Orthogonality
Matrix Computations — CPSC 5006 E

Julien Dompierre

Department of Mathematics and Computer Science
Laurentian University

Sudbury, November 7, 2009

Outline

Orthogonality

- Orthogonality.
- Gram-Schmidt.
- QR decomposition.
- Section 5.2 of the textbook.

Orthogonal Vectors

Definition

We say that two non zero vectors are **orthogonal** if the angle between them is a right angle.

Theorem

Two non zero vectors u and v are orthogonal if and only if their scalar product $u^T v = (u, v) = u \cdot v = 0$.

Corollary

The zero vector is orthogonal to any vector.

Orthogonal Set

Definition

A set of vectors $\{u_1, u_2, ..., u_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $u_i \cdot u_j = 0$ for all $i \neq j$.

Note. There are ${}_pC_2 = p!/((p-2)!2!)$ combinations to verify if you want to prove that a set of p vectors is an orthogonal set.

Example of an Orthogonal Set

Show that $\{u_1, u_2, u_3\}$ is an orthogonal set, where

$$u_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad u_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}.$$

There are ${}_3C_2 = 3!/((3-2)!\,2!) = 3$ combinations to verify, namely $\{u_1, u_2\}$, $\{u_1, u_3\}$ and $\{u_2, u_3\}$.

$$\begin{aligned} u_1 \cdot u_2 &= u_1^T u_2 = 3(-1) + 1(2) + 1(1) = 0 \\ u_1 \cdot u_3 &= u_1^T u_3 = 3(-1/2) + 1(-2) + 1(7/2) = 0 \\ u_2 \cdot u_3 &= u_2^T u_3 = -1(-1/2) + 2(-2) + 1(7/2) = 0 \end{aligned}$$

Review: Linear Combination

Definition

Let $v_1, v_2, \ldots, v_p$ be vectors in a vector space V. We say that y, a vector in V, is a **linear combination** of $v_1, v_2, \ldots, v_p$ if there exist scalars $c_1, c_2, \ldots, c_p$ such that y can be written as

$$y = c_1 v_1 + c_2 v_2 + \cdots + c_p v_p.$$

The scalars $c_1, c_2, \ldots, c_p$ are called the **weights**.

Review: Linear Dependence and Independence

Definition

(a) The set of vectors $\{v_1, v_2, \ldots, v_p\}$ in a vector space V is said to be **linearly dependent** if there exist scalars $c_1, c_2, \ldots, c_p$ not all zero such that

$$c_1 v_1 + c_2 v_2 + \cdots + c_p v_p = 0.$$

(b) The set of vectors $\{v_1, v_2, \ldots, v_p\}$ in a vector space V is said to be **linearly independent** if

$$c_1 v_1 + c_2 v_2 + \cdots + c_p v_p = 0$$

can only be satisfied when

$$c_1 = 0, c_2 = 0, \ldots, c_p = 0.$$

Orthogonal Set and Linear Independence

Theorem

If $S = \{u_1, u_2, ..., u_p\}$ is an orthogonal set of non zero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

Proof.

If $0 = c_1 u_1 + \cdots + c_p u_p$ for the scalars $c_1, ..., c_p$, then

$$\begin{aligned}
0 &= 0 \cdot u_1 = (c_1 u_1 + c_2 u_2 + \cdots + c_p u_p) \cdot u_1 \\
&= c_1(u_1 \cdot u_1) + c_2(u_2 \cdot u_1) + \cdots + c_p(u_p \cdot u_1) \\
&= c_1(u_1 \cdot u_1)
\end{aligned}$$

because u_1 is orthogonal to $u_2, ..., u_p$. Since u_1 is non zero, $u_1 \cdot u_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, ..., c_p$ must be zero. Thus S is linearly independent. $\square$

Orthogonal Basis

Definition

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

Orthogonal Basis and Linear Combination

Theorem

Let $\{u_1, u_2, \ldots, u_p\}$ *be an orthogonal basis for a subspace* W *of* $\mathbb{R}^n$*. For each* y *in* W*, the weights in linear combination*

$$y = c_1 u_1 + c_2 u_2 + \cdots + c_p u_p$$

are given by

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j}, \quad \text{for } j = 1, \ldots, p.$$

Proof.

The orthogonality of $\{u_1, u_2, \ldots, u_p\}$ shows that

$$y \cdot u_j = (c_1 u_1 + c_2 u_2 + \cdots + c_p u_p) \cdot u_j = c_j (u_j \cdot u_j).$$

Since $u_j \cdot u_j$ is not zero, the equation above can be solved for c_j. □

Example: Linear Combination for Non Orthogonal Basis

Let $S = \{u_1, u_2, u_3\}$ be an orthogonal basis for $\mathbb{R}^3$ as in the previous example. The vector $y = [6, 1, -8]^T$ can be expressed as a linear combination of the vectors in S.

$$c_1 \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -1 & -1/2 \\ 1 & 2 & -2 \\ 1 & 1 & 7/2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$$

$$\left[\begin{array}{ccc|c} 3 & -1 & -1/2 & 6 \\ 1 & 2 & -2 & 1 \\ 1 & 1 & 7/2 & -8 \end{array}\right] \approx \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

What is the cost?

Example: Linear Combination for Orthogonal Basis

Let $S = \{u_1, u_2, u_3\}$ be an orthogonal basis for $\mathbb{R}^3$ as in the previous example. The vector $y = [6, 1, -8]^T$ can be expressed as a linear combination of the vectors in S.

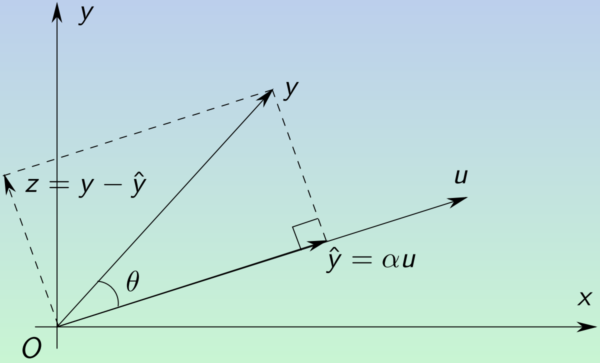
$$\begin{array}{ccc} y \cdot u_1 = 11, & y \cdot u_2 = -12, & y \cdot u_3 = -33, \\ u_1 \cdot u_1 = 11, & u_2 \cdot u_2 = 6, & u_3 \cdot u_3 = 33/2. \end{array}$$

Using orthogonality of the basis, we get

$$\begin{aligned} y &= \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 + \frac{y \cdot u_3}{u_3 \cdot u_3} u_3 \\ &= \frac{11}{11} u_1 + \frac{-12}{6} u_2 + \frac{-33}{33/2} u_3 \\ &= u_1 - 2u_2 - 2u_3 \end{aligned}$$

What is the cost?

Let u and y be vectors in $\mathbb{R}^n$ with angle θ between them. The vector $\hat{y}$ tells us "how much" of y is pointing in the direction of u. We call $\hat{y} = \alpha u$ the **orthogonal projection of y onto u**. The vector $z = y - \hat{y}$ is the **component of y orthogonal to u**.

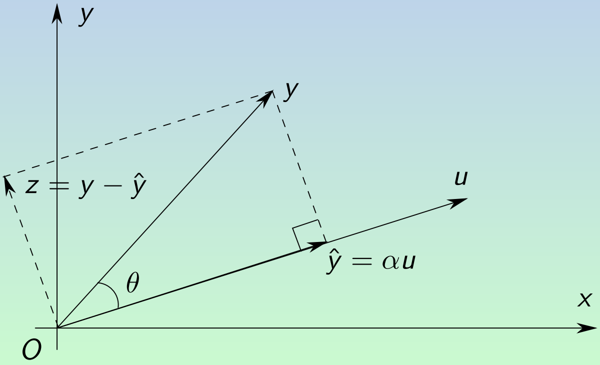


Projection of One Vector onto Another Vector

Definition

The projection of a vector y onto a non zero vector u in $\mathbb{R}^n$ is denoted $\text{proj}_u y$ and is defined by

$$\hat{y} = \text{proj}_u y = \left(\frac{y \cdot u}{u \cdot u}\right) u$$



Orthonormal Set

Definition

A set $\{u_1, u_2, ..., u_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. if W is the subspace spanned by such a set, then $\{u_1, u_2, ..., u_p\}$ is an orthonormal basis for W.

The simplest example of an orthonormal set is the standard basis $\{e_1, e_2, ..., e_n\}$ for $\mathbb{R}^n$.

Orthonormal Columns of a Matrix

Theorem

An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

Proof.

See next slide. $\square$

Definition

An **orthogonal matrix** (or **unitary** matrix) is a square invertible matrix U such that $U^{-1} = U^T$. Such a matrix has orthonormal columns (and so a better name might be *orthonormal matrix*). Such a matrix must have orthonormal rows, too.

Proof of Theorem of the Previous Slide

Proof.

To simplify the proof, suppose that U has only three columns, each a vector in $\mathbb{R}^m$. The proof in the general case is essentially the same. Let $U = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}$ and compute

$$U^T U = \begin{bmatrix} u_1^T \\ u_2^T \\ u_3^T \end{bmatrix} \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{bmatrix}$$

The entries in the matrix are inner products, using the transpose notation. The columns of U are orthogonal if and only if

$$u_1^T u_2 = u_2^T u_1 = 0, \quad u_1^T u_3 = u_3^T u_1 = 0, \quad u_2^T u_3 = u_3^T u_2 = 0.$$

The columns of U all have unit length if and only if

$$u_1^T u_1 = 1, \quad u_2^T u_2 = 1, \quad u_3^T u_3 = 1.$$

The Problem

Notation: $V = [v_1, \ldots, v_n]$ is the matrix with column-vectors $v_1, \ldots, v_n$.

Problem: Given a system $X = [x_1, \ldots, x_n]$, compute an orthonormal system $Q = [q_1, \ldots, q_n]$ which spans the same space as the columns of X.

Theorem

Given a basis $\{x_1, x_2, ..., x_p\}$ *for a subspace* W *of* $\mathbb{R}^n$, *define*

$$\begin{aligned}
q_1 &= x_1, \\
q_2 &= x_2 - \frac{x_2 \cdot q_1}{q_1 \cdot q_1} q_1, \\
q_3 &= x_3 - \frac{x_3 \cdot q_1}{q_1 \cdot q_1} q_1 - \frac{x_3 \cdot q_2}{q_2 \cdot q_2} q_2, \\
\vdots &= \vdots \\
q_p &= x_p - \frac{x_p \cdot q_1}{q_1 \cdot q_1} q_1 - \frac{x_p \cdot q_2}{q_2 \cdot q_2} q_2 - \cdots - \frac{x_p \cdot q_{p-1}}{q_{p-1} \cdot q_{p-1}} q_{p-1}.
\end{aligned}$$

Then $\{q_1, q_2, ..., q_p\}$ *is an* **orthogonal** *basis for* W. *In addition*

$$span\{q_1, q_2, ..., q_p\} = span\{x_1, x_2, ..., x_p\} \quad for \quad 1 \leq k \leq p.$$

The Gram-Schmidt Process

Theorem

Given a basis $\{x_1, x_2, \ldots, x_p\}$ *for a subspace* W *of* $\mathbb{R}^n$*, define*

$$\begin{aligned}
q_1 &= x_1, \quad q_1 = q_1/\|q_1\| \\
q_2 &= x_2 - (x_2 \cdot q_1)q_1, \quad q_2 = q_2/\|q_2\| \\
q_3 &= x_3 - (x_3 \cdot q_1)q_1 - (x_3 \cdot q_2)q_2, \quad q_3 = q_3/\|q_3\| \\
\vdots &= \vdots \\
q_p &= x_p - (x_p \cdot q_1)q_1 - (x_p \cdot q_2)q_2 - \cdots - (x_p \cdot q_{p-1})q_{p-1}, \\
&\quad\quad q_p = q_p/\|q_p\|
\end{aligned}$$

Then $\{q_1, q_2, \ldots, q_p\}$ *is an* **orthonormal** *basis for* W*. In addition*

$$span\{q_1, q_2, \ldots, q_p\} = span\{x_1, x_2, \ldots, x_p\} \quad for \quad 1 \leq k \leq p.$$

The Gram-Schmidt Process

Theorem

Given a basis $\{x_1, x_2, \ldots, x_p\}$ *for a subspace* W *of* $\mathbb{R}^n$*, define*

$$\begin{aligned}
q_1 &= x_1, \\
q_2 &= x_2 - proj_{q_1}(x_2) \\
q_3 &= x_3 - proj_{q_1}(x_3) - proj_{q_2}(x_3), \\
\vdots &= \vdots \\
q_p &= x_p - proj_{q_1}(x_p) - proj_{q_2}(x_p) - \cdots - proj_{q_{p-1}}(x_p).
\end{aligned}$$

Then $\{q_1, q_2, \ldots, q_p\}$ *is an orthogonal basis for* W*. In addition*

$$span\{q_1, q_2, \ldots, q_p\} = span\{x_1, x_2, \ldots, x_p\} \quad for \quad 1 \leq k \leq p.$$

The Gram-Schmidt Process

Theorem

Given a basis $\{x_1, x_2, \dots, x_p\}$ *for a subspace* W *of* $\mathbb{R}^n$*, define*

$$\begin{aligned} q_1 &= x_1, \\ q_2 &= x_2 - proj_{W_1}(x_2), \quad W_1 = span\{q_1\} \\ q_3 &= x_3 - proj_{W_2}(x_3), \quad W_2 = span\{q_1, q_2\} \\ \vdots &= \vdots \\ q_p &= x_p - proj_{W_{p-1}}(x_p), \quad W_{p-1} = span\{q_1, q_2, \dots, q_{p-1}\}. \end{aligned}$$

Then $\{q_1, q_2, \dots, q_p\}$ *is an orthogonal basis for* W*. In addition*

$$span\{q_1, q_2, \dots, q_p\} = span\{x_1, x_2, \dots, x_p\} \quad \text{for} \quad 1 \leq k \leq p.$$

Classical Gram-Schmidt (p. 231)

Algorithm 1 Classical Gram-Schmidt. Given $A = [\, a_1 \ \cdots \ a_n \,] \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = n$, the following algorithm computes the factorization $A = QR$ where $Q = [\, q_1 \ \cdots \ q_n \,] \in \mathbb{R}^{m\times n}$ has orthonormal columns and $R \in \mathbb{R}^{n\times n}$ is upper triangular.

```
for j = 1 : n
    Set q = a_j
    for i = 1 : j − 1
        Compute r_ij = q · q_i
    end
    for i = 1 : j − 1
        Compute q = q − r_ij q_i
    end
    Compute r_jj = ||q||_2
    q_j = q / r_jj
end
```

Example: Orthonormalize the system of vectors

$$X = [x_1, x_2, x_3] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 4 \end{bmatrix}$$

Answer:

$$q_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}; \quad \hat{q}_2 = x_2 - (x_2, q_1)q_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - 1 \times \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$$

$$\hat{q}_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}; \quad q_2 = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

Example

$$\hat{q}_3 = x_3 - (x_3, q_1)q_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 4 \end{bmatrix} - 2 \times \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ -2 \\ 3 \end{bmatrix}$$

$$\hat{q}_3 = \hat{q}_3 - (\hat{q}_3, q_2)q_2 = \begin{bmatrix} 0 \\ -1 \\ -2 \\ 3 \end{bmatrix} - (-1) \times \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -2.5 \\ 2.5 \end{bmatrix}$$

$$\|\hat{q}_3\|_2 = \sqrt{13} \rightarrow q_3 = \frac{\hat{q}_3}{\|\hat{q}_3\|_2} = \frac{1}{\sqrt{13}} \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -2.5 \\ 2.5 \end{bmatrix} \approx \begin{bmatrix} 0.1387 \\ -0.1387 \\ -0.6934 \\ 0.6934 \end{bmatrix}$$

Lines 4, 7 and 9 show that

$$x_j = r_{1j} q_1 + r_{2j} q_2 + \ldots + r_{jj} q_j$$

If $X = [x_1, x_2, \ldots, x_n]$, $Q = [q_1, q_2, \ldots, q_n]$, and if R is the $n \times n$ upper triangular matrix

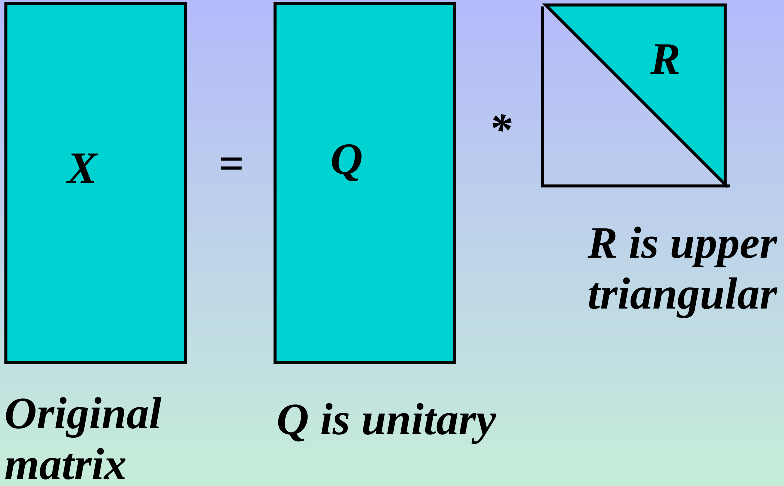
$$R = \{r_{ij}\}_{i,j=1,\ldots,n}$$

then the above relation can be written as

$$X = QR$$

R is upper triangular, Q is unitary. This is called the **QR factorization** of X.

QR Decomposition



Another Decomposition:

A matrix X, with linearly independent columns, is the product of a unitary matrix Q and a upper triangular matrix R.

Classical Gram-Schmidt – Matrix Form (p. 231)

Algorithm 2 Classical Gram-Schmidt. Given $A = [\, a_1 \;\; \cdots \;\; a_n \,] \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = n$, the following algorithm computes the factorization $A = QR$ where $Q = [\, q_1 \;\; \cdots \;\; q_n \,] \in \mathbb{R}^{m\times n}$ has orthonormal columns and $R \in \mathbb{R}^{n\times n}$ is upper triangular.

$R(1,1) = \|A(:,1)\|_2$
$Q(:,1) = A(:.1)/R(1,1)$
for $j = 2 : n$
  $R(1:j-1,j) = Q(:,1:j-1)^T A(:,j)$
  $z = A(:,j) - Q(:,1:j-1)\, R(1:j-1,j)$
  $R(j,j) = \|z\|_2$
  $Q(:,j) = z/R(j,j)$
end

Better algorithm: Modified Gram-Schmidt

Algorithm 3 Modified Gram-Schmidt. Given $A = [\, a_1 \ \cdots \ a_n \,] \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = n$, the following algorithm computes the factorization $A = QR$ where $Q = [\, q_1 \ \cdots \ q_n \,] \in \mathbb{R}^{m\times n}$ has orthonormal columns and $R \in \mathbb{R}^{n\times n}$ is upper triangular.

```
for j = 1, ..., n
    Define q̂ := a_j
    for i = 1, ..., j − 1
        r_ij := (q̂, q_i)
        q̂ := q̂ − r_ij q_i
    end
    Compute r_jj := ‖q̂‖_2
    q_j := q̂ / r_jj
end
```

- **for** $j = 1, \ldots, n$
  - Define $\hat{q} := a_j$
  - **for** $i = 1, \ldots, j-1$
    - $r_{ij} := (\hat{q}, q_i)$
    - $\hat{q} := \hat{q} - r_{ij} q_i$
  - **end**
  - Compute $r_{jj} := \|\hat{q}\|_2$
  - $q_j := \hat{q}/r_{jj}$
- **end**

Only difference: inner product uses the accumulated subsum instead of original $\hat{q}$.

Error Analysis of Modified Gram-Schmidt

Modified Gram-Schmidt algorithm is much more stable than classical Gram-Schmidt in general. (A few examples easily show this)
Suppose MGS is applied to A yielding computed matrices $\hat{Q}$ and $\hat{R}$. Then there are constants c_i (depending on (m, n)) such that

$$A + E_1 = \hat{Q}\hat{R} \qquad \|E_1\|_2 \leq c_1 \; u \; \|A\|_2$$

$$\|\hat{Q}^T \hat{Q} - I\|_2 \leq c_2 \; u \; \kappa_2(A) + O((u\kappa_2(A))^2)$$

for a certain perturbation matrix E_1, and there exists an orthonormal matrix Q such that

$$A + E_2 = Q\hat{R} \qquad \|E_2(:,j)\|_2 \leq c_3 u \|A(:,j)\|_2$$

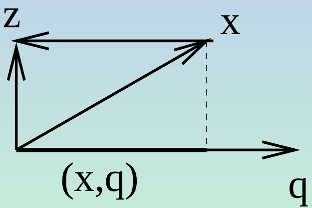
for a certain perturbation matrix E_2. (u is the machine epsilon).

Orthogonalization

The operations in lines 4 and 5 can be written as

$$\hat{q} := ORTH(\hat{q}, q_i)$$

Where $ORTH(x, q)$ denotes the operation of orthogonalizing a vector x against a unit vector q.



Result of $z = ORTH(x, q)$

Unitary Matrix

For this example: compute $Q^T Q$.

Result is the identity matrix.

Recall: For any unitary matrix Q, we have

$$Q^T Q = I$$

(In complex case: $Q^H Q = I$).

Consequence: For an $n \times n$ unitary matrix $\boxed{Q^{-1} = Q^T}$

Application: another method for solving linear systems.

$$Ax = b$$

A is an $n \times n$ non singular matrix. Compute its QR factorization.

Multiply both sides by $Q^T \rightarrow Q^T QRx = Q^T b \rightarrow$

$$Rx = Q^T b$$

<u>Method:</u>

Compute the QR factorization of A, $A = QR$.

Solve the upper triangular system $Rx = Q^T b$.
(Cost??)