Lecture 21: Orthogonality and Bases

MAT2040 Linear Algebra

Recall coordinates:

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p\}$ be a basis for a subspace W of $\mathbb{R}^n$, and $P = \begin{bmatrix} \mathbf{b}_1, & \mathbf{b}_2, & \dots, & \mathbf{b}_p \end{bmatrix}$.

To compute the $\mathcal{B}$-coordinates of $\mathbf{x} \in W$ we have to solve the system of linear equations

$$P[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}.$$

Recall coordinates:

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ be a basis for a subspace W of $\mathbb{R}^n$, and $P = \begin{bmatrix} \mathbf{b}_1, & \mathbf{b}_2, & \ldots, & \mathbf{b}_p \end{bmatrix}$.

To compute the $\mathcal{B}$-coordinates of $\mathbf{x} \in W$ we have to solve the system of linear equations

$$P[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}.$$

A lot of work in general.

Recall coordinates:

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ be a basis for a subspace W of $\mathbb{R}^n$, and $P = \begin{bmatrix} \mathbf{b}_1, & \mathbf{b}_2, & \ldots, & \mathbf{b}_p \end{bmatrix}$.

To compute the $\mathcal{B}$-coordinates of $\mathbf{x} \in W$ we have to solve the system of linear equations

$$P[\mathbf{x}]_{\mathcal{B}} = \mathbf{x}.$$

A lot of work in general.

What if $\mathbf{b}_i$ is **orthogonal** to $\mathbf{b}_j$ for every $i \neq j$? (The set $\mathcal{B}$ is then called an **orthogonal set**.)

Theorem 21.1 (Coordinates with respect to Orthogonal Basis)

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ *be a basis for a subspace* W *of* $\mathbb{R}^n$*, that is also an orthogonal set. For every* $\mathbf{x} \in W$ *the weights in the linear combination*

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p$$

are given by

$$c_i = \frac{\mathbf{x} \cdot \mathbf{b}_i}{\mathbf{b}_i \cdot \mathbf{b}_i}.$$

Theorem 21.1 (Coordinates with respect to Orthogonal Basis)

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ *be a basis for a subspace* W *of* $\mathbb{R}^n$*, that is also an orthogonal set. For every* $\mathbf{x} \in W$ *the weights in the linear combination*

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p$$

are given by

$$c_i = \frac{\mathbf{x} \cdot \mathbf{b}_i}{\mathbf{b}_i \cdot \mathbf{b}_i}.$$

We will call such $\mathcal{B}$ an **orthogonal basis** for W.

Example 21.2

Let $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} \right\}$ and $\mathbf{x} = \begin{bmatrix} 4 \\ -7 \\ -2 \\ 15 \end{bmatrix}$.

Compute $[\mathbf{x}]_{\mathcal{B}}$.

Theorem 21.3 (Projection on Subspace with Orthogonal Basis)

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Let $\mathbf{x} \in \mathbb{R}^n$, and let $\hat{\mathbf{x}}$ be the orthogonal projection of $\mathbf{x}$ on W.

Then $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p$$

where

$$c_i = \frac{\mathbf{x} \cdot \mathbf{b}_i}{\mathbf{b}_i \cdot \mathbf{b}_i}.$$

Recall:

Theorem 19.23 (Orthogonal Decomposition Theorem)

If $\mathbf{x} \in \mathbb{R}^n$ *and* S *is a subspace of* $\mathbb{R}^n$*, then* $\mathbf{x}$ *can be* **uniquely** *expressed as* $\mathbf{x} = \hat{\mathbf{x}} + \mathbf{z}$*, where* $\hat{\mathbf{x}} \in S$ *and* $\mathbf{z} \in S^{\perp}$*.*

Example 21.4

Let $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ -1 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} \right\}$ and $\mathbf{x} = \begin{bmatrix} 6 \\ 2 \\ 2 \\ 1 \end{bmatrix}$.

Compute the orthogonal projection of $\mathbf{x}$ on $\operatorname{Span} \mathcal{B}$.

This would be even nicer if $\mathbf{b}_i \cdot \mathbf{b}_i = 1$ for every $i = 1, 2, \ldots, p$.

This would be even nicer if $\mathbf{b}_i \cdot \mathbf{b}_i = 1$ for every $i = 1, 2, \ldots, p$.

Easy! If we have an orthogonal set, we can just scale the vectors in the set so that they are unit length vectors!

This would be even nicer if $\mathbf{b}_i \cdot \mathbf{b}_i = 1$ for every $i = 1, 2, \ldots, p$.

Easy! If we have an orthogonal set, we can just scale the vectors in the set so that they are unit length vectors!

Definition 21.5

An orthogonal set $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ where $\|\mathbf{b}_i\| = 1$ or all $i = 1, 2, \ldots, p$ is called an **orthonormal set**.

This would be even nicer if $\mathbf{b}_i \cdot \mathbf{b}_i = 1$ for every $i = 1, 2, \ldots, p$.

Easy! If we have an orthogonal set, we can just scale the vectors in the set so that they are unit length vectors!

Definition 21.5

An orthogonal set $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ where $\|\mathbf{b}_i\| = 1$ or all $i = 1, 2, \ldots, p$ is called an **orthonormal set**.

An orthonormal set $\mathcal{B}$ that is also a basis, will be referred to as an **orthonormal basis**.

Theorem 21.6 (Coordinates with respect to and Projection on Subspace with Orthonormal Basis)

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$. For every $\mathbf{x} \in W$ the weights in the linear combination

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p$$

are given by

$$c_i = \mathbf{x} \cdot \mathbf{b}_i.$$

Theorem 21.6 (Coordinates with respect to and Projection on Subspace with Orthonormal Basis)

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ *be an orthonormal basis for a subspace* W *of* $\mathbb{R}^n$*. For every* $\mathbf{x} \in W$ *the weights in the linear combination*

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p$$

are given by

$$c_i = \mathbf{x} \cdot \mathbf{b}_i.$$

For every $\mathbf{y} \in \mathbb{R}^n$ *the orthogonal projection of* $\mathbf{y}$ *on* W *is given by*

$$\hat{\mathbf{y}} = t_1\mathbf{b}_1 + t_2\mathbf{b}_2 + \cdots + t_p\mathbf{b}_p$$

where

$$t_i = \mathbf{y} \cdot \mathbf{b}_i.$$

So if you have an orthonormal basis for a subspace W, calculating coordinates with respect to this basis are simple inner product calculations.

So if you have an orthonormal basis for a subspace W, calculating coordinates with respect to this basis are simple inner product calculations.

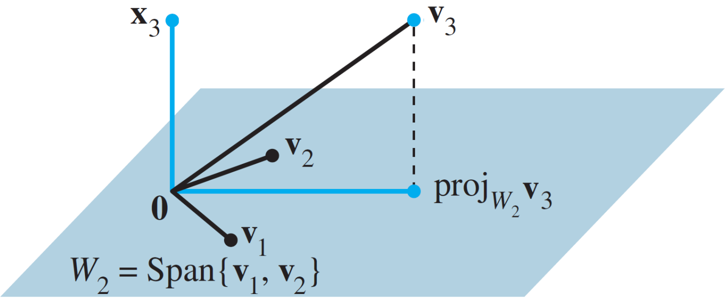
And even finding the orthogonal projection on W can just be done by simple inner product calculations.

So if you have an orthonormal basis for a subspace W, calculating coordinates with respect to this basis are simple inner product calculations.

And even finding the orthogonal projection on W can just be done by simple inner product calculations.

I hear you thinking: "This is very exciting, but do these nice orthonormal bases exist for any subspace? And if they exist, how can we find them?"

Idea of how to change basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ to orthogonal basis:



(picture from David Lay, Linear Algebra.)

$$\mathbf{x}_3 = \mathbf{v}_3 - \text{proj}_{W_2}\mathbf{v}_3$$

$\mathbf{x}_3$ is orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$!

Note that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3\}$ are linearly independent vectors in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and must therefore be a basis for this subspace.

Note that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3\}$ are linearly independent vectors in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and must therefore be a basis for this subspace.

We can therefore use $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3\}$ instead of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ as a basis, and then repeat this process for $\mathbf{v}_2$ (for example).

Note that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3\}$ are linearly independent vectors in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and must therefore be a basis for this subspace.

We can therefore use $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{x}_3\}$ instead of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ as a basis, and then repeat this process for $\mathbf{v}_2$ (for example).

(And making an orthogonal basis orthonormal can be done by just scaling the vectors, to make them have unit length.)

Instead of projecting every basis vector on the subspace spanned by the other vectors, we will actually do something smarter:

Instead of projecting every basis vector on the subspace spanned by the other vectors, we will actually do something smarter:

We will build the orthonormal basis one vector at the time, making sure that our basis is orthonormal at every step.

Instead of projecting every basis vector on the subspace spanned by the other vectors, we will actually do something smarter:

We will build the orthonormal basis one vector at the time, making sure that our basis is orthonormal at every step.

When we add a new vector to the orthonormal basis, we just subtract the projection of the vector on the subspace spanned by the current orthonormal basis to make sure the new (larger) basis is still orthogonal (and we scale the vector to have unit length).

Instead of projecting every basis vector on the subspace spanned by the other vectors, we will actually do something smarter:

We will build the orthonormal basis one vector at the time, making sure that our basis is orthonormal at every step.

When we add a new vector to the orthonormal basis, we just subtract the projection of the vector on the subspace spanned by the current orthonormal basis to make sure the new (larger) basis is still orthogonal (and we scale the vector to have unit length).

Calculations are easy, only requiring the calculation of inner products — see Theorem 21.6!

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W.

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W.

Step 1: $\mathbf{u}_1 = \dfrac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1$.

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W.

Step 1: $\mathbf{u}_1 = \dfrac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1$.

Note that $\{\mathbf{u}_1\}$ is an orthonormal basis for $\text{Span}\{\mathbf{x}_1\}$.

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W.

Step 1: $\mathbf{u}_1 = \dfrac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1$.

Step 2: $\mathbf{u}_2 = \dfrac{1}{\|\mathbf{y}_2\|}\mathbf{y}_2$ where $\mathbf{y}_2 = \mathbf{x}_2 - (\mathbf{x}_2 \cdot \mathbf{u}_1)\mathbf{u}_1$.

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W.

Step 1: $\mathbf{u}_1 = \dfrac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1$.

Step 2: $\mathbf{u}_2 = \dfrac{1}{\|\mathbf{y}_2\|}\mathbf{y}_2$ where $\mathbf{y}_2 = \mathbf{x}_2 - (\mathbf{x}_2 \cdot \mathbf{u}_1)\mathbf{u}_1$.

$\hat{\mathbf{x}}_2 = (\mathbf{x}_2 \cdot \mathbf{u}_1)\mathbf{u}_1$ is the orthogonal projection of $\mathbf{x}_2$ on Span$\{\mathbf{u}_1\}$ (Theorem 21.6).

We therefore want to add $\mathbf{y}_2 = \mathbf{x}_2 - \hat{\mathbf{x}}_2$ to the basis — but we have to make it unit length.

Note that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for Span$\{\mathbf{x}_1, \mathbf{x}_2\}$.

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W.

Step 3: $\mathbf{u}_3 = \dfrac{1}{\|\mathbf{y}_3\|}\mathbf{y}_3$ where $\mathbf{y}_3 = \mathbf{x}_3 - (\mathbf{x}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{x}_3 \cdot \mathbf{u}_2)\mathbf{u}_2$.

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W.

Step 3: $\mathbf{u}_3 = \dfrac{1}{\|\mathbf{y}_3\|}\mathbf{y}_3$ where $\mathbf{y}_3 = \mathbf{x}_3 - (\mathbf{x}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{x}_3 \cdot \mathbf{u}_2)\mathbf{u}_2$.

$\hat{\mathbf{x}}_3 = (\mathbf{x}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x}_3 \cdot \mathbf{u}_2)\mathbf{u}_2$ is the orthogonal projection of $\mathbf{x}_3$ on $\mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ (Theorem 21.6).

We therefore want to add $\mathbf{y}_3 = \mathbf{x}_3 - \hat{\mathbf{x}}_3$ to the basis — after scaling. Note that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis.

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W.

Step k: $\mathbf{u}_k = \dfrac{1}{\|\mathbf{y}_k\|}\mathbf{y}_k$ where

$$\mathbf{y}_k = \mathbf{x}_k - (\mathbf{x}_k \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{x}_k \cdot \mathbf{u}_2)\mathbf{u}_2 - \cdots - (\mathbf{x}_k \cdot \mathbf{u}_{k-1})\mathbf{u}_{k-1}.$$

The Gram-Schmidt Process

Input: a basis $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$ for a subspace W.

Output: an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ for subspace W.

Step k: $\mathbf{u}_k = \dfrac{1}{\|\mathbf{y}_k\|}\mathbf{y}_k$ where

$$\mathbf{y}_k = \mathbf{x}_k - (\mathbf{x}_k \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{x}_k \cdot \mathbf{u}_2)\mathbf{u}_2 - \cdots - (\mathbf{x}_k \cdot \mathbf{u}_{k-1})\mathbf{u}_{k-1}.$$

$\hat{\mathbf{x}}_k = (\mathbf{x}_k \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x}_k \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{x}_k \cdot \mathbf{u}_{k-1})\mathbf{u}_{k-1}$ is the orthogonal projection of $\mathbf{x}_3$ on $\mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{k-1}\}$ (Theorem 21.6).

We therefore want to add $\mathbf{y}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$ to the basis — but we have to make it unit length. Note that $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ is an orthonormal basis for $\mathrm{Span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$.

Theorem 21.7

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ be a basis for a subspace W.
The Gram-Schmidt Process produces an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W after p steps. The vector $\mathbf{u}_k$ is computed in step k as:

$$\mathbf{u}_k = \frac{1}{\|\mathbf{y}_k\|}\mathbf{y}_k, \text{ where}$$

$$\mathbf{y}_k = \mathbf{x}_k - (\mathbf{x}_k \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{x}_k \cdot \mathbf{u}_2)\mathbf{u}_2 - \cdots - (\mathbf{x}_k \cdot \mathbf{u}_{k-1})\mathbf{u}_{k-1}.$$

Theorem 21.7

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ be a basis for a subspace W.
The Gram-Schmidt Process produces an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ for subspace W after p steps. The vector $\mathbf{u}_k$ is computed in step k as:

$$\mathbf{u}_k = \frac{1}{\|\mathbf{y}_k\|}\mathbf{y}_k, \text{ where}$$

$$\mathbf{y}_k = \mathbf{x}_k - (\mathbf{x}_k \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{x}_k \cdot \mathbf{u}_2)\mathbf{u}_2 - \cdots - (\mathbf{x}_k \cdot \mathbf{u}_{k-1})\mathbf{u}_{k-1}.$$

($\mathbf{y}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$, where $\hat{\mathbf{x}}_k$ is the orthogonal projection of $\mathbf{x}_k$ on $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}\}$.)

Example 21.8

Find an orthonormal basis for Col A, where

$$A = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}.$$

Let's try to understand the Gram-Schmidt Process a little better when we perform the Gram-Schmidt Process on the columns of a matrix A.

Let's try to understand the Gram-Schmidt Process a little better when we perform the Gram-Schmidt Process on the columns of a matrix A.

It (iteratively) defines

$$\mathbf{u}_k = \lambda_k(\mathbf{a}_k - (\mathbf{a}_k \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{a}_k \cdot \mathbf{u}_2)\mathbf{u}_2 - \cdots - (\mathbf{a}_k \cdot \mathbf{u}_{k-1})\mathbf{u}_{k-1}),$$

where λ_k is some (positive) real number (making sure $\|\mathbf{u}_k\| = 1$).

Let's try to understand the Gram-Schmidt Process a little better when we perform the Gram-Schmidt Process on the columns of a matrix A.

It (iteratively) defines

$$\mathbf{u}_k = \lambda_k(\mathbf{a}_k - (\mathbf{a}_k \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{a}_k \cdot \mathbf{u}_2)\mathbf{u}_2 - \cdots - (\mathbf{a}_k \cdot \mathbf{u}_{k-1})\mathbf{u}_{k-1}),$$

where λ_k is some (positive) real number (making sure $\|\mathbf{u}_k\| = 1$).

Note that we can also solve this equation for $\mathbf{a}_k$, and we see that we can interpret this equation as "$\mathbf{a}_k$ can be written as a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$".

If we let Q be the matrix with columns equal to the vectors in the orthonormal basis, then the Gram-Schmidt Process tell us that column i of A can be written as a linear combination of the first i columns of Q!

If we let Q be the matrix with columns equal to the vectors in the orthonormal basis, then the Gram-Schmidt Process tell us that column i of A can be written as a linear combination of the first i columns of Q!

I.e., for every $i = 1, 2, \ldots, n$ there exist r_{ij} for $j = 1, 2, \ldots, i$ so that

$$\mathbf{a}_i = r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i.$$

If we let Q be the matrix with columns equal to the vectors in the orthonormal basis, then the Gram-Schmidt Process tell us that column i of A can be written as a linear combination of the first i columns of Q!

I.e., for every $i = 1, 2, \dots, n$ there exist r_{ij} for $j = 1, 2, \dots, i$ so that

$$\mathbf{a}_i = r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i.$$

In matrix times vector notation...

$$\mathbf{a}_i = r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i = Q \begin{bmatrix} r_{1i} \\ r_{2i} \\ \vdots \\ r_{ii} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

And therefore

$$A = \begin{bmatrix} \mathbf{a}_1, & \mathbf{a}_2, & \ldots, & \mathbf{a}_n \end{bmatrix} = Q \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix}.$$

Theorem 21.9

Let A be an $m \times n$ matrix of rank n. The Gram-Schmidt process can be used to find a decomposition $A = QR$, where Q is an $m \times n$ matrix with orthonormal column vectors that are a basis for $\operatorname{Col} A$, and R is an upper triangular $n \times n$ matrix, with positive diagonal entries.

Theorem 21.9

Let A be an $m \times n$ matrix of rank n. The Gram-Schmidt process can be used to find a decomposition $A = QR$, where Q is an $m \times n$ matrix with orthonormal column vectors that are a basis for $\operatorname{Col} A$, and R is an upper triangular $n \times n$ matrix, with positive diagonal entries.

In fact, $r_{11} = \|\mathbf{a}_1\|$ and $r_{ii} = \|\mathbf{y}_i\|$ for all $i = 2, \ldots, n$ (where $\mathbf{y}_i$ as in the description of the algorithm), and

$$r_{ij} = \mathbf{a}_j \cdot \mathbf{u}_i$$

for all $i = 1, 2, \ldots, j-1$ and $j = 2, 3, \ldots, n$.

Example 21.10

Find a QR-decomposition of

$$A = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}.$$

Theorem 21.11

Let A be an $m \times n$ matrix of rank n. The least squares solution of $A\mathbf{x} = \mathbf{b}$ is the same as the solution to

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

Theorem 21.11

Let A be an $m \times n$ matrix of rank n. The least squares solution of $A\mathbf{x} = \mathbf{b}$ is the same as the solution to

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

Note that you can use back substitution to find the solution, because R is upper triangular!

Example 21.12

Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}$$

and

$$\mathbf{b} = \begin{bmatrix} 0 \\ 2 \\ 2 \\ 2 \end{bmatrix}.$$

Theorem 21.13

If Q is a matrix with orthonormal columns, then $Q^T Q = I$.

Summarizing:

- First decomposition you learned: $A = LU$.
  - L is lower triangular, U is upper triangular.
  - Good for solving a system of equations $A\mathbf{x} = \mathbf{b}$.
- Second decomposition: $A = QR$.
  - Q has orthonormal columns, R is upper triangular with positive diagonal.
  - Good for finding least squared solution to (inconsistent) $A\mathbf{x} = \mathbf{b}$, or (equivalently) finding projection of $\mathbf{b}$ on $\operatorname{Col} A$.

Note that Q's columns span $\operatorname{Col} A$.

Note that Q's columns span $\operatorname{Col} A$.

We assumed that A's columns are linearly independent (they are the basis of some subspace). What happens (in the Gram-Schmidt process) if this is not the case?