Optimal Estimation Methods

(Lecture 3 – Least Squares Estimation)

Dr. John L. Crassidis

University at Buffalo – State University of New York
Department of Mechanical & Aerospace Engineering
Amherst, NY 14260-4400
johnc@buffalo.edu
http://www.buffalo.edu/~johnc

University at Buffalo The State University of New York

Introduction (i)

- Estimation combines many topics, as discussed before
  - Optimization (parameter), control theory (optimal control), stochastic processes (probability and statistics)
- Three quantities of interest
  - True value (truth, note: unknown!!)
  - Estimated value (our best estimate of truth)
  - Measured value (our sensed quantity)

Both x and v are unknown!

$$\begin{array}{ccccc} \text{measured value} & = & \text{true value} & + & \text{measurement error} \\ \tilde{x} & = & x & + & v \end{array}$$

$\hat{x} + e = x + v$

$$\begin{array}{ccccc} \text{measured value} & = & \text{estimated value} & + & \text{residual error} \\ \tilde{x} & = & \hat{x} & + & e \end{array}$$

- Measured value is “modeled” by true value with some error
  - This is just a number that is obtained from a sensor
  - We never now the truth but we generally have some statistics on the measurement error through sensor calibration

Introduction (ii)

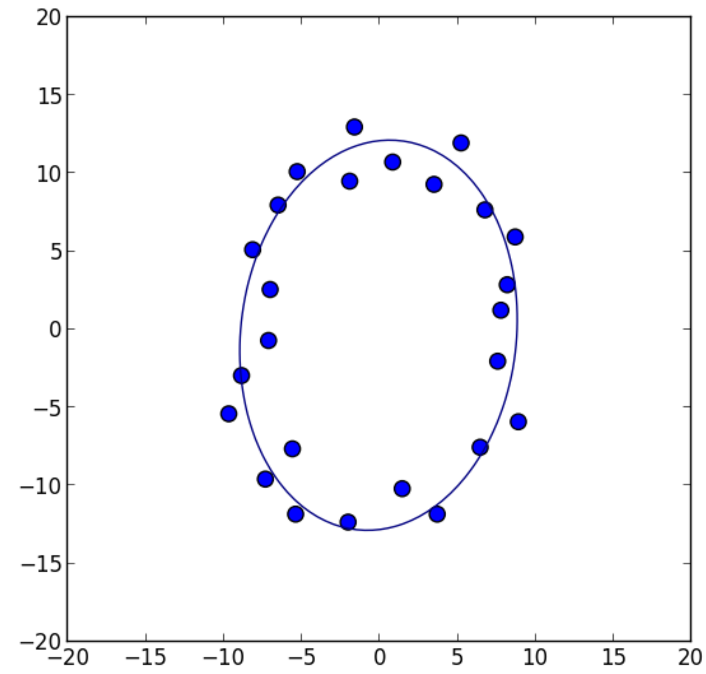
- The least-squares method is usually credited to Carl Friedrich Gauss (1809), but it was actually first published by Adrien-Marie Legendre earlier (1805)
  - “Least squares” means that the overall solution minimizes the sum of the squares of the errors in the results of every single equation
  - Legendre demonstrated the new method by analyzing the same data as Laplace for the shape of the Earth
  - Gauss published his method of accurately calculating the orbits of celestial bodies
    - Predicted the future location of the newly discovered asteroid Ceres that was lost after 40 days of observations
    - Astronomer Franz Xaver von Zach relocated Ceres using Gauss’ approach (note that Gauss was only 24 years old at the time!)
    - He went beyond Legendre and succeeded in connecting the method of least squares with the principles of probability and to the normal distribution

Introduction (iii)

- Many practical applications
  - Determining the damping properties of a fluid-filled damper as a function of temperature
  - Identification of aircraft dynamic and static aerodynamic coefficients
  - Orbit and attitude determination ← need 4 GPS satellites
  - Position determination using triangulation
  - Modal identification of vibratory systems
  - Even modern control strategies, for instance certain adaptive controllers, use the least squares approximation to update model parameters in the control system
  - Many others
- Foundation for modern dynamic filters
  - The "Kalman filter" can be derived using a least squares approach
- In Gauss' original approach vectors and matrices were not used, but it's much easier to use them, which will be done here

Introduction (iv)

- The most common application is in data fitting
  - The best fit in the least-squares sense minimizes the sum of squared residuals, a residual being the difference between an observed value and the fitted value provided by a model with unknown parameters
  - Over-deterministic case
    - Number of observations is greater than the number of unknowns
  - Under-deterministic case
    - Number of observations is less than the number of unknowns
    - No solution possible
  - Deterministic case
    - Number of observations is equal to the number of unknowns
    - Leads to a simple solution



Conic fitting a set of points using least-squares approximation

An "Experiment" (i)

- Say we have a set of measurements

Time	0	1	2	3	4	5
$\tilde{y}$	0.01	0.087	0.173	0.263	0.346	0.440
$\hat{x}$	∞	4.98	4.95	5.03	4.96	5.05

- Given model: $\tilde{y} = x \sin t + v$
- Want to estimate x (remember v is unknown)
- Only one unknown, so why not estimate x at each time using

$$\hat{x} = \frac{\tilde{y}}{\sin t}$$

- What happened?
  - At $t = 0$ no estimate is possible
  - Different estimates at different times, which is due to the measurement errors

An “Experiment” (ii)

- This is clearly an over-deterministic case
  - More measurements than unknowns
- Why not average the estimates from each time to obtain the final estimate?
  - Ignore $t = 0$ since the solution is ∞

$$\hat{x} = \frac{4.98 + 4.95 + 5.03 + 4.96 + 5.05}{5} = 4.994$$

  - True value was 5
  - Averaged value is “better” than the individual ones
- Is there a more mathematically rigorous way?
  - The answer is yes
  - Least squares estimation

Linear Least Squares (i)

- Measurement equation

Basis

$$\underset{m\times 1}{\tilde{\mathbf{y}}} = \underset{m\times n}{H}\underset{n\times 1}{\hat{\mathbf{x}}} + \underset{m\times 1}{\mathbf{e}}$$

where

$$\tilde{\mathbf{y}} = \begin{bmatrix}\tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_m\end{bmatrix}^T = \textit{measured } y\text{-values}$$

$$\mathbf{e} = \begin{bmatrix}e_1 & e_2 & \cdots & e_m\end{bmatrix}^T = \textit{residual errors}$$

$$\hat{\mathbf{x}} = \begin{bmatrix}\hat{x}_1 & \hat{x}_2 & \cdots & \hat{x}_n\end{bmatrix}^T = \textit{estimated } x\text{-values}$$

$$H = \begin{bmatrix} h_1(t_1) & h_2(t_1) & \cdots & h_n(t_1) \\ h_1(t_2) & h_2(t_2) & \cdots & h_n(t_2) \\ \vdots & \vdots & & \vdots \\ h_1(t_m) & h_2(t_m) & \cdots & h_n(t_m) \end{bmatrix}$$

Given from sensors

This is what we wish to find

Chosen basis functions

Must be linearly independent

Linear Least Squares (ii)

- Identical equation

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

$\tilde{y} - \hat{y} = e = Hx + v - H\hat{x} = \tilde{y} - H\hat{x}$

where

$$\underset{n\times 1}{\mathbf{x}} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T = \textit{true } x\text{-values}$$
$$\underset{m\times 1}{\mathbf{v}} = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix}^T = \textit{measurement errors}$$
$$\underset{m\times 1}{\hat{\mathbf{y}}} = \begin{bmatrix} \hat{y}_1 & \hat{y}_2 & \cdots & \hat{y}_m \end{bmatrix}^T = \textit{estimated } y\text{-values}$$
$$\underset{m\times 1}{\tilde{\mathbf{y}}} = \begin{bmatrix} \tilde{y}_1 & \tilde{y}_2 & \cdots & \tilde{y}_m \end{bmatrix}^T = \textit{measured } y\text{-values}$$

- Gauss selects, as an optimum choice for the unknown parameters, the particular $\hat{\mathbf{x}}$ that minimizes the sum square of the residual errors

$$J = \frac{1}{2}\mathbf{e}^T\mathbf{e}$$

½ factor to be discussed later

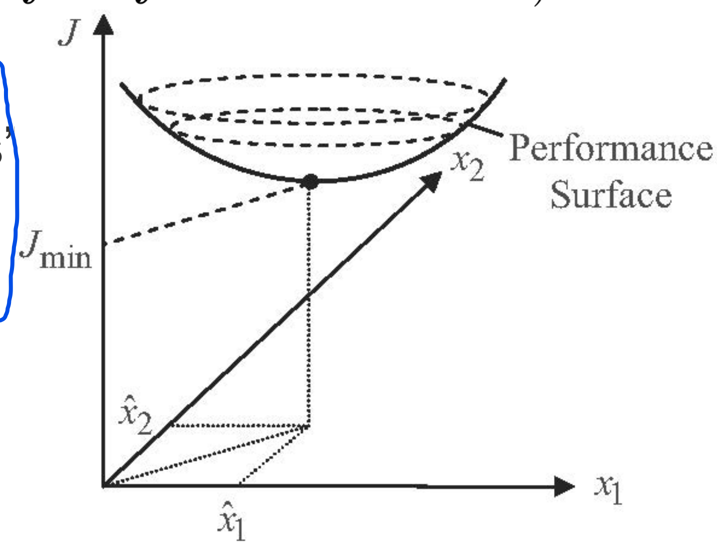
$$\mathbf{e} = \tilde{\mathbf{y}} - \hat{\mathbf{y}} = \tilde{\mathbf{y}} - H\,\hat{\mathbf{x}}$$

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

necessary condition

Gauss' "Normal Equations"

$$\nabla_{\hat{\mathbf{x}}} J \equiv \begin{bmatrix} \frac{\partial J}{\partial \hat{x}_1} \\ \vdots \\ \frac{\partial J}{\partial \hat{x}_n} \end{bmatrix} = \boxed{H^T H\hat{\mathbf{x}} - H^T\tilde{\mathbf{y}} = \mathbf{0}}$$



sufficient condition

$$\nabla^2_{\hat{\mathbf{x}}} J \equiv \frac{\partial^2 J}{\partial \hat{\mathbf{x}}\,\partial \hat{\mathbf{x}}^T} = H^T H \text{ must be positive definite}$$

All eigenvalues > 0

- Least squares solution

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

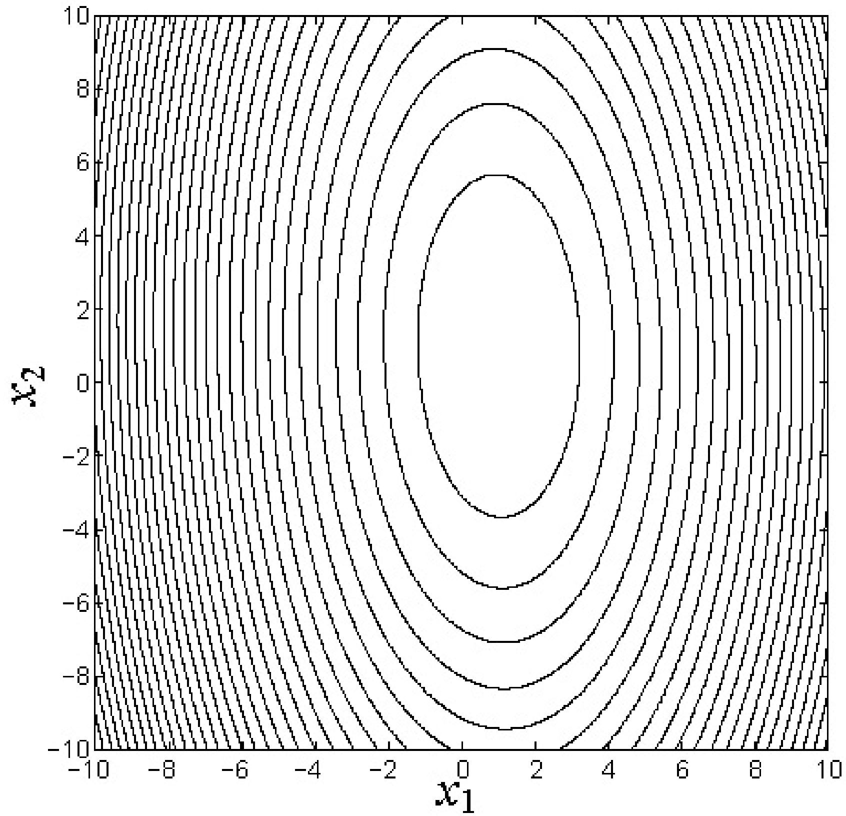
- Some observations
  - Matrix inverse is the same dimension as unknowns $m \times n$
  - Matrix inverse must exist: two conditions must be true
    - Number of unknowns must be less than or equal to number of measurements $m \geq n$
    - Must have at least n independent basis functions i.e. $H \neq \cos + \sin$
  - Note $H^T H$ must be positive definite from sufficient condition, which covers these two conditions $\lambda > 0$
  - For equal unknowns and equal measurements we simply have the following result $H^{-T} H^{-1} H^T = H^{-1}$

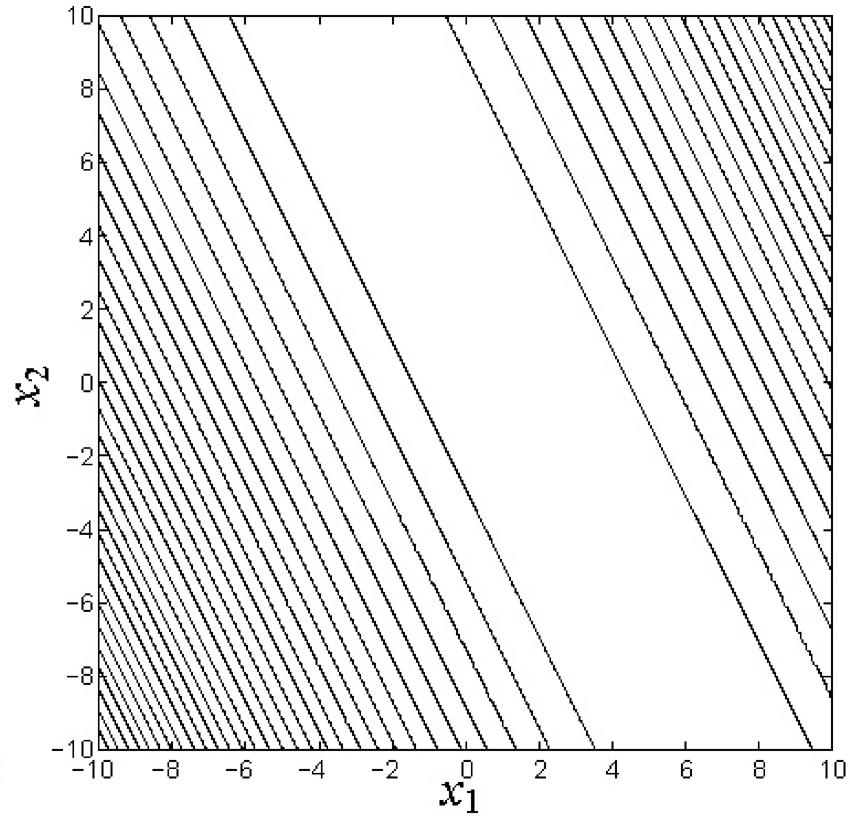
$$\hat{\mathbf{x}} = H^{-1} \tilde{\mathbf{y}}$$

    - This is for "square" systems $m = n$

Linear Least Squares (iii)

Contour Plots for Observable and Unobservable Solutions





Observable

$$H_1 = \begin{bmatrix} \sin t & 2\cos t \end{bmatrix}$$

Unobservable

$$H_2 = \begin{bmatrix} \sin t & 2\sin t \end{bmatrix}$$

linearly dependent

Last Experiment

- Measurements and model

Time	0	1	2	3	4	5
$\tilde{y}$	0.01	0.087	0.173	0.263	0.346	0.440

$$\tilde{y} = x \sin t + v$$

$$H = \begin{bmatrix} \sin 0 & \sin 1 & \sin 2 & \sin 3 & \sin 4 & \sin 5 \end{bmatrix}^T$$

$$H^T H = \sin^2 0 + \sin^2 1 + \sin^2 2 + \sin^2 3 + \sin^2 4 + \sin^2 5 = 0.016724$$

$$\begin{aligned} H^T \tilde{\mathbf{y}} &= 0.01 \sin 0 + 0.087 \sin 1 + 0.173 \sin 2 + 0.263 \sin 3 + 0.346 \sin 4 + 0.440 \sin 5 \\ &= 0.08380 \end{aligned}$$

$$\hat{\mathbf{x}} = (H^T H)^{-1} H^T \tilde{\mathbf{y}} = 59.79430758 \times 0.08380 = 5.011$$

- Not the same as the averaged approach result
- But no singularity for $t = 0$ (can use the first measurement)
  - More information used here than the averaged approach
- Note that we only need to take the inverse of a scalar here

Simple Example (i)

- Polynomial fitting problem
  - Say measurement is modeled by
  $$\tilde{y}_i = 0.3\sin(t_i) + 0.5\cos(t_i) + 0.1t_i + v_i$$
  - Generate measurements with time going from 0 to 10 seconds in 0.1 second intervals
    - Gives $m = 101$ total measurements
  - Measurement noise is Gaussian with variance given by 0.001 $= \sigma^2$
  - Assumed basis function matrix
  $$H = \begin{bmatrix} \sin(t_0) & \cos(t_0) & t_0 & \cos(t_0)\sin(t_0) & t_0^2 \\ \sin(t_1) & \cos(t_1) & t_1 & \cos(t_1)\sin(t_1) & t_1^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sin(t_m) & \cos(t_m) & t_m & \cos(t_m)\sin(t_m) & t_m^2 \end{bmatrix}$$
  - Note we have two "extra" basis functions
    - Estimated coefficients for these should be near zero

Simple Example (ii)

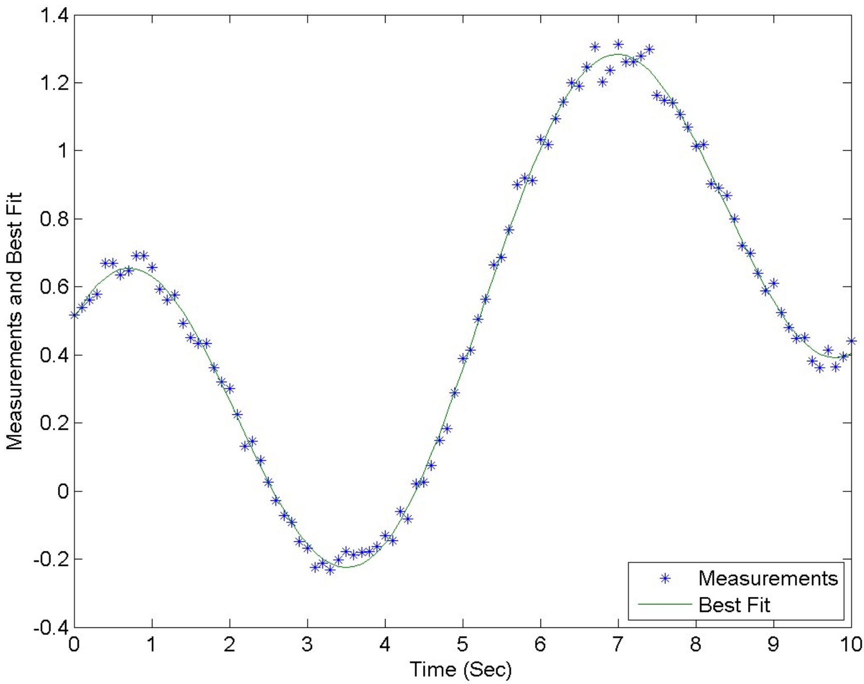
```
% Generate Measurements
t=[0:0.1:10]';m=length(t);
r=0.001;
ym=0.3*sin(t)+0.5*cos(t)+0.1*t+sqrt(r)*randn(m,1);

% Form Basis Function Matrix
h=[sin(t) cos(t) t cos(t).*sin(t) t.^2];

% Least Squares Estimate
xe=inv(h'*h)*h'*ym

% Estimated Output
ye=xe(1)*sin(t)+xe(2)*cos(t)+xe(3)*t+xe(4)*cos(t).*sin(t)+xe(5)*t.^2;

% Plot Results
plot(t,ym,'*',t,ye)
set(gca,'fontsize',12)
legend('Measurements','Best Fit','Location','SouthEast')
xlabel('Time (Sec)')
ylabel('Measurements and Best Fit')
```




```
xe =
0.3019
0.5072
0.1027
0.0012
-0.0003
```

Least Squares Example

- Estimate state parameters of simple linear system

excites all modes

$$y_{k+1} = \Phi y_k + \Gamma u_k$$

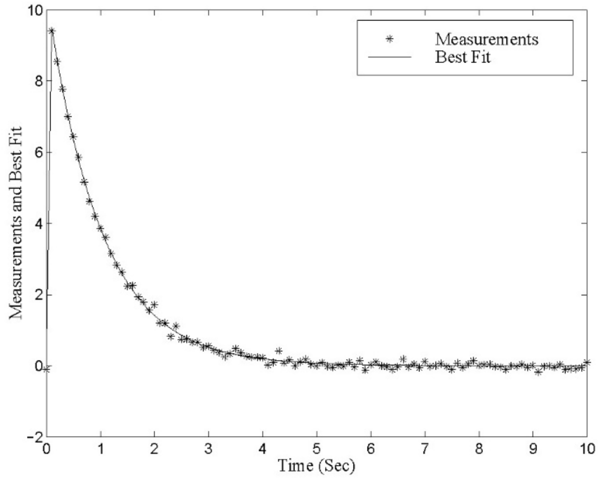
  - Impulse input, 101 measurements with $\Delta t = 0.1$ sec
  - Basis functions matrix given by

Noisy measurements are present here. Least squares is not really "optimum." Use Total Least Squares for optimum results.

$$H = \begin{bmatrix} \tilde{y}_1 & u_1 \\ \tilde{y}_2 & u_2 \\ \vdots & \vdots \\ \tilde{y}_{100} & u_{100} \end{bmatrix}, \quad \text{so} \quad \begin{bmatrix} \tilde{y}_2 \\ \tilde{y}_3 \\ \vdots \\ \tilde{y}_{101} \end{bmatrix} = H \begin{bmatrix} \hat{\Phi} \\ \hat{\Gamma} \end{bmatrix} + \begin{bmatrix} e_2 \\ e_3 \\ \vdots \\ e_{101} \end{bmatrix}$$

  - Results

$$\begin{bmatrix} \hat{\Phi} \\ \hat{\Gamma} \end{bmatrix} = \begin{bmatrix} 0.9048 \\ 0.0950 \end{bmatrix}, \quad \begin{bmatrix} \Phi \\ \Gamma \end{bmatrix} = \begin{bmatrix} 0.9048 \\ 0.0952 \end{bmatrix}$$



Another Example (i)

- Simple example $y = 3e^{-t}$
  - Choose e^{-t} as basis function
  - Measurements are

$$\tilde{y}_1(0) = 3, \quad \tilde{y}_2(3) = 0.15, \quad \text{with} \quad v_1 = v_2 = 0$$

  - Estimates

$$H = \begin{bmatrix} e^{-0} & e^{-3} \end{bmatrix}^T, \quad (H^T H)^{-1} = 0.9975, \quad H^T \tilde{\mathbf{y}} = 3.0075$$

$$\hat{x} = (0.9975)(3.0075) = 3$$

  - Add measurement noise $v_1 = v_2 = 0.01$

$$H^T \tilde{\mathbf{y}} = 3.018, \quad \hat{x} = (0.9975)(3.018) = 3.01 \quad \text{some error}$$

  - Say $v_1 = -0.0075$ and $v_2 = -0.15$

$$H^T \tilde{\mathbf{y}} = 3.0075, \quad \hat{x} = (0.9975)(3.0075) = 3$$

  - Got the right answer even with noise! What happened?
    - We got lucky, the noise cancelled the errors out

Another Example (ii)

- Choose $v_1 = v_2 = 0$ but choose e^{-2t} as basis function
- Estimates

$$H = \begin{bmatrix} e^{-0} & e^{-6} \end{bmatrix}^T = \begin{bmatrix} 1 & 0.0025 \end{bmatrix}^T, \quad (H^T H)^{-1} = 1$$

$$\hat{x} = (1)(3.0075) = 3.0075, \quad \text{looks good, right?}$$

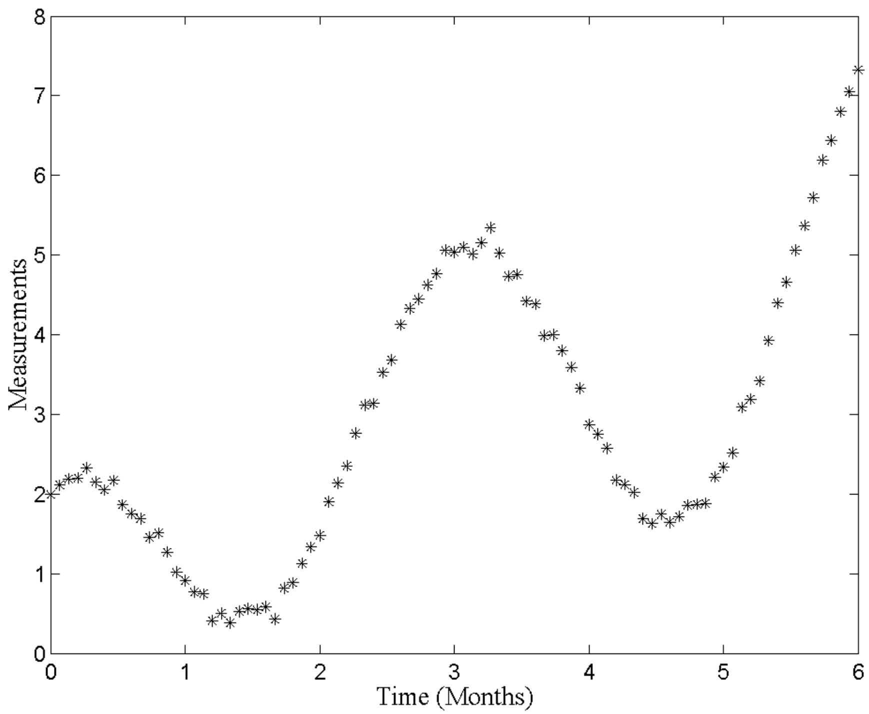
- Look at measurements and estimates at the two times

t	$\tilde{y}$	$\hat{y}$
0	3	3.0075
3	0.15	0.00745

  - First estimate is ok
  - Second one is way off from the measurement though, even through there are no measurement errors
  - This is due to the incorrect basis function
  - But, it is still the best estimate given the basis function
- Obviously, basis functions are important

Curve Fitting (i)

- Want to fit a model to the data shown below



- Assume two models

$$\text{Model 1:} \quad y_1(t) = c_1 t + c_2 \sin(t) + c_3 \cos(2t)$$

$$\text{Model 2:} \quad y_2(t) = d_1(t+2) + d_2 t^2 + d_3 t^3$$

- After applying least squares we find the following estimates for the coefficients of each model

$$(\hat{c}_1, \hat{c}_2, \hat{c}_3) = (0.9967, 0.9556, 2.0030)$$

$$(\hat{d}_1, \hat{d}_2, \hat{d}_3) = (0.6721, -0.1303, 0.0210)$$

- How well did we do?
  - Check residuals $e_1(t)$ and $e_2(t)$

$$\hat{y}_1(t) = \hat{c}_1 t + \hat{c}_2 \sin(t) + \hat{c}_3 \cos(2t)$$

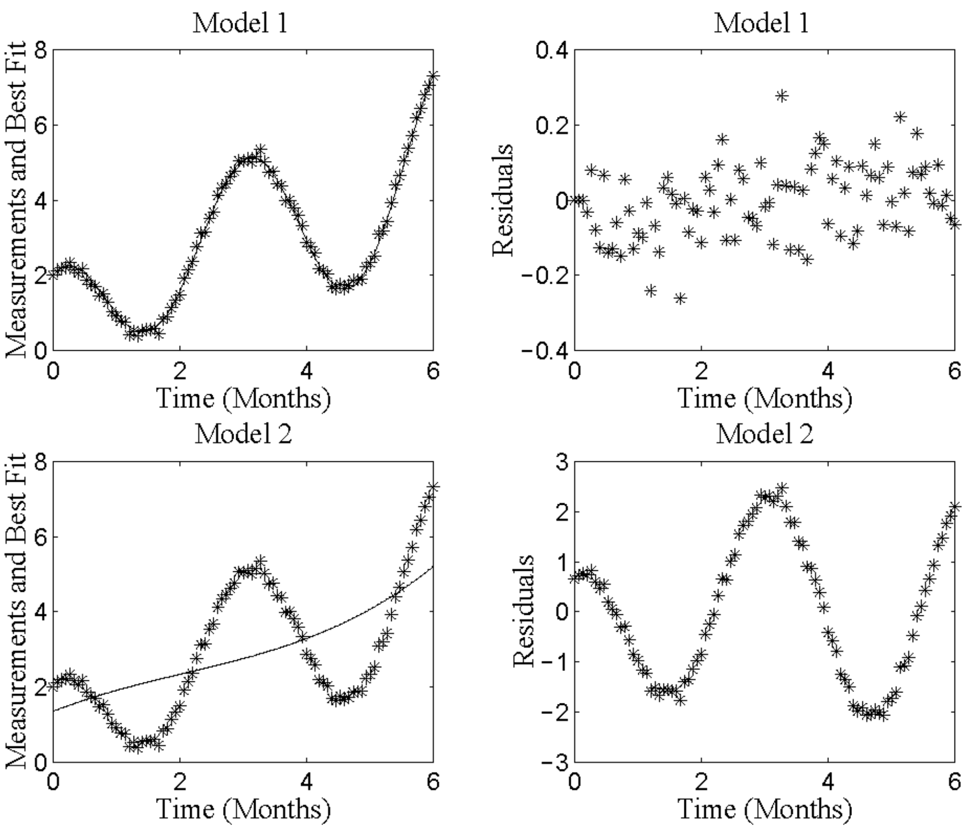
$$e_1(t) = \tilde{y}_1(t) - \hat{y}_1(t) \qquad \tilde{y}_1 = \tilde{y}_2$$

$$\hat{y}_2(t) = \hat{d}_1(t+2) + \hat{d}_2 t^2 + \hat{d}_3 t^3$$

$$e_2(t) = \tilde{y}_2(t) - \hat{y}_2(t)$$

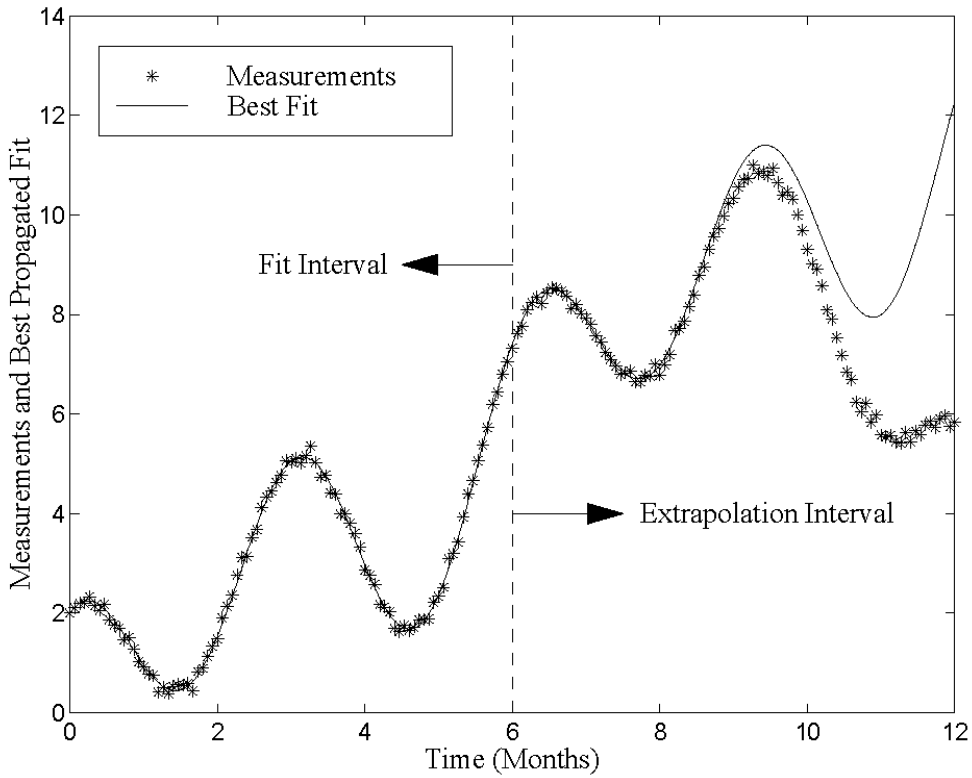
  - Note that these residuals can always be computed (i.e., they are known)

Curve Fitting (ii)



- Model 1 is able to obtain the best fit with the determined coefficients
  - Model 1 sample mean is 1×10^{-5} and the sample standard deviation is 0.0921
  - Model 2 sample mean is 1×10^{-5} and the sample standard deviation is 1.3856
- Model 1’s residual errors are “random” in appearance, while Model 2’s best fit failed to predict significant trends in the data
  - Having no reason to suspect that systematic errors are present in the measurements or in Model 1, we conclude that Model 1 is better

Extrapolation



- A plot of Model 1's predictions superimposed on the measured data over a 12-month period is shown to the left
- What happened?
  - The extrapolation interval matched the data for a while but then diverges
  - The synthetic measurements were actually generated using

$$\tilde{y}(t) = t + \sin(t) + 2\cos(2t) - \frac{0.4e^t}{1 \times 10^4} + v(t)$$

  - Model used for least-squares fitting only included the first three terms

Weighted Least Squares (i)

- Suppose we wish to provide more “trust” in one measurement versus another
  - For example, for attitude determination, Sun sensor is more accurate than a three-axis magnetometer (TAM) so we wish to weight Sun sensor more than TAM
  - Use weighted least squares
  - Minimize the following

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

  where W is the $m \times m$ weighting matrix (does not need to be strictly diagonal, but it is symmetric in general)
    - Note this weights each measurement
  - Residual is again given by

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

Weighted Least Squares (ii)

- Necessary and sufficient conditions become
  - Note it is assumed here that W is symmetric

necessary condition

$$\nabla_{\hat{\mathbf{x}}} J = H^T W H \hat{\mathbf{x}} - H^T W \tilde{\mathbf{y}} = \mathbf{0}$$

sufficient condition

$$\nabla^2_{\hat{\mathbf{x}}} J = H^T W H \text{ must be positive definite}$$

- Leads to weighted least squares solution

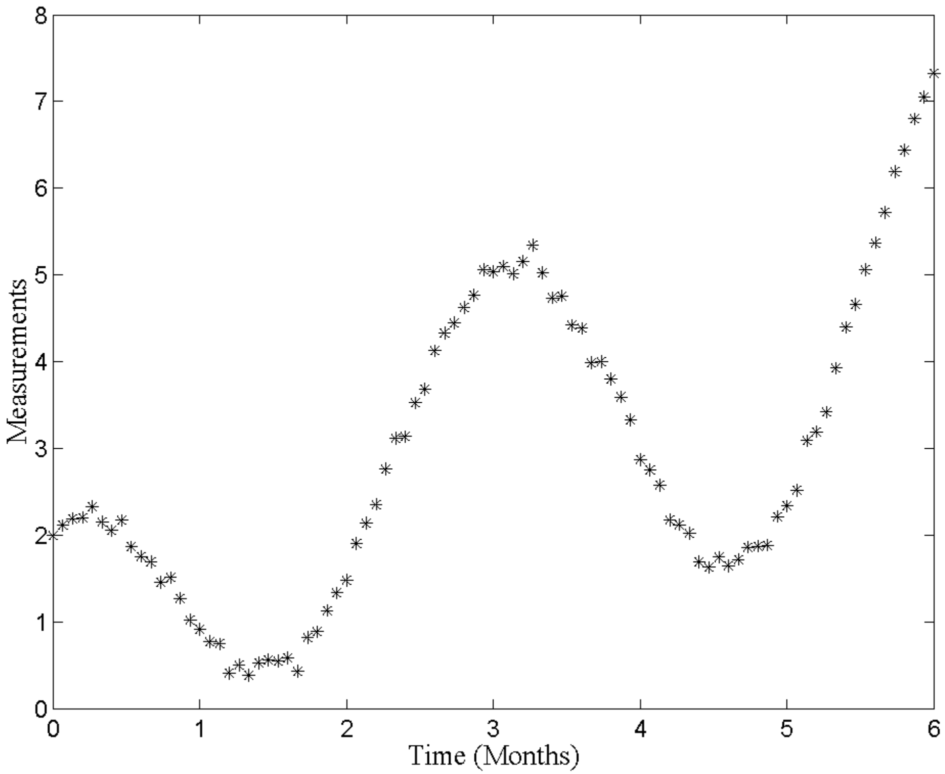
$$\boxed{\hat{\mathbf{x}} = (H^T W H)^{-1} H^T W \tilde{\mathbf{y}}}$$

If $W = I$, then you get least squares (unweighted)

- From sufficient condition we see that W must be positive definite, otherwise it violates the sufficient condition
- Optimal method to choose weight discussed later

Example (i)

- Employ a subset of 31 measurements from the 91 measurements shown below



- Here, the first three measurements are known to contain smaller measurement errors than the remaining measurements
- Toward this end, the structure of the weighting matrix now becomes

$$W = \mathrm{diag}\begin{bmatrix} w & w & w & 1 & \cdots & 1 \end{bmatrix}$$

- If $w = 1$ then the standard least squares solution is given
  - In fact any scalar times the identity matrix for W produces the same solution as standard (un-weighted) least squares

Example (ii)

w	$\hat{\mathbf{x}}$	constraint residual norm
1×10^0	$(1.0278, 0.8750, 1.9884)$	3.21×10^{-2}
1×10^1	$(1.0388, 0.8675, 2.0018)$	1.17×10^{-2}
1×10^2	$(1.0258, 0.8923, 2.0049)$	7.87×10^{-3}
1×10^5	$(0.9047, 1.0949, 2.0000)$	5.91×10^{-5}
1×10^7	$(0.9060, 1.0943, 2.0000)$	1.10×10^{-5}
1×10^{10}	$(0.9932, 1.0068, 2.0000)$	4.55×10^{-7}
1×10^{15}	$(0.9970, 1.0030, 2.0000)$	0.97×10^{-9}

- One can see that the residual constraint error (i.e., the computed norm of the measurements minus the estimates for the first three observations) decreases as more weight is used
  - However, this does not generally guarantee that the estimates are closer to their true values of 1, 1, and 2
  - The interaction of the basis function plays an important role
  - Still, if the weight is sufficiently large, the estimates are indeed closer to their true values, as expected
  - In this simulation, the first three measurements were obtained with no measurement errors

Constrained Least Squares (i)

- Suppose the original observations in partition naturally into the sub-systems

$$\begin{bmatrix} \tilde{\mathbf{y}}_1 \\ \cdots \\ \tilde{\mathbf{y}}_2 \end{bmatrix} = \begin{bmatrix} H_1 \\ \cdots \\ H_2 \end{bmatrix} \hat{\mathbf{x}} + \begin{bmatrix} \mathbf{e}_1 \\ \cdots \\ \mathbf{0} \end{bmatrix} \qquad \text{or} \qquad \begin{aligned} \tilde{\mathbf{y}}_1 &= H_1 \hat{\mathbf{x}} + \mathbf{e}_1 \\ \tilde{\mathbf{y}}_2 &= H_2 \hat{\mathbf{x}} \end{aligned}$$

where

$$\begin{aligned} \tilde{\mathbf{y}}_1 &= \text{an } m_1 \times 1 \text{ vector of measured } y\text{-values} \\ H_1 &= \text{an } m_1 \times n \text{ basis function matrix corresponding} \\ &\quad\ \text{with the measured } y\text{-values} \\ \mathbf{e}_1 &= \text{an } m_1 \times 1 \text{ vector of residual errors} \\ \tilde{\mathbf{y}}_2 &= \text{an } m_2 \times 1 \text{ vector of perfectly measured } y\text{-values} \\ H_2 &= \text{an } m_2 \times n \text{ basis function matrix corresponding} \\ &\quad\ \text{with the perfectly measured } y\text{-values} \end{aligned}$$

- Further assume that $n \geq m_2$ and $n \leq m_1$

overconstrained

Constrained Least Squares (ii)

- The absence of the residual error matrix $\mathbf{e}_2$ reflects the fact that $H_2\mathbf{x}$ is required to equal the second measurement set exactly
- Thus, we can formulate the problem as a constrained minimization problem by minimizing (with an added weight matrix here)

$$J = \frac{1}{2}\mathbf{e}_1^T W_1 \mathbf{e}_1 = \frac{1}{2}(\tilde{\mathbf{y}}_1 - H_1\hat{\mathbf{x}})^T W_1 (\tilde{\mathbf{y}}_1 - H_1\hat{\mathbf{x}})$$

subject to

$$\tilde{\mathbf{y}}_2 - H_2\hat{\mathbf{x}} = \mathbf{0} \qquad \text{(constraint)}$$

- Use Lagrange multiplier approach to form augmented function

$$J = \frac{1}{2}\left[\tilde{\mathbf{y}}_1^T W_1 \tilde{\mathbf{y}}_1 - 2\tilde{\mathbf{y}}_1^T W_1 H_1 \hat{\mathbf{x}} + \hat{\mathbf{x}}^T (H_1^T W_1 H_1)\hat{\mathbf{x}}\right] + \boldsymbol{\lambda}^T(\tilde{\mathbf{y}}_2 - H_2\hat{\mathbf{x}})$$

Constrained Least Squares (iii)

- Necessary conditions

$$\nabla_{\hat{\mathbf{x}}} J = -H_1^T W_1 \tilde{\mathbf{y}}_1 + (H_1^T W_1 H_1)\hat{\mathbf{x}} - H_2^T \boldsymbol{\lambda} = \mathbf{0} \qquad (1)$$

$$\nabla_{\boldsymbol{\lambda}} J = \tilde{\mathbf{y}}_2 - H_2 \hat{\mathbf{x}} = \mathbf{0}, \quad \rightarrow \tilde{\mathbf{y}}_2 = H_2 \hat{\mathbf{x}} \qquad (2)$$

- Solving Eq. (1) for the estimate gives

$$\hat{\mathbf{x}} = (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{\mathbf{y}}_1 + (H_1^T W_1 H_1)^{-1} H_2^T \boldsymbol{\lambda}$$

- Substituting this $\tilde{x}$ equation into Eq. (2) and solving for $\boldsymbol{\lambda}$ gives

$$\boldsymbol{\lambda} = \left[H_2 (H_1^T W_1 H_1)^{-1} H_2^T\right]^{-1} \left[\tilde{\mathbf{y}}_2 - H_2 (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{\mathbf{y}}_1\right]$$

- Substituting this equation into the estimate equation gives

$$\boxed{\hat{\mathbf{x}} = \bar{\mathbf{x}} + K(\tilde{\mathbf{y}}_2 - H_2 \bar{\mathbf{x}})} \qquad (3)$$

update form

where

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

$$\bar{\mathbf{x}} = (H_1^T W_1 H_1)^{-1} H_1^T W_1 \tilde{\mathbf{y}}_1 \qquad (4)$$

unconstrained solution

Constrained Least Squares (iv)

- Note that Eq. (4) is least squares estimate of $\mathbf{x}$ in the absence of the constraint equations
- The second term in Eq. (3) is an additive correction in which the "gain matrix" K multiplies the constraint residual prior to the correction
  - This general "update form" in Eq. (3) is seen often in estimation theory and is therefore an important result
- Note that a weighted least squares approach with "large" weights on the perfect measurements can be used instead, which gives numerically the same solution as constrained least squares

set to ∞

- If $m_2 = n$ then H_2 is square, which leads directly to

$$K = H_2^{-1}, \quad \text{and} \quad \hat{\mathbf{x}} = H_2^{-1}\tilde{\mathbf{y}}_2$$

  - This shows that the solution is dependent on the perfectly measured values and H_2 only, which is the same result obtained using a square H matrix in the standard least squares solution
  - Thus, the solution is unaffected by an arbitrary number of erroneous measurements (we don't need them, which makes sense)

- Employ a subset of 31 measurements from the previous weighted least squares example
  - Study three cases for various perfect measurement cases

$$
\begin{aligned}
&\text{case 1:} \quad \tilde{\mathbf{y}}_1 = \begin{bmatrix}\tilde{y}_2 & \tilde{y}_3 & \cdots & \tilde{y}_{31}\end{bmatrix}^T, \quad \tilde{\mathbf{y}}_2 = y_1 \\
&\text{case 2:} \quad \tilde{\mathbf{y}}_1 = \begin{bmatrix}\tilde{y}_3 & \tilde{y}_4 & \cdots & \tilde{y}_{31}\end{bmatrix}^T, \quad \tilde{\mathbf{y}}_2 = \begin{bmatrix}y_1 & y_2\end{bmatrix}^T \\
&\text{case 3:} \quad \tilde{\mathbf{y}}_1 = \begin{bmatrix}\tilde{y}_4 & \tilde{y}_5 & \cdots & \tilde{y}_{31}\end{bmatrix}^T, \quad \tilde{\mathbf{y}}_2 = \begin{bmatrix}y_1 & y_2 & y_3\end{bmatrix}^T
\end{aligned}
$$

  - Results are given below

case	$\bar{\mathbf{x}}$	$\hat{\mathbf{x}}$
1	$(1.0261, 0.8766, 1.9869)$	$(1.0406, 0.8629, 2.0000)$
2	$(1.0233, 0.8789, 1.9840)$	$(0.9039, 1.0901, 2.0000)$
3	$(1.0192, 0.8820, 1.9793)$	$(0.9970, 1.0030, 2.0000)$

Example (ii)

- We see that when one perfect measurement is used (case 1), the solution is not substantially improved over conventional least squares since $\bar{\mathbf{x}} \approx \hat{\mathbf{x}}$
- However, when two perfect measurements are used (case 2), the estimates are closer to their true values
- When three perfect measurements are used (case 3), which implies that $m_2 = n$, the estimates are even closer to their true values
  - In fact, the estimates are identical within several significant digits to the case of $w = 1 \times 10^{15}$ in the weighted least squares example
  - Shows what we discussed before about using weighted least squares instead of constrained least squares to obtain nearly identical solutions
  - Note, were it not for the unaccounted error term $-0.4e^t/1\times10^4$ in the simulated measurements, these would be found to agree exactly with the true coefficients $(1, 1, 2)$