
Numerical Analysis Hilary Term 2018
Lecture 6: QR Factorization

Definition: a square real matrix Q is **orthogonal** if $Q^T = Q^{-1}$. This is true if, and only if, $Q^T Q = I = Q Q^T$.

Example: the permutation matrices P in LU factorization with partial pivoting are orthogonal.

Proposition. The product of orthogonal matrices is an orthogonal matrix.

Proof. If S and T are orthogonal, $(ST)^T = T^T S^T$ so

$$(ST)^T(ST) = T^T S^T ST = T^T (S^T S) T = T^T T = I.$$

Definition: The **scalar (dot)(inner) product** of two vectors

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

in $\mathbb{R}^n$ is

$$x^T y = y^T x = \sum_{i=1}^n x_i y_i \in \mathbb{R}$$

Definition: Two vectors $x, y \in \mathbb{R}^n$ are **orthogonal** if $x^T y = 0$. A set of vectors $\{u_1, u_2, \dots, u_r\}$ is an **orthogonal set** if $u_i^T u_j = 0$ for all $i, j \in \{1, 2, \dots, r\}$ such that $i \neq j$.

Lemma. The columns of an orthogonal matrix Q form an orthogonal set, which is moreover an orthonormal basis for $\mathbb{R}^n$.

Proof. Suppose that $Q = [q_1 \ q_2 \ \dots \ q_n]$, i.e., q_j is the j th column of Q . Then

$$Q^T Q = I = \begin{bmatrix} q_1^T \\ q_2^T \\ \dots \\ q_n^T \end{bmatrix} [q_1 \ q_2 \ \dots \ q_n] = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

Comparing the (i, j) th entries yields

$$q_i^T q_j = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases}$$

Note that the columns of an orthogonal matrix are of length 1 as $q_i^T q_i = 1$, so they form an orthonormal set $\iff$ they are linearly independent (check!) $\implies$ they form an

orthonormal basis for $\mathbb{R}^n$ as there are n of them. □

Lemma. If $u \in \mathbb{R}^n$, P is n -by- n orthogonal and $v = Pu$, then $u^T u = v^T v$.

Proof. See problem sheet.

Definition: The **outer product** of two vectors x and $y \in \mathbb{R}^n$ is

$$xy^T = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_n \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_n \end{bmatrix},$$

an n -by- n matrix (notation: $xy^T \in \mathbb{R}^{n \times n}$). More usefully, if $z \in \mathbb{R}^n$, then

$$(xy^T)z = xy^T z = x(y^T z) = \left(\sum_{i=1}^n y_i z_i \right) x.$$

Definition: For $w \in \mathbb{R}^n$, $w \neq 0$, the **Householder** matrix $H(w) \in \mathbb{R}^{n \times n}$ is the matrix

$$H(w) = I - \frac{2}{w^T w} ww^T.$$

Proposition. $H(w)$ is an orthogonal matrix.

Proof.

$$\begin{aligned} H(w)H(w)^T &= \left(I - \frac{2}{w^T w} ww^T \right) \left(I - \frac{2}{w^T w} ww^T \right) \\ &= I - \frac{4}{w^T w} ww^T + \frac{4}{(w^T w)^2} w(w^T w)w^T \\ &= I. \end{aligned} \quad \square$$

Lemma. Given $u \in \mathbb{R}^n$, there exists a $w \in \mathbb{R}^n$ such that

$$H(w)u = \begin{bmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{bmatrix} \equiv v,$$

say, where $\alpha = \pm \sqrt{u^T u}$.

Remark: Since $H(w)$ is an orthogonal matrix for any $w \in \mathbb{R}$, $w \neq 0$, it is necessary for the validity of the equality $H(w)u = v$ that $v^T v = u^T u$, i.e., $\alpha^2 = u^T u$; hence our choice of $\alpha = \pm \sqrt{u^T u}$.

Proof. Take $w = \gamma(u - v)$, where $\gamma \neq 0$. Recall that $u^T u = v^T v$. Thus,

$$\begin{aligned} w^T w &= \gamma^2(u - v)^T(u - v) = \gamma^2(u^T u - 2u^T v + v^T v) \\ &= \gamma^2(u^T u - 2u^T v + u^T u) = 2\gamma u^T(\gamma(u - v)) \\ &= 2\gamma w^T u. \end{aligned}$$

So

$$H(w)u = \left(I - \frac{2}{w^T w} w w^T \right) u = u - \frac{2w^T u}{w^T w} w = u - \frac{1}{\gamma} w = u - (u - v) = v.$$

□

☐

Now if u is the first column of the n -by- n matrix A ,

$$H(w)A = \left[\begin{array}{c|ccc} \alpha & \times & \cdots & \times \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array} \right], \text{ where } \times = \text{general entry.}$$

Similarly for B , we can find $\hat{w} \in \mathbb{R}^{n-1}$ such that

$$H(\hat{w})B = \left[\begin{array}{c|ccc} \beta & \times & \cdots & \times \\ \hline 0 & & & \\ \vdots & & C & \\ 0 & & & \end{array} \right]$$

and then

$$\left[\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ 0 & \hline & H(\hat{w}) & & \\ \vdots & & & \\ 0 & & & \end{array} \right] H(w)A = \left[\begin{array}{cc|cccc} \alpha & \times & \times & \dots & \times \\ 0 & \beta & \times & \dots & \times \\ 0 & 0 & \hline & & & & \\ 0 & 0 & & C & \\ \vdots & \vdots & & & \\ 0 & 0 & & & \end{array} \right].$$

Note

$$\begin{bmatrix} 1 & 0 \\ 0 & H(\hat{w}) \end{bmatrix} = H(w_2), \quad \text{where } w_2 = \begin{bmatrix} 0 \\ \hat{w} \end{bmatrix}.$$

Thus if we continue in this manner for the $n - 1$ steps, we obtain

$$\underbrace{H(w_{n-1}) \cdots H(w_3)H(w_2)H(w)}_{Q^T} A = \begin{bmatrix} \alpha & \times & \cdots & \times \\ 0 & \beta & \cdots & \times \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \gamma \end{bmatrix} = (\top).$$

The matrix Q^T is orthogonal as it is the product of orthogonal (Householder) matrices, so we have constructively proved that

Theorem. Given any square matrix A , there exists an orthogonal matrix Q and an upper triangular matrix R such that

$$A = QR$$

Notes: 1. This could also be established using the Gram–Schmidt Process.

2. If u is already of the form $(\alpha, 0, \dots, 0)^T$, we just take $H = I$.

3. It is not necessary that A is square: if $A \in \mathbb{R}^{m \times n}$, then we need the product of (a) $m-1$ Householder matrices if $m \leq n \implies$

$$(\boxed{}) = A = QR = (\boxed{})(\overline{})$$

or (b) n Householder matrices if $m > n \implies$

$$(\boxed{}) = A = QR = (\boxed{})(\overline{}).$$

Another useful family of orthogonal matrices are the **Givens rotation** matrices:

$$J(i, j, \theta) = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & c & s & \\ & & -s & c & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix} \quad \begin{array}{l} \leftarrow i\text{th row} \\ \leftarrow j\text{th row} \end{array}$$

$\begin{array}{cc} \uparrow & \uparrow \\ i & j \end{array}$

where $c = \cos \theta$ and $s = \sin \theta$.

Exercise: Prove that $J(i, j, \theta)J(i, j, \theta)^T = I$ —obvious though, since the columns form an orthonormal basis.

Note that if $x = (x_1, x_2, \dots, x_n)^T$ and $y = J(i, j, \theta)x$, then

$$\begin{aligned} y_k &= x_k \text{ for } k \neq i, j \\ y_i &= cx_i + sx_j \\ y_j &= -sx_i + cx_j \end{aligned}$$

and so we can ensure that $y_j = 0$ by choosing $x_i \sin \theta = x_j \cos \theta$, i.e.,

$$\tan \theta = \frac{x_j}{x_i} \text{ or equivalently } s = \frac{x_j}{\sqrt{x_i^2 + x_j^2}} \text{ and } c = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}. \quad (1)$$

Thus, unlike the Householder matrices, which introduce lots of zeros by pre-multiplication, the Givens matrices introduce a single zero in a chosen position by pre-multiplication. Since (1) can always be satisfied, we only ever think of Givens matrices $J(i, j)$ for a specific vector or column with the angle chosen to make a zero in the j th position, e.g., $J(1, 2)x$ tacitly implies that we choose $\theta = \tan^{-1} x_2/x_1$ so that the second entry of $J(1, 2)x$ is zero. Similarly, for a matrix $A \in \mathbb{R}^{m \times n}$, $J(i, j)A := J(i, j, \theta)A$, where $\theta = \tan^{-1} a_{ji}/a_{ii}$, i.e., it is the i th column of A that is used to define θ so that $(J(i, j)A)_{ji} = 0$.

We shall return to these in a later lecture.