Outline

0. Syllabus

1. What is estimation theory?

2. Least squares approximation — a curve fitting example

3. Linear least squares – how to estimate parameters

4. Least squares approximation — linear system of equations example

5. Weighted least squares

6. Constrained least squares

Legredre

Gauss

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

1. What is estimation ?

e.g.1: temperature of GUG 305

① $\tilde{y}_1 = 70F$

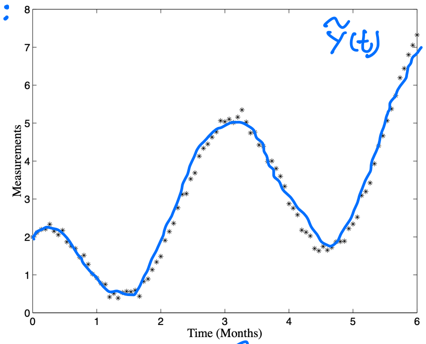
② $\tilde{y}_1 = 70F$ $\quad \tilde{y}_2 = 71$ $\quad \tilde{y}_3 = 68$ $\quad \tilde{y}_4 = 75$

taking average

③ add weight

? Can we know the true value of the temperature ?

e.g. 2:



Stock

? predict the price

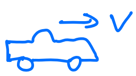
$\hat{y}(t)$

some important quantities

$\tilde{x}$ $\tilde{y}$ $\tilde{z}$: measured data

$\hat{x}$ $\hat{y}$ $\hat{z}$: estimated variable

x y z : true value

e.g.3 

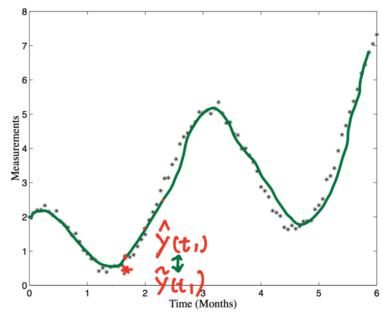
$\tilde{v}(t)$ measurements

dynamics $\dot{v} = \frac{dv}{dt} = a$

} Kalman filter

Prof. Mehran Mesbahi : Estimation is a process of finding the best interpretation / representation of the practical world based on all the information you gathered

2. A curve fitting example (Least $\boxed{\text{Squares}}$ Approximation)



measurements of a stock

$\boxed{\text{Goal}}$: develop a mathematical model in order to predict the price in 6 months

parameter estimation

Apply LS to obtain optimal parameter in a given mathematical model

Several quantities:

$\tilde{x}$: measurement value $\qquad x$: true value (never know)

$\hat{x}$: estimated value

Relations in these quantities:

$\tilde{x} = x + v \qquad v$: measurement error (stochastic properties)

$\tilde{x} = \hat{x} + e \qquad e$: residual error

minimize the difference between $\tilde{x}$ and $\hat{x}$ which is the residual error

Requirements: $\mu \le 0.0075$

$\sigma^2 \le 0.125$

sample mean $\quad \mu = \frac{1}{m} \sum_{i=1}^{m} \underbrace{[\tilde{y}(t_i) - \hat{y}(t_i)]}_{[e_i]}$

sample standard deviation

$$\sigma^2 = \frac{1}{m-1} \sum_{i=1}^{m} \underbrace{\{[\tilde{y}(t_i) - \hat{y}(t_i)] - \mu\}^2}_{\{e_i - \mu\}^2}$$

Model 1: $\hat{y}_1(t) = c_1 t + c_2 \sin t + c_3 \cos(2t)$

Model 2: $\hat{y}_2(t) = d_1(t+2) + d_2 t^2 + d_3 t^3$

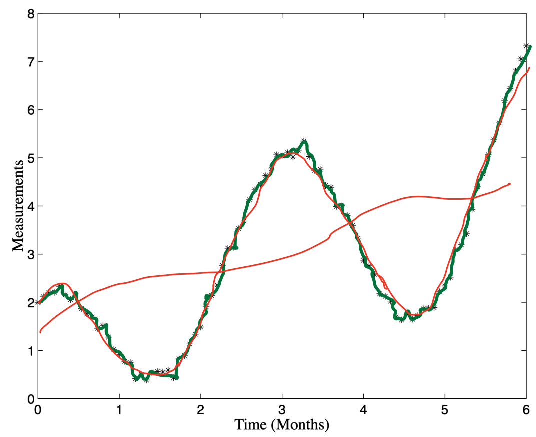
Apply LS to get $c_1 \sim c_3$ / $d_1 \sim d_3$

t, $\sin t$, $\cos(2t)$ / $(t+2)$, t^2, t^3 are basis functions

Principle of LS:

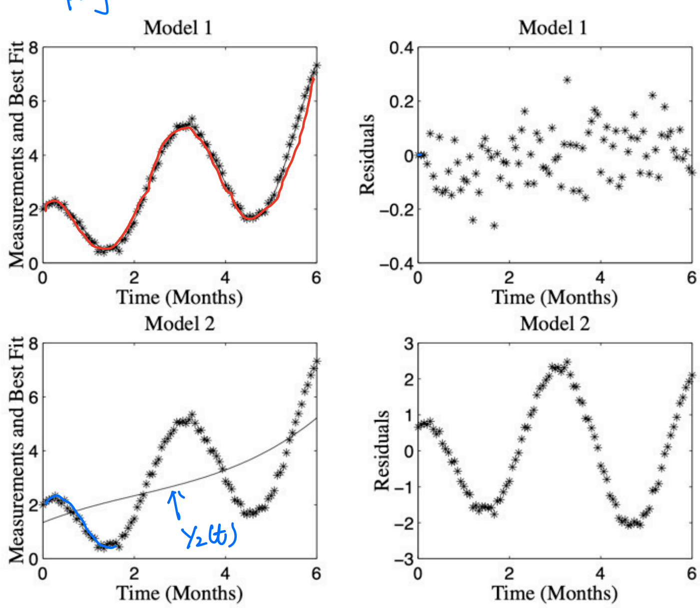
minimize $\sum_{i=1}^{m} \underbrace{\left[\tilde{y}(t_i) - \hat{y}(t_i)\right]}_{e_i \text{ residual error}}{}^2$

(m: total number of measurements)



$\hat{Y}_1(t) = 2t + 3\sin t + 5\cos(2t)$ red line

plug $t = 0, 0.1, \cdots 6$



Model 1: $\mu = 1 \times 10^{-5}$

$\sigma^2 = 0.0921$

Model 2: $\mu = 1 \times 10^{-5}$

$\sigma^2 = 1.3856$

3. Linear least squares approximation

Suppose we have a set of measured values, $\tilde{y}_j$

$\{\tilde{y}_1, t_1 ; \tilde{y}_2, t_2 ; \cdots ; \tilde{y}_m, t_m\}$

and a proposed mathematical model

$$y(t) = \sum_{i=1}^{n} x_i h_i(t) \qquad m \geq n$$

where m: total number of measurements (data points)

n: total number of unknown variables to be estimated

$h_i(t) \in \{h_1(t), h_2(t), \cdots h_n(t)\}$ basic functions

Goal: select the optimum x-values based upon a measure of "how well" the proposed model predicts the measurements

$$\tilde{y}_j \equiv \tilde{y}(t_j) = \sum_{i=1}^{n} x_i h_i(t_j) + v_j$$

$$\hat{y}_j \equiv \hat{y}(t_j) = \sum_{i=1}^{n} \hat{x}_i h_i(t_j)$$

j: time index

i: unknown varable index

e.g. $\hat{y}(t) = C_1 t + C_2 \sin t + C_3 \cos(2t)$

$C_1, C_2, C_3 \Rightarrow x_1, x_2, x_3$

$h_1(t) = t$

$h_2(t) = \sin t \quad h_3(t) = \cos(2t)$

$$\tilde{y}_j = \hat{y}_j + e_j$$

$$= \sum_{i=1}^{n} \hat{x}_i h_i(t_j) + e_j$$

$$\tilde{\vec{y}} = \begin{bmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_m \end{bmatrix} \qquad \hat{\vec{x}} = \begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix} \qquad \vec{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix}$$

$$H = \begin{bmatrix} h_1(t_1) & h_2(t_1) & \cdots & h_n(t_1) \\ h_1(t_2) & h_2(t_2) & \cdots & h_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(t_m) & h_2(t_m) & \cdots & h_n(t_m) \end{bmatrix} \qquad H \in \mathbb{R}^{m \times n}$$

$$\tilde{\vec{y}} = H\hat{\vec{x}} + \vec{e}$$ compact form ($\vec{e}$: residual errors)

$$\vec{e} = \tilde{\vec{y}} - H\hat{\vec{x}}$$

principles of LS: select $\hat{x}$ that minimize the sum square of the residual errors

$$\min \; J = \frac{1}{2}(e_1^2 + e_2^2 + \cdots + e_m^2) = \frac{1}{2}\vec{e}^T\vec{e}$$

J is a scalar

$$J(\vec{x}) = \frac{1}{2}\vec{e}^T\vec{e}$$
$$= \frac{1}{2}(\tilde{\vec{y}} - H\hat{\vec{x}})^T(\tilde{\vec{y}} - H\hat{\vec{x}})$$
$$= \frac{1}{2}(\tilde{\vec{y}}^T\tilde{\vec{y}} - 2\tilde{\vec{y}}^T H\hat{\vec{x}} + \hat{\vec{x}}^T H^T H\hat{\vec{x}})$$

Necessary condition

$$\nabla_{\hat{\vec{x}}} J = \frac{\partial J}{\partial \hat{\vec{x}}} = \begin{bmatrix} \frac{\partial J}{\partial \hat{x}_1} \\ \frac{\partial J}{\partial \hat{x}_2} \\ \vdots \\ \frac{\partial J}{\partial \hat{x}_n} \end{bmatrix} = 0$$

$$\underbrace{\nabla_{\hat{\vec{x}}} J}_{\text{Jacobian}} = H^T H\hat{\vec{x}} - H^T\tilde{\vec{y}} = 0 \Rightarrow H^T H\hat{\vec{x}} = H^T\tilde{\vec{y}}$$
$$\Rightarrow \hat{\vec{x}} = (H^T H)^{-1} H^T \tilde{\vec{y}}$$

if $H^T H$ is invertible

Sufficient condition

$$\underbrace{\nabla^2_{\hat{\vec{x}}} J}_{\text{Hessian}} = \frac{\partial^2 J}{\partial \hat{\vec{x}}^2} = \frac{\partial}{\partial \hat{\vec{x}}}\left(\frac{\partial J}{\partial \hat{\vec{x}}}\right) = H^T H \succ 0$$

* Positive Definite (matrix $A \in \mathbb{R}^{n\times n}$) **Definition**

for any vector $u \in \mathbb{R}^{n\times 1} \neq 0$, $u^T A u > 0$

* A is Positive definite $\Leftrightarrow$ all eigenvalues of A are positive

Statement: $H^T H$ is always positive semidefinite ($H^T H \succeq 0$)

proof: for any vector $\vec{v} \in \mathbb{R}^{n\times 1}$, let $\vec{u} = H\vec{v}$, then $\vec{u}^T\vec{u} = \vec{v}^T H^T H\vec{v}$

$$\|u\|_2^2 = u^T u \geq 0 \quad \forall\; u \in \mathbb{R}^{m\times 1}$$

hence $\vec{v}^T H^T H\vec{v} \geq 0 \;\; \forall\; v \in \mathbb{R}^{n\times 1} \Rightarrow H^T H \succeq 0$

$H^T H \geq 0 \rightarrow$ if H has rank n, then $H^T H > 0$

If H has rank n, then its null space is $\{0\}$, hence the only $\vec{v}$ that makes $H\vec{v}=0$ is $\vec{v}=0$. For all $\vec{v} \neq 0$, $\vec{v} H^T H \vec{v} > 0$

minimize

$$\left.\begin{aligned} J(\hat{\vec{x}}) &= \tfrac{1}{2}\vec{e}^T\vec{e} \\ &= \tfrac{1}{2}(\tilde{\vec{y}} - H\hat{\vec{x}})^T(\tilde{\vec{y}} - H\hat{\vec{x}}) \\ &= \tfrac{1}{2}(\tilde{\vec{y}}^T\tilde{\vec{y}} - 2\tilde{\vec{y}}^T H\hat{\vec{x}} + \hat{\vec{x}}^T H^T H \hat{\vec{x}}) \end{aligned}\right\} \text{objective}$$

Necessary condition

$$\nabla_{\hat{\vec{x}}} J = \frac{\partial J}{\partial \hat{\vec{x}}} = \begin{bmatrix} \frac{\partial J}{\partial \hat{x}_1} \\ \frac{\partial J}{\partial \hat{x}_2} \\ \vdots \\ \frac{\partial J}{\partial \hat{x}_n} \end{bmatrix} = 0$$

$$\underbrace{\nabla_{\hat{\vec{x}}} J}_{\text{Jacobian}} = H^T H \hat{\vec{x}} - H^T \tilde{\vec{y}} = 0 \Rightarrow H^T H \hat{\vec{x}} = H^T \tilde{\vec{y}}$$

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

if $H^T H$ is invertible

Sufficient condition

$$\underbrace{\nabla^2_{\hat{\vec{x}}} J}_{\text{Hessian}} = \frac{\partial^2 J}{\partial \hat{\vec{x}}^2} = \frac{\partial}{\partial \hat{\vec{x}}}\left(\frac{\partial J}{\partial \hat{\vec{x}}}\right) = \underline{\underline{H^T H > 0}}$$

we need $\hat{\vec{x}}$ to satisfy both necessary conditions and saffi cient condition

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

optimal solution to $\min J(\hat{\vec{x}})$

4. Weighted least squares approximation

unweighted one : $J = \min \frac{1}{2} e^T e$

weighted one : $J = \min \frac{1}{2} (w_1 e_1^2 + w_2 e_2^2 + \cdots + w_m e_m^2)$

$= \min \frac{1}{2} e^T W e$

(W is a diagonal matrix)

$$W = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_m \end{bmatrix}$$

$$\nabla_{\hat{\vec{x}}} J = H^T W H \hat{\vec{x}} - H^T W \tilde{\vec{y}} = 0 \quad \text{(necessary)}$$

$$\nabla^2_{\hat{\vec{x}}} J = H^T W H \quad \text{(sufficient)} \quad \succ 0$$

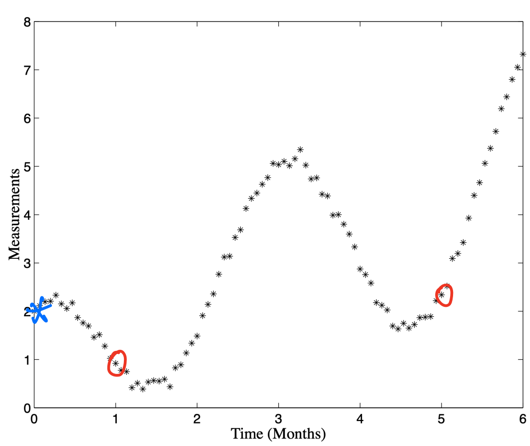
$$\boxed{\hat{\vec{x}} = (H^T W H)^{-1} H^T W \tilde{\vec{y}}}$$

5. Constrained least squares approximation

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

$$\begin{cases} \tilde{\vec{y}}_1 = H_1 \hat{x} + e_1 \\ \tilde{\vec{y}}_2 = H_2 \hat{x} \quad \text{(perfect measurement)} \end{cases}$$

$m_1 \geq n \qquad m_2 \leq n$



$\tilde{\vec{y}}_2 = \{ \tilde{y}(t=1), \tilde{y}(t=2), \ldots \tilde{y}(t=6) \}$

require $\hat{\vec{y}}_2 = H_2 \hat{\vec{x}} = \tilde{\vec{y}}_2$

Problem statement for optimization set up

$$\min \quad J = \frac{1}{2} e_1^T W_1 e_1 \qquad \text{objective}$$

$$\text{subject to} \quad \tilde{\vec{Y}}_2 = H_2 \hat{X} \qquad \text{constraint } 1.$$

$\hat{x}$ is the varable we need to select to optimize J

Using the method of Lagrangian multipliers

$$J = \frac{1}{2} e_1^T W_1 e + \underline{\vec{\lambda}^T (\tilde{\vec{y}} - H_2 \hat{x})}$$

In this new J, we now have two variables $\hat{x}$ and λ

$$\nabla_{\hat{\vec{x}}} J = (H_1^T W_1 H_1)\hat{\vec{x}} - H_1^T W_1 \tilde{\vec{Y}}_1 - H_2^T \lambda = 0 \quad ①$$

$$\nabla_{\vec{\lambda}} J = \underline{\tilde{\vec{Y}}_2 - H_2 \hat{X}} = 0 \qquad \text{Constraint } ②$$

1) From ① $\Rightarrow$ $\underline{(H_1^T W_1 H_1)}\hat{\vec{x}} = H_1^T W_1 \tilde{\vec{Y}}_1 + \underline{H_2^T \vec{\lambda}}$

$$\hat{\vec{x}} = \underbrace{(H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{\vec{Y}}_1}_{\text{term } ⑤} + (H_1^T W_1 H_1)^{-1} H_2^T \vec{\lambda}$$

2) plug $\hat{\vec{x}}$ into ②

$$\tilde{\vec{Y}}_2 = H_2 \hat{\vec{X}}$$

$$\hat{\vec{Y}}_2 = \underbrace{H_2 (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{\vec{Y}}_1}_{\text{term } ③} + \underbrace{H_2 (H_1^T W_1 H_1)^{-1} H_2^T \vec{\lambda}}_{\text{term } ④}$$

$$\underline{H_2 (H_1^T W_1 H_1)^{-1} H_2^T \vec{\lambda}} = \tilde{\vec{Y}}_2 - H_2 (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{\vec{Y}}_1$$

$$\vec{\lambda} = \left[H_2 (H_1^T W_1 H_1)^{-1} H_2^T\right]^{-1} \left\{ \tilde{\vec{Y}}_2 - \underline{H_2 (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{\vec{Y}}_1} \right]$$

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

$$K = (H_1^T W_1 H_1)^{-1} H_2^T \left[H_2 (H_1^T W_1 H_1)^{-1} H_2^T \right]^{-1}$$

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