SDM366 Optimal Control and Estimation

Lecture Note 3
Least Squares and Basic System Identification

- ChatGPT :
- data driven

Analysis

In-depth understanding the Least squares paradigm plays fundamental role in control-Learning

Prof. Wei Zhang
Southern University of Science and Technology

zhangw3@sustech.edu.cn
https://www.wzhanglab.site/

Outline

- **Least-Squares Problem Formulation**
- Solution to Linear Least-Squares Problems
- Linear Least-Squares Examples
- Applications to System ID
- Nonlinear Least Squares

- Last lecture: obtain discrete-time linear state space model from
  - physical process
  - given continuous time state space model
  - given nonlinear state space model

- The goal of this lecture note:
  - learn how to build model based on observed input-output data pairs
    - General case beyond the scope of this course
    - Focus on special case, where first obtain transfer function model from input-output data pairs, and then obtain the corresponding state space model

  - Main method: **Least Squares**

Least-Squares Problem Formulation:

- Measurement Equation:

measurement

$$y = g(\theta) + v$$

hidden parameter

  - $y \in R^m$: measurements data
  - $\theta \in \Theta \subseteq R^n$: parameter to be estimated, where Θ is the constraint set for feasible parameters
  - $v \in R^m$: unknown measurement noise

e.g. $g(\theta) = NN(\theta)$

  - $g: R^n \to R^m$: known (possibly) nonlinear function relates θ with measurement y

$y = 2\theta + v$

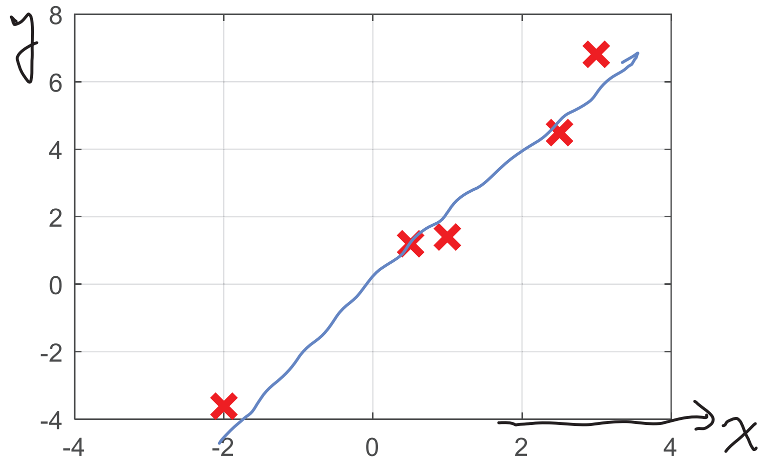
$g(\theta) = 2\theta$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} \sin\theta_1 \\ e^{\theta_2} \\ \theta_1 - \theta_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$g(\theta)$

▪ Least Squares Example:

i	1	2	3	4	5
x	1	0.5	−2	3	2.5
y	1.4	1.2	−3.6	6.8	4.5



- Line fitting: $\underbrace{y = ax+b+v}$, we need to estimate $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$ to "fit" data

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_5 \end{bmatrix}}_{y} = \begin{bmatrix} ax_1+b \\ ax_2+b \\ \vdots \\ ax_5+b \end{bmatrix} + v \Rightarrow y = \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_5 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\theta}}_{g(\theta) = H \cdot \theta} + v$$

$$y_i = ax_i^2 + bx_i + c = \underbrace{[\, x_i^2 \quad x_i \quad 1 \,]} \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{\theta}$$

Least-Squares Problem Formulation:

- **Problem Statement**: Find the best parameter in the constraint set Θ that minimizes the difference between the model and the measured data

$$\min_{\theta\in\Theta} J(\theta) = \min_{\theta\in\Theta} ||y - g(\theta)||^2 \quad \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} - \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_m \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \qquad g(\theta) = \begin{bmatrix} g_1(\theta) \\ g_2(\theta) \\ \vdots \\ g_m(\theta) \end{bmatrix} \triangleq \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_m \end{bmatrix}$$

$$= \min_{\theta\in\Theta} \left((y - g(\theta))^T (y - g(\theta)) \right)$$

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \qquad = \min_{\theta\in\Theta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$$

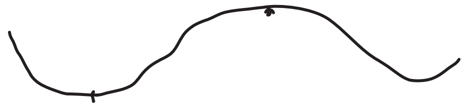
$g_i(\theta)$

- **Linear Least Squares**: $g(\theta) = H\theta$, where $H \in R^{m\times n}$ is a given deterministic matrix

Outline

- Least-Squares Problem Formulation
- **Solution to Linear Least-Squares Problems**
- Linear Least-Squares Examples
- Applications to System ID
- Nonlinear Least Squares

Optimization of multivariable function



- 1^{st} –order necessary condition for optimality of $J(\theta)$

e.g. 1-dim $\theta \in \mathbb{R}$: $J(\theta) = J(\theta_0) + \left(\frac{\partial J}{\partial \theta}\right)\Big|_{\theta=\theta_0} (\theta - \theta_0) + H.O.T.$

If θ_0 is minimizers $\Rightarrow \frac{\partial J}{\partial \theta}\Big|_{\theta=\theta_0} = 0$

because if ① $\frac{\partial J}{\partial \theta}\Big|_{\theta=\theta_0} > 0$, $\hat{\theta} = \theta_0 - \varepsilon \Rightarrow J(\hat{\theta}) < J(\theta_0)$

② $\dots < 0$, $\hat{\theta} = \theta_0 + \varepsilon \Rightarrow \dots\dots$

- Matrix calculus:
  - If $f: R^n \to R^m$, then $\frac{\partial f}{\partial x}(x) = Df(x) \triangleq \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & & \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in$ m×n matrix

$$f(x) = \begin{bmatrix} f_1(x_1, x_2 \cdots x_n) \\ f_2(x_1, \cdots x_n) \\ \vdots \\ f_m(x_1, \cdots x_n) \end{bmatrix} \qquad = \left[\frac{\partial f_i}{\partial x_j}\right]_{i,j}$$

e.g. $f(x) = \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \\ f_3(x_1, x_2) \end{bmatrix}$, $f: \mathbb{R}^2 \to \mathbb{R}^3$; $df = \frac{\partial f}{\partial x} \cdot dx = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \vdots & \vdots \\ \frac{\partial f_3}{\partial x_1} & \frac{\partial f_3}{\partial x_2} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}$

Optimization of multivariable function

R^1

- **Gradient**: For scaler valued multivariate function $f: R^n \to R$, its gradient is defined as:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{n\times 1}$$

$$\frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1} \ \cdots \ \frac{\partial f}{\partial x_n}\right]$$

$$\nabla f(x) = \left(\frac{\partial f}{\partial x}\right)^T$$

$$df = (\nabla f(x))^T \cdot dx = \langle \nabla f(x), dx \rangle$$

- For $f: R^n \to R$, notational convention $\nabla f(x) = \left(\frac{\partial f}{\partial x}(x)\right)^T$
- Some references use $\frac{\partial f}{\partial x}$ to denote gradient

$e.g.\ d = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

- Directional derivative: $Df(x; d) = \lim_{\alpha \to 0} \frac{f(x+\alpha d) - f(x)}{\alpha} = g'(0)$

where $g(\alpha) \triangleq f(x+\alpha d)$

$\Downarrow$ calculus of variation

$$\frac{\partial f}{\partial x_1} = Df\left(x; \begin{bmatrix} 1 \\ \vdots \end{bmatrix}\right)$$

- Some calculus examples:
  - $f(x) = Ax$

assume: $f: \mathbb{R}^n \to \mathbb{R}^m$

$A \in \mathbb{R}^{m\times n}$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$\frac{\partial f}{\partial x} \triangleq \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \ddots & \\ \vdots & & \ddots \end{bmatrix} = A$$

$$f_1(x) = a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n$$

$$\frac{\partial f_1}{\partial x_1} = a_{11}, \quad \frac{\partial f_1}{\partial x_2} = a_{12}$$

  - $f(x) = x^T A\, x$

assume: $A \in \mathbb{R}^{n\times n}$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $\Rightarrow f: \mathbb{R}^n \to \mathbb{R}$

Question: $\frac{\partial f}{\partial x} = \left[\frac{\partial f}{\partial x_1} \;\; \frac{\partial f}{\partial x_2} \;\; \cdots \;\; \frac{\partial f}{\partial x_n} \right]$

By definition:

$$f(x) = \sum_i \sum_j a_{ij} x_i x_j$$

pick k

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k}\left[\cancel{\sum_{i\neq k}\sum_{j\neq k} a_{ij} x_i x_j} + \sum_{i\neq k} a_{ik} x_i x_k + \sum_{j\neq k} a_{kj} x_k x_j + a_{kk} x_k x_k \right]$$

$$= \sum_{i=1}^{n} a_{ik} x_i + \sum_{j=1}^{n} a_{kj} x_j$$

$$\left(\frac{\partial f}{\partial x}\right)^T = \nabla f(x) = A^T x + Ax$$

If $A = A^T$, i.e. A is symmetric, $\nabla f(x) = 2Ax$

  - **Exercise**: compute $\frac{\partial f}{\partial x}(x)$, where $x \in R^n$, and $f(x) = x^T x \cdot x$

- Derivation of linear least square solutions
  - Normal equation: L.S. $\min_{\theta \in \mathbb{R}^n} J(\theta) = \min_{\theta \in \mathbb{R}^n} \| y - H\theta \|^2$

$H \in \mathbb{R}^{m \times n}$

$J: \mathbb{R}^n \to \mathbb{R}$, $\quad J(\theta) = (y - H\theta)^T (y - H\theta) = y^T y - y^T H\theta - \theta^T H^T y + \theta^T H^T H \theta$

(Qustion: $\theta^T H^T y \,?\!= y^T H \theta$, similar to $\begin{bmatrix} 3 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix}$

$= \theta^T H^T H \theta - 2 y^T H \theta + y^T y$

Note:
$H^T H$: symmetric matrix
$(H^T H)^T = H^T H$

$\nabla J(\theta) = 2 H^T H \theta - 2 H^T y = 0 \leftarrow$ vector $\in \mathbb{R}^n$

$\Rightarrow (H^T H)\theta = H^T y \Leftarrow$ normal equation

- Solution with full rank H : $H \in \mathbb{R}^{m \times n}$; $m > n$

- Normal equation: $H^T H \theta = H^T y$

① If H is full rank $(\text{rank}(H) = n) \xRightarrow{\text{V.F.Y.}} \underbrace{H^T H}_{n \times n}$ is nonsingular

$$\Rightarrow \theta_{LS} = (H^T H)^{-1} H^T y$$

② If H is not full rank $\Rightarrow H^T H$ singular

e.g. $H^T H = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, $H^T y = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$, $\hat{\theta}_{LS} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -2 \\ 1 \end{bmatrix}$

infinitely many solution

$J(\theta) = \|y - H\theta\|^2$

- Geometric interpretation of linear least squares

· For any $\theta \in \mathbb{R}^n$, $H\theta$ is a linear combination of columns of H

① · $y \in col(H)$, we can find θ^* such that $y = H\theta^* \Rightarrow J(\theta^*) = 0$

· $\theta_{LS} = \underline{(H^T H)^{-1} H^T y}$, if $y \in col(H)$, $J(\theta_{LS}) = 0$

· i.e., we need show $\underline{y - H \cdot (H^T H)^{-1} H^T y} = 0$, for $y \in col(H)$

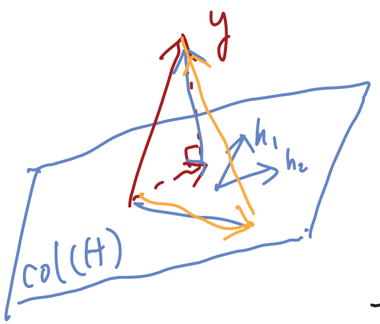
· pf: $y \in col(H) \Rightarrow y = H \cdot \beta$, for some β

$\Rightarrow \quad H\beta - H\cancel{(H^T H)^{-1} H^T H}\beta = 0$

② $y \notin col(H)$, no exact solution to $y = H\theta$., need to find minimum distance solution

Geometrically, e.g. $H = [h_1, h_2]$, L.S. tries to find the θ_{LS}

to minimize $\|y - H\theta\|$. Intuitively, $\underline{H\theta_{LS}}$ should be the projection



of y onto col(H)

let's verify: we need to show $(y - H\theta_{LS}) \perp col(H)$

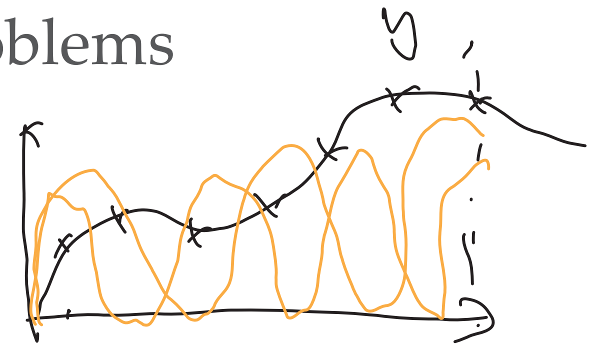
$\Leftrightarrow (y - H\theta_{LS})^T H\beta = 0, \ \forall \beta$

plug in : $\left[y - H\left((H^TH)^{-1}H^Ty\right)\right]^T \cdot H\beta$

$= y^TH\beta - y^TH\cancel{(H^TH)}^{-1}\cancel{H^TH}\beta = 0$

Outline

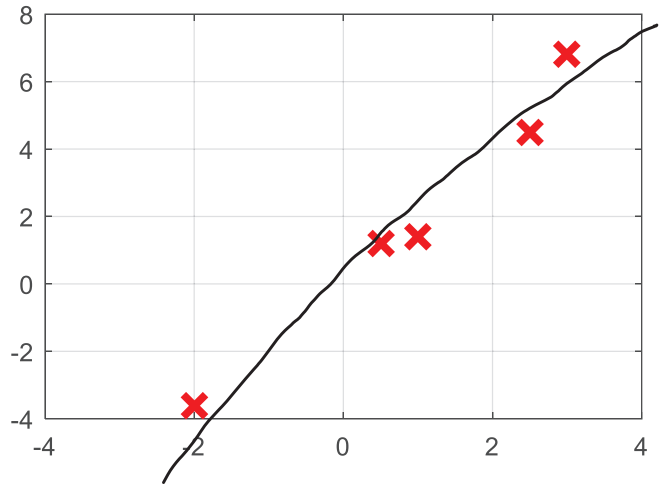
- Least-Squares Problem Formulation
- Solution to Linear Least-Squares Problems
- **Linear Least-Squares Examples**
- Applications to System ID
- Nonlinear Least Squares



$y \approx H\theta$

Linear Least Squares Example:

i	1	2	3	4	5
x	1	0.5	−2	3	2.5
y	1.4	1.2	−3.6	6.8	4.5



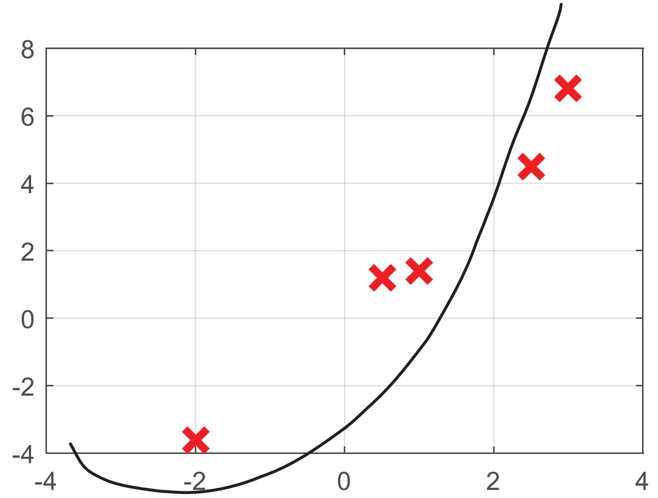
- Assume $y = \alpha x + \beta$, $\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_5 \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_5 & 1 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\theta} + v$$

$$\theta_{LS} = (H^T H)^{-1} H^T y = ?$$

$$y = \begin{bmatrix} 1.4 \\ 1.2 \\ -3.6 \\ 6.8 \\ 4.5 \end{bmatrix}, \quad H = \begin{bmatrix} 1 & 1 \\ 0.5 & 1 \\ -2 & 1 \\ 3 & 1 \\ 2.5 & 1 \end{bmatrix}$$

- Change hypothesis, assume $y \approx be^{ax}$



use the same data, find L.S. estimate of (a,b)

$\theta = \begin{bmatrix} a \\ b \end{bmatrix}$. method 1: $\min_\theta \left\| \begin{bmatrix} y_1 \\ \vdots \\ y_5 \end{bmatrix} - \begin{bmatrix} be^{ax_1} \\ be^{ax_2} \\ \vdots \\ be^{ax_5} \end{bmatrix} \right\|^2$

method 2: $y \approx be^{ax}$

take log $\log y \approx \log b + ax \Rightarrow$ $\underbrace{\begin{bmatrix} \log y_1 \\ \log y_2 \\ \vdots \\ \log y_5 \end{bmatrix}}_{\tilde{y}} \begin{matrix} \approx c + ax_1 \\ \approx c + ax_2 \\ \vdots \\ \approx c + ax_5 \end{matrix}$

c

$\tilde{y} \approx \underbrace{\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_5 & 1 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} a \\ c \end{bmatrix}}_{\tilde{\theta}}$

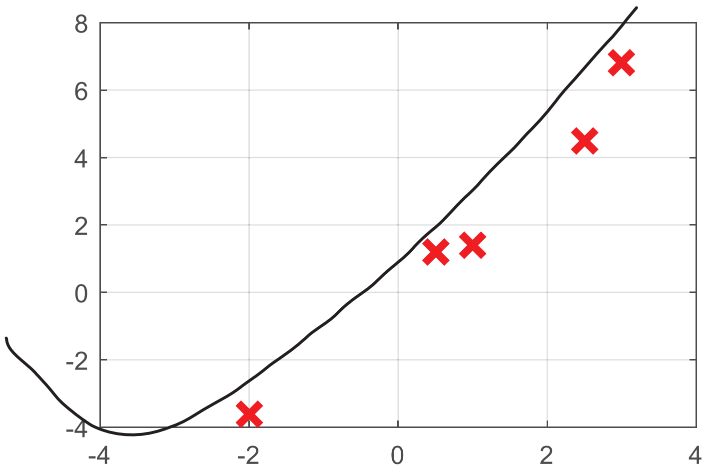
$\tilde{\theta}_{LS} = (H^T H)^{-1} H^T \tilde{y}$

$\begin{bmatrix} a_{LS} \\ c_{LS} \end{bmatrix}$

$b_{LS} = e^{c_{LS}}$

- Change hypothesis, assume that $y = \alpha_0 + \alpha_1 x + \alpha_2 x^2$

- same data find L.S. estimate of $\alpha_0, \alpha_1, \alpha_2$



$$\theta = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

$$y_1 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2$$

$$y_2 = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_2^2$$

$$\vdots$$

$$y_5 = \alpha_0 + \alpha_1 x_5 + \alpha_2 x_5^2$$

$$\Rightarrow \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_5 \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_5 & x_5^2 \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\theta}$$

$$\theta_{LS} = (H^T H)^{-1} H^T y$$

Outline

- Least-Squares Problem Formulation
- Solution to Linear Least-Squares Problems
- Linear Least-Squares Examples
- **Applications to System ID**
- Nonlinear Least Squares

$\|y - S(\theta)\|^2$

$\Downarrow$

$H\theta$

$H\hat{\theta}_{LS}$ is projection of y onto col(H)

$u \to \boxed{\text{sys}} \to y$ static $y = h(u)$; dynamic system $(y, u) \sim$ ODE

Application to System Identification for Linear Systems

- ARX(p, q) model :(Autoregressive with exogenous input)

$$\begin{cases} \underbrace{y(k) + \alpha_1 y(k-1) + \cdots + \alpha_p y(k-p)}_{\text{autoregressive terms}} \\ = \underbrace{\beta_0 u(k) + \beta_1 u(k-1) + \cdots + \beta_q u(k-q) + v(k)}_{\text{moving average}} \end{cases}$$

  - $v(k)$: noise signal

e.g. $y(k) + y(k-1) = 3u(k) - 2u(k-3) + v(k) \Rightarrow \text{ARX}(1,3)$

  - Model parameter: $\theta = [\alpha_1, \ldots, \alpha_p, \beta_0, \beta_1, \ldots, \beta_q]^T$

$y(k) + 2y(k-1) = 3u(k) - \beta u(k-2) + v(k) \Longrightarrow$ unknown parameter β

  - One-step predictor:

$$\hat{y}(k|\theta) = -\alpha_1 y(k-1) - \cdots - \alpha_p y(k-p)$$
$$+\beta_0 u(k) + \beta_1 u(k-1) + \cdots + \beta_q u(k-q)$$

Given parameter θ, we expect to see an output $\underbrace{\hat{y}(k|\theta)}_{S(\theta)}$

let $y(k)$ be the real measurement

we want to find θ such that $\|\hat{y}(k|\theta) - y(k)\|^2$ is small

$u \rightarrow \boxed{\text{sys}} \rightarrow y$

- System ID problem for ARX model:
  - Given data pairs $\underbrace{\{(u(k), y(k))\}}_{\text{data}}{}_{k \leq N}$, find the parameter vector θ that minimizes cost:
    - $J(\theta) = \sum_{k=1}^{N} ||\hat{y}(k|\theta) - y(k)||^2$

$$J(\theta) = \sum_{k=1}^{N} || \hat{y}(k|\theta) - y(k) ||^2$$

optimization variable

Our goal is to $\min_{\theta}$ $J(\theta)$ $\Rightarrow \hat{\theta}_{LS}$

- Formulate as least square problem:

given data set, $(u_1, y_1), (u_2, y_2), \dots, (u_m, y_m)$ # of data pairs

- "Regressor:" at time k

$$\hat{y}(k|\theta) = -\alpha_1 y(k-1) - \alpha_2 y(k-2) \cdots - \alpha_p y(k-p) + \beta_0 u(k) + \beta_1 u(k-1) + \cdots + \beta_q u(k-q)$$

$$= \underbrace{\left[-y(k-1) \;\; -y(k-2) \;\; \cdots \;\; -y(k-p) \;\; u(k) \;\; u(k-1) \;\cdots\; u(k-q)\right]}_{\text{is called "regressor"} \,\triangleq\, \phi^T(k)} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \\ \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{bmatrix}}_{\theta}$$

$$= \phi^T(k)\,\theta$$

Note: For $\phi(k)$ to be well defined, we need $k > \max\{p, q\}$

- Derivation continued

Let's denote $k_0 = \max\{p, q\} + 1$, $\quad y(k_0) = \phi^T(k_0)\theta + v(k)$

$$
\underbrace{\begin{bmatrix} y(k_0) \\ y(k_0+1) \\ \vdots \\ y(m) \end{bmatrix}}_{y}
\begin{matrix} = \phi^T(k_0)\theta + v(k_0) \\ = \phi^T(k_0+1)\theta + v(k_0) \\ \\ = \phi^T(m)\theta + v(m) \end{matrix}
$$

$$
y = \underbrace{\begin{bmatrix} \phi^T(k_0) \\ \phi^T(k_0+1) \\ \vdots \\ \phi^T(m) \end{bmatrix}}_{H} \theta + v
\qquad \Rightarrow \hat{\theta}_{LS} = (H^T H)^{-1} H^T y
$$

$Y(z) = H(z)U(z)$

zero-state response

- **System ID Example I:**

$U(z) \longrightarrow \boxed{G(z)} \xrightarrow{Y(z)}$, where

$U(z) \Leftrightarrow u(k)$

$Y(z) \Leftrightarrow y(k)$

$G(z) = \frac{(z^2+b)}{z^3+az}$, find best estimate for a, b, given data set $(u_1, y_1), (u_2, y_2), \dots, (u_{20}, y_{20})$

1° Find ARX model, $G(z) = \frac{z^{-1}+bz^{-3}}{1+az^{-2}}$. $Y(z) = G(z)H(z)$

$$\Rightarrow (1+az^{-2})Y(z) = (z^{-1}+bz^{-3})U(z)$$

Inverse z-transform

$$\Rightarrow y(k) + ay(k-2) = u(k-1) + bu(k-3)$$

Let $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$, ARX(2,3) , $k_0 = 3+1 = 4$

$$y(4) = -ay(2) + u(3) + bu(1) \not\Rightarrow [-y(2) \quad u(3) \quad u(1)] \begin{bmatrix} a \\ 1 \\ b \end{bmatrix}$$

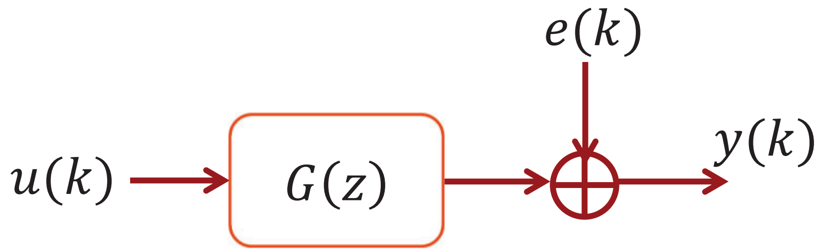
$$\Rightarrow \underbrace{y(4) - u(3)} = [-y(2) \quad u(1)] \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\Rightarrow \underbrace{\begin{bmatrix} y(4) - u(3) \\ y(5) - u(4) \\ \vdots \\ y(20) - u(19) \end{bmatrix}} = \underbrace{\begin{bmatrix} -y(2) & u(1) \\ -y(3) & u(2) \\ \vdots & \vdots \\ -y(18) & u(17) \end{bmatrix}} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\theta} + v$$

$$\hat{\theta}_{LS} = (H^T H)^{-1} H^T \tilde{y}$$

ŷ: H

- **System ID Example 2:**



- $G(z) = \frac{z-1}{z-a}$, where a is an unknown scalar
- Data: $u(1) = 1, u(2) = \frac{1}{2}, u(3) = 1, y(1) = 2, y(2) = 1, y(3) = 2$

V.F.I?

Outline

- Least-Squares Problem Formulation
- Solution to Linear Least-Squares Problems
- Linear Least-Squares Examples
- Applications to System ID
- **Nonlinear Least Squares**

Linear L.S.:

assumption 1: $\Theta \in \mathbb{R}^n$ constraint

2. $g(\theta) = H\theta$

$\Rightarrow$ Analytical solution

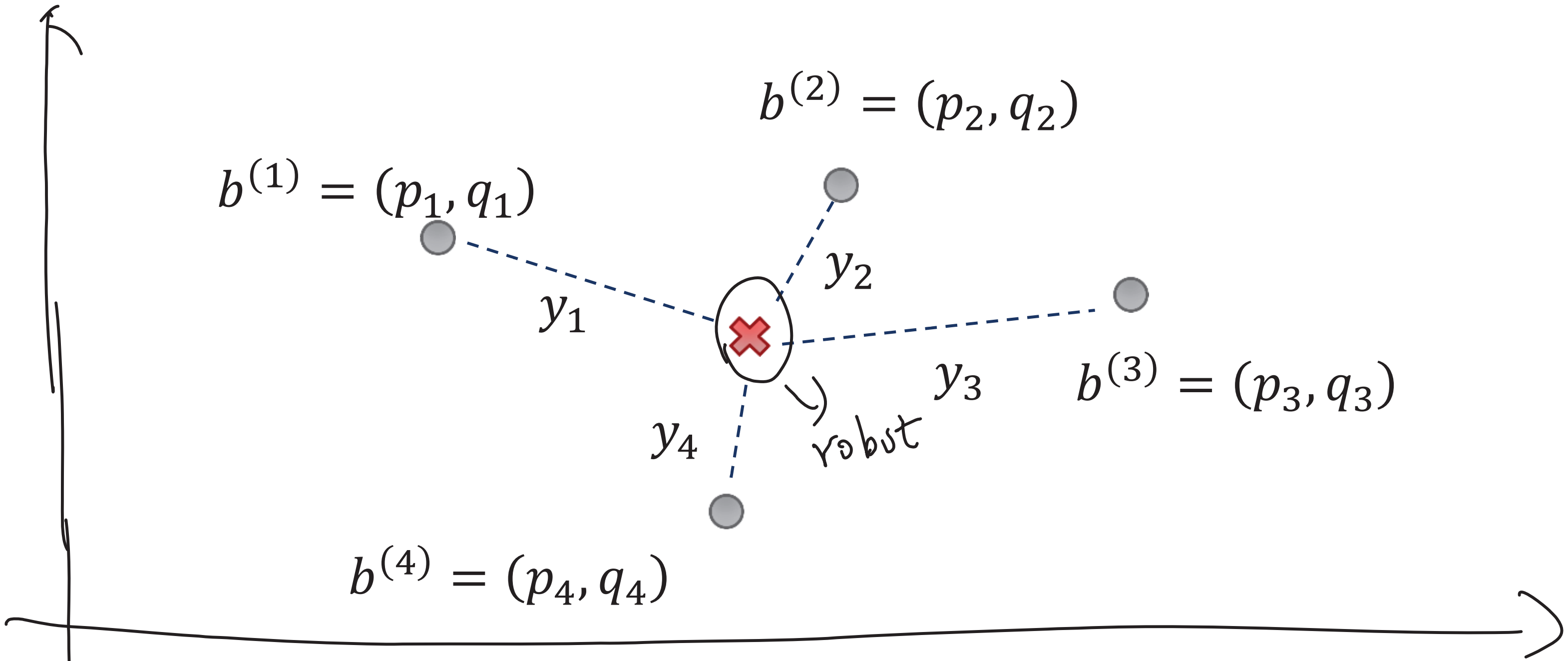
- Nonlinear Least Squares:

$$\min_{\theta\in\Theta} J(\theta) = \underbrace{\|y - g(\theta)\|^2}_{\min_{\theta\in\Theta}}$$

  - For general nonlinear function $g(\theta)$, analytical solution to the above optimization is not available
    "Neural Network"
  - Numerical optimization algorithms can be used to find the optimizer

$$\theta^* = \underline{argmin_{\theta\in\Theta} J(\theta)}$$

- Nonlinear Least Square Example: Navigation by range measurement:



- : beacons with known positions $b^{(i)} = (p_i, q_i)$ ← known and given
- : target with unknown position $\theta = (\theta_1, \theta_2)$
- y_i : known measured distance or range from beacon i:

  typical assumption: $y_i = \underbrace{\|b^{(i)} - \theta\|}_{g_i(\theta)} + v_i = \sqrt{(p_i-\theta_1)^2 + (q_i-\theta_2)^2} + v_i$

- Given measurements $y_1, y_2, \dots, y_m$, find the best target location θ

- We can choose cost function: $J(\theta) = \sum_{i=1}^{m} \Big(y_i - \overbrace{\left\|b^{(i)} - \theta\right\|}^{g_i(\theta),\ \hat{y}_i(\theta)}\Big)^2$

$$\Rightarrow \hat{\theta}_{LS} = \underset{\theta}{\operatorname{argmin}}\ J(\theta)$$

- Coding Example