Section 5: Inner products spaces

Matthew Macauley

School of Mathematical & Statistical Sciences
Clemson University
http://www.math.clemson.edu/~macaule/

Math 8530, Advanced Linear Algebra

Overview

Up until now, much of our previous theory has been algebraic in flavor. What's been missing is a metric.

In this section, we will study vector spaces where we also have a notion of length.

As a result, this part of the class will contain more analysis, and less algebra.

In regular Euclidean space, we have standard concepts such as length and angle.

These allow us to speak of orthogonality, and to project vectors onto other vectors, or onto subspaces.

All of this is made possible by the dot product:

$$\langle x, y \rangle := x \cdot y = (x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = x_1 y_1 + \cdots + x_n y_n.$$

This works because the dot product is a symmetric bilinear form with an additional property.

In this section, we will abstract this notion to the concept of an inner product.

Until we say otherwise, we will assume that X is an n-dimensional vector space over $\mathbb{R}$.

Euclidean geometry

The length or norm of $x \in X$, denoted $||x||$, is the distance from x to $0 \in X$.

By the Pythagorean theorem, $||x|| = \sqrt{x_1^2 + \cdots + x_n^2}$. Clearly, $||x||^2 = \langle x, x \rangle$.

Since the dot product is symmetric and bilinear:

$$\begin{aligned}\langle x+y, x+y \rangle &= \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \\ &= ||x||^2 + 2\langle x, y\rangle + ||y||^2 \\ &= ||x+y||^2.\end{aligned}$$

Likewise,

$$\begin{aligned}\langle x-y, x-y \rangle &= \langle x, x\rangle - 2\langle x, y\rangle + \langle y, y\rangle \\ &= ||x||^2 - 2\langle x, y\rangle + ||y||^2 \\ &= ||x-y||^2.\end{aligned}$$

Remarks

- This is independent of the choice of basis (coordinate system)
- Geometrically, we understand $||x||$, $||y||$, and $||x-y||$, but not $\langle x, y\rangle$...yet.

How the dot product defines angles

To understand $\langle x, y\rangle$, we'll pick a special x and y.

Given any basis ("coordinate system") $x_1, \dots, x_n$:

1. Let x be a scalar of x_1. Then $x = (||x||, 0, \dots, 0)$.
2. Let $y \in \text{Span}(x_1, x_2)$. Then $y = (||y|| \cos\theta,\ ||y|| \sin\theta,\ 0, \dots, 0)$.

The dot product of x and y is thus

$$\langle x, y\rangle = (||x||, 0, \dots, 0) \cdot (||y|| \cos\theta, ||y|| \sin\theta, 0, \dots, 0) = ||x||\, ||y||\, \cos\theta.$$

We can characterize the angle between x and y as

$$\cos\theta = \frac{\langle x, y\rangle}{||x||\, ||y||}.$$

We can also derive the law of cosines:

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

Remark

One requirement for generalizing Euclidean space will be that $-1 \leq \cos\theta \leq 1$, i.e.,

$$-1 \leq \frac{\langle x, y\rangle}{||x||\, ||y||} \leq 1.$$

Fundamental properties of Euclidean space

Cauchy-Schwarz inequality

For all $x, y \in \mathbb{R}^n$,

$$|\langle x, y \rangle| \leq ||x|| \cdot ||y||,$$

and equality holds if and only if x and y are scalar multiples of each other.

Triangle inequality

For all $x, y \in \mathbb{R}^n$,

$$||x + y|| \leq ||x|| + ||y||.$$

Corollary 5.1

For any $x \in \mathbb{R}^n$,

$$||x|| = \max \left\{ \langle x, y \rangle : ||y|| = 1 \right\}.$$

Generalizing the dot product

The dot product on $\mathbb{R}^n$ gives us a notion of:

- *length*: $||x|| = \sqrt{\langle x, x\rangle}$
- *angle*: $\cos\theta = \dfrac{\langle x, y\rangle}{||x||\,||y||}$

But there's nothing special about the dot product, other than it's a symmetric bilinear form that is additionally positive-definite:

$$\langle x, x\rangle > 0, \qquad \text{for all } x \neq 0.$$

Definition

An inner product on a real vector space X is a symmetric positive-definite bilinear form

$$\langle -, -\rangle \colon X \times X \longrightarrow \mathbb{R}.$$

A vector space endowed with an inner product is an inner product space.

Key point

Everything we've done thus far (Cauchy-Schwarz, triangle inequality, etc.) works for a general inner product spaces.

Examples & non-examples

Let's explore some examples, and see what works and what doesn't.

- $X = \mathbb{R}^2$ with inner product

$$\langle a_1 e_1 + a_2 e_2,\ b_1 e_1 + b_2 e_2 \rangle = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = 2a_1 b_1 + a_1 b_2 + a_2 b_1 + 2a_2 b_2.$$

- $X = \mathbb{R}^2$ with inner product

$$\langle a_1 e_1 + a_2 e_2,\ b_1 e_1 + b_2 e_2 \rangle = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = a_1 b_1 + 2a_1 b_2 + 2a_2 b_1 + a_2 b_2.$$

- $X = \mathbb{R}^2$ with inner product

$$\langle a_1 e_1 + a_2 e_2,\ b_1 e_1 + b_2 e_2 \rangle = a_1 b_2 + a_2 b_1.$$

- $X = \mathrm{Hom}(X, Y)$ with inner product

$$\langle A, B \rangle = \mathrm{tr}(B^T A) = \sum_{i,j} a_{ij} b_{ij}.$$

- $X = \mathcal{C}[a, b]$, the space of continuous functions $f \colon [a, b] \to \mathbb{R}$ with inner product

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

Orthogality

Thinking of an inner product space as a generalization of Euclidean space, the concept of orthogonal is the analogue of perpendicular.

Definition

Two vectors $x, y \in X$ are orthogonal if $\langle x, y \rangle = 0$. We write $x \perp y$.

Pythagorean theorem

If $x \perp y$, then $||x + y||^2 = ||x||^2 + ||y||^2$.

Why orthogonal bases are nice

Let $x_1, \dots, x_n$ be an orthogonal basis (not necessarily orthonormal).

Given $v \in X$, we can write

$$v = a_1 x_1 + \cdots + a_n x_n.$$

We can find a formula for a_i by applying the linear map $\langle -, x_i \rangle$ to both sides:

$$a_i = \frac{\langle v, x_i \rangle}{\langle x_i, x_i \rangle}.$$

Remark

We can project x onto a vector $u \in X$ by defining

$$\operatorname{proj}_u x = \frac{\langle x, u \rangle}{\langle u, u \rangle}, \qquad \operatorname{Proj}_u x = \frac{\langle x, u \rangle}{\langle u, u \rangle} u.$$

Definition

The vectors $x_1, \dots, x_k$ in X is orthonormal if

$$\langle x_i, x_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Orthonormal bases

Key idea

- Orthogonal is the abstract version of "*perpendicular*."
- Orthonormal means "*perpendicular and unit length*."

Orthonormal bases are really desirable!

If $x_1, \dots, x_n$ is an orthonormal basis, $x = \sum_{i=1}^{n} a_i x_i$, and $y = \sum_{i=1}^{n} b_i x_i$, then

- $a_i = \mathrm{proj}_{x_i} x = \langle x, x_i \rangle$
- $\langle x, y \rangle = \sum_{i=1}^{n} a_i b_i$
- $||x||^2 = \sum_{i=1}^{n} a_i^2$.

Remark

If the columns of a matrix A are orthonormal, then $A^T A = I$.

Examples of orthogonality

Let's compare what orthogonality means in several inner product spaces:

1. $X = \mathbb{R}^n$, with the standard dot product.
2. $X = \mathbb{R}^2$, with inner product

$$\langle a_1e_1 + a_2e_2, b_1e_1 + b_2e_2 \rangle = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = 2a_1b_1 + a_1b_2 + b_1a_2 + 2a_2b_2.$$

Next, for fun, we'll do a quick high-level tour of how orthogonality arises in differential equations, involving:

1. Fourier series
2. Sturm-Liouville theory

Fourier series

Consider the space $X = \text{Per}_{2\pi}(\mathbb{R})$ of 2π-periodic piecewise functions, with the inner product

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) g(x)\, dx.$$

The set

$$\left\{ \tfrac{1}{\sqrt{2}},\ \cos x,\ \cos 2x, \dots \right\} \cup \left\{ \sin x,\ \sin 2x,\ \dots \right\}.$$

is an orthonormal basis w.r.t. to this inner product.

Thus, we can write each $f(x) \in \text{Per}_{2\pi}$ *uniquely* as

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,$$

where

$$a_n = \text{proj}_{\cos nx}\left(f\right) = \langle f, \cos nx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx\, dx$$

$$b_n = \text{proj}_{\sin nx}\left(f\right) = \langle f, \sin nx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx\, dx$$

Remark

There are technical details that need to be addressed regarding infinite sums and convergence, but those are beyond the scope of this class.

Legendre polynomials

The following is an eigenvalue problem $Ly = \lambda y$, on $(-1,1)$:

$$-\frac{d}{dx}\left[(1-x^2)\frac{d}{dx}y\right] = \lambda y.$$

The eigenvalues are $\lambda_n = n(n+1)$, $n \in \mathbb{N}$, and the eigenfunctions solve Legendre's equation:

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0.$$

For each n, one solution is a degree-n "Legendre polynomial"

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left[(x^2-1)^n\right].$$

They are orthogonal with respect to the inner product $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$.

It can be checked that

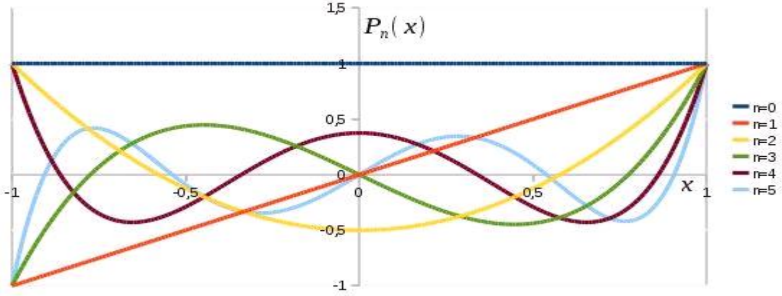
$$\langle P_m, P_n\rangle = \int_{-1}^{1} P_m(x)P_n(x)\,dx = \frac{2}{2n+1}\delta_{mn}.$$

By orthogonality, every function f, continuous on $-1 < x < 1$, can be expressed using Legendre polynomials:

$$f(x) = \sum_{n=0}^{\infty} c_n P_n(x), \qquad \text{where} \quad c_n = \frac{\langle f,\, P_n\rangle}{\langle P_n,\, P_n\rangle} = (n+\tfrac{1}{2})\langle f,\, P_n\rangle.$$

Legendre polynomials

$$P_0(x) = 1$$
$$P_1(x) = x$$
$$P_2(x) = \tfrac{1}{2}(3x^2 - 1)$$
$$P_3(x) = \tfrac{1}{2}(5x^3 - 3x)$$
$$P_4(x) = \tfrac{1}{8}(35x^4 - 30x^2 + 3)$$
$$P_5(x) = \tfrac{1}{8}(63x^5 - 70x^3 + 15x)$$
$$P_6(x) = \tfrac{1}{8}(231x^6 - 315x^4 + 105x^2 - 5)$$
$$P_7(x) = \tfrac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x)$$



Chebyshev polynomials

The following is a "weighted" eigenvalue problem $Ly = \lambda w(x) y$ on $[-1,1]$:

$$-\frac{d}{dx}\left[\sqrt{1-x^2}\,\frac{d}{dx}y\right] = \lambda \frac{1}{\sqrt{1-x^2}}\,y.$$

The eigenvalues are $\lambda_n = n^2$ for $n \in \mathbb{N}$, and the eigenfunctions solve Chebyshev's equation:

$$(1-x^2)y'' - xy' + n^2 y = 0.$$

For each n, one solution is a degree-n "Chebyshev polynomial," defined recursively by

$$T_0(x) = 1, \qquad T_1(x) = x, \qquad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

They are orthogonal with respect to the inner product $\langle f, g\rangle = \displaystyle\int_{-1}^{1} \frac{f(x)g(x)}{\sqrt{1-x^2}}\,dx$.

It can be checked that

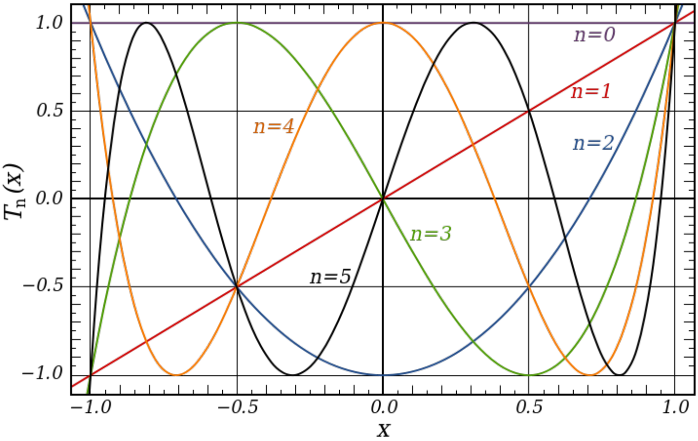
$$\langle T_m, T_n\rangle = \int_{-1}^{1} \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}}\,dx = \begin{cases} \frac{1}{2}\pi\delta_{mn} & m \neq 0, n \neq 0 \\ \pi & m = n = 0 \end{cases}$$

By orthogonality, every function $f(x)$, continuous for $-1 < x < 1$, can be expressed using Chebyshev polynomials:

$$f(x) \sim \sum_{n=0}^{\infty} c_n T_n(x), \qquad \text{where} \quad c_n = \frac{\langle f,\ T_n\rangle}{\langle T_n,\ T_n\rangle} = \frac{2}{\pi}\langle f,\ T_n\rangle, \ \text{ if } n > 0.$$

Chebyshev polynomials (of the first kind)

$$T_0(x) = 1 \qquad T_4(x) = 8x^4 - 8x^2 + 1$$
$$T_1(x) = x \qquad T_5(x) = 16x^5 - 20x^3 + 5x$$
$$T_2(x) = 2x^2 - 1 \qquad T_6(x) = 32x^6 - 48x^4 + 18x^2 - 1$$
$$T_3(x) = 4x^3 - 3x \qquad T_7(x) = 64x^7 - 112x^5 + 56x^3 - 7x$$



Constructing an orthonormal basis

Recall that X is an n-dimensional inner product space over $\mathbb{R}$.

We just saw why having an orthogonal (or even better: orthonormal) basis is very convenient.

Now, we'll see how to *construct* an orthogonal basis.

Gram-Schmidt process

Given an arbitrary basis $x_1, \dots, x_n$, construct an orthonormal basis $q_1, \dots, q_n$ for which $q_k \in \mathrm{Span}(x_1, \dots, x_k)$.

Remark

In matrix form, this leads to the QR factorization:

$$A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \langle x_1, q_1 \rangle & \langle x_2, q_1 \rangle & \langle x_3, q_1 \rangle & \cdots \\ 0 & \langle x_2, q_2 \rangle & \langle x_3, q_2 \rangle & \cdots \\ 0 & 0 & \langle x_3, q_3 \rangle & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = QR.$$

Identifying a space with its dual

Earlier in this class, we found it helpful to think of dual vectors $\ell \in X'$ as row vectors.

Going forward, it will be helpful to canonically identify these elements with vectors in X.

However, *the isomorphism will depend on the inner product*.

Proposition 5.2

Every linear function $\ell \in X'$ can be written as

$$\ell(x) = \langle x, y \rangle, \qquad \text{for some fixed } y \in X.$$

Corollary 5.3

For any fixed $y \in X$, the mapping

$$R_y \colon X \longrightarrow X', \qquad R_y \colon y \longmapsto \langle -, y \rangle$$

is an isomorphism. There is an analogous isomorphism

$$L_x \colon X \longrightarrow X', \qquad L_x \colon x \longmapsto \langle x, - \rangle.$$

Orthogonal complements

Definition

Let Y be a subspace of X. The orthogonal complement of Y is the set

$$Y^{\perp} := \{x \in X \mid \langle x, y \rangle = 0, \ \ \forall y \in Y\}.$$

Proposition 5.4

For any subspace Y of X, we have $X = Y \oplus Y^{\perp}$.

Examples of orthogonal complements

Let's return to several familiar examples.

1. $X = \mathbb{R}^n$, with the standard dot product.

2. $X = \mathbb{R}^2$, with inner product

$$\langle a_1 e_1 + a_2 e_2, b_1 e_1 + b_2 e_2 \rangle = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = 2a_1 b_1 + a_1 b_2 + b_1 a_2 + 2a_2 b_2.$$

3. $V = \mathrm{Hom}(X, Y)$ with inner product

$$\langle A, B \rangle = \mathrm{tr}(B^T A) = \sum_{i,j} a_{ij} b_{ij}.$$

4. $X = \mathrm{Per}_{2\pi}(\mathbb{R})$, the 2π-periodic functions, with the inner product

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) g(x) \, dx.$$

Orthogonal projection

If $X = Y \oplus Y^\perp$, then the map

$$P_Y \colon X \longrightarrow X, \qquad P_Y \colon y + y^\perp \longmapsto y$$

is the orthogonal projection of X onto Y.

Proposition 5.5 (exercise)

The orthogonal projection map P_Y is linear and idempotent (i.e., $P_Y^2 = P_Y$), and hence diagonalizable.

Proposition 5.6

The orthogonal projection map $P_Y \colon X \longrightarrow X$ sends $x \in X$ to

$$P_Y(x) = \arg\min \big\{ ||x - y|| : y \in Y \big\}.$$

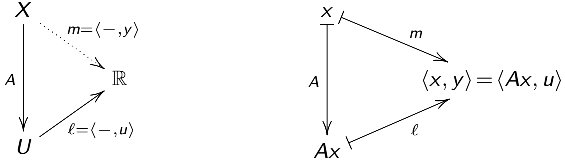
The transpose vs. the adjoint

Consider a linear map $A\colon X \to U$ between real inner product spaces.

The transpose of $A\colon X \to U$ is a linear map $A'\colon U' \to X'$ satisfying

$$(A'\ell, x) = (\ell, Ax), \qquad x \in X,\ \ell \in U'.$$

In the picture below, $A'\colon \ell \mapsto m$.



If we identify X and U with their duals via $y \mapsto \langle -, y\rangle$, the transpose $\langle -, u\rangle \mapsto \langle -, y\rangle$ defines a map $u \mapsto y$ called the adjoint of A, denoted A^*.

Key idea

Given a linear map $A\colon X \to U$,

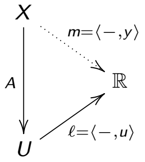
- the transpose $A'\colon U' \to X'$ maps $\ell \mapsto m$, independent of an inner product,
- the adjoint $A^*\colon U \to X$ maps $u \mapsto y$, and depends on the inner product structure.

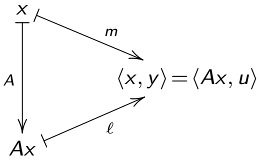
Formal definition of the adjoint

Definition

Let $A\colon X \to U$ be a linear map between real inner product spaces. The adjoint of A is the unique map $A^*\colon U \to X$ such that

$$\underbrace{\langle x, A^* u\rangle}_{\text{inner product in } X} = \underbrace{\langle Ax, u\rangle}_{\text{inner product in } U}.$$





Basic properties of adjoints

Proposition 5.7

Let $A, B\colon X \to U$ and $C\colon U \to V$ be linear maps between real inner product spaces.

(i) $(A+B)^* = A^* + B^*$

(ii) $(CA)^* = A^* C^*$

(iii) If A is bijective, then $(A^{-1})^* = (A^*)^{-1}$

(iv) $(A^*)^* = A$

(v) The matrix representations of A and A^* are transposes of each other.

Adjoints and the four subspaces

Proposition 5.8 (HW)

Let $A\colon X \to U$ be a linear maps between finite-dimensional inner product spaces. Then

(a) $N_{A^*} = R_A^\perp$

(b) $R_{A^*} = N_A^\perp$

(c) $N_A = R_{A^*}^\perp$

(d) $R_A = N_{A^*}^\perp$.

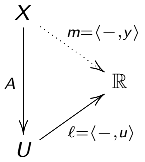
Together, this tells us that

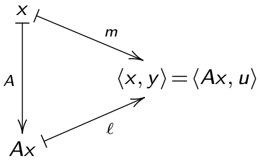
- $X = R_{A^*} \oplus N_A$ "*the orthogonal complement of the row space is the nullspace*"
- $U = R_A \oplus N_{A^*}$ "*the orthogonal complement of the column space is the left nullspace*"

Self-adjointness

Recall that the adjoint of A is the map $A^*\colon U \to X$ such that

$$\underbrace{\langle x, A^* u\rangle}_{\text{inner product in } X} = \underbrace{\langle Ax, u\rangle}_{\text{inner product in } U}.$$





Definition

A linear map $A\colon X \to U$ is self-adjoint if $A^* = A$.

Proposition 5.9

The linear maps A^*A and AA^* are self-adjoint.

Projections and orthogonal

Recall that if $X = Y \oplus Y^\perp$, then the map

$$P_Y \colon X \longrightarrow X, \qquad P_Y \colon y + y^\perp \longmapsto y$$

is the orthogonal projection of X onto Y.

Proposition 5.10

Orthogonal projections are self-adjoint.

Some books define a projection to be any linear map $P \colon X \to X$ such that $P^2 = P$.

It is not hard to show that $X = R_P \oplus N_P$.

Exercise (HW)

A projection $P \colon X \to X$ is an orthogonal projection if and only if it is self-adjoint.

More on the map A^*A

Lemma 5.11

The maps A and A^*A have the same nullspace.

Suppose A is an $m \times n$ matrix ($m > n$) with linearly independent columns. Then:

- the columns of A are a *basis* for the range (column space) of A
- A^*A is invertible.

The map A^*A and projection

The fact that $N_{A^*A} = N_A$, and the following, is the crux of the least squares method of finding the "best fit line."

Corollary 5.12

Consider an underdetermined system $Ax = b$, where $A\colon X \to U$ has trivial nullspace. The (unique) vector x that minimizes $||Ax - b||^2$ is the solution to $A^*Az = A^*b$.

An example of least squares

Let's find the "best fit line" $a_0 + a_1 x$ through the points $(1,1)$, $(2,2)$, and $(3,2)$ in $\mathbb{R}^2$.

The projection map $A(A^*A)^{-1}A^*$

Key idea

Let $y_1, \dots, y_k$ be a basis for Y, and $A = [y_1\ y_2\ \cdots\ y_k]$. Then

$$A(A^*A)^{-1}A^*$$

is the orthogonal projection matrix onto Y.

Isometries

Roughly speaking, an isometry is a distance-preserving map.

Definition

Let X be an inner product space. A function $A\colon X \to X$ is an isometry if

$$||Ax - Ay|| = ||x - y||, \qquad \text{for all } x, y \in X.$$

Examples

The following are all isometries of $\mathbb{R}^n$:

1. any translation
2. any rotation
3. any reflection
4. any compositions of these.

The isometries of X form a group ... but that's not a group we're all that interested in.

Orthogonal maps

Given any isometry, one can compose it with a translation to get an isometry that fixes 0.

Conversely, *any* isometry can be decomposed into one that fixes 0, followed by a translation.

Definition

An isometry $A\colon X \to X$ fixing 0 is said to be orthogonal.

The orthogonal maps on X form a group called the orthogonal group, denoted $O(X)$.

If $X = \mathbb{R}^n$, we denote this by $O(n)$ or O_n.

We will say that a matrix orthogonal if it represents an orthogonal linear map.

Remark

A matrix A is orthogonal if and only if its columns are orthonormal. That is, if $A^T A = I$.

Next, we'll show that all orthogonal maps are linear.

Properties of orthogonal maps

Theorem 5.13

Let $A\colon X \to X$ be orthogonal.

(i) A is linear

(ii) $A^*A = I$ (and conversely)

(iii) A is invertible, and A^{-1} is an isometry

(iv) $\det A = \pm 1$.

Key point

The geometric meaning of this theorem is that any map fixing 0 that preserves distances is linear, preserves angles, and preserves volume.

Definition

The subgroup of $O(X)$ of maps with determinant 1 is the special orthogonal group, denoted $SO(X)$.

Elements in $SO(X)$ describe rotations.

The norm of a linear map

The norm of a vector measures its size, or magnitude.

The set $\text{Hom}(X, U)$ of linear maps is a vector space. So what is the norm of $A\colon X \to U$?

The determinant is one way to measure the "size" of a linear map. However, this won't work, because

1. it is only defined when $X = U$,
2. it cannot be a norm, as there are nonzero linear maps with determinant zero.

There are a number of approaches that will work. Two reasonable ones are

1. the norm arising from the inner product $\langle A, B\rangle := \text{tr}(B^* A)$,
2. the largest factor that A can stretch a vector.

Let's recall the following definition from real analysis.

Definition

The supremum of a bounded subset $S \subseteq \mathbb{R}$, is its least upper bound. This always exists, and is denoted $\sup S$.

Moreover, if S is closed (contains all of its limit points), then $\sup S = \max S$.

Frobenius and induced norms

We can define an inner product on $\text{Hom}(X, U)$ by

$$\langle A, B \rangle = \text{tr}(B^* A).$$

Naturally, this gives us a definition of the norm of a linear map.

Definition

Let X and U be vector spaces. The Frobenius norm of $A\colon X \to U$ is

$$||A|| = \sqrt{\text{tr}(A^* A)} = \sqrt{\sum_{i,j} |a_{ij}|^2}.$$

This does *not* depend on any inner product structure of X or U.

Alternatively, we can define $||A||$ as the largest factor that A stretches a (nonzero) vector by.

Clearly, this depends on the inner products (and hence norms) on X and U.

Definition

Let X and U be inner product spaces. The induced norm of $A\colon X \to U$ is

$$||A|| := \sup_{||x||=1} ||Ax|| = \sup_{x \neq 0} \frac{||Ax||}{||x||}.$$

Properties of the induced norm

Henceforth, we will use the induced norm, unless otherwise stated.

Proposition 5.14

For any linear map $A\colon X \to U$,

(i) $||Az|| \leq ||A|| \cdot ||z||$, for all $z \in X$.

(ii) $||A|| = \sup_{||x||=||v||=1} \langle Ax, v\rangle$.

Properties of the induced norm

Proposition 5.15

Given linear maps $A, B\colon X \to U$ and $C\colon U \to V$,

(i) $||kA|| = |k| \cdot ||A||$

(ii) $||A + B|| \leq ||A|| + ||B||$

(iii) $||CA|| \leq ||C|| \cdot ||A||$

(iv) $||A^*|| = ||A||$.

Open sets and invertible maps

Let X be a vector space with a norm. For $x_0 \in X$ and $r > 0$, define the ball of radius r, centered at x_0 to be

$$B_r(x_0) = \{x \in X : ||x - x_0|| < r\}.$$

A subset $U \subseteq X$ is open if for every $u \in U$, there is some $r > 0$ for which $B_r(u) \subseteq U$.

The following implies that the subset of invertible maps is open.

Theorem 5.16

Let $A \colon X \to U$ be invertible, and suppose $B \colon X \to U$

$$||A - B|| < \frac{1}{||A^{-1}||}.$$

Then B is invertible.

Other norms

Definition

Let X and U be a vector spaces over R. A norm on $\text{Hom}(X, U)$ is a function

$$|| \cdot || : \text{Hom}(X, U) \longrightarrow \mathbb{R}$$

such that

1. $||kA|| = |k| \cdot ||A||$
2. $||A + B|| \leq ||A|| + ||B||$
3. $||A|| > 0$ for $A \neq 0$.

If $X = U$, then a norm is submultiplicative if

$$||AB|| \leq ||A|| \cdot ||B||.$$

Sequences of real and complex numbers

Definition

A sequence $\{a_k\}$ of numbers:

1. converges to a limit a if $|a_k - a| \to 0$. We write $\lim_{k \to \infty} a_k = a$.
2. is Cauchy if $|a_k - a_j| \to 0$ as $j, k \to \infty$.
3. is bounded if for some $R \geq 0$, every $|a_k| < R$.

The real (and complex) numbers are complete: every Cauchy sequence converges.

They are also locally compact: every bounded sequence contains a convergent subsequence.

Goal

Extend these properties from numbers to finite-dimensional inner product spaces.

Sequences of vectors

Definition

A sequence $\{x_k\}$ of vectors:

1. converges to a limit x if $||x_k - x|| \to 0$. We write $\lim_{k\to\infty} x_k = x$.
2. is Cauchy if $||x_k - x_j|| \to 0$ as $j, k \to \infty$.
3. is bounded if for some $R \geq 0$, every $||x_k|| < R$.

Completeness of inner product spaces

Proposition 5.17

Every finite-dimensional inner product space is complete.

Local compactness of inner product spaces

Proposition 5.18

Let X be an inner product space. Then X is locally compact if and only if $\dim X < \infty$.

Real vs. complex vector spaces

We have primarily been dealing with $\mathbb{R}$-vector spaces. Things are a little different over $\mathbb{C}$.

Let's compare the notion of *norm* for real vs. complex numbers.

- For any real number $x \in \mathbb{R}$, its norm (distance from 0) is $|x| = \sqrt{x^2} \in \mathbb{R}$.
- For any complex number $z = a + bi \in \mathbb{C}$, its norm (distance from 0) is defined by

$$|z| = \sqrt{z\overline{z}} = \sqrt{(a+bi)(a-bi)} = \sqrt{a^2+b^2}.$$

Let's now go from $\mathbb{R}$ and $\mathbb{C}$ to $\mathbb{R}^2$ and $\mathbb{C}^2$.

- For any vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, its norm (distance from 0) is

$$||x|| = \sqrt{\langle x, x \rangle} = \sqrt{x^T x} = \sqrt{x_1^2 + x_2^2}.$$

- For any $z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \in \mathbb{C}^2$, with $z_1 = a + bi$, $z_2 = c + di$, its norm is defined by

$$||z|| = \sqrt{\langle z, z \rangle} := \sqrt{\overline{z}^T z} = \sqrt{|z_1|^2 + |z_2|^2}.$$

For example, let's compute the norms of $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2$ and $z = \begin{bmatrix} i \\ i \end{bmatrix} \in \mathbb{C}^2$.

Complex dot product

Definition

If X is a finite-dimensional vector space over $\mathbb{C}$, then define the complex dot product as

$$\langle z, w\rangle = w^H z := \overline{w}^T z = \begin{bmatrix} \overline{w_1} & \overline{w_2} & \cdots & \overline{w_n} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}.$$

Here, H stands for Hermitian.

The norm of a vector $z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$ in $\mathbb{C}^n$ is thus defined by

$$||z||^2 = \langle z, z\rangle = \overline{z}^T z = \begin{bmatrix} \overline{z_1} & \overline{z_2} & \cdots & \overline{z_n} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2.$$

Just like how we abstracted the dot product to a real inner product, we can abstract the complex dot product to a complex inner product.

Complex inner products and sesquilinear forms

Definition

A complex inner product space is a vector space X over $\mathbb{C}$ endowed with a map

$$\langle\ ,\ \rangle : X \times X \longrightarrow \mathbb{C}$$

satisfying

(i) $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ and $\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$

(ii) $\langle ku, v\rangle = k\langle u, v\rangle$ "*linear in the 1st coordinate*"

(iii) $\langle u, kv\rangle = \overline{k}\langle u, v\rangle$ "*antilinear in the 2nd coordinate*"

(iv) $\overline{\langle v, u\rangle} = \langle u, v\rangle$ "*Hermitian*"

(v) $\langle u, u\rangle > 0$ if $u \neq 0$, "*positive-definite*"

for all $u, v, w \in X$ and $k \in \mathbb{C}$.

Conditions (i)–(iii) are called sesquilinear. [Latin prefix *sesqui-* means "one and a half".]

A map satisfying (i)–(iv) is called a symmetric sesquilinear, or complex Hermitian form.

Adjoints and orthogonality in complex spaces

Let X and U be complex inner product spaces.

For any vectors x and y,

$$||x+y||^2 = ||x||^2 + \langle x,y\rangle + \langle y,x\rangle + ||y||^2 = ||x||^2 + 2\Re\langle x,y\rangle + ||y||^2.$$

Most results for real spaces carry over to complex spaces; just replace T with H.

The adjoint of a linear map $A\colon X \to U$ is the map $A^*\colon U \to X$ such that

$$\langle x, A^*u\rangle = \langle Ax, u\rangle, \qquad \forall x \in X,\ u \in U.$$

Proposition

With respect to the complex dot product $\langle z,w\rangle = w^H z$, the adjoint of $A\colon X \to U$ is its conjugate transpose, $A^* = A^H := \overline{A}^T$.

Two vectors x, y are orthogonal if $\langle x,y\rangle = 0$. The vectors $x_1, \dots, x_k$ in X are orthonormal if

$$\langle x_i, x_j\rangle = x_j^H x_i = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Unitary maps

Recall that an isometry of a real inner product space fixing 0 is called orthogonal.

An isometry of a complex inner product space fixing 0 is called unitary.

The matrix A is orthogonal if $A^T A = I$, and unitary if $A^H A = I$.

Note that

- orthogonal means $A^* = A^{-1}$ in an $\mathbb{R}$-vector space
- unitary means $A^* = A^{-1}$ in a $\mathbb{C}$-vector space.

Proposition

Let $U \colon X \to X$ be unitary.

(i) U is linear

(ii) $U^* U = I$ (and conversely)

(iii) U is invertible, and U^{-1} is an isometry

(iv) $|\det U| = 1$.

The unitary maps form the unitary group, denoted $U(n)$ or U_n. The special unitary group $SU(n)$ are those with determinant 1.

Complex Fourier series

Consider the space $X = \mathrm{Per}_{2\pi}(\mathbb{C})$ of 2π-periodic complex-valued functions.

We can define an inner product as

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx.$$

The set

$$\Big\{e^{inx} \mid n \in \mathbb{Z}\Big\} = \Big\{\ldots, e^{-2ix}, e^{-ix}, 1, e^{ix}, e^{2ix}, \ldots\Big\}$$

is an orthonormal basis w.r.t. to this inner product.

Thus, we can write each $f(x) \in \mathrm{Per}_{2\pi}(\mathbb{C})$ *uniquely* as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx} = c_0 + \sum_{n=1}^{\infty} c_n e^{inx} + c_{-n}e^{-inx},$$

where

$$c_n = \mathrm{proj}_{e^{inx}}\big(f\big) = \langle f, e^{inx}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\,dx.$$