The Gram-Schmidt Process

D. H. Luecking

MASC

10 April 2024

Since orthonormal bases are so important, we need a straightforward way to obtain them.

Since orthonormal bases are so important, we need a straightforward way to obtain them.

We've actually seen how to do this, but we didn't dwell too much on it, except to use it in the proof that every subspace of an inner product space has an orthonormal basis.

Since orthonormal bases are so important, we need a straightforward way to obtain them.

We've actually seen how to do this, but we didn't dwell too much on it, except to use it in the proof that every subspace of an inner product space has an orthonormal basis.

Theorem (Gram-Schmidt)

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ *is an independent set in an inner product space* V *then there exists and orthonormal set* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ *such that*
$\operatorname{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r) = \operatorname{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r)$.

Since orthonormal bases are so important, we need a straightforward way to obtain them.

We've actually seen how to do this, but we didn't dwell too much on it, except to use it in the proof that every subspace of an inner product space has an orthonormal basis.

Theorem (Gram-Schmidt)

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ is an independent set in an inner product space V then there exists and orthonormal set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ such that $\mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r) = \mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r)$.

The process takes r steps, and at the end of step j the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_j$ have the same span as $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j$.

Since orthonormal bases are so important, we need a straightforward way to obtain them.

We've actually seen how to do this, but we didn't dwell too much on it, except to use it in the proof that every subspace of an inner product space has an orthonormal basis.

Theorem (Gram-Schmidt)

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ is an independent set in an inner product space V then there exists and orthonormal set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ such that $\mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r) = \mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r)$.

The process takes r steps, and at the end of step j the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_j$ have the same span as $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j$. So, by the last step the set is found.

Since orthonormal bases are so important, we need a straightforward way to obtain them.

We've actually seen how to do this, but we didn't dwell too much on it, except to use it in the proof that every subspace of an inner product space has an orthonormal basis.

Theorem (Gram-Schmidt)

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ *is an independent set in an inner product space* V *then there exists and orthonormal set* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ *such that* $\mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r) = \mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r)$.

The process takes r steps, and at the end of step j the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_j$ have the same span as $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j$. So, by the last step the set is found. If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ is a basis for a subspace S of V, then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ is an orthonormal basis for S.

Since orthonormal bases are so important, we need a straightforward way to obtain them.

We've actually seen how to do this, but we didn't dwell too much on it, except to use it in the proof that every subspace of an inner product space has an orthonormal basis.

Theorem (Gram-Schmidt)

If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ *is an independent set in an inner product space* V *then there exists and orthonormal set* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ *such that* $\mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r) = \mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r)$.

The process takes r steps, and at the end of step j the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_j$ have the same span as $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_j$. So, by the last step the set is found. If $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ is a basis for a subspace S of V, then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ is an orthonormal basis for S.

When computing by hand, it is quite a bit easier to first create an orthogonal set $\mathbf{v}_1, \ldots, \mathbf{v}_r$ and then normalize them at the end by $\mathbf{u}_j = (1/\left\|\mathbf{v}_j\right\|)\mathbf{v}_j$.

The steps

- Step 1: set $\mathbf{v}_1 = \mathbf{x}_1$.

The steps

- Step 1: set $\mathbf{v}_1 = \mathbf{x}_1$.
- Step 2: set $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1$, where $\mathbf{p}_1$ is the projection of $\mathbf{x}_2$ onto the span of $\mathbf{v}_1$.

The steps

- Step 1: set $\mathbf{v}_1 = \mathbf{x}_1$.
- Step 2: set $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1$, where $\mathbf{p}_1$ is the projection of $\mathbf{x}_2$ onto the span of $\mathbf{v}_1$.

  That is $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1$.

The steps

- Step 1: set $\mathbf{v}_1 = \mathbf{x}_1$.
- Step 2: set $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1$, where $\mathbf{p}_1$ is the projection of $\mathbf{x}_2$ onto the span of $\mathbf{v}_1$.

  That is $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1$.
- Step 3: set $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2$ where $\mathbf{p}_2$ is the projection of $\mathbf{x}_3$ onto $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$.

The steps

- Step 1: set $\mathbf{v}_1 = \mathbf{x}_1$.
- Step 2: set $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1$, where $\mathbf{p}_1$ is the projection of $\mathbf{x}_2$ onto the span of $\mathbf{v}_1$. That is $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1$.
- Step 3: set $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2$ where $\mathbf{p}_2$ is the projection of $\mathbf{x}_3$ onto $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. That is

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2$$

The steps

- Step 1: set $\mathbf{v}_1 = \mathbf{x}_1$.
- Step 2: set $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1$, where $\mathbf{p}_1$ is the projection of $\mathbf{x}_2$ onto the span of $\mathbf{v}_1$. That is $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1$.
- Step 3: set $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2$ where $\mathbf{p}_2$ is the projection of $\mathbf{x}_3$ onto $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. That is
$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2$$
- Steps $k = 4$ thru r: set $\mathbf{v}_k = \mathbf{x}_k - \mathbf{p}_{k-1}$ where $\mathbf{p}_{k-1}$ is the projection of $\mathbf{x}_k$ onto $\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_{k-1})$.

The steps

- Step 1: set $\mathbf{v}_1 = \mathbf{x}_1$.
- Step 2: set $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1$, where $\mathbf{p}_1$ is the projection of $\mathbf{x}_2$ onto the span of $\mathbf{v}_1$. That is $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1$.
- Step 3: set $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2$ where $\mathbf{p}_2$ is the projection of $\mathbf{x}_3$ onto $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. That is
$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2$$
- Steps $k = 4$ thru r: set $\mathbf{v}_k = \mathbf{x}_k - \mathbf{p}_{k-1}$ where $\mathbf{p}_{k-1}$ is the projection of $\mathbf{x}_k$ onto $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1})$. That is
$$\mathbf{v}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} \frac{\langle \mathbf{x}_k, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} \mathbf{v}_i$$

The steps

- Step 1: set $\mathbf{v}_1 = \mathbf{x}_1$.
- Step 2: set $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1$, where $\mathbf{p}_1$ is the projection of $\mathbf{x}_2$ onto the span of $\mathbf{v}_1$. That is $\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1$.
- Step 3: set $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2$ where $\mathbf{p}_2$ is the projection of $\mathbf{x}_3$ onto $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$. That is
$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2$$
- Steps $k = 4$ thru r: set $\mathbf{v}_k = \mathbf{x}_k - \mathbf{p}_{k-1}$ where $\mathbf{p}_{k-1}$ is the projection of $\mathbf{x}_k$ onto $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1})$. That is
$$\mathbf{v}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} \frac{\langle \mathbf{x}_k, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} \mathbf{v}_i$$

Then step $r + 1$: set $\mathbf{u}_j = (1/\left\|\mathbf{v}_j\right\|)\mathbf{v}_j$ for each j.

Example: find an orthonormal basis for the span of

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix},\ \mathbf{x}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix},\ \mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Example: find an orthonormal basis for the span of

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix},\ \mathbf{x}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix},\ \mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

Example: find an orthonormal basis for the span of

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}. \qquad \mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ -2/3 \\ 1 \\ 1/3 \end{pmatrix}.$$

Example: find an orthonormal basis for the span of

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix}. \qquad \mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ -2/3 \\ 1 \\ 1/3 \end{pmatrix}.$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} - \frac{2/3}{5/3} \begin{pmatrix} 1/3 \\ -2/3 \\ 1 \\ 1/3 \end{pmatrix} = \begin{pmatrix} -4/5 \\ 3/5 \\ 3/5 \\ 1/5 \end{pmatrix}.$$

Since $\|\mathbf{v}_1\| = \sqrt{3}$, $\|\mathbf{v}_2\| = \sqrt{15}/3$ and $\|\mathbf{v}_3\| = (\sqrt{35})/5$ we get the orthonormal basis

$$\mathbf{u}_1 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 0 \\ 1/\sqrt{3} \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 1/\sqrt{15} \\ -2/\sqrt{15} \\ 3/\sqrt{15} \\ 1/\sqrt{15} \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} -4/\sqrt{35} \\ 3/\sqrt{35} \\ 3/\sqrt{35} \\ 1/\sqrt{35} \end{pmatrix}.$$

Since $\|\mathbf{v}_1\| = \sqrt{3}$, $\|\mathbf{v}_2\| = \sqrt{15}/3$ and $\|\mathbf{v}_3\| = (\sqrt{35})/5$ we get the orthonormal basis

$$\mathbf{u}_1 = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 0 \\ 1/\sqrt{3} \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 1/\sqrt{15} \\ -2/\sqrt{15} \\ 3/\sqrt{15} \\ 1/\sqrt{15} \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} -4/\sqrt{35} \\ 3/\sqrt{35} \\ 3/\sqrt{35} \\ 1/\sqrt{35} \end{pmatrix}.$$

A further simplification: since the $\mathbf{v}$s don't have to have norm equal to 1, we can replace $\mathbf{v}_2$ by any multiple of it. So taking $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -2 \\ 3 \\ 1 \end{pmatrix}$ instead of what we used, the next step simplifies to:

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{15} \begin{pmatrix} 1 \\ -2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -4/5 \\ 3/5 \\ 3/5 \\ 1/5 \end{pmatrix}$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{3}\begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{15}\begin{pmatrix} 1 \\ -2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -4/5 \\ 3/5 \\ 3/5 \\ 1/5 \end{pmatrix}$$

and then we could also use $\mathbf{v}_3 = \begin{pmatrix} -4 \\ 3 \\ 3 \\ 1 \end{pmatrix}$.

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{2}{3}\begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{15}\begin{pmatrix} 1 \\ -2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -4/5 \\ 3/5 \\ 3/5 \\ 1/5 \end{pmatrix}$$

and then we could also use $\mathbf{v}_3 = \begin{pmatrix} -4 \\ 3 \\ 3 \\ 1 \end{pmatrix}$.

Because the norms are, respectively, $\sqrt{3}$, $\sqrt{15}$, and $\sqrt{35}$, we will end up with the same orthonormal basis.

The Gram-Schmidt process can be applied to a dependent set, but it will sometimes produce $\mathbf{v}_j = \mathbf{0}$. In that case we discard $\mathbf{x}_j$ and continue with the rest.

Sometimes one of the $\mathbf{x}_j$ is orthogonal to all the previous $\mathbf{x}_i$. In that case we get $\mathbf{p}_{j-1} = \mathbf{0}$ and $\mathbf{v}_j = \mathbf{x}_j$.

Example: Find an orthonormal basis for $\mathbb{R}^3$ Using the Gram-Schmidt on the set

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \ \mathbf{x}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Sometimes one of the $\mathbf{x}_j$ is orthogonal to all the previous $\mathbf{x}_i$. In that case we get $\mathbf{p}_{j-1} = \mathbf{0}$ and $\mathbf{v}_j = \mathbf{x}_j$.

Example: Find an orthonormal basis for $\mathbb{R}^3$ Using the Gram-Schmidt on the set

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Sometimes one of the $\mathbf{x}_j$ is orthogonal to all the previous $\mathbf{x}_i$. In that case we get $\mathbf{p}_{j-1} = \mathbf{0}$ and $\mathbf{v}_j = \mathbf{x}_j$.

Example: Find an orthonormal basis for $\mathbb{R}^3$ Using the Gram-Schmidt on the set

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{0}{1}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Sometimes one of the $\mathbf{x}_j$ is orthogonal to all the previous $\mathbf{x}_i$. In that case we get $\mathbf{p}_{j-1} = \mathbf{0}$ and $\mathbf{v}_j = \mathbf{x}_j$.

Example: Find an orthonormal basis for $\mathbb{R}^3$ Using the Gram-Schmidt on the set

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{x}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \frac{0}{1}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{p}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{1}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1/2 \\ 1/2 \end{pmatrix}$$

So, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

So, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

The transition matrix from the basis $\mathcal{B} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ to the basis $\mathcal{C} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ has a particularly nice form.

So, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

The transition matrix from the basis $\mathcal{B} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ to the basis $\mathcal{C} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ has a particularly nice form.

The transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{E}$ is $A = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{pmatrix}$.

So, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

The transition matrix from the basis $\mathcal{B} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ to the basis $\mathcal{C} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ has a particularly nice form.

The transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{E}$ is $A = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{pmatrix}$. The transition matrix from $\mathcal{C}$ to the standard basis $\mathcal{E}$ is $Q = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}$.

So, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

The transition matrix from the basis $\mathcal{B} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ to the basis $\mathcal{C} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ has a particularly nice form.

The transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{E}$ is $A = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{pmatrix}$. The transition matrix from $\mathcal{C}$ to the standard basis $\mathcal{E}$ is $Q = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}$.

So the transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{C}$ is $Q^T A = (\mathbf{u}_i^T \mathbf{x}_j)$. But if $i > j$ we know that $\mathbf{u}_i$ is orthogonal to $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}) = \mathrm{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$.

So, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

The transition matrix from the basis $\mathcal{B} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ to the basis $\mathcal{C} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ has a particularly nice form.

The transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{E}$ is $A = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{pmatrix}$. The transition matrix from $\mathcal{C}$ to the standard basis $\mathcal{E}$ is $Q = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}$.

So the transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{C}$ is $Q^T A = (\mathbf{u}_i^T \mathbf{x}_j)$. But if $i > j$ we know that $\mathbf{u}_i$ is orthogonal to $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}) = \mathrm{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$. That is, $\mathbf{u}_i^T \mathbf{x}_j = 0$ when $i > j$. This means that $Q^T A$ is upper triangular.

So, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

The transition matrix from the basis $\mathcal{B} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ to the basis $\mathcal{C} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ has a particularly nice form.

The transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{E}$ is $A = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{pmatrix}$. The transition matrix from $\mathcal{C}$ to the standard basis $\mathcal{E}$ is $Q = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}$.

So the transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{C}$ is $Q^T A = (\mathbf{u}_i^T \mathbf{x}_j)$. But if $i > j$ we know that $\mathbf{u}_i$ is orthogonal to $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}) = \mathrm{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$. That is, $\mathbf{u}_i^T \mathbf{x}_j = 0$ when $i > j$. This means that $Q^T A$ is upper triangular.

Something like this happens in every Gram-Scmidt process. Start with an $n \times k$ matrix A, whose columns are independent vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\mathbb{R}^n$.

So, $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, $\mathbf{u}_3 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$.

The transition matrix from the basis $\mathcal{B} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3]$ to the basis $\mathcal{C} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ has a particularly nice form.

The transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{E}$ is $A = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{pmatrix}$. The transition matrix from $\mathcal{C}$ to the standard basis $\mathcal{E}$ is $Q = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}$.

So the transition matrix from $\mathcal{B}$ to the standard basis $\mathcal{C}$ is $Q^T A = (\mathbf{u}_i^T \mathbf{x}_j)$. But if $i > j$ we know that $\mathbf{u}_i$ is orthogonal to $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}) = \mathrm{Span}(\mathbf{x}_1, \ldots, \mathbf{x}_{i-1})$. That is, $\mathbf{u}_i^T \mathbf{x}_j = 0$ when $i > j$. This means that $Q^T A$ is upper triangular.

Something like this happens in every Gram-Scmidt process. Start with an $n \times k$ matrix A, whose columns are independent vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ in $\mathbb{R}^n$. Apply Gram-Schmidt to them to get orthonormal vectors $\mathbf{q}_1, \ldots, \mathbf{q}_k$.

If we let $Q = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_k \end{pmatrix}$ then $R = Q^T A$ is an upper triangular $k \times k$ matrix and $A = QR$ where Q is an $n \times k$ with orthonormal columns. This is called the QR-factorization of A.

If we let $Q = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_k \end{pmatrix}$ then $R = Q^T A$ is an upper triangular $k \times k$ matrix and $A = QR$ where Q is an $n \times k$ with orthonormal columns. This is called the QR-factorization of A.

Note that R is a square matrix. Also its diagonals are $\mathbf{q}_j^T \mathbf{x}_j$. These are not zero because $\mathbf{q}_j = (1/\|\mathbf{v}_j\|)\mathbf{v}_j$ where $\mathbf{v}_j$ has the form

$$\mathbf{v}_j = \mathbf{x}_j - (\text{a linear combination of } \mathbf{v}_i, i < j)$$

If we let $Q = \begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_k \end{pmatrix}$ then $R = Q^T A$ is an upper triangular $k \times k$ matrix and $A = QR$ where Q is an $n \times k$ with orthonormal columns. This is called the QR-factorization of A.

Note that R is a square matrix. Also its diagonals are $\mathbf{q}_j^T \mathbf{x}_j$. These are not zero because $\mathbf{q}_j = (1/\|\mathbf{v}_j\|)\mathbf{v}_j$ where $\mathbf{v}_j$ has the form

$$\mathbf{v}_j = \mathbf{x}_j - (\text{a linear combination of } \mathbf{v}_i, i < j)$$

which means

$$\mathbf{x}_j = \mathbf{v}_j + (\text{a linear combination of } \mathbf{v}_i, i < j)$$

and that says that $\mathbf{q}_j^T \mathbf{x}_j = (1/\|\mathbf{v}_j\|)\mathbf{v}_j^T \mathbf{v}_j = \|\mathbf{v}_j\| \neq 0$.

Examples: our first Gram-Schmidt example gives us

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{15} & -4/\sqrt{35} \\ 1/\sqrt{3} & -2/\sqrt{15} & 3/\sqrt{35} \\ 0 & 3/\sqrt{15} & 3/\sqrt{35} \\ 1/\sqrt{3} & 1/\sqrt{15} & 1/\sqrt{35} \end{pmatrix}$$

and

$$Q^T A = R = \begin{pmatrix} 3/\sqrt{3} & 2/\sqrt{3} & 2/\sqrt{3} \\ 0 & 5/\sqrt{15} & 2/\sqrt{15} \\ 0 & 0 & 7/\sqrt{35} \end{pmatrix}$$

Our second Gram-Schmidt example gives us:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Our second Gram-Schmidt example gives us:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

and

$$Q^T A = R = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{pmatrix}$$

The QR factorization of A allows a simplified calculation of least squares problems.

The QR factorization of A allows a simplified calculation of least squares problems.

Solving $A\mathbf{x} = \mathbf{b}$ with least squares would normally require solving $A^T A\mathbf{x} = A^T\mathbf{b}$.

The QR factorization of A allows a simplified calculation of least squares problems. Solving $A\mathbf{x} = \mathbf{b}$ with least squares would normally require solving $A^T A\mathbf{x} = A^T\mathbf{b}$. The multiplication by A^T is designed to kill off the part of $\mathbf{b}$ orthogonal to the column space of A.

The QR factorization of A allows a simplified calculation of least squares problems.

Solving $A\mathbf{x} = \mathbf{b}$ with least squares would normally require solving $A^T A\mathbf{x} = A^T\mathbf{b}$. The multiplication by A^T is designed to kill off the part of $\mathbf{b}$ orthogonal to the column space of A.

But the columns of Q have the same span as the columns of A, and therefore $\mathcal{N}(A^T) = \mathcal{R}(A)^\perp = \mathcal{R}(Q)^\perp = \mathcal{N}(Q^T)$.

The QR factorization of A allows a simplified calculation of least squares problems.

Solving $A\mathbf{x} = \mathbf{b}$ with least squares would normally require solving $A^T A\mathbf{x} = A^T\mathbf{b}$. The multiplication by A^T is designed to kill off the part of $\mathbf{b}$ orthogonal to the column space of A.

But the columns of Q have the same span as the columns of A, and therefore $\mathcal{N}(A^T) = \mathcal{R}(A)^\perp = \mathcal{R}(Q)^\perp = \mathcal{N}(Q^T)$. So, multiplying by Q^T has the same effect. That is, $Q^T A\mathbf{x} = Q^T\mathbf{b}$ will be consistant.

The QR factorization of A allows a simplified calculation of least squares problems.

Solving $A\mathbf{x} = \mathbf{b}$ with least squares would normally require solving $A^T A\mathbf{x} = A^T\mathbf{b}$. The multiplication by A^T is designed to kill off the part of $\mathbf{b}$ orthogonal to the column space of A.

But the columns of Q have the same span as the columns of A, and therefore $\mathcal{N}(A^T) = \mathcal{R}(A)^\perp = \mathcal{R}(Q)^\perp = \mathcal{N}(Q^T)$. So, multiplying by Q^T has the same effect. That is, $Q^T A\mathbf{x} = Q^T\mathbf{b}$ will be consistant.

It is also particularly easy to solve because $Q^T A = R$ is upper triangular and $R\mathbf{x} = Q^T\mathbf{b}$ allows a relatively quick solution via back substitution.

Example (using the A, Q and R from earlier): solve

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Example (using the A, Q and R from earlier): solve

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Multiplying by Q^T gives

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 5/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Example (using the A, Q and R from earlier): solve

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Multiplying by Q^T gives

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 5/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

This is a little easier to handle if we multiply the second and third rows (equations) by $\sqrt{2}$:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}$$

Example (using the A, Q and R from earlier): solve

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Multiplying by Q^T gives

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 5/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

This is a little easier to handle if we multiply the second and third rows (equations) by $\sqrt{2}$:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}$$

This has solution $\mathbf{x} = \begin{pmatrix} 0 & 2 & 1 \end{pmatrix}^T$.