

• Please, fill out the course evaluation, which will be available starting from Nov 23 (due date is Dec 6).

• Last time → Orthonormal sets/bases
 ↳ $A^T \cdot A = I$ iff columns of A form an orthonormal set
 ↳ orthogonal projection of $\vec{y} \in \mathbb{R}^n$ onto a subspace $W \subseteq \mathbb{R}^n$.

Ex1: Let A be an $n \times n$ matrix whose columns form an orthonormal set in \mathbb{R}^n . Show that the rows of A form an orthonormal basis of \mathbb{R}^n .

↳ This follows immediately from $AA^T = I_n$ derived in the very end of the last lecture.

Ex2: Find the distance from $\vec{y} = \begin{pmatrix} 8 \\ 2 \\ 0 \end{pmatrix}$ to the subspace $W \subseteq \mathbb{R}^3$ spanned by $\vec{u}_1 = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$, $\vec{u}_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.

↳ Recall from [Lecture 24, Ex1] that \vec{u}_1 & \vec{u}_2 - orthogonal, hence, they form an orthogonal basis of W . Therefore,

$$\hat{\vec{y}} = \text{proj}_W \vec{y} = \frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1} \vec{u}_1 + \frac{\vec{y} \cdot \vec{u}_2}{\vec{u}_2 \cdot \vec{u}_2} \vec{u}_2 = \frac{14}{14} \vec{u}_1 + \frac{6}{3} \vec{u}_2 = \vec{u}_1 + 2\vec{u}_2 = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}$$

Thus: the distance from \vec{y} to W equals the length

$$\|\vec{y} - \hat{\vec{y}}\| = \left\| \begin{pmatrix} 8 \\ 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 5 \\ 1 \\ -4 \end{pmatrix} \right\| = \sqrt{42}$$

Now that we have learnt the importance of orthogonal/orthonormal sets/bases for all computations, a natural question arises:

Q: How to produce an orthogonal or orthonormal basis of a nonzero subspace $W \subseteq \mathbb{R}^n$?

The answer is given by the Gram-Schmidt process.

§6.4 The Gram-Schmidt Process

Let's start right away with the main result:

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

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

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

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

⋮

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

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

Moreover: $\text{Span}\{\vec{v}_1, \dots, \vec{v}_j\} = \text{Span}\{\vec{x}_1, \dots, \vec{x}_j\}$ for any $1 \leq j \leq k$.

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

$$\vec{v}_3 = \vec{x}_3 - \text{proj}_{\text{span}\{\vec{v}_1, \vec{v}_2\}} \vec{x}_3$$

⋮

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

hence the above claim is almost obvious!

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 $\vec{u}_i := \frac{\vec{v}_i}{\|\vec{v}_i\|}$, $1 \leq i \leq k$.

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

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

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

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

Lecture #25

EX4 (see Ex 10 on p. 381 of the textbook): Find an orthogonal basis for the column space of

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

$$\vec{v}_1 = \vec{x}_1 = \begin{pmatrix} -1 \\ 3 \\ 1 \\ 1 \end{pmatrix}$$

$$\vec{v}_2 = \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 = \begin{pmatrix} 6 \\ -8 \\ -2 \\ -4 \end{pmatrix} - \frac{-36}{12} \begin{pmatrix} -1 \\ 3 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

$$\vec{v}_3 = \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2 = \begin{pmatrix} 6 \\ 3 \\ 6 \\ -3 \end{pmatrix} - \frac{6}{12} \begin{pmatrix} -1 \\ 3 \\ 1 \\ 1 \end{pmatrix} - \frac{30}{12} \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 3 \\ -1 \end{pmatrix}$$

EX5: Same question but for $B = \begin{pmatrix} -1 & 6 & 3 \\ 3 & -8 & 1 \\ 1 & -2 & 1 \\ 1 & -4 & -1 \end{pmatrix}$

$$\text{As in Ex 4, we find } \vec{v}_1 = \begin{pmatrix} -1 \\ 3 \\ 1 \\ 1 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

$$\text{Now, } \vec{v}_3 = \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix} - 0 \cdot \vec{v}_1 - \frac{12}{12} \begin{pmatrix} 3 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \vec{0}. \text{ Hence, we ignore } \vec{v}_3 \text{ at all!}$$

So: $\{\vec{v}_1, \vec{v}_2\}$ is an orthogonal basis of $\text{Col } B$.

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 $\text{Col} A$, 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.

Ex 6: 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 $\text{Col} A$ (check it consists of 3 v's)

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

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

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

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

□

§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

$$\|\vec{b} - A\hat{\vec{x}}\| \leq \|\vec{b} - A\vec{x}\| \text{ for any } \vec{x} \in \mathbb{R}^n$$

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

Key Observation: $A\hat{\vec{x}}$ must coincide with proj_{Col A} $\vec{b} =: \hat{\vec{b}}$
orthogonal projection of \vec{b} onto $\text{Col } A$.

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

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

↑ 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

$$A^T A \vec{x} = A^T \vec{b}$$

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:

- The equation $A\vec{x} = \vec{b}$ has a unique least-squares solution for any $\vec{b} \in \mathbb{R}^m$
- The columns of A are linearly independent
- The matrix $A^T A$ is invertible

When these statements are true, the least-squares solution $\hat{\vec{x}}$ is:

$$\hat{\vec{x}} = (A^T A)^{-1} A^T \vec{b}$$

Lecture #25

Ex 7 (see Ex 6.5.10 on p. 388): Find the orthogonal projection of \vec{b} onto $\text{Col } A$ and a least-squares solution of $A\vec{x} = \vec{b}$ for

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

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

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

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

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

Remark 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 $\text{Col } A$ first.

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

Remark 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. $\hat{\vec{x}} = \begin{pmatrix} 3 \\ 1/2 \end{pmatrix}$.

Lecture #25

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

Claim: If the columns of an $m \times n$ matrix A are lin. indep, then the least-squares solution of the equation $A\vec{x} = \vec{b}$ is:

$$\hat{\vec{x}} = R^{-1} Q^T \vec{b},$$

where $A = QR$ is a QR factorization of A .

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