Lecture 6

Least Squares

Ye Ding (丁 烨)

Email: y.ding@sjtu.edu.cn

School of Mechanical Engineering

Shanghai Jiao Tong University

Least Squares

- **References for Least Squares**

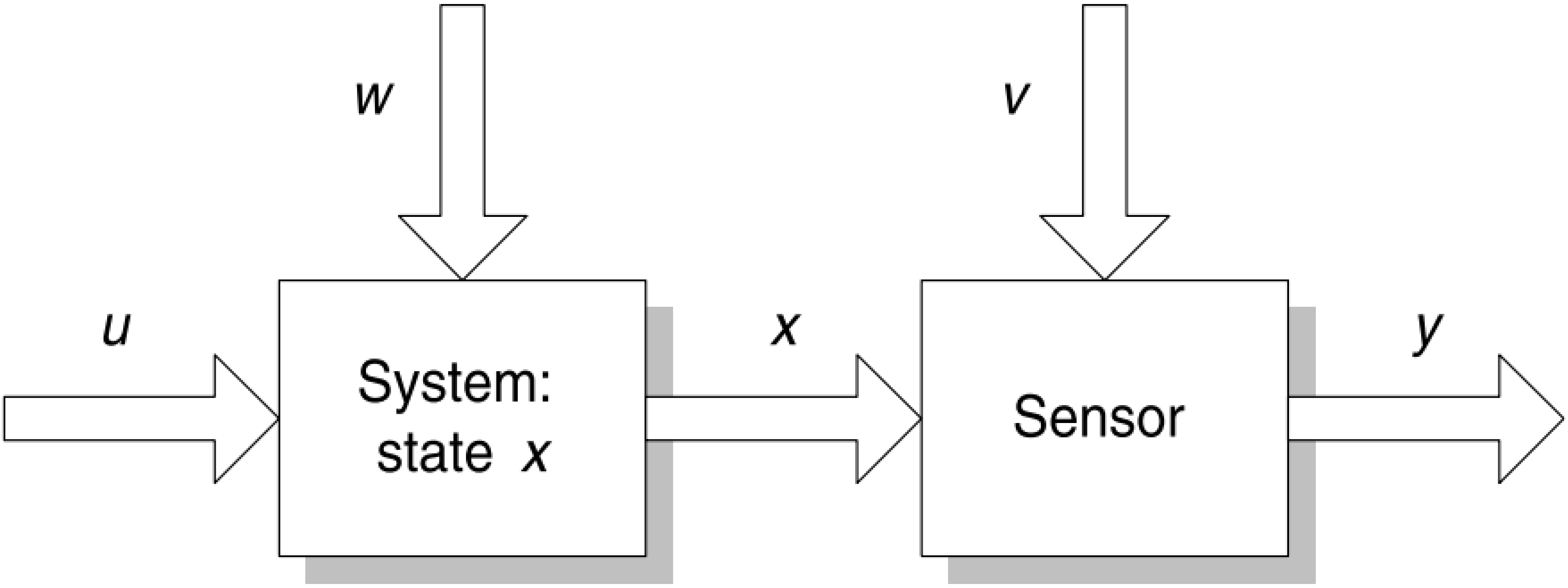
[1] Cleve Moler, Numerical Computing with MATLAB, Society for Industrial and Applied Mathematics, 2004. Chapter 5

[2] Timothy Sauer, Numerical analysis (2nd ed.), Pearson Education, 2012. Chapter 4

[3] Stephen Boyd & Lieven Vandenberghe, Introduction to Applied Linear Algebra, Cambridge University Press, 2018.

Least Squares

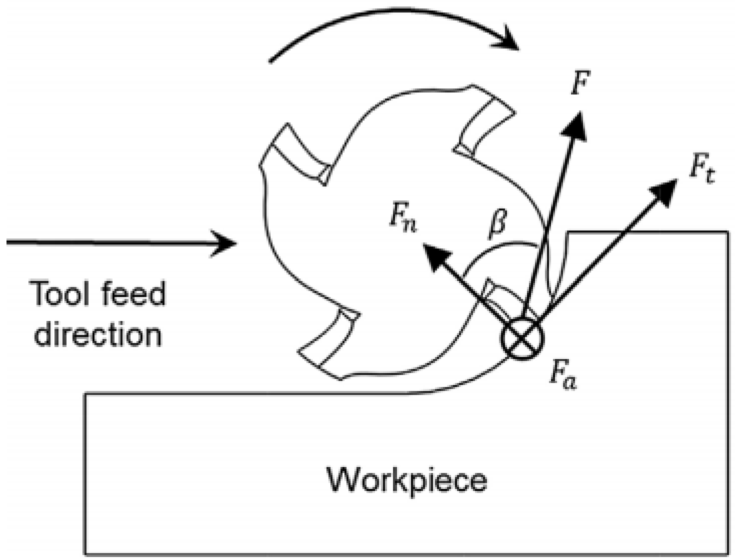
- **Motivation: System Parameter Identification**

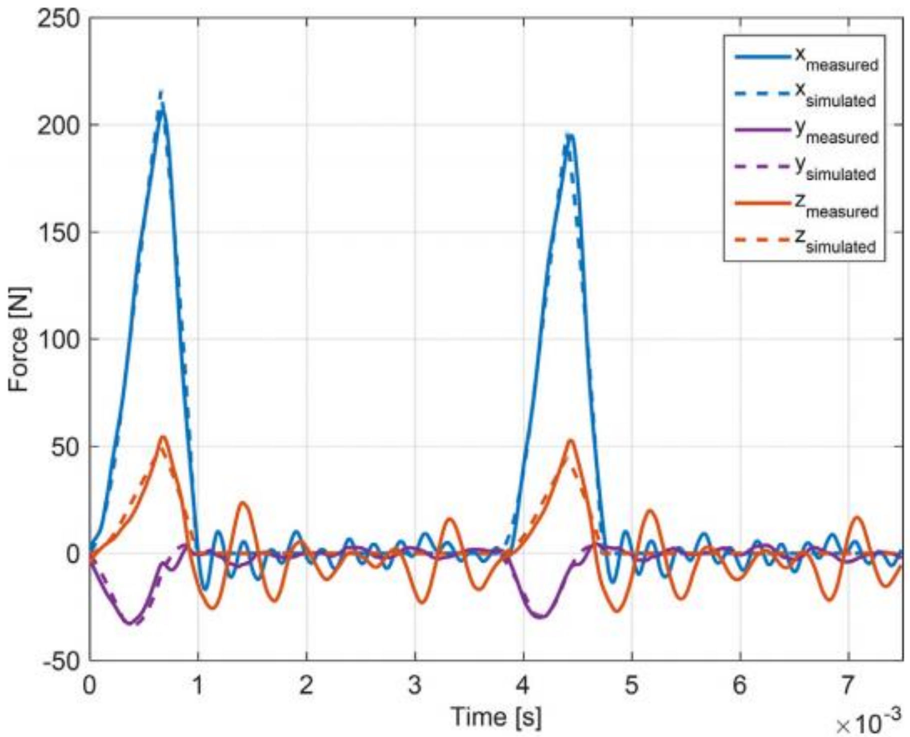


$$G(s) = \frac{K_M}{a_2 s^2 + a_1 s + a_0} e^{-sh}$$

Least Squares

- **Motivation: Identification of force model coefficients**





$$\bar{F}_x = \left\{ \frac{N_t b f_t}{8\pi} \left[-k_{tc} \cos(2\phi) + k_{nc} (2\phi - \sin(2\phi)) \right] \right.$$

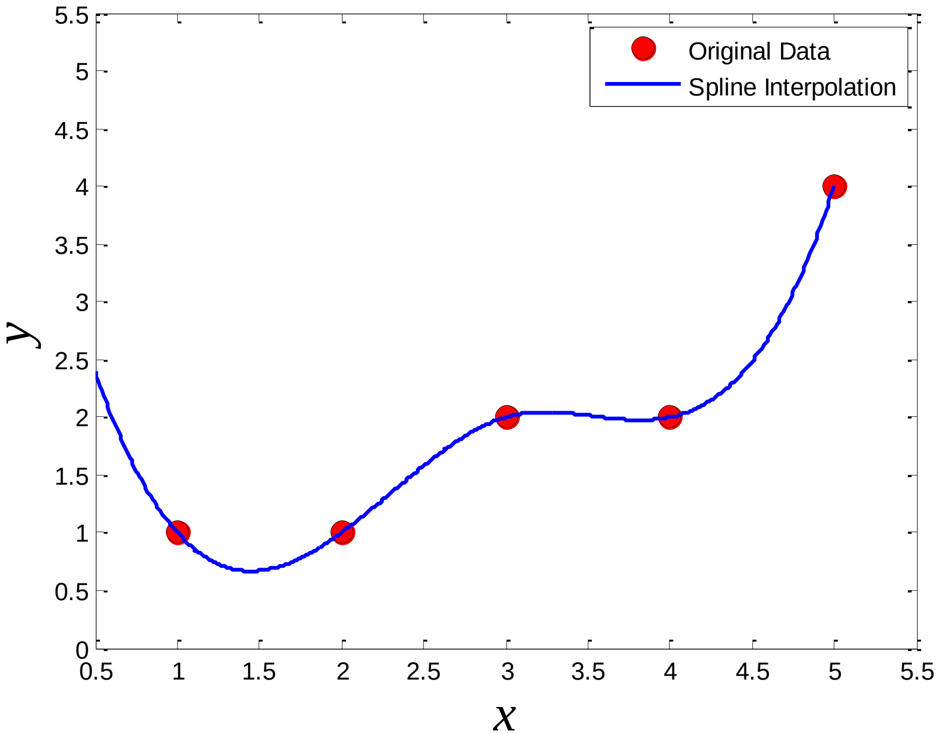
$$\left. + \frac{N_t b}{2\pi} \left[k_{te} \sin(\phi) - k_{ne} \cos(\phi) \right] \right\}_{\phi_s}^{\phi_e} \Rightarrow k_{tc},\ k_{nc},\ k_{te},\ k_{ne}$$

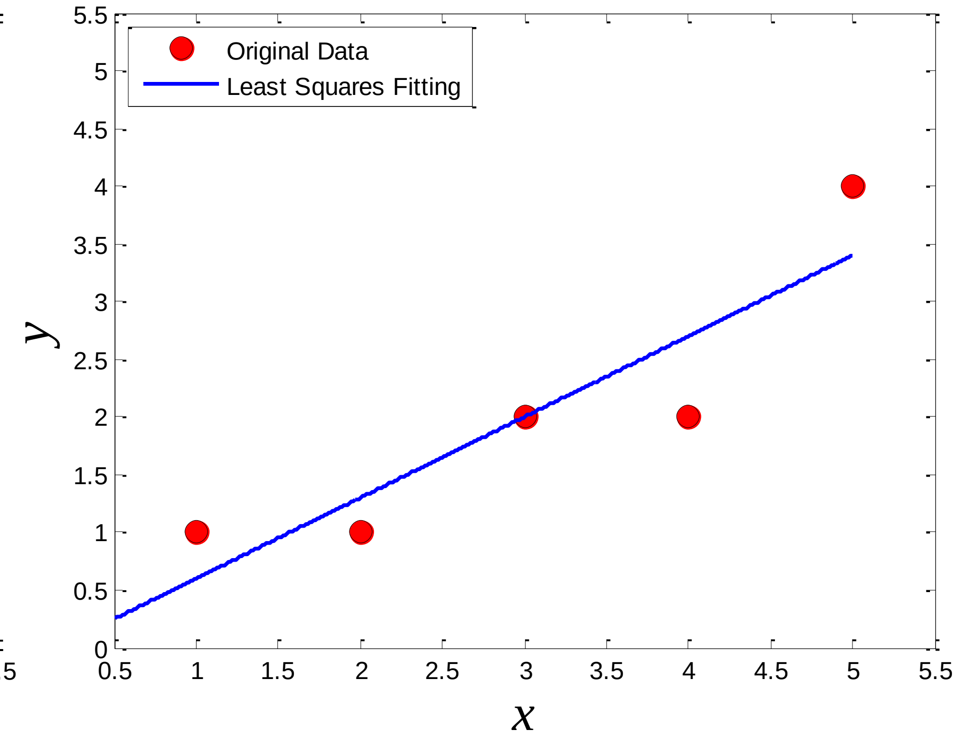
Least Squares

- **Motivation: Interpolation VS. Least Squares Fitting**

Consider the experimental data below and find a way to best represent it:

x	1	2	3	4	5
y	1	1	2	2	4





Least Squares

- **Motivation: Interpolation VS. Least Squares Fitting**
  - **Interpolation means fitting some function to given data so that the function has the same values as the given data.**
  - **Least Squares Fitting is to derive a single curve that represents the general trend of the data. The curve is designed to follow *the pattern of the points taken as a group*. Any individual data point may be incorrect, we make no effort to intersect every point.**

Least Squares

- **Linear Least Squares**
  - **The Normal Equation**
  - **QR Factorization**
- **Nonlinear Least Squares**
  - **Gauss–Newton Method**
  - **Levenberg–Marquardt Method**

Linear Least Squares

The Normal Equation: Basic Idea

Consider the following three equations in two unknowns:

$$x_1 + x_2 = 2$$

$$x_1 - x_2 = 1$$

$$x_1 + x_2 = 3$$

$$\Downarrow \quad Ax = b$$

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Linear Least Squares

The Normal Equation: Basic Idea

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}}_{\mathbf{b}}$$

$$\Updownarrow$$

$$\boxed{x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \triangleq \mathbf{b}$$

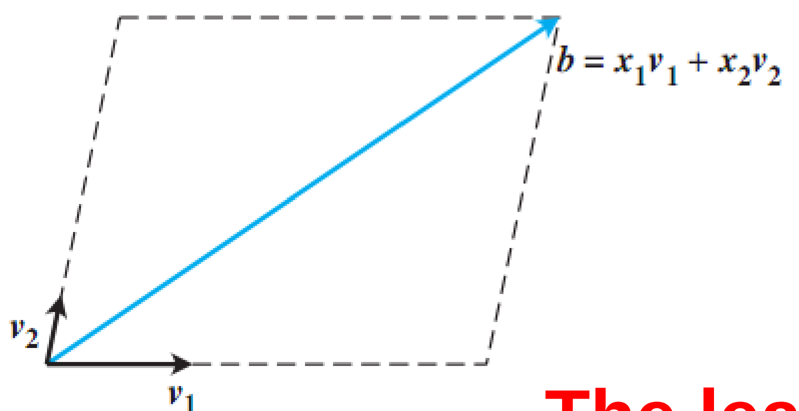
Combinations of two three-dimensional vectors

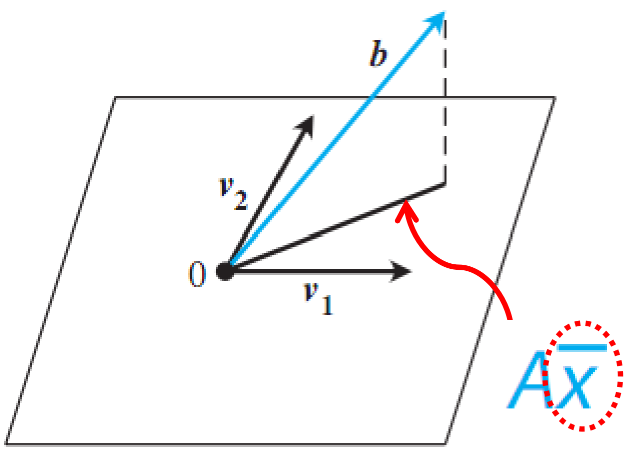
Linear Least Squares

The Normal Equation: Basic Idea

$$\boxed{x_1 \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{v}_1} + x_2 \underbrace{\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}}_{\mathbf{v}_2}} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \triangleq \mathbf{b}$$

Combinations of two three-dimensional vectors v_1 and v_2





The least squares solution

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Linear Least Squares

- **The Normal Equation: Basic Idea**

Search for a formula for $\overline{x}$

$$(b - A\overline{x}) \perp \{Ax \mid x \in R^n\}$$

$$\Downarrow$$

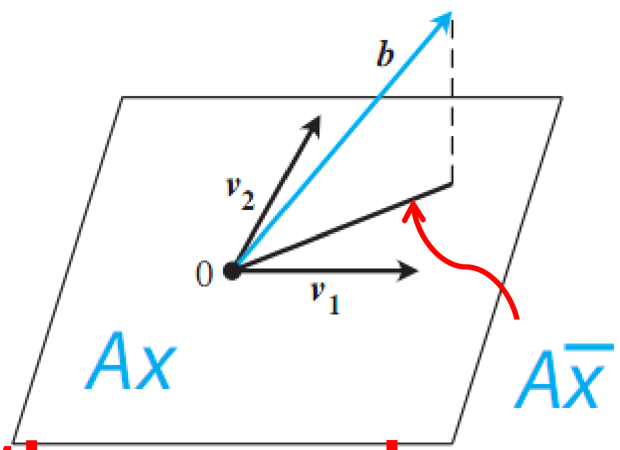
$$(Ax)^T(b - A\overline{x}) = 0 \text{ for all } x \text{ in } R^n$$

$$\Downarrow$$

$$A^T(b - A\overline{x}) = 0$$

$$\Downarrow$$

$$\boxed{A^T A\overline{x} = A^T b}$$



the normal equations

Linear Least Squares

The Normal Equation: Example 1

Consider the following three equations in two unknowns:

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3 \end{aligned} \Rightarrow A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

Linear Least Squares

The Normal Equation: Example 1

Consider the following three equations in two unknowns:

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3 \end{aligned} \Rightarrow A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

The normal equations:

$$\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix} \Rightarrow \overline{x} = \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix}$$

Linear Least Squares

The Normal Equation: Example 1

Substituting the least squares solution into the original problem yields:

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} 2.5 \\ 1 \\ 2.5 \end{bmatrix} \neq \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

The residual of the least squares solution:

$$r = b - A\overline{x} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 1 \\ 2.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0.0 \\ 0.5 \end{bmatrix}$$

Linear Least Squares

- **The Normal Equation: the size of the residual**

The 2-norm of the residual vector:

$$||r||_2 = \sqrt{r_1^2 + \cdots + r_m^2}$$

The squared error:

$$\mathrm{SE} = r_1^2 + \cdots + r_m^2$$

The root mean squared error:

$$\mathrm{RMSE} = \sqrt{\mathrm{SE}/m} = \sqrt{\left(r_1^2 + \cdots + r_m^2\right)/m}$$

Linear Least Squares

- **The Normal Equation: the size of the residual**

The 2-norm of the residual vector:

$$||r||_2 = \sqrt{r_1^2 + \cdots + r_m^2}$$

The least squares solution of a linear system of equations *Ax* = *b* minimizes the Euclidean norm of the residual

$$||Ax - b||_2$$

Linear Least Squares

- **The Normal Equation: General Procedure**

Given a set of *m* data points $(t_1, y_1), \dots, (t_m, y_m)$.

STEP 1. Choose a model. Identify a parameterized model, such as $y = c_1 + c_2 t$, $y = c_1 + c_2 t + c_3 t^2$

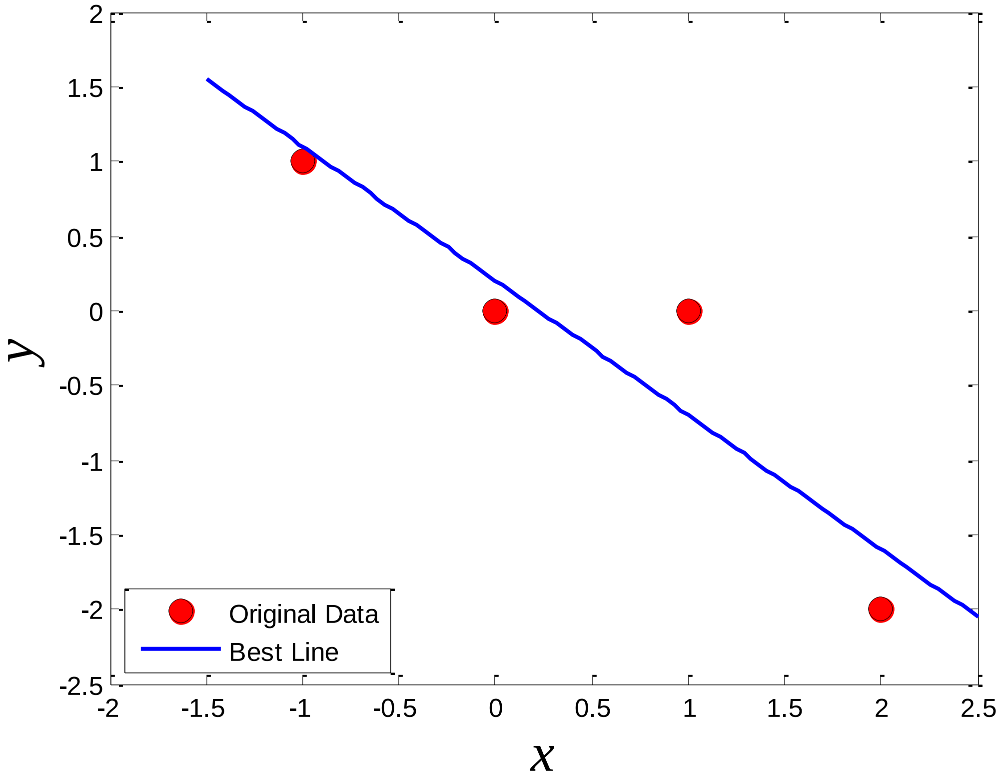
STEP 2. Force the model to fit the data. Substitute the data points into the model, resulting in a system $Ax = b$

STEP 3. Solve the normal equations. The least squares solution will be found as the solution to the system of normal equations $A^T A x = A^T b$

Linear Least Squares

- **The Normal Equation: Example 2**

Find the best line and best parabola for the four data points (−1,1), (0,0), (1,0), (2,−2).



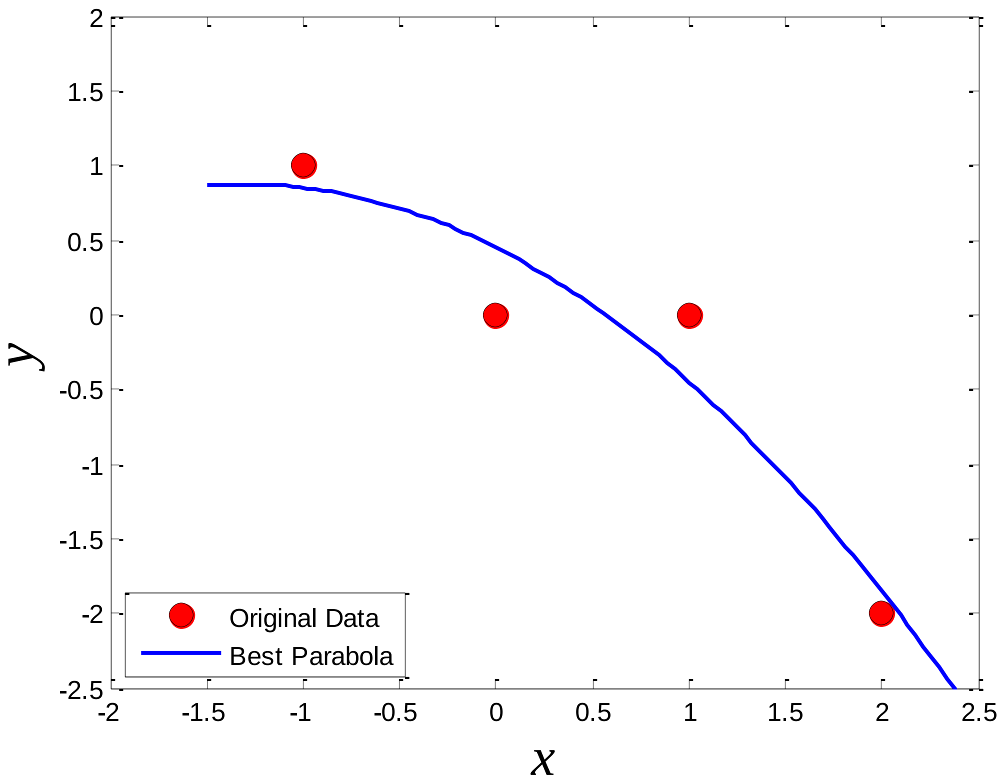
$$y = c_1 + c_2 t$$
$$= 0.2 - 0.9t$$

The squared error:
SE = 0.7

Linear Least Squares

The Normal Equation: Example 2

Find the best line and best parabola for the four data points (−1,1), (0,0), (1,0), (2,−2).



$$y = c_1 + c_2 t + c_3 t^2$$
$$= \mathbf{0.45 - 0.65}\boldsymbol{t} - \boldsymbol{0.25t^2}$$

The squared error:
SE = 0.45

Linear Least Squares

- **The Normal Equation: Example 3**

Let $x_1 = 2.0$, $x_2 = 2.2$, $x_3 = 2.4$,..., $x_{11} = 4.0$ be equally spaced points in [2,4], and set

$$y_i = 1 + x_i + x_i^2 + x_i^3 + x_i^4 + x_i^5 + x_i^6 + x_i^7$$

for $1 \le i \le 11$.

Use the normal equations to find the least squares polynomial

$$P(x) = c_1 + c_2 x + \cdots + c_8 x^7$$

fitting the (x_i, y_i).

Linear Least Squares

The Normal Equation: Example 3

x_NE = (A'*A)\(A'*b)

x_backslash = A\b

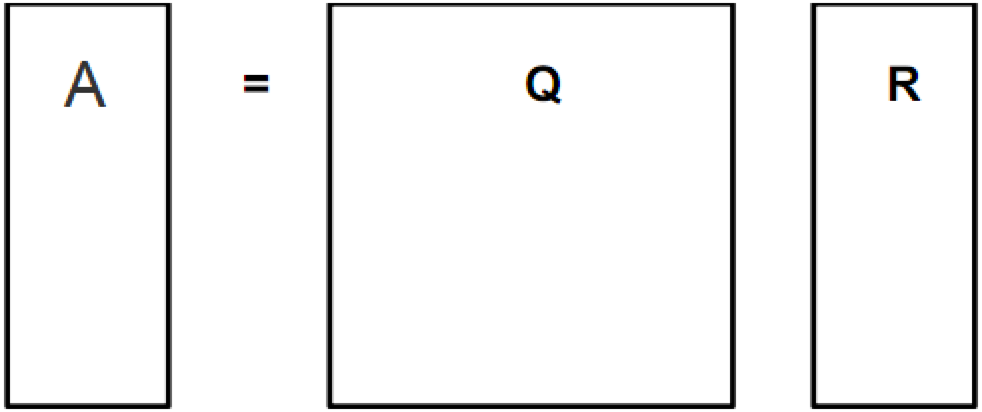
```
x_NE =
  4.985448776320641
 -8.842392139312931
 11.312052465639942
 -4.942514793003778
  3.034575762766727
  0.586051762312912
  1.046351797098351
  0.997795828974419
```

```
x_backslash =
  1.000000015989120
  0.999999965715620
  1.000000030766503
  0.999999985089027
  1.000000004186126
  0.999999999326902
  1.000000000056217
  0.999999999998201
```

Linear Least Squares

- **QR Factorization: Basic Idea**

To find the least squares solution without forming the normal equations.



R is the same size as A and Q is a square matrix with as many rows as A. The letter Q is a substitute for the letter O in orthogonal and the letter R is for right triangular matrix.

Linear Least Squares

- **QR Factorization: Basic Idea**

To find the least squares solution without forming the normal equations.

To minimize $||Ax - b||_2$

$$\Downarrow$$

$$||QRx - b||_2$$

$$||Qx||_2 = ||x||_2 \dashrightarrow \Downarrow$$

$$||Rx - Q^T b||_2$$

QR Factorization

- **Least squares by QR factorization: Procedure**

Given the $m \times n$ inconsistent system

$$Ax = b,$$

find the full QR factorization $A = QR$ and set

$$\begin{aligned}\hat{R} &= \text{upper } n \times n \text{ submatrix of } R \\ \hat{d} &= \text{upper } n \text{ entries of } d = Q^T b\end{aligned}$$

Solve $\hat{R}\overline{x} = \hat{d}$ for least squares solution $\overline{x}$.

QR Factorization

- **Least squares by QR factorization: Example 3**

Let x_1 = 2.0, x_2 = 2.2, x_3 = 2.4,..., x_{11} = 4.0 be equally spaced points in [2,4], and set

$$y_i = 1 + x_i + x_i^2 + x_i^3 + x_i^4 + x_i^5 + x_i^6 + x_i^7$$

for $1 \le i \le 11$.

Use the normal equations to find the least squares polynomial

$$P(x) = c_1 + c_2 x + \cdots + c_8 x^7$$

fitting the (x_i, y_i).

QR Factorization

- **Least squares by QR factorization: Example 3**

x_NE = (A'*A)\(A'*b)

[Q,R]=qr(A);d = Q' * b;
x_qr = R(1:8,:)\d(1:8)

x_NE =	x_qr =
4.985448776320641	0.999999837023661
-8.842392139312931	1.000000399264445
11.312052465639942	0.999999585059873
-4.942514793003778	1.000000237157606
3.034575762766727	0.999999919481018
0.586051762312912	1.000000016242281
1.046351797098351	0.999999998197170
0.997795828974419	1.000000000084963

QR Factorization

- **How QR factorization? Basic Idea**

$$(A_1|\cdots|A_n) = (q_1|\cdots|q_m)\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$

A **is** $\boldsymbol{m \times n}$

Q is an orthogonal square matrix with $\boldsymbol{m \times m}$

R is the upper triangular matrix with $\boldsymbol{m \times n}$

QR Factorization

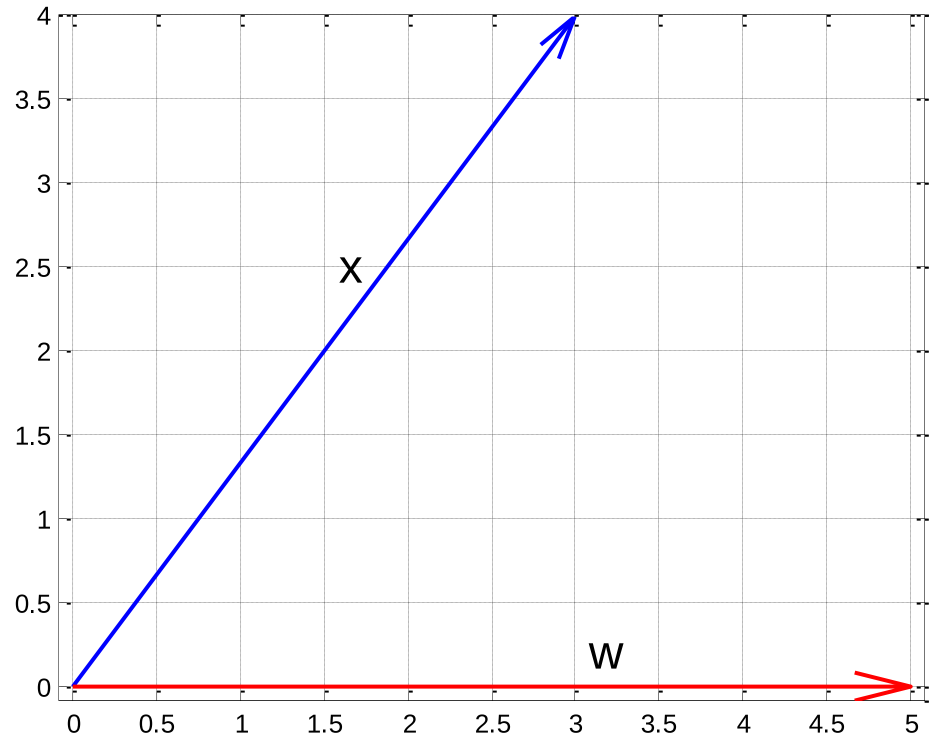
How QR factorization? A Simple Example

Given a vector *x* in the plane, to relocate it to a vector *w* of equal length, e.g.,

$$x = [3, 4] \text{ and } w = [5, 0]$$

How to find a matrix *H*, such that

$$Hx = w$$



QR Factorization

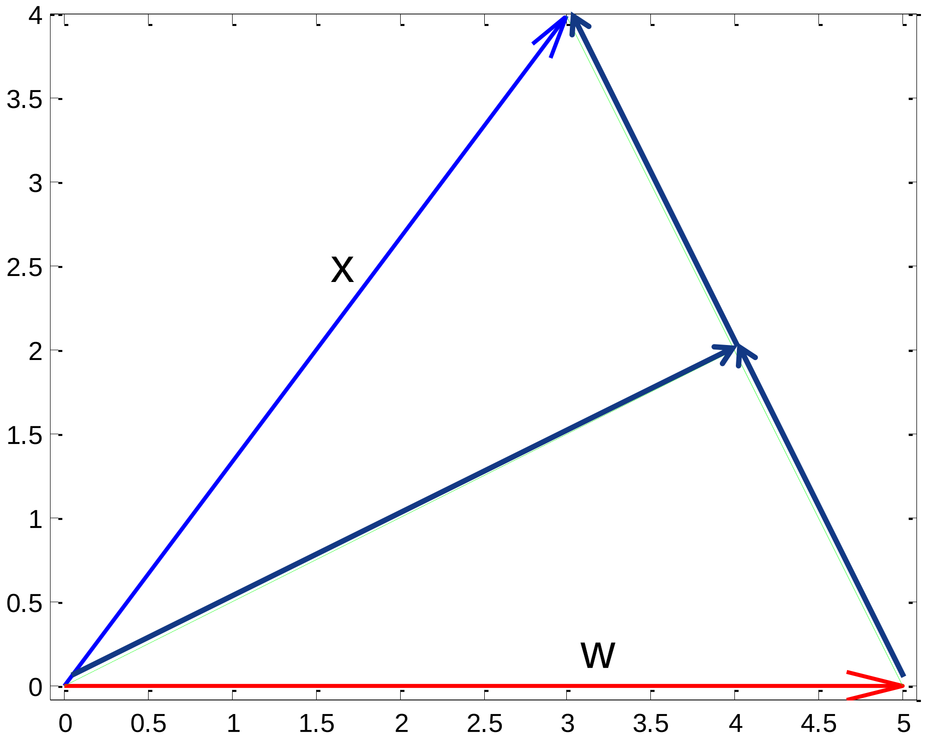
- **How QR factorization? A Simple Example**

Given a vector x in the plane, to relocate it to a vector w of equal length, e.g.,

$$x = [3, 4] \text{ and } w = [5, 0]$$

How to find a matrix H, such that

$$Hx = w$$



QR Factorization

How QR factorization? A Simple Example

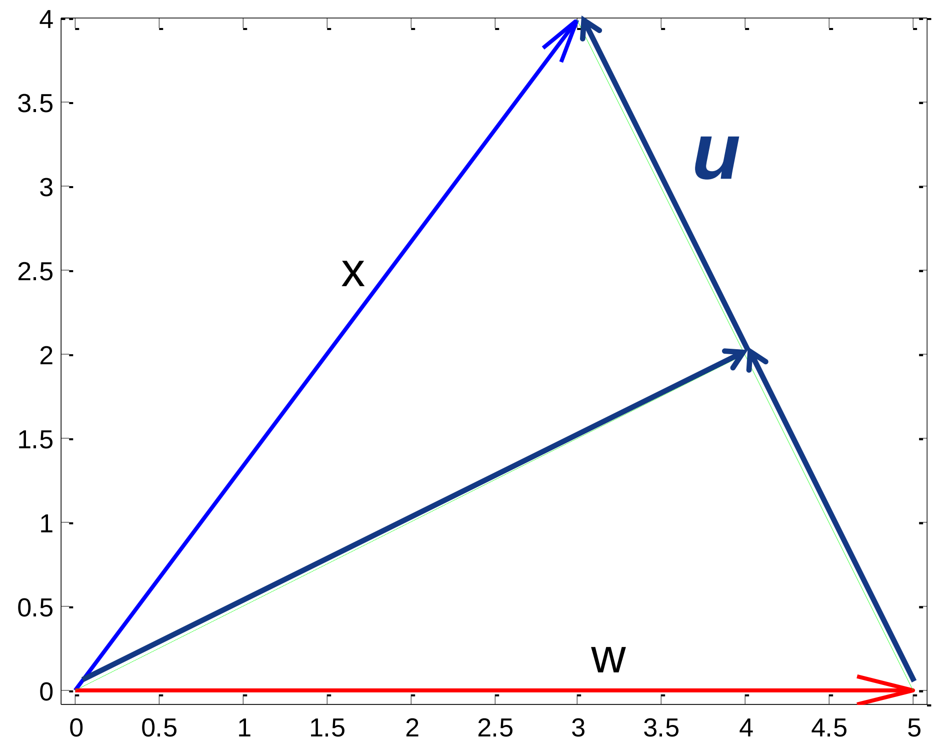
Given a vector x in the plane, to relocate it to a vector w of equal length, e.g.,

$$x = [3, 4] \text{ and } w = [5, 0]$$

$$\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{w}$$

$$\boldsymbol{e} = \frac{\boldsymbol{v}}{\sqrt{\boldsymbol{v}^T \boldsymbol{v}}}$$

$$\boldsymbol{u} = \boldsymbol{e}\left(\boldsymbol{e} \bullet \boldsymbol{x}\right) = \frac{\boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}}\left(\boldsymbol{v}^T \boldsymbol{x}\right)$$



QR Factorization

How QR factorization? A Simple Example

Given a vector x in the plane, to relocate it to a vector w of equal length, e.g.,

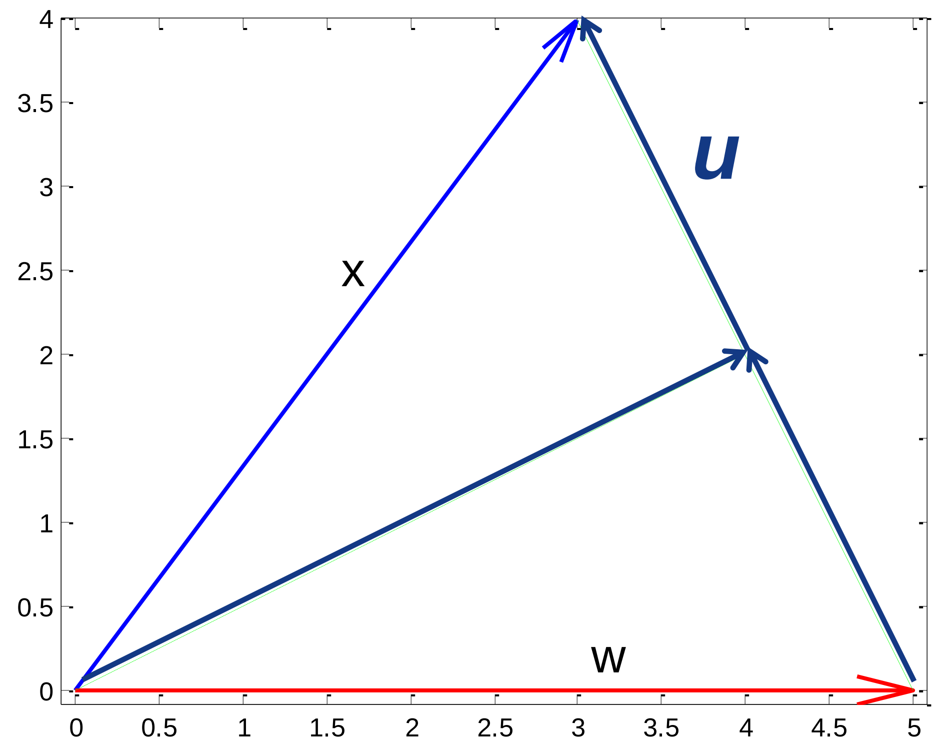
$x = [3, 4]$ and $w = [5, 0]$

$$\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{w}$$

$$\boldsymbol{u} = \frac{\boldsymbol{v}\boldsymbol{v}^T}{\boldsymbol{v}^T\boldsymbol{v}}\boldsymbol{x} \triangleq \boldsymbol{P}\boldsymbol{x}$$

$$\boldsymbol{x} - 2\boldsymbol{P}\boldsymbol{x} = \boldsymbol{w}$$

$$\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{P}$$



QR Factorization

- **How QR factorization? A Simple Example**

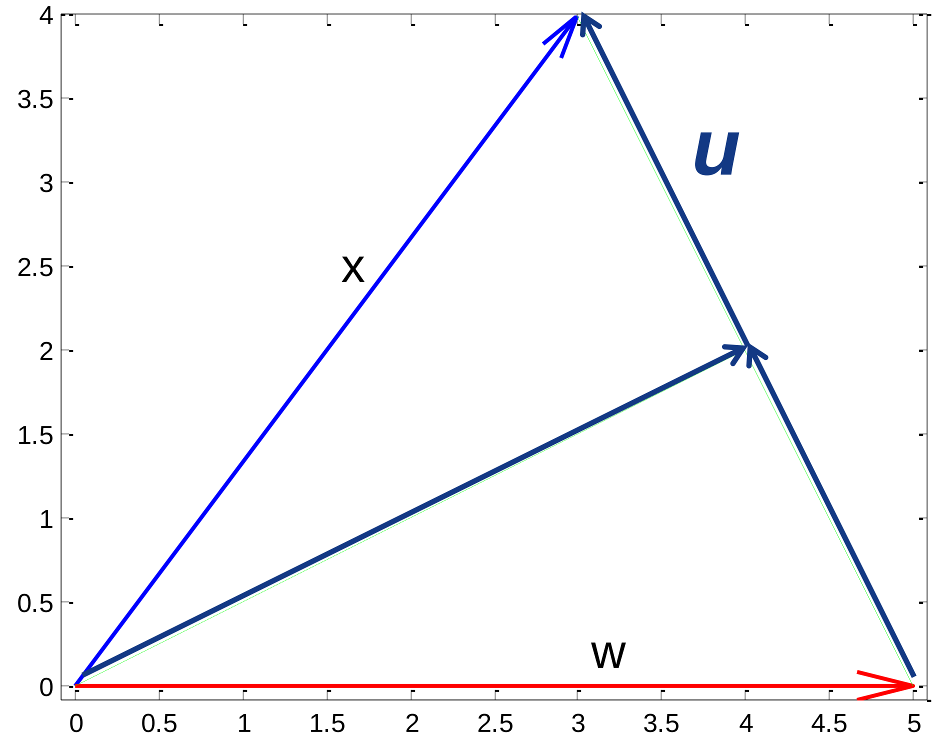
Given a vector *x* in the plane, to relocate it to a vector *w* of equal length, e.g.,

$$x = [3, 4] \text{ and } w = [5, 0]$$

$$P = \begin{bmatrix} 0.2 & -0.4 \\ -0.4 & 0.8 \end{bmatrix}$$

$$H = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix}$$

$$Hx = \begin{bmatrix} 5 \\ 0 \end{bmatrix} = w$$



QR Factorization

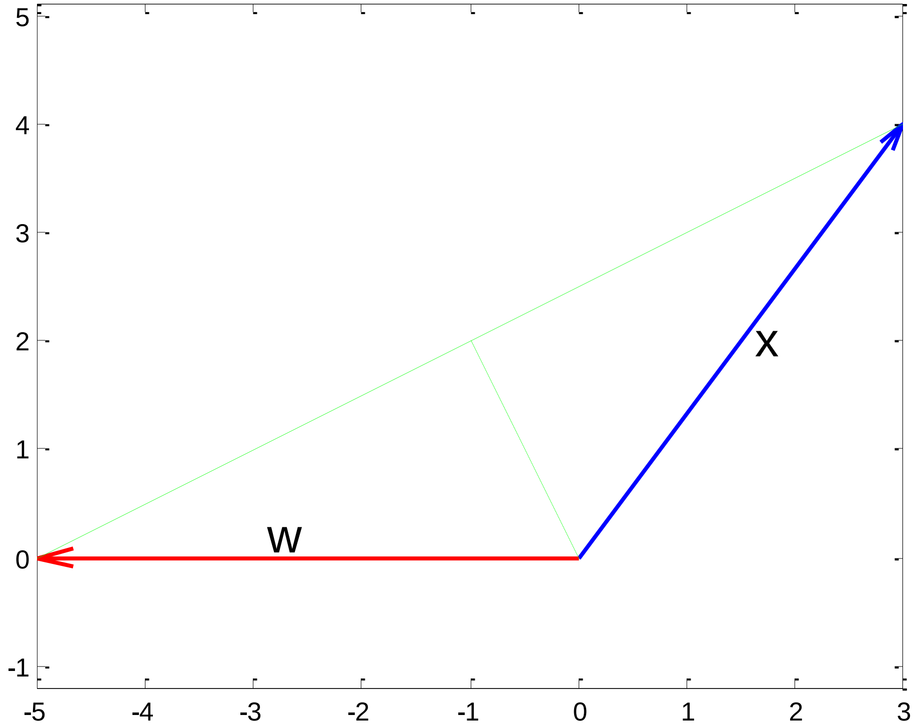
- **How QR factorization? A Simple Example**

Given a vector x in the plane, to relocate it to a vector w of equal length, e.g.,

$$x = [3, 4], \quad w = [-5, 0]$$

$$H = \begin{bmatrix} -0.6000 & -0.8000 \\ -0.8000 & 0.6000 \end{bmatrix}$$

$$Hx = \begin{bmatrix} -5 \\ 0 \end{bmatrix} = w$$



QR Factorization

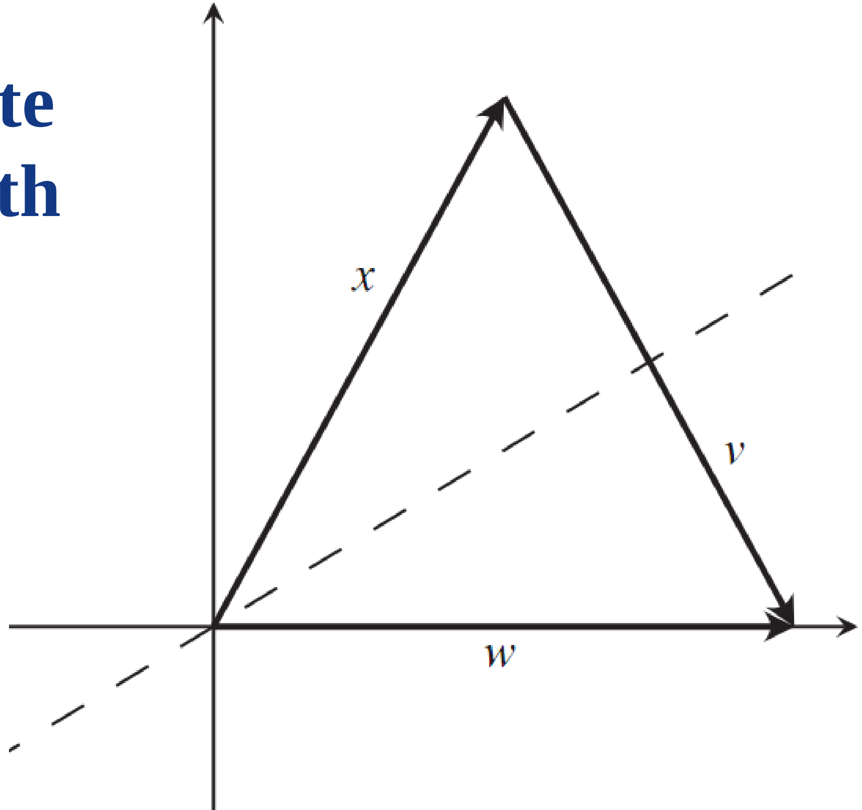
- **How QR factorization? Householder reflector**

A Householder reflector is an orthogonal matrix that reflects all m-vectors through an $m-1$ dimensional plane.

Given a vector x, to relocate to a vector w of equal length

Householder reflectors:

$$Hx = w$$



QR Factorization

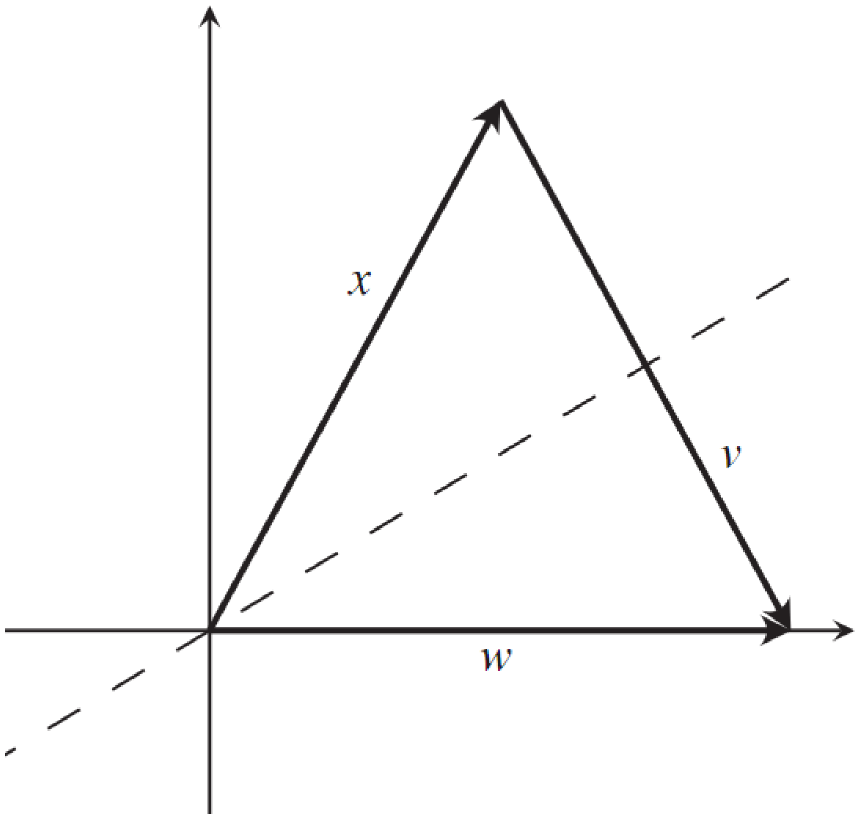
- **Householder Reflector: Basic Idea**

Assume that *x* and *w* are vectors of the same Euclidean length,

$$||x||_2 = ||w||_2$$

Then $w - x$ and $w + x$ are perpendicular.

$$(w - x)^T (w + x)$$
$$= ||w||^2 - ||x||^2$$
$$= 0$$



QR Factorization

Householder Reflector: Basic Idea

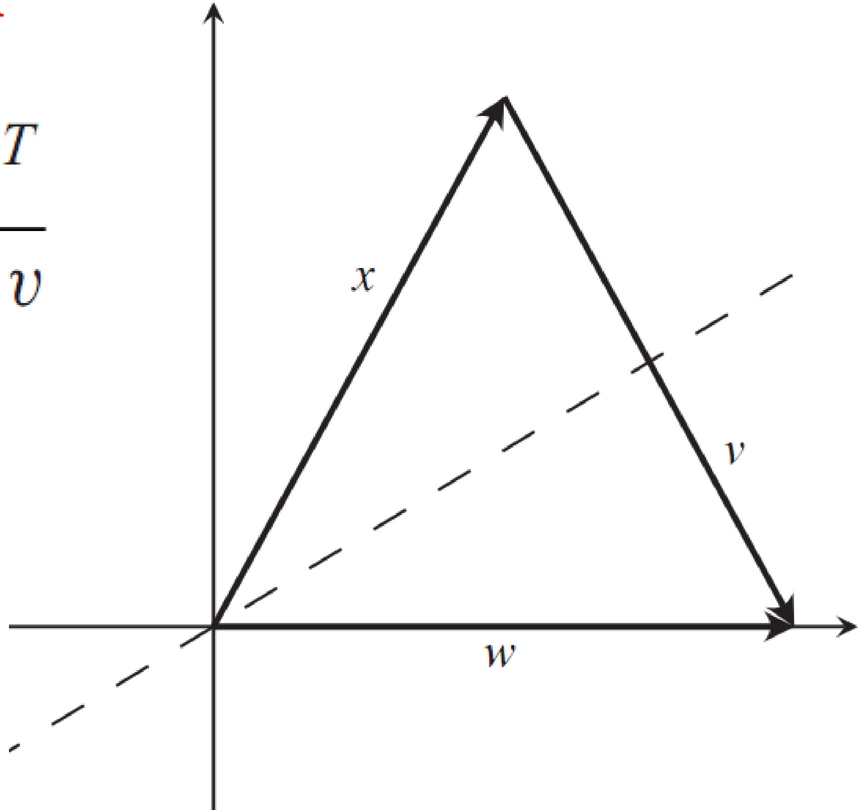
Assume that *x* and *w* are vectors of the same Euclidean length, $||x||_2 = ||w||_2$

Define the vector $v = w - x$

Projection matrix $P = \frac{vv^T}{v^T v}$

Householder reflector

$$H = I - 2P$$



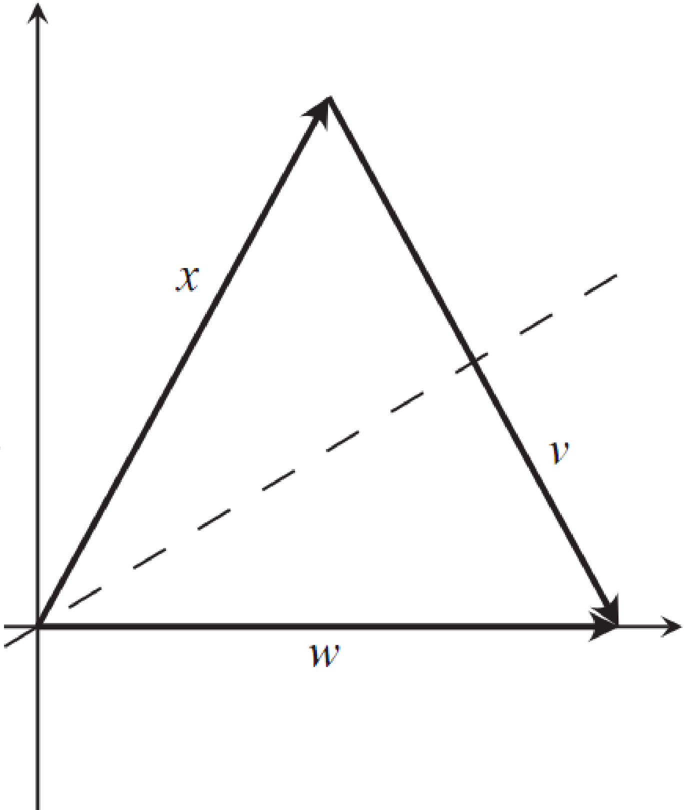
QR Factorization

- **Householder Reflector: Basic Idea**

Assume that *x* and *w* are vectors of the same Euclidean length, $||x||_2 = ||w||_2$

Projection matrix $P = \frac{vv^T}{v^Tv}$

$$\begin{aligned} Hx &= x - 2Px \\ &= w - v - \frac{2vv^Tx}{v^Tv} \\ &= w - v - \frac{vv^Tx}{v^Tv} - \frac{vv^T(w-v)}{v^Tv} \\ &= w - \frac{vv^T(w+x)}{v^Tv} \\ &= w \end{aligned}$$



QR Factorization

- **Householder Reflector: Theory**

Assume that *x* and *w* are vectors of the same Euclidean length,

$$||x||_2 = ||w||_2$$

and define the vector $v = w - x$.
Then

$$H = I - 2vv^T / v^T v$$

is a symmetric and orthogonal matrix and $Hx = w$.

QR Factorization

Householder Reflector: Theory

Projection matrix $P = \dfrac{vv^T}{v^T v}$, $P^2 = P$

$$H = I - 2vv^T / v^T v$$

is a symmetric and orthogonal matrix and *Hx* = *w*, i.e.,

$$\begin{aligned} H^T H = HH &= (I - 2P)(I - 2P) \\ &= I - 4P + 4P^2 \\ &= I. \end{aligned}$$

QR Factorization

QR Factorization Using Householder Reflector

Given a matrix *A*, to write it in the form *A* = *QR*

Let x_1 be the first column of *A*
Let $w = \pm(||x_1||_2, 0, \ldots, 0)$ be a vector along the first coordinate axis of identical Euclidean length

Householder reflector H_1

$$H_1 A = H_1 \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$$

QR Factorization

- **QR Factorization Using Householder Reflector**

Given a matrix A, to write it in the form $A = QR$

Let x_2 be the lower $m-1$ entries in column 2 of H_1A

Let w_2 be $\pm(||x_2||_2, 0, \ldots, 0)$

Householder reflector $\hat{H}_2$

$$\underbrace{\left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & \hat{H}_2 & \\ 0 & & & \end{array}\right)}_{H_2} \left(\begin{array}{c|cc} \times & \times & \times \\ \hline 0 & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{array}\right) = \left(\begin{array}{c|cc} \times & \times & \times \\ \hline 0 & \times & \times \\ 0 & 0 & \times \\ 0 & 0 & \times \end{array}\right)$$

$H_2 H_1 A$

QR Factorization

QR Factorization Using Householder Reflector

Given a matrix *A*, to write it in the form *A* = *QR*

One more step gives

Householder reflector $\hat{H}_3$

$$\underbrace{\left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & & \\ 0 & 0 & & \hat{H}_3 \end{array}\right)}_{H_3}\left(\begin{array}{cc|c} \times & \times & \times \\ 0 & \times & \times \\ \hline 0 & 0 & \times \\ 0 & 0 & \times \end{array}\right) = \left(\begin{array}{cc|c} \times & \times & \times \\ 0 & \times & \times \\ \hline 0 & 0 & \times \\ 0 & 0 & 0 \end{array}\right)$$

$$H_3 H_2 H_1 A = R$$

QR Factorization

- **QR Factorization Using Householder Reflector**

Given a matrix *A*, to write it in the form *A* = *QR*

One more step gives

Householder reflector $\hat{H}_3$

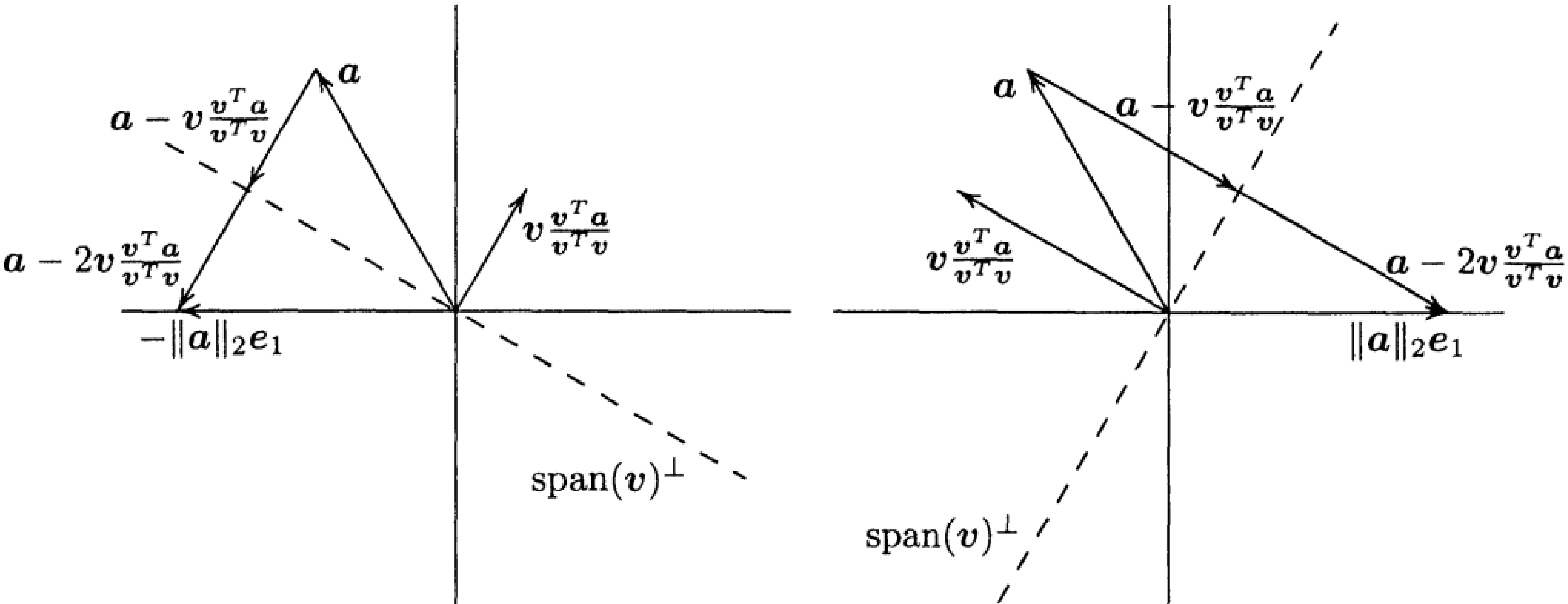
$$\left(\begin{array}{cc:cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hdashline 0 & 0 & & \\ 0 & 0 & & \hat{H}_3 \end{array}\right)\left(\begin{array}{cc:c} \times & \times & \times \\ 0 & \times & \times \\ \hdashline 0 & 0 & \times \\ 0 & 0 & \times \end{array}\right) = \left(\begin{array}{cc:c} \times & \times & \times \\ 0 & \times & \times \\ \hdashline 0 & 0 & \times \\ 0 & 0 & 0 \end{array}\right)$$

$$Q = H_1 H_2 H_3 \quad \Leftarrow \quad H_3 H_2 H_1 A = \boldsymbol{R}$$

QR Factorization

- **Householder Reflector: More Details**

Geometric Interpretation of Householder Reflector



For numerical stability, the sign is often chosen to be the opposite of the sign of the first component of *a* to avoid the possibility of subtracting nearly equal numbers when forming *v*.

QR Factorization

Householder QR Factorization: Procedure

Given a matrix *A*, to write it in the form *A* = *QR*

for $k = 1$ to n { loop over columns }
  $\alpha_k = -\text{sign}(a_{kk})\sqrt{a_{kk}^2 + \cdots + a_{mk}^2}$ { compute Householder vector for current col }
  $\boldsymbol{v}_k = [0 \ \cdots \ 0 \ a_{kk} \ \cdots \ a_{mk}]^T - \alpha_k \boldsymbol{e}_k$
  $\beta_k = \boldsymbol{v}_k^T \boldsymbol{v}_k$
  if $\beta_k = 0$ then { skip current column if it's already zero }
    continue with next k
  for $j = k$ to n { apply transformation to remaining submatrix }
    $\gamma_j = \boldsymbol{v}_k^T \boldsymbol{a}_j$
    $\boldsymbol{a}_j = \boldsymbol{a}_j - (2\gamma_j/\beta_k)\boldsymbol{v}_k$
  end
end

QR Factorization

QR Factorization: Example 4

Use Householder reflectors to find the QR factorization of

$$A = \begin{bmatrix} 3 & 1 \\ 4 & 3 \end{bmatrix} \Rightarrow A = \begin{bmatrix} 0.6 & 0.8 \\ 0.8 & -0.6 \end{bmatrix} \begin{bmatrix} 5 & 3 \\ 0 & -1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1/3 & -14/15 & -2/15 \\ 2/3 & 1/3 & -2/3 \\ 2/3 & 2/15 & 11/15 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix}$$

Nonlinear Least Squares

Motivation: Example 5

Use model linearization to find the best least squares exponential fit

$$y = c_1 e^{c_2 t}$$

to the following world automobile supply data:

year	cars ($\times 10^6$)
1950	53.05
1955	73.04
1960	98.31
1965	139.78
1970	193.48
1975	260.20
1980	320.39

Nonlinear Least Squares

- **Motivation: Example 5**

$$y = c_1 e^{c_2 t}$$

$\Downarrow$ **applying the natural logarithm**

$$\ln y = \ln(c_1 e^{c_2 t}) = \ln c_1 + c_2 t$$

$\Downarrow$ $c_1 = e^k$

$$\ln y = k + c_2 t$$

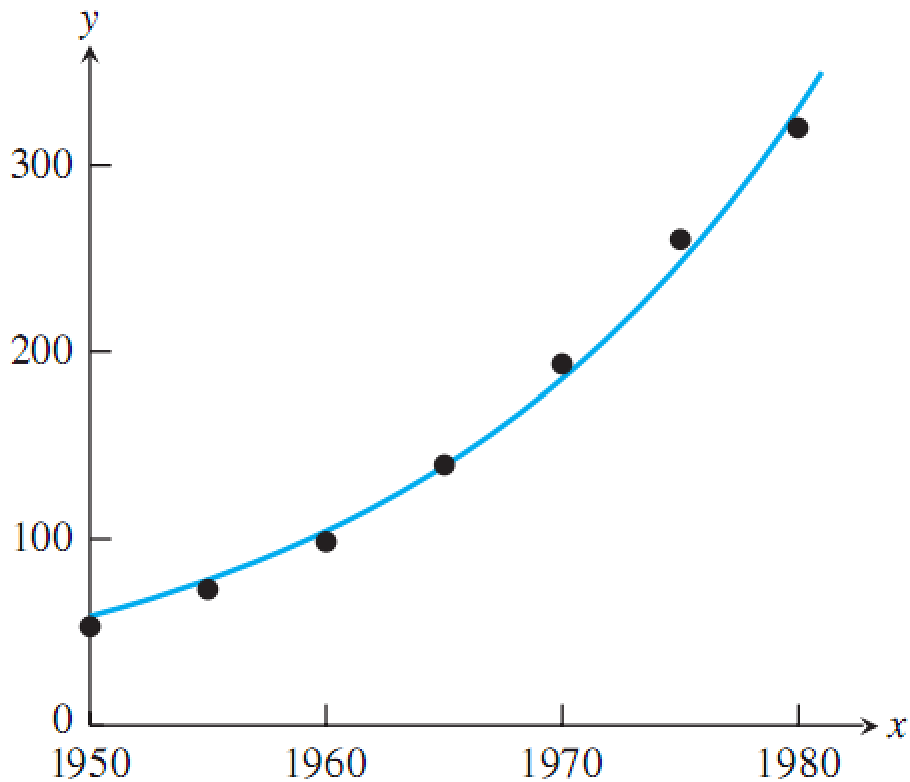
Now both coefficients k and c_2 are linear in the model.

Nonlinear Least Squares

- **Motivation: Example 5**

Use model linearization to find the best least squares exponential fit

$$y = c_1 e^{c_2 t}$$



$$y = 54.03e^{0.06152t}$$

The RMSE in log space:
RMSE = 0.0357.

Nonlinear Least Squares

Gauss–Newton Method

Consider the system of *m* equations in *n* unknowns

$$r_1(x_1, \ldots, x_n) = 0$$

$$\vdots$$

$$r_m(x_1, \ldots, x_n) = 0$$

The sum of the squares of the errors is represented by the function

$$E(x_1, \ldots, x_n) = \frac{1}{2}(r_1^2 + \cdots + r_m^2) = \frac{1}{2} r^T r$$

where $r = [r_1, \ldots, r_m]^T$.

Nonlinear Least Squares

- **Review of the derivative of a vector-valued function**

Let $f(x_1, ..., x_n)$ be a scalar-valued function of n variables. The derivative of f is the vector-valued function

$$Df(x_1, \ldots, x_n) = \left[f_{x_1}, \ldots, f_{x_n} \right] \qquad (1 \times n)$$

Let $F(x_1, \ldots, x_n) = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_n(x_1, \ldots, x_n) \end{bmatrix}$

be a vector-valued function of n variables.

The Jacobian of F: $DF(x_1, \ldots, x_n) = \begin{bmatrix} Df_1 \\ \vdots \\ Df_n \end{bmatrix}$

Nonlinear Least Squares

- **Review of the derivative of a vector-valued function**

Vector dot product rule

$$D(u^T v) = v^T Du + u^T Dv$$

Matrix/vector product rule

$$D(Av) = A \cdot Dv + \sum_{i=1}^{n} v_i Da_i,$$

where a_i denotes the ith column of A.

Nonlinear Least Squares

Gauss–Newton Method

The sum of the squares of the errors is represented by the function

$$E(x_1, \ldots, x_n) = \frac{1}{2}(r_1^2 + \cdots + r_m^2) = \frac{1}{2}r^T r$$

To minimize *E*, we set the derivative of *E* to zero

$$0 = F(x) = DE(x) = D\left(\frac{1}{2}r(x)^T r(x)\right) = r(x)^T Dr(x).$$

$1 \times m \quad m \times n$

where *Dr*(*x*) is the Jacobian matrix of *r*(*x*) .

Nonlinear Least Squares

- **Newton's Method for Nonlinear Equations: Review**

Systems of Equations:

$$\mathbf{f}(\mathbf{x}) = \mathbf{0}$$

Nonlinear Least Squares

- **Newton's Method for Nonlinear Equations: Review**

Newton's method for the $n \times n$ nonlinear system $\mathbf{f}(\mathbf{x}) = 0$

Step 1: Choose a starting point $\mathbf{x}_0$; ε_1, ε_2; let $k = 0$;

Step 2: Calculate $\mathbf{f}(\mathbf{x}_k)$ and $\mathbf{J}(\mathbf{x}_k)$;

Step 3: Solve the linear system $\mathbf{J}(\mathbf{x}_k)\,\delta = -\,\mathbf{f}(\mathbf{x}_k)$;

Step 4: Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \delta$;

Step 5: Check $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| < \varepsilon_1$ and $\|\mathbf{f}(\mathbf{x}_k)\| < \varepsilon_2$; if they are satisfied, stop and $\mathbf{r} = \mathbf{x}_{k+1}$, else, go to Step 6;

Step 6: Set $k = k + 1$, go to Step 2.

Nonlinear Least Squares

- **Gauss–Newton Method**

To use the Newton's Method

$$0 = F(x) = DE(x) = D\left(\frac{1}{2}r(x)^T r(x)\right) = r(x)^T Dr(x).$$

$$\Downarrow$$

$$F(x)^T = (r^T Dr)^T = (Dr)^T r$$

$$\Downarrow \quad \textbf{the Jacobian}$$

$$DF(x)^T = D((Dr)^T r) = (Dr)^T \cdot Dr + \sum_{i=1}^{m} r_i Dc_i$$

Nonlinear Least Squares

Gauss–Newton Method

To use the Newton's Method

$$DF(x)^T = D((Dr)^T r) = (Dr)^T \cdot Dr + \sum_{i=1}^{m} r_i Dc_i$$

where c_i is the ith column of Dr.

The Hessian matrix:

$$Dc_i = H_{r_i} = \begin{bmatrix} \frac{\partial^2 r_i}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 r_i}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 r_i}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 r_i}{\partial x_n \partial x_n} \end{bmatrix}$$

Nonlinear Least Squares

- **Gauss–Newton Method: General Procedure**

  To minimize

$$r_1(x)^2 + \cdots + r_m(x)^2.$$

Set $x^0 =$ initial vector,

for $k = 0, 1, 2, \ldots$

$$A = Dr(x^k)$$

$$A^T A v^k = -A^T r(x^k)$$ **%The normal equation**

$$x^{k+1} = x^k + v^k$$

end

Nonlinear Least Squares

- **Gauss–Newton Method: Example 5 revisited**

Use the Gauss–Newton Method to fit the world automobile supply data of Example 5 with a (nonlinearized) exponential model.

year	cars ($\times 10^6$)
1950	53.05
1955	73.04
1960	98.31
1965	139.78
1970	193.48
1975	260.20
1980	320.39

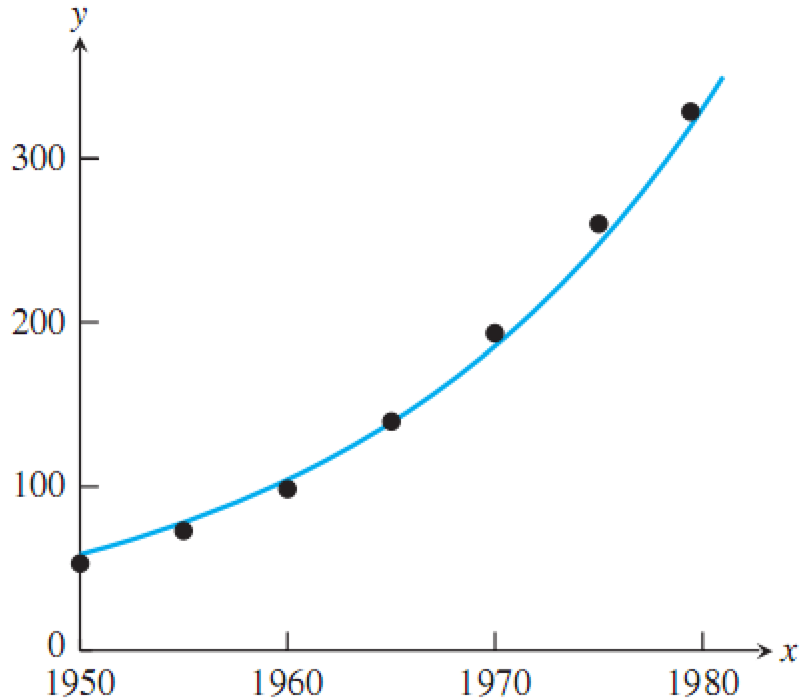
$$y = f_c(t) = c_1 e^{c_2 t}$$

$$r = \begin{bmatrix} c_1 e^{c_2 t_1} - y_1 \\ \vdots \\ c_1 e^{c_2 t_m} - y_m \end{bmatrix}$$

Nonlinear Least Squares

- **Gauss–Newton Method: Example 5 revisited**

Use the Gauss–Newton Method to fit the world automobile supply data of Example 5 with a (nonlinearized) exponential model.



$$Dr = \begin{bmatrix} e^{c_2 t_1} & c_1 t_1 e^{c_2 t_1} \\ \vdots & \vdots \\ e^{c_2 t_m} & c_1 t_m e^{c_2 t_m} \end{bmatrix}$$

$$y = 58.51 e^{0.05772 t}$$

Nonlinear Least Squares

- **Levenberg–Marquardt Method: Motivation**

Least squares minimization is especially challenging when the coefficient matrix turns out

Many plausible model definitions yield poorly conditioned *Dr* matrices.

$$A^T A v^k = -A^T r(x^k)$$

$$\Downarrow$$

$$(A^T A + \underline{\lambda \operatorname{diag}(A^T A)})v^k = -A^T r(x^k)$$

regularization term

Nonlinear Least Squares

- **Levenberg–Marquardt Method: Procedure**

  To minimize

$$r_1(x)^2 + \cdots + r_m(x)^2.$$

Set x^0 = initial vector, λ = constant
for $k = 0, 1, 2, \ldots$

$$A = Dr(x^k)$$
$$(A^T A + \lambda \operatorname{diag}(A^T A))v^k = -A^T r(x^k)$$
$$x^{k+1} = x^k + v^k$$

end

Nonlinear Least Squares

- **Levenberg–Marquardt Method: Example 6**

Use Levenberg–Marquardt method to fit the model

$$y = c_1 e^{-c_2(t-c_3)^2}$$

to the data points

$$(t_i, y_i) = \{(1,3), (2,5), (2,7), (3,5), (4,1)\}$$

$$r = \begin{bmatrix} c_1 e^{-c_2(t_1-c_3)^2} - y_1 \\ \vdots \\ c_1 e^{-c_2(t_5-c_3)^2} - y_5 \end{bmatrix}$$

Nonlinear Least Squares

- **Levenberg–Marquardt Method: Example 6**

Use Levenberg–Marquardt method to fit the model

$$y = c_1 e^{-c_2(t-c_3)^2}$$

to the data points

$$(t_i, y_i) = \{(1,3), (2,5), (2,7), (3,5), (4,1)\}$$

$$Dr = \begin{bmatrix} e^{-c_2(t_1-c_3)^2} & -c_1(t_1-c_3)^2 e^{-c_2(t_1-c_3)^2} & 2c_1c_2(t_1-c_3)e^{-c_2(t_1-c_3)^2} \\ \vdots & \vdots & \vdots \\ e^{-c_2(t_5-c_3)^2} & -c_1(t_5-c_3)^2 e^{-c_2(t_5-c_3)^2} & 2c_1c_2(t_5-c_3)e^{-c_2(t_5-c_3)^2} \end{bmatrix}$$

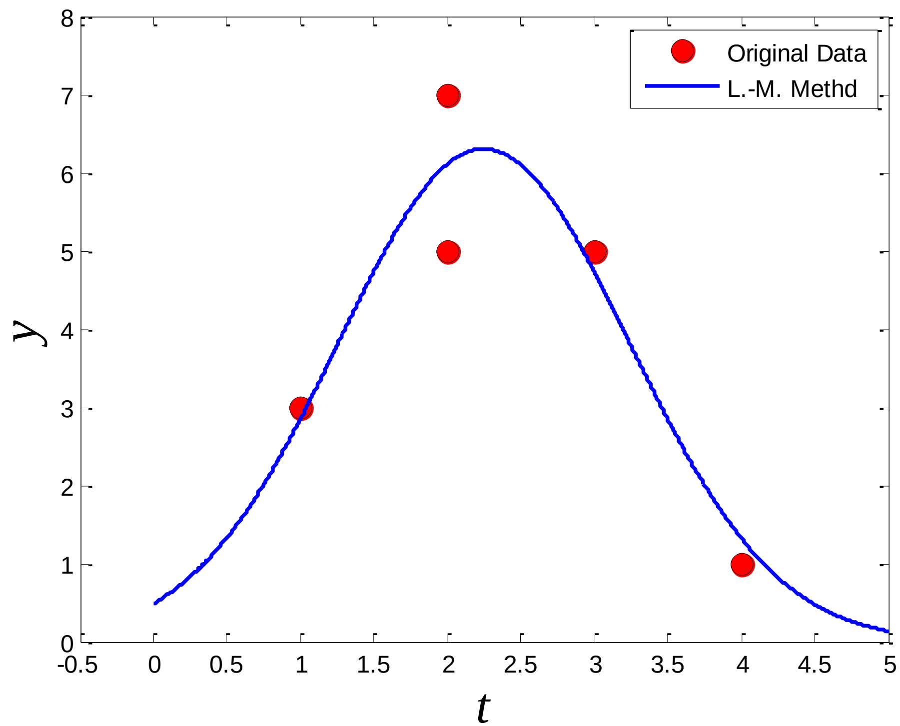
with initial guess $(c_1, c_2, c_3) = (1, 1, 1)$ and λ fixed at 50

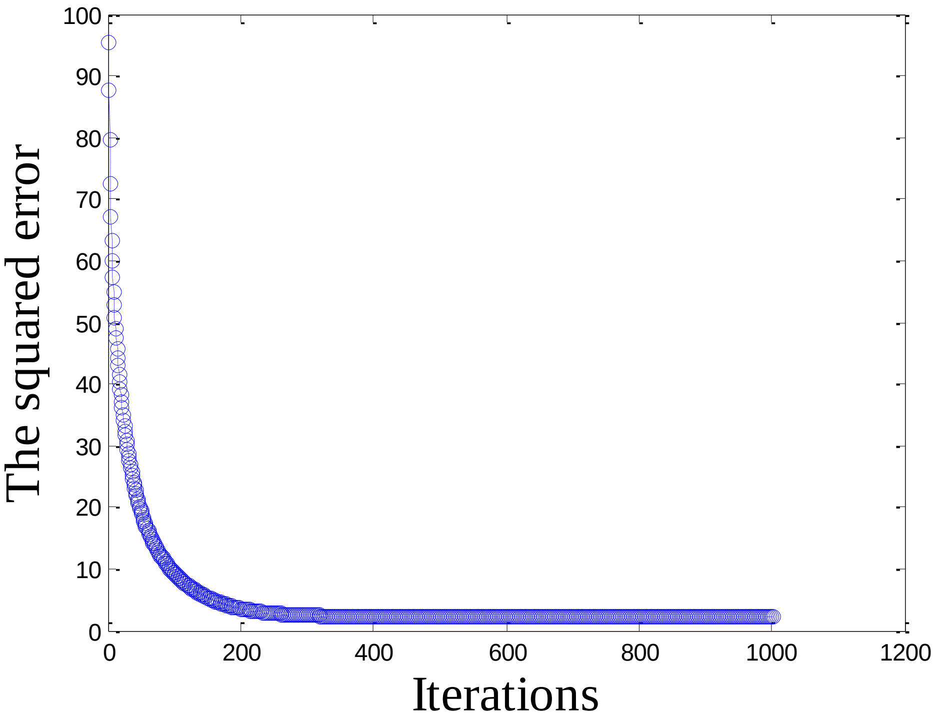
Nonlinear Least Squares

Levenberg–Marquardt Method: Example 6

Use Levenberg–Marquardt method to fit the model

$$y = c_1 e^{-c_2(t-c_3)^2}$$





$$y = 6.301e^{-0.5088(t-2.249)^2}$$

MATLAB Built-in Functions

- **MATLAB Built-in Functions for Least Squares**
  - ✓ **Polynomial curve fitting: *polyfit***
  - ✓ **Solve nonlinear least-squares (nonlinear data-fitting) problems: *lsqnonlin***
  - ✓ **Solve nonlinear curve-fitting (data-fitting) problems in least-squares sense: *lsqcurvefit***
  - ✓ **Find minimum of unconstrained multivariable function using <u>derivative-free</u> method: *fminsearch***

Summary

This lecture introduces a number of methods for least squares fitting problems.

- ❒ **Linear Least Squares**
  - ✓ **The Normal Equation**
  - ✓ **QR Factorization**
- ❒ **Nonlinear Least Squares**
  - ✓ **Gauss–Newton Method**
  - ✓ **Levenberg–Marquardt Method**

Thank You !