MATH 323

Linear Algebra

**Lecture 24a:
Orthogonality in inner product spaces.
The Gram-Schmidt process.**

Norm

The notion of *norm* generalizes the notion of length of a vector in $\mathbb{R}^n$.

Definition. Let V be a vector space. A function $\alpha : V \to \mathbb{R}$, usually denoted $\alpha(\mathbf{x}) = \|\mathbf{x}\|$, is called a **norm** on V if it has the following properties:

(i) $\|\mathbf{x}\| \geq 0$, $\|\mathbf{x}\| = 0$ only for $\mathbf{x} = \mathbf{0}$ (positivity)

(ii) $\|r\mathbf{x}\| = |r| \|\mathbf{x}\|$ for all $r \in \mathbb{R}$ (homogeneity)

(iii) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality)

A **normed vector space** is a vector space endowed with a norm. The norm defines a distance function on the normed vector space: $\mathrm{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.

Examples. $V = \mathbb{R}^n$, $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

- $\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|)$.

- $\|\mathbf{x}\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)^{1/p}$, $p \geq 1$.

Examples. $V = C[a, b]$, $f : [a, b] \to \mathbb{R}$.

- $\|f\|_\infty = \max\limits_{a \leq x \leq b} |f(x)|$.

- $\|f\|_p = \left(\displaystyle\int_a^b |f(x)|^p \, dx\right)^{1/p}$, $p \geq 1$.

Inner product

The notion of *inner product* generalizes the notion of dot product of vectors in $\mathbb{R}^n$.

Definition. Let V be a vector space. A function $\beta : V \times V \to \mathbb{R}$, usually denoted $\beta(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$, is called an **inner product** on V if it is positive, symmetric, and bilinear. That is, if

(i) $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ only for $\mathbf{x} = \mathbf{0}$ (positivity)

(ii) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ (symmetry)

(iii) $\langle r\mathbf{x}, \mathbf{y} \rangle = r \langle \mathbf{x}, \mathbf{y} \rangle$ (homogeneity)

(iv) $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$ (distributive law)

An **inner product space** is a vector space endowed with an inner product.

Examples. $V = \mathbb{R}^n$.

- $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$.

- $\langle \mathbf{x}, \mathbf{y} \rangle = d_1 x_1 y_1 + d_2 x_2 y_2 + \cdots + d_n x_n y_n$,
where $d_1, d_2, \ldots, d_n > 0$.

Examples. $V = C[a, b]$.

- $\langle f, g \rangle = \displaystyle\int_a^b f(x) g(x) \, dx$.

- $\langle f, g \rangle = \displaystyle\int_a^b f(x) g(x) w(x) \, dx$,

where w is bounded, piecewise continuous, and
$w > 0$ everywhere on $[a, b]$.

Norms induced by inner products

Theorem Suppose $\langle \mathbf{x}, \mathbf{y} \rangle$ is an inner product on a vector space V. Then $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ is a norm.

Examples. • The length of a vector in $\mathbb{R}^n$,
$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2},$$

is the norm induced by the dot product

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

• The norm $\|f\|_2 = \left(\int_a^b |f(x)|^2 \, dx \right)^{1/2}$ on the

vector space $C[a, b]$ is induced by the inner product

$$\langle f, g \rangle = \int_a^b f(x) g(x) \, dx.$$

Orthogonality

Let V be an inner product space with an inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\| \cdot \|$. We say that vectors $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal** if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

Definition. A nonempty set $S \subset V$ of nonzero vectors is called an **orthogonal set** if all vectors in S are mutually orthogonal. That is, $\mathbf{0} \notin S$ and $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ for any $\mathbf{x}, \mathbf{y} \in S$, $\mathbf{x} \neq \mathbf{y}$.

An orthogonal set $S \subset V$ is called **orthonormal** if $\| \mathbf{x} \| = 1$ for any $\mathbf{x} \in S$.

Example. The standard basis $\mathbf{e}_1 = (1, 0, 0, \ldots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \ldots, 0)$, $\ldots$, $\mathbf{e}_n = (0, 0, 0, \ldots, 1)$ in $\mathbb{R}^n$. It is an orthonormal set.

Example

- $V = C[-\pi, \pi]$, $\langle f, g \rangle = \displaystyle\int_{-\pi}^{\pi} f(x)g(x)\, dx$.

$f_1(x) = \sin x$, $f_2(x) = \sin 2x$, ..., $f_n(x) = \sin nx$, ...

$$\langle f_m, f_n \rangle = \int_{-\pi}^{\pi} \sin(mx)\sin(nx)\, dx = \begin{cases} \pi & \text{if } m = n, \\ 0 & \text{if } m \neq n. \end{cases}$$

Thus the set $\{f_1, f_2, f_3, \dots\}$ is orthogonal but not orthonormal.

It is orthonormal with respect to a scaled inner product
$$\langle\langle f, g \rangle\rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)g(x)\, dx.$$

Orthogonality $\implies$ linear independence

Theorem Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are nonzero vectors that form an orthogonal set. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent.

Proof: Suppose $t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_k \mathbf{v}_k = \mathbf{0}$ for some $t_1, t_2, \ldots, t_k \in \mathbb{R}$.

Then for any index $1 \leq i \leq k$ we have

$$\langle t_1 \mathbf{v}_1 + t_2 \mathbf{v}_2 + \cdots + t_k \mathbf{v}_k, \mathbf{v}_i \rangle = \langle \mathbf{0}, \mathbf{v}_i \rangle = 0.$$

$$\implies \ t_1 \langle \mathbf{v}_1, \mathbf{v}_i \rangle + t_2 \langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + t_k \langle \mathbf{v}_k, \mathbf{v}_i \rangle = 0$$

By orthogonality, $t_i \langle \mathbf{v}_i, \mathbf{v}_i \rangle = 0 \implies t_i = 0.$

Orthonormal bases

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be an orthonormal basis for an inner product space V.

Theorem Let $\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$ and $\mathbf{y} = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_n\mathbf{v}_n$, where $x_i, y_j \in \mathbb{R}$. Then

(i) $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$,

**(ii)** $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.

Proof: (ii) follows from (i) when $\mathbf{y} = \mathbf{x}$.

$$\langle \mathbf{x}, \mathbf{y} \rangle = \left\langle \sum_{i=1}^{n} x_i\mathbf{v}_i, \sum_{j=1}^{n} y_j\mathbf{v}_j \right\rangle = \sum_{i=1}^{n} x_i \left\langle \mathbf{v}_i, \sum_{j=1}^{n} y_j\mathbf{v}_j \right\rangle$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{i=1}^{n} x_i y_i.$$

Orthogonal projection

Theorem Let V be an inner product space and V_0 be a finite-dimensional subspace of V. Then any vector $\mathbf{x} \in V$ is uniquely represented as $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p} \in V_0$ and $\mathbf{o} \perp V_0$.

The component $\mathbf{p}$ is called the **orthogonal projection** of the vector $\mathbf{x}$ onto the subspace V_0.
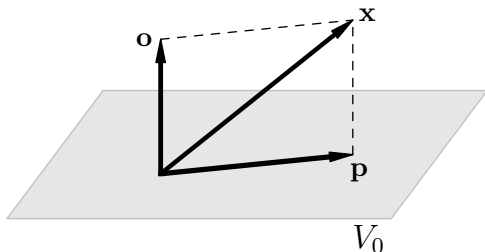
The projection $\mathbf{p}$ is closer to $\mathbf{x}$ than any other vector in V_0. Hence the distance from $\mathbf{x}$ to V_0 is $\|\mathbf{x} - \mathbf{p}\| = \|\mathbf{o}\|$.

Let V be an inner product space. Let $\mathbf{p}$ be the orthogonal projection of a vector $\mathbf{x} \in V$ onto a finite-dimensional subspace V_0.

If V_0 is a one-dimensional subspace spanned by a vector $\mathbf{v}$ then $\mathbf{p} = \dfrac{\langle \mathbf{x}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v}$.

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthogonal basis for V_0 then
$$\mathbf{p} = \frac{\langle \mathbf{x}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 + \frac{\langle \mathbf{x}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 + \cdots + \frac{\langle \mathbf{x}, \mathbf{v}_n \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle} \mathbf{v}_n.$$

Indeed, $\langle \mathbf{p}, \mathbf{v}_i \rangle = \displaystyle\sum_{j=1}^{n} \frac{\langle \mathbf{x}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \langle \mathbf{v}_j, \mathbf{v}_i \rangle = \frac{\langle \mathbf{x}, \mathbf{v}_i \rangle}{\langle \mathbf{v}_i, \mathbf{v}_i \rangle} \langle \mathbf{v}_i, \mathbf{v}_i \rangle = \langle \mathbf{x}, \mathbf{v}_i \rangle$

$\implies \langle \mathbf{x} - \mathbf{p}, \mathbf{v}_i \rangle = 0 \implies \mathbf{x} - \mathbf{p} \perp \mathbf{v}_i \implies \mathbf{x} - \mathbf{p} \perp V_0.$

Coordinates relative to an orthogonal basis

Theorem If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthogonal basis for an inner product space V, then

$$\mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 + \frac{\langle \mathbf{x}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{x}, \mathbf{v}_n \rangle}{\langle \mathbf{v}_n, \mathbf{v}_n \rangle}\mathbf{v}_n$$

for any vector $\mathbf{x} \in V$.

Corollary If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthonormal basis for an inner product space V, then

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{x}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{x}, \mathbf{v}_n \rangle \mathbf{v}_n$$

for any vector $\mathbf{x} \in V$.

The Gram-Schmidt orthogonalization process

Let V be a vector space with an inner product.
Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a basis for V. Let
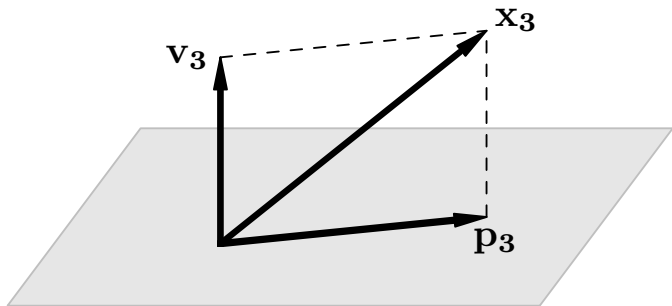
$$\mathbf{v}_1 = \mathbf{x}_1,$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1,$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2,$$
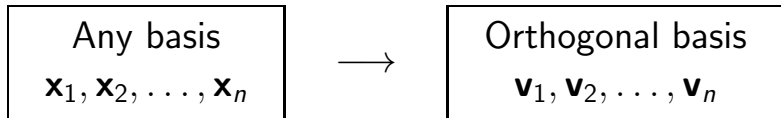
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\mathbf{v}_n = \mathbf{x}_n - \frac{\langle \mathbf{x}_n, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 - \cdots - \frac{\langle \mathbf{x}_n, \mathbf{v}_{n-1} \rangle}{\langle \mathbf{v}_{n-1}, \mathbf{v}_{n-1} \rangle} \mathbf{v}_{n-1}.$$

Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthogonal basis for V.

$$\text{Span}(\mathbf{v}_1, \mathbf{v}_2) = \text{Span}(\mathbf{x}_1, \mathbf{x}_2)$$

$$\boxed{\begin{array}{c}\text{Any basis} \\ \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\end{array}} \longrightarrow \boxed{\begin{array}{c}\text{Orthogonal basis} \\ \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\end{array}}$$

Properties of the Gram-Schmidt process:

• $\mathbf{v}_k = \mathbf{x}_k - (\alpha_1 \mathbf{x}_1 + \cdots + \alpha_{k-1} \mathbf{x}_{k-1})$, $1 \le k \le n$;

• the span of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is the same as the span of $\mathbf{x}_1, \ldots, \mathbf{x}_k$;

• $\mathbf{v}_k$ is orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$;

• $\mathbf{v}_k = \mathbf{x}_k - \mathbf{p}_k$, where $\mathbf{p}_k$ is the orthogonal projection of the vector $\mathbf{x}_k$ on the subspace spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$;

• $\|\mathbf{v}_k\|$ is the distance from $\mathbf{x}_k$ to the subspace spanned by $\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}$.

Normalization

Let V be a vector space with an inner product.
Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthogonal basis for V.

Let $\mathbf{w}_1 = \dfrac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$, $\mathbf{w}_2 = \dfrac{\mathbf{v}_2}{\|\mathbf{v}_2\|}, \ldots, \mathbf{w}_n = \dfrac{\mathbf{v}_n}{\|\mathbf{v}_n\|}$.

Then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ is an orthonormal basis for V.

Theorem Any finite-dimensional vector space with an inner product has an orthonormal basis.

Remark. An infinite-dimensional vector space with an inner product may or may not have an orthonormal basis.

Orthogonalization / Normalization

An alternative form of the Gram-Schmidt process combines orthogonalization with normalization.

Suppose $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a basis for an inner product space V. Let

$$\mathbf{v}_1 = \mathbf{x}_1, \quad \mathbf{w}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|},$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{w}_1 \rangle \mathbf{w}_1, \quad \mathbf{w}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|},$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \langle \mathbf{x}_3, \mathbf{w}_1 \rangle \mathbf{w}_1 - \langle \mathbf{x}_3, \mathbf{w}_2 \rangle \mathbf{w}_2, \quad \mathbf{w}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|},$$

....

$$\mathbf{v}_n = \mathbf{x}_n - \langle \mathbf{x}_n, \mathbf{w}_1 \rangle \mathbf{w}_1 - \cdots - \langle \mathbf{x}_n, \mathbf{w}_{n-1} \rangle \mathbf{w}_{n-1},$$
$$\mathbf{w}_n = \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|}.$$

Then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ is an orthonormal basis for V.

Problem. Let Π be the plane spanned by vectors $\mathbf{x}_1 = (1, 1, 0)$ and $\mathbf{x}_2 = (0, 1, 1)$.

(i) Find the orthogonal projection of the vector $\mathbf{y} = (4, 0, -1)$ onto the plane Π.

(ii) Find the distance from $\mathbf{y}$ to Π.

First we apply the Gram-Schmidt process to the basis $\mathbf{x}_1, \mathbf{x}_2$:

$\mathbf{v}_1 = \mathbf{x}_1 = (1, 1, 0)$,

$\mathbf{v}_2 = \mathbf{x}_2 - \dfrac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 = (0, 1, 1) - \dfrac{1}{2}(1, 1, 0) = (-1/2, 1/2, 1)$.

Now that $\mathbf{v}_1, \mathbf{v}_2$ is an orthogonal basis for Π, the orthogonal projection of $\mathbf{y}$ onto Π is

$\mathbf{p} = \dfrac{\langle \mathbf{y}, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle} \mathbf{v}_1 + \dfrac{\langle \mathbf{y}, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle} \mathbf{v}_2 = \dfrac{4}{2}(1, 1, 0) + \dfrac{-3}{3/2}(-1/2, 1/2, 1)$

$= (2, 2, 0) + (1, -1, -2) = (3, 1, -2)$.

The distance from $\mathbf{y}$ to Π is $\|\mathbf{y} - \mathbf{p}\| = \|(1, -1, 1)\| = \sqrt{3}$.

Problem. Approximate the function $f(x) = e^x$ on the interval $[-1, 1]$ by a quadratic polynomial.

The best approximation would be a polynomial $p(x)$ that minimizes the distance relative to the uniform norm:
$$\|f - p\|_\infty = \max_{|x| \leq 1} |f(x) - p(x)|.$$

However there is no analytic way to find such a polynomial. Instead, one can find a **"least squares"** approximation that minimizes the integral norm
$$\|f - p\|_2 = \left(\int_{-1}^{1} |f(x) - p(x)|^2 \, dx \right)^{1/2}.$$

The norm $\| \cdot \|_2$ is induced by the inner product
$$\langle g, h \rangle = \int_{-1}^{1} g(x)h(x)\, dx.$$

Therefore $\|f - p\|_2$ is minimal if p is the orthogonal projection of the function f on the subspace $\mathcal{P}_3$ of quadratic polynomials.

We should apply the Gram-Schmidt process to the polynomials $1, x, x^2$, which form a basis for $\mathcal{P}_3$. This would yield an orthogonal basis p_0, p_1, p_2. Then
$$p(x) = \frac{\langle f, p_0 \rangle}{\langle p_0, p_0 \rangle} p_0(x) + \frac{\langle f, p_1 \rangle}{\langle p_1, p_1 \rangle} p_1(x) + \frac{\langle f, p_2 \rangle}{\langle p_2, p_2 \rangle} p_2(x).$$