Chapter 4: Least Squares

Liangda Fang

Dept. of Computer Science
Jinan University

Motivation

- In Chapter 2, we study how to solve a linear equation.

Motivation

- In Chapter 2, we study how to solve a linear equation.
- What to do if the equation is inconsistent?
- For example,

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3. \end{aligned}$$

Outline

Outline

The least squares solution

$$\begin{aligned} x_1 + x_2 &= 2 \\ x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3. \end{aligned}$$

- The above linear equation is inconsistent.
- We can find a vector $\bar{x}$ that comes the closest to being a solution.

The least squares solution

$$
\begin{aligned}
x_1 + x_2 &= 2 \\
x_1 - x_2 &= 1 \\
x_1 + x_2 &= 3.
\end{aligned}
$$

- The above linear equation is inconsistent.
- We can find a vector $\bar{x}$ that comes the closest to being a solution.
- Question: How to define the vector $\bar{x}$?

The least squares solution

The matrix form of the inconsistent system is as follows:

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

The least squares solution

The matrix form of the inconsistent system is as follows:

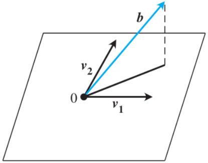
$$x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}.$$

Any $m \times n$ system $Ax = b$ can be viewed as a vector equation

$$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b.$$

b is a linear combination of the columns v_i of A, with coefficients $x_1, \ldots, x_n$.

The least squares solution



- $v_1 x_1 + v_2 x_2$ forms a plane inside R^3;
- b is out of the plane;
- There is no solution satisfying $v_1 x_1 + v_2 x_2 = b$;
- A special vector $\bar{x}$ in the plane $v_1 x_1 + v_2 x_2$ is closest to b;
- $b - v_1 \bar{x}_1 - v_2 \bar{x}_2$ is perpendicular to the plane $v_1 x_1 + v_2 x_2$.

The least squares solution

- A: a $m \times n$ matrix A;
- b: m-dimensional vector;
- The least squares solution $\bar{x}$ satisfies $(b - A\bar{x}) \perp \{Ax \mid x \in R^n\}$.

The least squares solution

1. $(b - A\bar{x}) \perp \{Ax \mid x \in R^n\}$;

The least squares solution

1. $(b - A\bar{x}) \perp \{Ax \mid x \in R^n\}$;
2. $(Ax)^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;

The least squares solution

1. $(b - A\bar{x}) \perp \{Ax \mid x \in R^n\}$;
2. $(Ax)^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;
3. $x^\top A^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;

The least squares solution

1. $(b - A\bar{x}) \perp \{Ax \mid x \in R^n\}$;
2. $(Ax)^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;
3. $x^\top A^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;
4. The vector $A^\top (b - A\bar{x})$ is perpendicular to every vector $x \in R^n$;

The least squares solution

1. $(b - A\bar{x}) \perp \{Ax \mid x \in R^n\}$;
2. $(Ax)^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;
3. $x^\top A^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;
4. The vector $A^\top (b - A\bar{x})$ is perpendicular to every vector $x \in R^n$;
5. $A^\top (b - A\bar{x}) = 0$;

The least squares solution

1. $(b - A\bar{x}) \perp \{Ax \mid x \in R^n\}$;
2. $(Ax)^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;
3. $x^\top A^\top (b - A\bar{x}) = 0$ for all $x \in R^n$;
4. The vector $A^\top (b - A\bar{x})$ is perpendicular to every vector $x \in R^n$;
5. $A^\top (b - A\bar{x}) = 0$;
6. Reduce obtaining the least squares solution to $Ax = b$ to solving the equation $A^\top A\bar{x} = A^\top b$;
7. $A^\top A\bar{x} = A^\top b$ is called the **normal equations** of $Ax = b$.

The least squares solution

Theorem

Let A be a $m \times n$ matrix and b m-dimensional vector. Let $\bar{x}$ be the solution of $A^\top Ax = A^\top b$. Then, the minimum value of $\|Ax - b\|_2$ is achieved when $x = \bar{x}$.

The least squares solution

Example

- $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$;
- $A^\top A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$;
- $A^\top b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$;
- The normal equation is $\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$;
- $\bar{x} = \begin{bmatrix} \frac{7}{4} & \frac{3}{4} \end{bmatrix}^\top$.

The least squares solution

Example

- $A\bar{x} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{7}{4} \\ \frac{3}{4} \end{bmatrix} = \begin{bmatrix} 2.5 \\ 1 \\ 2.5 \end{bmatrix} \neq \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$
- The residual: $r = d - A\bar{x} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 1 \\ 2.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0 \\ 0.5 \end{bmatrix}$

Three ways to express the size of residual

1. 2-norm: $\|r\|_2 = \sqrt{r_1^2 + \cdots + r_m^2}$;
2. The squared error (SE): $r_1^2 + \cdots + r_m^2$;
3. The root mean squared error (RMSE): $\sqrt{\frac{r_1^2+\cdots+r_m^2}{m}} = \frac{\|r\|_2}{\sqrt{m}}$.

Three ways to express the size of residual

1. 2-norm: $\|r\|_2 = \sqrt{r_1^2 + \cdots + r_m^2}$;
2. The squared error (SE): $r_1^2 + \cdots + r_m^2$;
3. The root mean squared error (RMSE): $\sqrt{\frac{r_1^2+\cdots+r_m^2}{m}} = \frac{\|r\|_2}{\sqrt{m}}$.

Example

1. $\|r\|_2 = \sqrt{0.5^2 + 0^2 + (-0.5)^2} = \sqrt{0.5} \approx 0.707$;
2. SE: $0.5^2 + 0^2 + (-0.5)^2 = 0.5$;
3. RMSE: $\sqrt{\frac{0.5^2+0^2+(-0.5)^2}{3}} = \frac{1}{\sqrt{6}} \approx 0.408$.

Fitting models to data

- Remind that the problem which Chapter 3 focuses on.

Fitting models to data

- Remind that the problem which Chapter 3 focuses on.
- How to generate a polynomial given data points?

Fitting models to data

- Remind that the problem which Chapter 3 focuses on.
- How to generate a polynomial given data points?
- Generally, given n points, a degree $n-1$ polynomial is generated.

Fitting models to data

- Remind that the problem which Chapter 3 focuses on.
- How to generate a polynomial given data points?
- Generally, given n points, a degree $n-1$ polynomial is generated.
- Actually, we hope that the function is simpler, *e.g.*, $c_0 + c_1 x$, $c_0 + c_1 x + c_2 x^2$, $c_0 + c_1 \sin(x) + c_2 \cos(x)$.

Fitting models to data

Example

Find the best line for the four data points $(-1,1),(0,0),(1,0),(2,-2)$.

1. Choose the function $c_0 + c_1 x$;
2. Force the model to fit the data yields $\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -2 \end{bmatrix}$;
3. The normal equations are $\begin{bmatrix} 4 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} -1 \\ -5 \end{bmatrix}$;
4. $y = c_0 + c_1 x = 0.2 - 0.9x$ is the best line.

Fitting models to data

Example

1. The residuals are

x	y	line	error
-1	1	1.1	-0.1
0	0	0.2	-0.2
1	0	-0.7	0.7
2	-2	-1.6	-0.4

2. SE: $(-0.1)^2 + (-0.2)^2 + 0.7^2 + (-0.4)^2 = 0.7$;
3. RMSE: $\sqrt{\frac{0.7}{4}} \approx 0.418$.

Fitting models to data

Example

Find the best parabola for the four data points $(-1,1),(0,0),(1,0),(2,-2)$.

1. Choose the function $c_0 + c_1 x + c_2 x^2$;
2. Force the model to fit the data yields $\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -2 \end{bmatrix}$;
3. The normal equations are $\begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 8 \\ 6 & 8 & 18 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} -1 \\ -5 \\ -7 \end{bmatrix}$;
4. $y = c_0 + c_1 x + c_2 x^2 = 0.45 - 0.65x - 0.25x^2$ is the best parabola.

Fitting models to data

Example

1. The residuals are

x	y	line	error
-1	1	0.85	0.15
0	0	0.45	-0.45
1	0	-0.45	0.45
2	-2	-1.85	-0.15

2. SE: $0.15^2 + (-0.45)^2 + 0.45^2 + (-0.15)^2 = 0.45 < 0.7$ (SE of $0.2 - 0.9x$);

3. RMSE: $\sqrt{\frac{0.45}{4}} \approx 0.335 < 0.418$ (RMSE of $0.2 - 0.9x$).

Outline

Periodic data

- The sampling data may be sometimes periodic, *e.g.*, the daily or yearly temperatures in a city.
- But the polynomial is not a periodic function.
- We need some periodic functions, *e.g.*, sine and cosine functions.

Periodic data

Example

Fit the recorded temperatures in Washing, D.C., on Jan. 1, 2001 as listed in the following table, to a periodic model:

time of day	y	temp(C)
0:00	0	-2.2
3:00	$\frac{1}{8}$	-2.8
6:00	$\frac{1}{4}$	-6.1
9:00	$\frac{3}{8}$	-3.9
12:00	$\frac{1}{2}$	0.0
15:00	$\frac{5}{8}$	1.1
18:00	$\frac{3}{4}$	-0.6
21:00	$\frac{7}{8}$	-1.1

Periodic data

Example

Firstly, we choose the model $y = c_1 + c_2 \cos 2\pi t + c_3 \sin 2\pi t$, and substitute the date into the model.

$$\begin{aligned}
c_1 + c_2 \cos 2\pi(0) + c_3 \sin 2\pi(0) &= -2.2 \\
c_1 + c_2 \cos 2\pi(\frac{1}{8}) + c_3 \sin 2\pi(\frac{1}{8}) &= -2.8 \\
c_1 + c_2 \cos 2\pi(\frac{1}{4}) + c_3 \sin 2\pi(\frac{1}{4}) &= -6.1 \\
c_1 + c_2 \cos 2\pi(\frac{3}{8}) + c_3 \sin 2\pi(\frac{3}{8}) &= -3.9 \\
c_1 + c_2 \cos 2\pi(\frac{1}{2}) + c_3 \sin 2\pi(\frac{1}{2}) &= 0.0 \\
c_1 + c_2 \cos 2\pi(\frac{5}{8}) + c_3 \sin 2\pi(\frac{5}{8}) &= 1.1 \\
c_1 + c_2 \cos 2\pi(\frac{3}{4}) + c_3 \sin 2\pi(\frac{3}{4}) &= -0.6 \\
c_1 + c_2 \cos 2\pi(\frac{7}{8}) + c_3 \sin 2\pi(\frac{7}{8}) &= -1.1
\end{aligned}$$

Periodic data

Example

We get the inconsistent equation $Ax = b$ where

$$A = \begin{bmatrix} 1 & \cos 0 & \sin 0 \\ 1 & \cos\frac{\pi}{4} & \sin\frac{\pi}{4} \\ 1 & \cos\frac{\pi}{2} & \sin\frac{\pi}{2} \\ 1 & \cos\frac{3}{4}\pi & \sin\frac{3}{4}\pi \\ 1 & \cos\pi & \sin\pi \\ 1 & \cos\frac{5}{4}\pi & \sin\frac{5}{4}\pi \\ 1 & \cos\frac{3}{2}\pi & \sin\frac{3}{2}\pi \\ 1 & \cos\frac{7}{4}\pi & \sin\frac{7}{4}\pi \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & 0 & 1 \\ 1 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ 1 & 0 & -1 \\ 1 & \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \text{ and } b = \begin{bmatrix} -2.2 \\ -2.8 \\ -6.1 \\ -3.9 \\ 0.0 \\ 1.1 \\ -0.6 \\ -1.1 \end{bmatrix}.$$

Periodic data

Example

- The normal equations $A^\top A c = A^\top b$ are
$$\begin{bmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \end{bmatrix};$$

Periodic data

Example

- The normal equations $A^\top Ac = A^\top b$ are
$$\begin{bmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \end{bmatrix};$$
- $c = \begin{bmatrix} -1.95 & -0.7445 & -2.5594 \end{bmatrix}^\top;$

Periodic data

Example

- The normal equations $A^\top Ac = A^\top b$ are
$$\begin{bmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \end{bmatrix};$$
- $c = \begin{bmatrix} -1.95 & -0.7445 & -2.5594 \end{bmatrix}^\top;$
- The model: $y = -1.95 - 0.7445 \cos 2\pi t - 2.5594 \sin 2\pi t;$

Periodic data

Example

- The normal equations $A^\top Ac = A^\top b$ are
$$\begin{bmatrix} 8 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \end{bmatrix};$$
- $c = \begin{bmatrix} -1.95 & -0.7445 & -2.5594 \end{bmatrix}^\top$;
- The model: $y = -1.95 - 0.7445 \cos 2\pi t - 2.5594 \sin 2\pi t$;
- The RMSE: 1.063.

Periodic data

Example

Now, consider the improved model

$y = c_1 + c_2 \cos 2\pi t + c_3 \sin 2\pi t + c_4 \cos 4\pi t.$

$$
\begin{aligned}
c_1 + c_2 \cos 2\pi(0) + c_3 \sin 2\pi(0) + c_4 \cos 4\pi(0) &= -2.2 \\
c_1 + c_2 \cos 2\pi(\frac{1}{8}) + c_3 \sin 2\pi(\frac{1}{8}) + c_4 \cos 4\pi(\frac{1}{8}) &= -2.8 \\
c_1 + c_2 \cos 2\pi(\frac{1}{4}) + c_3 \sin 2\pi(\frac{1}{4}) + c_4 \cos 4\pi(\frac{1}{4}) &= -6.1 \\
c_1 + c_2 \cos 2\pi(\frac{3}{8}) + c_3 \sin 2\pi(\frac{3}{8}) + c_4 \cos 4\pi(\frac{3}{8}) &= -3.9 \\
c_1 + c_2 \cos 2\pi(\frac{1}{2}) + c_3 \sin 2\pi(\frac{1}{2}) + c_4 \cos 4\pi(\frac{1}{2}) &= 0.0 \\
c_1 + c_2 \cos 2\pi(\frac{5}{8}) + c_3 \sin 2\pi(\frac{5}{8}) + c_4 \cos 4\pi(\frac{5}{8}) &= 1.1 \\
c_1 + c_2 \cos 2\pi(\frac{3}{4}) + c_3 \sin 2\pi(\frac{3}{4}) + c_4 \cos 4\pi(\frac{3}{4}) &= -0.6 \\
c_1 + c_2 \cos 2\pi(\frac{7}{8}) + c_3 \sin 2\pi(\frac{7}{8}) + c_4 \cos 4\pi(\frac{7}{8}) &= -1.1
\end{aligned}
$$

Periodic data

Example

- The normal equations $A^\top A c = A^\top b$ are

$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \\ 4.5 \end{bmatrix};$$

Periodic data

Example

- The normal equations $A^\top A c = A^\top b$ are
$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \\ 4.5 \end{bmatrix};$$
- $c = \begin{bmatrix} -1.95 & -0.7445 & -2.5594 & 1.125 \end{bmatrix}^\top;$

Periodic data

Example

- The normal equations $A^\top A c = A^\top b$ are
$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \\ 4.5 \end{bmatrix};$$
- $c = \begin{bmatrix} -1.95 & -0.7445 & -2.5594 & 1.125 \end{bmatrix}^\top$;
- The model:
$y = -1.95 - 0.7445 \cos 2\pi t - 2.5594 \sin 2\pi t + 1.125 \cos 4\pi t$;

Periodic data

Example

- The normal equations $A^\top Ac = A^\top b$ are
$$\begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} -15.6 \\ -2.9778 \\ -10.2376 \\ 4.5 \end{bmatrix};$$

- $c = \begin{bmatrix} -1.95 & -0.7445 & -2.5594 & 1.125 \end{bmatrix}^\top;$

- The model:
$y = -1.95 - 0.7445 \cos 2\pi t - 2.5594 \sin 2\pi t + 1.125 \cos 4\pi t;$

- The RMSE: 0.705.

- This model is better than the last one since its RMSE is less than that of the latter.

Outline

Data linearization

- Exponential growth is very popular.
- To describe the growth, we need the exponential model $y = c_1 e^{c_2(t-t_1)}$.
- The model cannot be directly fit by least squares because c_2 does not appear linearly in the model equation.
- We can linear this model as $\ln y = \ln(c_1 e^{c_2(t-t_1)}) = \ln c_1 + c_2(t - t_1)$.

Data linearization

Example

The number of transistors on Intel CPU since the early 1970s is given in the following table. Fit the model $y = c_1 e^{c_2(t-1970)}$ to the data.

CPU	year	transistors
4004	1971	2,250
8008	1972	2,500
8080	1974	5,000
8086	1978	29,000
286	1982	120,000
386	1985	275,000
486	1989	1,180,000
Pentium	1993	3,100,000
Pentium II	1997	7,500,000
Pentium III	1999	24,000,000
Pentium 4	2000	42,000,000
Itanium	2002	220,000,000
Itanium 2	2003	410,000,000

Data linearization

Example

Substitute the date into the linearized model $c_1' + c_2(t - 1970) = \ln y$ where $c_1' = \ln c_1$:

$$
\begin{aligned}
c_1' + c_2(1) &= \ln 2250 \\
c_1' + c_2(2) &= \ln 2500 \\
c_1' + c_2(4) &= \ln 5000 \\
c_1' + c_2(8) &= \ln 29000 \\
&\vdots \\
c_1' + c_2(33) &= \ln 410000000
\end{aligned}
$$

Data linearization

Example

We get the inconsistent equation $Ax = b$ where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \\ 1 & 8 \\ \vdots & \vdots \\ 1 & 33 \end{bmatrix} \text{ and } b = \begin{bmatrix} \ln 2250 \\ \ln 2500 \\ \ln 5000 \\ \ln 29000 \\ \vdots \\ \ln 410000000 \end{bmatrix}.$$

Data linearization

Example

- The normal equations $A^\top Ac = A^\top b$ are
$$\begin{bmatrix} 13 & 235 \\ 235 & 5927 \end{bmatrix} = \begin{bmatrix} c_1' \\ c_2 \end{bmatrix} = \begin{bmatrix} 176.90 \\ 3793.23 \end{bmatrix};$$

Data linearization

Example

- The normal equations $A^\top Ac = A^\top b$ are
$\begin{bmatrix} 13 & 235 \\ 235 & 5927 \end{bmatrix} = \begin{bmatrix} c_1' \\ c_2 \end{bmatrix} = \begin{bmatrix} 176.90 \\ 3793.23 \end{bmatrix}$;
- $c_1' \approx 7.197$ and $c_2 \approx 0.3546$;
- $c_1 = e^{c_1'} \approx 1335.3$;

Data linearization

Example

- The normal equations $A^\top Ac = A^\top b$ are
$$\begin{bmatrix} 13 & 235 \\ 235 & 5927 \end{bmatrix} = \begin{bmatrix} c_1' \\ c_2 \end{bmatrix} = \begin{bmatrix} 176.90 \\ 3793.23 \end{bmatrix};$$
- $c_1' \approx 7.197$ and $c_2 \approx 0.3546$;
- $c_1 = e^{c_1'} \approx 1335.3$;
- The model: $y = 1335.3e^{0.3546t}$;

Data linearization

Example

- The normal equations $A^\top Ac = A^\top b$ are
$\begin{bmatrix} 13 & 235 \\ 235 & 5927 \end{bmatrix} = \begin{bmatrix} c_1' \\ c_2 \end{bmatrix} = \begin{bmatrix} 176.90 \\ 3793.23 \end{bmatrix}$;
- $c_1' \approx 7.197$ and $c_2 \approx 0.3546$;
- $c_1 = e^{c_1'} \approx 1335.3$;
- The model: $y = 1335.3e^{0.3546t}$;
- The doubling time for the law is $\ln 2/c_2 \approx 1.95$ years;
- Gordon C. Moore, cofounder of Intel, predicted in 1965 that over the ensuing decade, computing power would double every 2 years.

Outline

Conditioning of least squares

- The least squares problem of $Ax = b$ reduces to solving the normal equation $A^\top Ax = A^\top b$.
- The condition number of $A^\top A$ may be too large.

Conditioning of least squares

- The least squares problem of $Ax = b$ reduces to solving the normal equation $A^\top Ax = A^\top b$.
- The condition number of $A^\top A$ may be too large.

Example

- $x_1, \ldots, x_{11}$: equally spaced points in $[2, 4]$;
- $y_i = 1 + x_i + x_i^2 + x_i^3 + x_i^4 + x_i^5 + x_i^6 + x_i^7$ for $1 \leq i \leq 11$;
- Find the least squares polynomial $P(x) = c_0 + c_1 x + \cdots + c_7 x^7$ fitting the (x_i, y_i);

Conditioning of least squares

Example

- Substituting the data points into the model $P(x)$ yields $Ac = y$ as follows:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^7 \\ 1 & x_2 & x_2^2 & \cdots & x_2^7 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11} & x_{11}^2 & \cdots & x_{11}^7 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_7 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_1 1 \end{bmatrix}$$

- The condition number of $A^\top A$ is 1.4359×10^{19};

Conditioning of least squares

Example

- Substituting the data points into the model $P(x)$ yields $Ac = y$ as follows:

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^7 \\ 1 & x_2 & x_2^2 & \cdots & x_2^7 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{11} & x_{11}^2 & \cdots & x_{11}^7 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_7 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_1 1 \end{bmatrix}$$

- The condition number of $A^\top A$ is 1.4359×10^{19};
- The correct least squares solution is $c_0 = \cdots = c_7 = 1$;
- The solution computed by Matlab is $c_0 = 1.5134$, $c_1 = -0.2644$, $c_2 = 2.3211$, $c_3 = 0.2408$, $c_4 = 1.2592$, $c_5 = 0.9474$, $c_6 = 1.0059$, $c_7 = 0.9997$.

Outline

Gram-Schmidt orthogonalization

- Given n linear independent vectors $\{A_1, \cdots, A_n\}$;
- Generate a set $\{q_1, \cdots, q_n\}$ of vectors satisfying:
  1. Each q_i is a unit vector: $\|q_i\|_2 = 1$ for $1 \leq i \leq n$;
  2. The set are pairwise perpendicular: $q_i^\top q_j = 0$ for $i \neq j$;
  3. Two subspaces are equal: $(A_1, \cdots, A_n) = (q_1, \cdots, q_n)$.

Gram-Schmidt orthogonalization

y_i: an auxiliary vector which is perpendicular to $q_1, \cdots q_{i-1}$.

Gram-Schmidt orthogonalization

y_i: an auxiliary vector which is perpendicular to $q_1, \cdots q_{i-1}$.

1. Let $y_1 = A_1$;

Gram-Schmidt orthogonalization

y_i: an auxiliary vector which is perpendicular to $q_1, \cdots q_{i-1}$.

1. Let $y_1 = A_1$;
2. Normalize y_1: $q_1 = \frac{y_1}{\|y_1\|_2}$, *i.e.*, $q_1^\top q_1 = 1$;

Gram-Schmidt orthogonalization

y_i: an auxiliary vector which is perpendicular to $q_1, \cdots q_{i-1}$.

1. Let $y_1 = A_1$;
2. Normalize y_1: $q_1 = \frac{y_1}{\|y_1\|_2}$, *i.e.*, $q_1^\top q_1 = 1$;
3. Acquire y_2 via orthogonalization: $y_2 = A_2 - q_1 q_1^\top A_2$;
    - $q_1^\top y_2 = q_1^\top (A_2 - q_1 q_1^\top A_2) = q_1^\top A_2 - q_1^\top A_2 = 0.$

Gram-Schmidt orthogonalization

y_i: an auxiliary vector which is perpendicular to $q_1, \cdots q_{i-1}$.

1. Let $y_1 = A_1$;
2. Normalize y_1: $q_1 = \frac{y_1}{\|y_1\|_2}$, *i.e.*, $q_1^\top q_1 = 1$;
3. Acquire y_2 via orthogonalization: $y_2 = A_2 - q_1 q_1^\top A_2$;
    - $q_1^\top y_2 = q_1^\top (A_2 - q_1 q_1^\top A_2) = q_1^\top A_2 - q_1^\top A_2 = 0$.
4. Normalize y_2: $q_2 = \frac{y_2}{\|y_2\|_2}$;

Gram-Schmidt orthogonalization

y_i: an auxiliary vector which is perpendicular to $q_1, \cdots q_{i-1}$.

1. Let $y_1 = A_1$;
2. Normalize y_1: $q_1 = \frac{y_1}{\|y_1\|_2}$, *i.e.*, $q_1^\top q_1 = 1$;
3. Acquire y_2 via orthogonalization: $y_2 = A_2 - q_1 q_1^\top A_2$;
   - $q_1^\top y_2 = q_1^\top (A_2 - q_1 q_1^\top A_2) = q_1^\top A_2 - q_1^\top A_2 = 0.$
4. Normalize y_2: $q_2 = \frac{y_2}{\|y_2\|_2}$;
5. $\vdots$
6. Acquire y_j: $y_j = A_j - q_1(q_1^\top A_j) - \cdots - q_{j-1}(q_{j-1}^\top A_j)$;

Gram-Schmidt orthogonalization

y_i: an auxiliary vector which is perpendicular to $q_1, \cdots q_{i-1}$.

1. Let $y_1 = A_1$;
2. Normalize y_1: $q_1 = \frac{y_1}{\|y_1\|_2}$, *i.e.*, $q_1^\top q_1 = 1$;
3. Acquire y_2 via orthogonalization: $y_2 = A_2 - q_1 q_1^\top A_2$;
   - $q_1^\top y_2 = q_1^\top (A_2 - q_1 q_1^\top A_2) = q_1^\top A_2 - q_1^\top A_2 = 0.$
4. Normalize y_2: $q_2 = \frac{y_2}{\|y_2\|_2}$;
5. $\vdots$
6. Acquire y_j: $y_j = A_j - q_1(q_1^\top A_j) - \cdots - q_{j-1}(q_{j-1}^\top A_j)$;
7. Normalize y_j: $q_j = \frac{y_j}{\|y_j\|_2}$;

Gram-Schmidt orthogonalization

y_i: an auxiliary vector which is perpendicular to $q_1, \cdots q_{i-1}$.

1. Let $y_1 = A_1$;
2. Normalize y_1: $q_1 = \frac{y_1}{\|y_1\|_2}$, *i.e.*, $q_1^\top q_1 = 1$;
3. Acquire y_2 via orthogonalization: $y_2 = A_2 - q_1 q_1^\top A_2$;
   - $q_1^\top y_2 = q_1^\top (A_2 - q_1 q_1^\top A_2) = q_1^\top A_2 - q_1^\top A_2 = 0.$
4. Normalize y_2: $q_2 = \frac{y_2}{\|y_2\|_2}$;
5. $\vdots$
6. Acquire y_j: $y_j = A_j - q_1(q_1^\top A_j) - \cdots - q_{j-1}(q_{j-1}^\top A_j)$;
7. Normalize y_j: $q_j = \frac{y_j}{\|y_j\|_2}$;
8. Until q_n is computed.

Gram-Schmidt orthogonalization

Theorem

q_i and q_j are perpendicular for $i \neq j$.

Gram-Schmidt orthogonalization

Theorem

q_i and q_j are perpendicular for $i \neq j$.

Proof.

- Assume that $i < j \leq n$.
- It is sufficient to prove that $q_i^\top y_j = 0$.

Gram-Schmidt orthogonalization

Theorem

q_i and q_j are perpendicular for $i \neq j$.

Proof.

- Assume that $i < j \leq n$.
- It is sufficient to prove that $q_i^\top y_j = 0$.
- Base case ($j = 2$): $q_1^\top y_2 = 0$.

Gram-Schmidt orthogonalization

Theorem

q_i and q_j are perpendicular for $i \neq j$.

Proof.

- Assume that $i < j \leq n$.
- It is sufficient to prove that $q_i^\top y_j = 0$.
- Base case ($j = 2$): $q_1^\top y_2 = 0$.
- Inductive step: assume that $q_i^\top y_l = 0$ for $i \neq l$ and $i, l < k$.
- We now prove that $q_i^\top y_l = 0$ for $i \neq l$ and $i, l < k + 1$.
- It is sufficient to prove that $q_i^\top y_k = 0$.

Gram-Schmidt orthogonalization

Proof.

$$
\begin{aligned}
q_i^\top y_k &= q_i^\top (A_k - q_1 q_1^\top A_k - \cdots q_i q_i^\top A_k - \cdots - q_{k-1} q_{k-1}^\top A_k) \\
&= q_i^\top A_k - q_i^\top q_1 q_1^\top A_k - \cdots q_i^\top q_i q_i^\top A_k - \cdots - q_i^\top q_{k-1} q_{k-1}^\top A_k \\
&= q_i^\top A_k - q_i^\top q_i (q_i^\top A_k) \\
&= q_i^\top A_k - q_i^\top A_k \\
&= 0
\end{aligned}
$$

Reduced QR factorization

- What is the relation between $A_1, \ldots, A_n$ and $q_1, \ldots, q_n$?

Reduced QR factorization

- What is the relation between $A_1, \ldots, A_n$ and $q_1, \ldots, q_n$?
- $(A_1 | \cdots | A_n) = (q_1 | \cdots | q_n) \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}.$
- $r_{jj} = \|y_j\|_2$ and $r_{ij} = q_i^\top A_j$;
- $A_j = r_{1j} q_1 + \cdots + r_{j-1,j} q_{j-1} + r_{jj} q_j$;

Reduced QR factorization

- What is the relation between $A_1, \dots, A_n$ and $q_1, \dots, q_n$?
- $$(A_1|\cdots|A_n) = (q_1|\cdots|q_n)\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}.$$
- $r_{jj} = \|y_j\|_2$ and $r_{ij} = q_i^\top A_j$;
- $A_j = r_{1j}q_1 + \cdots + r_{j-1,j}q_{j-1} + r_{jj}q_j$;
- QR is the **reduced QR factorization** of A.

Reduced QR factorization

Example

Find the reduced QR factorization of $A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$.

- $y_1 = A_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$;
- $r_{11} = \|y_1\|_2 = \sqrt{1^2 + 2^2 + 2^2} = 3$;
- $q_1 = \frac{y_1}{\|y_1\|_2} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$;

Reduced QR factorization

Example

- $y_2 = A_2 - q_1 q_1^\top A_2 = \begin{bmatrix} -4 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} -4 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{14}{3} \\ \frac{5}{3} \\ \frac{2}{3} \end{bmatrix}$;
- $r_{12} = q_1^\top A_2 = 2$ and $r_{22} = \|y_2\|_2 = 5$;
- $q_2 = \frac{y_2}{\|y_2\|_2} = \frac{1}{5} \begin{bmatrix} -\frac{14}{3} \\ \frac{5}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix}$;
- $A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ & r_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{14}{15} \\ \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{2}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} = QR.$

Full QR factorization

- In reduced QR factorization:
  1. Q is $m \times n$;
  2. R is $n \times n$.
- In full QR factorization:
  1. Q is $m \times m$;
  2. R is $m \times n$.
- $$(A_1|\cdots|A_n) = (q_1|\cdots|q_n|\cdots|q_m)\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}.$$
- QR is the **full QR factorization** of A.

Full QR factorization

Example

Find the full QR factorization of $A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$.

- Construct a vector $A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ linearly independent of A_1 and A_2;
- $y_3 = A_3 - q_1 q_1^\top A_3 - q_2 q_2^\top A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} - (-\frac{14}{15}) \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix} = \frac{2}{225} \begin{bmatrix} 2 \\ 10 \\ -11 \end{bmatrix}$;

Full QR factorization

Example

- $q_3 = \frac{y_3}{\|y_3\|_2} = \begin{bmatrix} \frac{2}{15} \\ \frac{10}{15} \\ -\frac{11}{15} \end{bmatrix}$;
-

$$\begin{aligned} A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} &= \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ & r_{22} \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{3} & -\frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{15} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} \\ &= QR \end{aligned}$$

Classical Gram-Schmidt orthogonalization

Algorithm 1: Classical Gram-Schmidt orthogonalization

Input: A: an $m \times n$ matrix
Output: Q: an orthogonal matrix
R: an upper triangular matrix s.t. $A = QR$
for $j = 1, 2, \ldots, n$ **do**
  $y = A_j$
  for $i = 1, 2, \ldots, j-1$ **do**
    $r_{ij} = q_i^\top A_j$
    $y = y - r_{ij} q_i$
  $r_{jj} = \|y\|_2$
  $q_j = \frac{y}{r_{jj}}$

Its application to least squares

Definition

A square matrix Q is **orthogonal** if $Q^\top = Q^{-1}$.

Its application to least squares

Definition

A square matrix Q is **orthogonal** if $Q^\top = Q^{-1}$.

Proposition

A square matrix is orthogonal iff its columns are pairwise orthogonal unit vectors.

Its application to least squares

Proof.

Let $Q = (q_1 | \cdots | q_n)$.

Its application to least squares

Proof.

Let $Q = (q_1 | \cdots | q_n)$.

- Q is orthogonal $\iff$

Its application to least squares

Proof.

Let $Q = (q_1 | \cdots | q_n)$.

- Q is orthogonal $\iff$
- $(q_1 | \cdots | q_n)^\top (q_1 | \cdots | q_n) = I \iff$

Its application to least squares

Proof.

Let $Q = (q_1|\cdots|q_n)$.

- Q is orthogonal $\Longleftrightarrow$
- $(q_1|\cdots|q_n)^\top(q_1|\cdots|q_n) = I \Longleftrightarrow$
- $\begin{bmatrix} q_1^\top q_1 & q_1^\top q_2 & \cdots & q_1^\top q_n \\ q_2^\top q_1 & q_2^\top q_2 & \cdots & q_2^\top q_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^\top q_1 & q_n^\top q_2 & \cdots & q_n^\top q_n \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \Longleftrightarrow$

Its application to least squares

Proof.

Let $Q = (q_1|\cdots|q_n)$.

- Q is orthogonal $\iff$
- $(q_1|\cdots|q_n)^\top (q_1|\cdots|q_n) = I \iff$
- $$\begin{bmatrix} q_1^\top q_1 & q_1^\top q_2 & \cdots & q_1^\top q_n \\ q_2^\top q_1 & q_2^\top q_2 & \cdots & q_2^\top q_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^\top q_1 & q_n^\top q_2 & \cdots & q_n^\top q_n \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \iff$$
- $$q_i^\top q_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Application to least squares

Lemma

If Q is an orthogonal $m \times m$ matrix and x is an m-dimensional vector, then $\|Qx\|_2 = \|x\|_2$.

An orthogonal matrix Q does not change the length of vector x.

Application to least squares

Lemma

If Q is an orthogonal $m \times m$ matrix and x is an m-dimensional vector, then $\|Qx\|_2 = \|x\|_2$.

An orthogonal matrix Q does not change the length of vector x.

Proof.

$\|Qx\|_2^2 = (Qx)^\top (Qx) = x^\top Q^\top Qx = x^\top x = \|x\|_2^2$

Application to least squares

- Least squares: minimize $\|Ax - b\|_2$;
- Minimize $\|QRx - b\|_2$ $(A = QR)$;
- Minimize $\|Rx - Q^\top b\|_2$ $(\|Rx - Q^\top b\|_2 = \|QRx - b\|_2)$.

Application to least squares

$Rx - Q^\top b = e$ where $\|e\|_2$ is the error between Ax and b.

$$
\begin{bmatrix}
r_{11} & r_{12} & \cdots & r_{1n} \\
 & r_{22} & \cdots & r_{2n} \\
 & & \ddots & \vdots \\
 & & & r_{nn} \\
0 & 0 & \cdots & 0 \\
\vdots & & & \vdots \\
0 & 0 & \cdots & 0
\end{bmatrix}
\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}
-
\begin{bmatrix} d_1 \\ \vdots \\ d_n \\ d_{n+1} \\ \vdots \\ d_m \end{bmatrix}
=
\begin{bmatrix} e_1 \\ \vdots \\ e_n \\ e_{n+1} \\ \vdots \\ e_m \end{bmatrix}
$$

- $d = Q^\top b$;
- $\bar{x}$: the solution of the upper part;
- $\|e\|_2^2 = d_{n+1}^2 + \cdots + d_m^2$: the least squares error.

Application to least squares

Example

- Solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$;

Application to least squares

Example

- Solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$;
- $A = QR = \begin{bmatrix} \frac{1}{3} & -\frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{15} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix}$;

Application to least squares

Example

- Solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$;
- $A = QR = \begin{bmatrix} \frac{1}{3} & -\frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{15} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix}$;
- $Rx = Q^\top b$

Application to least squares

Example

- Solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$;
- $A = QR = \begin{bmatrix} \frac{1}{3} & -\frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{15} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix}$;
- $Rx = Q^\top b$
- $\begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ -\frac{14}{15} & \frac{1}{3} & \frac{2}{15} \\ \frac{2}{15} & \frac{2}{3} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \\ 3 \end{bmatrix}$;

Application to least squares

Example

- Solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$;
- $A = QR = \begin{bmatrix} \frac{1}{3} & -\frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{15} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix}$;
- $Rx = Q^\top b$
- $\begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ -\frac{14}{15} & \frac{1}{3} & \frac{2}{15} \\ \frac{2}{15} & \frac{2}{3} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \\ 3 \end{bmatrix}$;
- Equating the upper parts yields $\begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \end{bmatrix}$;

Application to least squares

Example

- Solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$;
- $A = QR = \begin{bmatrix} \frac{1}{3} & -\frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{15} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix}$;
- $Rx = Q^\top b$
- $\begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ -\frac{14}{15} & \frac{1}{3} & \frac{2}{15} \\ \frac{2}{15} & \frac{2}{3} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \\ 3 \end{bmatrix}$;
- Equating the upper parts yields $\begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \end{bmatrix}$;
- The least squares solution: $\bar{x}_1 = 3.8$ and $\bar{x}_2 = 1.8$;

Application to least squares

Example

- Solve the least squares problem $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix}$;
- $A = QR = \begin{bmatrix} \frac{1}{3} & -\frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{15} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix}$;
- $Rx = Q^\top b$
- $\begin{bmatrix} 3 & 2 \\ 0 & 5 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ -\frac{14}{15} & \frac{1}{3} & \frac{2}{15} \\ \frac{2}{15} & \frac{2}{3} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} -3 \\ 15 \\ 9 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \\ 3 \end{bmatrix}$;
- Equating the upper parts yields $\begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 9 \end{bmatrix}$;
- The least squares solution: $\bar{x}_1 = 3.8$ and $\bar{x}_2 = 1.8$;
- The least squares error: $\|e\|_2 = \|(0, 0, -3)\|_2 = 3$.

A problem of Gram-Schmidt orthogonalization

Gram-Schmidt algorithm sometimes has the accuracy problem.

A problem of Gram-Schmidt orthogonalization

Gram-Schmidt algorithm sometimes has the accuracy problem.

Example

Let $A = \begin{bmatrix} 1 & 1 & 1 \\ \delta & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & \delta \end{bmatrix}$ where $\delta = 10^{-10}$.

- $y_1 = A_1 = \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix}$;
- $r_{11} = \|y_1\|_2 = \sqrt{1+\delta^2} = \sqrt{1+10^{-20}} = 1$ after rounding;
- $q_1 = \frac{y_1}{r_{11}} = \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix}$.

A problem of Gram-Schmidt orthogonalization

Example

- $y_2 = A_2 - q_1 q_1^\top A_2 = \begin{bmatrix} 1 \\ 0 \\ \delta \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\delta \\ \delta \\ 0 \end{bmatrix}$;
- $r_{22} = \|y_2\|_2 = \sqrt{(-\delta)^2 + \delta^2} = \sqrt{2}\delta$;
- $q_2 = \frac{y_2}{r_{22}} = \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \\ 0 \end{bmatrix}$.

A problem of Gram-Schmidt orthogonalization

Example

- $y_3 = A_3 - q_1 q_1^\top A_3 - q_2 q_2^\top A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \delta \end{bmatrix} - \begin{bmatrix} 1 \\ \delta \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\delta \\ 0 \\ \delta \end{bmatrix};$
- $r_{33} = \|y_3\|_2 = \sqrt{(-\delta)^2 + \delta^2} = \sqrt{2}\delta;$
- $q_3 = \frac{y_3}{r_{33}} = \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{bmatrix}.$
- $q_2^\top q_3 = \begin{bmatrix} 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ 0 \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \frac{1}{2} \neq 0.$

Outline

Householder reflectors

Another alternative way to QR factorization based on **Householder reflectors**.

1. Deliver better orthogonality;
2. Has lower memory requirements.

Householder reflectors

Definition (Householder reflectors)

Let v be an n-dimensional **unit** vector.
Then the matrix $H = I - 2vv^\top$ is called a **Householder reflector**.

Householder reflectors

Definition (Householder reflectors)

Let v be an n-dimensional **unit** vector.
Then the matrix $H = I - 2vv^\top$ is called a **Householder reflector**.

Proposition

A Householder reflector, $H = I - 2vv^\top$, is symmetric and orthogonal.

Householder reflectors

Definition (Householder reflectors)

Let v be an n-dimensional **unit** vector.
Then the matrix $H = I - 2vv^\top$ is called a **Householder reflector**.

Proposition

A Householder reflector, $H = I - 2vv^\top$, is symmetric and orthogonal.

Proof.

- Symmetric: $$\begin{aligned} H^\top &= (I - 2vv^\top)^\top = I^\top - 2(vv^\top)^\top \\ &= I - 2[(v^\top)^\top v^\top] = I - 2vv^\top = H \end{aligned}$$

Householder reflectors

Definition (Householder reflectors)

Let v be an n-dimensional **unit** vector.
Then the matrix $H = I - 2vv^\top$ is called a **Householder reflector**.

Proposition

A Householder reflector, $H = I - 2vv^\top$, is symmetric and orthogonal.

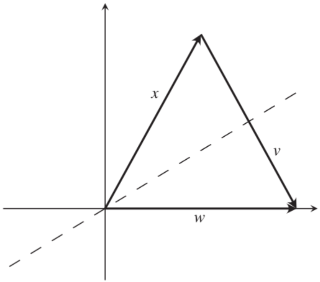
Proof.

- Symmetric: $$\begin{aligned} H^\top &= (I - 2vv^\top)^\top = I^\top - 2(vv^\top)^\top \\ &= I - 2[(v^\top)^\top v^\top] = I - 2vv^\top = H \end{aligned}$$
- Orthogonal: $$\begin{aligned} HH^\top &= (I - 2vv^\top)(I - 2vv^\top) \\ &= I - 2vv^\top - 2vv^\top + 4vv^\top vv^\top \\ &= I - 4vv^\top + 4vv^\top = I \end{aligned}$$

Householder reflectors

A Householder reflector can be used to

- project an n dimensional vector on an $n-1$ dimensional plane;
- but not change the length of the vector.



Householder reflectors

Lemma

Let x and w be two vectors with $\|x\|_2 = \|w\|_2$.
Then $w - x$ and $w + x$ are perpendicular.

Householder reflectors

Lemma

Let x and w be two vectors with $\|x\|_2 = \|w\|_2$.
Then $w - x$ and $w + x$ are perpendicular.

Proof.

$$(w-x)^\top (w+x) = w^\top w - x^\top w + w^\top x - x^\top x = \|w\|_2 - \|x\|_2 = 0$$

Householder reflectors

Theorem

Given the following:

- x *and* w*: two vectors with* $\|x\|_2 = \|w\|_2$*;*
- $u = w - x$ *and* $v = \frac{u}{\|u\|_2}$*;*
- $H = I - 2vv^\top$.

Then $Hx = w$ *and* $Hw = x$*.*

Householder reflectors

Proof.

Firstly, we prove that $Hx = w$.

$$\begin{aligned} Hx &= x - 2vv^\top x \\ &= w - u - 2\frac{uu^\top x}{\|u\|_2^2} \\ &= w - \frac{uu^\top u}{\|u\|_2^2} - \frac{uu^\top x}{\|u\|_2^2} - \frac{uu^\top (w-u)}{\|u\|_2^2} \\ &= w - \frac{uu^\top (w+x)}{\|u\|_2^2} \\ &= w - \frac{u(w-x)^\top (w+x)}{\|u\|_2^2} = w. \end{aligned}$$

Householder reflectors

Proof.

Firstly, we prove that $Hx = w$.

$$
\begin{aligned}
Hx &= x - 2vv^\top x \\
&= w - u - 2\frac{uu^\top x}{\|u\|_2^2} \\
&= w - \frac{uu^\top u}{\|u\|_2^2} - \frac{uu^\top x}{\|u\|_2^2} - \frac{uu^\top (w-u)}{\|u\|_2^2} \\
&= w - \frac{uu^\top (w+x)}{\|u\|_2^2} \\
&= w - \frac{u(w-x)^\top (w+x)}{\|u\|_2^2} = w.
\end{aligned}
$$

Secondly, we prove that $Hw = x$.

$$
\begin{aligned}
Hx &= w \\
H^{-1}Hx &= H^{-1}w \\
x &= H^\top w
\end{aligned}
$$

Householder reflectors

Example

- Let $x = \begin{bmatrix} 3 & 4 \end{bmatrix}^\top$ and $w = \begin{bmatrix} 5 & 0 \end{bmatrix}^\top$.
- Find a Householder reflector H s.t. $Hx = w$.

Householder reflectors

Example

- Let $x = \begin{bmatrix} 3 & 4 \end{bmatrix}^\top$ and $w = \begin{bmatrix} 5 & 0 \end{bmatrix}^\top$.
- Find a Householder reflector H s.t. $Hx = w$.
- Set $u = w - x = \begin{bmatrix} 5 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$ and $v = \frac{u}{\|u\|_2} = \begin{bmatrix} \frac{\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} \end{bmatrix}$.

Householder reflectors

Example

- Let $x = \begin{bmatrix} 3 & 4 \end{bmatrix}^\top$ and $w = \begin{bmatrix} 5 & 0 \end{bmatrix}^\top$.
- Find a Householder reflector H s.t. $Hx = w$.
- Set $u = w - x = \begin{bmatrix} 5 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$ and $v = \frac{u}{\|u\|_2} = \begin{bmatrix} \frac{\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} \end{bmatrix}$.
- $H = I - 2vv^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} \frac{\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}}{5} & -\frac{2\sqrt{5}}{5} \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix}$.

Householder reflectors

Example

- Let $x = \begin{bmatrix} 3 & 4 \end{bmatrix}^\top$ and $w = \begin{bmatrix} 5 & 0 \end{bmatrix}^\top$.
- Find a Householder reflector H s.t. $Hx = w$.
- Set $u = w - x = \begin{bmatrix} 5 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$ and $v = \frac{u}{\|u\|_2} = \begin{bmatrix} \frac{\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} \end{bmatrix}$.
- $H = I - 2vv^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} \frac{\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}}{5} & -\frac{2\sqrt{5}}{5} \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix}$.
- $Hx = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix} = w$.

Householder reflectors

Example

- Let $x = \begin{bmatrix} 3 & 4 \end{bmatrix}^\top$ and $w = \begin{bmatrix} 5 & 0 \end{bmatrix}^\top$.
- Find a Householder reflector H s.t. $Hx = w$.
- Set $u = w - x = \begin{bmatrix} 5 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$ and $v = \frac{u}{\|u\|_2} = \begin{bmatrix} \frac{\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} \end{bmatrix}$.
- $H = I - 2vv^\top = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2 \begin{bmatrix} \frac{\sqrt{5}}{5} \\ -\frac{2\sqrt{5}}{5} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}}{5} & -\frac{2\sqrt{5}}{5} \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix}$.
- $Hx = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix} = w$.
- $Hw = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} = x$.

Full QR Factorization based on Householder reflectors

Given an $m \times n$ matrix

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

1. Let $x_1 = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \end{bmatrix}^\top$;

Full QR Factorization based on Householder reflectors

Given an $m \times n$ matrix

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

1. Let $x_1 = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \end{bmatrix}^\top$;
2. Let $w_1 = \begin{bmatrix} \mathrm{sgn}(x_{11})\|x_1\|_2 & 0 & \cdots & 0 \end{bmatrix}^\top$;

Full QR Factorization based on Householder reflectors

Given an $m \times n$ matrix

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

1. Let $x_1 = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \end{bmatrix}^\top$;
2. Let $w_1 = \begin{bmatrix} \text{sgn}(x_{11})\|x_1\|_2 & 0 & \cdots & 0 \end{bmatrix}^\top$;
3. So $u_1 = w_1 - x_1$, $v_1 = \frac{u_1}{\|u_1\|_2}$ and $H_1 = I - 2v_1v_1^\top$;

Full QR Factorization based on Householder reflectors

Given an $m \times n$ matrix

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

1. Let $x_1 = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \end{bmatrix}^\top$;
2. Let $w_1 = \begin{bmatrix} \mathrm{sgn}(x_{11})\|x_1\|_2 & 0 & \cdots & 0 \end{bmatrix}^\top$;
3. So $u_1 = w_1 - x_1$, $v_1 = \frac{u_1}{\|u_1\|_2}$ and $H_1 = I - 2v_1v_1^\top$;
4. $$B = H_1 A = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$
   where $b_{11} = -\mathrm{sgn}(x_{11})\|x_1\|_2$.

Full QR Factorization based on Householder reflectors

1. Let $x_2 = \begin{bmatrix} b_{22} & b_{23} & \cdots & b_{2m} \end{bmatrix}^\top$;
2. Let $w_2 = \begin{bmatrix} \text{sgn}(x_{21})\|x_2\|_2 & 0 & \cdots & 0 \end{bmatrix}^\top$;
3. So $u_2 = w_2 - x_2$, $v_2 = \frac{u_2}{\|u_2\|_2}$ and $\hat{H}_2 = I - 2v_2v_2^\top$;
4. $$C = H_2 B = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \hat{H}_2 & \\ 0 & & & \end{bmatrix} B = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & c_{22} & c_{23} & \cdots & c_{2n} \\ 0 & 0 & c_{33} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & c_{m3} & \cdots & c_{mn} \end{bmatrix}$$
   where $c_{22} = \text{sgn}(x_{21})\|x_2\|_2$ and $c_{11} = b_{11}$.
5. $c_{1i} = b_{1i}$ for $1 \le i \le n$.

Full QR Factorization based on Householder reflectors

1. Let $x_n = \begin{bmatrix} b_{nn} & \cdots & b_{nm} \end{bmatrix}^\top$;
2. Let $w_n = \begin{bmatrix} \text{sgn}(x_{n1})\|x_n\|_2 & 0 & \cdots & 0 \end{bmatrix}^\top$;
3. So $u_n = w_n - x_n$, $v_n = \frac{u_n}{\|u_n\|_2}$ and $\hat{H}_n = I - 2v_nv_n^\top$;
4. $$R = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & 0 & \cdots & 0 \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \hat{H}_n & \\ 0 & \cdots & 0 & & \end{bmatrix} H_{n-1} \cdots H_1 A = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & r_{nn} \\ 0 & \cdots & \cdots & 0 \\ \vdots & 0 & 0 & \vdots \end{bmatrix}.$$

Full QR Factorization based on Householder reflectors

Finally, $Q = H_1 \cdots H_n$ and $R = H_n \cdots H_1 A$.

Full QR factorization based on Householder reflectors

Example

Use Householder reflectors to compute the full QR factorization of

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

- $x_1 = \begin{bmatrix} 1 & 2 & 2 \end{bmatrix}^\top$ and $w_1 = \begin{bmatrix} \|x_1\|_2 & 0 & 0 \end{bmatrix}^\top = \begin{bmatrix} 3 & 0 & 0 \end{bmatrix}^\top$;
- $u_1 = w_1 - x_1 = \begin{bmatrix} 3 & 0 & 0 \end{bmatrix}^\top - \begin{bmatrix} 1 & 2 & 2 \end{bmatrix}^\top = \begin{bmatrix} 2 & -2 & -2 \end{bmatrix}^\top$;
- $v_1 = \frac{u_1}{\|u_1\|_2} = \begin{bmatrix} \frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3} \end{bmatrix}^\top$;
- $H_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - 2\begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$;
- $H_1 A = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & -3 \\ 0 & -4 \end{bmatrix}.$

Full QR factorization based on Householder reflectors

Example

- $x_2 = \begin{bmatrix} -3 & -4 \end{bmatrix}^\top$ and $w_2 = \begin{bmatrix} -\|x_2\|_2 & 0 \end{bmatrix}^\top = \begin{bmatrix} -5 & 0 \end{bmatrix}^\top$;
- $u_2 = w_2 - x_2 = \begin{bmatrix} -5 & 0 \end{bmatrix}^\top - \begin{bmatrix} -3 & -4 \end{bmatrix}^\top = \begin{bmatrix} -2 & 4 \end{bmatrix}^\top$;
- $v_2 = \frac{u_2}{\|u_2\|_2} = \begin{bmatrix} -\frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \end{bmatrix}^\top$;
- $\hat{H}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2\begin{bmatrix} \frac{2}{5} & -\frac{4}{5} \\ -\frac{4}{5} & \frac{8}{5} \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix}$;
- $R = H_2(H_1 A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{5} & \frac{4}{5} \\ 0 & \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & -3 \\ 0 & -4 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & -5 \\ 0 & 0 \end{bmatrix}$.

Full QR factorization based on Householder reflectors

Example

- $Q = H_1 H_2 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{5} & \frac{4}{5} \\ 0 & \frac{4}{5} & -\frac{3}{5} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{15} & -\frac{11}{15} \end{bmatrix}$
- $A = QR = \begin{bmatrix} \frac{1}{3} & \frac{14}{15} & \frac{2}{15} \\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & -\frac{2}{15} & -\frac{11}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & -5 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}.$

Conclusions

- Solve the inconsistent problem $Ax = b$;
- Stimulate functions via the least squares method;
- Reduce it to solving $A^\top Ax = A^\top b$;
- Reduce it $Ax = b$ to solving $Rx = Q^\top b$ via QR Factorization $A = QR$;
- Use Householder reflectors to implement QR Factorization.

Thank you!