Outline

0. Review

1. Least squares approximation — overdetermined linear system of equations

2. Linear sequential estimation

3. Nonlinear least squares approximation

4. Basis functions

5. A brief probability review

6. Homework discussion if we have time.

* lecture notes are inspired by Prof. Kristi Morgansen, Prof. Steve Brunton, Dr. Dan Caldedrone

0. Review

Least squares approximation

given measurements $\{\tilde{y}(t_j)\}, j \in \{1, \cdots m\}$ and a proposed

model $y(t) = \sum_{i=1}^{n} x_i h_i(t)$, $m \geq n$ $\quad \vec{y} = Hx$

$$\hat{\vec{x}} = (H^T H)^{-1} H^T \hat{\tilde{y}}$$

Weighted least squares approximation

$$\hat{\vec{x}} = (H^T W H)^{-1} H^T W \tilde{\vec{y}}$$

Constrained least squares approximation

$$\hat{x} = \bar{x} + K(\tilde{y}_2 - H_2 \bar{x})$$

$$K = (H_1^T W H_1)^{-1} H_2^T [H_2 (H_1^T W_1 H)^{-1} H_2^T]^{-1}$$

$$\bar{x} = (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{y}_1$$

Note: parameter estimation

the assumption is y is a linear function in x

No dynamics $(\dot{x} = \frac{dx}{dt} = Ax)$

1. Overdetermined linear system of equations 1.6

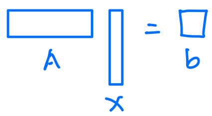
– a least square approximation example

$Ax = b \quad A \in \mathbb{R}^{m \times n} \quad b \in \mathbb{R}^{m \times 1} \quad x$ is unknown

If A is square ($m = n$) and nonsingular, then $x = A^{-1}b$ (unique)

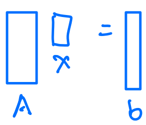
If A is non-square

- underdetermined, A is short, fat ($n > m$)



usually there will be infinite many solutions

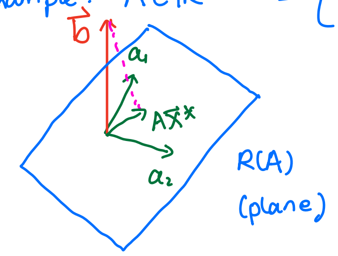
- overdetermined, A is tall and skinny ($n < m$)



usually there is no solution

Goal: find a best approximation, $\vec{x}^*$ such that $A\vec{x}^*$ is as close as possible to $\vec{b}$

Example: $A \in \mathbb{R}^{3 \times 2} = [\, a_1 \;\; a_2 \,]$



purple vector is perpendicular to $R(A)$

$\min \|\vec{b} - A\vec{x}\|_2^2$ 2-norm^2

$A\vec{x} - \vec{b}$

$\Rightarrow \left.\begin{array}{l} A\vec{x} - \vec{b} \perp a_1 \\ A\vec{x} - \vec{b} \perp a_2 \end{array}\right\}$

$\langle a_1, A\vec{x} - \vec{b} \rangle = 0 \Rightarrow a_1^T(A\vec{x} - \vec{b}) = 0$

$a_2^T(A\vec{x} - \vec{b}) = 0$

$A^T(A\vec{x} - \vec{b}) = 0$

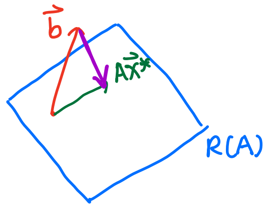
$\underline{\vec{x}^* = (A^TA)^{-1}A^T\vec{b}}$

$\vec{x}^*$ is the least squares approximation to $\min \|A\vec{x} - \vec{b}\|$

General case:

The closest vector in $R(A)$ to $\vec{b}$ is the projection of $\vec{b}$ onto $R(A)$

$A\vec{x}^* = \text{proj}_{R(A)} \vec{b}$ $\qquad$ $\vec{x}^*$ is the least squares approximation



$$A\vec{x}^* - \vec{b} = \underline{\text{proj}_{R(A)} \vec{b} - \vec{b}} \in R(A)^\perp$$

$$R(A)^\perp = N(A^T)$$

$$A^T(A\vec{x}^* - \vec{b}) = 0$$

$$\Rightarrow \vec{x}^* = (A^T A)^{-1} A^T \vec{b}$$

$\boxed{\text{minimum norm}}$: $\quad \min \|\vec{x}\|_2^2$

for underdetermined system $\qquad$ s.t. $A\vec{x} = \vec{b}$

Solution: Lagrangian $\mathcal{L} = \|\vec{x}\|_2^2 + \lambda^T(\vec{b} - A\vec{x})$

$$\frac{\partial \mathcal{L}}{\partial x} = 2\vec{x} - A^T\lambda = 0 \Rightarrow \vec{x} = \frac{1}{2}A^T\lambda$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \vec{b} - A\vec{x} = 0 \Rightarrow \vec{b} = A\vec{x}$$

$$\vec{b} = \frac{1}{2}AA^T\lambda \Rightarrow \lambda = 2(AA^T)^{-1}\vec{b}$$

$$\boxed{\vec{x} = A^T(AA^T)^{-1}\vec{b}}$$

2. Linear sequential estimation

Suppose we have two subsets

$\vec{\tilde{y}}_1 \in \mathbb{R}^{m_1 \times 1}$ $\quad \vec{\tilde{y}}_2 \in \mathbb{R}^{m_2 \times 1}$

$$\vec{\tilde{y}}_1 = H_1 \vec{x} + V_1 \quad \Rightarrow \quad \underbrace{\vec{\hat{x}}}_{\vec{\hat{x}}_1} = (H_1^T W_1 H_1)^{-1} H_1^T W_1 \vec{\tilde{y}}_1 \quad ①$$

$$\vec{\tilde{y}}_2 = H_2 \vec{x} + V_2$$

First, consider $\vec{\tilde{y}}_1$ and $\vec{\tilde{y}}_2$ simulaneously

$$\vec{\tilde{y}} = H \vec{x} + V$$

where $\vec{\tilde{y}} = \begin{bmatrix} \vec{\tilde{y}}_1 \\ \vec{\tilde{y}}_2 \end{bmatrix}$ $\quad H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}$ $\quad V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$ $\quad W = \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix}$

$$\underbrace{\vec{\hat{x}}}_{\vec{\hat{x}}_2} = (H^T W H)^{-1} H^T W \vec{\tilde{y}} \quad ②$$

$$= \underbrace{(H_1^T W_1 H_1 + H_2^T W_2 H_2]^{-1}}_{\star\ n \times n} (H_1^T W_1 \vec{\tilde{y}}_1 + H_2^T W_2 \vec{\tilde{y}}_2)$$

Let $\begin{cases} P_1 = [H_1^T W_1 H_1]^{-1} \\ P_2 = [H_1^T W_1 H_1 + H_2^T W_2 H_2]^{-1} \end{cases}$

then $P_2^{-1} = H_1^T W_1 H_1 + H_2^T W_2 H_2 = P_1^{-1} + H_2^T W_2 H_2 \quad ③$

we have $\vec{\hat{x}}_1 = (H_1^T W_1 H_1)^{-1} H_1^T W_1 \vec{\tilde{y}}_1 = P_1 H_1^T W_1 \vec{\tilde{y}}_1$

$$P_1^{-1} \vec{\hat{x}}_1 = H_1^T W_1 \vec{\tilde{y}}_1 \quad ④$$

$$\vec{\hat{x}}_2 = P_2 (H_1^T W_1 \vec{\tilde{y}}_1 + H_2^T W_2 \vec{\tilde{y}}_2)$$

From ③, we have $P_1^{-1} = P_2^{-1} - H_2^T W_2 H_2 \quad ⑤$

substitute ⑤ into ④

$$(P_2^{-1} - H_2^T W_2 H_2) \vec{\hat{x}}_1 = H_1^T W_1 \vec{\tilde{y}}_1 \quad ⑥$$

$$\hat{\vec{x}}_2 = P_2 (H_1^T W_1 \tilde{\vec{y}}_1 + H_2^T W_2 \tilde{\vec{y}}_2)$$

$$= P_2 (P_2^{-1} - H_2^T W_2 H_2) \hat{\vec{x}}_1 + P_2 H_2^T W_2 \tilde{\vec{y}}_2$$

$$= P_2 P_2^{-1} \hat{\vec{x}}_1 - \underline{P_2 H_2^T W_2} H_2 \hat{\vec{x}}_1 + \underline{P_2 H_2^T W_2} \tilde{\vec{y}}_2$$

$$= \hat{\vec{x}}_1 + P_2 H_2^T W_2 (\tilde{\vec{y}}_2 - H_2 \hat{\vec{x}}_1)$$

$$= \hat{\vec{x}}_1 + K (\tilde{\vec{y}}_2 - H_2 \hat{\vec{x}}_1)$$

K: Kalman gain matrix

$$K = P_2 H_2^T W_2$$

$$\hat{\vec{x}}_{k+1} = \hat{\vec{x}}_k + K_{k+1} (\tilde{\vec{y}}_{k+1} - H_{k+1} \hat{\vec{x}}_k)$$

Information matrix recursion

$$K_{k+1} = P_{k+1} H_{k+1}^T W_{k+1}$$

$$P_{k+1}^{-1} = \underbrace{P_k^{-1} + H_{k+1}^T W_{k+1} H_{k+1}}_{n \times n} \quad ☆ \quad 7$$

read textbook P21 - P23

$$P_{k+1} = P_k - P_k H_{k+1}^T \underbrace{(\overbrace{H_{k+1} P_k H_{k+1}^T + W_{k+1}^{-1}}^{m_{k+1}})^{-1}}_{\text{can be less than } n} H_{k+1} P_k$$

$$K_{k+1} = P_k H_{k+1}^T [I - (H_{k+1} P_k H_{k+1}^T + W_{k+1}^{-1})^{-1} H_{k+1} P_k H_{k+1}^T] W_{k+1}$$

Let $Q = (H_{k+1} P_k H_{k+1}^T + W_{k+1}^{-1})^{-1}$

$$K_{k+1} = P_k H_{k+1}^T [Q Q^{-1} - Q H_{k+1} P_k H_{k+1}^T] W_{k+1}$$

$$= P_k H_{k+1}^T Q [H_{k+1} \cancel{P_k H_{k+1}^T} + W_{k+1}^{-1} - H_{k+1} \cancel{P_k H_{k+1}^T}] W_{k+1}$$

$$= P_k H_{k+1}^T Q$$

$$= P_k H_{k+1}^T (H_{k+1} P_k H_{k+1}^T + W_{k+1}^{-1})^{-1}$$

Covariance recursion form

$$\hat{\vec{x}}_{k+1} = \hat{\vec{x}}_k + K_{k+1} (\tilde{\vec{y}}_{k+1} - H_{k+1} \hat{\vec{x}}_k)$$

$$K_{k+1} = P_k H_{k+1}^T (H_{k+1} P_k H_{k+1}^T + W_{k+1}^{-1})^{-1}$$

$$P_{k+1} = [I - K_{k+1} H_{k+1}] P_k$$

3. Nonlinear least squares approximation

previously, proposed math model $y(t) = \sum_{i=1}^{n} x_i h_i(t) = Hx$

Now let's consider the following model

$\tilde{\vec{y}} = f(\underset{\text{true}}{\vec{x}}) + \vec{v}_{\text{measurement error}}$ y is a nonlinear function

$\hat{\vec{y}} = f(\hat{\vec{x}})$

$\underline{\vec{e} = \tilde{\vec{y}} - \hat{\vec{y}} \equiv \Delta\vec{y}}$ $\tilde{\vec{y}} = \hat{\vec{y}} + \vec{e} = f(\hat{\vec{x}}) + \vec{e}$

min $\underset{\text{objective}}{J} = \frac{1}{2}\vec{e}^T W \vec{e} = \frac{1}{2}[\tilde{\vec{y}} - f(\hat{\vec{x}})]^T W [\tilde{\vec{y}} - f(\hat{\vec{x}})]$

usually hard to get a closed-form solution

start estimate x_c (guess)

$\hat{\vec{x}} = \vec{x}_c + \Delta x$

linearize $f(\hat{\vec{x}})$ about $\vec{x}_c$

$f(\hat{\vec{x}}) \approx f(x_c) + \frac{\partial f}{\partial \vec{x}_c}\Delta x = \underline{f(x_c) + H\Delta x}$

$\Delta y = \tilde{\vec{y}} - \hat{\vec{y}} \approx \tilde{\vec{y}} - f(\vec{x}_c) - H\Delta\vec{x}$

$= \Delta y_c - H\Delta\vec{x}$

$J = \frac{1}{2}\vec{e}^T W \vec{e} = \frac{1}{2}[\underline{\tilde{\vec{y}} - f(\hat{\vec{x}})}]^T W [\tilde{\vec{y}} - f(\hat{\vec{x}})]$

$J_p \approx \frac{1}{2}(\Delta y_c - H\Delta\vec{x})^T W (\Delta y_c - H\Delta x)$

$\Delta x = (H^T W H)^{-1} H^T W \Delta y_c$

$\hat{\vec{x}} = \underline{x_c + \Delta x}$

Model $f(x)$

$\downarrow$

compute $H = \dfrac{\partial f}{\partial x_c}$ $\quad \dfrac{\partial f}{\partial x}\Big|_{x_c}$

$\downarrow$

$\Delta y_c = \vec{\tilde{y}} - f(\vec{x}_c)$

$J_p = \frac{1}{2} \Delta y_c^T W \Delta y_c$

$\downarrow$

$\Delta x = (H^T W H)^{-1} H^T W \Delta y_c$

$\downarrow$

test stop condition $\xrightarrow{\text{Yes}}$ stop

$\downarrow$ No

$x_c\,(\text{new}) = x_c + \Delta x$

stop condition: $\quad \delta J = \dfrac{|J_i - J_{i-1}|}{J_i} \leq \dfrac{\varepsilon}{\|W\|}$

4. A brief probability review (Appendix C)

* probability is a measure of uncertainty, takes value between 0 and 1

* random variable (wiki)

  Prof. Lorig (canvas)

  A random variable X is a measurable function $X: \Omega \to E$ from a set of possible outcomes Ω to a measurable space E ($\mathbb{R}$)

  a) discrete random variable example

  coin toss (a fair)

  $\Omega = \{heads, tails\}$

  $$X(w) = \begin{cases} 1 & \text{if } w = heads \\ 0 & \text{if } w = tails \end{cases}$$

  probability mass function

  $P_X(x) = P(X = x)$

  function   probability

  $P_X(1) = 0.5$

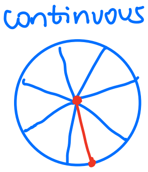
  $P_X(0) = 0.5$

  If X is distributed as a Bernoulli random variable with a parameter q, $X \sim Ber(q)$

  $$P_X(x) = \begin{cases} 1-q & x=0 \\ q & x=1 \end{cases}$$

  b) continuous random variable

  tire

  reference line : connect the valve stem to the center point.

  Let X be the angle between this line and the horizon

  X takes value from $[0, 360)$

  infinite many number

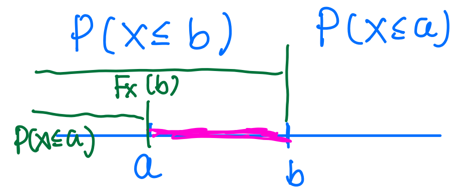
$P(X = 10°) = 0$

$P(X \in [10°, 20°]) = \frac{20° - 10°}{360°} = \frac{1}{36}$

Distribution function of a real-valued random variable is given as $F_X(x) = P(X \leq x)$

the probability that the random variable X takes on a value less than or equal to x

$$\underbrace{P(a < X < b)} = F_X(b) - F_X(a)$$



$$\underbrace{\lim_{x \to \infty} F_X(x) = 1}_{\text{covers all outcomes}} \qquad \underbrace{\lim_{x \to -\infty} F_X(x) = 0}_{\text{no event}}$$

probability density function

$$f(x) = \frac{dF(x)}{dx}$$

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

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

example : uniform distribution $X \sim U([a, b])$

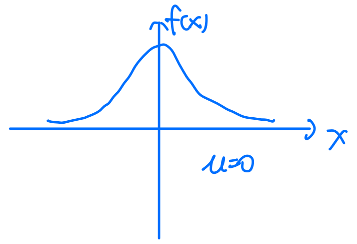
$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

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

example 2: Gaussian (Normal) distribution $X \sim N(\mu, \sigma^2)$

with mean $\mu \in \mathbb{R}$ and variance σ^2

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Moments.

1. First moment (expected value)

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

2. Second moment (variance)

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

Conditional probability

A measure of the probability of an event occurring given that another event has already occurred

$$\underset{\text{probability}}{P(A|B)} = \frac{P(A \cap B)}{P(B)} \qquad P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Bayes' theorem

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$

Example : a blood test to detect the presence of a banned drug.

* 99% sensitive.

  the test will produce 99% true positive results for the real drug user

* 99% specific ☆

  the test will produce 99% true negative results for non-drug user

Suppose 0.5% of people are user ☆

? : the probability that a randomly selected individual who tested positive is a user ?

$$\underset{A}{P(\text{user}} \mid \underset{B}{+)} = \frac{P(\overset{B}{+} \mid \overset{A}{\text{user}})\, P(\text{user})}{P(+)}$$

$$= \frac{0.99 \times 0.005}{\underset{0.99}{P(+\mid \text{user})}\,\underset{0.005}{P(\text{user})} + \underset{0.001}{P(+\mid \text{non-user})}\,\underset{0.995}{P(\text{non-user})}} ☆$$

$\approx 0.332 = 33\%$