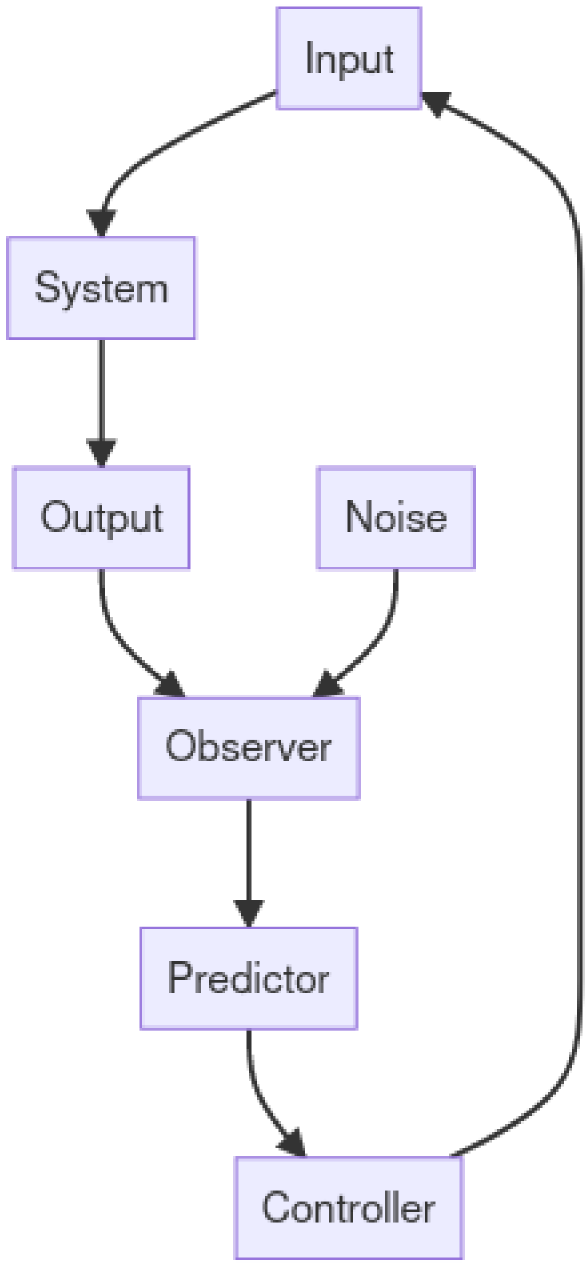
Lecture 1



- Observer = what happened
- Predictor = what will happen
- Control = how to make it happen

Estimate a constant

- There is noise in the system
- $\frac{1}{N}\sum_{k=1}^{N} X_k$
- How big should we make N?

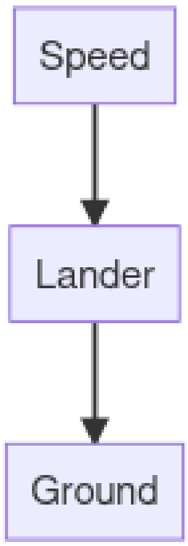
Model

- $X_{m+1} = X_m$
- X = state
- $Y_m = X_m + noise$
- Can assume that the noise has a Gaussian distribution
- We care about the variance

Systems : Models

- $X_{m+1} = A(X_m 1, u_m)$ input
- $y_m = h(X_m)$ output
- This is said to be stochastic in nature

Planet Lander Example:



- $x(t)$ = position = 0

- $\dot{x}(t)$ = speed = 0

- cost |u|

- You estimate to stop from smashing into the ground
- We can then make the problem optimal by minimizing cost

$$\begin{array}{cccc} 0 & 0 & 0 & G \\ 0 & [] & 0 & 0 \\ R & 0 & 0 & 0 \end{array}$$

- The observer is a Kalman filter
- Optimal control
- Based on the bellman equation

Probability

- u = universe of all possible outcomes
- A is the set we want
- $P(A) = \frac{A}{u}$
- O <= P(A) <= 1
- B is some other set of outcomes and we want the intersection of B and A
- $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$
- Bayesian reasoning

Lecture 2

Probability Partitioning

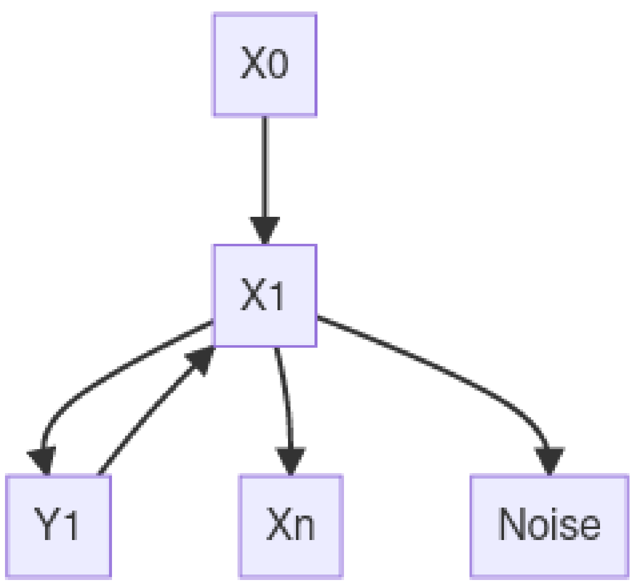
$$\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}$$

- $A_k \cap A = \phi$
- $VA_k = u$
- $P(c) = \sum P(c|A_k) * A_k$

Hero Example

- 1/100 people is a hero
- 90% accurate, 10% false positive
- 80% accurate, 20% false negative
- P(H | X) = 0.9
- P(H | !X) = 0.2
- P(X) = 0.01
- $P(X|H) = \frac{P(H|X)*P(X)}{P(H)}$
- $P(X|H) = \frac{0.9*0.01}{0.01*0.90.99*0.2} = 0.05 = 5$

Bayesian Reasoning



- $P(X|Y) = \int XP(X|Y)dx$

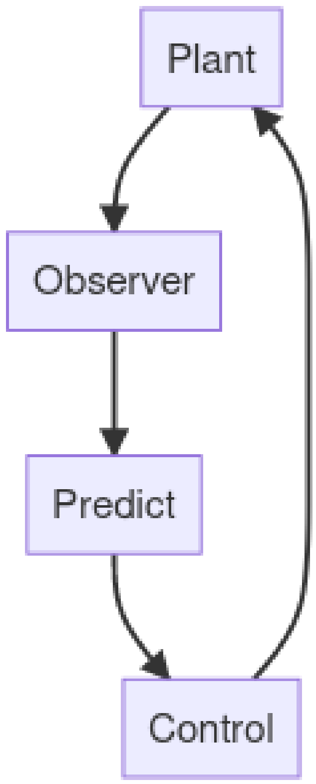
Hallway Robot

- 0 1 2 3 4 5 6 7 8 9
- 0 1 1 0 0 0 1 0 0 0
- $P_1(x) = 0 $\frac{1}{3}$ $\frac{1}{3}$ 0 0 0 $\frac{1}{3}$ 0 0 0
- $P_0(x) = $\frac{1}{7}$ 0 0 $\frac{1}{7}$ $\frac{1}{7}$ $\frac{1}{7}$ 0 $\frac{1}{7}$ $\frac{1}{7}$ $\frac{1}{7}$
- $X_0 = 0.1 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1 \quad 0.1$
- $P(X|1) = \frac{P(1|X)P(x)}{P(1)}$

Move the robot

- $X_1(x) = 0 0 $\frac{1}{3}$ $\frac{1}{3}$ 0 0 0 $\frac{1}{3}$ 0 0
- $P(X|1) = \frac{P(G|1)P(X)}{P(1)}$

Plant Example



- x' = A(x, u)
- y = u(x)
- X(T) = 0
- $X(\text{inf}) = 0$
- $J = \int_0^{\text{inf}} cost(x, u)dt$
- Arg min J u
- Minimize control input cost that approaches final state

Race Track Example

- Find the series of values to mimize cost
- $Min \sum ||\widetilde{X_k} - X_k||$
- $Min \sum ||Robot - Markers||$
- $\widetilde{X_{k+1}} = A\widetilde{X_k} + Bu_k$

Lecture 3

Control

Models:
* $x' = f(x, u)$
* $y = h(x)$

Discrete time:
* $y'' = -y + u'$ lander
* pos =
* $x_1(m+1) = x_1(m) + \Delta t x_2(m)$
* $x_2(m+1) = x_2(m) + \Delta t(-g + h)$
* $h(x) = x_1$

Linear systems

- $x_{m+1} = Ax_m + Bu_m$
- $y_{m+1} = Cx_m$
- $x_{m+1} = \begin{matrix} 1 & \Delta t \\ 1 & 0 \end{matrix} * \begin{matrix} 0 \\ \Delta t \end{matrix}$

Learning (Supervised)

- $data(x^k, y^k)$
- x^k are uncorrelated
- $F_\theta(x) = y$

- θ = parameter to learn
- Arg min $\theta \sum_{k=1}^{N} ||F_\theta(x^k) - y^k||$ = error
- Convex shape, where you iterate to find the global minimum
- $x_u = x + \alpha A'(x)$

Tracking Example

- $data(x^k, y^k)$
- Find a control input for a car to follow a path
- $x'' = u'' = F = ma$
- $$\begin{matrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{matrix} * \begin{matrix} 1 & 1 \\ g & 0 \\ - & - \\ - & - \end{matrix} * \begin{matrix} 0 \\ 1 \\ - \\ - \end{matrix}$$
- $x_{m+1} = Ax_m + Bu_m$
- $y_m = Cx_m$
- Arg min $\theta \sum_{k=0}^{N} ||yk - y^k||$ = error
- u <- u - error

Control Solution / Control Matrix

- $x_{m+1} = Ax_m + Bu_m$
- $y_m = Cx_m$
- x_0
- $x_1 = Ax_0 + Bu_0$
- $x_2 = A^2x_0 + ABu_0 + Bu_1$
- $x_N = AˆN x_0 + . . . + B u_{N-1}

- $$\begin{array}{cccc} Ax_0 & CB & u_0 & y^1 \\ Ax_1 & CAB * CB & u_1 & y^2 \\ Ax_2 & + CA^2BCABCB * & u_2 & = y^3 \\ ... & ... & ... & ... \\ A^N x_C & CA^{N+1}B... & u_{N-1} & y^k \end{array}$$

Lecture 4

Linear Algebra

- Vector space
- $V over \Re(C|)$
- $x, y \Xi V$
- $\lambda x' \Xi V$
- $\lambda_1 x + lambda_2 y \Xi V$
- $\lambda_1 lambda_2 \Xi \Re$
- $X \Xi V \exists y x + y = 0$

Apples and Oranges Example

- $A, B \Xi V$
- A = apples, B = oranges
- $\lambda_1 A = \lambda_2 B$
- $\lambda_1 = \lambda_2 = 0$
- This allows you to solve by super position because of the linear independence
- $X \Xi V$
- $\lambda_1(A) + \lambda_2(B) = x = 3A + B$
- $\lambda_1(A + B) + \lambda_2(B) = x = 3A - 2B$

- $\lambda_1(A+B)+\lambda_2(2A+2B)=$ x = Not possible

- $3A+2B=3//2$

- $\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} * \begin{matrix} \lambda_1 \\ lambda_2 \end{matrix} = x$

- $\lambda_1 \begin{matrix} 1 \\ 1 \end{matrix} + \lambda_2 \begin{matrix} 0 \\ 1 \end{matrix} = x$

- $\begin{matrix} 1 & 0 \\ 1 & 1 \end{matrix} * \begin{matrix} \lambda_1 \\ \lambda_2 \end{matrix} = (\begin{matrix} 3 \\ 1 \end{matrix})$

Finite Dimension Solution

- $\Re^M = Finite$

- e^{imw_o}

- $X(m) = \sum_{-\inf}^{\inf} \delta(m-k) * x(k)$

- Vector space
  - metric "measure"

  - inner product

  - topology

  - completeness

Metric

- Distance

- $d(x,y) >= 0$

- $d(x,x) = 0$

- $d(x,x) <= d(y,z) + d(z,y)$

- Norms

  - $||X||_2 = sqrt(\sum_{k=1}^{N} X_k^2)$

  - $d(x,y) = ||x-y||_2$

- $||x||_p = (\sum X_k^p)^{\frac{1}{P}}$
- $||x||_2^2 = \sum x_k^2$

Scalar Product

- Inner product
- $< x, y >= \sum x_i y_i$
- $< x, y >= |x||y|cos\alpha$

Linear Functions

- $V = f(x)$
- $V- > \Re$
- $Givenf(x)V- > \Re$
- $Linearf(\lambda_1 x + lambda_2 y)$
- $= \lambda_1 f(x) + \lambda_2 f(y)$
- $Existsc\Xi\Re^N$
- $f(x, y) = C^T x$
- $f(x, y) = 2x + 3y$
- $C = \begin{matrix} 2 \\ 3 \end{matrix}$
- Dual allows you to compare contorllers
  - $\Re^N- > \Re$

Linear Transforms

- $X_{m+1} = AX_m = Bu_m$
- $X_m- > 0$

- $\frac{||AX||}{||X|||}$
- $||X^* - x_m|| = ||AX^* - AX||$
- $X^* = fixedpoint$

Lecture 5

Euclidean Space

- $\Re^m$
- m equations
- m < n, x + y = 1 = Space / undeterminant
- m = n there is one solution
- m > n there are more parameters than solutions

Model Fitting

- Model -> Functional
- $F_\theta(x), (x^k, y^k)$
- $\theta, x = parameters$
- $\frac{ArgMin}{\theta} \sum_{k=1}^{N} ||F_\theta(x^k) - y^k||^2$
- $F_\theta(x) = \sum_{m=0}^{N} \theta_m \phi_m(x)$
- $\phi_m(x) = x^m$
- $\phi_m(x) = e^{-iwm}$
- Basis
  - $\lambda_1 \phi_m(x) = \lambda_2 \phi_k(x)$
  - $\lambda_1 = lambda_2 = 0$
  - $m = k$
  - $\phi_0 = \begin{matrix} 1 \\ 0 \end{matrix}$

- $\phi_1 = \begin{matrix} 0 \\ 1 \end{matrix}$
- $\begin{matrix} 2 \\ 3 \end{matrix} = 2\phi_0 + 3\phi_1$

- $F_\theta(x^k) = y^k$
- $\theta_0\phi_0(x^1) + \theta_1\phi_1(x^1) + \theta_2\phi_2(x^1) = y^1$
- $$\begin{matrix} \phi_0(x^1) & phi_1(x^1) & ... & \phi_m(x^1) & \theta_0 & y^1 \\ \phi_0(x^2) & phi_1(x^2) & ... & \phi_m(x^2) * & \theta_1 & * y^2 \\ \phi_0(x^N) & phi_1(x^N) & ... & \phi_m(x^N) & theta_m & y^k \end{matrix}$$
- $||A\theta 0y||2$
- k = horizontal, n = vertical
- N » K, significantly larger

Linear Algebra

- $||x||^2 = X^TX$
- Norm Scalar Product
- $||x||^2 = \sum x_i^2$
- $\sum x_i^2 = variance$
- $\frac{ArgMin}{\theta}||A\theta - y||^2$
- $(A\theta - y)^t(A\theta - y)$
- $\theta^TA^TA\theta - \theta^TA^Ty - y^TA\theta + y^Ty$
  - $(AB)^T = B^TA^T$
  - $X^TY = Y^TX$
  - $Y^tA\theta = A^T\theta^Ty$
  - $\theta 6TA^TA\theta - 2\theta^TA^Ty + y^Ty$

- $\frac{d}{d\theta} 2A^TA\theta - 2A^Ty = 0$
- $A^TA\theta = A^Ty =$ normal equations
- $A\theta = y$

Example

- $F(x) = ax + b$
- (x^k, y^k)
  - $ax^1 + b = y^1$
  - $ax^2 + b = y^2$

$$\begin{matrix} 1 & x^1 \\ 1 & x^2 \\ 1 & x^N \end{matrix} * \begin{matrix} b \\ a \end{matrix} * \begin{matrix} y^1 \\ y^2 \\ y^N \end{matrix}$$

- $A^TA = \begin{matrix} N & \sum x^k \\ \sum x^k & \sum {x^k}^2 \end{matrix}$
- $(\begin{matrix} a & b \\ c & d \end{matrix})^{-1} = \frac{1}{det(A)} * (\begin{matrix} d & -b \\ -c & a \end{matrix})$
- $N \sum (x^k)^2 - (\sum x_k)^2$
- $a = \frac{N(\sum x^k y^k) - (\sum x^k)(\sum y^k)}{N \sum (x^k)^2 - \sum (x_k)^2}$

Example 2

- $f(x) = ax + b$
- (1, 2) (1, 3) (2, 5)
- $\begin{matrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{matrix} * \begin{matrix} b \\ a \end{matrix} = \begin{matrix} 1 \\ 3 \\ 5 \end{matrix}$
- $\begin{matrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{matrix} * \begin{matrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{matrix} = A^TA$
- $\begin{matrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{matrix} * \begin{matrix} 1 \\ 3 \\ 5 \end{matrix} = A^Ty$

Lecture 6

Model fitting

- $(x^k, y^k) Given$
- k = 1 . . . N
- $\frac{Argmin}{\theta} \sum_{k=1} N||F_\theta(x^k) - y^k||^2$
- $F_\theta(x) = \sum_{k=0}^{M} \theta_k]phi_k(x)$
- Linear is a combination of basis elements
- $A_\theta = y$
- $$A = \begin{matrix} \theta_0(x^1)d_1(x^1) & ... & \theta_m(x^1) \\ ... & ... & ... \\ \theta_0(x^N) & ... & \theta_m(x^N) \end{matrix}$$
- $A^T A\theta = A^T Y$
- Numerically use a QR factorization

Plant Model

- $ x_m -> Plant -> y_m$
- not online
- not a filter
- infinite memory filter
- Example Question:
  - (x^k, y^k) -> F(x) = ax$
  - $J_a(x) = \sum_{k=1} N(ax^k - y^k)^2$
  - $= a^2{x^k}^2 - 2ax^ky^k + {y^k}^2$

- $\frac{d}{da} - 2ax^{k^2} - 2x^k y^k = 0$
- $ax^{k^2} = x^k y^k$
- $a = \frac{\sum x^k y^k}{\sum x^{k^2}}$

Trajectory Example

- $f(x) = ax^2 + bx + c$
- $\frac{d}{dx} = 2ax + b$
- $x = \frac{b}{2a}$ MAX
- $$\begin{matrix} 1 & x^1 & {x^1}^2 & c \\ 1 & x^2 & {x^2}^2 * b \\ 1 & x^3 & {x^3}^2 & a \end{matrix}$$

AI

- \$(x^k, y^k)
- With model
  - $\frac{ArgMin}{\theta} \sum ||F_\theta(x^k) - y^k||^2$
  - Iterate and update your θ
- Gradient search
  - $\theta_{m+1} = \theta_m + u\frac{dJ}{d\theta}$
  - $J = \sum \frac{1}{2}(F_\theta(x) - y)^2$
  - $F_\theta(x) = ax + b$ <- Line
  - $\frac{dJ}{da} = (ax + b - y)^2 * x$
  - $\frac{dJ}{db} = (ax + b - y)^2 * 1$

Lecture 7

Lecture 8

Random Variables

- $X = randomvariable$
- $X \Xi 1, 2, 3, 4, 5, 6$
- $P(X = z) = \frac{1}{6}$
- P(X) = probability density distribution
- $\sum_x P_x(X = x) = 1$
- $E(x) = \sum_x xP(X = x)$
- X pull samples
- Pull N samples
  - $u(x) = \frac{1}{N} \sum_{k=1}^{N} x_k$
  - Monte Carlo Simulation

Binning

- Pull Samples
- By binning you simulate the probability density function

Continous Case

- $X \Xi \Re$
- $P(0.55 <= X <= 0.6)$
- $\int P(x)dx = 1$
- $\int xP(x)dx = E(x)$
- $E(E(x) - x)^2$
- $= E(E(x)^2 - 2xE(x) + x^2)$

- $E(x)^2 - 2xE(x)E(x) + E(x)^2)$
- $E(x^2) - E(x)^2$
- Var X, E(x) = 0, No DC
- $Var(x) = E((E(x) - x)^2)$
- $= E(x^2) = 0$ - mean
- $PullX_k$
- $Var(x) = \frac{1}{N} \sum x_k^2$

Gauss or Normal Distribution

- $P(x) = \frac{1}{sqrt(2\pi\sigma^2)} e^{\frac{-1}{2} * \frac{(x-u)^2}{\sigma^2}}$
- $\lim \sigma -> 0$
- $P(x) = \delta(x)$
- $f(a) = \int f(x)\delta(x-a)dx$

Several Random Variables

- X, Y
- $Plot(X_k, Y_k)$
- $CO - VAR(x, y) = (E(x) - x|E(y) - y)$
- $CO - VAR(x, x) = VAR(x)$
- E(x) = 0, E(y) = 0
  - Can be accomplished with a DC filter
  - CO-VAR(x,y) = E(x * y)
  - Monte Carlo
    - $\frac{1}{N} \sum x_k y_k$
    - $x_k = cos(w_0 km)$

Lecture 9

Stochastic Process

- $E(x) = \int x P_X(x) dx$
- $E(x) linear$
- $E(\lambda_1 + lambda_2 y)$
- $= \lambda_1 E(x) + \lambda_2 E(y)$
- E(c) = c

Variance

- $VAR(x) = E((E(x) - x)^2)$
- $= E(x^2) - E(x)^2$
- $VAR(x) >= 0, VAR(c) = 0$ bi-linear quadractic
- $VAR(ax) = a^2 VAR(x)$
- $CO - VAR(x, y) = E((E(x) - x) * (E(y) - y))$
- $VAR(x + y) = VAR(X) + VAR(Y) + 2CO - VAR(x, y)$

2 Dimension Example

- $(\begin{matrix} X \\ Y \end{matrix}) * (X \quad Y)^T$
- $= (\begin{matrix} X^2 & YX \\ YX & Y^2 \end{matrix})$
- $= (\begin{matrix} VAR(X) & CO - VAR(X, Y) \\ CO - VAR(X, Y) & VAR(Y) \end{matrix})$ = covariance matrix
- $$\begin{matrix} x \\ y * x \quad y \quad z \\ z \end{matrix} = (\begin{matrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{matrix})$$
- The above matrix is symmetric

Noise

- $Whitenoise \frac{1}{\epsilon(t)}$
- All frequencies have same probability
- Gaussian noise shaped like gauss
- Pink noise $\frac{12DB}{octave}$

Stochastic (Model Fitting)

- Model Y = aX + b

Equations

- $Y_k = ax_k + b + \epsilon_k$
- $\epsilon_k = N(0, \sigma_x^2)$
- $VAR = \sigma_x^2$
-
$$CO - VAR(\epsilon_k, \epsilon_l) = \begin{matrix} 0 & k! = l \\ \sigma_x^2 & k = l \end{matrix}$$

Monte Carlo

- $\frac{1}{N} \sum_{k=1}^{N} y_k$
- $= \frac{1}{N} \sum_{k=1}^{N} aX_k + \epsilon_k$
- $= \frac{1}{N} \sum_{k=1}^{N} aX_k + \frac{1}{N} \sum_{k=1}^{N} \epsilon_k$

Ensemble Averaging

- By adding together all the noise, due to the noise being guassian the noise is equal to 0. So that gives you a meaningful measurement.

Maximum Likelihood (Estimator)

- $Y_k = aX_k + b$ -> Guass

- $P(Y_k - aX_kb)$ -> Want to maximize the probability
- $L = \prod_{k=1}^{N} P(y_k - ax_k - b)$
- $\frac{MaxL}{a,b} = \prod P(y_k - ax_k - b)$
- $Y_k - aX_k - b = \epsilon_k$
- $\epsilon_k = N(0, \sigma_x^2)$
- $\prod e^{\frac{\frac{-1}{2}(y_k - ax_k - b)^2}{\sigma_x^2}}$
- Maximize L
- First pull log
- Max(f) = Max(log(f))
- $e^{\sum_{k=1}^{N} \frac{\frac{-1}{2}(y_k - ax_k - b)^2}{\sigma_x^2}}$
- Pull Log
- $\frac{-1}{2sigma_x^2} \sum (y_k - ax_k - b)^2$ -> same as before

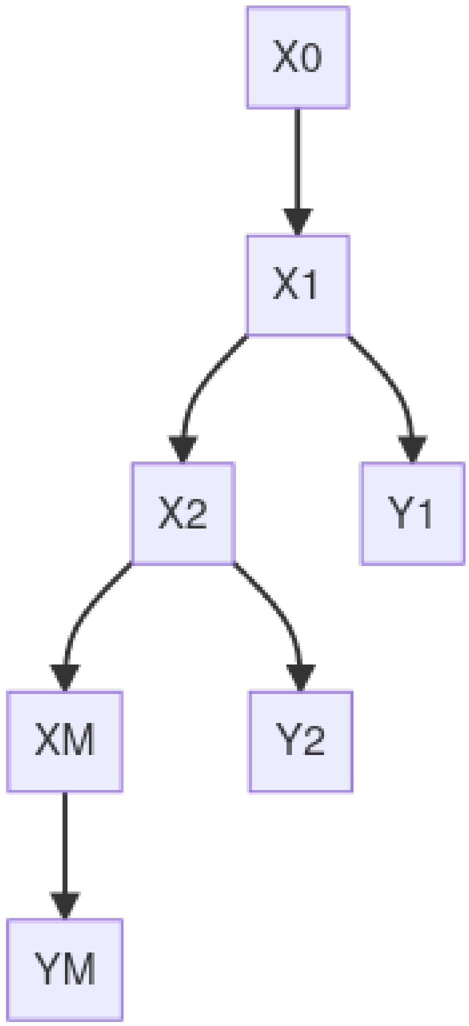
Stochastic System

- $x_{m+1} = f(x_m) = \epsilon_k$
- $y_m = h(x_m) + u_k$

Lecture 10

Stochastic Systems

- Markov Chains



- $x_{m+1} = f(x_m)$ -> deterministic
- $x_{m+1} = f(x_m) + \epsilon_m$
- $P(x_{m+1}|x_m)$
- $E(x_{m+1} = \int xP(x_{m+1})x_m dx)$
- $P(y_m|x_m)$

State Space Control

- $x' = f(x, u)$
- $y = h(x)$
- These equations are constraints on the system

Performance Measurement

- $J = \int_0^{\inf} g(x, x', u)dt$ = cost function

Particle Example

- f(x)
- g(x) = 0
- $L = g(x, u) + \lambda f(x - u) - x'$
- $\frac{dL}{dx}$
- $\frac{dL}{d\lambda}$
- $x_{m+1} = Ax_m + Bu_m + \epsilon_m$
- $y_m = Cx_m$
- $J = \sum_{m=0}^{T-1}(x_m^T Qx_m + u_m^T Ru_m) + x_T^T Q_T x_T$
- Infinite horizon -> for stability, T = inf
- Finite -> positioning
- $x^T y = scalarproduct = x^T Qy$
- $x^T x = ||x||$

Optimal Control

Controlability

- $x_{m+1} = (\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix})x_m + (\begin{matrix} 1 \\ 1 \end{matrix})u_m$

- $y(u) = (1, 1)x_m$
- Pick some x such that, go from 0 to any x
- (2, 2) works but (2, 3) doesn't because the system only works on the diagonal
- $x_{m+1} = Ax_m + B_m$
  - $x_0 = (\begin{matrix}0\\0\end{matrix})$
  - $x_1 = Bu(0)$
  - $x_2 = ABu(0) + Bu(1)$
  - $B, AB, A^2B, ...$
  - If the matrix is full rank it is controllable

Observable

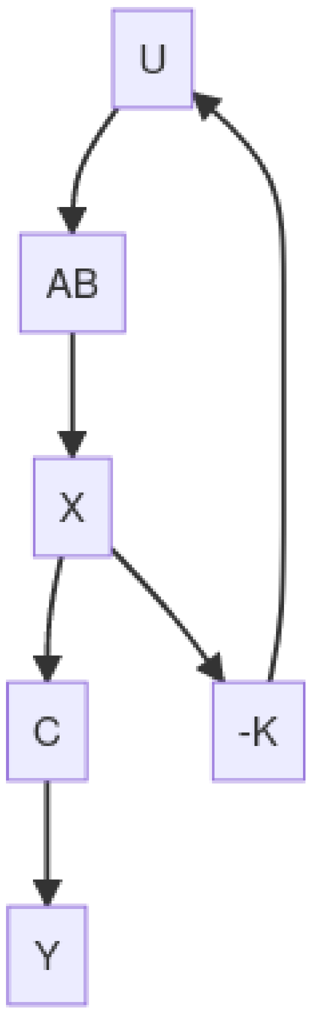
- Reconstruct x(0) from observing $x_1, x_2, x_3, ...x_m$
- $x_{m+1} = (\begin{matrix}1 & 0\\1 & 0\end{matrix})x + (\begin{matrix}1\\1\end{matrix})u$
- y(u) = (0, 1) x
- This system is not observable because you can't view all states

Lecture 11

State Space Control

- $x_{m+1} = Ax_m + Bu_m$
- $y_m = Cx_m$
  - No noise
  - Controllable
  - Observable

- Full state feedback control



- $u_m = -kx_m$
- $x_{m+1} = Ax_m - BKx_m$
- $= (A - BK|x_m)$
- We get to pick the value of K

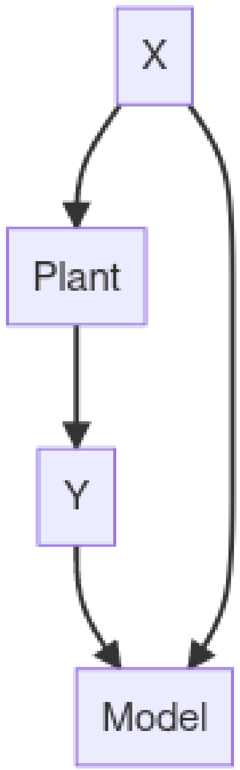
- $x_m = 0$
- $x_{m+1} = A * 0 - KB0$
- $x_{m+1} = (A - BK)x_m$
- $||x_{m+1}|| = ||(A - BK)x_m||$
- $||x_{m+1}|| <= ||A - BK||||x_m||$

Control Example

- $y'' + y' + y = u$
- $x_{m+!} = Ax_m$
- Want to control to y*
- $u = K(x^* - x)$

Observer

- Estimates the current state

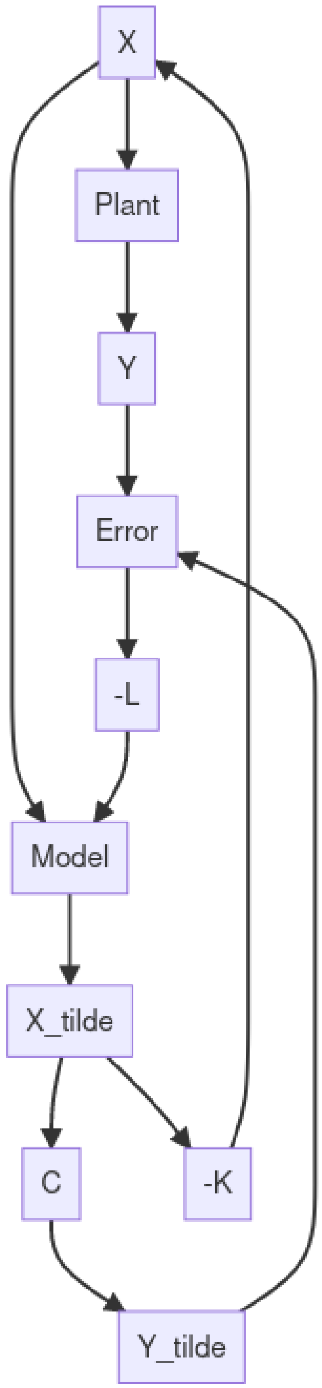


- Model means that you know A, B, C
- $\widetilde{x_{m+1}} = A\widetilde{x_m} + Bu_m$
- As long as you know X_0
- $\widetilde{x_0} = x_0$
- $\widetilde{x_1} = Ax_m + Bu_m$
- This is open loop control

Construct an observer

- Control observation error to zero
- $\widetilde{x_{m+1}} = A\widetilde{x_m} + Bu_m + L(y_m - \widetilde{y_m})$

29


X
Plant
Y
Error
-L
Model
X_tilde
C
-K
Y_tilde

Plant

- $x_{m+1} = Ax_m + Bu_m$
- $y_m = Cx_m$

Model

- $\widetilde{x_{m+1}} = A\widetilde{x_m} + Bu_m + L(y_m - \widetilde{y_m})$
- $\widetilde{y_m} = C\widetilde{x_m}$

Error Analysis

- $error = x_{m+1} - \widetilde{x_{m+1}}$
- $Ax_m = Bu_m - A\widetilde{x_m} - Bu_m - L(y_m - \widetilde{y_m}) - cx_m + c\widetilde{x_m}$
- $= (A - Lc)(x_m - \widetilde{x_m})$
- Pick L so poles are in the unit circle
- (A - Bk) = luneberg observer

Lecture 12

Observer

- Run a second system in parallel and try to estimate the internal state
- $\widetilde{x_{m+1}} = A\widetilde{x_m} + Bu_m = L(y_m - \widetilde{y_m})$
- Pick the L such that it goes ot 0

Plant

- $x_{m+1} = Ax_m + Bu_m$
- $u_m = -kx_m$
- $= -k\widetilde{x_m}$
- $= x_m - e_m$ where $e_m = x_m - \widetilde{x_m}$

- e_m approaches 0, x_m approaches 0

Learning and Fusion

- $\widetilde{x_{m+1}} = A\widetilde{x_m} + L(y_m \widetilde{y_m})$

- Want a finite memory filter because we don't want to remember all previous values.
- Predict -> Observe -> Update
- $S - N = \frac{1}{N}\sum_{k=1}^{N} y_k$

- $S - N = \frac{1}{N}\sum_{k=1}^{N-1} y_k + \frac{1}{N}y_N$

- $\frac{N-1}{N}\frac{1}{N-1}\sum_{k=1}^{N-1} y_k + \frac{1}{N}y_N$

- $S_{N-1} + \frac{1}{N}(y_m - S_{N-1})$

Example

- Everything is gauss noise

- $VAR(x) = \sigma_x^2$

- How do you use fuse 2 guassian observations

- x,y -> E(x), E(y)
  - Choose 0 <= K <= 1

  - $kx + (1-k)y$

  - Want to minimize the variance

  - $VAR(kx + (1-k)y)$

  - $E((E(ax+by) - ax - by)^2)$

  - $E((aE(x) + bE(y) - ax - by)^2)$

  - $(a(E(x) - x) + b(E(y) - y))^2$

  - $VAR(ax+by) = a^2 VAR(x) + b^{@} VAR(y) + 2abCO - VAR(x,y)$

- $k^2\sigma_x^2 + (1-k)^2\sigma_y^2$

- Minimize
  - $k^2\sigma_x^2 + \sigma_y^2 - k\sigma_y^2 + k^2\sigma_y^2$
  - $\frac{d}{dk} = 2k\sigma_x^2 - \sigma_y^2 + 2k\sigma_y^2 = 0$
  - $k = \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2}$

Lecture 13

Predictor & Observer

- Use the prediction and observation to determine state
- $P(x_m|x_{m-1})$ = prediction
- $P(x_m|y_m)$ = measurement of state
- Guassian assumption
- $E(x_m) = \widetilde{x}_m$
- $VAR(x_m) = P_m$
- $Given\widetilde{x}_{m-1}P_{m-1}$
- $Predict\widetilde{x}_m^+P_m^+$
- $Observey_m$
- Update to get $\widetilde{x}_mP_m$

Kalman Filter

- Scalar
- $x_{m+1} = ax_m + v_m$
- $y_m = cx_m + q_m$
- $N(0, \sigma_v^2)$
- $N(0, \sigma_q^2)$

- $\widetilde{x}_m = E(x_m)$
- $P_m = VAR(x_m - \widetilde{x}_m)$

1. Predict -> given $\widetilde{x}_{m-1} P_{m-1}$
    - $\widetilde{x}_m^+ = A\widetilde{x}_{m-1}$
    - $\widetilde{P}_m^+ = VAR(x_m - \widetilde{x}_m^+)$
    - $= VAR(Ax_{m-1} + v_m - A\widetilde{x}_{m-1})$
    - $= a^2 VAR(x_{m-1} - \widetilde{x}_{m-1}) + VAR(v_m) + 2CO - VAR(x_{m-1}, -\widetilde{x}_{m-1}, v_m)$
    - $= a^2 P_{,-1} + \sigma_v^2$
    - $\widetilde{x}_m^+ = a\widetilde{x}_{m-1}$
    - $P_m^+ = a^2 P_{m-1} + \sigma_v^2$
2. Observe y_m
    - $\widetilde{x}_m = \widetilde{x}_m^+ + k(y_m - \widetilde{y}_m)$
    - $\widetilde{x}_m = \widetilde{x}_m^+ + kcx_m - kc\widetilde{x}_m^+ + kq_m$
    - $P_m = VAR(x_m - \widetilde{x_m}) = E((x_m - \widetilde{x}_m)^2)$
    - $x_m - \widetilde{x}_m = x_m - \widetilde{x}_m^+ - k(cx_m + q - \widetilde{y}_m)$
    - $x_m - \widetilde{x}_m^+ - kcx_m - kc\widetilde{x}_m^+ - kq_m$
    - $(1 - kc)(x_m - \widetilde{x}_m^+) - kq_m$
    - $(ax + b)2$
        - $a = (1 - kc)$
        - $a = (x_m - \widetilde{x}_m^+)$
        - $k = -kq_m$

- $E((1 - kc)^2 (x_m \ widetildex_m^+) + k^2 q_m^2 + 2(a - kc)(x - \widetilde{x}_m^+))$
- $(1 - kc)^2 E((x_m - \widetilde{x}_m^+)^2) = k^2 E(q_m^2)$

- $(1-2kc+k^2c^2)P_m^+ + k^2\sigma_q^2$
- $\frac{d}{dk} = -2cP_m^+ + 2kc^2P_m^+ + 2k\sigma_q^2 = 0$
- $k = \frac{cP_m^+}{c^2P_m^+ + \sigma_q^2}$
- $\widetilde{x}_m = \widetilde{x}_m^+ + k(y_m - c\widetilde{x}_m^+)$
- $P_m = (1-kc)P_m^+$

Scalar Kalman

- $Given a, c, \sigma_v^2 = model uncertainty, \sigma_q^2 = observation uncertainty$
- $Start at \widetilde{x}_0 with P_0$
- Predict
  - $\widetilde{x}_m^+ = ax_{m-1}$
  - $P_m^+ = a^2P_{m-1} + \sigma_v^2$
- Update y_m
  - $\widetilde{x}_m = \widetilde{x}_m^+ + k(y_m - c\widetilde{x}_m^+)$
  - $P_m = (1-kc)P_m^+$
  - $k = \frac{cP_m^+}{c^2P_m^+ + \sigma_q^2}$

Lecture 14

Kalman

- $x_{m+1} = Ax_m = w_m$
- $y_m = cx_m + v_m$
- $E(w) = 0, E(v) = 0$
- $VAR(w) = Q$

- XX^T = CO-VAR matrix
- $E(ww^T) = Q, E(vv^T) = R$

Observer

- $\widetilde{x}_m = E(x_m)$
- $P_m = VAR(x_m - E(x_m)$
- $= E((x_m - \widetilde{x}_m)(x_m - \widetilde{x}_m)^T)$

Predict

- $\widetilde{x}_m = A\widetilde{x}_{m-1}$
- $P_m^+ = E((x_m - \widetilde{x}_m^+)(x_m - \widetilde{x}_m^+)^T)$
- $E(A()()^T A^T + 2A()v_m^T + v_m v_m^T)$
- $E(A()()^T A^T = v_m v_m^T)$
- $P_m^+ = AP_{m-1}A^T + Q$
- $\widetilde{x}_m^+ = A\widetilde{x}_{m-1}$

Update

- Observe y_m
- $\widetilde{x}_m = \widetilde{x}_m^+ = k(y - m - c\widetilde{x}_m^+)$
- $\widetilde{x}_m = \widetilde{x}_m^+ + kcx_m + kv_m - kc\widetilde{x}_m^+$
- $= (I + kc)(x_m - \widetilde{x}_m^+) + kv_m$
- Minimize P choosing K
- $E((x_m - \widetilde{x}_m)(x_m - widetildex_m)^T)$
- $\widetilde{x}_m = \widetilde{x}_m^+ + kcx_m + kv_m - kc\widetilde{x}_m^+$
- $x_m - \widetilde{x}_m = x_m - \widetilde{x}_m^+ - kcx_m - kv_m + kc\widetilde{x}_m^+$

- $(I-kc)(x_m-\widetilde{x}_m^+)-kv_m$
- $aa^T+ba^T=ab^T+bb^T$
- $(I-kc)(x_m-\widetilde{x}_m^+)(x_m-\widetilde{x}_m^+)^T(I-kc)^T+kv_mv_m^Tk^T$

Kalman Steps

- Given $\widetilde{x}_0, P_0, Q, R$
- $\widetilde{x}^+(m)=A\widetilde{x}(m-1)$
- $P^+(m)=AP(m-1)A^T+Q$
- Update with y(m)
  - $k=\frac{P^+(m)C^T}{cP^+(m)C^T+R}$
  - $\widetilde{x}_m=\widetilde{x}_m^++k(y(m)-c\widetilde{x}^+(m))$
  - $P(m)=(1-kc)P^+(m)$

Lecture 15

Bayesian Reasoning

Linear Least Sqaures

State Space Control Principles

- Controllable
- Observable
- Luneberg observer

Regression

Expected Value, Variance

- Matrix version
- Scalar version

Fusion (optimal linear mix) + Scalar Kalman

How to Practice

Bayesian Reasoning Example

- P(A | B) -> B has happened
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- Partitioning is when you split into sections
  - $\cup B_k = u$
  - $B_k \cap b_l$ = Disjoint
  - $P(k) = \sum P(A|B_k)P(K_k)$

Linear Least Squares Example

- (x^k, y^k)
- Linear Model F(x) = $\sum \alpha_k J_k(x)$
- $\frac{Minimize}{\alpha_1 \alpha_n} \sum_{k=1}^{M} ||F(x^k) - y^k||$
- Given (x^k, y^k), k=1 … N
- Model is f(x) = aX
- $\begin{bmatrix} x^1 \\ x^2 \\ x^N \end{bmatrix} * \begin{bmatrix} a \end{bmatrix} = \begin{bmatrix} y^1 \\ y^2 \\ y^N \end{bmatrix}$
- $A^T Ax = A^T y$
- $\begin{bmatrix} x^1 & ... & x^n \end{bmatrix} * \begin{bmatrix} x^1 \\ ... \\ x^N \end{bmatrix} a = \begin{bmatrix} x^1 & ... & x^n \end{bmatrix} y$
- $\sum x_k^2 a = \sum x_k y_k$

- $a = \frac{\sum x_k^2}{\sum x_k y_k}$

Probability Example

- Random Var X -> x = scalar
- Random Vector (X_1, X_2) = matrix
- E(x) -> Definition u(x) = $\frac{1}{N} \sum x^k$
- $VAR(x) = E((E(x) - x)^2)$
- $CO - VAR(x, y) = E((E(x) - x)(E(y) - y))$

Scalar Example

- x is a random variable, C is a constant
- VAR(x + C) = VAR(x)
- f(x) is a linear mapping
- $\Re - > \Re$
- I know the variance of x.
- What is the VAR(f(x))
- f(x) = ax
- VAR(ax) = aˆ2 VAR(x)

Vector Example

- $VAR(X) = E((E(x) - x)(E(x) - x)^T)$
- $(AB)^T = B^T A^T$
- $X^T y = y^T X$

State Space Control Example

- Controllable, Observable
- $x_{m+1} = Ax_m + Bu_m$
- $y_m = Cx_m$