

The Gram-Schmidt orth. via Orthogonal Projectors

Note Title

LECTURE 10

To understand the behavior of the classical GS algorithm and to discuss the better version, let's view the classical GS alg. using ortho. projectors.

$A \in \mathbb{R}^{m \times n}$, $m \geq n$, full rank
i.e., $\text{rank}(A) = n$.

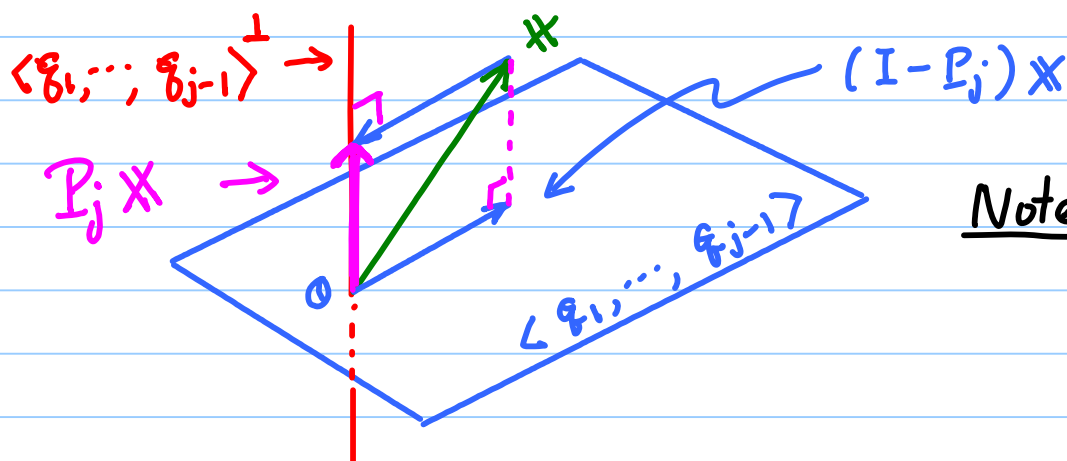
$$\begin{aligned} \xi_1 &= \frac{a_1}{r_{11}}, \quad \xi_2 = \frac{a_2 - r_{12}\xi_1}{r_{22}}, \quad \dots, \quad \xi_n = \frac{a_n - \sum_{i=1}^{n-1} r_{in}\xi_i}{r_{nn}} \\ &\quad \updownarrow \quad \quad \quad \updownarrow \quad \quad \quad \updownarrow \\ \xi_1 &= \frac{P_1 a_1}{\|P_1 a_1\|}, \quad \xi_2 = \frac{P_2 a_2}{\|P_2 a_2\|}, \quad \dots, \quad \xi_n = \frac{P_n a_n}{\|P_n a_n\|} \end{aligned}$$

where $P_j = \text{Ortho. proj. onto}$

$$\langle \xi_1, \dots, \xi_{j-1} \rangle^\perp$$

$$j = 1, 2, \dots, n$$

orthogonal complement
to $\langle \xi_1, \dots, \xi_{j-1} \rangle$



Note: $P_1 = I$.

$$\mathbb{R}^m = \langle \xi_1, \dots, \xi_{j-1} \rangle \oplus \langle \xi_1, \dots, \xi_{j-1} \rangle^\perp$$

$$= \underbrace{\text{null}(P_j)}_{\dim = j-1} \oplus \underbrace{\text{range}(P_j)}_{\dim = m-(j-1)}$$

LECTURE 10

Note: $\xi_j \perp \langle \xi_1, \dots, \xi_{j-1} \rangle$,

$$\xi_j \in \langle a_1, \dots, a_j \rangle,$$

and $\|\xi_j\| = 1$, by construction.

Now let $\hat{Q}_{j-1} := [\xi_1 \cdots \xi_{j-1}] \in \mathbb{R}^{m \times (j-1)}$

Then clearly, $\underline{P_j} = I - \hat{Q}_{j-1} \hat{Q}_{j-1}^T$, $j > 1$

$$P_1 = I.$$

* Modified Gram-Schmidt Algorithm

Recall the CGS algorithm:

for $j = 1:n$

$$\left\{ \begin{array}{l} v_j = a_j \\ \text{for } i = 1:j-1 \\ \quad \left\{ \begin{array}{l} r_{ij} = \xi_i^T a_j \\ v_j = v_j - r_{ij} \xi_i \end{array} \right. \\ r_{jj} = \|v_j\| \\ \xi_j = v_j / r_{jj} \end{array} \right.$$

This part
computes
 $\underline{P_j} a_j$ and
store it as v_j

Since $\text{rank}(P_j) = m - (j-1)$,
rank of $P_j \downarrow$ as $j \uparrow$, which is
not good. Also, numerical error
accumulates in the inner "for" loop.

LECTURE 10

The modified GS (MGS) algorithm
"uses the fresh material immediately
rather than waiting to avoid staleness."

What I mean above is :

to use

$$\begin{cases} P_j = P_{\perp f_{j-1}} P_{\perp f_{j-2}} \cdots P_{\perp f_1} ; j > 1 \\ P_1 = I \end{cases}$$

Note that each $P_{\perp f_i}$ has rank $m-1$

$P_{\perp f_i}$ = the complementary projection
to $\underline{P_{f_i}}$

$$= \underline{I - f_i f_i^T}$$

Mathematically,

$$v_j = P_j a_j = (I - \hat{Q}_{j-1} \hat{Q}_{j-1}^T) a_j$$

and

$$\begin{aligned} v_j &= P_j a_j = P_{\perp f_{j-1}} \cdots P_{\perp f_1} a_j \\ &= (I - f_{j-1} f_{j-1}^T) \cdots (I - f_1 f_1^T) a_j \end{aligned}$$

are equivalent. But the sequence
of arithmetic operations are different.
The MGS computes and updates:

$$v_j^{(1)} = a_j \rightarrow f_1$$

$$v_j^{(2)} = P_{\perp f_1} v_j^{(1)} = v_j^{(1)} - f_1 f_1^T v_j^{(1)} \rightarrow f_2$$

$$v_j^{(3)} = P_{\perp f_2} v_j^{(2)} = v_j^{(2)} - f_2 f_2^T v_j^{(2)} \rightarrow f_3$$

$\vdots$

LECTURE 10

$$\begin{aligned} \vdots \\ \psi_j^{(j)} &= \mathbb{P}_{\perp \xi_{j-1}} \psi_j^{(j-1)} = \psi_j^{(j-1)} - \xi_{j-1} \xi_{j-1}^T \psi_j^{(j-1)} \\ &\rightarrow \psi_j \rightarrow \xi_j \end{aligned}$$

This process is applied for $j=1, \dots, n$

Algorithm (Modified Gram-Schmidt)

```

for i = 1 : n
     $\psi_i = a_i$ 
    for j = i+1 : n
         $r_{ij} = \xi_i^T \psi_j$ 
         $\psi_j = \psi_j - r_{ij} \xi_i$ 
    end
     $r_{ii} = \|\psi_i\|$ 
     $\xi_i = \psi_i / r_{ii}$ 
end
    
```

Note the order of r_{ij} computation

$i :$	①	r_{11}	r_{12}	r_{13}	$\dots$	r_{1n}	→
	②		r_{22}	r_{23}	$\dots$	r_{2n}	→
	③			r_{33}	$\dots$	r_{3n}	→
	$\vdots$					$\vdots$	
	④					r_{nn}	→

LECTURE 10

Let's consider the previous example.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{bmatrix} \quad \begin{array}{l} \varepsilon : \text{small} \\ \text{s.t. } \varepsilon^2 \text{ can} \\ \text{be ignored.} \end{array}$$

Now apply the MGS algorithm!

$$v_j^{(1)} = a_j, \quad j=1, 2, 3$$

$$r_{11} = \|v_1^{(1)}\| = \sqrt{1+\varepsilon^2} \approx 1.$$

$$g_1 = v_1^{(1)} / r_{11} = [1 \ \varepsilon \ 0 \ 0]^T$$

Now immediately compute r_{12}, r_{13} :

$$\left\{ \begin{array}{l} r_{12} = g_1^T v_2^{(1)} = [1 \ \varepsilon \ 0 \ 0] \begin{bmatrix} 1 \\ 0 \\ \varepsilon \\ 0 \end{bmatrix} = 1 \\ v_2^{(2)} = v_2^{(1)} - r_{12} g_1 \\ \quad = \begin{bmatrix} 1 \\ 0 \\ \varepsilon \\ 0 \end{bmatrix} - 1 \cdot \begin{bmatrix} 1 \\ \varepsilon \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\varepsilon \\ \varepsilon \\ 0 \end{bmatrix} \\ r_{13} = g_1^T v_3^{(1)} = [1 \ \varepsilon \ 0 \ 0] \begin{bmatrix} 1 \\ 0 \\ 0 \\ \varepsilon \end{bmatrix} = 1 \\ v_3^{(2)} = v_3^{(1)} - r_{13} g_1 \\ \quad = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \varepsilon \end{bmatrix} - 1 \cdot \begin{bmatrix} 1 \\ \varepsilon \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\varepsilon \\ 0 \\ \varepsilon \end{bmatrix} \end{array} \right.$$

$$r_{22} = \|v_2^{(2)}\| = \sqrt{2} \varepsilon$$

$$g_2 = v_2^{(2)} / r_{22} = [0 \ -1/\sqrt{2} \ 1/\sqrt{2} \ 0]^T$$

Now immediately compute r_{23} :

$$\left\{ \begin{array}{l} r_{23} = g_2^T v_3^{(2)} = [0 \ -1/\sqrt{2} \ 1/\sqrt{2} \ 0] \begin{bmatrix} 0 \\ -\varepsilon \\ 0 \\ \varepsilon \end{bmatrix} = \frac{\varepsilon}{\sqrt{2}} \end{array} \right.$$

LECTURE 10

$$\begin{aligned} \psi_3^{(3)} &= \psi_3^{(2)} - r_{23} \phi_2 \\ &= \begin{bmatrix} 0 \\ -\varepsilon \\ 0 \\ \varepsilon \end{bmatrix} - \frac{\varepsilon}{\sqrt{2}} \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\varepsilon/2 \\ -\varepsilon/2 \\ \varepsilon \end{bmatrix} \end{aligned}$$

$$r_{33} = \|\psi_3^{(3)}\| = \varepsilon \cdot \sqrt{(-\frac{1}{2})^2 + (-\frac{1}{2})^2 + 1^2} = \sqrt{\frac{3}{2}} \varepsilon$$

$$\phi_3 = \psi_3^{(3)} / r_{33} = \begin{bmatrix} 0 \\ -1/\sqrt{6} \\ -1/\sqrt{6} \\ \sqrt{2/3} \end{bmatrix}$$

Hence

$$\hat{Q} = \begin{bmatrix} 1 & 0 & 0 \\ \varepsilon & -1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 0 & \sqrt{2/3} \end{bmatrix} \quad \hat{R} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & \sqrt{2} \varepsilon & \varepsilon/\sqrt{2} \\ 0 & 0 & \sqrt{3/2} \varepsilon \end{bmatrix}$$

Notice that $A = \hat{Q} \hat{R}$ holds.

Moreover,

$$\hat{Q}^T \hat{Q} = \begin{bmatrix} 1 + \varepsilon^2 & -\varepsilon/\sqrt{2} & -\varepsilon/\sqrt{6} \\ -\varepsilon/\sqrt{2} & 1 & 0 \\ -\varepsilon/\sqrt{6} & 0 & 1 \end{bmatrix}$$

This is
closer to
 $I_{3 \times 3}$

$\approx$

$$\begin{bmatrix} 1 & -\varepsilon/\sqrt{2} & -\varepsilon/\sqrt{6} \\ -\varepsilon/\sqrt{2} & 1 & 0 \\ -\varepsilon/\sqrt{6} & 0 & 1 \end{bmatrix}$$

than
this

Compare this with the CGS result:

$$\hat{Q}^T \hat{Q} \approx \begin{bmatrix} 1 & -\varepsilon/\sqrt{2} & -\varepsilon/\sqrt{2} \\ -\varepsilon/\sqrt{2} & 1 & 1/2 \\ -\varepsilon/\sqrt{2} & 1/2 & 1 \end{bmatrix}$$

LECTURE 10

- Note: The best algorithm for QR factorization is the so-called "Householder Triangularization" which will be discussed in the next lecture.

★ Application to the LS problem

Recall the solution $\mathbf{x} \in \mathbb{R}^n$ of the LS problem: $\|\mathbf{b} - A\mathbf{x}\|^2 \rightarrow \min$ where $A \in \mathbb{R}^{m \times n}$, $m \geq n$, $\mathbf{b} \in \mathbb{R}^m$, satisfies the normal eqn.

$$A^T A \mathbf{x} = A^T \mathbf{b}.$$

(Suppose A is full rank)

Now plug in the reduced QR fact.

of $A = \hat{Q} \hat{R}$

$$\Leftrightarrow \hat{R}^T \underbrace{\hat{Q}^T \hat{Q}}_{=I} \hat{R} \mathbf{x} = \hat{R}^T \hat{Q}^T \mathbf{b}$$

$$\Leftrightarrow \hat{R}^T \hat{R} \mathbf{x} = \hat{R}^T \hat{Q}^T \mathbf{b}$$

Now notice that $\hat{R}^T$ is the same on both sides and it's

nonsingular. So we can remove it to get

If A is full rank $\rightarrow$

$$\underline{\hat{R} \mathbf{x} = \hat{Q}^T \mathbf{b}}$$

LECTURE 10

So, the LS solution via QR proceeds:

- (1) Compute reduced QR of A .
- (2) Compute $y = \hat{Q}^T b$.
- (3) Solve $\hat{R}x = y$

Note that $\hat{R}$: upper triangular helps solve this system (3)

→ back substitution

$$\begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & r_{n-1,n-1} & r_{n-1,n} \\ 0 & \dots & 0 & r_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{bmatrix}$$

Start solving $r_{nn}x_n = y_n$

$$x_n = y_n / r_{nn}$$

then go backward:

$$\begin{aligned} r_{n-1,n-1}x_{n-1} + r_{n-1,n}x_n &= y_{n-1} \\ \Rightarrow x_{n-1} &= \frac{1}{r_{n-1,n-1}} (y_{n-1} - r_{n-1,n}x_n) \end{aligned}$$

Next solve for $x_{n-2}, \dots$, up to x_1 //
Direct consequence of the GS procedure!

Existence & Uniqueness of QR

Thm Every $A \in \mathbb{R}^{m \times n}$ ($m \geq n$) has a full QR factorization, hence also $\hat{Q}\hat{R}$.

Thm Each $A \in \mathbb{R}^{m \times n}$ ($m \geq n$) of full rank matrix has a unique $\hat{Q}\hat{R}$ with $r_{ii} > 0$.