Outline

0. Review of lecture 2 and questions

1. More about probability theory

2. Minimum variance estimation without a priori

3. Minimum variance estimation with a priori

4. Unbiased estimation

0. Review

* Overdetermined linear system of equations

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

  Underdetermined linear system of equations

  $\min \ \|x\|_2^2$

  $\text{s.t.} \quad Ax = b$

* Linear sequential estimation

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

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

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

* Nonlinear least squares approximation

* Probability review

  random variable

  Bayes' theorem

* Basis function (Read Section 1.5)

HW 2

1.17 Nonlinear least squares approximation

Note: 1. Generating data: $\vec{b}_j = A_j r_j + c + \varepsilon_j$, $\tilde{y} = \vec{b}_j^T \vec{b}_j - r_j^T r_j$

Nonlinear model: $\hat{\tilde{y}}_j = 2 \vec{b}_j^T \hat{c} - \hat{c}^T \hat{c}$

$= f(\hat{c})$

Stop condition: ε very small number

1.18 Compare linear & nonlinear

P3 Application of Bayes' Theorem

P4 Sample space and distribution

P5 Probability density functions

P6 Extra credit

1. More about probability review

1.1 Moments

The k-th moment of a real-valued continuous function $f_X(x)$ about a value c is

$$\mu_k = \int_{-\infty}^{\infty} (x-c)^k f(x)\,dx$$

$\boxed{\text{Expected value of } X}$ (first moment about 0)

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

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx \quad \text{expected value of } g(X)$$

Properties: $E[X+Y] = E[X] + E[Y]$

$E[aX] = aE[X]$

$\boxed{\text{Variance of } X}$ (second moment about μ)

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

σ is standard deviation

$$\begin{aligned} E[(X-E[X])^2] &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - E[2XE[X]] + E[E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Covariance for two jointly distributed random variable X, Y

$$cov(X,Y) = E[(X-E[X])(Y-E[Y])]$$

$$cov(X,X) = Var(X)$$

1.2 Marginal distribution

Joint distribution

a) discrete random variables X and Y

$$p_{X,Y} = P(X=x, Y=y)$$

↑ probability mass function

$$p_{X_1, X_2, \cdots X_n} = P(X_1 = x_1, X_2 = x_2, \cdots X_n = x_n)$$

If X and Y are independent

$$P(X=x, Y=y) = P(X=x)P(Y=y)$$

b) Continuous case

$$F_{X,Y}(x,y) = P(X \le x, Y \le y) \quad \text{distribution function}$$

$$F_{X_1, X_2, \cdots X_n} = P(X_1 \le x_1, X_2 \le x_2, \cdots X_n \le x_n)$$

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \, \partial y}$$

probability density function

$$f_X(x) = \frac{dF_X(x)}{dx}$$

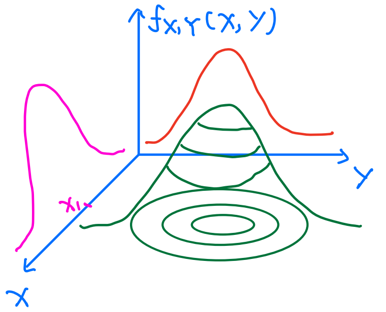
$$\int_x \int_y f_{X,Y}(x,y)\, dy\, dx = 1$$

$$\int_{x_1} \int_{x_2} \cdots \int_{x_n} f_{X_1, X_2, \dots X_n}(x_1, x_2, \cdots x_n)\, dx_n \cdots dx_1 = 1$$

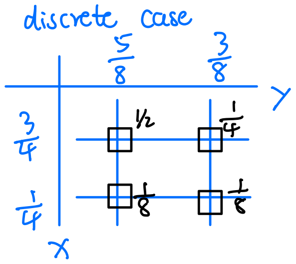
Marginal distribution

If the joint probability density function of random variable X and Y is $f_{X,Y}(x,y)$, the marginally probability density function of X is given

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$$



discrete case



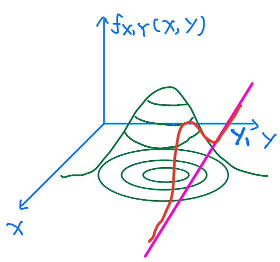
X $\Omega = \{1, 2\}$

Y $\Omega = \{3, 4\}$

$$P_X(X=1) = \sum_{j=1}^{n} P_{X,Y}(x=1, y_j)$$

$$P_Y(Y=3) = \sum_{i=1}^{m} P_{X,Y}(x_i, Y=1)$$

Conditional probability



$$f(x \mid Y = y_1) = \frac{f(x, c)}{\int_{-\infty}^{\infty} f(x, c)\, dx}$$

$$f_{X|Y} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Independent case $f_{X|Y} = f_X(x)$

$$\int_x f_{X|Y}\, dx = 1$$

2. Minimum variance estimation without a priori

$y = Hx$ proposed model $m \geq n$

$\tilde{\vec{y}} = H\vec{x} + \vec{v}$ $\hat{\vec{y}} = H\hat{\vec{x}}$

In Least squares approximation residual error

$J = \min \ (\tilde{\vec{y}} - \hat{\vec{y}})^T W (\tilde{\vec{y}} - \hat{\vec{y}})$

Now we estimate x as a linear function of $\tilde{\vec{y}}$

$\hat{\vec{x}} = M\tilde{\vec{y}} + \vec{n}$ we don't know M and $\vec{n}$ yet

The minimum variance of "optimum" M and $\vec{n}$ is that the variance of all n estimates $\hat{x}_i$ from their respective true value is minimized

$J_i = \frac{1}{2} E\{(\hat{x}_i - x_i)^2\}$ ← minimize for $\forall x$

$\hat{x}_i$ is a random variable

Consider the special case of perfect measurements $(v=0)$

$\tilde{\vec{y}} \equiv H\vec{x}$

This perfect measurement should result $\underline{\hat{\vec{x}} = \vec{x}}$

$\hat{\vec{x}} = M\tilde{\vec{y}} + \vec{n} = MH\vec{x} + \vec{n}$

For the perfect measurement case

$\hat{\vec{x}} = \vec{x} = MH\vec{x}$

$\vec{n}$ needs to be zero.

and $MH = I$

★ $\boxed{\hat{\vec{x}} = M\tilde{\vec{y}}}$ s.t $MH = I \in \mathbb{R}^{n\times n}$

$\hat{\vec{x}} \in \mathbb{R}^{n\times 1}$ $H \in \mathbb{R}^{m\times n}$ $M \in \mathbb{R}^{n\times m}$

$$M = \begin{bmatrix} - M_1 - \\ - M_2 - \\ \vdots \\ - M_n - \end{bmatrix} \quad I = \begin{bmatrix} - I_1 - \\ - I_2 - \\ \vdots \\ - I_n - \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & & 0 \\ 0 & \cdots & & & 0 & 1 \end{bmatrix}$$

$M_i H = I_i$

$H^T M_i^T = I_i^T$

$\hat{\vec{x}} = M\tilde{\vec{y}}$

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{bmatrix} = \begin{bmatrix} - M_1 - \\ - M_2 - \\ \vdots \\ - M_n - \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$(m \geq n)$

m: number of measurement

n: number of unknown parameters

$\hat{x}_i = M_i \tilde{\vec{y}}$; $\quad \tilde{\vec{y}} = H\vec{x} + \vec{v}$

$M_i H \vec{x} = I_i \vec{x}$
$= x_i$

$$J_i = \frac{1}{2} E\{ (\hat{x}_i - x_i)^2 \}$$
$$= \frac{1}{2} E\{ (M_i \tilde{\vec{y}} - x_i)^2 \}$$
$$= \frac{1}{2} E\{ [M_i (H\vec{x} + \vec{v}) - x_i]^2 \}$$
$$= \frac{1}{2} E\{ (M_i H \vec{x} + M_i \vec{v} - x_i)^2 \}$$
$$= \frac{1}{2} E\{ (x_i + M_i \vec{v} - x_i)^2 \}$$
$$= \frac{1}{2} E\{ (M_i \vec{v})^2 \}$$
$$= \frac{1}{2} E\{ M_i V V^T M_i^T \}$$
$$= \frac{1}{2} M_i E\{ V V^T \} M_i^T$$

Define the covariance matrix of measurement errors

$cov(V) = R = E\{VV^T\}$ known

$J_i = \min \frac{1}{2} M_i R M_i^T$

s.t. $H^T M_i^T = I_i^T$ $\quad$ M_i is unknown

Lagrangian $J_i = \frac{1}{2} M_i R M_i^T + \underset{\lambda_i}{\lambda_i^T} (I_i^T - H^T M_i^T)$

Lagrangian multiplier

The necessary condition

$\nabla_{M_i^T} J = R M_i^T - H\lambda_i = 0 \Rightarrow M_i = \lambda_i^T H^T R^{-1}$

$M_i^T = R^{-1} H \lambda_i$

$\nabla_{\lambda_i} J = I_i^T - H^T M_i^T = 0$

$$I_i^T - H^T R^{-1} H \lambda_i = 0$$

$$\Rightarrow \lambda_i^T = I_i^T (H^T R^{-1} H)^{-1}$$

plug λ_i^T into $M_i = \lambda_i^T H^T R^{-1}$

$$M_i = I_i^T (H^T R^{-1} H)^{-1} H^T R^{-1}$$

$$M = (H^T R^{-1} H)^{-1} H^T R^{-1}$$

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

Minimum variance estimation

Unbiased

$$W = R^{-1}$$

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

Least squares estimation

3. Minimum variance estimation with a priori

Assume the observation model

$\tilde{\vec{y}} = H\vec{x} + \vec{v}$ given $R = E[vv^T]$

$\vec{v}$ is a random variable.

$\hat{\vec{x}}$ is a random variable

A priori state estimate are given as the sum of the true state $\vec{x}$ and the errors in the a priori estimate w $(E[w]=0)$

$\hat{\vec{x}}_a = \vec{x} + w$ $\quad \hat{\vec{x}}_a$ is known $\quad Q = E[ww^T]$ known

We desire to estimate $\vec{x}$ as a linear combination of $\tilde{\vec{y}}$ and $\hat{\vec{x}}_a$

$\hat{\vec{x}} = M\tilde{\vec{y}} + N\hat{\vec{x}}_a + \vec{n}$

The minimum variance definition of "optimum"

$J_i = \frac{1}{2} E\{(\hat{x}_i - x_i)^2\}$

compact form: $J = \frac{1}{2} Tr[E\{(\hat{\vec{x}} - \vec{x})(\hat{\vec{x}} - \vec{x})^T\}]$

For the perfect measurement $(\vec{v}=0)$

$\tilde{\vec{y}} = Hx$

for the perfect a priori state estimate $(\vec{w}=0)$

$\hat{\vec{x}}_a = \vec{x}$

$\hat{\vec{x}} = M\tilde{\vec{y}} + N\hat{\vec{x}}_a + \vec{n}$

$= \vec{x} = MH\vec{x} + N\vec{x} + \vec{n}$

$= (MH + N)\vec{x} + \vec{n}$

$\Rightarrow \vec{n} = 0$

$MH + N = I$

Now the desired estimator has the form

$\hat{\vec{x}} = M\tilde{\vec{y}} + N\hat{\vec{x}}_a$

$\frac{1}{2}Tr[E\{(\hat{\vec{x}}-\vec{x})(\hat{\vec{x}}-\vec{x})^T\}]$

$=\frac{1}{2}Tr[E\{(M\tilde{\vec{y}}+N\hat{\vec{x}}_a-\vec{x})(M\tilde{\vec{y}}+N\hat{\vec{x}}_a-\vec{x})^T\}]$

$=\frac{1}{2}Tr[E\{(MH\vec{x}+M\vec{v}+N\vec{x}+N\vec{w}-\vec{x})(*)^T\}]$

$=\frac{1}{2}Tr[E\{(M\vec{v}+N\vec{w})(M\vec{v}+N\vec{w})^T\}]$

Assumption: $\vec{v}$ and $\vec{w}$ are uncorrolated,

$E[\vec{v}w^T]=E[\vec{w}\vec{v}^T]=0$

$J=\frac{1}{2}Tr(MRM^T+NQN')$

Optimization problem

$J=\min\frac{1}{2}Tr[E\{(\hat{\vec{x}}-\vec{x})(\hat{\vec{x}}-\vec{x})^T\}]$

s.t. $MH+N=I$

Lagrangian

$J=\frac{1}{2}Tr(MRM^T+NQN^T)+Tr[\Lambda(I-MH-N)]$

Necessary conditions

$$\left.\begin{aligned}\nabla_M J &= MR-\Lambda^T H^T=0\\ \nabla_N J &= NQ-\Lambda^T=0\\ \nabla_\Lambda J &= I-MH-N=0\end{aligned}\right\}$$

$\Lambda^T=(H^TR^{-1}H+Q^{-1})^{-1}$

$M=(H^TR^{-1}H+Q^{-1})^{-1}H^TR^{-1}$

$N=(H^TR^{-1}H+Q^{-1})^{-1}Q^{-1}$

$\hat{\vec{x}}=M\tilde{\vec{y}}+N\hat{\vec{x}}_a$

$=(H^TR^{-1}H+Q^{-1})^{-1}(H^TR^{-1}\tilde{\vec{y}}+Q^{-1}\hat{\vec{x}}_a)$

$R\to\infty$ $\qquad$ $\hat{\vec{x}}=\hat{\vec{x}}_a$ $\qquad$ $\tilde{\vec{y}}=H\vec{x}+\vec{v}$

$R^{-1}\to 0$ $\qquad$ $R=E[\vec{v}\vec{v}^T]$

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

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

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

☆ estimation from the first set of measurements $\tilde{\vec{y}}_1$

$H_1^T W_1 H_1 = Q^{-1}$ $\quad H_1 = I$

$\hat{\vec{x}}_a = Q H_1^T W_1 \tilde{\vec{y}}_1 \quad \Rightarrow \quad H_1^T W_1 \tilde{\vec{y}}_1 = Q^{-1} \hat{\vec{x}}_a$

$\hat{\vec{x}}$ sequential

$$= (\underline{H_1^T W_1 H_1} + H_2^T W_2 H_2]^{-1} (\underline{H_1^T W_1 \tilde{\vec{y}}_1} + H_2^T W_2 \tilde{\vec{y}}_2)$$

$$= [Q^{-1} + H_2^T W_2 H_2]^{-1} (Q^{-1} \hat{\vec{x}}_a + H_2^T W_2 \tilde{\vec{y}}_2)$$

$$(H^T R^{-1} H + Q^{-1})^{-1} \; (H^T R^{-1} \tilde{\vec{y}} + Q^{-1} \hat{\vec{x}}_a)$$

4. Unbiased estimator

An estimator is said to be "unbiased estimator" of x

if $E\{\hat{\vec{x}}(\tilde{\vec{y}})\} = \vec{x}$

Why MVE is unbiased ?

proof $E\{\hat{\vec{x}}(\tilde{\vec{y}})\} = E\{M \tilde{\vec{y}}\}$

$\hat{\vec{x}}$ is a function of $\tilde{\vec{y}}$

$$= E\{M(H\vec{x} + V)\}$$

$$= E\{\underbrace{MH}_{=I}\vec{x}\} + \underbrace{E\{\cancel{M\vec{V}}\}}_{\to 0} \quad \text{zero-mean}$$

$$= E\{\vec{x}\} = \vec{x}$$

$$E\{\hat{\vec{x}}(\tilde{\vec{y}})\} = \vec{x}$$