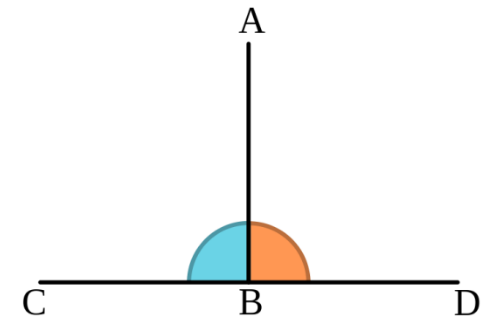
3 Orthogonality （正交性）



3.4

ORTHOGONAL BASES AND GRAM-SCHMIDT （正交基; Gram-Schmidt正交化）

- Orthogonal Matrices
- Rectangular Matrices with Orthonormal Columns
- Gram-Schmidt Orthogonalization
- QR factorization

For $\mathbf{R}^n$, the standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ satisfies

$\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{e}_j = 1$ or 0 depending on $i = j$ or $i \neq j$, respectively.

They are called orthonormal.

In general, the vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ are called **orthonormal (标准正交, 单位正交)** if

$$\boldsymbol{q}_i^T \boldsymbol{q}_j = \begin{cases} 0 & \text{whenever } i \neq j, \\ 1 & \text{whenever } i = j. \end{cases}$$

giving the orthogonality; (两两正交)

giving the normalization. (单位长度)

We will see:

1. What happens with a matrix with orthonormal columns – square (*orthogonal matrix*, 正交矩阵) or rectangular;
2. A subspace always has an orthonormal basis, and it can be constructed in a simple way out of any basis whatsoever.

What method can *convert a skewed set of axes into a perpendicular set* – *Gram-Schmidt orthogonalization* (Gram-Schmidt 正交化).

I. Orthogonal Matrices (正交矩阵)

Definition 1 If $\boldsymbol{Q}$ (square or rectangular) has orthonormal columns, then $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$:

$$\begin{bmatrix} - & q_1^{\mathrm{T}} & - \\ - & q_2^{\mathrm{T}} & - \\ & \vdots & \\ - & q_n^{\mathrm{T}} & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdot & 0 \\ 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 1 \end{bmatrix} = I.$$

An orthogonal matrix is a square matrix with orthonormal columns.

(正交矩阵是方阵, 它的列是标准正交的向量组)

Then $\boldsymbol{Q}^{\mathrm{T}}$ is $\boldsymbol{Q}^{-1}$.

(For orthogonal matrices, ***the transpose is the inverse***.)

Notes:

An orthogonal matrix is square. (正交矩阵特指满足上述条件的方阵)

$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$ even if $\boldsymbol{Q}$ is rectangular. But then $\boldsymbol{Q}^{\mathrm{T}}$ is only a left-inverse.

Example 1

$$\boldsymbol{Q} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \qquad \boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Rotation

$\boldsymbol{Q}$ is an orthogonal matrix.

The matrix $\boldsymbol{Q}^{\mathrm{T}}$ is just as much an orthogonal matrix as $\boldsymbol{Q}$.

Example 2

Any permutation matrix is an orthogonal matrix.

$$\boldsymbol{P}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{P}_1^{-1} = \boldsymbol{P}_1^{\mathrm{T}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

$\boldsymbol{P}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ is also an orthogonal matrix and takes $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ to $\begin{bmatrix} z \\ y \\ x \end{bmatrix}$.

$\boldsymbol{P}_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ reflects $\begin{bmatrix} x \\ y \end{bmatrix}$ into $\begin{bmatrix} y \\ x \end{bmatrix}$. ***Reflection***

Example 3 Decide whether the following matrices are orthogonal matrices:

$$\textbf{(1)}\ \begin{bmatrix} 1 & -1/2 & 1/3 \\ -1/2 & 1 & 1/2 \\ 1/3 & 1/2 & -1 \end{bmatrix}; \qquad \textbf{(2)}\ \begin{bmatrix} 1/9 & -8/9 & -4/9 \\ -8/9 & 1/9 & -4/9 \\ -4/9 & -4/9 & 7/9 \end{bmatrix};$$

$$\textbf{(3)}\ \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}.$$

Theorem 1 (*nice properties of orthogonal matrices*)

Multiplication by any $n \times n$ orthogonal matrix $\boldsymbol{Q}$ preserves lengths:

$$\|\boldsymbol{Qx}\| = \|\boldsymbol{x}\| \text{ for every vector } \boldsymbol{x}\in\mathbf{R}^n.$$

It also preserves inner products and angles.

(向量的内积、长度及向量间的夹角都保持不变)

Proof $\|\boldsymbol{Qx}\| = \|\boldsymbol{x}\|$ because $(\boldsymbol{Qx})^{\mathrm{T}}(\boldsymbol{Qx}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Qx} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$.

And since $(\boldsymbol{Qx})^{\mathrm{T}}(\boldsymbol{Qy}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Qy} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$, $\boldsymbol{Q}$ also preserves inner products and thus the angles.

Remark: This property is shared by *rotations* and *reflections*, and in fact by every orthogonal matrix. It is not shared by projections, which are not orthogonal (some projections are not even invertible).

Theorem 2 Let $\boldsymbol{A}$, $\boldsymbol{B}$ be $n\times n$ orthogonal matrices, then

(1) $\boldsymbol{A}^{-1}= \boldsymbol{A}^{\mathrm{T}}$;

(2) $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$;

(3) $\boldsymbol{A}^{-1}$ (i.e., $\boldsymbol{A}^{\mathrm{T}}$) is also an orthogonal matrix;

(4) $\boldsymbol{AB}$ is also an orthogonal matrix.

Proof

(1) $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}= \boldsymbol{I}$, then $\boldsymbol{A}^{-1}= \boldsymbol{A}^{\mathrm{T}}$. ($\boldsymbol{A}$ is full rank)

(2) $\boldsymbol{A}^{-1}= \boldsymbol{A}^{\mathrm{T}} \Rightarrow \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.

(3) $(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}= \boldsymbol{I}$, so $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}^{-1}$ is also an orthogonal matrix.

That is: ***The rows of a square matrix are orthonormal whenever the columns are.*** (正交矩阵的**行向量组**也是 $\mathbf{R}^n$ 的标准正交基.)

(4) $(\boldsymbol{AB})^{\mathrm{T}}(\boldsymbol{AB}) = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}})(\boldsymbol{AB}) =\boldsymbol{B}^{\mathrm{T}}\boldsymbol{I}\boldsymbol{B}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}=\boldsymbol{I}$,

so $\boldsymbol{AB}$ is also an orthogonal matrix.

If we have a basis, then any vector is a combination of the basis vectors. This is exceptionally simple for an *orthonormal basis*. The problem is *to find the coefficients of the basis vectors*:

Problem: *Write* $\boldsymbol{b}$ *as a combination:* $\boldsymbol{b} = x_1\boldsymbol{q}_1 + x_2\boldsymbol{q}_2 + \ldots + x_n\boldsymbol{q}_n$.

Try to find the coordinates $x_1, x_2, \ldots, x_n$.

Trick: Multiply both sides of the equation by $\boldsymbol{q}_i^{\mathrm{T}}$.

The only term *survives* is $x_i\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i$. The other terms *die* of orthogonality. That is,

$$\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{b} = x_i\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i = x_i.$$

Therefore,

$$\boldsymbol{b} = (\boldsymbol{q}_1^T\boldsymbol{b})\boldsymbol{q}_1 + (\boldsymbol{q}_2^T\boldsymbol{b})\boldsymbol{q}_2 + \ldots + (\boldsymbol{q}_n^T\boldsymbol{b})\boldsymbol{q}_n.$$

Geometrically, every vector $\boldsymbol{b}$ *is the sum of its one-dimensional projections onto the lines through the* $\boldsymbol{q}$*'s.*

(**$\boldsymbol{b}$** 在标准正交基 $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_n$ 下的坐标的第 i 个分量是 $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{b}$, 即$\boldsymbol{b}$ 在 $\boldsymbol{q}_i$ 上的投影.)

II. Rectangular Matrices with Orthonormal Columns (列向量标准正交的长方形矩阵)

Problem:

Solve $\boldsymbol{Qx} = \boldsymbol{b}$, where $\boldsymbol{Q}$ is an m by n matrix ($m > n$), with n orthonormal vectors $\boldsymbol{q}_i$ as the columns of $\boldsymbol{Q}$.

Then we cannot expect to solve $\boldsymbol{Qx} = \boldsymbol{b}$ exactly.

We solve it by least squares.

Tip: Orthonormal columns should make the problem simple.

The key is to notice that *we still have* $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$. (So $\boldsymbol{Q}^{\mathrm{T}}$ is still the ***left-inverse*** of $\boldsymbol{Q}$.)

The normal equations: $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$

Therefore, $\hat{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$.

Note: When $\boldsymbol{Q}$ is square, then $\hat{\boldsymbol{x}}$ is the exact solution;

If $\boldsymbol{Q}$ is rectangular, then $\hat{\boldsymbol{x}}$ is the least squares solution.

If $\boldsymbol{Q}$ has orthonormal columns, the least-squares problem becomes easy:

$\boldsymbol{Qx} = \boldsymbol{b}$	rectangular system (m>n) with no solution for most $\boldsymbol{b}$.
$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\hat{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$	normal equation for the best $\hat{\boldsymbol{x}}$ — in which $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$.
$\hat{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$	$\hat{\boldsymbol{x}}_i$ is $\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{b}$.
$\boldsymbol{p} = \boldsymbol{Q}\hat{\boldsymbol{x}}$	the projection of $\boldsymbol{b}$ is $(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_1 + (\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_2 + \ldots + (\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_n$.
$\boldsymbol{p} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$	the projection matrix is $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$.

($\boldsymbol{P} = \boldsymbol{Q}\left(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\right)^{-1}\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$,
is an $m \times m$ projection matrix;
while $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$ is an $n \times n$ identity matrix)

Example 4 The following case is simple but typical.

Suppose we project a point $\boldsymbol{b} = (x,y,z)^{\mathrm{T}}$ onto the x-y plane.

Its projection is $\boldsymbol{p} = (x,y,0)^{\mathrm{T}}$,

and this is the sum of the separate projections onto the x- and y-axes:

$$\boldsymbol{q}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \text{ and } (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_1 = \begin{bmatrix}x\\0\\0\end{bmatrix}; \quad \boldsymbol{q}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \text{ and } (\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_2 = \begin{bmatrix}0\\y\\0\end{bmatrix}.$$

The overall projection matrix is

$$\boldsymbol{P} = \boldsymbol{q}_1\boldsymbol{q}_1^{\mathrm{T}} + \boldsymbol{q}_2\boldsymbol{q}_2^{\mathrm{T}} = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}, \quad \text{and} \quad \boldsymbol{P}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}x\\y\\0\end{bmatrix}.$$

Remark: *Projection onto a plane = sum of projections onto orthonormal* $\boldsymbol{q}_1$ *and* $\boldsymbol{q}_2$.

III. Gram-Schmidt Orthogonalization (Gram-Schmidt正交化)

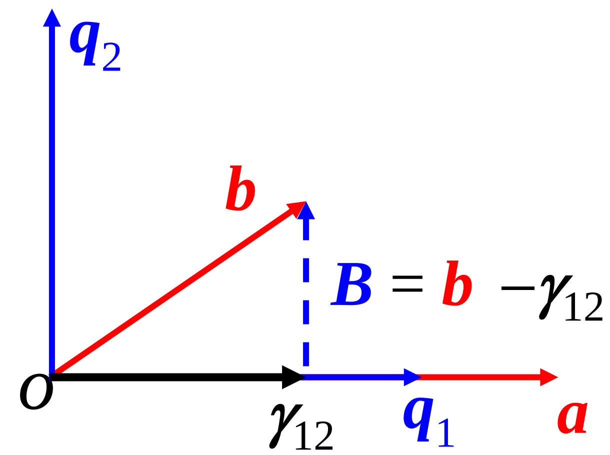
Convert independent vectors into orthonormal vectors

In $\mathbf{R}^n$, we try to make the independent vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ orthonormal.
(由$\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$出发, 构造出一组标准正交向量 $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, $\boldsymbol{q}_3$.)

Let $\boldsymbol{q}_1 = \boldsymbol{a}/\|\boldsymbol{a}\|$ to make it a unit vector.

Find the projection of $\boldsymbol{b}$ onto $\boldsymbol{q}_1$ (which is the direction of $\boldsymbol{a}$):

$$\boldsymbol{\gamma}_{12} = \left(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}\right)\boldsymbol{q}_1$$



That component has to be subtracted:
Take $\boldsymbol{B} = \boldsymbol{b} - \boldsymbol{\gamma}_{12} = \boldsymbol{b} - \left(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}\right)\boldsymbol{q}_1$,
and $\boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$.

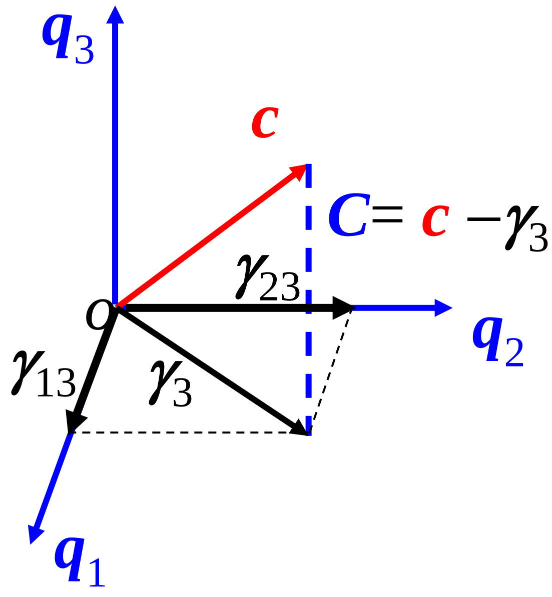
The vector $\boldsymbol{c}$ will not be in the plane of $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$, which is the plane of $\boldsymbol{a}$ and $\boldsymbol{b}$.

However, it may have a component in that plane, and that has to be subtracted:

$$\boldsymbol{C} = \boldsymbol{c} - \boldsymbol{\gamma}_3 = \boldsymbol{c} - \left(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c}\right)\boldsymbol{q}_1 - \left(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c}\right)\boldsymbol{q}_2,$$

and $\boldsymbol{q}_3 = \boldsymbol{C}/\|\boldsymbol{C}\|$.

When there is a fourth vector in $\mathbf{R}^n$ ($n \geq 4$) , we subtract away its components in the directions of $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, $\boldsymbol{q}_3$.



The idea of the Gram-Schmidt process:

to subtract from every new vector its components in the directions that are already settled.

(从每一个新的向量中扣除其在已经确定了的方向上的投影分量)

The Gram-Schmidt process:

starts with *independent* vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and ends with *orthonormal* vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$.

At step *j* it subtracts from $\boldsymbol{a}_j$ its components in the directions $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_{j-1}$ that are already settled:

$$\boldsymbol{A}_j = \boldsymbol{a}_j - (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_j)\boldsymbol{q}_1 - \cdots - (\boldsymbol{q}_{j-1}^{\mathrm{T}}\boldsymbol{a}_j)\boldsymbol{q}_{j-1}.$$

Then $\boldsymbol{q}_j$ is the unit vector $\boldsymbol{A}_j / \|\boldsymbol{A}_j\|$.

Example 5 Suppose the independent vectors are $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$:

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{c} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

Then

$$\boldsymbol{q}_1 = \frac{\boldsymbol{a}}{\sqrt{2}};$$

$$\boldsymbol{B} = \boldsymbol{b} - \left(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}\right)\boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \boldsymbol{q}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix};$$

$$\boldsymbol{C} = \boldsymbol{c} - \left(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c}\right)\boldsymbol{q}_1 - \left(\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c}\right)\boldsymbol{q}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \sqrt{2} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix} - \sqrt{2} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{q}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Remark:

1. $\boldsymbol{Q} = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \boldsymbol{q}_3] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix}$ is an orthogonal matrix.

Example 5 Suppose the independent vectors are $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$:

$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{c} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

Remark:

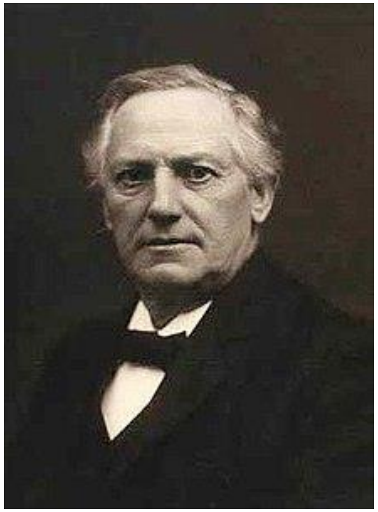
2. It is easier to compute the orthogonal vectors without forcing their lengths to equal one. They can be normalized to unit vector at the end by dividing by their lengths. (先正交化, 再单位化)

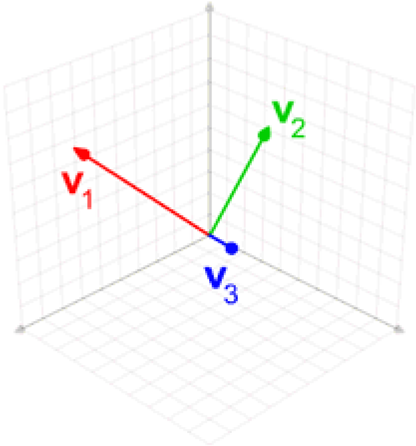
$\boldsymbol{A} = \boldsymbol{a}$;

$$\boldsymbol{B} = \boldsymbol{b} - \left(\frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\right)\boldsymbol{A} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix};$$

$$\boldsymbol{C} = \boldsymbol{c} - \left(\frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\right)\boldsymbol{A} - \left(\frac{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}}\right)\boldsymbol{B} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - 2\begin{bmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

Then we have $\boldsymbol{q}_1 = \boldsymbol{A}/\|\boldsymbol{A}\|$, $\boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$, $\boldsymbol{q}_3 = \boldsymbol{C}/\|\boldsymbol{C}\|$.





Jørgen Pedersen Gram

Erhard Schmidt

The modified Gram-Schmidt process being executed on three linearly independent, non-orthogonal vectors of a basis for $\mathbf{R}^3$

https://en.wikipedia.org/wiki/Gram%E2%80%93Schmidt_process

https://www.geogebra.org/m/stqjb5er

IV. QR Factorization (QR分解)

The Gram-Schmidt process:

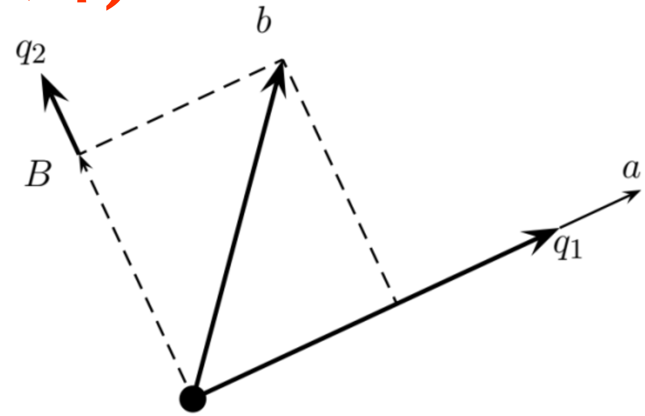
starts with *independent* vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ and ends with *orthonormal* vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$.

Let $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]$,

$$\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n].$$

The matrices $\boldsymbol{A}$ and $\boldsymbol{Q}$ are *m* by *n* when the *n* vectors are in *m*-dimensional space,

and there has to be a third matrix that connects them: $\boldsymbol{R}$.



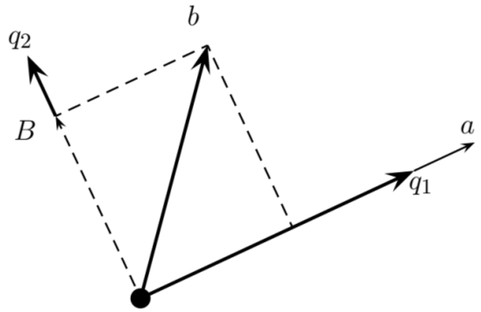
For example,

$$\boldsymbol{A} = [\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}], \;\; \boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3].$$

The idea is to write the $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ as combinations of the $\boldsymbol{q}$'s:

$$\boldsymbol{a} = (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a})\boldsymbol{q}_1,$$
$$\boldsymbol{b} = (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_1 + (\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_2,$$
$$\boldsymbol{c} = (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c})\boldsymbol{q}_1 + (\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c})\boldsymbol{q}_2 + (\boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c})\boldsymbol{q}_3.$$

$$\boldsymbol{a} = (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a})\boldsymbol{q}_1,$$
$$\boldsymbol{b} = (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_1 + (\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_2,$$
$$\boldsymbol{c} = (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c})\boldsymbol{q}_1 + (\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c})\boldsymbol{q}_2 + (\boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c})\boldsymbol{q}_3.$$



If we express that in matrix form we have ***the new factorization*** $\boldsymbol{A} = \boldsymbol{QR}$:

$$\boldsymbol{A} = [\boldsymbol{a} \quad \boldsymbol{b} \quad \boldsymbol{c}] = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \boldsymbol{q}_3]\begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ & & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix} = \boldsymbol{QR}.$$

- $\boldsymbol{R}$ is *upper triangular* because of the way Gram-Schmidt was done. (The first vectors $\boldsymbol{a}$ and $\boldsymbol{q}_1$ fell on the same line. Then $\boldsymbol{q}_1$, $\boldsymbol{q}_2$ were in the same plane as $\boldsymbol{a}$, $\boldsymbol{b}$. The third vectors $\boldsymbol{c}$ and $\boldsymbol{q}_3$ were not involved until step 3.)
- *You see the lengths of* $\boldsymbol{a}$, $\boldsymbol{B}$, $\boldsymbol{C}$ *on the diagonal of* $\boldsymbol{R}$.
- $\boldsymbol{Q}$ has orthonormal columns.

Example 5 (Continued) Suppose the independent vectors are $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$:

$$\boldsymbol{a} = \begin{bmatrix}1\\0\\1\end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix}1\\0\\0\end{bmatrix}, \qquad \boldsymbol{c} = \begin{bmatrix}2\\1\\0\end{bmatrix}.$$

And we have $\boldsymbol{Q} = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \boldsymbol{q}_3] = \begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\\ 0 & 0 & 1\\ 1/\sqrt{2} & -1/\sqrt{2} & 0\end{bmatrix}.$

The whole factorization is

$$\boldsymbol{A} = \begin{bmatrix}1 & 1 & 2\\0 & 0 & 1\\1 & 0 & 0\end{bmatrix} = \begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\\ 0 & 0 & 1\\ 1/\sqrt{2} & -1/\sqrt{2} & 0\end{bmatrix}\begin{bmatrix}\sqrt{2} & 1/\sqrt{2} & \sqrt{2}\\ & 1/\sqrt{2} & \sqrt{2}\\ & & 1\end{bmatrix} = \boldsymbol{QR}.$$

$$[\boldsymbol{a} \quad \boldsymbol{b} \quad \boldsymbol{c}] = [\boldsymbol{q}_1 \quad \boldsymbol{q}_2 \quad \boldsymbol{q}_3]\begin{bmatrix}\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c}\\ & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c}\\ & & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c}\end{bmatrix}$$

Note: Another way to find $\boldsymbol{R}$? $\quad \boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}.$

Theorem 3 Every *m* by *n* matrix *with independent columns* can be factored into $\boldsymbol{A} = \boldsymbol{QR}$. The columns of $\boldsymbol{Q}$ are orthonormal, and $\boldsymbol{R}$ is upper triangular and invertible.

When $m = n$ and all matrices are square, $\boldsymbol{Q}$ becomes an orthogonal matrix. (*Any invertible matrix can be factorized as a product of an orthogonal matrix and an upper triangular matrix.*)

For least-squares problem: $Ax = b$

$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ becomes easier: $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$.

The normal equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ simplifies to a triangular system: $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\widehat{\boldsymbol{x}} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$

$$\Rightarrow \quad \boldsymbol{R}\widehat{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$$

This is just back-substitution because $\boldsymbol{R}$ is triangular.

Key words:

Orthogonal matrices

Rectangular Matrices with Orthonormal Columns

Gram-Schmidt orthogonalization

QR factorization

Homework

See Blackboard