Orthogonal Projection Perspective

- Back to the case of invertible $C = A^T A$, consider the orthogonal projection of $\mathbf{b}$ onto Col A as
$$\hat{\mathbf{b}} = f(\mathbf{b}) = A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b} = C\mathbf{b}$$
where $C = A(A^T A)^{-1} A^T$.

- One can see that the orthogonal projection is actually a **linear transformation** $f(\mathbf{b}) = C\mathbf{b}$ where the standard matrix is defined as $C = A(A^T A)^{-1} A^T$.

- What if A has orthonormal columns? (More in the next slides.)

Orthogonal and Orthonormal Sets

- **Definition**: A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is an **orthogonal set** if each pair of distinct vectors from the set is orthogonal That is, if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

- **Definition**: A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is an **orthonorm al set** if it is an orthogonal set of unit vectors.

- Is an orthogonal (or orthonormal) set also a linearly independent set? What about its converse?

Orthogonal and Orthonormal Basis

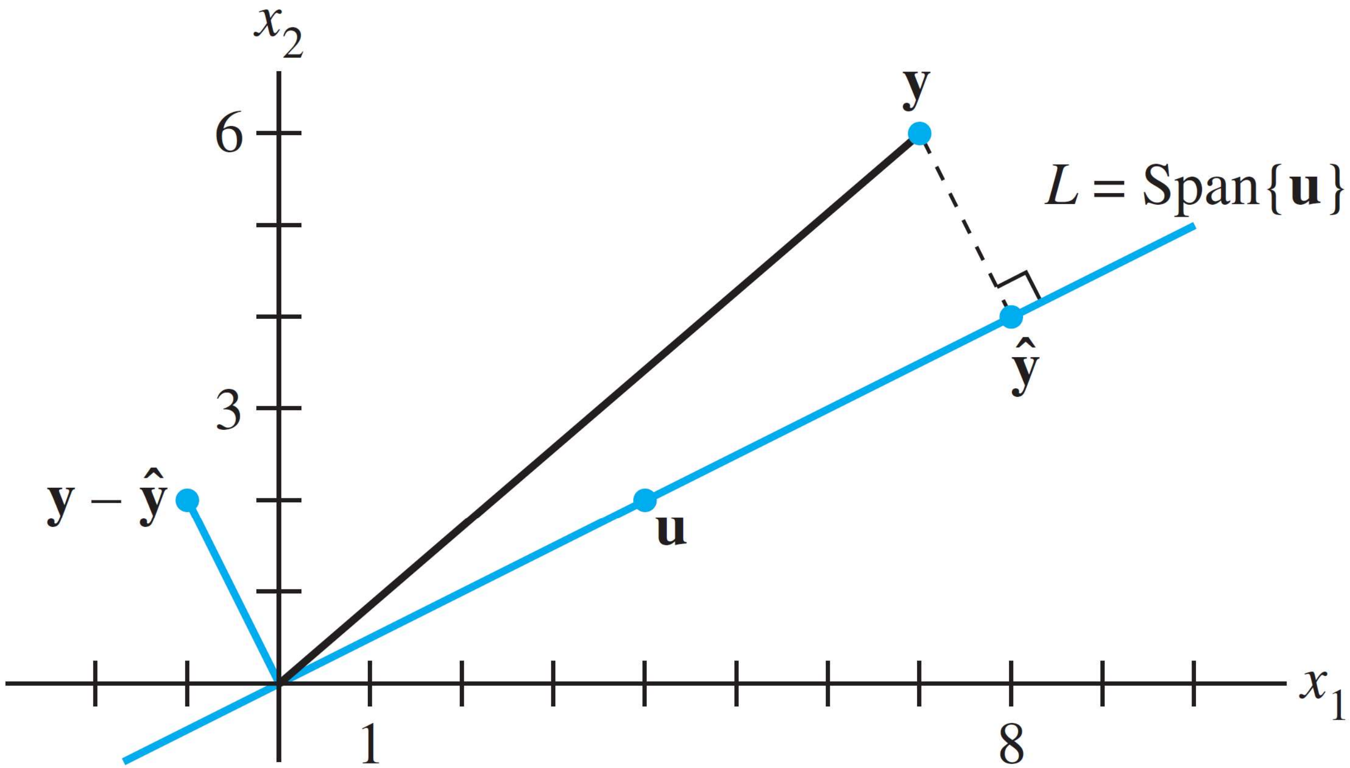
- Consider basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of a p-dimensional subspace W in $\mathbb{R}^n$.
- Can we make it as an orthogonal (or orthonormal) basis?
  - Yes, it can be done by Gram–Schmidt process. → QR factorization.
- Given the orthogonal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ of W,
  let's compute the orthogonal projection of $\mathbf{y} \in \mathbb{R}^n$ onto W.

Orthogonal Projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto Line

- Consider the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto one-dimensional subspace

- $\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}$

- If $\mathbf{u}$ is a unit vector, $\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = (\mathbf{y} \cdot \mathbf{u})\mathbf{u}$



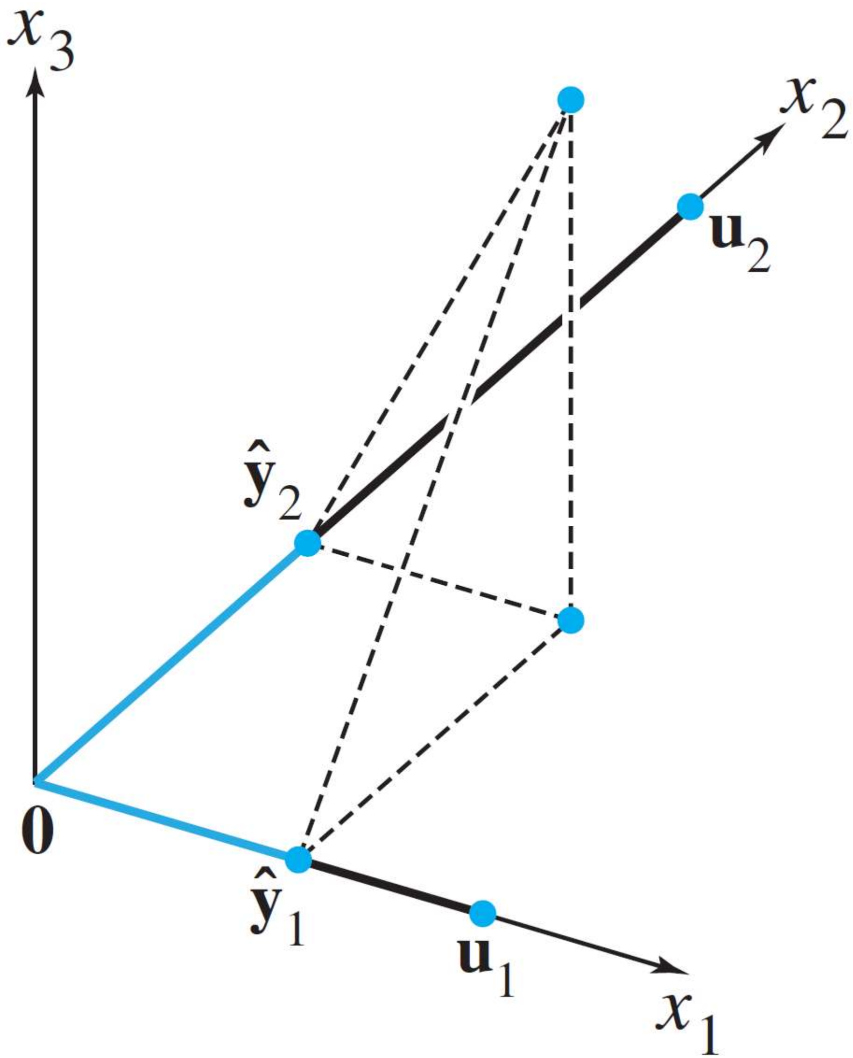
Orthogonal Projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto Plane

- Consider the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto two-dimensional subspace W

- $\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2$

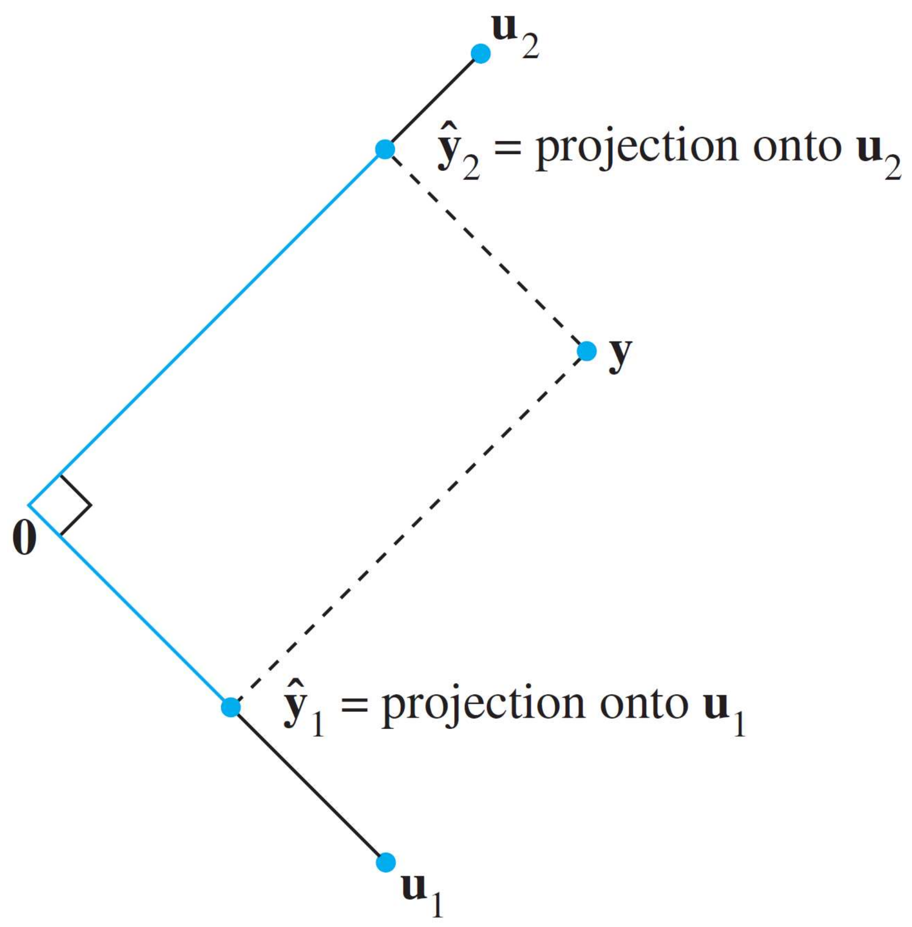
- If $\mathbf{u}_1$ and $\mathbf{u}_2$ are unit vectors,
  $\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2$

- Projection is done independently on each orthogonal basis vector.



Orthogonal Projection when $\mathbf{y} \in W$

- Consider the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto two-dimensional subspace W, where $\mathbf{y} \in W$
- $\hat{\mathbf{y}} = \mathrm{proj}_L \mathbf{y} = \mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2$
- If $\mathbf{u}_1$ and $\mathbf{u}_2$ are unit vectors, $\hat{\mathbf{y}} = \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2$
- The solution is the same as before. Why?



Transformation: Orthogonal Projection

- Consider a transformation of orthogonal projection $\hat{\mathbf{b}}$ of $\mathbf{b}$, given orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ of a subspace W:

$$\begin{aligned}
\hat{\mathbf{b}} = f(\mathbf{b}) &= (\mathbf{b} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{b} \cdot \mathbf{u}_2)\mathbf{u}_2 \\
&= (\mathbf{u}_1^T\mathbf{b})\mathbf{u}_1 + (\mathbf{u}_2^T\mathbf{b})\mathbf{u}_2 \\
&= \mathbf{u}_1(\mathbf{u}_1^T\mathbf{b}) + \mathbf{u}_2(\mathbf{u}_2^T\mathbf{b}) \\
&= (\mathbf{u}_1\mathbf{u}_1^T)\mathbf{b} + (\mathbf{u}_2\mathbf{u}_2^T)\mathbf{b} \\
&= (\mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T)\mathbf{b} \\
&= [\mathbf{u}_1 \quad \mathbf{u}_2]\begin{bmatrix}\mathbf{u}_1^T \\ \mathbf{u}_2^T\end{bmatrix}\mathbf{b} = UU^T\mathbf{b} = C\mathbf{b} \Rightarrow \text{linear transformation!}
\end{aligned}$$

Orthogonal Projection Perspective

- Let's verify the following, when $A = U = [\mathbf{u}_1 \quad \mathbf{u}_2]$ has orthonormal columns:
  Back to the case of invertible $C = A^T A$, consider the orthogon al projection of $\mathbf{b}$ onto Col A as
$$\hat{\mathbf{b}} = A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b} = f(\mathbf{b})$$

- $C = A^T A = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{bmatrix} [\mathbf{u}_1 \quad \mathbf{u}_2] = I$. Thus,
$$\hat{\mathbf{b}} = A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \mathbf{b} = A(I)^{-1} A^T \mathbf{b} = AA^T \mathbf{b} = UU^T \mathbf{b}$$

Further Study

- Least-squares derivation from maximum likelihood perspectiv
  e
  (via Gaussian distribution)
  - Kevin Murphy, “Machine Learning: A Probabilistic Perspective,” Ch7.2

- Orthogonal projection and QR decomposition
  - Lay Ch6.2, Ch.6.3, Ch6.4

Gram-Schmidt Orthogonalization

- **Example 1**: Let $W = \mathrm{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.
  Construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

- **Solution:** Let $\mathbf{v}_1 = \mathbf{x}_1$. Next, Let $\mathbf{v}_2$ the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$, i.e.,

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

- The set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for W.

Gram-Schmidt Orthogonalization

- **Example 2**: Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Then

$\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is clearly linearly independent and thus a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

Gram-Schmidt Orthogonalization

- **Solution:**
- **Step 1.** Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \mathrm{Span}\{\mathbf{x}_1\} = \mathrm{Span}\{\mathbf{v}_1\}$.
- **Step 2.** Let $\mathbf{v}_2$ be the vector produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is, let

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathrm{proj}_{W_1}\mathbf{x}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

- $\mathbf{v}_2$ is the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$, and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace W_2 spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$.

Gram-Schmidt Orthogonalization

- **Step 2′ (optional).** If appropriate, scale $\mathbf{v}_2$ to simplify later computations, e.g.,

$$\mathbf{v}_2 = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \longrightarrow \mathbf{v}_2' = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Gram-Schmidt Orthogonalization

- **Step 3.** Let $\mathbf{v}_3$ be the vector produced by subtracting from $\mathbf{x}_3$ its projection onto the subspace W_2. Use the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2'\}$ to compute this projection onto W_2:

$$\mathrm{proj}_{W_2}\mathbf{x}_3 = \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_3 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_3 \cdot \mathbf{v}_2'}\mathbf{v}_2' = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

- Then $\mathbf{v}_3$ is the component of $\mathbf{x}_3$ orthogonal to W_2, namely,

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathrm{proj}_{W_2}\mathbf{x}_3 = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

Gram-Schmidt Orthogonalization

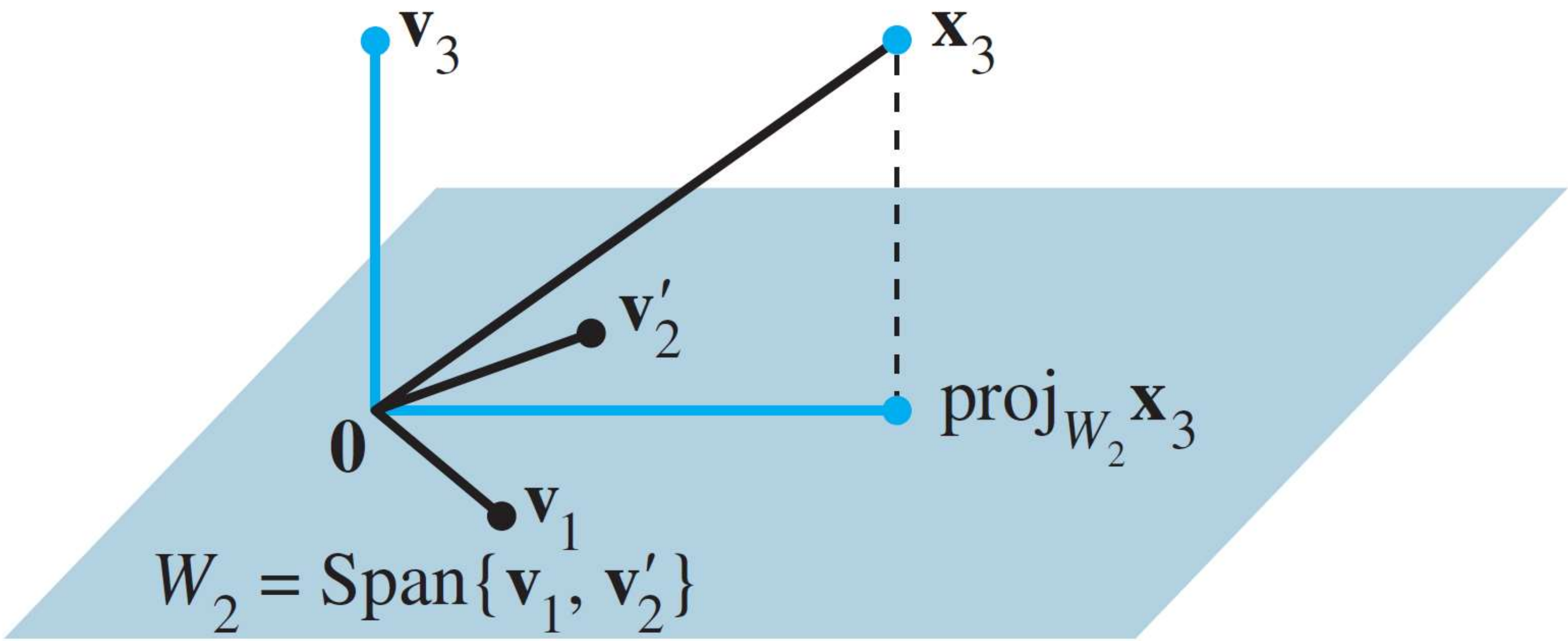


FIGURE 2 The construction of $\mathbf{v}_3$ from $\mathbf{x}_3$ and W_2.

Figure from Lay Ch6.4

QR Factorization

- If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, 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 its diagonal.

Computing QR Factorization

- **Step 1 (Construction of Q):** The columns of A form a basis for $\mathrm{Col}\,A$ since they are linearly independent. Let these columns be $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. Then, we can construct the orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ for $\mathrm{Col}\,A$ by the Gram-Schmidt process described by Theorem 11. Using this basis, we can construct Q as

$$Q = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n]$$

Computing QR Factorization

- **Step 2 (Construction of R):** From (1) in Theorem 11, for $k = 1, \dots, n$, $\mathbf{x}_k$ is in $\mathrm{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} = \mathrm{Span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$. Therefore, there exist constants $r_{1k}, \dots, r_{kk}$ such that

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0 \cdot \mathbf{u}_{k+1} + \cdots + 0 \cdot \mathbf{u}_n$$

- We can always make $r_{kk} \geq 0$ because if $r_{kk} < 0$, then we can multiply both r_{kk} and $\mathbf{u}_k$ by -1. Using this linear combination representation, we can construct $\mathbf{r}_k$, the k-th column of R, as

$$\mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Computing QR Factorization

- That is, $\mathbf{x}_k = Q\mathbf{r}_k$ for $k = 1, \ldots, n$. Let $R = [\mathbf{r}_1 \quad \cdots \quad \mathbf{r}_n]$. Then,
$$A = [\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_n] = [Q\mathbf{r}_1 \quad \cdots \quad Q\mathbf{r}_n] = QR$$

- The fact that R is invertible follows easily from the fact that the columns of A are linearly independent (Exercise 19). Since R is clearly upper triangular (from the previous slide) and invertible, the diagonal entries r_{kk}'s should be nonzero. By combining this with the fact that $r_{kk} \geq 0$, r_{kk}'s must be positive.

Example: QR Factorization

- **Example 4:** Find a QR factorization of $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.
- **Solution:** Let $A = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3]$. We first obtain $\mathbf{v}_1 = \mathbf{x}_1$ and its normalized vector is $\mathbf{u}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$.
- Thus, $\mathbf{x}_1 = 2\mathbf{u}_1$, which gives us $\mathbf{r}_{11}$ =2, i.e., $\mathbf{r}_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$.

Example: QR Factorization

- Next, we obtain $\mathbf{v}_3$ as $\mathbf{v}_3 = \mathbf{x}_3 - \text{proj}_{W_2}\mathbf{x}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - 1\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} - \frac{2}{\sqrt{12}}\begin{bmatrix} -3/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$ and its normalized vector $\mathbf{u}_2$ as $\mathbf{u}_2 = \begin{bmatrix} 0 \\ -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$.
- Thus, $\mathbf{x}_3 = 1\mathbf{u}_1 + \frac{2}{\sqrt{12}}\mathbf{u}_2 + \frac{2}{\sqrt{6}}\mathbf{u}_3$, i.e., $\mathbf{r}_3 = \begin{bmatrix} 1 \\ 2/\sqrt{12} \\ 2/\sqrt{6} \end{bmatrix}$.

Example: QR Factorization

- In conclusion, $Q = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3] = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}$

  and $R = [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3] = \begin{bmatrix} 2 & -3/2 & 1 \\ 0 & -3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{bmatrix}$.