In The Name of God

Sharif University of Technology
Department of Aerospace Engineering

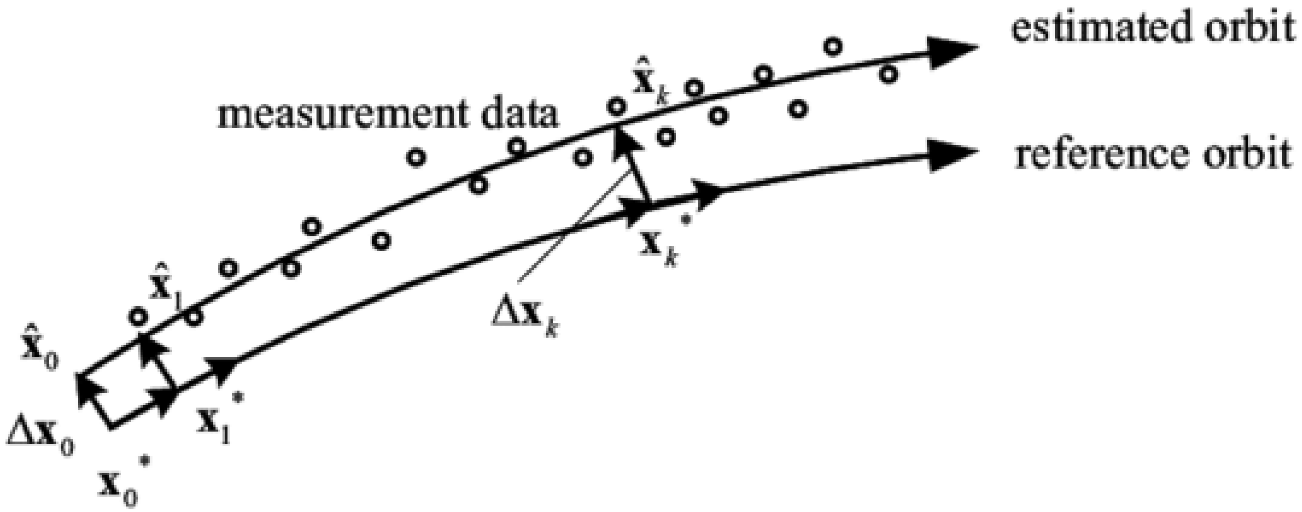
45-766: Optimal Control II

Instructor: Maryam Kiani

CH#2: Least Squares Estimation

Least Squares Estimation

- Least squares can be used in a wide variety of categorical applications including: curve fitting of data, parameter estimation, etc.
- Three quantities of interest for any variable or parameter in estimation:
  - true value
  - measured value
  - estimated value



A Curve Fitting Example

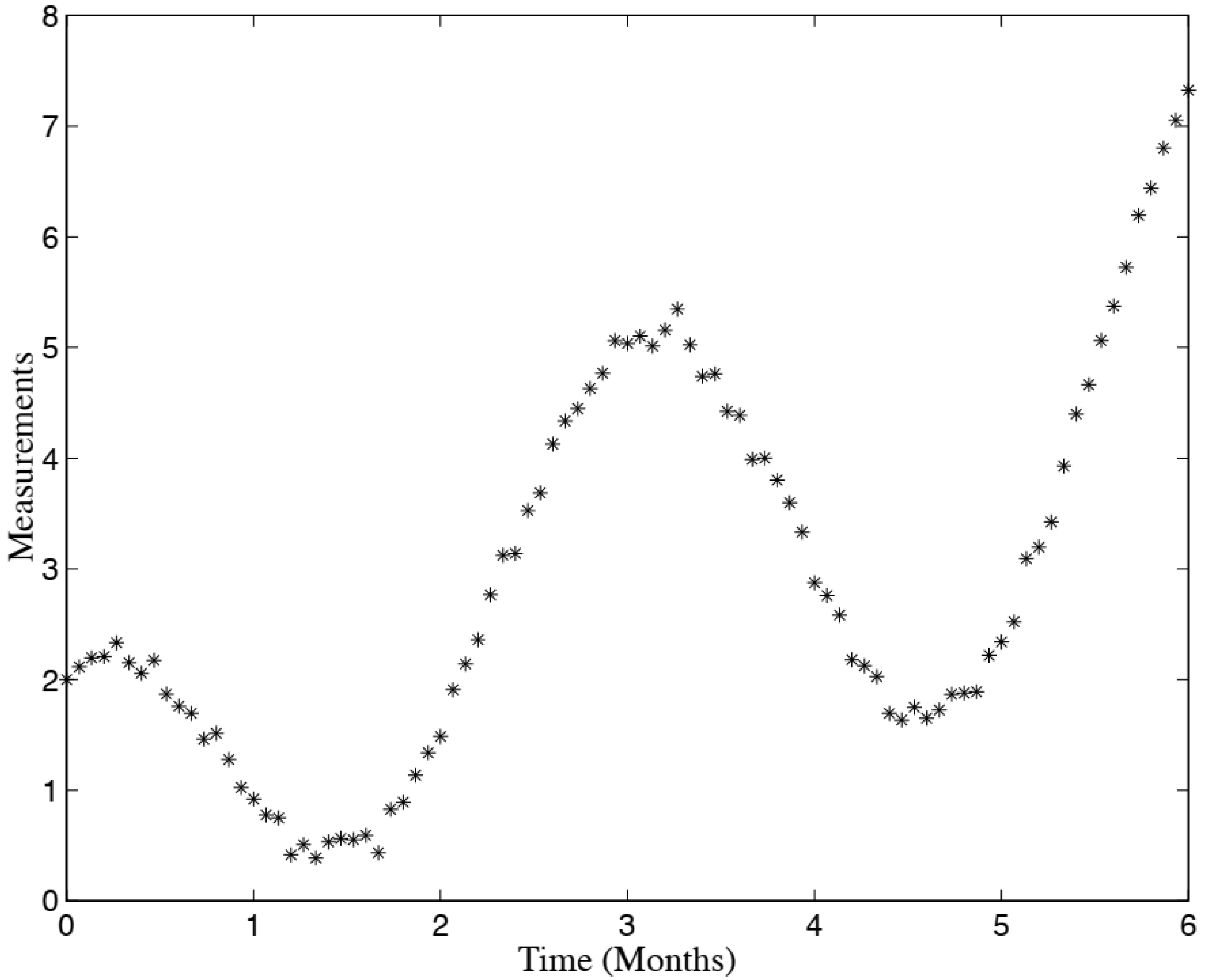
Model 1: $z_1(t) = c_1 t + c_2 \sin(t) + c_3 \cos(2t)$

Model 2: $z_2(t) = d_1(t+2) + d_2 t^2 + d_3 t^3$

- The process of fitting curves, such as Models 1 and 2, to measured data is known in statistics as *regression*.

$$\mathbf{z} = H\mathbf{x} + \boldsymbol{\upsilon}$$

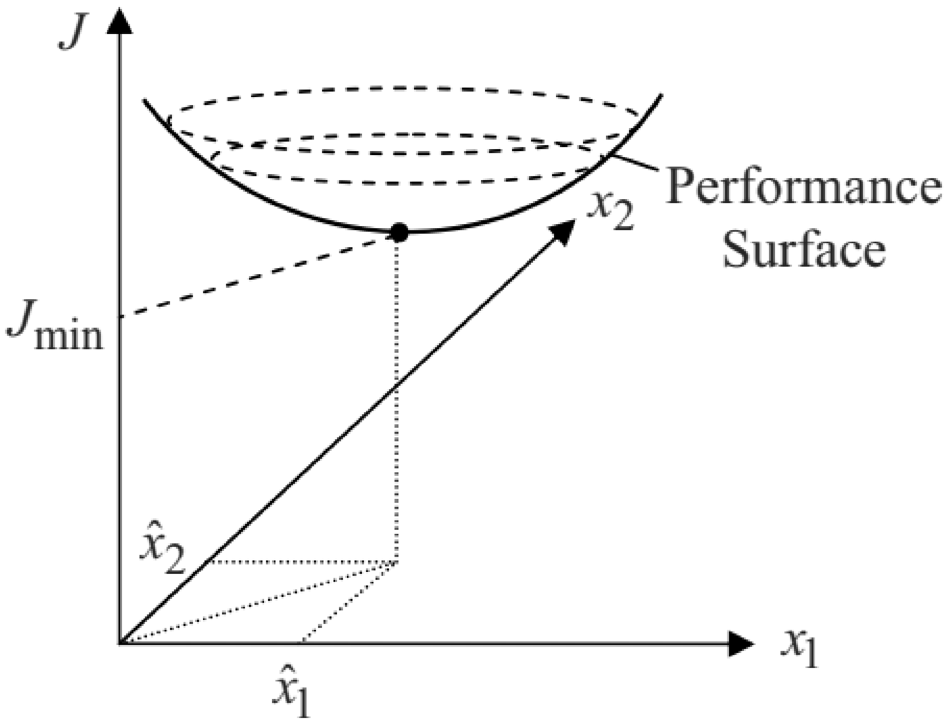
$$\hat{\mathbf{z}} = H\hat{\mathbf{x}}$$



Linear Least Squares Estimation

$$\mathbf{z} = H\mathbf{x} + \boldsymbol{\upsilon}, \qquad \hat{\mathbf{z}} = H\hat{\mathbf{x}}$$

$$J = \frac{1}{2}(\mathbf{z} - \hat{\mathbf{z}})^T(\mathbf{z} - \hat{\mathbf{z}}) = \frac{1}{2}(\mathbf{z} - H\hat{\mathbf{x}})^T(\mathbf{z} - H\hat{\mathbf{x}})$$



- Necessary condition:

$$\frac{\partial J}{\partial \hat{\mathbf{x}}} = -H^T\mathbf{z} + H^T H\hat{\mathbf{x}} = 0 \rightarrow \hat{\mathbf{x}} = (H^T H)^{-1} H^T \mathbf{z}$$

- Sufficient condition:

$$\frac{\partial^2 J}{\partial \hat{\mathbf{x}} \partial \hat{\mathbf{x}}^T} = H^T H > 0$$

Least Squares Estimation

$$Cov(\hat{\mathbf{x}}) \equiv E[(\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))(\hat{\mathbf{x}} - E(\hat{\mathbf{x}}))^T]$$

$$= E\left\{(H^T H)^{-1} H^T (\mathbf{z} - \hat{\mathbf{z}})(\mathbf{z} - \hat{\mathbf{z}})^T H (H^T H)^{-1}\right\}$$

$$= (H^T H)^{-1} H^T \underbrace{E(\boldsymbol{\upsilon}\boldsymbol{\upsilon}^T)}_{R} H (H^T H)^{-1}$$

Example

- $z_i = 0.3\sin(t_i) + 0.5\cos(t_i) + 0.1t_i + \upsilon_i, \qquad \upsilon \sim N(0, \sqrt{0.001})$

  $\hat{z}_i = a\sin(t_i) + b\cos(t_i) + ct_i$

$$H = \begin{bmatrix} \sin(t_0) & \cos(t_0) & t_0 \\ \sin(t_1) & \cos(t_1) & t_1 \\ \vdots & \vdots & \vdots \\ \sin(t_{100}) & \cos(t_{100}) & t_{100} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$\Rightarrow \hat{\mathbf{x}} = [0.3019 \quad 0.5072 \quad 0.1027]^T$$

Example

- Some applications
  - TAM Calibration: $\boldsymbol{B}_m = \boldsymbol{CH} + \boldsymbol{b} + \boldsymbol{v}$
  - Aerodynamic parameter Estimation

$$C_m = C_{m_0} + C_{m_\alpha}\alpha + C_{m_M}M + C_{m_{\alpha M}}\alpha M + v_m$$

$$C_m = \frac{1}{\bar{q}S\bar{c}}\left[I_y\dot{q} + (I_x - I_z)pr + I_{xz}(p^2 - r^2) - I_p\Omega_p r\right]$$

Least Squares Approximation

- QR method

$$(H^T H)\hat{\mathbf{x}} = H^T \mathbf{z} \xrightarrow{H = \overset{\text{orthogonal}}{Q}R} (R^T Q^T Q R)\hat{\mathbf{x}} = R^T Q^T \mathbf{z}$$

$$R^T R\hat{\mathbf{x}} = R^T Q^T \mathbf{z} \longrightarrow R\hat{\mathbf{x}} = Q^T \mathbf{z} = \mathbf{y}$$

- If n=3

$$\begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \hat{x}_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \Rightarrow \begin{aligned} \hat{x}_3 &= \frac{y_3}{R_{33}} \\ \hat{x}_2 &= \frac{1}{R_{22}}(y_2 - R_{23}\hat{x}_3) \\ \hat{x}_3 &= \frac{1}{R_{11}}(y_1 - R_{12}\hat{x}_2 - R_{13}\hat{x}_3) \end{aligned}$$

Least Squares Approximation

- SVD (Singular Value Decomposition) method

$$(H^T H)\hat{\mathbf{x}} = H^T \mathbf{z} \xrightarrow{H = USV^T} (VSU^T USV^T)\hat{\mathbf{x}} = VSU^T \mathbf{z}$$

$$SV^T \hat{\mathbf{x}} = U^T \mathbf{z} \longrightarrow \hat{\mathbf{x}} = VS^{-1}U^T \mathbf{z}$$

Weighted Least Square Estimation

- $\mathbf{z} = H\mathbf{x} + \boldsymbol{\upsilon},\ \ \boldsymbol{\upsilon} \sim (0, R)$

  $\hat{\mathbf{z}} = H\hat{\mathbf{x}}$

  $$J = \frac{1}{2}(\mathbf{z} - \hat{\mathbf{z}})^T W (\mathbf{z} - \hat{\mathbf{z}}) = \frac{1}{2}(\mathbf{z} - H\hat{\mathbf{x}})^T W (\mathbf{z} - H\hat{\mathbf{x}})$$

  $$\frac{\partial J}{\partial \hat{\mathbf{x}}} = 0 \to \hat{\mathbf{x}} = (H^T W H)^{-1} H^T W \mathbf{z}$$

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

Nonlinear Least Square Estimation

- The method of NLS was originally developed by Gauss and employed to determine planetary orbits (during the early 1800s) from telescope measurements of the “line of sight angles” to the planets.

- It is a generalization of Newton’s root solving method for finding *x*-values satisfying $z - h(x) = 0$.

- $\mathbf{z} = h(\mathbf{x}) + \boldsymbol{\upsilon}, \;\; \boldsymbol{\upsilon} \sim (0, R)$

  $\hat{\mathbf{z}} = h(\hat{\mathbf{x}})$

  $$J = \frac{1}{2}(\mathbf{z} - \hat{\mathbf{z}})^T W (\mathbf{z} - \hat{\mathbf{z}}) = \frac{1}{2}(\mathbf{z} - h(\hat{\mathbf{x}}))^T W (\mathbf{z} - h(\hat{\mathbf{x}}))$$

Nonlinear Least Square Estimation

- Assume $\hat{\mathbf{x}} = \mathbf{x}_c + \Delta\mathbf{x}$

$$h(\hat{\mathbf{x}}) \approx h(\mathbf{x}_c) + H\Delta\mathbf{x} \quad where \quad H = \left.\frac{\partial h}{\partial \mathbf{x}}\right|_{\mathbf{x}_c}$$

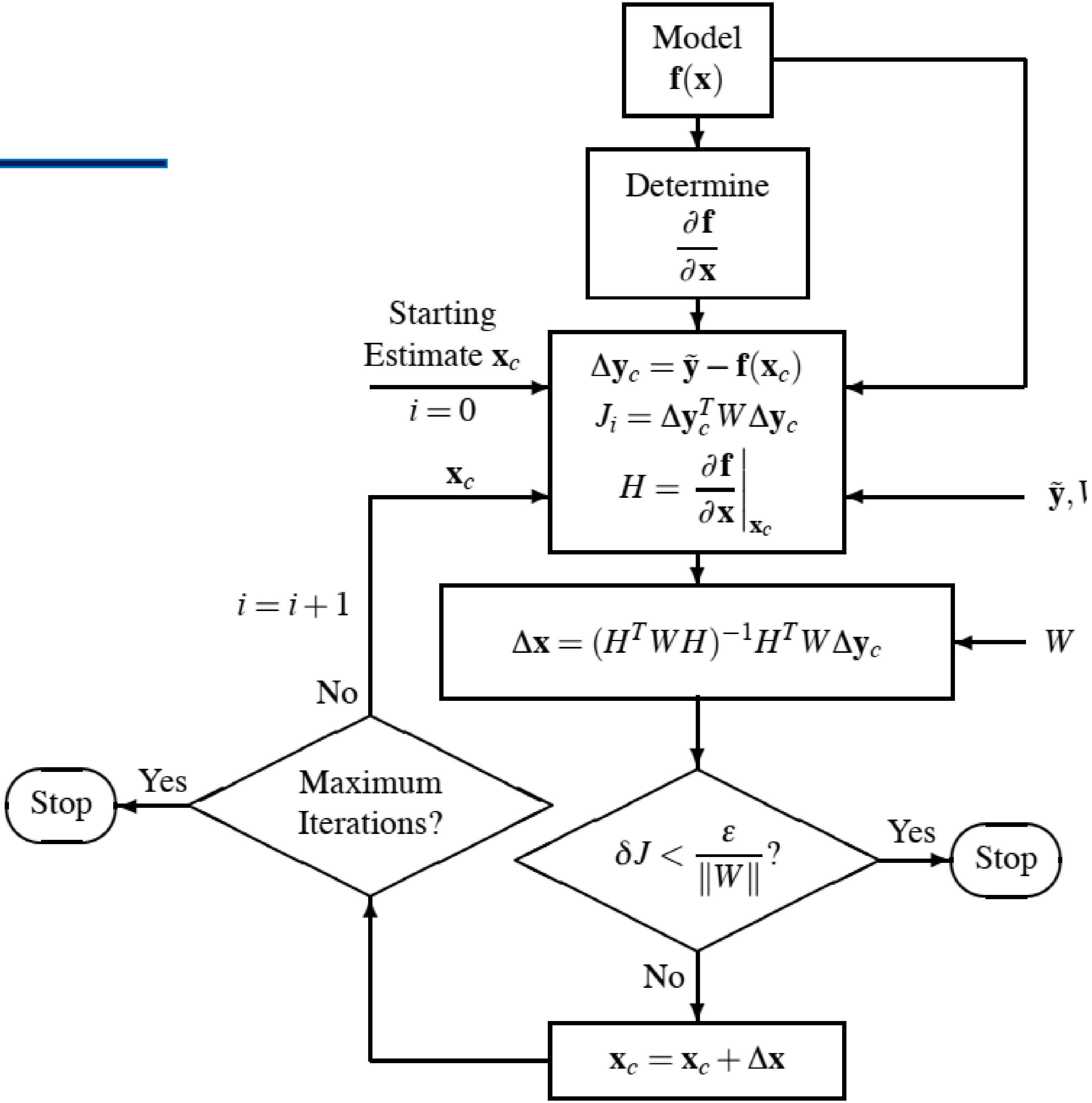
$$J = \frac{1}{2}(\mathbf{z} - h(\mathbf{x}_c) - H\Delta\mathbf{x})^T W(\mathbf{z} - h(\mathbf{x}_c) - H\Delta\mathbf{x})$$

$$= \frac{1}{2}(\Delta\mathbf{z}_c - H\Delta\mathbf{x})^T W(\Delta\mathbf{z}_c - H\Delta\mathbf{x})$$

$$\Rightarrow \Delta\mathbf{x} = (H^T W H)^{-1} H^T W \Delta\mathbf{z}_c$$

$$while \;\; \delta J = \frac{|J_i - J_{i-1}|}{J_i} < \frac{\varepsilon}{\|W\|}$$

NLS



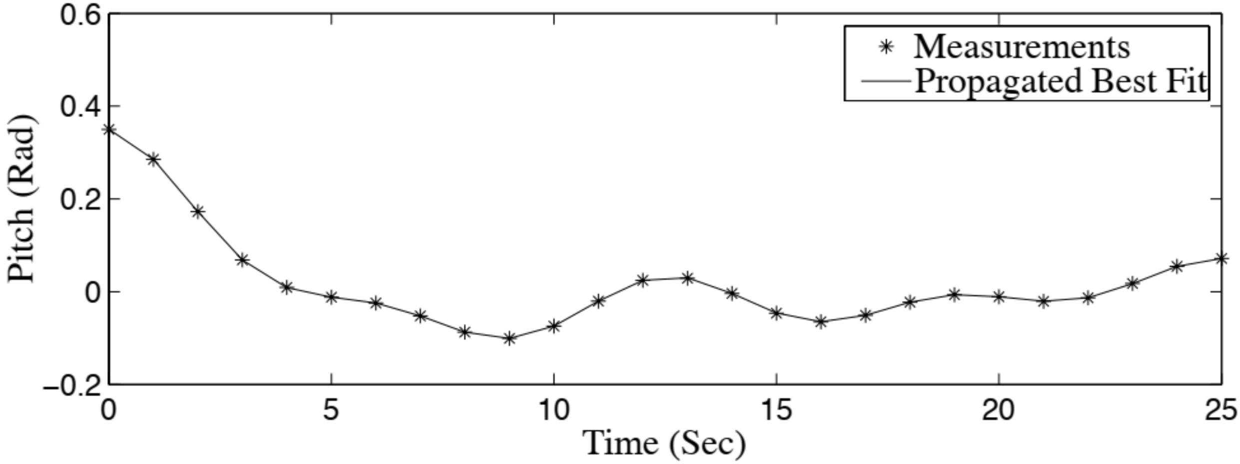
Model
$\mathbf{f}(\mathbf{x})$
Determine
$\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$
Starting
Estimate $\mathbf{x}_c$
$i = 0$
$\Delta\mathbf{y}_c = \tilde{\mathbf{y}} - \mathbf{f}(\mathbf{x}_c)$
$J_i = \Delta\mathbf{y}_c^T W \Delta\mathbf{y}_c$
$H = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}_c}$
$\mathbf{x}_c$
$\tilde{\mathbf{y}}$,
$i = i + 1$
$\Delta\mathbf{x} = (H^T W H)^{-1} H^T W \Delta\mathbf{y}_c$
W
No
Yes
Stop
Maximum
Iterations?
$\delta J < \frac{\varepsilon}{\|W\|}$?
Yes
Stop
No
$\mathbf{x}_c = \mathbf{x}_c + \Delta\mathbf{x}$

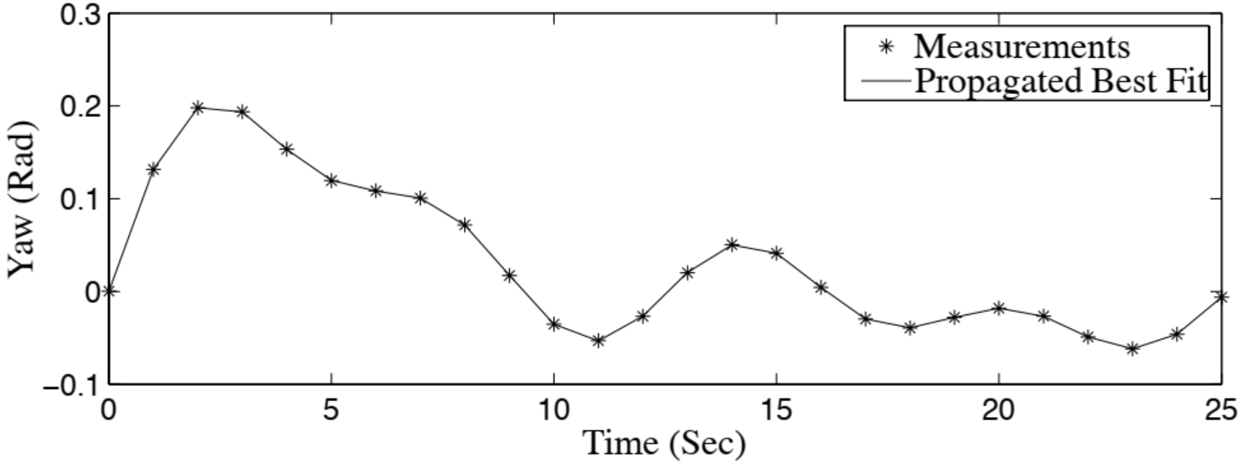
Example

- Under certain approximations, the pitch (θ) and yaw (ψ) attitude dynamics of an inertially and aerodynamically symmetric projectile can be modeled as

$$\theta(t) = k_1 e^{\lambda_1 t}\cos(\omega_1 t + \delta_1) + k_2 e^{\lambda_2 t}\cos(\omega_2 t + \delta_2) + k_3 e^{\lambda_3 t}\cos(\omega_3 t + \delta_3) + k_4$$

$$\psi(t) = k_1 e^{\lambda_1 t}\sin(\omega_1 t + \delta_1) + k_2 e^{\lambda_2 t}\sin(\omega_2 t + \delta_2) + k_3 e^{\lambda_3 t}\sin(\omega_3 t + \delta_3) + k_5$$





Example

- $\overset{(14\times 1)}{\mathbf{x}} = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 & k_5 & \lambda_1 & \lambda_2 & \lambda_3 & \omega_1 & \omega_2 & \omega_3 & \delta_1 & \delta_2 & \delta_3 \end{bmatrix}^T$

$$\overset{(52\times 14)}{H} = \begin{bmatrix} \left.\dfrac{\partial\theta(0)}{\partial x_1}\right|_{\mathbf{x}_c} & \cdots & \left.\dfrac{\partial\theta(0)}{\partial x_{14}}\right|_{\mathbf{x}_c} \\ \left.\dfrac{\partial\psi(0)}{\partial x_1}\right|_{\mathbf{x}_c} & \cdots & \left.\dfrac{\partial\psi(0)}{\partial x_{14}}\right|_{\mathbf{x}_c} \\ \vdots & & \vdots \\ \left.\dfrac{\partial\theta(25)}{\partial x_1}\right|_{\mathbf{x}_c} & \cdots & \left.\dfrac{\partial\theta(25)}{\partial x_{14}}\right|_{\mathbf{x}_c} \\ \left.\dfrac{\partial\psi(25)}{\partial x_1}\right|_{\mathbf{x}_c} & \cdots & \left.\dfrac{\partial\psi(25)}{\partial x_{14}}\right|_{\mathbf{x}_c} \end{bmatrix}$$

Parameter	Iteration Number				
	0	1	2	$\cdots$	5
k_1	0.5000	0.1852	0.1975		0.1999
k_2	0.2500	0.1075	0.1012		0.0997
k_3	0.1250	0.0567	0.0505		0.0500
k_4	0.0000	−0.0006	0.0001		0.0002
k_5	0.0000	−0.0018	−0.0005		0.0001
λ_1	−0.1500	−0.1234	−0.0954		−0.0998
λ_2	−0.0600	−0.0661	−0.0585		−0.0497
λ_3	−0.0300	−0.0398	−0.0338		−0.0250
ω_1	0.2600	0.2490	0.2471		0.2500
ω_2	0.5500	0.5300	0.4955		0.4999
ω_3	0.9500	0.9697	1.0068		0.9998
δ_1	0.0100	0.0344	0.0143		0.0010
δ_2	0.0100	−0.0447	0.0051		0.0001
δ_3	0.0100	0.0024	−0.0570		−0.0001

Recursive Least Square Estimation

- How to recursively compute the weighted least squares estimate of a parameter?

- Suppose we have $\hat{\mathbf{x}}_k$ after (k-1) measurements, and we obtain new measurement $\mathbf{z}_k$

$$\mathbf{z}_k = H_k \mathbf{x} + \boldsymbol{\upsilon}_k$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + K_k(\mathbf{z}_k - H_k \hat{\mathbf{x}}_{k-1})$$

- $\tilde{\mathbf{x}}_k = \mathbf{x} - \hat{\mathbf{x}}_k$

$$\begin{aligned} E(\tilde{\mathbf{x}}_k) &= E(\mathbf{x} - \hat{\mathbf{x}}_k) = E[\mathbf{x} - \hat{\mathbf{x}}_{k-1} - K_k(\mathbf{z}_k - H_k \hat{\mathbf{x}}_{k-1})] = \\ &= E[\tilde{\mathbf{x}}_{k-1} - K_k(H_k \mathbf{x} + \boldsymbol{\upsilon}_k - H_k \hat{\mathbf{x}}_{k-1})] = E[(I - K_k H_k)\tilde{\mathbf{x}}_{k-1} - K_k \boldsymbol{\upsilon}_k] \\ &= (I - K_k H_k) E(\tilde{\mathbf{x}}_{k-1}) - K_k E(\boldsymbol{\upsilon}_k) \end{aligned}$$

Recursive Least Square Estimation

- $J_k = E[(x_1 - \hat{x}_{1,k})^2 + \ldots + (x_n - \hat{x}_{n,k})^2]$

  $= E[\tilde{x}_{1,k}^2 + \ldots + \tilde{x}_{n,k}^2] = E[\tilde{\mathbf{x}}_k^T \tilde{\mathbf{x}}_k] = E[Tr(\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^T)] = Tr[E(\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^T)] = Tr(P_k)$

- $P_k = E(\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k{}^T) = E\{[(I - K_k H_k)\tilde{\mathbf{x}}_{k-1} - K_k \boldsymbol{\upsilon}_k][\ldots]^T\}$

  $= (I - K_k H_k) E(\tilde{\mathbf{x}}_{k-1} \tilde{\mathbf{x}}_{k-1}{}^T)(I - K_k H_k)^T + K_k E(\boldsymbol{\upsilon}_k \boldsymbol{\upsilon}_k{}^T) K_k{}^T$

  $= (I - K_k H_k) P_{k-1} (I - K_k H_k)^T + K_k R_k K_k{}^T$

- $\dfrac{\partial J_k}{\partial K_k} = 2(I - K_k H_k) P_{k-1} (-H_k)^T + 2 K_k R_k = 0$

  $\rightarrow K_k = P_{k-1} H_k{}^T (H_k P_{k-1} H_k^T + R_k)^{-1}$

Recursive Least Square Estimation

- $\mathbf{z}_k = H_k \mathbf{x} + \boldsymbol{\upsilon}_k$

$$K_k = P_{k-1} H_k^T (H_k P_{k-1} H_k^T + R_k)^{-1}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + K_k (\mathbf{z}_k - H_k \hat{\mathbf{x}}_{k-1})$$

$$P_k = (I - K_k H_k) P_{k-1} (I - K_k H_k)^T + K_k R_k K_k^T$$

Alternative Estimator Forms

- $S_k = H_k P_{k-1} H_k^T + R_k \rightarrow K_k = P_{k-1} H_k^T S_k^{-1}$

$$P_k = (I - K_k H_k) P_{k-1} (I - K_k H_k)^T + K_k R_k K_k^T$$

$$P_k = (I - P_{k-1} H_k^T S_k^{-1} H_k) P_{k-1} (I - P_{k-1} H_k^T S_k^{-1} H_k)^T + (P_{k-1} H_k^T S_k^{-1}) R_k (P_{k-1} H_k^T S_k^{-1})^T$$

$$\begin{aligned}
P_k &= P_{k-1} - P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} - P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} + \\
&\quad P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} + P_{k-1} H_k^T S_k^{-1} R_k S_k^{-1} H_k P_{k-1} \\
P_k &= P_{k-1} - 2 P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} + P_{k-1} H_k^T S_k^{-1} S_k S_k^{-1} H_k P_{k-1} \\
&= P_{k-1} - 2 P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} + P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} \\
&= P_{k-1} - P_{k-1} H_k^T S_k^{-1} H_k P_{k-1} \\
P_k &= P_{k-1} - K_k H_k P_{k-1} \\
&= (I - K_k H_k) P_{k-1}
\end{aligned}$$

Alternative Estimator Forms

- $$\begin{aligned} P_k &= P_{k-1} - P_{k-1}H_k^T S_k^{-1} H_k P_{k-1} \\ &= P_{k-1} - P_{k-1}H_k^T (H_k P_{k-1} H_k^T + R_k)^{-1} H_k P_{k-1} \end{aligned}$$

$$P_k^{-1} = [P_{k-1} - P_{k-1}H_k^T (H_k P_{k-1} H_k^T + R_k)^{-1} H_k P_{k-1}]^{-1}$$

Remember matrix inversion lemma

$$\begin{aligned} P_k^{-1} &= P_{k-1}^{-1} + P_{k-1}^{-1} P_{k-1} H_k^T \left[(H_k P_{k-1} H_k^T + R_k) - \right. \\ &\quad \left. H_k P_{k-1} P_{k-1}^{-1} (P_{k-1} H_k^T)\right]^{-1} H_k P_{k-1} P_{k-1}^{-1} \\ &= P_{k-1}^{-1} + H_k^T R_k^{-1} H_k \end{aligned}$$

$$P_k = \left[P_{k-1}^{-1} + H_k^T R_k^{-1} H_k\right]^{-1}$$

Alternative Estimator Forms

- Matrix inversion lemma

$$F = [A + BCD]^{-1}$$

where

$$\begin{aligned} F &= \text{an arbitrary } n \times n \text{ matrix} \\ A &= \text{an arbitrary } n \times n \text{ matrix} \\ B &= \text{an arbitrary } n \times m \text{ matrix} \\ C &= \text{an arbitrary } m \times m \text{ matrix} \\ D &= \text{an arbitrary } m \times n \text{ matrix} \end{aligned}$$

$$\Rightarrow \quad F = A^{-1} - A^{-1}B\left(DA^{-1}B + C^{-1}\right)^{-1}DA^{-1}$$

Alternative Estimator Forms

- $K_k = P_{k-1}H_k^T(H_kP_{k-1}H_k^T + R_k)^{-1}$

$$K_k = P_kP_k^{-1}P_{k-1}H_k^T(H_kP_{k-1}H_k^T + R_k)^{-1}$$

$$K_k = P_k(P_{k-1}^{-1} + H_k^TR_k^{-1}H_k)P_{k-1}H_k^T(H_kP_{k-1}H_k^T + R_k)^{-1}$$

$$K_k = P_k(H_k^T + H_k^TR_k^{-1}H_kP_{k-1}H_k^T)(H_kP_{k-1}H_k^T + R_k)^{-1}$$

$$K_k = P_kH_k^T(I + R_k^{-1}H_kP_{k-1}H_k^T)(H_kP_{k-1}H_k^T + R_k)^{-1}$$

$$\begin{aligned} K_k &= P_kH_k^TR_k^{-1}(R_k + H_kP_{k-1}H_k^T)(H_kP_{k-1}H_k^T + R_k)^{-1} \\ &= P_kH_k^TR_k^{-1} \end{aligned}$$

Example

- $$\begin{aligned} y_k &= x_1 + 0.99^{k-1} x_2 + v_k \\ &= \begin{bmatrix} 1 & 0.99^{k-1} \end{bmatrix} x + v_k \end{aligned}$$

where v_k is the measurement noise, which is a zero-mean random variable with a variance of $R = 0.01$. Suppose that $x_1 = 10$ and $x_2 = 5$. Further suppose that your initial estimates are $\hat{x}_1 = 8$ and $\hat{x}_2 = 7$, with an initial estimation-error variance P_0 that is equal to the identity matrix.