A matrix Q has orthonormal columns $\iff$ $Q^TQ = I$

Why? Let $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ be the columns of Q. By the way matrix multiplication works, the entries of Q^TQ are dot products of these columns:

$$\begin{bmatrix} - & \boldsymbol{q}_1^T & - \\ - & \boldsymbol{q}_2^T & - \\ & \vdots & \end{bmatrix} \begin{bmatrix} | & | & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

Hence, $Q^TQ = I$ if and only if $\boldsymbol{q}_i^T\boldsymbol{q}_j = 0$ (that is, the columns are orthogonal), for $i \neq j$, and $\boldsymbol{q}_i^T\boldsymbol{q}_i = 1$ (that is, the columns are normalized).

Example 72. $Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}$ obtained from Example 70 satisfies $Q^TQ = I$.

The QR decomposition

Just like the LU decomposition encodes the steps of Gaussian elimination, the QR decomposition encodes the steps of Gram–Schmidt.

(QR decomposition) Every $m \times n$ matrix A of rank n can be decomposed as $A = QR$, where

- Q has orthonormal columns, $(m \times n)$
- R is upper triangular and invertible. $(n \times n)$

How to find Q and R?

- Gram–Schmidt orthonormalization on (columns of) A, to get (columns of) Q
- $R = Q^TA$
  Why? If $A = QR$, then $Q^TA = Q^TQR$ which simplifies to $R = Q^TA$ (since $Q^TQ = I$).

The decomposition $A = QR$ is unique if we require the diagonal entries of R to be positive (and this is exactly what happens when applying Gram–Schmidt).

Practical comment. Actually, no extra work is needed for computing R. All of its entries have been computed during Gram–Schmidt.

Variations. We can also arrange things so that Q is an $m \times m$ **orthogonal** matrix (this means Q is square and has orthonormal columns) and R a $m \times n$ upper triangular matrix. This is a tiny bit more work (and not required for many applications): we need to complement "our" Q with additional orthonormal columns and add corresponding zero rows to R. For square matrices this makes no difference.

Armin Straub
straub@southalabama.edu

Example 73. Determine the QR decomposition of $A=\begin{bmatrix} 1 & 1\\ 1 & -1\\ 1 & 1 \end{bmatrix}$.

Solution. The first step is Gram–Schmidt orthonormalization on the columns of A. We then use the resulting orthonormal vectors (they need to be normalized!) as the columns of Q.

We already did Gram–Schmidt in Example 70: from that work, we have $Q=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6}\\ 1/\sqrt{3} & -2/\sqrt{6}\\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}$.

Hence, $R=Q^TA=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}\begin{bmatrix} 1 & 1\\ 1 & -1\\ 1 & 1 \end{bmatrix}=\begin{bmatrix} \sqrt{3} & 1/\sqrt{3}\\ 0 & 4/\sqrt{6} \end{bmatrix}$.

Comment. The entries of R have actually all been computed during Gram–Schmidt, so that, if we pay attention, we could immediately write down R (no extra work required). Looking back at Example 70, can you see this?

Check. Indeed, $QR=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6}\\ 1/\sqrt{3} & -2/\sqrt{6}\\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}\begin{bmatrix} \sqrt{3} & 1/\sqrt{3}\\ 0 & 4/\sqrt{6} \end{bmatrix}=\begin{bmatrix} 1 & 1\\ 1 & -1\\ 1 & 1 \end{bmatrix}$ equals A.

Example 74. Determine the QR decomposition of $A=\begin{bmatrix} 0 & 2 & 1\\ 3 & 1 & 1\\ 0 & 0 & 1\\ 0 & 0 & 1 \end{bmatrix}$.

Solution. The first step is Gram–Schmidt orthonormalization on the columns of A. We then use the resulting orthonormal vectors as the columns of Q.

We already did Gram–Schmidt in Example 71: from that work, we have $Q=\begin{bmatrix} 0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1/\sqrt{2}\\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}$.

Hence, $R=Q^TA=\begin{bmatrix} 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 0 & 2 & 1\\ 3 & 1 & 1\\ 0 & 0 & 1\\ 0 & 0 & 1 \end{bmatrix}=\begin{bmatrix} 3 & 1 & 1\\ 0 & 2 & 1\\ 0 & 0 & \sqrt{2} \end{bmatrix}$.

Comment. As commented earlier, the entries of R have actually all been computed during Gram–Schmidt, so that, if we pay attention, we could immediately write down R (no extra work required). Looking back at Example 71, can you see this?

Letting Sage do the work for us.

```
Sage] A = matrix(QQbar, [[0,2,1],[3,1,1],[0,0,1],[0,0,1]])
Sage] A.QR(full=false)
```

$$\left(\begin{bmatrix} 0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0.7071067811865475?\\ 0 & 0 & 0.7071067811865475? \end{bmatrix}, \begin{bmatrix} 3 & 1 & 1\\ 0 & 2 & 1\\ 0 & 0 & 1.414213562373095? \end{bmatrix}\right)$$

Comment. Can you figure out what happens if you omit the `full=false`? Check out the comment under **Variations** in the statement of the QR decomposition. On the other hand, the `QQbar` is telling Sage to compute with algebraic numbers (instead of just rational numbers); if omitted, it would complain that square roots are not available

Example 75. (extra) Determine the QR decomposition of $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & -5 \\ 0 & 3 & 6 \end{bmatrix}$.

Solution. We first apply Gram–Schmidt orthonormalization to the columns of A. For a variation, like a computer, we normalize after each step (rather than normalize at the end):

- $\boldsymbol{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, so that $\boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.
- $\boldsymbol{b}_2 = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} - \left(\begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} \cdot \boldsymbol{q}_1 \right) \boldsymbol{q}_1 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$, so that $\boldsymbol{q}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.
- $\boldsymbol{b}_3 = \begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix} - \left(\begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix} \cdot \boldsymbol{q}_1 \right) \boldsymbol{q}_1 - \left(\begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix} \cdot \boldsymbol{q}_2 \right) \boldsymbol{q}_2 = \begin{bmatrix} 0 \\ -5 \\ 0 \end{bmatrix}$, so that $\boldsymbol{q}_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$.

Therefore, $Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$. Finally, $R = Q^T A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & -5 \\ 0 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}$.

In conclusion, we have found the QR decomposition: $\underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 0 & 0 & -5 \\ 0 & 3 & 6 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix}}_{R}$

Comment. As noted before, we actually could write down R without any additional computation. Indeed, realize that the second column of R, that is $[2, 3, 0]^T$ means that

$$\text{2nd col of } A = 2\boldsymbol{q}_1 + 3\boldsymbol{q}_2.$$

Which we already knew from our computation of $\boldsymbol{q}_2$! Also, by construction, we know that the second column of A is a linear combination of $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ only, and that $\boldsymbol{q}_3$ enters the story later on. This corresponds to the fact that R is always upper triangular.

Letting Sage do the work for us.

```
Sage] A = matrix(QQbar, [[1,2,4], [0,0,-5], [0,3,6]])
Sage] A.QR()
```

$$\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 6 \\ 0 & 0 & 5 \end{bmatrix} \right)$$

Comment. The `QQbar` is telling Sage to compute with algebraic numbers (instead of just rational numbers); in general, if omitted, it would complain that square roots are not available (because the matrices Q and R typically involve square roots). Here, we are lucky that square roots didn't creep in.

Example 76. (extra) Find the QR decomposition of $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

Solution. (final answer only) $A = QR$ with $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$ and $R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & \frac{1}{\sqrt{2}} & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}$.

Example 77. One practical application of the QR decomposition is solving systems of linear equations.

$$\begin{aligned} A\boldsymbol{x}=\boldsymbol{b} \quad &\iff \quad QR\boldsymbol{x}=\boldsymbol{b} && \text{(now, multiply with } Q^T \text{ from the left)} \\ &\implies \quad R\boldsymbol{x}=Q^T\boldsymbol{b} \end{aligned}$$

The last system is triangular and can be solved by back-substitution.

A couple of comments are in order:

- If A is $n\times n$ and invertible, then the "$\implies$" is actually a "$\iff$".
- The equation $R\boldsymbol{x}=Q^T\boldsymbol{b}$ is always consistent! (Recall that R is invertible.)

  Indeed, if A is not $n\times n$ or not invertible, then $R\boldsymbol{x}=Q^T\boldsymbol{b}$ gives the least squares solutions!

  Why? $A^TA\hat{\boldsymbol{x}}=A^T\boldsymbol{b} \iff \underbrace{(QR)^TQR}_{=R^TQ^TQR}\hat{\boldsymbol{x}}=(QR)^T\boldsymbol{b} \iff R^TR\hat{\boldsymbol{x}}=R^TQ^T\boldsymbol{b} \iff R\hat{\boldsymbol{x}}=Q^T\boldsymbol{b}$

  [For the last step we need that R is invertible, which is always the case when A is $m\times n$ of rank n.]
- So, how does the QR way of solving linear systems compare to our beloved Gaussian elimination (LU)?

  It turns out that QR is a little slower than LU but makes up for it in "numerical stability".

  What does that mean? When computing numerically, we use floating point arithmetic and approximate each number by an expression of the form $0.1234\cdot 10^{-16}$. A certain (fixed) number of bits is used to store the part 0.1234 (here, 4 decimal places of accuracy) as well as the exponent -16.

  Now, here is something terrible that can happen in numerical computations: mathematically, the quantities x and $(x+1)-1$ are exactly the same. However, numerically, they might not. Take, for instance, $x=0.1234\cdot 10^{-6}$. Then, to an accuracy of 4 decimal places, $x+1=0.1000\cdot 10^1$, so that $(x+1)-1=0.0000$. But $x\neq 0$. We completely lost all the information about x.

  To be numerically stable, an algorithm must avoid issues like that.

$\hat{\boldsymbol{x}}$ is a least squares solution of $A\boldsymbol{x}=\boldsymbol{b}$

$\iff R\hat{\boldsymbol{x}}=Q^T\boldsymbol{b}$ (where $A=QR$)

Example 78. (review) In Example 17, we diagonalized $A=\begin{bmatrix} 4 & 0 & 2 \\ 2 & 2 & 2 \\ 1 & 0 & 3 \end{bmatrix}$ as $A=PDP^{-1}$.

We found that one choice for P and D is $P=\begin{bmatrix} 2 & 0 & -1 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, $D=\begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

Spell out what that tells us about A!

Solution. The diagonal entries $5, 2, 2$ of D are the eigenvalues of A.
The columns of P are corresponding eigenvectors of A.

- $\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ is a 5-eigenvector of A (that is, $A\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}=5\begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$).
- The 2-eigenspace of A is 2-dimensional. A basis is $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$.

Example 79. Diagonalize the **symmetric matrix** $A=\begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix}$ as $A=PDP^{-1}$.

Review. Recall that a matrix A is symmetric if $A^T=A$.

Solution. We let Sage do the work for us:

```
Sage] A = matrix([[8,-6,2],[-6,7,-4],[2,-4,3]])
Sage] A.eigenmatrix_right()
```

$$\left(\begin{bmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 & 1 \\ -1 & \frac{1}{2} & 2 \\ \frac{1}{2} & -1 & 2 \end{bmatrix}\right)$$

This ouput shows that A is diagonalizable as $A=PDP^{-1}$ with $D=\begin{bmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $P=\begin{bmatrix} 1 & 1 & 1 \\ -1 & \frac{1}{2} & 2 \\ \frac{1}{2} & -1 & 2 \end{bmatrix}$.

Just to make sure. This means that the eigenvalues of A are $15, 3, 0$ (the diagonal entries of D).

Moreover, we have that $\begin{bmatrix} 1 \\ -1 \\ \frac{1}{2} \end{bmatrix}$ is a 15-eigenvector, $\begin{bmatrix} 1 \\ \frac{1}{2} \\ -1 \end{bmatrix}$ is a 3-eigenvector, and $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ is a 0-eigenvector.

Important observation. Note that the eigenspaces of A are orthogonal to each other here.
The **spectral theorem** says that this is true for all symmetric matrices A.

Example 80. Diagonalize the symmetric matrix $A=\begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix}$ as $A=PDP^T$.

Solution. By the previous example, we can diagonalize A as $\tilde{P}\,D\,\tilde{P}^{-1}$ with $\tilde{P}=\begin{bmatrix} 2 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & -2 & 2 \end{bmatrix}$ and $D=\begin{bmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

(To avoid fractions, we just scaled the first two columns of $\tilde{P}$, which are eigenvectors.)
Note that the columns of $\tilde{P}$ are orthogonal (this is due the spectral theorem). If we normalize them (they all have length $\sqrt{2^2+2^2+1}=3$), then we obtain the orthogonal matrix $P=\frac{1}{3}\begin{bmatrix} 2 & 2 & 1 \\ -2 & 1 & 2 \\ 1 & -2 & 2 \end{bmatrix}$.
Since $P^{-1}=P^T$, we now have $A=PDP^T$.

Example 81.

(a) Determine the eigenspaces of the symmetric matrix $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$.

(b) Diagonalize A as $A = PDP^T$.

Solution.

(a) The characteristic polynomial is $\begin{vmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{vmatrix} = \lambda^2 - 5\lambda = \lambda(\lambda - 5)$, and so A has eigenvalues $5, 0$.

The 5-eigenspace is $\operatorname{null}\left(\begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix}\right)$ has basis $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

The 0-eigenspace is $\operatorname{null}\left(\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\right)$ has basis $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

Important observation. The 5-eigenvector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and the 0-eigenvector $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ are orthogonal!

(b) Note that a usual diagonalization is of the form $A = PDP^{-1}$.

We need to choose P so that $P^{-1} = P^T$, which means that P must be **orthogonal** (meaning orthonormal columns). [Choosing such a P is only possible if the eigenspaces of A are orthogonal.]

Hence, we normalize the two eigenvectors to $\frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\frac{1}{\sqrt{5}}\begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

With $P = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}$, we then have $A = PDP^T$.

Armin Straub
straub@southalabama.edu