23 October 2018

- Reminder RLS is not on Exam 1.
- Review: $y_i = C_i x + e_i \quad i = 1, 2, \ldots$

$x \in \mathbb{R}^n$ constant and unknown

$e_i \in \mathbb{R}^m$ unknown and unmeasured

$y_i \in \mathbb{R}^m$ is **measured**

$C_i = m \times n$ matrix is known from a "model"

$S_i > 0$ (pos. def) $m \times m$ matrix "weight"

$$\hat{x}_k := \underset{x \in \mathbb{R}^n}{\arg\min} \sum_{i=1}^{k} \underbrace{(y_i - C_i x)^\top}_{e_i^\top} S_i \underbrace{(y_i - C_i x)}_{e_i}$$

(a) **Basic Version:**

- Initialization Step: Choose n such that Q_n is invertible (full rank)

$$Q_n := \sum_{i=1}^{n} C_i^\top S_i C_i$$

$$\Gamma_n := \sum_{i=1}^{n} C_i^\top S_i y_i$$

$$\hat{x}_n := (Q_n)^{-1} \Gamma_n$$

- Recursion: For $n \leq k < N$

$Q_{k+1} = n \times n$

$$Q_{k+1} := Q_k + C_{k+1}^\top S_{k+1} C_{k+1}$$

$$K_{k+1} := (Q_{k+1})^{-1} C_{k+1}^\top S_{k+1}$$

$$\hat{x}_{k+1} := \hat{x}_k + K_{k+1} \underbrace{(y_{k+1} - \underbrace{C_{k+1}\hat{x}_k}_{\hat{y}_{k+1}})}_{\text{"new information"}}$$

(b) **Improved Version Using the Matrix Inversion Lemma:**

- Initialization Step: Choose n such that Q_n is invertible (full rank)

$$
\begin{aligned}
Q_n &:= \sum_{i=1}^{n} C_i^\top S_i C_i \\
P_n &:= (Q_n)^{-1} \\
\Gamma_n &:= \sum_{i=1}^{n} C_i^\top S_i y_i \\
\hat{x}_n &:= P_n \Gamma_n
\end{aligned}
$$

- Recursion: For $n \le k < N$

$$
\begin{aligned}
P_{k+1} &= P_k - P_k C_{k+1}^\top \overbrace{[S_{k+1}^{-1} + C_{k+1} P_k C_{k+1}^\top]}^{m \times m}{}^{-1} C_{k+1} P_k. \\
K_{k+1} &:= P_{k+1} C_{k+1}^\top S_{k+1} \\
\hat{x}_{k+1} &:= \hat{x}_k + K_{k+1} \left(y_{k+1} - C_{k+1} \hat{x}_k \right)
\end{aligned}
$$

- How to Derive the Riccati Equation? It comes from the Matrix Inversion Lemma

$$
\begin{aligned}
Q_{k+1} &= Q_k + C_{k+1}^\top S_{k+1} C_{k+1} \\
Q_{k+1}^{-1} &= \left(Q_k + C_{k+1}^\top S_{k+1} C_{k+1} \right)^{-1} \\
&= Q_k^{-1} - Q_k^{-1} C_{k+1}^\top \left[S_{k+1}^{-1} + C_{k+1} Q_k^{-1} C_{k+1}^\top \right]^{-1} C_{k+1} Q_k^{-1} \\
P_k &:= Q_k^{-1} \\
P_{k+1} &= P_k - P_k C_{k+1}^\top \left[S_{k+1}^{-1} + C_{k+1} P_k C_{k+1}^\top \right]^{-1} C_{k+1} P_k.
\end{aligned}
$$

- Jacopo Francesco Riccati (1676-1754) `http://en.wikipedia.org/wiki/Jacopo_Riccati`

- HW 06 compares batch processing to RLS to RLS+MIL
- Recursive estimators so far
  - HW 05 moving window
  - RLS

Today: Underdetermined Equations (Quickly)

- $Ax = b$, $\text{rank}(A) < \dim(x)$

$\therefore$ Exist many solutions Idea: Choose the solution of smallest norm

We suppose the inner product on $\mathbb{R}^n$ is defined by $< x, y >= x^\top S y$, where $S > 0$ is an $n \times n$ positive definite matrix. We denote the corresponding norm by $\|x\|_S := (x^\top S x)^{1/2}$.

Let $Ax = b$, where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A = m \times n$, $n > m$, and $\text{rank}(A) = m$. In other words, we are assuming the rows of A are linearly independent instead of the columns of A are linearly independent.

Def. If $\forall b_0 \in \mathbb{R}^m, \exists x_0 \in \mathbb{R}^n$, such that $b_0 = Ax_0$, $b = Ax$ is consistent.

Fact: If $\text{rank}(A) =$ the number of rows, then the equation $b = Ax$ is consistent.

Fact: Suppose x_0 is such that $b_0 = Ax_0$, and $V = \{x \in \mathbb{R}^n | b = Ax\}$ is the set of solutions. Then, $V = x_0 + \mathcal{N}(A)$, where $\mathcal{N}(A) = \{x \in \mathbb{R}^n | Ax = 0\}$ is the null space of A. Therefore, V is the translate of a subspace. We can also say that V is an "affine" space.

Theorem: If the rows of A are linearly independent, then

$$\hat{x} := \underset{x \in V}{\text{argmin}} \|x\|_S = \underset{Ax=b}{\text{argmin}} \|x\|_S = \underset{Ax=b}{\text{argmin}} (x^\top S x)^{\frac{1}{2}}$$

exists, is unique, and is given by

$$\hat{x} = S^{-1} A^\top \left(A S^{-1} A^\top \right)^{-1} b.$$

Proof is developed in HW.

$S > 0$, $\|x\|_S^2 := \langle x, x \rangle := x^\top S x$

Kalman Filter Motivation

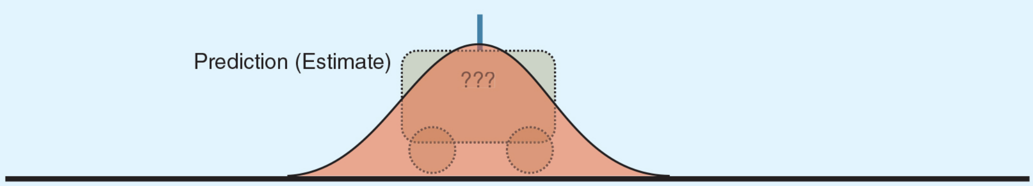
x_0 = Initial Condition

Dynamics

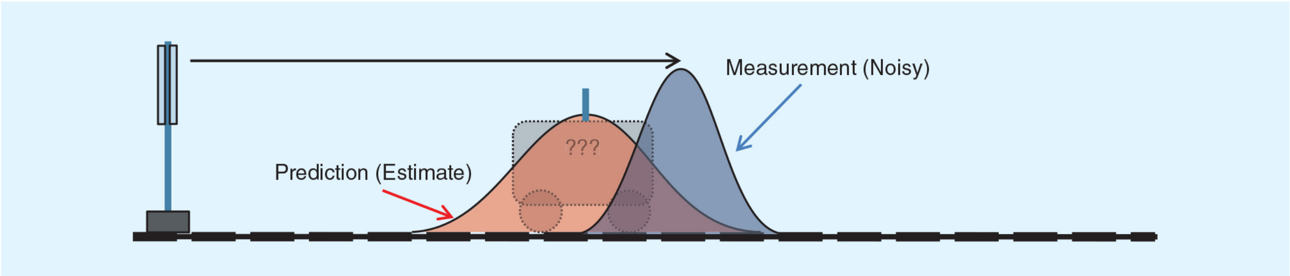
$\dot{v} = 0$ + "wind-gust noise"

[FIG2] The initial knowledge of the system at time $t = 0$. The red Gaussian distribution represents the pdf providing the initial

Uncertainty typically grows as time increases



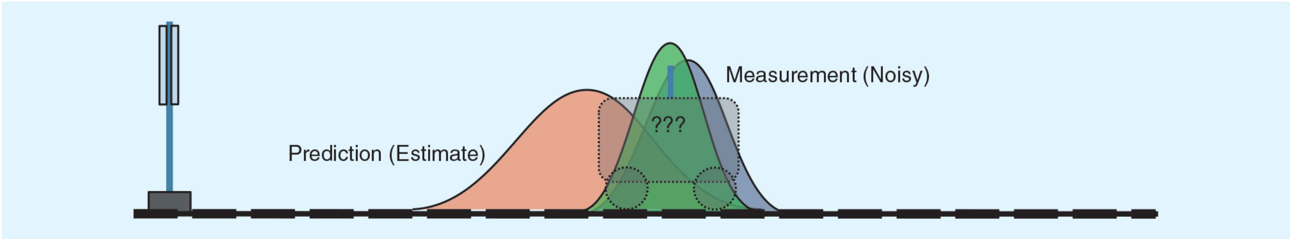
[FIG3] **Here, the prediction of the location of the train at time $t = 1$ and the level of uncertainty in that prediction is shown. The confidence in the knowledge of the position of the train has decreased, as we are not certain if the train has undergone any accelerations or decelerations in the intervening period from $t = 0$ to $t = 1$.**



[FIG4] Shows the measurement of the location of the train at time $t = 1$ and the level of uncertainty in that noisy measurement, represented by the blue Gaussian pdf. The combined knowledge of this system is provided by multiplying these two pdfs together.

$x(t)$ has a "probability distribution"

$y(t)$ is not perfect = "probability distribution"



[FIG5] Shows the new pdf (green) generated by multiplying the pdfs associated with the prediction and measurement of the train's location at time $t = 1$. This new pdf provides the best estimate of the location of the train, by fusing the data from the prediction and the measurement.

Big Question: How to "fuse" (= combine) the two quantities to have a "best estimate" $\hat{x}(t)$???

RLS: $A_k = I$, $G_k = 0$

V_k = deterministic = no stochastic or random model.

$$x_{k+1} = A_k x_k + G_k w_k$$
$$y_k = C_k x_k + v_k$$

The Kalman Filter

Definition of Terms:

$$\begin{aligned}
\widehat{x}_{k|k} &:= \mathcal{E}\{x_k | y_0, \cdots, y_k\} \\
P_{k|k} &:= \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \cdots, y_k\}
\end{aligned}$$

$$\begin{aligned}
\widehat{x}_{k+1|k} &:= \mathcal{E}\{x_{k+1} | y_0, \cdots, y_k\} \\
P_{k+1|k} &:= \mathcal{E}\{(x_{k+1} - \widehat{x}_{k+1|k})(x_{k+1} - \widehat{x}_{k+1|k})^\top | y_0, \cdots, y_k\}
\end{aligned}$$

Initial Conditions:

$$\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \quad \text{and} \quad P_{0|-1} := P_0 = \text{cov}(x_0)$$

For $k \geq 0$

Measurement Update Step:

$$\begin{aligned}
K_k &= P_{k|k-1} C_k^\top \left(C_k P_{k|k-1} C_k^\top + Q_k \right)^{-1} \\
&\quad \text{(Kalman Gain)} \\
\widehat{x}_{k|k} &= \widehat{x}_{k|k-1} + K_k \left(y_k - C_k \widehat{x}_{k|k-1} \right) \\
P_{k|k} &= P_{k|k-1} - K_k C_k P_{k|k-1}
\end{aligned}$$

Time Update or Prediction Step:

$$\begin{aligned}
\widehat{x}_{k+1|k} &= A_k \widehat{x}_{k|k} \\
P_{k+1|k} &= A_k P_{k|k} A_k^\top + G_k R_k G_k^\top
\end{aligned}$$

End of For Loop (Just stated this way to emphasize the recursive nature of the filter)

First Probability Review

$f : \mathbb{R} \to \mathbb{R}$ is a probability density for a continuous (real-valued) random variable X if

a) $\forall x \in X, \quad f(x) \geq 0$

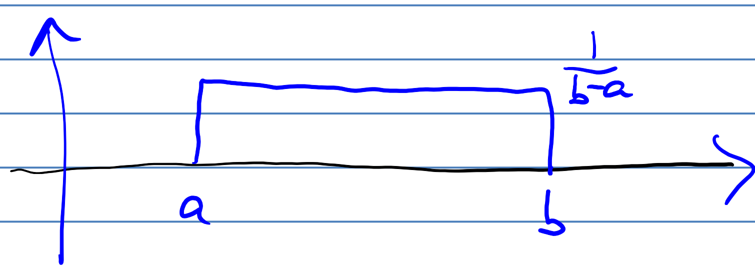
b) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$

c) $\forall \; x_1 < x_2$, then the probality that X takes values between x_1 and x_2 is $\displaystyle P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)\,dx$

Examples

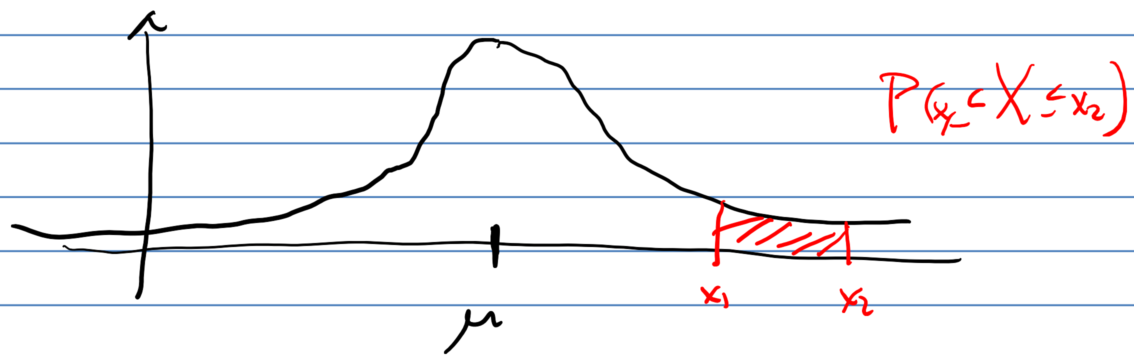
1. Uniform Random Variable has density on $[a,b]$, $a<b$

$$f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$$



② Normal or Gaussian Random Variable

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



Mean Value :

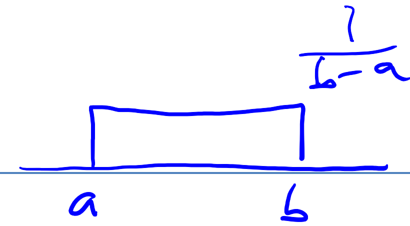
$$\mu := E\{X\} := \int_{-\infty}^{\infty} x f(x)\, dx$$

AKA (Also known as) Average Value or Expected Value.

In general let $g: \mathbb{R} \to \mathbb{R}$ be a function,

$$E\{g(X)\} := \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

Expectation operator.

Examples ① $f(x) \sim$ 

$$\mu = E\{X\} = \frac{b+a}{2}$$

② $f(x) \sim N(\mu, \sigma^2)$

$$\mu = E\{X\}$$

Variance of a Random Variable

(spread about the mean value)

$$\mathrm{var}(X) := E\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

(2-norm squared for random variables)

Examples

① Uniform R.V. on $[a,b]$

$$\mathrm{var}\{X\} = E\{(X-\mu)^2\}, \quad \mu = \frac{b+a}{2}$$

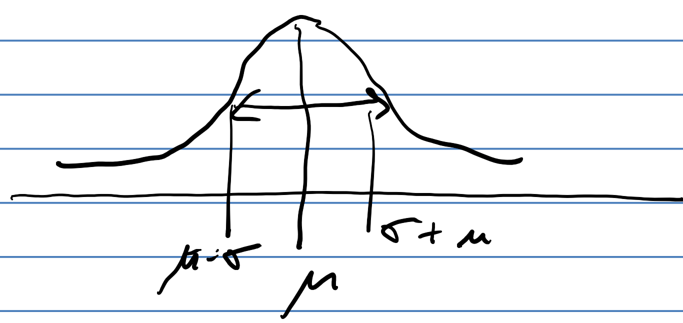
$$= \int_a^b \left(x - \frac{b+a}{2}\right)^2 \frac{1}{b-a}\, dx$$

$= $ (check this at home)

$= \frac{1}{2}(b-a)^2$

(2) Normal density $N(\mu, \sigma^2)$

$$E\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \sigma^2$$



Random Vectors

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}, \quad X_i \text{ is a random variable}$$

its density $f_X(x_1, x_2, \ldots, x_n) \geq 0$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$$

Mean Vector

$$\mu := E\{X\} := \begin{bmatrix} E\{X_1\} \\ E\{X_2\} \\ \vdots \\ E\{X_n\} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}$$

$$E\{X_i\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i \, f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n$$

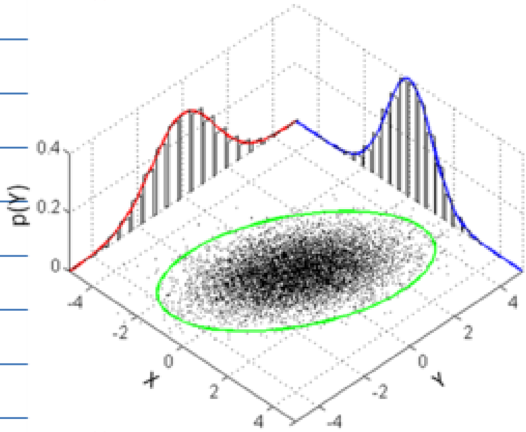
Covariance Matrix

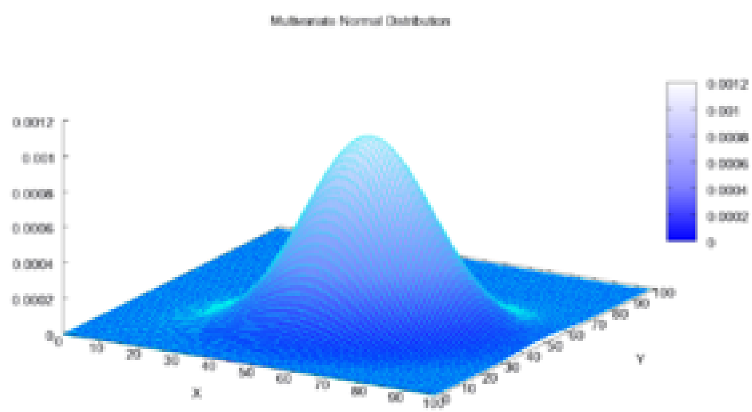
$$Q := E\{ \underbrace{(X-\mu)}_{n\times 1} \underbrace{(X-\mu)^T}_{1\times n} \} = n\times n$$

$$[Q]_{ij} = E\{ (X_i - \mu_i)(X_j - \mu_j) \} = 1\times 1$$

Fact $Q \geq 0$ (always at least pos. semi definite)

In our work, we will always assume $Q > 0$





Q big $\Rightarrow$ lots of uncertainty

Q^{-1} is then small

Q^{-1} is called the Information Matrix

$$Q = \begin{bmatrix} 10 & 0 \\ 0 & 0.1 \end{bmatrix} \quad Q^{-1} = \begin{bmatrix} \frac{1}{10} & 0 \\ 0 & 10 \end{bmatrix}$$

$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, we have 100 times more information in X_2 than in X_1.

$$Q = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = O^T \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} O$$

where $O = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

Define a new random variable by $Y = OX$

Assume $\mu = E\{X\} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$\begin{aligned} \mathrm{cov}\{Y\} &= E\{YY^T\} = E\{OXX^TO^T\} \\ &= O\,E\{XX^T\}O^T \\ &= O\,Q\,O^T = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

$$Y = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}$$

$$\mathrm{var}(X) = E\{ \underbrace{(X-\mu)^T}_{1\times n} \underbrace{(X-\mu)}_{n\times 1} \} = 1\times n$$

$$= \mathrm{trace}(\Sigma)$$

$= \text{an important norm}^2 \text{ for us!}$