Orthogonality

Stephen Boyd and Sanjay Lall
Revision by Amirhossein Afsharrad

EE263
Stanford University

Orthonormal set of vectors

$$U^T U = \begin{bmatrix} u_1^T \\ \vdots \\ u_k^T \end{bmatrix} \begin{bmatrix} u_1 & \cdots & u_k \end{bmatrix} = \begin{array}{c|c} & j \\ \hline i & \begin{bmatrix} & & \\ & u_i^T u_j & \\ & & \end{bmatrix} \end{array}$$

set of vectors $\{u_1, \ldots, u_k\} \subset \mathbb{R}^n$ is

- *normalized* if $||u_i|| = 1$, $i = 1, \ldots, k$
  (u_i are called *unit vectors* or *direction vectors*)
- *orthogonal* if $u_i \perp u_j$ for $i \neq j$ $\quad \langle u_i, u_j \rangle = 0 = u_i^T u_j$
- *orthonormal* if both

slang: we say '$u_1, \ldots, u_k$ are orthonormal vectors' but orthonormality (like independence) is a property of a *set* of vectors, not vectors individually

in terms of $U = [u_1 \ \cdots \ u_k]$, orthonormal means

$$U^T U = I_k$$

$$u_i^T u_j = \begin{cases} 0 & i \neq j \\ \|u_i\|^2 = 1 & i = j \end{cases}$$

Orthonormality

$u_1^T(\alpha_1 u_1 + \cdots + \alpha_k u_k) = 0$

$\alpha_1 \underbrace{\|u_1\|^2}_{1} + \alpha_2 \underbrace{u_1^T u_2}_{0} + \cdots \to \alpha_1 = 0$

U has a left inverse (i.e. U^T) $\Rightarrow$ **an orthonormal set of vectors is independent**

- to see this, multiply $Ux = 0$ by U^T
- hence $\{u_1, \ldots, u_k\}$ is an *orthonormal basis* for

$$\mathbf{span}(u_1, \ldots, u_k) = \mathbf{range}(U)$$

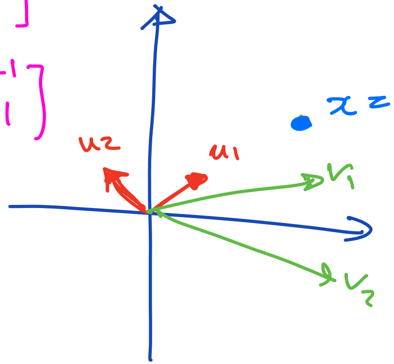
- **warning**: if $k < n$ then $UU^T \neq I$ (since its rank is at most k)

  (more on this matrix later ...)

Orthonormal basis for $\mathbb{R}^n$

$u_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$u_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$



$x = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

A matrix U is called *orthogonal* if

$$U \text{ is square and } U^\mathsf{T} U = I$$

- the set of columns $u_1, \ldots, u_n$ is an orthonormal *basis* for $\mathbb{R}^n$
- (you'd think such matrices would be called *orthonormal*, not *orthogonal*)
- it follows that $U^{-1} = U^\mathsf{T}$, and hence also $UU^\mathsf{T} = I$, *i.e.*,

$$\sum_{i=1}^{n} u_i u_i^\mathsf{T} = I$$

$x = \alpha_1 u_1 + \alpha_2 u_2$

$\alpha_1 = u_1^\mathsf{T} x$

$\alpha_2 = u_2^\mathsf{T} x$

$x = \beta_1 v_1 + \beta_2 v_2$

Expansion in orthonormal basis

$x = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$

$\alpha_i = ? \rightarrow u_i^T x = \alpha_1 \underbrace{u_i^T u_1}_{0} + \cdots + \alpha_i \overbrace{u_i^T u_i}^{1} + \cdots + \alpha_n \underbrace{u_i^T u_n}_{0}$

$\boxed{\alpha_i = u_i^T x}$

$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} u_1^T x \\ \vdots \\ u_n^T x \end{bmatrix} = U^T x$

suppose U is orthogonal, so $x = UU^\top x$, *i.e.*,

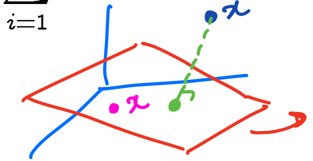
$x = U \underbrace{U^T x}_{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}}$

$$x = \sum_{i=1}^{n} (u_i^\top x) u_i$$

- $u_i^\top x$ is called the *component* of x in the direction u_i
- $a = U^\top x$ *resolves* x into the vector of its u_i components
- $x = Ua$ *reconstitutes* x from its u_i components
- $x = Ua = \sum_{i=1}^{n} a_i u_i$ is called the $(u_i$-$)$ *expansion* of x

$x \in \text{range}(U) \rightarrow x = Uz \; \{ \; U \underbrace{U^T U}_{I} z = Uz = x$

$x = U\alpha = \alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n$



$U \in \mathbb{R}^{n \times k}$, $k < n$, u_i's are orthonormal

when $UU^T x = x$?

$x \in \text{range}(U)$

Orthonormal basis in practice

$$f(x) = \sum_t \alpha_t \cos(tx) + \beta_t \sin(tx)$$

$$\langle f, g \rangle = \int fg$$

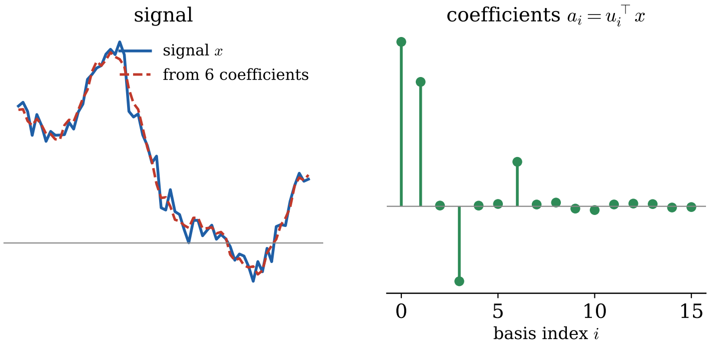
in an orthonormal basis the coordinates are just inner products,

$$a_i = u_i^\top x$$

no matrix needs to be inverted, and energy is preserved (Parseval):

$$||x||^2 = \sum_i a_i^2$$

- many signals have a few large coefficients in the right basis: keep those, drop the rest
- Fourier and DCT, wavelets, and PCA all supply such orthonormal bases



Complementary subspaces

$Q = [Q_1 \;\; Q_2]$, $Q^TQ = QQ^T = I$ $T = S^{\perp}$

$I = QQ^T = [Q_1 \;\; Q_2]\begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} = Q_1Q_1^T + Q_2Q_2^T$ $x \in T,\ y \in S \Rightarrow x^Ty = 0$

if $Q = [\overbrace{Q_1}^{k} \;\; \overbrace{Q_2}^{n-k}]$ and Q is orthogonal then **range**(Q_1) and **range**(Q_2) are called *complementary subspaces*, because

$[u_1 \cdots u_k]\,[u_{k+1} \cdots u_n]$

$$\boxed{\textbf{range}(Q_2) = \textbf{range}(Q_1)^{\perp}}$$

$x \in \text{range}(Q_1)$ $y \in \text{range}(Q_2)$

$x = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ $y = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha_{k+1} \\ \vdots \\ \alpha_n \end{bmatrix}$

T/F Question : $x \in \mathbb{R}^n \rightarrow x \in \text{range}(Q_1) \cup \text{range}(Q_2)$ F

- they are orthogonal *i.e.*, every vector in the first subspace is orthogonal to every vector in the second subspace
- every vector in $\mathbb{R}^m$ can be expressed as a sum of two vectors, one from each subspace
- each subspace is the *orthogonal complement* of the other

$x \in \mathbb{R}^n \rightarrow x = y + z$

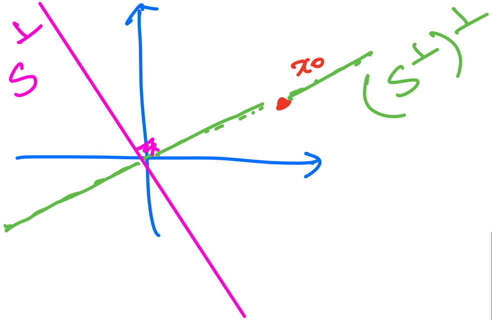
$y \in \text{range}(Q_1)$

$z \in \text{range}(Q_2)$

$\Rightarrow x = Q_1\underbrace{(Q_1^Tx)}_{\in \mathbb{R}^k} + Q_2\underbrace{(Q_2^Tx)}_{\in \mathbb{R}^{n-k}}$ $\rightarrow x = \overbrace{(Q_1Q_1^T + Q_2Q_2^T)}^{I}x$

Complementary subspaces

$S = \{x_0\}$



$(S^\perp)^\perp = S$ ✗

if S subspace ✓

o.w. $(S^\perp)^\perp = \text{span}(S)$

$$\textbf{range}(Q_2) = \textbf{range}(Q_1)^\perp$$

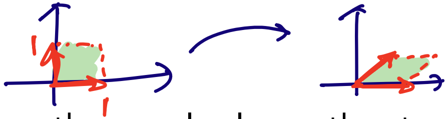
- $\textbf{range}\, Q_2 \subset (\textbf{range}\, Q_1)^\perp$ because $Q_1^T Q_2 = 0$
- to show $\textbf{range}\, Q_2 \supset (\textbf{range}\, Q_1)^\perp$, suppose $x \in (\textbf{range}\, Q_1)^\perp$, then $Q_1^T x = 0$, and since $x = Q_1 Q_1^T x + Q_2 Q_2^T x$ we have $x = Q_2 Q_2^T x$ and so $x \in \textbf{range}\, Q_2$

Geometric interpretation

$\|v\|^2 = v^T v$

$\|\alpha_1 u_1 + \alpha_2 u_2 + \cdots\|^2 = \alpha_1^2\|u_1\|^2 + \alpha_2^2\|u_2\|^2 + \cdots + \alpha_1\alpha_2 \underbrace{u_1^T u_2}_{0} + \cdots$

$$\|Uz\|^2 = (Uz)^T(Uz) = z^T \underbrace{(U^T U)}_{I} z = z^T z = \|z\|^2$$



if U has orthonormal columns then transformation $w = Uz$

- preserves *norm* of vectors, *i.e.*, $||Uz|| = ||z||$
- preserves *angles* between vectors, *i.e.*, $\angle(Uz, U\tilde{z}) = \angle(z, \tilde{z})$
- we say U is *isometric*, it preserves distances

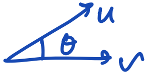
$U \in \mathbb{R}^{2\times 2}$, U is orthogonal

U is always either a rotation or a reflection

$\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}$

$\begin{bmatrix}\cos & \sin \\ \sin & -\cos\end{bmatrix}$

$$u, v \to \cos\theta = \frac{u^T v}{\|u\|\,\|v\|} \rightsquigarrow Uz_1, Uz_2 \Rightarrow \cos\theta = \frac{(Uz_1)^T(Uz_2)}{\underbrace{\|Uz_1\|}_{\|z_1\|}\,\underbrace{\|Uz_2\|}_{\|z_2\|}} = \frac{z_1^T \overbrace{U^T U}^{I} z_2}{\|z_1\|\,\|z_2\|} = \frac{z_1^T z_2}{\|z_1\|\,\|z_2\|}$$

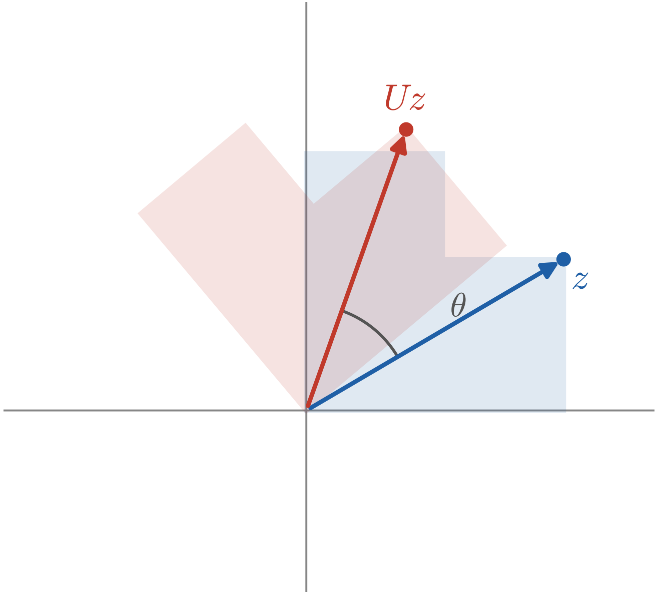


Application: orthogonal maps preserve geometry

orthogonal maps rotate or reflect without changing lengths or angles

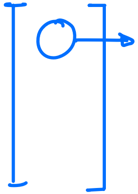
- **rotary position embeddings (RoPE):** transformers rotate query and key vectors by an angle set by position, preserving norm and encoding relative position in the angle
- **orthogonal initialization:** keeps signal and gradient norms stable through deep networks, avoiding vanishing or exploding values
- **random projections:** near-orthogonal maps cut dimension while preserving distances (Johnson–Lindenstrauss)

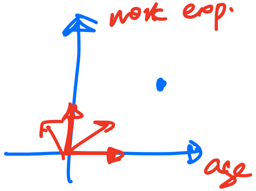
$\|Uz\| = \|z\|$, angles preserved



Near-orthogonality in high dimensions

$x = \alpha_1 u_1 + \dots + \alpha_{1000,000} u_{1,000,000}$





$|u^T v| < 0.01$

take two random directions in $\mathbb{R}^d$: as d grows they are almost surely nearly orthogonal, $u^\top v \approx 0$

so a high-dimensional space holds *many more than d* almost-orthogonal directions

- models exploit this *superposition*: a d-dimensional representation can store far more than d features as nearly-orthogonal directions
- nearly-orthogonal features barely interfere, so each can be read out almost independently