Numerical Analysis

Raphael Hauser
with thanks to Endre Süli

Oxford Mathematical Institute

HT 2019

QR Factorization

Definition

A square real matrix Q is **orthogonal** if $Q^{\mathrm{T}} = Q^{-1}$.

QR Factorization

Definition

A square real matrix Q is **orthogonal** if $Q^{\mathrm{T}} = Q^{-1}$.

Q is orthogonal if, and only if, $Q^{\mathrm{T}}Q = I = QQ^{\mathrm{T}}$.

QR Factorization

Definition

A square real matrix Q is **orthogonal** if $Q^{\mathrm{T}} = Q^{-1}$.

Q is orthogonal if, and only if, $Q^{\mathrm{T}}Q = I = QQ^{\mathrm{T}}$.

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

Proposition

The product of orthogonal matrices is an orthogonal matrix.

Proposition

The product of orthogonal matrices is an orthogonal matrix.

Proof. Note that $(ST)^{\mathrm{T}} = T^{\mathrm{T}} S^{\mathrm{T}}$. Further, as S and T are orthogonal matrices,

$$(ST)^{\mathrm{T}}(ST) = T^{\mathrm{T}} S^{\mathrm{T}} ST = T^{\mathrm{T}}(S^{\mathrm{T}} S)T = T^{\mathrm{T}} T = I. \qquad \square$$

Definition

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

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

in $\mathbb{R}^n$ is

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

Definition

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

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

in $\mathbb{R}^n$ is

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

Definition

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

Lemma

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

Lemma

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

Proof. Suppose that $Q = [q_1 \;\; q_2 \;\vdots\; q_n]$, i.e., q_j is the jth column of Q. Then

$$Q^{\mathrm{T}}Q = I = \begin{pmatrix} q_1^{\mathrm{T}} \\ q_2^{\mathrm{T}} \\ \cdots \\ q_n^{\mathrm{T}} \end{pmatrix} [q_1 \;\; q_2 \;\vdots\; q_n] = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Lemma

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

Proof. Suppose that $Q = [q_1 \;\; q_2 \;\; \vdots \;\; q_n]$, i.e., q_j is the jth column of Q. Then

$$Q^{\mathrm{T}}Q = I = \begin{pmatrix} q_1^{\mathrm{T}} \\ q_2^{\mathrm{T}} \\ \cdots \\ q_n^{\mathrm{T}} \end{pmatrix} [q_1 \;\; q_2 \;\; \vdots \;\; q_n] = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Comparing the (i,j)th entries yields

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

Note that the columns of an orthogonal matrix are of length 1 as $q_i^{\mathrm{T}} q_i = 1$, so they form an orthonormal set $\implies$ they are linearly independent (check!) $\implies$ they form an orthonormal basis for $\mathbb{R}^n$ as there are n of them. $\square$

Lemma

If $u \in \mathbb{R}^n$, Q is an $n \times n$ orthogonal matrix and $v = Qu$, then $u^{\mathrm{T}} u = v^{\mathrm{T}} v$.

Proof. See problem sheet. □

Definition

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

$$xy^{\mathrm{T}} = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_n \end{pmatrix},$$

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

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

Definition

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

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

Definition

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

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

Proposition

$H(w)$ is an orthogonal matrix.

Definition

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

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

Proposition

$H(w)$ is an orthogonal matrix.

Proof.

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

□

Lemma

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

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

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

Proof. Take $w = \gamma(u - v)$, where $\gamma \neq 0$.

Proof. Take $w = \gamma(u - v)$, where $\gamma \neq 0$. Recall that since $H(w)$ is orthogonal, $u^{\mathrm{T}}u = v^{\mathrm{T}}v$.

Proof. Take $w = \gamma(u - v)$, where $\gamma \neq 0$. Recall that since $H(w)$ is orthogonal, $u^{\mathrm{T}}u = v^{\mathrm{T}}v$. Then

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

Proof. Take $w = \gamma(u - v)$, where $\gamma \neq 0$. Recall that since $H(w)$ is orthogonal, $u^{\mathrm{T}}u = v^{\mathrm{T}}v$. Then

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

So

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

□

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

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

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

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

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

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

Thus,

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

Thus,

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

Note that

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

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

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

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

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

The matrix Q^{T} is orthogonal as it is the product of orthogonal (Householder) matrices.

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

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

The matrix Q^{T} is orthogonal as it is the product of orthogonal (Householder) matrices.

So we have constructively proved the following result.

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

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

The matrix Q^{T} is orthogonal as it is the product of orthogonal (Householder) matrices.

So we have constructively proved the following result.

Theorem

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

$$A = QR$$

Notes:

Notes:

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

Notes:

1. This could also be established using the Gram–Schmidt Process.
2. If u is already of the form $(\alpha,\ 0,\ \cdots,\ 0)^{\mathrm{T}}$, we just take $H = I$.

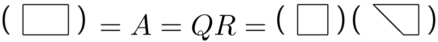
Notes:

1. This could also be established using the Gram–Schmidt Process.
2. If u is already of the form $(\alpha,\ 0,\ \cdots,\ 0)^{\mathrm{T}}$, 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 $m \leq n \Longrightarrow$

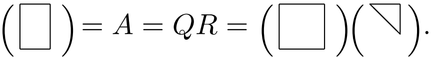
$$(\ \square\ ) = A = QR = (\ \square\ )(\ \diagdown\!\rceil\ )$$

Notes:

1. This could also be established using the Gram–Schmidt Process.
2. If u is already of the form $(\alpha,\ 0,\ \cdots,\ 0)^{\mathrm{T}}$, 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 $m \leq n \implies$



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



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

$$J(i,j,\theta) = \begin{pmatrix} 1 & & & & & & \\ & \cdot & & & & & \\ & & c & & s & & \\ & & & \cdot & & & \\ & & -s & & c & & \\ & & & & & \cdot & \\ & & & & & & 1 \end{pmatrix} \begin{matrix} \\ \\ \leftarrow \;\; i\text{th row} \\ \\ \leftarrow \;\; j\text{th row} \\ \\ \\ \end{matrix}$$

$$\begin{matrix} \uparrow & & \uparrow \\ i & & j \end{matrix}$$

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

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

Exercise: Prove that $J(i,j,\theta)J(i,j,\theta)^{\mathrm{T}} = I$— obvious though, since the columns form an orthonormal basis.
Note that if $x = (x_1,\ x_2,\ \ldots, x_n)^{\mathrm{T}}$ 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.,

$$\tan\theta = \frac{x_j}{x_i} \text{ or equivalently } s = \frac{x_j}{\sqrt{x_i^2 + x_j^2}} \text{ and } c = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}. \quad (6.1)$$

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.

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 (6.1) 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 jth 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.

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 (6.1) 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 jth 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 ith column of A, which is used to define θ so that $(J(i,j)A)_{ji} = 0$.

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 (6.1) 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 jth 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 ith column of A, which is used to define θ so that $(J(i,j)A)_{ji} = 0$. We shall return to these in a later lecture.