Lecture #25

§6.4 The Gram-Schmidt Process

Let's start right away with the main result:

Claim (Gram-Schmidt): Given a basis $\{\vec{x}_1, \ldots, \vec{x}_k\}$ for a nonzero subspace W of $\mathbb{R}^n$, define:

$$\vec{v}_1 := \vec{x}_1$$

$$\vec{v}_2 := \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1$$

$$\vec{v}_3 := \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2$$

$$\vdots$$

$$\vec{v}_k := \vec{x}_k - \frac{\vec{x}_k \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \frac{\vec{x}_k \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2 - \cdots - \frac{\vec{x}_k \cdot \vec{v}_{k-1}}{\vec{v}_{k-1} \cdot \vec{v}_{k-1}} \vec{v}_{k-1}$$

Then: $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is an orthogonal basis of W.

Moreover: $\mathrm{Span}\{\vec{v}_1, \ldots, \vec{v}_j\} = \mathrm{Span}\{\vec{x}_1, \ldots, \vec{x}_j\}$ for any $1 \le j \le k$.

Note: $\vec{v}_2 = \vec{x}_2 - \mathrm{proj}_{\vec{v}_1} \vec{x}_2$

$\vec{v}_3 = \vec{x}_3 - \mathrm{proj}_{\mathrm{span}\{\vec{v}_1, \vec{v}_2\}} \vec{x}_3$ hence the above claim is almost obvious!

$\vdots$

$\vec{v}_k = \vec{x}_k - \mathrm{proj}_{\mathrm{span}\{\vec{v}_1, \ldots, \vec{v}_{k-1}\}} \vec{x}_k$

Remark: To get an orthonormal basis of W, you first find an orthogonal basis by the above algorithm, and then normalize each basis vector, i.e. set $\vec{u}_i := \frac{\vec{v}_i}{\|\vec{v}_i\|}$, $1 \le i \le k$.

Ex 1: Find an orthonormal basis of $W = \mathrm{span}\left\{ \underset{\vec{x}_2}{\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}}, \underset{\vec{x}_1}{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} \right\} \subseteq \mathbb{R}^3$.

$\vec{v}_1 = \vec{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \Rightarrow \vec{u}_1 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}$

$\vec{v}_2 = \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - 2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \Rightarrow \vec{u}_2 = \begin{pmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}$

So: the above $\vec{u}_1$ & $\vec{u}_2$ form an orthonormal basis of W ■

Lecture #25

Ex 2 (see Ex 10 on p. 381 of the textbook): Find an orthogonal basis for the column space of

$$A = \begin{pmatrix} -1 & 6 & 6 \\ 3 & -8 & 3 \\ 1 & -2 & 6 \\ 1 & -4 & -3 \end{pmatrix}$$

► $\vec{v_1} = \vec{x_1} = \begin{pmatrix} -1 \\ 3 \\ 1 \\ 1 \end{pmatrix}$

$$\vec{v_2} = \vec{x_2} - \frac{\vec{x_2}\cdot\vec{v_1}}{\vec{v_1}\cdot\vec{v_1}}\vec{v_1} = \begin{pmatrix} 6 \\ -8 \\ -2 \\ -4 \end{pmatrix} - \frac{-36}{12}\begin{pmatrix} -1 \\ 3 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

$$\vec{v_3} = \vec{x_3} - \frac{\vec{x_3}\cdot\vec{v_1}}{\vec{v_1}\cdot\vec{v_1}}\vec{v_1} - \frac{\vec{x_3}\cdot\vec{v_2}}{\vec{v_2}\cdot\vec{v_2}}\vec{v_2} = \begin{pmatrix} 6 \\ 3 \\ 6 \\ -3 \end{pmatrix} - \frac{6}{12}\begin{pmatrix} -1 \\ 3 \\ 1 \\ 1 \end{pmatrix} - \frac{30}{12}\begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 3 \\ -1 \end{pmatrix}$$

■

Ex 3: Same question but for $B = \begin{pmatrix} -1 & 6 & 3 \\ 3 & -8 & 1 \\ 1 & -2 & 1 \\ 1 & -4 & -1 \end{pmatrix}$

► As in Ex 4, we find $\vec{v_1} = \begin{pmatrix} -1 \\ 3 \\ 1 \\ 1 \end{pmatrix}$, $\vec{v_2} = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix}$

Now, $\vec{v_3} = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix} - 0\cdot\vec{v_1} - \frac{12}{12}\begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \vec{0}$. Hence, we ignore $\vec{v_3}$ at all!

So: $\{\vec{v_1}, \vec{v_2}\}$ – an orthogonal basis of Col B. ■

Rem 1: If you start from a spanning set $\vec{x_1}, \dots, \vec{x_k}$ of W (not a basis) then you apply Gram-Schmidt the same way, but each time you get some $\vec{v_i} = \vec{0}$, you SKIP it!

Rem 2: If some $\vec{v_i}$ involve fractions, it is convenient to multiply by the common denominator of its coordinates to end up with a better looking vector

Lecture #25

> **Claim** (QR factorization): If A is an $m \times n$ matrix with linearly independent columns then A can be factored as $A = Q \cdot R$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\mathrm{Col}A$, and R is an $n \times n$ upper triangular invertible matrix with positive entries on diagonal

Let's illustrate the algorithm of finding Q, R on the above example.

Ex4: Find a QR factorization of $A = \begin{pmatrix} -1 & 6 & 6 \\ 3 & -8 & 3 \\ 1 & -2 & 6 \\ 1 & -4 & -3 \end{pmatrix}$

- **Step 1**: Find an orthonormal basis of $\mathrm{Col}\,A$ (& check it consists of 3 els)

  According to Ex 4, we can just take normalized basis we found there:

$$\vec{u}_1 = \begin{pmatrix} -1/\sqrt{12} \\ 3/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \end{pmatrix}, \quad \vec{u}_2 = \begin{pmatrix} 3/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \\ -1/\sqrt{12} \end{pmatrix}, \quad \vec{u}_3 = \begin{pmatrix} -1/\sqrt{12} \\ -1/\sqrt{12} \\ 3/\sqrt{12} \\ -1/\sqrt{12} \end{pmatrix}$$

So: $$Q = \underbrace{\begin{pmatrix} -1/\sqrt{12} & 3/\sqrt{12} & -1/\sqrt{12} \\ 3/\sqrt{12} & 1/\sqrt{12} & -1/\sqrt{12} \\ 1/\sqrt{12} & 1/\sqrt{12} & 3/\sqrt{12} \\ 1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} \end{pmatrix}}$$

Step 2: One could find all constants expressing $\vec{x}_j = r_{1j}\vec{u}_1 + \ldots + r_{jj}\vec{u}_j$.

But instead, we can also use the fact $Q^T Q = I_3$ to derive $R = Q^T Q R = Q^T A$, so that

$$R = \begin{pmatrix} -1/\sqrt{12} & 3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -1/\sqrt{12} \\ -1/\sqrt{12} & -1/\sqrt{12} & 3/\sqrt{12} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} -1 & 6 & 6 \\ 3 & -8 & 3 \\ 1 & -2 & 6 \\ 1 & -4 & -3 \end{pmatrix} = \underbrace{\begin{pmatrix} \sqrt{12} & -3\sqrt{12} & \frac{1}{2}\sqrt{12} \\ 0 & \sqrt{12} & \frac{5}{2}\sqrt{12} \\ 0 & 0 & \sqrt{12} \end{pmatrix}}$$

Upshot: $\boxed{R = Q^T A}$

Lecture #25

§6.5 Least-Squares Problems

In real life, often the linear systems $A\vec{x}=\vec{b}$ are inconsistent, but we want some best approximation of $\vec{b}$ by vectors of the form $A\vec{x}$.

Def: If A is an $m\times n$ matrix and $\vec{b}\in\mathbb{R}^m$, a least-squares solution of $A\vec{x}=\vec{b}$ is an $\hat{\vec{x}}\in\mathbb{R}^n$ such that
$$\|\vec{b}-A\hat{\vec{x}}\|\le\|\vec{b}-A\vec{x}\| \quad \text{for any } \vec{x}\in\mathbb{R}^n$$

Remark: This can be rephrased as looking for the point in Col A that is closest to the vector $\vec{b}$.

Key Observation: $A\hat{\vec{x}}$ must coincide with $\boxed{\underbrace{\text{proj}_{\text{Col}A}\,\vec{b}}_{\text{orthogonal projection of } \vec{b} \text{ onto Col} A} =: \hat{\vec{b}}}$

Note: As $\hat{\vec{b}}\in \text{Col}A$, the system $A\hat{\vec{x}}=\hat{\vec{b}}$ is compatible!

Note that $\vec{b}-A\hat{\vec{x}}=\vec{b}-\hat{\vec{b}}$ is orthogonal to each column $\vec{a}_j$ of the matrix A

$\Rightarrow A^T(\vec{b}-A\hat{\vec{x}})=0 \Rightarrow \boxed{A^TA\hat{\vec{x}}=A^T\vec{b}}$

the normal equations for $A\vec{x}=\vec{b}$.

Claim: The set of least-squares solutions of $A\vec{x}=\vec{b}$ coincides with the nonempty set of solutions of the normal equations
$$A^TA\vec{x}=A^T\vec{b}$$

In general, it may happen that $\hat{\vec{x}}$ is not unique. However, we have:

Claim: Let A be an $m\times n$ matrix. The following are equivalent:

a) The equation $A\vec{x}=\vec{b}$ has a unique least-squares solution for any $\vec{b}\in\mathbb{R}^m$

b) The columns of A are linearly independent

c) The matrix A^TA is invertible

When these statements are true, the least-squares solution $\hat{\vec{x}}$ is:
$$\hat{\vec{x}}=(A^TA)^{-1}A^T\vec{b}$$

Lecture #25

Ex 5 (see Ex 6.5.10 on p. 388): Find the orthogonal projection of $\vec{b}$ onto Col A and a least-squares solution of $A\vec{x} = \vec{b}$ for

$$\vec{b} = \begin{pmatrix} 3 \\ -1 \\ 5 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{pmatrix}$$

▸ $$A^T A = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 4 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 24 \end{pmatrix}$$

$$A^T \vec{b} = \begin{pmatrix} 1 & -1 & 1 \\ 2 & 4 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 9 \\ 12 \end{pmatrix}$$

So: $\begin{pmatrix} 3 & 0 \\ 0 & 24 \end{pmatrix} \hat{\vec{x}} = \begin{pmatrix} 9 \\ 12 \end{pmatrix} \implies \hat{\vec{x}} = \begin{pmatrix} 3 \\ 1/2 \end{pmatrix}$

Thus: $\hat{\vec{b}} = A\hat{\vec{x}} = \begin{pmatrix} 4 \\ -1 \\ 4 \end{pmatrix}$ ■

Rmk 1: The above proof is actually much quicker than if you first found $\hat{\vec{b}}$ and then $\hat{\vec{x}}$, as the former would need to find an orthogonal basis of Col A first

Rmk 2: While $\hat{\vec{b}}$ is unique, $\hat{\vec{x}}$ is not necessarily.

Rmk 3: In the above example, the columns of A were orthogonal! In such a case we know how to compute $\hat{\vec{b}}$ right away:

$$\hat{\vec{b}} = \frac{\vec{b}\cdot\vec{a_1}}{\vec{a_1}\cdot\vec{a_1}}\vec{a_1} + \frac{\vec{b}\cdot\vec{a_2}}{\vec{a_2}\cdot\vec{a_2}}\vec{a_2} = 3\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + \frac{12}{24}\begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ -1 \\ 4 \end{pmatrix}$$

Moreover, $\hat{\vec{x}}$ can be also read off from above by looking at coefficients of $\vec{a_1}, \vec{a_2}$, i.e. $\hat{\vec{x}} = \begin{pmatrix} 3 \\ 1/2 \end{pmatrix}$

Finally, when columns of A are lin. indep, there is another approach to finding the least-squares solution:

$\boxed{\begin{array}{l}\text{Claim: If the columns of an } m\times n \text{ matrix } A \text{ are lin. indep, then} \\ \text{the least-squares solution of the equation } A\vec{x}=\vec{b} \text{ is:} \\ \qquad \hat{\vec{x}} = R^{-1}Q^{T}\vec{b}, \\ \text{where } A=QR \text{ is a } QR \text{ factorization of } A.\end{array}}$

Note: this crucially relies on the equality $QQ^{T}\vec{b}=\hat{\vec{b}}=\mathrm{proj}_{\mathrm{Col}\,A}\vec{b}$ established last time (as columns of Q form an orthonormal set).