Orthogonality

Stephen Boyd and Sanjay Lall
Revision by Amirhossein Afsharrad

EE263
Stanford University

Orthonormal set of vectors

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 $i \neq j$
- *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

$$\boxed{U^\mathsf{T} U = I_k}$$

Orthonormality

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

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

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

  (more on this matrix later . . .)

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

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.*,

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

Expansion in orthonormal basis

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

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

- $u_i^\top x$ 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

Orthonormal basis in practice

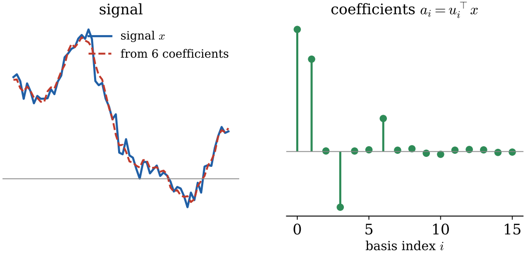
in an orthonormal basis the coordinates are just inner products,

$$a_i = u_i^\top x$$

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

if $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ and Q is orthogonal then **range**(Q_1) and **range**(Q_2) are called *complementary subspaces*, because

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

- 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

Complementary subspaces

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

- $\textbf{range}\, Q_2 \subset (\textbf{range}\, Q_1)^\perp$ 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

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

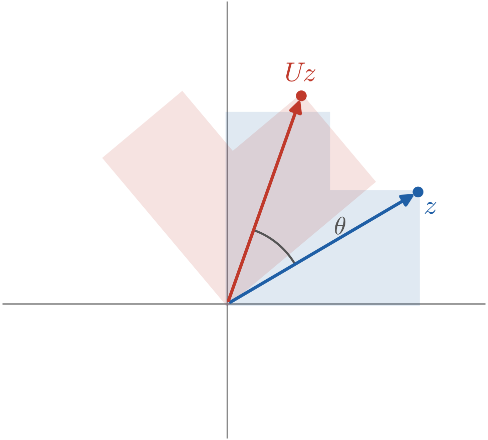
- preserves *norm* of vectors, $i.e.$, $||Uz|| = ||z||$
- preserves *angles* between vectors, $i.e.$, $\angle(Uz, U\tilde{z}) = \angle(z, \tilde{z})$
- we say U is *isometric*, it preserves distances

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

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