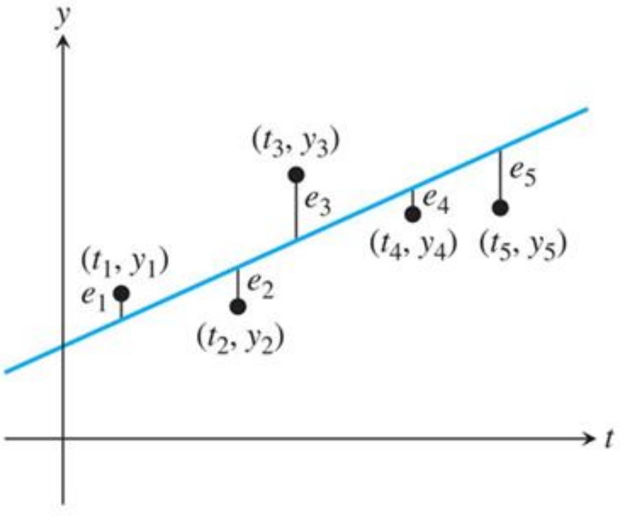
QR Factorization

Mei-Chen Yeh

Review: Normal equations for least squares

Given an inconsistent system



$$A\underline{x} = \underline{b},$$

solve

$$A^T A\underline{\tilde{x}} = A^T \underline{b}$$

for the least squares solution $\underline{\tilde{x}}$ that minimizes the Euclidean length of the residual $\underline{r} = \underline{b} - A\underline{\tilde{x}}$.

Review: Fitting data

$$A^T A \underline{\tilde{x}} = A^T \underline{b}$$

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

1. Choose a model. Example: $y = c_1 + c_2 t$
2. Force the model to fit the data
   - Let the unknown x represents the model parameters
   - #unknowns: #model parameters
   - #equations: m
3. Solve the normal equations
   - $A^T A \underline{\tilde{x}} = A^T \underline{b}$

$$A^T A \underline{\tilde{x}} = A^T \underline{b}$$

- How accurately can be the least squares solution $\underline{\tilde{x}}$ be determined?

Example

- $x_1 = 2.0$, $x_2 = 2.2$, $x_3 = 2.4$, …, $x_{11} = 4.0$
- $y_i = 1 + x_i + x_i^2 + \ldots + x_i^7$

Find the least squared polynomial $P(x) = c_1 + c_2 x + \ldots + c_8 x^7$ fitting (x_i, y_i)

- What are the coefficients c_i?

$x_1 = 2.0$, $x_2 = 2.2$, $x_3 = 2.4$, …, $x_{11} = 4.0$

$y_i = 1 + x_i + x_i^2 + \ldots + x_i^7$

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

Least squares solution:

#unknown?

#equations?

$$\underbrace{\begin{bmatrix} 1 & \cdots & x_1^7 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & x_{11}^7 \end{bmatrix}}_{11\text{x}8} \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_8 \end{bmatrix}}_{8\text{x}1} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{11} \end{bmatrix}}_{11\text{x}1}$$

Get $\underline{c}$ =
1.5134
-0.2644
2.3211
0.2408
0.9474
1.0059
0.9997

Solve the normal equation with MATLAB (double precision)

*Solving the normal equations in double precision **cannot** deliver an accurate value for the least squares solution!*

Conditioning of least squares

- Recall that we compute cond(A) for error estimation on solving $A^{\mathrm{T}}\underline{x} = \underline{b}$

$$\text{error magnification factor} = \frac{\text{relative forward error}}{\text{relative backward error}} = \text{cond}(A)$$

$$= \|A\| \times \|A^{-1}\|$$

Back to the example

$$\boxed{A^{\mathrm{T}} A \underline{\tilde{x}} = A^{\mathrm{T}} \underline{b}}$$

- cond($A^{\mathrm{T}}A$) = 1.4359e+019
  - Too large to deal with in double precision arithmetic
  - The normal equations are ill-conditioned!
- Remedy: avoid forming $A^{\mathrm{T}}A$

Today

- QR factorization
- Gram-Schmidt orthogonalization

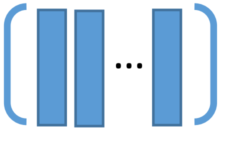
Analogy to LU factorization

- Solving matrix equations: LU factorization
- What are the benefits using LU?
- Solving least squares: **QR factorization**

Preliminaries

- Orthogonal set
  - A set of vectors in which $v_i^T v_j = 0$ whenever $i \neq j$.
  - Example: {[1, 1, 1]^T, [2, 1, -3]^T, [4, -5, 1]^T}
- Orthonormal set
  - An orthogonal set of ***unit vectors***.

    a vector of length 1
  - $\|v\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = 1$
  - Example: {[0, 0, 1]^T, [0, 1, 0]^T, [1, 0, 0]^T}
  - Normalizing a vector? $u/\|u\|_2$
- Orthogonal matrix
  - The column vectors form an orthonormal set.
  - $Q^{-1} = Q^T$



QR Factorization: Output

- Given a matrix A (m x n)
- Reduced QR factorization

q_i: mutually perpendicular unit vectors (orthonormal set)

$$\underbrace{(A_1|\cdots|A_n)}_{m \times n} = \underbrace{\boxed{(q_1|\cdots|q_n)}}_{m \times n} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}}_{n \times n}, \tag{4.26}$$

- Full QR factorization

$$\underbrace{(A_1|\cdots|A_n)}_{m \times n} = \underbrace{\boxed{(q_1|\cdots|q_m)}}_{m \times m} \underbrace{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}}_{m \times n}. \tag{4.27}$$

QR for solving least squares

QR vs. Normal equations?

Avoid computing A^TA

- $A = QR$
  - Q is orthogonal $\implies Q^{-1} = Q^T$
  - R is upper-triangular
- Given an m x n inconsistent system $A\underline{x} = \underline{b}$
  - $A\underline{x} = \underline{b}$
  - $QR\underline{x} = \underline{b}$
  - $Q^{-1}QR\underline{x} = Q^{-1}\underline{b} = Q^T\underline{b}$
  - $R\underline{x} = Q^T\underline{b}$ directly using back substitution to solve $\underline{x}$ ☺

$$\underset{m \times n}{(A_1|\cdots|A_n)} = \underset{m \times n}{(q_1|\cdots|q_n)} \underset{n \times n}{\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}}, \qquad (4.26)$$

Gram-Schmidt method

Orthogonalizes a set of vectors

Input: n linearly independent input vectors (A_i)

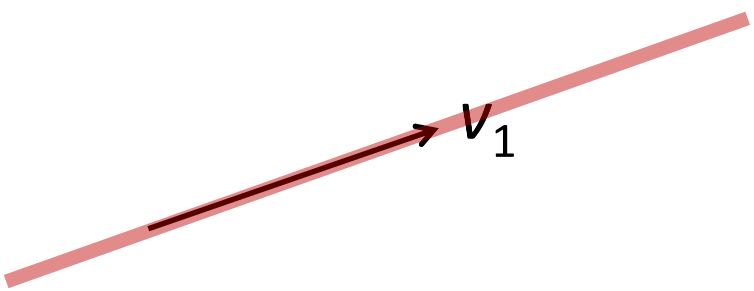
Output: n mutually perpendicular unit vectors spanning the same space (q_i)

$$(A_1|\cdots|A_n) = \boxed{(q_1|\cdots|q_n)} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}, \tag{4.26}$$

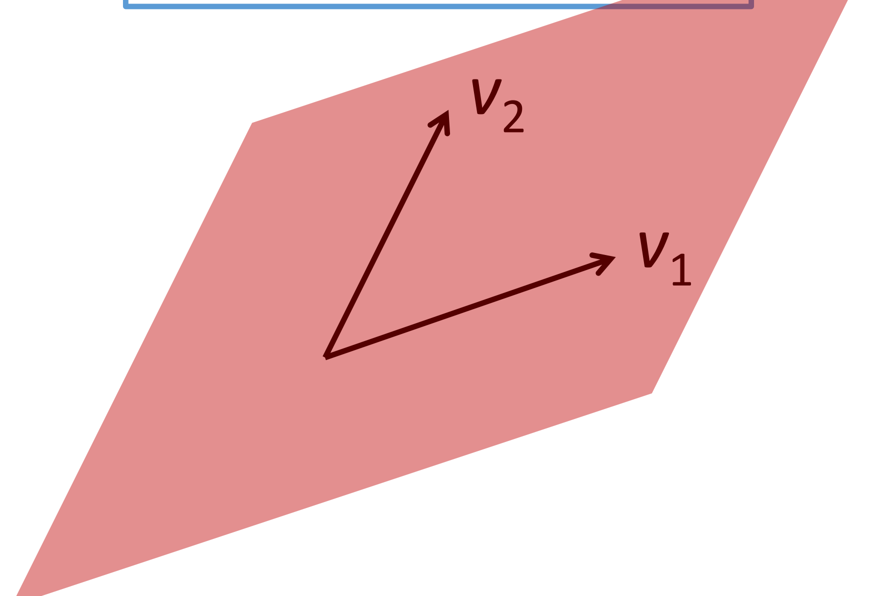
Preliminaries

- Span: The set of all linear combinations of $v_1, \ldots, v_n$ is the *span* of $v_1, \ldots, v_n$.
- Examples

1-dimensional case



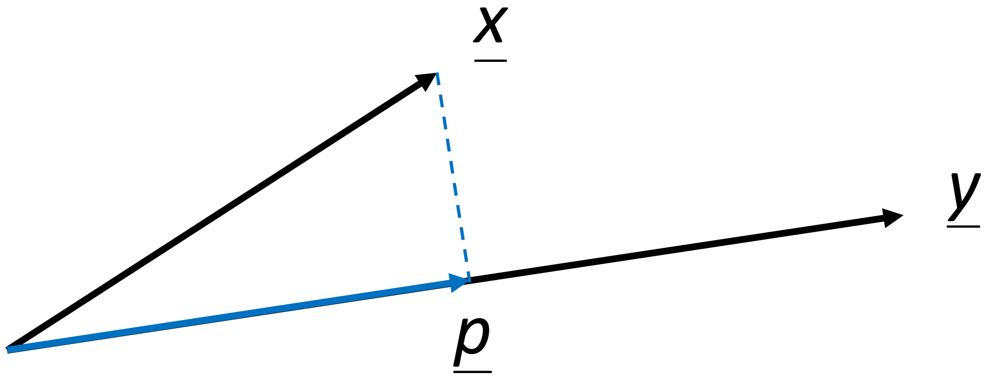
2-dimensional case



Preliminaries

- Vector project of $\underline{x}$ onto $\underline{y}$:

$$\underline{p} = (x^T y)\frac{y}{\|y\|_2}$$



Gram-Schmidt method

- https://www.khanacademy.org/math/linear-algebra/alternate_bases/orthonormal_basis/v/linear-algebra-the-gram-schmidt-process

Gram-Schmidt method

Input: n linearly independent input vectors $\{A_i\}$

Output: n mutually perpendicular unit vectors spanning the same space $\{q_i\}$

- 1st unit vector: $y_1 = A_1 \quad \text{and} \quad q_1 = \frac{y_1}{||y_1||_2}.$
- 2nd unit vector:

$$y_2 = A_2 - q_1(q_1^T A_2), \quad \text{and} \quad q_2 = \frac{y_2}{||y_2||_2}.$$

- *j*-th unit vector:

$$y_j = A_j - q_1(q_1^T A_j) - q_2(q_2^T A_j) - \ldots - q_{j-1}(q_{j-1}^T A_j) \quad \text{and} \quad q_j = \frac{y_j}{||y_j||_2}.$$

Example

Orthogonal?

- Find the **reduced** QR factorization of A = $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$
- 3-dimension, 2 column vectors
- Solution:

$$y_1 = A_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \qquad r_{11} = ||y_1||_2 = \sqrt{1^2 + 2^2 + 2^2} = 3,$$

$$q_1 = \frac{y_1}{||y_1||_2} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}$$

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

- Then, find the 2nd unit vector

$$y_2 = A_2 - \underbrace{q_1 q_1^T A_2}_{r_{12}} = \begin{bmatrix} -4 \\ 3 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} 2 = \begin{bmatrix} -\frac{14}{3} \\ \frac{5}{3} \\ \frac{2}{3} \end{bmatrix}$$

$$q_2 = \frac{y_2}{\underbrace{||y_2||_2}_{r_{22}}} = \frac{1}{5} \begin{bmatrix} -\frac{14}{3} \\ \frac{5}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ \frac{2}{15} \end{bmatrix}$$

Orthogonal?

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 1/3 & -14/15 \\ 2/3 & 1/3 \\ 2/3 & 2/15 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix} = QR$$

q_1 q_2

$$r_{11} = ||y_1||_2 = \sqrt{1^2 + 2^2 + 2^2} = 3,$$

$$r_{12} = q_1^T A_2 = 2$$

$$r_{22} = ||y_2||_2 = 5$$

$$\boxed{\begin{aligned} r_{jj} &= \|y\|_2 \\ r_{ij} &= q_i^T A_j \end{aligned}}$$

from reduced to full QR

- Find the **full** QR factorization of A = $\begin{bmatrix} 1 & -4 \\ 2 & 3 \\ 2 & 2 \end{bmatrix}$
- Add a 3rd vector $A_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ (arbitrary)

$$y_3 = A_3 - q_1 q_1^T A_3 - q_2 q_2^T A_3$$

$$= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \frac{1}{3} - \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ -\frac{2}{15} \end{bmatrix} \left(-\frac{14}{15}\right) = \frac{2}{225} \begin{bmatrix} 2 \\ 10 \\ -11 \end{bmatrix}$$

$$y_3 = A_3 - q_1 q_1^T A_3 - q_2 q_2^T A_3$$

$$= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \frac{1}{3} - \begin{bmatrix} -\frac{14}{15} \\ \frac{1}{3} \\ -\frac{2}{15} \end{bmatrix} \left(-\frac{14}{15}\right) = \frac{2}{225} \begin{bmatrix} 2 \\ 10 \\ -11 \end{bmatrix}$$

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

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

q_3

Analogy

- LU factorization → recording the information of Gaussian elimination
- QR factorization → ?

  recording the orthogonalization of a matrix!

The ill-conditioned A^TA example

- $x_1 = 2.0$, $x_2 = 2.2$, $x_3 = 2.4$, ..., $x_{11} = 4.0$
- $y_i = 1 + x_i + x_i^2 + \ldots + x_i^7$
- Find the least squared polynomial $P(x) = c_1 + c_2x + \ldots + c_8x^7$ fitting (x_i, y_i)
- What are the coefficients c_i?

$x_1 = 2.0, x_2 = 2.2, x_3 = 2.4, \ldots, x_{11} = 4.0$

$y_i = 1 + x_i + x_i^2 + \ldots + x_i^7$

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

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^7 \\ 1 & x_2 & x_2^2 & \cdots & x_2^7 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{11} & x_{11}^2 & \cdots & x_{11}^7 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_8 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{11} \end{bmatrix}$$

```
>> x = (2+(0:10)/5)';
>> y = 1+x+x.^2+x.^3+x.^4+x.^5+x.^6+x.^7;
>> A = [x.^0 x x.^2 x.^3 x.^4 x.^5 x.^6 x.^7];
>> [Q, R] = qr(A);
>>  Q'*y;
>> c = R(1:8, 1:8)\b(1:8);
```

Get $\underline{c}$ =

1.000

1.000

1.000

1.000

1.000

1.000

1.000

程式練習

And, please upload your program on moodle.

- Apply Gram-Schmidt orthogonalization to find the full QR factorization of **one** of the matrices:

$$\begin{bmatrix} 4 & 0 \\ 3 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 4 & 8 & 1 \\ 0 & 2 & -2 \\ 3 & 6 & 7 \end{bmatrix}$$