Most important ideas:

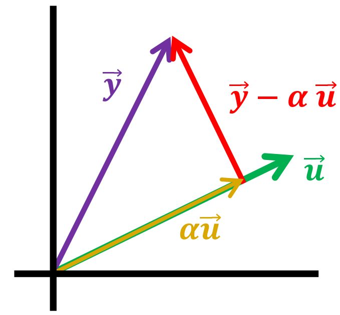
- Projection of one vector onto another vector.
- Orthogonal vectors.
- Orthonormal vectors; orthogonal matrix, including that its transpose is its inverse (whhaaaat?)

Reminder: Suppose $f(x) = a + bx + cx^2$ (where $c > 0$). Then $f(x)$ is minimized where:

$\boldsymbol{f'(x) = b + 2cx = 0}$ **which occurs when** $\boldsymbol{x = -\frac{b}{2c}}$.

Suppose $f(\alpha) = a + b\,\alpha + c\,\alpha^2$ (where $c > 0$). Then $f(\alpha)$ is minimized where:

$\boldsymbol{f'(\alpha) = b + 2c\alpha = 0}$ **which occurs when** $\boldsymbol{\alpha = -\frac{b}{2\alpha}}$.



Goal: Given vectors $\vec{u}$ and $\vec{y}$, find the point/vector $\alpha\vec{u}$ on/along $\vec{u}$ that is closest to $\vec{y}$.

Note that the vector between $\vec{y}$ and $\alpha\vec{u}$ is $\vec{y} - \alpha\vec{u}$. So another way of thinking of this problem is that we to find the value of α for which we **minimize the distance from** $\boldsymbol{\vec{y}}$ **to** $\boldsymbol{\alpha\vec{u}}$**, that is, the size of** $\boldsymbol{\vec{y} - \alpha\vec{u}}$:

$$\|\vec{y} - \alpha\vec{u}\| = \sqrt{(\vec{y} - \alpha\vec{u}) \cdot (\vec{y} - \alpha\vec{u})}$$

Minimizing $\|\vec{y} - \alpha\vec{u}\|$ is equivalent to minimizing $\|\vec{y} - \alpha\vec{u}\|^2$. Let

$$f(\alpha) = \|\vec{y} - \alpha\vec{u}\|^2 = (\vec{y} - \alpha\vec{u}) \cdot (\vec{y} - \alpha\vec{u})$$
$$= \vec{y} \cdot \vec{y} - \alpha\vec{u} \cdot \vec{y} - \alpha\vec{u} \cdot \vec{y} + \alpha^2\vec{u} \cdot \vec{u} = \vec{y} \cdot \vec{y} - 2\alpha\vec{u} \cdot \vec{y} + \alpha^2\vec{u} \cdot \vec{u}$$

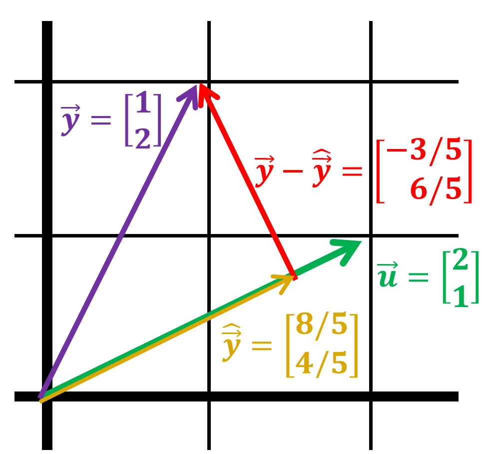
Then $\boldsymbol{f(\alpha)}$ **is minimized where:**

$$\boldsymbol{f'(\alpha) = 0 - 2\,\vec{u} \cdot \vec{y} + 2\,\alpha\,\vec{u} \cdot \vec{u} = 0} \text{ **which occurs when** } \boldsymbol{\alpha = \frac{\vec{y} \cdot \vec{u}}{\vec{u} \cdot \vec{u}}}.$$

Example 1: $\boldsymbol{\vec{u}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\boldsymbol{\vec{y}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Then the point $\boldsymbol{\widehat{\vec{y}}}$ on $\vec{u}$ that is closest to $\vec{y}$ is known as **the projection of** $\boldsymbol{\vec{y}}$ **onto** $\boldsymbol{\vec{u}}$:

$$\boldsymbol{\widehat{\vec{y}} = proj_{\vec{u}}\vec{y} = \alpha\,\vec{u} = \frac{\vec{y}\ \cdot\ \vec{u}}{\vec{u}\ \cdot\ \vec{u}}\,\vec{u} = \frac{4}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8/5 \\ 4/5 \end{bmatrix}}.$$

Note that $\boldsymbol{\vec{y} - \widehat{\vec{y}}} \perp \boldsymbol{\vec{u}}$ since $\begin{bmatrix} -3/5 \\ 6/5 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 0$.



Alternate (and very important!) approach: find the value of $\boldsymbol{\alpha}$ **so that** $\boldsymbol{\vec{y} - \alpha\vec{u}}$ **is orthogonal to** $\boldsymbol{\vec{u}}$**. That is, choose** $\boldsymbol{\alpha}$ **so that** $\boldsymbol{(\vec{y} - \alpha\vec{u}) \cdot \vec{u} = 0}$**:**

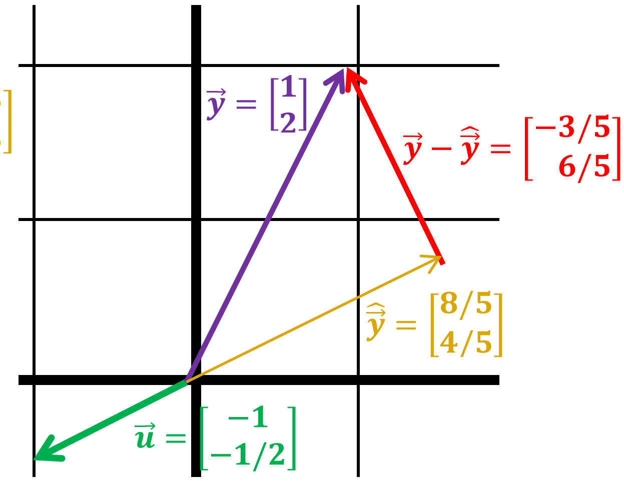
$$\boldsymbol{\vec{y} \cdot \vec{u} - \alpha\vec{u} \cdot \vec{u} = 0 \quad \Rightarrow \quad \alpha = \frac{\vec{y} \cdot \vec{u}}{\vec{u} \cdot \vec{u}}}$$

Next, suppose $\vec{u}$ is some other vector but in the *same direction* as the $\vec{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ in Example 1.

Example 2: If $\vec{\boldsymbol{u}} = \begin{bmatrix} -1 \\ -1/2 \end{bmatrix}$ and $\vec{\boldsymbol{y}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, then

$\hat{\vec{y}} = proj_{\vec{u}}\vec{y} = \frac{\vec{y}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\vec{u} = \frac{2}{5/4}\begin{bmatrix} -1 \\ -1/2 \end{bmatrix} = -\frac{8}{5}\begin{bmatrix} -1 \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 8/5 \\ 4/5 \end{bmatrix}$

as found before.



If $\hat{\vec{\boldsymbol{y}}} = \frac{\vec{\boldsymbol{y}}\cdot\vec{\boldsymbol{u}}}{\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{u}}}\vec{\boldsymbol{u}}$, then $\vec{\boldsymbol{y}} - \hat{\vec{\boldsymbol{y}}}$ is orthogonal to $\vec{\boldsymbol{u}}$:

$$\left(\vec{\boldsymbol{y}} - \hat{\vec{\boldsymbol{y}}}\right)\cdot\vec{\boldsymbol{u}} = \left(\vec{\boldsymbol{y}} - \frac{\vec{y}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\vec{\boldsymbol{u}}\right) = \vec{\boldsymbol{y}}\cdot\vec{\boldsymbol{u}} - \frac{\vec{y}\cdot\vec{u}}{\vec{u}\cdot\vec{u}}\vec{\boldsymbol{u}}\cdot\vec{\boldsymbol{u}} = \mathbf{0}.$$

Note that $\vec{y}\cdot\vec{u}$ and $\vec{u}\cdot\vec{u}$ are simply numbers.

Definition: $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$ is an <u>orthogonal set</u> if every vector in the set is orthogonal to every other vector in the set. That is, $\vec{u}_i \cdot \vec{u}_j = 0$ for $i \neq j$. (What about if $i = j$? Recall that $\vec{u}\cdot\vec{u} \geq 0$ and that $\vec{u}\cdot\vec{u} = 0$ if and only if $\vec{u} = \vec{0}$.)

Example 3: $\vec{u}_1 = \begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix}$, $\vec{u}_3 = \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix}$.

$\vec{u}_1 \cdot \vec{u}_2 = -3 - 6 - 3 + 12 = 0.$

$\vec{u}_1 \cdot \vec{u}_3 = \ \ 9 - 16 + 7 + 0 = 0.$

$\vec{u}_2 \cdot \vec{u}_3 = -3 + 24 - 21 + 0 = 0.$

What does it look like for two vectors in R^4 to be orthogonal? I don't know what anything in four dimensions looks like! Luckily, it doesn't matter at all whether we can visualize this.

Orthogonal is even better than linearly independent. (Is this similar to how being infamous is even better than famous in "The Three Amigos.") **In fact, any set of (non-zero) orthogonal vectors is linearly independent. But being linearly independent does not necessarily mean the vectors are orthogonal.** Let's look at Theorem 4 and its proof on page 338.

Example 4: $\vec{u}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$. These vectors are orthogonal and linearly independent.

$\vec{u}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$. These vectors are linearly independent, but not orthogonal.

Extremely useful: If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$ is an orthogonal basis for W and if $\vec{y} \in W$, then

$$\vec{y} = \frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\,\vec{u}_1 + \frac{\vec{y}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\,\vec{u}_2 + \cdots + \frac{\vec{y}\cdot\vec{u}_p}{\vec{u}_p\cdot\vec{u}_p}\,\vec{u}_p$$

Intuition: $\frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\,\vec{u}_1$ is the part of $\vec{y}$ that comes from $\vec{u}_1$.

$\frac{\vec{y}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\,\vec{u}_2$ is the part of $\vec{y}$ that comes from $\vec{u}_2$. And so on.

Let's look at Theorem 5 and its proof on page 339.

Example 5 (compare to Example 1): If $\vec{u}_1 = \begin{bmatrix}2\\1\end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix}-1\\2\end{bmatrix}$, $\vec{y} = \begin{bmatrix}1\\2\end{bmatrix}$, then

$$\vec{y} = \frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\,\vec{u}_1 + \frac{\vec{y}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\,\vec{u}_2 = \frac{4}{5}\begin{bmatrix}2\\1\end{bmatrix} + \frac{3}{5}\begin{bmatrix}-1\\2\end{bmatrix} = \begin{bmatrix}8/5\\4/5\end{bmatrix} + \begin{bmatrix}-3/5\\6/5\end{bmatrix}.$$

Example 6: $\vec{u}_1 = \begin{bmatrix}3\\-2\\1\\3\end{bmatrix}$, $\vec{u}_2 = \begin{bmatrix}-1\\3\\-3\\4\end{bmatrix}$, $\vec{u}_3 = \begin{bmatrix}3\\8\\7\\0\end{bmatrix}$. Suppose $\vec{y} = \begin{bmatrix}14\\-4\\16\\1\end{bmatrix}$.

Then $\vec{y} = \frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\,\vec{u}_1 + \frac{\vec{y}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\,\vec{u}_2 + \frac{\vec{y}\cdot\vec{u}_3}{\vec{u}_3\cdot\vec{u}_3}\,\vec{u}_3$

$$= \frac{69}{23}\begin{bmatrix}3\\-2\\1\\3\end{bmatrix} + \frac{-70}{35}\begin{bmatrix}-1\\3\\-3\\4\end{bmatrix} + \frac{122}{122}\begin{bmatrix}3\\8\\7\\0\end{bmatrix} = 3\begin{bmatrix}3\\-2\\1\\3\end{bmatrix} + (-2)\begin{bmatrix}-1\\3\\-3\\4\end{bmatrix} + 1\begin{bmatrix}3\\8\\7\\0\end{bmatrix} = \begin{bmatrix}9\\-6\\3\\9\end{bmatrix} + \begin{bmatrix}2\\-6\\6\\-8\end{bmatrix} + \begin{bmatrix}3\\8\\7\\0\end{bmatrix}$$

Note that the three vectors $\vec{u}_1, \vec{u}_2, \vec{u}_3$ do not span R^4. (How do we know this?). But I chose a vector $\vec{y}$ that is in their span, that is, that can be built (as a linear combination) from them. One case we'll deal with in the next book section is when $\vec{y}$ actually cannot be built from them.

Caution: the above formula is *not true* if the basis is *not orthogonal*.

Example 7: If $\vec{u}_1 = \begin{bmatrix}1\\2\end{bmatrix}, \vec{u}_2 = \begin{bmatrix}3\\4\end{bmatrix}, \vec{y} = \begin{bmatrix}5\\6\end{bmatrix}$. So $\vec{y}$ can be built (is a linear combination) from $\vec{u}_1$ and $\vec{u}_2$, but

$$\frac{\vec{y}\cdot\vec{u}_1}{\vec{u}_1\cdot\vec{u}_1}\,\vec{u}_1 + \frac{\vec{y}\cdot\vec{u}_2}{\vec{u}_2\cdot\vec{u}_2}\,\vec{u}_2 = \frac{17}{5}\begin{bmatrix}1\\2\end{bmatrix} + \frac{39}{25}\begin{bmatrix}3\\4\end{bmatrix} \neq \begin{bmatrix}5\\6\end{bmatrix}.$$

Definition: Even better than orthogonal is orthonormal, which means both orthogonal and normalized (length 1): $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}$ is orthonormal if every vector in the set is orthogonal to every other vector in the set and if every vector is of length 1:

$$\vec{u}_i\cdot\vec{u}_j\ (= \vec{u}_i^T\vec{u}_j\,) = \begin{cases}1 & \text{if } i = j\\ 0 & \text{if } i \neq j\end{cases}$$

Example 8: $\vec{u}_1 = \frac{1}{\sqrt{23}}\begin{bmatrix}3\\-2\\1\\3\end{bmatrix}$, $\vec{u}_2 = \frac{1}{\sqrt{35}}\begin{bmatrix}-1\\3\\-3\\4\end{bmatrix}$, $\vec{u}_3 = \frac{1}{\sqrt{122}}\begin{bmatrix}3\\8\\7\\0\end{bmatrix}$.

We saw in Example 6 that these vectors are orthogonal, and we have now normalized them (divided each one by its own length) to make them each length 1. For example,

$$\vec{u}_1 \cdot \vec{u}_1 = \frac{9}{23} + \frac{4}{23} + \frac{1}{23} + \frac{9}{23} = \frac{23}{23} = 1\text{, and similarly for } \vec{u}_2 \text{ and } \vec{u}_3.$$

A matrix $U = [\vec{u}_1\ \vec{u}_2\ \cdots\ \vec{u}_p]$ with orthonormal columns is an orthogonal matrix, or unitary matrix, hence the letter U. (I would call it an *orthonormal* matrix, but I'm not in charge.)

Example 9: $U = \begin{bmatrix}1/\sqrt{3} & -1/\sqrt{2}\\ 1/\sqrt{3} & 0\\ 1/\sqrt{3} & 1/\sqrt{2}\end{bmatrix}$. Notice that the columns are orthonormal.

Notice that $U^T U = \begin{bmatrix}1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\\ -1/\sqrt{2} & 0 & 1/\sqrt{2}\end{bmatrix}\begin{bmatrix}1/\sqrt{3} & -1/\sqrt{2}\\ 1/\sqrt{3} & 0\\ 1/\sqrt{3} & 1/\sqrt{2}\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$.

In general, $\boldsymbol{U^T U = [\vec{u}_1\ \vec{u}_2\ \cdots\ \vec{u}_p]^T[\vec{u}_1\ \vec{u}_2\ \cdots\ \vec{u}_p] = \begin{bmatrix}\vec{u}_1^T\\ \vec{u}_2^T\\ \vdots\\ \vec{u}_p^T\end{bmatrix}[\vec{u}_1\ \vec{u}_2\ \cdots\ \vec{u}_p]}$

$$\boldsymbol{= \begin{bmatrix}\vec{u}_1^T\vec{u}_1 & \vec{u}_1^T\vec{u}_2 & \cdots & \vec{u}_1^T\vec{u}_p\\ \vec{u}_2^T\vec{u}_1 & \vec{u}_2^T\vec{u}_2 & \cdots & \vec{u}_2^T\vec{u}_p\\ \vdots & \vdots & \ddots & \vdots\\ \vec{u}_p^T\vec{u}_1 & \vec{u}_p^T\vec{u}_2 & \cdots & \vec{u}_p^T\vec{u}_p\end{bmatrix} = \begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}}$$

Continuation of Example 9: add a third column which is orthogonal to the first two, and of size/length 1 (**there is only one such vector—why?**) to the two existing columns:

$$U = \begin{bmatrix}1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6}\\ 1/\sqrt{3} & 0 & -2/\sqrt{6}\\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6}\end{bmatrix}$$

$$U^T U = \begin{bmatrix}1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\\ -1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6}\end{bmatrix}\begin{bmatrix}1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6}\\ 1/\sqrt{3} & 0 & -2/\sqrt{6}\\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6}\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

So U^T is the inverse of U.

Recall from the Invertible Matrix Theorem for square matrix A, if $AB = I$ then it is also true that $BA = I$. **So if $\boldsymbol{U^T U = I}$ it must also be that $\boldsymbol{U U^T = I}$**, that is, $(U^T)^T U^T = I$. So U^T is also an orthogonal matrix (columns are orthonormal). That is, if the columns of U are orthonormal, then the columns of U^T (the rows of U) are also. This is a curious result.

$U^T = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}$. Is it obvious the columns are orthonormal? Nope.

We can check for orthogonality of columns , and we can check for normality (length 1).

Theorem 7 is interesting. In particular, if U is an orthogonal matrix, i.e. $U^TU = I$, then:

$$\|\boldsymbol{U\vec{x}}\| = \sqrt{(\boldsymbol{U\vec{x}})^T(\boldsymbol{U\vec{x}})} = \sqrt{\boldsymbol{\vec{x}^T U^T U \vec{x}}} = \sqrt{\boldsymbol{\vec{x}^T I \vec{x}}} = \sqrt{\boldsymbol{\vec{x}^T \vec{x}}} = \|\boldsymbol{\vec{x}}\|$$
$$(\boldsymbol{U\vec{x}}) \cdot (\boldsymbol{U\vec{y}}) = (\boldsymbol{U\vec{x}})^T(\boldsymbol{U\vec{y}}) = \boldsymbol{\vec{x}^T U^T U \vec{y}} = \boldsymbol{\vec{x}^T I \vec{y}} = \boldsymbol{\vec{x}^T \vec{y}} = \boldsymbol{\vec{x} \cdot \vec{y}}$$

Property (a) means that multiplying a vector by a orthogonal matrix doesn't change its size. A unitary matrix transforms (e.g. rotates, reflects, etc.) a vector, but doesn't change its size. Note: all rotation matrices are unitary, but not all unitary matrices are rotation matrices. Property (b) means that multiplying by an orthogonal matrix doesn't change the orthogonality of two vectors: if they were orthogonal before multiplying by $\boldsymbol{U}$, then they still are, and if they were not, they are still are not. These are unusual and useful properties.

Example: Suppose $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$, $\vec{x} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} -4 \\ 10 \end{bmatrix}$. Notice $\vec{x} \cdot \vec{y} = 0$.

So $\|\vec{x}\| = \sqrt{29}$

and $\|\vec{y}\| = \sqrt{116}$

$A\vec{x} = \begin{bmatrix} 9 \\ 23 \end{bmatrix}$ and $\|A\vec{x}\| = \sqrt{610}$

$A\vec{y} = \begin{bmatrix} 16 \\ 28 \end{bmatrix}$ and $\|A\vec{y}\| = \sqrt{1040}$

$U\vec{x} = \begin{bmatrix} \frac{7}{\sqrt{2}} \\ \frac{3}{\sqrt{2}} \end{bmatrix}$ and $\|U\vec{x}\| = \sqrt{29}$

$U\vec{y} = \begin{bmatrix} \frac{6}{\sqrt{2}} \\ -\frac{14}{\sqrt{2}} \end{bmatrix}$ and $\|U\vec{y}\| = \sqrt{116}$

Notice that $(U\vec{x}) \cdot (U\vec{y}) = 0$ but $(A\vec{x}) \cdot (A\vec{y}) \neq 0$.

One final note: all eigenvalues of any unitary matrix are of size 1. The eigenvalues of the above 2×2 U are ± 1, and the eigenvalues of the 3×3 matrix from the previous page are 1 and $\approx -.00720072 \pm .999974i$, the size of which is $\sqrt{(-.00720072)^2 + (\pm .999974)^2} \approx 1$.