N(0, \sigma_\varepsilon))$



Dynamic Prediction

- At t_2 we know $\hat{x}(t_2)$ $\sigma^2(t_2)$
- At t_3 after the change, before an observation.

$$\hat{x}(t_3^-) = \hat{x}(t_2) + u[t_3 - t_2]$$

$$\sigma^2(t_3^-) = \sigma^2(t_2) + \sigma_\varepsilon^2[t_3 - t_2]$$

- Next, we correct this prediction with the observation at time t_3.

Dynamic Correction

- At time t_3 we observe z_3 with variance σ_3^2
- Combine prediction with observation.

$$\hat{x}(t_3) = \hat{x}(t_3^-) + K(t_3)(z_3 - \hat{x}(t_3^-))$$

$$\sigma^2(t_3) = (1 - K(t_3))\sigma^2(t_3^-)$$

$$K(t_3) = \frac{\sigma^2(t_3^-)}{\sigma^2(t_3^-) + \sigma_3^2}$$

Qualitative Properties

$$\hat{x}(t_3) = \hat{x}(t_3^-) + K(t_3)(z_3 - \hat{x}(t_3^-))$$

$$K(t_3) = \frac{\sigma^2(t_3^-)}{\sigma^2(t_3^-) + \sigma_3^2}$$

- Suppose measurement noise σ_3^2 is large.
  - Then $K(t_3)$ approaches 0, and the measurement will be mostly ignored.
- Suppose prediction noise $\sigma^2(t_3^-)$ is large.
  - Then $K(t_3)$ approaches 1, and the measurement will dominate the estimate.

Kalman Filter

- Takes a stream of observations, and a dynamical model.
- At each step, a weighted average between
  - prediction from the dynamical model
  - correction from the observation.
- The Kalman gain $K(t)$ is the weighting,
  - based on the variances $\sigma^2(t)$ and σ_ε^2
- With time, $K(t)$ and $\sigma^2(t)$ tend to stabilize.

Simplifications

- We have only discussed a one-dimensional system.
  - Most applications are higher dimensional.
- We have assumed the state variable is observable.
  - In general, sense data give indirect evidence.

$$\dot{x} = F(x, u, \varepsilon_1) = u + \varepsilon_1$$
$$z = G(x, \varepsilon_2) = x + \varepsilon_2$$

- We will discuss the more complex case next.

Up To Higher Dimensions

- Our previous Kalman Filter discussion was of a simple one-dimensional model.
- Now we go up to higher dimensions:
  - State vector: $\mathbf{x} \in \Re^n$
  - Sense vector: $\mathbf{z} \in \Re^m$
  - Motor vector: $\mathbf{u} \in \Re^l$
- First, a little statistics.

Expectations

- Let x be a random variable.
- The expected value $E[x]$ is the mean:

$$E[x] = \int x \; p(x) \; dx \approx \bar{x} = \frac{1}{N} \sum_{1}^{N} x_i$$

  – The probability-weighted mean of all possible values. The sample mean approaches it.

- Expected value of a vector $\mathbf{x}$ is by component.

$$E[\mathbf{x}] = \bar{\mathbf{x}} = [\bar{x}_1, \cdots \bar{x}_n]^T$$

Variance and Covariance

- The variance is $E[\ (x\text{-}E[x])^2\ ]$

$$\sigma^2 = E[(x - \bar{x})^2] = \frac{1}{N}\sum_{1}^{N}(x_i - \bar{x})^2$$

- Covariance matrix is $E[\ (\mathbf{x}\text{-}E[\mathbf{x}])(\mathbf{x}\text{-}E[\mathbf{x}])^T\ ]$

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$$

  – Divide by N–1 to make the sample variance an *unbiased estimator* for the population variance.

Covariance Matrix

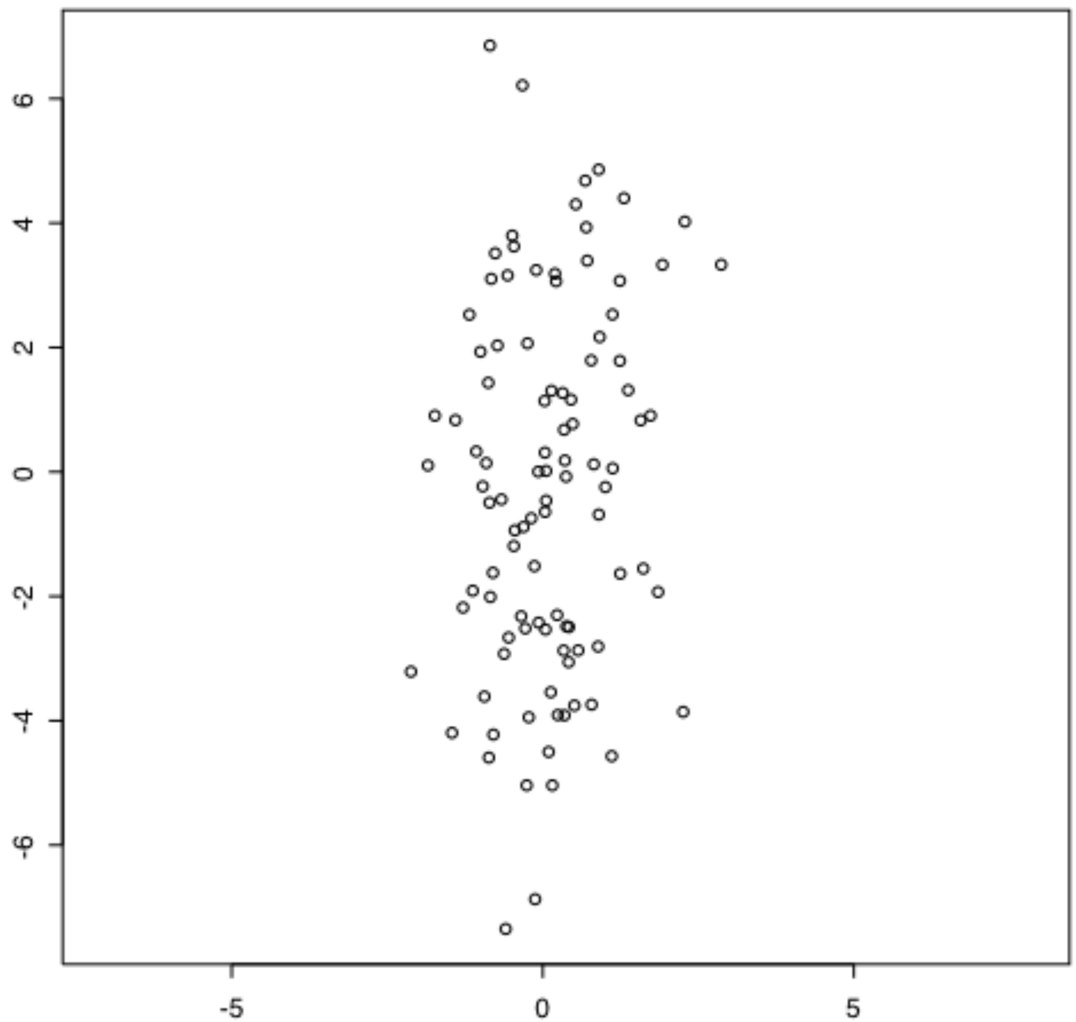
- Along the diagonal, C_{ii} are variances.
- Off-diagonal C_{ij} are essentially correlations.

$$\begin{bmatrix} C_{1,1} = \sigma_1^2 & C_{1,2} & & C_{1,N} \\ C_{2,1} & C_{2,2} = \sigma_2^2 & & \\ & & \ddots & \vdots \\ C_{N,1} & & \cdots & C_{N,N} = \sigma_N^2 \end{bmatrix}$$

Independent Variation

- x and y are Gaussian random variables (N=100)
- Generated with $\sigma_x=1 \quad \sigma_y=3$
- Covariance matrix:

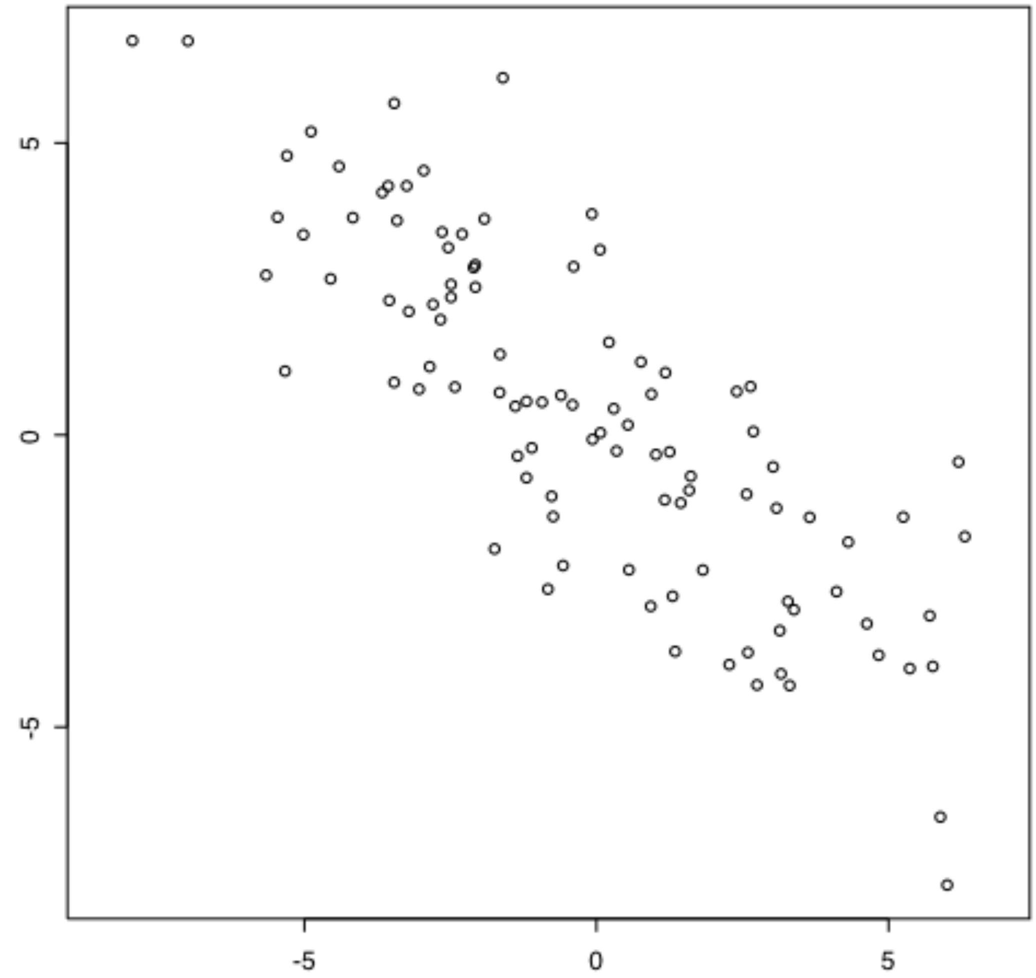
$$C_{xy} = \begin{bmatrix} 0.90 & 0.44 \\ 0.44 & 8.82 \end{bmatrix}$$



Dependent Variation

- c and d are random variables.
- Generated with $c=x+y \quad d=x-y$
- Covariance matrix:

$$C_{cd} = \begin{bmatrix} 10.62 & -7.93 \\ -7.93 & 8.84 \end{bmatrix}$$



Discrete Kalman Filter

- Estimate the state $\mathbf{x} \in \Re^n$ of a linear stochastic difference equation

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

  – process noise **w** is drawn from $N(0,\mathbf{Q})$, with covariance matrix **Q**.

- with a measurement $\mathbf{z} \in \Re^m$

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

  – measurement noise **v** is drawn from $N(0,\mathbf{R})$, with covariance matrix **R**.

- **A, Q** are $n \times n$. **B** is $n \times l$. **R** is $m \times m$. **H** is $m \times n$.

Estimates and Errors

- $\hat{\mathbf{x}}_k \in \Re^n$ is the estimated state at time-step k.
- $\hat{\mathbf{x}}_k^- \in \Re^n$ after prediction, before observation.
- Errors: $\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^-$

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$$

- Error covariance matrices:

$$\mathbf{P}_k^- = E[\mathbf{e}_k^- \mathbf{e}_k^{-T}]$$

$$\mathbf{P}_k = E[\mathbf{e}_k \mathbf{e}_k^T]$$

- Kalman Filter's task is to update $\hat{\mathbf{x}}_k$ $\mathbf{P}_k$

Time Update (Predictor)

- Update expected value of **x**

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

- Update error covariance matrix **P**

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}$$

- Previous statements were simplified versions of the same idea:

$$\hat{x}(t_3^-) = \hat{x}(t_2) + u[t_3 - t_2]$$

$$\sigma^2(t_3^-) = \sigma^2(t_2) + \sigma_\varepsilon^2[t_3 - t_2]$$

Measurement Update (Corrector)

- Update expected value

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-)$$

  – *innovation* is $\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-$

- Update error covariance matrix

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_k^-$$

- Compare with previous form

$$\hat{x}(t_3) = \hat{x}(t_3^-) + K(t_3)(z_3 - \hat{x}(t_3^-))$$

$$\sigma^2(t_3) = (1 - K(t_3))\sigma^2(t_3^-)$$

The Kalman Gain

- The optimal Kalman gain $\mathbf{K}_k$ is

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H}\mathbf{P}_k^- \mathbf{H}^T + \mathbf{R})^{-1}$$

$$= \frac{\mathbf{P}_k^- \mathbf{H}^T}{\mathbf{H}\mathbf{P}_k^- \mathbf{H}^T + \mathbf{R}}$$

- Compare with previous form

$$K(t_3) = \frac{\sigma^2(t_3^-)}{\sigma^2(t_3^-) + \sigma_3^2}$$

Extended Kalman Filter

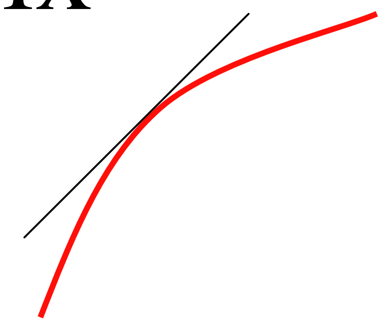
- Suppose the state-evolution and measurement equations are non-linear:

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1}$$

$$\mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{v}_k$$

  – process noise $\mathbf{w}$ is drawn from $N(0,\mathbf{Q})$, with covariance matrix $\mathbf{Q}$.
  – measurement noise $\mathbf{v}$ is drawn from $N(0,\mathbf{R})$, with covariance matrix $\mathbf{R}$.

The Jacobian Matrix



- For a scalar function $y=f(x)$,

$$\Delta y = f'(x)\Delta x$$

- For a vector function $\mathbf{y}=f(\mathbf{x})$,

$$\Delta \mathbf{y} = \mathbf{J}\Delta \mathbf{x} = \begin{bmatrix} \Delta y_1 \\ \vdots \\ \Delta y_n \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{bmatrix} \cdot \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix}$$

Linearize the Non-Linear

- Let $\mathbf{A}$ be the Jacobian of f with respect to $\mathbf{x}$.

$$\mathbf{A}_{ij} = \frac{\partial f_i}{\partial x_j}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1})$$

- Let $\mathbf{H}$ be the Jacobian of h with respect to $\mathbf{x}$.

$$\mathbf{H}_{ij} = \frac{\partial h_i}{\partial x_j}(\mathbf{x}_k)$$

- Then the Kalman Filter equations are almost the same as before!

EKF Update Equations

- Predictor step: $\hat{\mathbf{x}}_k^- = f(\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1})$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}$$

- Kalman gain: $\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H}\mathbf{P}_k^- \mathbf{H}^T + \mathbf{R})^{-1}$

- Corrector step: $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{z}_k - h(\hat{\mathbf{x}}_k^-))$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\mathbf{P}_k^-$$