MA1508E: LINEAR ALGEBRA FOR ENGINEERING

Lecture 10 Notes

References

1. Elementary Linear Algebra: Application Version, Section 5.1, 6.3-6.4
2. Linear Algebra with Application, Section 3.3, 8.1

5.3 Gram-Schmidt Process

We will now provide an algorithm to construct orthogonal and orthonormal basis for a subspace $V \subseteq \mathbb{R}^n$.

Let $V \subseteq \mathbb{R}^n$ be a subspace, and suppose $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$is a basis for V. Our goal is to first construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ (from S), and then by normalizing, we obtain an orthonormal basis

$$\left\{\mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \mathbf{w}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, ..., \mathbf{w}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right\}.$$

We will construct the $\mathbf{v}_i$ inductively, for $i = 1, ..., k$. We will first let $\mathbf{v}_1 = \mathbf{u}_1$. Then $\text{span}\{\mathbf{v}_1\} = \text{span}\{\mathbf{u}_1\}$ and $\{\mathbf{v}_1\}$ is an orthogonal set.

Suppose we have constructed an orthogonal set $\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\}$ such that $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\} = \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}\}$. We then find the projection of $\mathbf{u}_i$ onto the subspace $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\} = \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}\}$. From lecture 9, we know that since $\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\}$ is an orthogonal set, the projection is given by

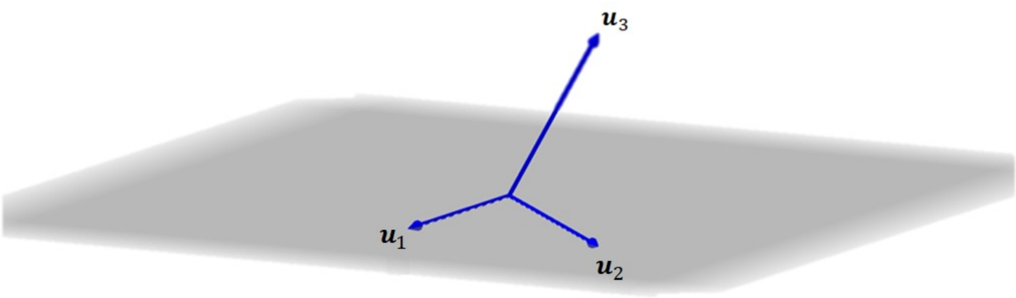
$$\left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_i}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 + \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_i}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 + \cdots + \left(\frac{\mathbf{v}_{i-1} \cdot \mathbf{u}_i}{\|\mathbf{v}_{i-1}\|^2}\right)\mathbf{v}_{i-1}.$$

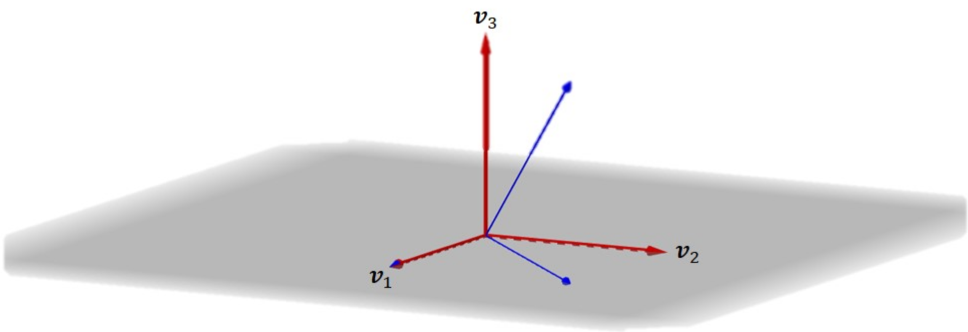
So, if we let

$$\mathbf{v}_i = \mathbf{u}_i - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_i}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_i}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_{i-1} \cdot \mathbf{u}_i}{\|\mathbf{v}_{i-1}\|^2}\right)\mathbf{v}_{i-1} \tag{4}$$

then $\mathbf{v}_i$ is the subtraction of $\mathbf{u}_i$ from its projection onto $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\} = \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}\}$. Thus, by construction, $\mathbf{v}_i$ is orthogonal to the subspace $\text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}\} = \text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}\}$ and so it is orthogonal to $\mathbf{v}_j$ for all $j < i$. Therefore $\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_i\}$ is an orthogonal set. By equation (4), $\mathbf{u}_i$ is a linear combination of $\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_i\}$, we have $\text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}, \mathbf{u}_i\} \subseteq \text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_i\}$. But since $\mathbf{v}_j$ is a linear combination of $\mathbf{v}_1, ..., \mathbf{v}_{i-1}$ for $j < i$, from equation (4), we can see that $\mathbf{v}_i$ is a linear combination of $\mathbf{u}_1, ..., \mathbf{u}_{i-1}, \mathbf{u}_i$. Thus $\text{span}\{\mathbf{u}_1, ..., \mathbf{u}_{i-1}, \mathbf{u}_i\} = \text{span}\{\mathbf{v}_1, ..., \mathbf{v}_{i-1}, \mathbf{v}_i\}$.

We will give a visualization of the Gram-Schmidt process in $\mathbb{R}^3$.





Theorem (Gram-Schmidt Process)**.** *Let* $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ *be a linearly independent set. Let*

$$
\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1 \\
\mathbf{v}_2 &= \mathbf{u}_2 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_2}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{u}_3 - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_3}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_3}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_i &= \mathbf{u}_i - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_i}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_i}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_{i-1} \cdot \mathbf{u}_i}{\|\mathbf{v}_{i-1}\|^2}\right)\mathbf{v}_{i-1} \\
&\vdots \\
\mathbf{v}_k &= \mathbf{u}_k - \left(\frac{\mathbf{v}_1 \cdot \mathbf{u}_k}{\|\mathbf{v}_1\|^2}\right)\mathbf{v}_1 - \left(\frac{\mathbf{v}_2 \cdot \mathbf{u}_k}{\|\mathbf{v}_2\|^2}\right)\mathbf{v}_2 - \cdots - \left(\frac{\mathbf{v}_{k-1} \cdot \mathbf{u}_k}{\|\mathbf{v}_{k-1}\|^2}\right)\mathbf{v}_{k-1}.
\end{aligned}
$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ *is an orthogonal set of nonzero vectors. Hence,*

$$\left\{\mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \mathbf{w}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, ..., \mathbf{w}_k = \frac{\mathbf{v}_k}{\|\mathbf{v}_k\|}\right\}$$

is an orthonormal set such that span$\{\mathbf{w}_1, ..., \mathbf{w}_k\}$ = *span*$\{\mathbf{u}_k, ..., \mathbf{u}_k\}$.

Exercise:

1. Show that $\mathbf{v}_i \neq \mathbf{0}$ for all $i = 1, ..., k$ (otherwise $\frac{1}{\|\mathbf{v}_i\|^2}$ is undefined).
2. What happens if S is not linearly independent?

Example. We will use the Gram-Schmidt process to convert $S = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \right\}$ into an orthonormal basis for $\mathbb{R}^3$.

$$
\begin{aligned}
\mathbf{v}_1 &= \begin{pmatrix}1\\2\\1\end{pmatrix}\\
\mathbf{v}_2 &= \begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{1+2+1}{1^2+2^2+1^2}\begin{pmatrix}1\\2\\1\end{pmatrix} = \frac{1}{3}\begin{pmatrix}1\\-1\\1\end{pmatrix}, \text{ let } \mathbf{v}_2 = \begin{pmatrix}1\\-1\\1\end{pmatrix} \text{ instead}\\
\mathbf{v}_3 &= \begin{pmatrix}1\\1\\2\end{pmatrix} - \frac{1+2+2}{1^2+2^2+1^2}\begin{pmatrix}1\\2\\1\end{pmatrix} - \frac{1-1+2}{1^2+(-1)^2+1^2}\begin{pmatrix}1\\-1\\1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}-1\\0\\1\end{pmatrix}\\
&\quad \text{let } \mathbf{v}_3 = \begin{pmatrix}-1\\0\\1\end{pmatrix} \text{ instead.}
\end{aligned}
$$

Why are we allowed to take $\mathbf{v}_2$ and $\mathbf{v}_3$ to be a multiple of the original vector found?

$$
\text{So } \left\{ \tfrac{1}{\sqrt{6}}\begin{pmatrix}1\\2\\1\end{pmatrix}, \tfrac{1}{\sqrt{3}}\begin{pmatrix}1\\-1\\1\end{pmatrix}, \tfrac{1}{\sqrt{2}}\begin{pmatrix}-1\\0\\1\end{pmatrix} \right\} \text{ is an orthonormal basis for } \mathbb{R}^3.
$$

5.4 Least Square Approximation

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Recall that $\mathbf{b} \in \text{Col}(\mathbf{A})$ if and only if $\mathbf{Ax} = \mathbf{b}$ is consistent.

More often, in real life application, the data collected or the mathematical model used might not be very accurate, and thus there might not be a solution to the system $\mathbf{Ax} = \mathbf{b}$. However, we might want to find an approximation (vis-à-vis dismissing the model or data) $\mathbf{b}'$ such that $\mathbf{Ax} = \mathbf{b}'$ is consistent, and $\mathbf{b}'$ is the "closest' to $\mathbf{b}$. In other words, we want a $\mathbf{b}' \in \text{Col}(\mathbf{A})$ in the column space of $\mathbf{A}$ that has the shortest distance to $\mathbf{b}$,

$$\|\mathbf{b}' - \mathbf{b}\| = d(\mathbf{b}', \mathbf{b}) \leq d(\mathbf{w}, \mathbf{b}) = \|\mathbf{w} - \mathbf{b}\|$$

for any $\mathbf{w} \in \text{Col}(\mathbf{A})$.

Since $\mathbf{Ax} = \mathbf{b}'$ is consistent, we can find a $\mathbf{u} \in \mathbb{R}^n$ such that $\mathbf{Au} = \mathbf{b}'$. Recall also that $\text{Col}(\mathbf{A}) = \{\ \mathbf{Av} \mid \mathbf{v} \in \mathbb{R}^n\ \}$. So, $\mathbf{Au} = \mathbf{b}'$ being the closest in the column space $\text{Col}(\mathbf{A})$ to $\mathbf{b}$ means

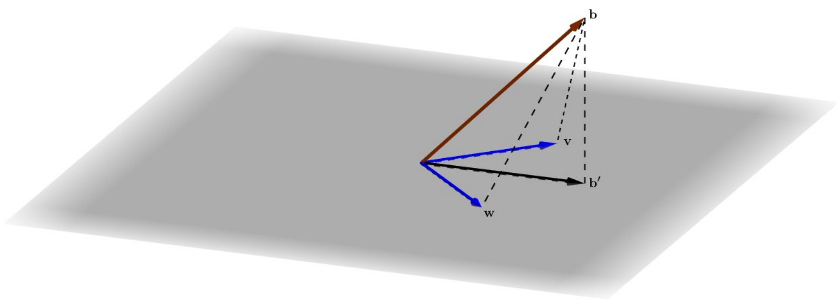
$$\|\mathbf{Au} - \mathbf{b}\| \leq \|\mathbf{Av} - \mathbf{b}\|$$

for any $v \in \mathbb{R}^n$. Then $\mathbf{u}$ is the best approximation to the system $\mathbf{Ax} = \mathbf{b}$. We will now give the formal definition.

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. A vector $\mathbf{u} \in \mathbb{R}^n$ is a least square solution to $\mathbf{Ax} = \mathbf{b}$ if for every vector $\mathbf{v} \in \mathbb{R}^n$,

$$\|\mathbf{Au} - \mathbf{b}\| \leq \|\mathbf{Av} - \mathbf{b}\|.$$

Geometrically, the vector $\mathbf{b}'$ in a subspace V closest to a vector $\mathbf{b}$ is the projection of $\mathbf{b}$ onto the subspace V.



Theorem. *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix and* $\mathbf{b} \in \mathbb{R}^m$. $\mathbf{u} \in \mathbb{R}^n$ *is a least square solution to* $\mathbf{Ax} = \mathbf{b}$ *if and only if* $\mathbf{Au}$ *is the projection of* $\mathbf{b}$ *onto the column space of* $\mathrm{Col}(\mathbf{A})$.

The previous theorem gives us an algorithm to find least square solutions. For $\mathbf{u}$ is a least square solution to $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{Au}$ is the projection of $\mathbf{b}$ onto $\mathrm{Col}(\mathbf{A})$, which is equivalent to $(\mathbf{Au} - \mathbf{b}) \perp \mathrm{Col}(\mathbf{A})$. Recall from lecture 9 that $\mathbf{w} \perp \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ if and only if $\mathbf{w} \in \mathrm{Null}(\mathbf{A}^T)$, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$. Hence, $\mathbf{u}$ is a least square solution of $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{A}^T(\mathbf{Au} - \mathbf{b}) = \mathbf{0}$, or $\mathbf{A}^T\mathbf{Au} = \mathbf{A}^T\mathbf{b}$. Thus, we arrive at the following theorem.

Theorem. *Let* $\mathbf{A}$ *be a* $m \times n$ *matrix and* $\mathbf{b} \in \mathbb{R}^m$. *A vector* $\mathbf{u} \in \mathbb{R}^n$ *is a least square solution to* $\mathbf{Ax} = \mathbf{b}$ *if and only if* $\mathbf{u}$ *is a solution to* $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$.

Remark.

1. Note that even though projection is unique, least square solution may not be unique.
2. The previous theorem seems to suggest that $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ is always consistent. Why is that so?

Example. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Then

$$\left(\begin{array}{cccc|c} 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right).$$

This shows that $\mathbf{b} \notin \mathrm{Col}(\mathbf{A})$. We will find a least square solution of $\mathbf{Ax} = \mathbf{b}$.

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 & 2 \\ 3 & 6 & 3 & 3 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 1 & 2 \end{pmatrix}, \quad \mathbf{A}^T\mathbf{b} = \begin{pmatrix} 2 \\ 4 \\ 2 \\ 2 \end{pmatrix}$$

$$\left(\begin{array}{cccc|c} 2 & 3 & 1 & 2 & 2 \\ 3 & 6 & 3 & 3 & 4 \\ 1 & 3 & 2 & 1 & 2 \\ 2 & 3 & 1 & 2 & 2 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 2/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

A general solution is $\begin{pmatrix} 0 \\ 2/3 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$, $s, t \in \mathbb{R}$. We may choose $s = 0 = t$, then $\mathbf{u} = \begin{pmatrix} 0 \\ 2/3 \\ 0 \\ 0 \end{pmatrix}$ is a least square solution. Then the projection of $\mathbf{b}$ onto the column space is $\mathbf{Au} = \frac{2}{3}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

Notice that $\begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ are in the nullspace of $\mathbf{A}$, and hence for any choice of s and t, the projection $\mathbf{Au}$ is unique.

The theorem above also gives us an alternative way to find orthgonal projection without having to (use Gram-Schmidt Process to) construct an orthonormal basis.

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a subspace and $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for V. Then the orthogonal projection of a vector $\mathbf{w} \in \mathbb{R}^n$ onto V is $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w}$, where $\mathbf{A} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{pmatrix}$.*

Proof. Since S is a basis for V, the columns of $\mathbf{A}$ are linearly independent, and thus $\mathbf{A}^T\mathbf{A}$ is invertible. Hence,

$$\mathbf{u} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w}$$

is the unique solution to $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{w}$. And hence, $\mathbf{Au} = \mathbf{A}((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w})$ is the projection of $\mathbf{w}$ onto $\mathrm{Col}(\mathbf{A}) = V$. $\square$

Example. Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \right\}$ and $V = \mathrm{span}(S)$. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 0 \end{pmatrix}$. Then the orthogonal projection of $\mathbf{w} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ onto V is

$$\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{w} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}^{-1} \begin{pmatrix} x+y \\ x-y \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x+y \\ x-y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

Indeed, since V is the xy-plane in $\mathbb{R}^3$.

6 Diagonalization

6.1 Eigenvalues and Eigenvectors

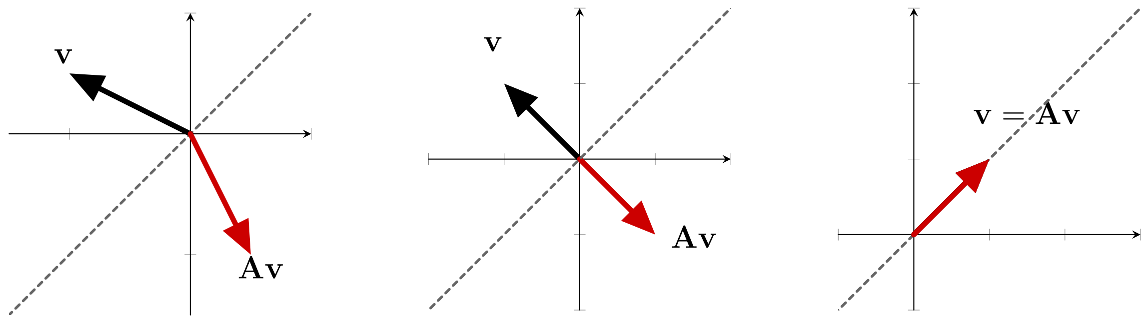
Let $\mathbf{A}$ be a square matrix of order n. Then notice that for any vector $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{Au}$ is also a vector in $\mathbb{R}^n$. So we may think of $\mathbf{A}$ as a map $\mathbb{R}^n \to \mathbb{R}^n$, taking a vector and transforming it to another vector in the same Euclidean space.

Example. 1.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Geometrically the matrix $\mathbf{A}$ reflects a vector along the line $x = y$.

$$\mathbf{A}\begin{pmatrix}-2\\1\end{pmatrix} = \begin{pmatrix}1\\-2\end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix}-1\\1\end{pmatrix} = \begin{pmatrix}1\\-1\end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}.$$

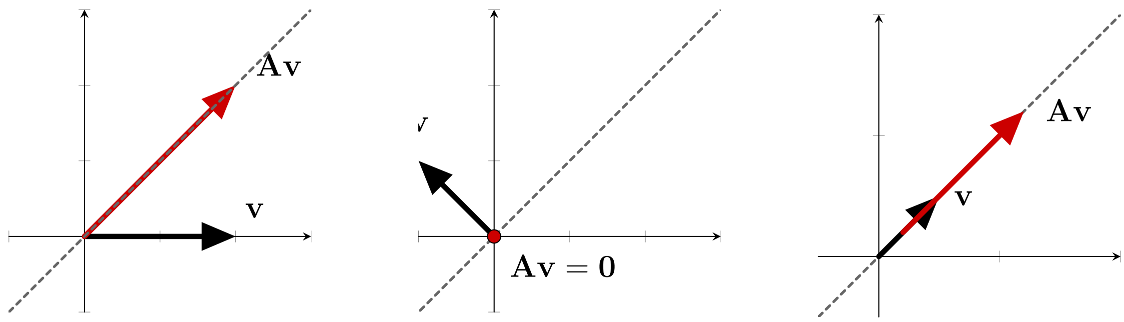


Observe that any vector on the line $x = y$ gets transform back to itself, and any vector along $x = -y$ line get transformed to the negative of itself.

2.
$$\mathbf{A} = \begin{pmatrix}1 & 1\\1 & 1\end{pmatrix}$$

The matrix $\mathbf{A}$ takes a vector and maps it to a vector along the line $x = y$ such that both coordinates in $\mathbf{Av}$ are the sum of the coordinates in $\mathbf{v}$.

$$\mathbf{A}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix}-1\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}2\\2\end{pmatrix}.$$



Observe that any vector $\mathbf{v}$ along the line $x = y$ is mapped to twice itself, $\mathbf{Av} = 2\mathbf{v}$, and it take any vector $\mathbf{v}$ along the line $x = -y$ to the origin, $\mathbf{Av} = \mathbf{0}$.

Let $\mathbf{A}$ be a square matrix of order n. A real number $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbf{A}$ if there is a nonzero vector $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$, such that $\mathbf{Av} = \lambda\mathbf{v}$. In this case, the nonzero vector $\mathbf{v}$ is called an eigenvector associated to λ. In other words, $\mathbf{A}$ transforms its eigenvectors by scaling it by a factor of the associated eigenvalue.

Remark. Note that for $\mathbf{v}$ to be an eigenvector associate to λ, necessarily $\mathbf{v} \neq \mathbf{0}$. Otherwise, $\mathbf{A0} = \lambda\mathbf{0}$ for any $\lambda \in \mathbb{R}$ and thus every number is an eigenvalue, which makes the definition pointless.

Example. 1. For $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Eigenvalue	Eigenvector
$\lambda = 1$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$\lambda = -1$	$\mathbf{v}_\lambda \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

2. For $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Eigenvalue	Eigenvector
$\lambda = 2$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$\lambda = 0$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Observe that if $\mathbf{v}$ is an eigenvector of $\mathbf{A}$ associated to eigenvalue λ, then for any $s \in \mathbb{R}$,

$$\mathbf{A}(s\mathbf{v}) = s(\mathbf{A}\mathbf{v}) = s(\lambda\mathbf{v}) = \lambda(s\mathbf{v}).$$

So for any $s \neq 0$, $s\mathbf{v}$ is also an eigenvector associated to eigenvalue λ. Next, $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ if and only if $\mathbf{0} = \lambda\mathbf{v} - \mathbf{A}\mathbf{v} = (\lambda\mathbf{I} - \mathbf{A})\mathbf{v}$. This means that λ is an eigenvalue of $\mathbf{A}$ if and only if the homogeneous system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

Theorem. *Let $\mathbf{A}$ be a square matrix of order n. $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbf{A}$ if and only if the homogeneous system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ has nontrivial solutions.*

Recall that the homogeneous system associated to a square matrix has nontrivial solutions if and only if the matrix is not invertible, which is equivalent to the determinant of the matrix begin 0. Hence, by the theorem above, λ is an eigenvalue of $\mathbf{A}$ if and only if $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$.

Lemma. *Let $\mathbf{A}$ be a square matrix of order n. Then $\det(x\mathbf{I} - \mathbf{A})$ is a polynomial of degree n.*

Example. Let $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an order 2 square matrix. Then the characteristic polynomial of $\mathbf{A}$ is

$$\begin{vmatrix} x-a & -b \\ -c & x-d \end{vmatrix} = (x-a)(x-d) - bc = x^2 - (a+d)x + ad - bc.$$

It is a degree 2 polynomial.

This means that λ is an eigenvalue of $\mathbf{A}$ if and only if it is a root of the polynomial $\det(x\mathbf{I} - \mathbf{A})$. This motivates the following definition.

Let $\mathbf{A}$ be a square matrix of order n, the characteristic polynomial of $\mathbf{A}$, denoted as $\text{char}(\mathbf{A})$, is the degree n polynomial

$$\text{char}(\mathbf{A}) = \det(x\mathbf{I} - \mathbf{A}).$$

So, we have the following theorem.

Theorem. *Let* $\mathbf{A}$ *be a square matrix of order* n. $\lambda \in \mathbb{R}$ *is an eigenvalue of* $\mathbf{A}$ *if and only if* λ *is a root of the characteristic polynomial* $\det(x\mathbf{I} - \mathbf{A})$.

Example.

1. $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x & -1 \\ -1 & x \end{vmatrix} = x^2 - 1.$$

So the eigenvalues of $\mathbf{A}$ are $\lambda = \pm 1$.

2. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$,

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & -1 \\ -1 & x-1 \end{vmatrix} = (x-1)^2 - 1 = x(x-2).$$

So the eigenvalues of $\mathbf{A}$ are $\lambda = 0$ and $\lambda = 2$.

3. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 1 \end{pmatrix}$,

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & 0 & 0 \\ 0 & x & 2 \\ 0 & -3 & x-1 \end{vmatrix} = (x-1)[x(x-1)-6)] = (x-1)(x+2)(x-3).$$

So the eigenvalues of $\mathbf{A}$ are $\lambda = 1$, $\lambda = -2$, and $\lambda = 3$.

Theorem. *A square matrix* $\mathbf{A}$ *is invertible if and only if* 0 *is not an eigenvalue of* $\mathbf{A}$.

Proof. 0 is an eigenvalue of $\mathbf{A} \Leftrightarrow 0$ is a root of the polynomial $\det(x\mathbf{I} - \mathbf{A}) \Leftrightarrow 0 = \det(0\mathbf{I} - \mathbf{A}) = \det(\mathbf{A}) \Leftrightarrow \mathbf{A}$ not invertible. □

We will add this to our list of equivalent statements for invertibility.

Theorem. *Let* $\mathbf{A}$ *be a square matrix of order* n. *The following statements are equivalent.*

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}$ *has a left inverse.*

(iii) $\mathbf{A}$ *has a right inverse.*

(iv) The reduced row-echelon form of $\mathbf{A}$ *is the identity matrix* $\mathbf{I}_n$.

(v) $\mathbf{A}$ *is a product of elementary matrices.*

(vi) The homogeneous linear system $\mathbf{A}\mathbf{x} = \mathbf{0}$ *has only the trivial solution, that is,* $\mathrm{Null}(\mathbf{A}) = \{\mathbf{0}\}$.

(vii) The linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ *has a unique solution for every* $\mathbf{b}$.

(viii) The determinant of $\mathbf{A}$ *is nonzero,* $\det(\mathbf{A}) \neq 0$.

(ix) The columns/rows of $\mathbf{A}$ *are linearly independent.*

(x) The columns/rows of $\mathbf{A}$ *spans* $\mathbb{R}^n$, $\text{Col}(\mathbf{A}) = \mathbb{R}^n/\text{Row}(\mathbf{A}) = \mathbb{R}^n$.

(xi) 0 *is not an eigenvalue of* $\mathbf{A}$.

Recall that the determinant of a triangular matrix is the product of the diagonal entries. Suppose $\mathbf{A}$ is a triangular matrix. Then $x\mathbf{I} - \mathbf{A}$ is also a triangular matrix. Hence, we have the following statement.

Lemma. *The eigenvaules of a triangular matrix are the diagonal entries.*

Proof. $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$

$$\Rightarrow \begin{vmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ 0 & x - a_{22} & \cdots & -a_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & x - a_{nn} \end{vmatrix} = (x - a_{11})(x - a_{22}) \cdots (x - a_{nn}).$$

□

Let λ be an eigenvalue of $\mathbf{A}$. The algebraic multiplicity of λ is the largest integer r_λ such that

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda)^{r_\lambda} p(x),$$

for some polynomial $p(x)$. Alternatively, r_λ is the positive integer such that in the above equation, λ is not a root of $p(x)$.

Suppose $\mathbf{A}$ is an order n square matrix such that $\det(x\mathbf{I} - \mathbf{A})$ can be factorize into linear factors completely. Then we can write

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k}$$

where $r_1 + r_2 + \cdots + r_k = n$, and $\lambda_, \lambda_2, ..., \lambda_k$ are the distinct eigenvalues of $\mathbf{A}$. Then the algebraic multiplicity of λ_i is r_i for $i = 1, ..., k$.

Example.

1. $\mathbf{A} = \mathbf{0}_n$. Then $\det(x\mathbf{I} - \mathbf{0}) = \det(x\mathbf{I}) = x^n$. The multiplicity of the eigenvalue 0 is n, $r_0 = n$.

2. $\mathbf{A} = \begin{pmatrix} 1 & 2 & 5 \\ 0 & 1 & -2 \\ 0 & 0 & 3 \end{pmatrix}$. Then $\det(x\mathbf{I} - \mathbf{A}) = (x - 1)^2(x - 3)$. The eigenvalue 1 has multiplicity 2, while the eigenvalue 3 has multiplicity 1, $r_1 = 2$, $r_3 = 1$. Observe that in general, the algebraic multiplicity of an eigenvalue λ of an triangular matrix $\mathbf{A}$ is the number of diagonal entries that takes the value λ.

3. $\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix}$. Then $\det(x\mathbf{I} - \mathbf{A}) = (x - 2)^2(x - 4)$. The eigenvalue 2 has multiplicity 2, while the eigenvalue 4 has multiplicity 1, $r_2 = 2$, $r_4 = 1$.

4. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$. Then $\det(x\mathbf{I} - \mathbf{A}) = (x-1)(x^2+1)$. The only real eigenvalue of $\mathbf{A}$ is 1 with multiplicity 1, $r_1 = 1$. The other eigenvalues are complex numbers (lecture 12).

Now suppose $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of $\mathbf{A}$ associated to the eigenvalue λ. Then for any real numbers $\alpha, \beta \in \mathbb{R}$,

$$\mathbf{A}(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha(\mathbf{A}\mathbf{u}) + \beta(\mathbf{A}\mathbf{v}) = \alpha(\lambda\mathbf{u}) + \beta(\lambda\mathbf{v}) = \lambda(\alpha\mathbf{u} + \beta\mathbf{v}).$$

This means that the set of all vectors that satisfies the relation $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ is closed under linearly combination. In particular, if $\alpha\mathbf{u} + \beta\mathbf{v} \neq \mathbf{0}$, then it is also an eigenvector associated to λ. This shows that the collection of all eigenvectors associated to an eigenvalue λ, together with the zero vector, form a subspace. This motivates the following definition.

The eigenspace associated to an eigenvalue λ of $\mathbf{A}$ is

$$E_\lambda = \{\ \mathbf{v} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\ \} = \mathrm{Null}(\lambda\mathbf{I} - \mathbf{A}).$$

The geometric multiplicity of an eigenvalue λ is the dimension of its associated eigenspace,

$$\dim(E_\lambda) = \mathrm{nullity}(\lambda\mathbf{I} - \mathbf{A}).$$

Example. $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$.

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & -1 & 0 \\ -1 & x-1 & -1 \\ 0 & 0 & x-2 \end{vmatrix} = (x-2)((x-1)^2 - 1) = x(x-2)^2.$$

The eigenvalues of $\mathbf{A}$ are 0 and 2, with algebraic multiplicities $r_0 = 1$ and $r_2 = 2$. Since 0 is an eigenvalue of $\mathbf{A}$, $\mathbf{A}$ is not invertible.

Now find the eigenspace.

$\lambda = 0$:

$$0\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0-1 & -1 & 0 \\ -1 & 0-1 & -1 \\ 0 & 0 & 0-2 \end{pmatrix} = \begin{pmatrix} -1 & -1 & 0 \\ -1 & -1 & -1 \\ 0 & 0 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

So $E_0 = \mathrm{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$.

$\lambda = 2$:

$$2\mathbf{I} - \mathbf{A} = \begin{pmatrix} 2-1 & -1 & 0 \\ -1 & 2-1 & 0 \\ 0 & 0 & 2-2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So $E_2 = \mathrm{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$.

So the geometric multiplicity of $\lambda = 0$ is $\dim(E_0) = 1$ and the geometric multiplicity of $\lambda = 2$ is $\dim(E_2) = 2$.

Appendix to Lecture 10

Theorem. *Let $V \subseteq \mathbb{R}^n$ be a subspace and $\mathbf{b} \in \mathbb{R}^n$. Let $\mathbf{b}_p \in V$ be the orthogonal projection of $\mathbf{b}$ onto V. Then for any $\mathbf{w} \in V$,*

$$\|\mathbf{b}_p - \mathbf{b}\| = d(\mathbf{b}_p, \mathbf{b}) \leq d(\mathbf{w}, \mathbf{b}) = \|\mathbf{w} - \mathbf{b}\|,$$

that is, the distance between $\mathbf{b}$ and the projection $\mathbf{b}_p$ is the minimum distance between $\mathbf{b}$ and any vector $\mathbf{w} \in V$.

Proof. Write $\mathbf{b} = \mathbf{b}_p + \mathbf{b}_n$, where $\mathbf{b}_p \in \mathrm{Col}(\mathbf{A})$ is the projection of $\mathbf{b}$ onto the column space of $\mathbf{A}$, and $\mathbf{b}_n$ is orthogonal to $\mathrm{Col}(\mathbf{A})$. Let $\mathbf{w}$ be any vector in V. Since $\mathbf{w}$ and $\mathbf{b}_p$ are in the subspace V and $\mathbf{b}_n \perp V$, $(\mathbf{w} - \mathbf{b}_p) \cdot \mathbf{b}_n = 0$. So

$$\begin{aligned} \|\mathbf{w} - \mathbf{b}\|^2 &= (\mathbf{w} - \mathbf{b}_p - \mathbf{b}_n) \cdot (\mathbf{w} - \mathbf{b}_p - \mathbf{b}_n) \\ &= (\mathbf{w} - \mathbf{b}_p) \cdot (\mathbf{w} - \mathbf{b}_p) - 2(\mathbf{w} - \mathbf{b}_p).\mathbf{b}_n + \mathbf{b}_n \cdot \mathbf{b}_n \\ &= \|\mathbf{w} - \mathbf{b}_p\|^2 + \|\mathbf{b}_n\|^2. \end{aligned} \tag{5}$$

So $\|\mathbf{w} - \mathbf{b}\| \geq \|\mathbf{b}_n\| = \|\mathbf{b}_p - \mathbf{b}\|$. □

Theorem. *Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. $\mathbf{u} \in \mathbb{R}^n$ is a least square solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{A}\mathbf{u}$ is the projection of $\mathbf{b}$ onto the column space of $\mathrm{Col}(\mathbf{A})$.*

Proof. ($\Rightarrow$) Suppose $\mathbf{u}$ is a least square solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. Since the projection $\mathbf{b}_p \in \mathrm{Col}(\mathbf{A})$ is in the column space, there must be a $\mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{v} = \mathbf{b}_p$. Then by equation (5),

$$\|\mathbf{A}\mathbf{u} - \mathbf{b}_p\|^2 + \|\mathbf{b}_n\|^2 = \|\mathbf{A}\mathbf{u} - \mathbf{b}\|^2 \leq \|\mathbf{A}\mathbf{v} - \mathbf{b}\|^2 = \|\mathbf{b}_p - \mathbf{b}\|^2 = \|\mathbf{b}_n\|^2,$$

which happens if and only if $\mathbf{A}\mathbf{u} = \mathbf{b}_p$, that is, $\mathbf{A}\mathbf{u}$ is the projection of $\mathbf{b}$ onto $\mathrm{Col}(\mathbf{A})$.

($\Leftarrow$) Suppose $\mathbf{A}\mathbf{u} = \mathbf{b}_p$ is the projection of $\mathbf{b}$ onto $\mathrm{Col}(\mathbf{A})$. Then by the theorem above, and the fact that $\mathrm{Col}(\mathbf{A}) = \{\ \mathbf{A}\mathbf{v} \mid \mathbf{v} \in \mathbb{R}^n\ \}$,

$$\|\mathbf{A}\mathbf{u} - \mathbf{b}\| = \|\mathbf{b}_p - \mathbf{b}\| \leq \|\mathbf{A}\mathbf{v} - \mathbf{b}\|$$

for any $\mathbf{v} \in \mathbb{R}^n$, that is, $\mathbf{u}$ is a least square solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

□