Orthonormal Bases and Orthogonal Matrices

- **In *orthogonal basis*, every vector is perpendicular to every other vector in the basis.**
- **The vectors $q_1, \dots, q_k$ are *orthonormal* if**

$$q_i^T q_j = \begin{cases} 0 & \text{whenever } i \neq j, \text{giving the orthogonality} \\ 1 & \text{whenever } i = j, \text{giving the normalization} \end{cases}$$

- ***Standard basis*: (1, 0,…,0), (0, 1,…, 0), …,(0,…, 0, 1)**
- **A matrix formed by orthonormal vectors as its columns is called matrix *Q***
- ***Q* with standard basis as its columns = *I***
- ***Orthogonal matrix*: a *square* matrix with *orthonormal* columns**
- **If the columns of *Q* are orthonormal then**

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

Therefore $Q^TQ=I$ and $Q^T=Q^{-1}$ if *Q* is square. *For orthogonal matrices, the transpose is the inverse.*

- **$Q^TQ=I$ even if *Q* is rectangular; Q^T is only a left-inverse then.**

Orthogonal Matrices

Example: rotation matrix:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad Q^T = Q^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Example: permutation matrix

If $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$ **then** $P^{-1} = P^T = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ **(*x*, *y*, *z*)→(*y*, *z*, *x*)**

- **Note: There is no such thing as "orthonormal matrix" and there is no name for rectangular matrix with orthonormal columns *Q*.**
- **Not every orthogonal matrix represents a rotation**
- $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ **(a reflection) is orthogonal but not a rotation**
- **Geometrically, <u>an orthogonal matrix is the *product of a rotation and a reflection.*</u>**
- **Multiplication by an orthogonal matrix *Q* preserves lengths and inner products (angles):**

$$\|Qx\| = \|x\|; (Qx)^T(Qy) = x^T Q^T Q y = x^T y$$

***Proof*:** $\|Qx\|^2 = (Qx)^T(Qx) = x^T Q^T Q x = x^T x = \|x\|^2$

- **Space rotation or reflection preserves • and || ||**

Expressing a Vector by Orthonormal Bases

- **Let a_i's be the basis. How to express b as a combination of a_i's:**

  Solving $Ax=b$ with a_i's as columns of A ⇒ Not an easy problem!
- **Write b as a combination of orthonormal q_i's:**

$$b = x_1 q_1 + x_2 q_2 + \cdots + x_n q_n$$

Multiply both sides by q_i^T **:** $q_i^T b = x_i q_i^T q_i \ \Rightarrow\ x_i = q_i^T b$

$\Rightarrow\ b = (q_1^T b) q_1 + (q_2^T b) q_2 + \cdots + (q_n^T b) q_n = q_1 q_1^T b + q_2 q_2^T b + \cdots + q_n q_n^T b$

- **The problem is identical to expressing b as linear combination of columns of Q: solving $Qx=b$:**

$$x = Q^{-1} b = Q^T b = \begin{bmatrix} - & q_1^T & - \\ & & \\ - & q_n^T & - \end{bmatrix} b = \begin{bmatrix} q_1^T b \\ \\ q_n^T b \end{bmatrix}$$

- **b is the sum of its one-dimensional projections onto the lines through the q_i's:** $\dfrac{q_i q_i^T}{q_i^T q_i} b = q_i q_i^T b$
- **Pythagoras:** $\|b\|^2 = (q_1^T b)^2 + (q_2^T b)^2 + \cdots + (q_n^T b)^2$ **remember** $\|b\|^2 = \left\|Q^T b\right\|^2$**?**
- **Sine $Q^T = Q^{-1}$ ⇒ $QQ^T = I$ ⇒ The rows of a square orthogonal matrix are orthonormal. Not trivial! At least not trivial geometrically but trivial induction from linear algebra.**

Rectangular Matrices with Orthonormal Columns

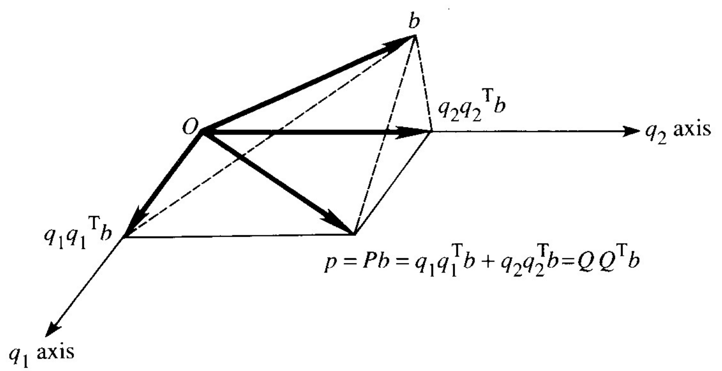
- **Rectangular Q only when $m>n$: no exact solution for $Qx=b$ ⇒ find the approximate** $\hat{x}$ **on the column space of Q ⇒**

  $Q^T Q \hat{x} = Q^T b \ \Rightarrow\ \hat{x} = Q^T b$ **(exact solution when Q is square and least square solution when Q is rectangular)**
- **If Q has orthonormal columns, then the least squares problem becomes easy:**

  $\hat{x} = Q^T b\ (\hat{x}_i = q_i^T b)$

  $p = Q\hat{x}$ **(projection of b onto column space of Q)**

  $p = QQ^T b$ **(so the projection matrix is $P=QQ^T$)**
- **Note:** $P = A(A^T A)^{-1} A^T = Q(Q^T Q)^{-1} Q^T = QQ^T$



Examples of Rectangular Matrix Q

- **Projecting b=(x, y, z) onto the x-y plane: p=(x, y, 0)**

  q_1=(1, 0, 0) $\Rightarrow$ $(q_1^T b)q_1 = (x, 0, 0)$;

  q_2=(0, 1, 0) $\Rightarrow$ $(q_2^T b)q_2 = (0, y, 0)$

  $$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- **In the problem of straight line fitting, if the *observed values of the independent variable have an average of zero*, then the columns in A are orthogonal. Example:**

  $$\begin{matrix} C + Dt_1 = y_1 \\ C + Dt_2 = y_2 \\ C + Dt_3 = y_3 \end{matrix} \quad \textbf{or} \quad \begin{bmatrix} 1 & -3 \\ 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Since (1, 1, 1)•(-3, 0, 3)^T=0 (two perpendicular columns), we can project y separately onto each column:

$$\hat{C} = \frac{[1 \quad 1 \quad 1][y_1 \quad y_2 \quad y_3]^T}{1^2 + 1^2 + 1^2}, \quad \hat{D} = \frac{[-3 \quad 0 \quad 3]\,[y_1 \quad y_2 \quad y_3]^T}{(-3)^2 + 0^2 + 3^2}$$

If [1,…,1]•t≠0 not orthogonal, instead of $y = C + Dt$**, we work with** $y = c + d(t - \bar{t})$**.**

Since we have $(1, \cdots, 1) \bullet (t_1 - \bar{t}, \cdots, t_m - \bar{t})^T = \sum t_i - \sum t_i = 0$ **orthogonal!**

$$\hat{d} = \frac{[(t_1 - \bar{t}) \quad \cdots \quad (t_m - \bar{t})][y_1 \quad \cdots \quad y_m]^T}{(t_1 - \bar{t})^2 + \cdots + (t_m - \bar{t})^2}$$

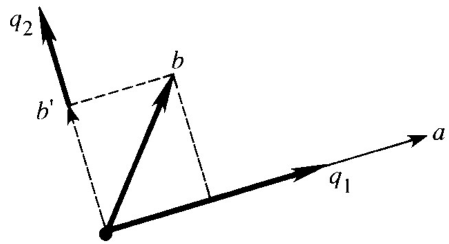
$\Rightarrow$ **Making columns perpendicular: Gram-Schmidt Process**

Gram-Schmidt Process

Given three independent vectors a, b, c, how to turn them into orthonormal vectors, q_1, q_2, q_3, that form the same space as formed by a, b and c.

q_1 can go in the direction of a $\Rightarrow$ $q_1 = a / \|a\|$

q_2 : subtract off the component of b, that belong to q_1.



$$b' = b - (q_1^T b)q_1 \;\Rightarrow\; q_2 = b' / \|b'\|$$

Similarly, $c' = c - (q_1^T c)q_1 - (q_2^T c)q_2 \;\Rightarrow\; q_3 = c' / \|c'\|$

(Subtracting from every new vector its components in the directions that are already settled)

- ***Gram-Schmidt Process*:**

  For n independent vectors a_1, a_2, …, a_n, at step j:

  $$a'_j = a_j - (q_1^T a_j)q_1 - \cdots - (q_{j-1}^T a_j)q_{j-1} \;\Rightarrow\; q_j = a'_j / \|a'_j\|$$

Calculations of Gram-Schmidt Process

Example:

$$a=\begin{bmatrix}1\\0\\1\end{bmatrix}, b=\begin{bmatrix}1\\0\\0\end{bmatrix}, c=\begin{bmatrix}2\\1\\0\end{bmatrix} \Rightarrow q_1=a/\sqrt{2}$$

$$b'=b-(q_1^T b)q_1=\begin{bmatrix}1\\0\\0\end{bmatrix}-\frac{1}{\sqrt{2}}\begin{bmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{bmatrix}=\begin{bmatrix}1/2\\0\\-1/2\end{bmatrix}$$

$$\Rightarrow q_2=b'/\|b'\|=b'/(1/\sqrt{2})=\begin{bmatrix}1/\sqrt{2}\\0\\-1/\sqrt{2}\end{bmatrix}$$

Easier Calculation: Projecting *b* onto *a* instead of q_1:

$$b'=b-\frac{a^T b}{a^T a}a \quad \textbf{and} \quad c'=c-\frac{a^T c}{a^T a}a-\frac{(b')^T c}{(b')^T b'}b'$$

In the previous example:

$$b'=\begin{bmatrix}1\\0\\0\end{bmatrix}-\frac{1}{2}\begin{bmatrix}1\\0\\1\end{bmatrix} \quad \textbf{and then} \quad c'=\begin{bmatrix}2\\1\\0\end{bmatrix}-\begin{bmatrix}1\\0\\1\end{bmatrix}-2\begin{bmatrix}1/2\\0\\-1/2\end{bmatrix}$$

Square roots enter only at the end !

A=*QR* Factorization

- **Factorization of *A* into *Q* times *R*, where *Q* is formed by the orthonormal columns: *A*=*QR***
- **Idea: Write columns of *A* as combinations of *Gram-Schmidt* q_i's:**

$$a=q_1(q_1^T a)$$

$$b=q_1(q_1^T b)+q_2(q_2^T b)$$

Similarly, $\quad c=q_1(q_1^T c)+q_2(q_2^T c)+q_3(q_3^T c)$

$$\Rightarrow \begin{bmatrix}| & | & |\\ q_1 & q_2 & q_3\\ | & | & |\end{bmatrix}\begin{bmatrix}q_1^T a & q_1^T b & q_1^T c\\ & q_2^T b & q_2^T c\\ & & q_3^T c\end{bmatrix}=\begin{bmatrix}| & | & |\\ a & b & c\\ | & | & |\end{bmatrix}\Rightarrow \boldsymbol{QR=A}$$

- **Every *m* by *n* matrix *A* with <u>linearly independent columns</u> can be factored into *A*=*QR*. The columns of *Q* are orthonormal and *R* is upper triangular and invertible. When *m*=*n* and all matrices are square, *Q* becomes an orthogonal matrix.**
- **How can orthogonalization help? Least Square Solutions**

$$\boldsymbol{A^T A=R^T Q^T QR=R^T R}$$

Normal Equations $A^T A\hat{x}=A^T b$ **becomes** $R^T R\hat{x}=R^T Q^T b$ **or** $R\hat{x}=Q^T b$. **Since *R* is triangular, the solution can be obtained quickly. The real cost is Gram-Schmidt (mn^2 opt. why?)**

Hilbert Space and Function

- ***Hilbert*** **Space: R^∞ with *finite length***

  v=(v_1, v_2, v_3, …) with $\|v\|^2 = v_1^2 + v_2^2 + v_3^2 + \cdots$ **finite**

  Example: (1, 1/2, 1/3. 1/4, …) is in Hilbert space and (1, 1, 1, …) is not.

- **In Hilbert space, the following properties are well kept:**

  1. $v \perp w$ **when** $v^T w = v_1 w_1 + v_2 w_2 + v_3 w_3 + \cdots = 0$
  2. **Schwarz Inequality** $\left|v^T w\right| \le \|v\| \, \|w\|$
  3. **Vectors can be turned into functions *f*(*x*)**

- **Function *f*(*x*)=sin *x*** ($0 \le x \le 2\pi$)**: we have infinitely many points of *f*(*x*) for *x* along the whole interval. sin *x* is thus like a vector with a whole continuum components. Length of such a vector?**

$$\|f\|^2 = \int_0^{2\pi} (f(x))^2 dx = \int_0^{2\pi} (\sin x)^2 dx = \pi$$

The length is finite: the *function* space is now *Hilbert* space

- ***Inner product*** **of two functions:**

$$(f, g) = \int_0^{2\pi} f(x) g(x) dx = \int_0^{2\pi} \sin x \cos x \, dx = 0$$

- **Two functions are orthogonal if $f^T g=0$ and are orthonormal after division by their lengths (in the case of sin and cos: ||sin *x*||=||cos *x*||=$\sqrt{\pi}$**

Function Spaces and Fourier Series

- ***Fourier Series*:**

$$y(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots$$

- **To compute the coefficient, say b_1:**

$$\int_0^{2\pi} y(x) \sin x dx = a_0 \int_0^{2\pi} \sin x dx + a_1 \int_0^{2\pi} \cos x \sin x dx + b_1 \int_0^{2\pi} (\sin x)^2 dx + \cdots$$

$$\Rightarrow b_1 = \frac{\int_0^{2\pi} y(x) \sin x dx}{\int_0^{2\pi} (\sin x)^2 dx} = \frac{(y, \sin x)}{(\sin x, \sin x)}$$ **(remember** $\hat{x} = \dfrac{b^T a}{a^T a}$ **?)**

<u>We are projecting *y* onto sin*x*!!</u>

- ***The Fourier series gives the coordinates of the vector y with respect to a set of infinitely many perpendicular axes***

- **How about polynomials? How to fit a polynomial function? 1, x, x^2, … are not orthogonal → Trouble!**

  It is virtually hopeless to solve $A^T A \hat{x} = A^T b$ **for the closest polynomial of degree ten.**

  ⇒ Gram-Schmidt orthogonalization!

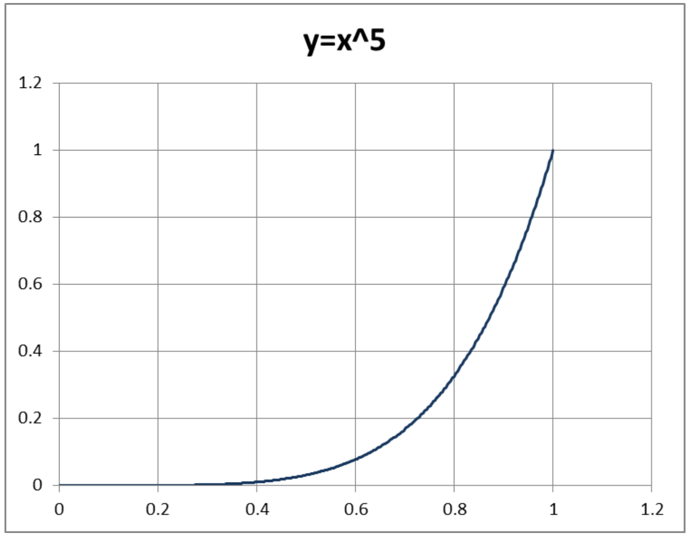
For $-1 \le x \le 1$**, since** $(1, x) = \int_{-1}^{1} x dx = 0$, $(x, x^2) = \int_{-1}^{1} x^3 dx = 0$

⇒ v_1=1 and v_2=*x* as the first two perpendicular axes

$$v_3 = x^2 - \frac{(1, x^2)}{(1,1)} 1 - \frac{(x, x^2)}{(x, x)} x = x^2 - \frac{\int_{-1}^{1} x^2 dx}{\int_{-1}^{1} 1 dx} = x^2 - \frac{1}{3}$$

This is called *Legendre polynomials*

Approximating $y=x^5$ by $C+Dx$ for $0 \le x \le 1$



Method 1: Solve $\begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = x^5 \qquad A^T A\hat{x} = A^T b$

$$\begin{bmatrix} (1,1) & (1,x) \\ (x,1) & (x,x) \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} (1,x^5) \\ (x,x^5) \end{bmatrix} \quad \textbf{or} \quad \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1/6 \\ 1/7 \end{bmatrix}$$

Method 2: Minimize $E^2 = \int_0^1 (x^5 - C - Dx)^2 \, dx$ **with respect to** C **and** D

Method 3: Apply Gram-Schmidt to replace x **by** $x-(1, x)/(1, 1)$ **(=** $x - 1/2 = x$**-average(**x**), which is orthogonal to 1). The best line now becomes:**

$$C + Dx = \frac{(x^5,1)}{(1,1)} 1 + \frac{(x^5, x-\frac{1}{2})}{(x-\frac{1}{2}, x-\frac{1}{2})}(x - \tfrac{1}{2}) = \tfrac{1}{6} + \tfrac{5}{7}(x - \tfrac{1}{2})$$