The Gram-Schmidt orth. via Orthogonal Projectors

Note Title

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.

$A \in \mathbb{R}^{m\times n}$, $m \ge n$, full rank

i.e., $\operatorname{rank}(A) = n$.

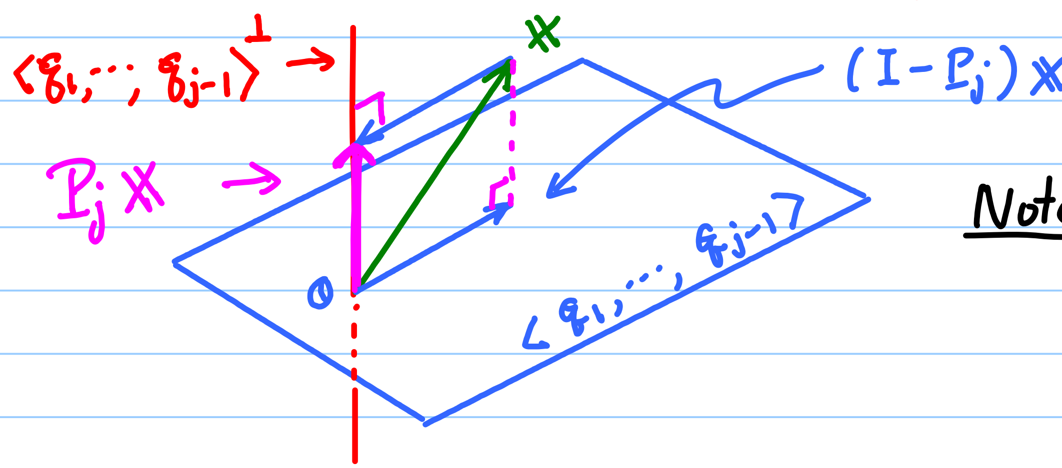
$$q_1 = \frac{a_1}{r_{11}}, \quad q_2 = \frac{a_2 - r_{12} q_1}{r_{22}}, \quad \cdots, \quad q_n = \frac{a_n - \sum_{i=1}^{n-1} r_{in} q_i}{r_{nn}}$$

$$\updownarrow \qquad\qquad \updownarrow \qquad\qquad \updownarrow$$

$$q_1 = \frac{P_1 a_1}{\|P_1 a_1\|}, \quad q_2 = \frac{P_2 a_2}{\|P_2 a_2\|}, \quad \cdots, \quad q_n = \frac{P_n a_n}{\|P_n a_n\|}$$

where P_j = ortho. proj. onto $\langle q_1, \cdots, q_{j-1} \rangle^{\perp}$

$j = 1, 2, \cdots, n$ — orthogonal complement to $\langle q_1, \cdots, q_{j-1} \rangle$



Note : $P_1 = I$.

$$\mathbb{R}^m = \langle q_1, \cdots, q_{j-1} \rangle \oplus \langle q_1, \cdots, q_{j-1} \rangle^{\perp}$$

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

Note: $q_j \perp \langle q_1, \cdots, q_{j-1} \rangle$,

$q_j \in \langle a_1, \cdots, a_j \rangle$,

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

Now let $\hat{Q}_{j-1} := [\, q_1 \; \cdots \; q_{j-1} \,] \in \mathbb{R}^{m \times (j-1)}$

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

$P_1 = I$.

☆ Modified Gram-Schmidt Algorithm

Recall the CGS algorithm:

$$
\begin{aligned}
&\text{for } j = 1 : n \\
&\quad v_j = a_j \\
&\quad \text{for } i = 1 : j-1 \\
&\qquad r_{ij} = q_i^T a_j \\
&\qquad v_j = v_j - r_{ij} q_i \\
&\quad r_{jj} = \| v_j \| \\
&\quad q_j = v_j / r_{jj}
\end{aligned}
$$

This part computes $P_j a_j$ and store it as v_j

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

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

what I mean above is:
to use

$$\begin{cases} P_j = P_{\perp q_{j-1}} P_{\perp q_{j-2}} \cdots P_{\perp q_1} \ ; \ j>1 \\ P_1 = I \end{cases}$$

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

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

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

Mathematically,

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

and

$$v_j = P_j a_j = P_{\perp q_{j-1}} \cdots P_{\perp q_1} a_j$$
$$= (I - q_{j-1} q_{j-1}^T) \cdots (I - q_1 q_1^T) a_j$$

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

$$\begin{cases} v_j^{(1)} = a_j \rightarrow q_1 \\ v_j^{(2)} = P_{\perp q_1} v_j^{(1)} = v_j^{(1)} - q_1 q_1^T v_j^{(1)} \rightarrow q_2 \\ v_j^{(3)} = P_{\perp q_2} v_j^{(2)} = v_j^{(2)} - q_2 q_2^T v_j^{(2)} \rightarrow q_3 \\ \vdots \end{cases}$$

$$\vdots$$

$$v_j^{(j)} = P_{\perp q_{j-1}} v_j^{(j-1)} = v_j^{(j-1)} - q_{j-1} q_{j-1}^T v_j^{(j-1)}$$

$$\longrightarrow \; = v_j \longrightarrow q_j$$

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

Algorithm (Modified Gram-Schmidt)

for $i = 1 : n$

  $v_i = a_i$

for $i = 1 : n$

$$\left\{\begin{array}{l} r_{ii} = \| v_i \| \\ q_i = v_i / r_{ii} \\ \text{for } j = i+1 : n \\ \quad \left\{\begin{array}{l} r_{ij} = q_i^T v_j \\ v_j = v_j - r_{ij} q_i \end{array}\right. \end{array}\right.$$

Note the order of r_{ij} computation

$$\begin{array}{llllll} i: & 1 & r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \rightarrow \\ & 2 & & r_{22} & r_{23} & \cdots & r_{2n} \rightarrow \\ & 3 & & & r_{33} & \cdots & r_{3n} \rightarrow \\ & \vdots & & & & & \vdots \\ & n & & & & & r_{nn} \rightarrow \end{array}$$

Let's consider the previous example.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \varepsilon & 0 & 0 \\ 0 & \varepsilon & 0 \\ 0 & 0 & \varepsilon \end{bmatrix}.$$

ε : small s.t. ε^2 can be ignored.

Now apply the MGS algorithm!

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

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

$q_1 = v_1^{(1)} / r_{11} = [1 \ \varepsilon \ 0 \ 0]^T$

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

$$\begin{cases} r_{12} = q_1^T v_2^{(1)} = [1 \ \varepsilon \ 0 \ 0] \begin{bmatrix} 1 \\ 0 \\ \varepsilon \\ 0 \end{bmatrix} = 1 \\ v_2^{(2)} = v_2^{(1)} - r_{12} q_1 \\ \quad = \begin{bmatrix} 1 \\ 0 \\ \varepsilon \\ 0 \end{bmatrix} - 1 \cdot \begin{bmatrix} 1 \\ \varepsilon \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\varepsilon \\ \varepsilon \\ 0 \end{bmatrix} \\ r_{13} = q_1^T v_3^{(1)} = [1 \ \varepsilon \ 0 \ 0] \begin{bmatrix} 1 \\ 0 \\ 0 \\ \varepsilon \end{bmatrix} = 1 \\ v_3^{(2)} = v_3^{(1)} - r_{13} q_1 \\ \quad = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \varepsilon \end{bmatrix} - 1 \cdot \begin{bmatrix} 1 \\ \varepsilon \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\varepsilon \\ 0 \\ \varepsilon \end{bmatrix} \end{cases}$$

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

$q_2 = v_2^{(2)} / r_{22} = [0 \ -1/\sqrt{2} \ 1/\sqrt{2} \ 0]^T$

Now immediately compute r_{23}:

$$\begin{cases} r_{23} = q_2^T v_3^{(2)} = [0 \ -1/\sqrt{2} \ 1/\sqrt{2} \ 0] \begin{bmatrix} 0 \\ -\varepsilon \\ 0 \\ \varepsilon \end{bmatrix} = \dfrac{\varepsilon}{\sqrt{2}} \end{cases}$$

$$v_3^{(3)} = v_3^{(2)} - r_{23} q_2$$

$$= \begin{bmatrix} 0 \\ -\varepsilon \\ 0 \\ \varepsilon \end{bmatrix} - \frac{\varepsilon}{\sqrt{2}} \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\varepsilon/2 \\ -\varepsilon/2 \\ \varepsilon \end{bmatrix}$$

$$r_{33} = \| v_3^{(3)} \| = \varepsilon \cdot \sqrt{(-\tfrac{1}{2})^2 + (\tfrac{1}{2})^2 + 1^2} = \sqrt{\tfrac{3}{2}}\,\varepsilon$$

$$q_3 = v_3^{(3)} / r_{33} = \begin{bmatrix} 0 \\ -1/\sqrt{6} \\ -1/\sqrt{6} \\ \sqrt{2/3} \end{bmatrix}$$

Hence

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

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 $I_{3\times 3}$

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

than this

Compare this with the CGS result:

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

//

- 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: $\|\mathbf{b} - A\mathbf{x}\|_2 \to \min$ 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 $A = \hat{Q}\hat{R}$

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

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

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

If A is full rank → nonsingular. So we can remove it to get

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

So, the LS solution via QR proceeds:

$$\begin{cases} (1) \text{ Compute reduced QR of } A. \\ (2) \text{ Compute } \mathbb{y} = \hat{Q}^T \mathbb{b}. \\ (3) \text{ Solve } \hat{R}\,\mathbb{x} = \mathbb{y} \end{cases}$$

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

$\Rightarrow$ back substitution

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

Start solving $r_{nn} x_n = y_n$

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

then go backward:

$$r_{n-1\,n-1} x_{n-1} + r_{n-1\,n} x_n = y_{n-1}$$

$$\Rightarrow \quad x_{n-1} = \frac{1}{r_{n-1\,n-1}} \left(y_{n-1} - r_{n-1\,n} x_n \right)$$

Next solve for $x_{n-2}, \cdots,$ 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 $r_{ii} > 0$.