4.3 Orthogonal Matrices

Definition: A square matrix $\mathbf{Q}$ is called an *orthogonal matrix* if it satisfies $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$. $(*)$

In particular, orthogonality implies $\mathbf{Q}^{-1} = \mathbf{Q}^T$.

Proposition: A matrix $\mathbf{Q}$ is orthogonal **iff** its columns form an orthonormal basis with respect to the Euclidean dot product on $\mathbb{R}^n$.

Proof: Let $\vec{u}_1, \ldots, \vec{u}_n$ be the columns of $\mathbf{Q}$. Then, $\vec{u}_1^T, \ldots, \vec{u}_n^T$ are the rows of the transposed matrix $\mathbf{Q}^T$.

The (i,j) entry of the product $\mathbf{Q}^T\mathbf{Q}$ is given as the product of: the i^{th} row of $\mathbf{Q}^T$, and the j^{th} column of $\mathbf{Q}$.

Thus, the orthogonality requirement $(*)$ implies $\vec{u}_i \cdot \vec{u}_j = u_i^T\vec{u}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$

which are precisely the conditions for $\vec{u}_i$ to form an orthonormal basis. ■

Concretely: Let's characterize *all* orthogonal $\mathbf{Q}^{2\times 2}$.

A 2×2 matrix $\mathbf{Q} = [\vec{x}_1\ \vec{x}_2] = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is orthogonal **iff** its columns $\vec{x}_1$, $\vec{x}_2$, form an orthonormal basis.

Equivalently, the requirement $\mathbf{Q}^T\mathbf{Q} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}$,

implies that its entries must satisfy the algebraic equations:

$$a^2 + c^2 = 1, \quad ab + cd = 0, \quad b^2 + d^2 = 1.$$

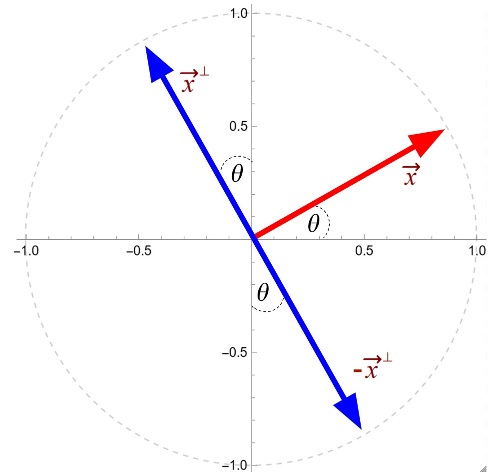
The first and last equations say that vectors (a, c) and (b, d) lie on the unit circle.

Therefore: $a = \cos\theta$, $c = \sin\theta$, $b = \cos\theta_2$, $d = \sin\theta_2$, for some θ, θ_2.

The remaining orthogonality condition above is now:

$$0 = ab + cd = \cos\theta\cos\theta_2 + \sin\theta\sin\theta_2 = \cos(\theta - \theta_2).$$

It implies θ and θ_2 differ by a right angle: $\theta_2 = \theta \pm \frac{\pi}{2}$.



Recharacterizing b, d in terms of θ, we have two cases:

$$b = -\sin\theta, \quad d = \cos\theta, \quad \text{or} \quad b = \sin\theta, \quad d = -\cos\theta.$$

Therefore, every 2×2 orthogonal matrix has the form:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ or } \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}. \qquad (*)$$

Lemma: An orthogonal $\mathbf{Q}$ has $|\mathbf{Q}| = \pm 1$.

Proof: Taking the determinant of $(*)$, and using the facts $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$ and $|\mathbf{A}^T| = |\mathbf{A}|$,

we have $1 = |\mathbf{I}| = |\mathbf{Q}^T\mathbf{Q}| = |\mathbf{Q}^T||\mathbf{Q}| = |\mathbf{Q}|^2$. ■

Definition: An orthogonal matrix is called *proper* or *special* if it has determinant +1.

An *improper* orthogonal matrix has determinant –1.

Proposition: The product of two orthogonal matrices is also orthogonal.

Proof: Let: $\mathbf{Q}_1^T\mathbf{Q}_1 = \mathbf{I} = \mathbf{Q}_2^T\mathbf{Q}_2$.

Want to show: $\mathbf{Q}_1\mathbf{Q}_2$ is orthogonal.

Observe that $(\mathbf{Q}_1\mathbf{Q}_2)^T(\mathbf{Q}_1\mathbf{Q}_2)$

$= \mathbf{Q}_2^T(\mathbf{Q}_1^T\mathbf{Q}_1)\mathbf{Q}_2 = \mathbf{Q}_2^T\mathbf{Q}_2 = \mathbf{I}$. ■

Example: True or false: a) If $\mathbf{Q}$ is an improper 2×2 orthogonal matrix, then $\mathbf{Q}^2 = \mathbf{I}$.

From $(*)$, $|\mathbf{Q}| = \begin{vmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{vmatrix} = -\cos^2\theta - \sin^2\theta = -1$. The other form gives $|\mathbf{Q}| = 1$.

Therefore all 2×2 improper orthogonal matrices take the form: $\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$.

And $\mathbf{Q}^2 = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}^2 = \begin{bmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}.$ ✓

So, true.

The QR Factorization

Now that we know about orthogonal matrices, we can recharacterize the Gram-Schmidt procedure as matrix factorization.

Let $\vec{w}_1, \ldots, \vec{w}_n$ be a basis of $\mathbb{R}^n$, and let $\vec{u}_1, \ldots, \vec{u}_n$ be the corresponding orthonormal basis that results from the Gram-Schmidt process.

Assemble nonsingular $n \times n$ matrices: $\mathbf{A} = \begin{bmatrix} \vec{w}_1 & \ldots & \vec{w}_n \end{bmatrix}$, $\mathbf{Q} = \begin{bmatrix} \vec{u}_1 & \ldots & \vec{u}_n \end{bmatrix}$.

Since the $\vec{u}_i$ form an orthonormal basis, **Q** is orthogonal.

Recall the Gram-Schmidt equation (see previous section):

$$\begin{aligned}
\vec{w}_1 &= r_{11}\vec{u}_1, \\
\vec{w}_2 &= r_{12}\vec{u}_1 + r_{22}\vec{u}_2, \\
\vec{w}_3 &= r_{13}\vec{u}_1 + r_{23}\vec{u}_2 + r_{33}\vec{u}_3, \\
\vdots \quad & \quad \vdots \qquad\quad \vdots \qquad\qquad \ddots \\
\vec{w}_n &= r_{1n}\vec{u}_1 + r_{2n}\vec{u}_2 + \ldots + r_{nn}\vec{u}_n, \qquad (**)
\end{aligned}$$

where $r_{ij} := \langle \vec{w}_j, \vec{u}_i \rangle$.

Also, recall the matrix multiplication formula, if $\mathbf{R} = \begin{bmatrix} \vec{r}_1 & \ldots & \vec{r}_n \end{bmatrix}$, then: $\mathbf{QR} = \begin{bmatrix} \mathbf{Q}\vec{r}_1 & \ldots & \mathbf{Q}\vec{r}_k \end{bmatrix}$.

The Gram-Schmidt equation can now be recast into an equivalent matrix form:

$$\mathbf{A} = \mathbf{QR}, \text{ where } \mathbf{R} =: \begin{bmatrix} \vec{r}_1 & \ldots & \vec{r}_n \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \ldots & r_{1n} \\ 0 & r_{22} & \ldots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & r_{nn} \end{bmatrix}. \qquad \text{(check this!)}$$

Only requirement on **A** is that its columns form a basis of $\mathbb{R}^n$ (nonsingular).

Theorem: Every nonsingular **A** can be factored, $\mathbf{A} = \mathbf{QR}$, into the product of an orthogonal matrix **Q** and an upper triangular matrix **R**. The factorization is unique if **R** has positive diagonal entries.

There is a more efficient algorithm to calculate the **QR** factorization for $\vec{w}_i$. The algorithm will rely on the following facts:

(1) Given an orthonormal basis $\vec{u}_i$, recall from a previous theorem that: $|\vec{w}| = \sqrt{\sum_{i=1}^{n} \langle \vec{w}, \vec{u}_i \rangle^2}$.

(2) Also, recall we defined $r_{ij} := \langle \vec{w}_j, \vec{u}_i \rangle$.

(3) Therefore, we have: $|\vec{w}_j|^2 = r_{1j}^2 + \ldots + r_{j-1,j}^2 + r_{jj}^2$ or $r_{jj} = \sqrt{|\vec{w}_j|^2 - (r_{1j}^2 + \ldots + r_{j-1,j}^2)}$.

(4) We see from $(**)$ that each $\vec{u}_i$ can be solved for: $\vec{u}_n = \frac{\vec{w}_n - \left(r_{1n}\vec{u}_1 + r_{2n}\vec{u}_2 + \ldots + r_{(n-1)n}\vec{u}_{n-1}\right)}{r_{nn}}$.

Let's learn the algorithm through the following example:

Example: Find the **QR** factorization of $\mathbf{A} = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix}$.

So the column vectors here are: $\vec{w}_1 := (0,-1,-1)$, $\vec{w}_2 = (1,1,1)$, and $\vec{w}_3 = (2,1,3)$.

In the end, we expect something of the form:

$$\vec{w}_1 = r_{11}\vec{u}_1,$$
$$\vec{w}_2 = r_{12}\vec{u}_1 + r_{22}\vec{u}_2,$$
$$\vec{w}_3 = r_{13}\vec{u}_1 + r_{23}\vec{u}_2 + r_{33}\vec{u}_3. \qquad (***)$$

The first step is to normalize $\vec{w}_1$: $r_{11} = |\vec{w}_1| = \sqrt{1^2 + 1^2} = \sqrt{2}$, so $\vec{u}_1 = \frac{\vec{w}_1}{|\vec{w}_1|} = \left(0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$.

From (2) above, we compute: $r_{12} = \langle \vec{w}_2, \vec{u}_1 \rangle = \left\langle (1,1,1), \left(0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \right\rangle = 0 - \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} = -\sqrt{2}$.

From (3): $r_{22} = \sqrt{|\vec{w}_2|^2 - r_{12}^2} = \sqrt{(1^2 + 1^2 + 1^2) - 2} = 1$.

Therefore, from (4): $\vec{u}_2 = \frac{\vec{w}_2 - r_{12}\vec{u}_1}{r_{22}} = \frac{(1,1,1) - (-\sqrt{2})\left(0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)}{1} = (1,0,0)$.

Working on the next vector, from (2): $r_{13} = \langle \vec{w}_3, \vec{u}_1 \rangle = \left\langle (2,1,3), \left(0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) \right\rangle = 0 - \frac{1}{\sqrt{2}} - \frac{3}{\sqrt{2}} = -2\sqrt{2}$.

Also: $r_{23} = \langle \vec{w}_3, \vec{u}_2 \rangle = \langle (2,1,3), (1,0,0) \rangle = 2$.

From (3): $r_{33} = \sqrt{|\vec{w}_3|^2 - r_{13}^2 - r_{23}^2} = \sqrt{(2^2 + 1^2 + 3^2) - 8 - 4} = \sqrt{2}$.

Therefore, from (4): $\vec{u}_3 = \frac{\vec{w}_3 - r_{13}\vec{u}_1 - r_{23}\vec{u}_2}{r_{33}} = \frac{(2,1,3) - (-2\sqrt{2})\left(0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) - 2(1,0,0)}{\sqrt{2}} = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

So: $\mathbf{Q} = \left[\vec{u}_1 \ \vec{u}_2 \ \vec{u}_3\right]$ and $\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & -\sqrt{2} & -2\sqrt{2} \\ 0 & 1 & 2 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$.

Checking my work: $\mathbf{QR} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & -\sqrt{2} & -2\sqrt{2} \\ 0 & 1 & 2 \\ 0 & 0 & \sqrt{2} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 1 & 1 \\ -1 & 1 & 3 \end{bmatrix} = [\vec{w}_1 \ \vec{w}_2 \ \vec{w}_3]$. ✓