ECE 586: Vector Space Methods
Chapter 3: Linear Algebra

Henry D. Pfister
Duke University

September 18th–, 2019

3.1: Fields

Consider a set F of objects with binary operations addition and mult.:

- For every pair of elements $s, t \in F$ then $(s + t) \in F$
- For every pair of elements $s, t \in F$ then $st \in F$

F forms a field if the two operations satisfy:

1. addition is commutative: $s + t = t + s \quad \forall s, t \in F$
2. addition is associative: $r + (s + t) = (r + s) + t \quad \forall r, s, t \in F$
3. for $s \in F$ there is a unique element $(-s) \in F$ such that $s + (-s) = 0$
4. multiplication is commutative: $st = ts \quad \forall s, t \in F$
5. multiplication is associative: $r(st) = (rs)t \quad \forall r, s, t \in F$
6. there is a unique non-zero element $1 \in F$ such that $s1 = s \quad \forall s \in F$
7. for $s \in F \setminus \{0\}$ there is a unique element $s^{-1} \in F$ such that $ss^{-1} = 1$
8. mult. distributes over addition: $r(s + t) = rs + rt \quad \forall r, s, t \in F$.

Example

The complex numbers with standard addition and mult. are a field.

Example

The set of integers with standard addition and mult. is not a field.

3.2: Matrices

Consider finding n scalars $x_1, \dots, x_n \in F$ that satisfy the conditions

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & y_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & y_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & y_m \end{array}$$

This defines a system of m linear equations in n unknowns $A\underline{x} = \underline{y}$, where $\underline{x} = (x_1, \dots, x_n)^T$, $\underline{y} = (y_1, \dots, y_m)^T$, and A is the matrix given by

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

3.2: Matrix Multiplication

Definition

Let A be an $m \times n$ matrix over F and let B be an $n \times p$ matrix over F. The matrix product AB is the $m \times p$ matrix C whose i, j entry is

$$c_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}.$$

Example

When j is fixed, one can eliminate i by grouping the elements of C and A into column vectors $\underline{c}_1, \dots, \underline{c}_p$ and $\underline{a}_1, \dots, \underline{a}_n$. For this case, we see that the j-th column of C is a linear combination of the columns of A,

$$\underline{c}_j = \sum_{r=1}^{n} \underline{a}_r b_{rj},$$

Also, grouping C and B into row vectors $\underline{c}_1, \dots, \underline{c}_m$ and $\underline{b}_1, \dots, \underline{b}_n$ gives

$$\underline{c}_i = \sum_{r=1}^{n} a_{ir} \underline{b}_r.$$

3.2: Echelon Forms

Definition

An matrix $A \in F^{m \times n}$ is in row echelon form if:

1. Any rows containing only zeros are below all non-zero rows, and
2. For non-zero rows, the leading coefficient (i.e., the first non-zero element from the left) is strictly to the right of the leading coefficient of the row above it.

Thus, the entries below the leading coefficient in a column are zero. A is in column echelon form if A^T is in row echelon form.

Definition

A matrix $A \in F^{m \times n}$ is in reduced row echelon form if it is:

1. in row echelon form with every leading coefficient equal to 1, and
2. every leading coefficient is the only non-zero element in its column.

A is reduced column echelon form if A^T is in reduced row echelon form.

3.2: Elementary Row Operations

Definition

An elementary row operation on an $m \times n$ matrix consists of

1. multiplying a row by a non-zero scalar,
2. swapping two rows, or
3. adding a non-zero scalar multiple of one row to another row.

An elementary column operation is the same but applied to the columns.

Lemma

For any $m \times n$ matrix A over F, there is an invertible $m \times m$ matrix P over F such that $R = PA$ is in reduced row echelon form.

Sketch of Proof.

To construct P, one applies Gaussian elimination to the augmented matrix $A' = [A \;\; I]$. The resulting augmented matrix $R' = [R \;\; P]$ (and, thus R) is in reduced row echelon form. Since elementary row operations defined by (invertible) matrix multiplies on the left side, one also finds that $R' = PA'$, $R = PA$, and P is invertible. $\square$

3.2: Foundation of Dimension

Lemma

Let A be an $m \times n$ matrix over F with $m < n$. Then, there exists a length-n column vector $\underline{x} \neq \underline{0}$ (over F) such that $A\underline{x} = \underline{0}$.

Proof.

- Apply steps below to concrete example on whiteboard.
- Use row reduction to compute the reduced row echelon form $R = PA$ of A, where P is invertible by previous lemma.
- Observe that the columns of R containing leading elements can be combined in a linear combination to cancel any other column of R.
- Thus, one can construct a vector $\underline{x}$ satisfying $R\underline{x} = \underline{0}$ and thus $A\underline{x} = P^{-1}R\underline{x} = \underline{0}$. □

3.3: Vector Spaces

Definition

A vector space consists of the following,

1. a **field F** of scalars
2. a **set V** of objects, called vectors
3. a **binary operation called vector addition**, which maps any pair of vectors $\underline{v}, \underline{w} \in V$ to a vector $\underline{v} + \underline{w} \in V$ such that
   1. addition is commutative: $\underline{v} + \underline{w} = \underline{w} + \underline{v}$
   2. addition is associative: $\underline{u} + (\underline{v} + \underline{w}) = (\underline{u} + \underline{v}) + \underline{w}$
   3. there is a unique vector $\underline{0} \in V$ such that $\underline{v} + \underline{0} = \underline{v}$, $\forall \underline{v} \in V$
   4. to each $\underline{v} \in V$ there is a unique vector $-\underline{v} \in V$ such that $\underline{v} + (-\underline{v}) = \underline{0}$
4. a **binary operation called scalar multiplication**, which maps any $s \in F$ and $\underline{v} \in V$ to a vector $s\underline{v} \in V$ such that
   1. $1\underline{v} = \underline{v}$, $\forall \underline{v} \in V$
   2. $(s_1 s_2)\underline{v} = s_1(s_2\underline{v})$
   3. $s(\underline{v} + \underline{w}) = s\underline{v} + s\underline{w}$
   4. $(s_1 + s_2)\underline{v} = s_1\underline{v} + s_2\underline{v}$.

3.3: Vector Space Examples

Example (Standard vector space for F^n)

Let F be a field, and let $V = F^n$ be the set of n-tuples $\underline{v} = (v_1, \ldots, v_n)$. If $\underline{w} = (w_1, \ldots, w_n) \in F^n$, the sum of $\underline{v}$ and $\underline{w}$ is defined by

$$\underline{v} + \underline{w} = (v_1 + w_1, \ldots, v_n + w_n).$$

The scalar product of $s \in F$ and $\underline{v} \in V$ is defined $s\underline{v} = (sv_1, \ldots, sv_n)$.

Example (Standard vector space of functions)

Let X be a non-empty set and let Y be a vector space over F. Consider the set V of all functions mapping X to Y. The vector addition of two functions $f, g \in V$ is the function from X into Y defined by

$$(f + g)(x) = f(x) + g(x) \quad \forall x \in X,$$

where the RHS uses vector addition from Y. The scalar product of $s \in F$ and the function $f \in V$ is the function sf defined by $(sf)(x) = sf(x)$ for all $x \in X$, where the RHS uses scalar multiplication from Y.

3.3.1: Subspaces

Definition

Let V be a vector space over F. A subspace of V is a subset $W \subset V$ which is itself a vector space over F.

Lemma

A non-empty subset $W \subset V$ is a subspace of V if and only if for every pair $\underline{w}_1, \underline{w}_2 \in W$ and every scalar $s \in F$ the vector $s\underline{w}_1 + \underline{w}_2 \in W$.

Sketch proof via inheritance from from V.

Example

Let A be an $m \times n$ matrix over F. The set of all $n \times 1$ column vectors V such that

$$\underline{v} \in V \implies A\underline{v} = \underline{0}$$

is a subspace of $F^{n\times 1}$.

3.3: Linear Combinations

Definition

A vector $\underline{w} \in V$ is said to be a linear combination of the vectors $\underline{v}_1, \ldots, \underline{v}_n \in V$ provided that there exist scalars $s_1, \ldots, s_n \in F$ such that

$$\underline{w} = \sum_{i=1}^{n} s_i \underline{v}_i.$$

Definition

Let U be a list (or set) of vectors in V. The span of U, denoted $\mathrm{span}(U)$, is defined to be the set of all finite linear combinations of vectors in U.

Example

For a vector space V, the span of any list (or set) of vectors in V form a subspace.

3.3.2: Linear Dependence and Independence

Definition

Let V be a vector space over F. A list of vectors $\underline{u}_1, \ldots, \underline{u}_n \in V$ is called linearly dependent if there are scalars $s_1, \ldots, s_n \in F$, not all equal to 0, such that

$$\sum_{i=1}^{n} s_i \underline{u}_i = 0.$$

A list that is not linearly dependent is called linearly independent.

Similarly, a subset $U \subset V$ is called linearly dependent if there is a finite list $\underline{u}_1, \ldots, \underline{u}_n \in U$ of distinct vectors that is linearly dependent. Otherwise, it is called linearly independent.

Example

For $V = \mathbb{R}^4$, vectors $\underline{v}_1 = (1, 1, 0, 0)$, $\underline{v}_2 = (0, 1, 1, 0)$, $\underline{v}_3 = (0, 0, 1, 1)$ are linearly independent because $\underline{u} = s_1\underline{v}_1 + s_2\underline{v}_2 + s_3\underline{v}_3$ is non-zero if $s_1 \neq 0$ or $s_3 \neq 0$ or $(s_2 \neq 0 \text{ and } s_1 = 0)$.

3.3.2: Basis

Definition

Let V be a vector space over F. Let $\mathcal{B} = \{\underline{v}_\alpha | \alpha \in A\}$ be a subset of linearly independent vectors from V such that every $\underline{v} \in V$ can be written as a finite linear combination of vectors from $\mathcal{B}$. Then, the set $\mathcal{B}$ is a Hamel basis for V. If V has a finite basis, it is called finite-dimensional.

From this, a basis decomposition $\underline{v} = \sum_{i=1}^{n} s_i \underline{v}_{\alpha_i}$ must be unique:

The difference between any two distinct decompositions produces a finite linear dependency in the basis and, hence, a contradiction.

Theorem

Every vector space has a Hamel basis.

3.3.2: Standard Basis

Example

Let F be a field and let $U \subset F^n$ be the subset consisting of the vectors $\underline{e}_1, \ldots, \underline{e}_n$ defined by

$$\begin{aligned} \underline{e}_1 &= (1, 0, \ldots, 0) \\ \underline{e}_2 &= (0, 1, \ldots, 0) \\ \vdots &= \qquad \vdots \\ \underline{e}_n &= (0, 0, \ldots, 1). \end{aligned}$$

For any $\underline{v} = (v_1, \ldots, v_n) \in F^n$, we have

$$\underline{v} = \sum_{i=1}^{n} v_i \underline{e}_i. \tag{1}$$

Thus, the collection $U = \{\underline{e}_1, \ldots, \underline{e}_n\}$ spans F^n. Since $\underline{v} = \underline{0}$ in (1) if and only if $v_1 = \cdots = v_n = 0$, U is linearly independent. Accordingly, the set U is a basis for $F^{n \times 1}$. This basis is termed the standard basis of F^n.

3.3.2: Dimension

Theorem

Let V be a finite-dimensional vector space that is spanned by a finite set of vectors $W = \{\underline{w}_1, \ldots, \underline{w}_n\}$. If $U = \{\underline{u}_1, \ldots, \underline{u}_m\} \subset V$ is a linearly independent set of vectors, then $m \leq n$.

Proof on whiteboard.

Definition

The dimension of a finite-dimensional vector space is the number of elements in any basis for V. It is denoted by $\dim(V)$.

3.3.2: Invertibility

Lemma

Let $A \in F^{n\times n}$ be an invertible matrix. Then, the columns of A form a basis for F^n. Similarly, the rows of A will also form a basis for F^n

Proof on whiteboard.

Theorem

Let A be an $n \times n$ matrix over F whose columns, denoted by $\underline{a}_1, \ldots, \underline{a}_n$, form a linearly independent set of vectors in F^n. Then A is invertible.

Proof on whiteboard.

3.4: Linear Transforms

Definition

Let V and W be vector spaces over a field F. A linear transform from V to W is a function T from V into W such that

$$T\left(s\underline{v}_1 + \underline{v}_2\right) = sT\underline{v}_1 + T\underline{v}_2$$

for all $\underline{v}_1$ and $\underline{v}_2$ in V and all scalars s in F.

Example

Let A be a fixed $m \times n$ matrix over F. The function T defined by $T\left(\underline{v}\right) = A\underline{v}$ is a linear transformation from $F^{n\times 1}$ into $F^{m\times 1}$.

Example

Let V be the space of continuous functions from $[0,1]$ to $\mathbb{R}$. Define T by

$$(Tf)(x) = \int_0^x f(t)dt.$$

Then, T is a linear transform from V to V because Tf is continuous.

3.4.2: Properties of Linear Transforms

Definition (Range)

For a linear transformation $T\colon V \to W$, the range of T is the set of all vectors $\underline{w} \in W$ such that $\underline{w} = T\underline{v}$ for some $\underline{v} \in V$. It is denoted by

$$\mathcal{R}(T) \triangleq \{\underline{w} \in W | \exists \underline{v} \in V \text{ s.t. } T\underline{v} = \underline{w}\} = \{T\underline{v} | \underline{v} \in V\}.$$

Definition (Nullspace)

For a linear transformation $T\colon V \to W$, the nullspace of T is the set of all vectors $\underline{v} \in V$ such that $T\underline{v} = \underline{0}$. We denote the nullspace of T by

$$\mathcal{N}(T) \triangleq \{\underline{v} \in V | T\underline{v} = \underline{0}\}.$$

Theorem

Let V, W be vector spaces over F and $\mathcal{B} = \{\underline{v}_\alpha | \alpha \in A\}$ be a Hamel basis for V. For each mapping $G\colon \mathcal{B} \to W$, there is a unique linear transformation $T\colon V \to W$ such that $T\underline{v}_\alpha = G\left(\underline{v}_\alpha\right)$ for all $\alpha \in A$.

Proof on whiteboard.

3.4.2: Rank and Nullity

Definition (Rank and Nullity)

Let V and W be vector spaces over a field F, and let T be a linear transformation from V into W. The rank of T is the dimension of the range of T and the nullity of T is the dimension of the nullspace of T.

Theorem (Rank-Nullity)

Let V and W be vector spaces over the field F and let T be a linear transformation from V into W. If V is finite-dimensional, then

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = \dim(V)$$

Proof on whiteboard.

Theorem

If A is an $m \times n$ matrix with entries in the field F, then

$$\text{row rank}(A) \triangleq \dim(\mathcal{R}(A^T)) = \dim(\mathcal{R}(A)) \triangleq \mathrm{rank}(A).$$

Proof on whiteboard.

3.5: Normed Vector Spaces

Let V be a vector space over the real numbers or the complex numbers.

Definition

A norm on vector space V is a real-valued function $\|\cdot\| : V \to \mathbb{R}$ that satisfies the following properties.

1. $\|\underline{v}\| \geq 0 \quad \forall \underline{v} \in V$; equality holds if and only if $\underline{v} = \underline{0}$
2. $\|s\underline{v}\| = |s| \, \|\underline{v}\| \quad \forall \underline{v} \in V, s \in F$
3. $\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\| \quad \forall \underline{v}, \underline{w} \in V$.

The concept of a norm is closely related to that of a metric. For instance, a metric can be defined from any norm.

Let $\|\underline{v}\|$ be a norm on vector space V, then the induced metric is

$$d\left(\underline{v}, \underline{w}\right) = \|\underline{v} - \underline{w}\|.$$

3.5: Examples of Normed Vector Spaces

Example (Standard Norms for Real/Complex Vector Spaces)

The following functions are examples of norms for $\mathbb{R}^n$ and $\mathbb{C}^n$:

1. the l^1 norm: $\|\underline{v}\|_1 = \sum_{i=1}^n |v_i|$
2. the l^p norm: $\|\underline{v}\|_p = \left(\sum_{i=1}^n |v_i|^p\right)^{\frac{1}{p}}, \quad p \in (1, \infty)$
3. the l^∞ norm: $\|\underline{v}\|_\infty = \max_{1,\ldots,n}\{|v_i|\}$

Example (Standard Norms for Real/Complex Function Spaces)

Similarly, for the vector space of functions from $[a, b]$ to $\mathbb{R}$ (or $\mathbb{C}$):

1. the L^1 norm: $\|f(t)\|_1 = \int_a^b |f(t)| dt$
2. the L^p norm: $\|f(t)\|_p = \left(\int_a^b |f(t)|^p dt\right)^{\frac{1}{p}}, \quad p \in (1, \infty)$
3. the L^∞ norm: $\|f(t)\|_\infty = \text{esssup}_{[a,b]}\{|f(t)|\}$

For infinite dimensional spaces, only vectors with finite norm are included.

3.5: Norms Versus Metrics

Example

Consider vectors in $\mathbb{R}^n$ with the euclidean metric

$$d(\underline{v}, \underline{w}) = \sqrt{(v_1 - w_1)^2 + \cdots + (v_n - w_n)^2}.$$

Recall the bounded metric given by

$$\bar{d}(\underline{v}, \underline{w}) = \min\{d(\underline{v}, \underline{w}), 1\}.$$

Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(\underline{v}) = \bar{d}(\underline{v}, \underline{0})$. Is the function f a norm?

By the properties of a metric, we have

1. $\bar{d}(\underline{v}, \underline{0}) \geq 0 \quad \forall \underline{v} \in V$; equality holds if and only if $\underline{v} = \underline{0}$
2. $\bar{d}(\underline{v}, \underline{0}) + \bar{d}(\underline{w}, \underline{0}) = \bar{d}(\underline{v}, \underline{0}) + \bar{d}(\underline{0}, \underline{w}) \geq \bar{d}(\underline{v}, \underline{w}) \quad \forall \underline{v}, \underline{w} \in V.$

However, $\bar{d}(s\underline{v}, \underline{0})$ is not always equal to $s\bar{d}(\underline{v}, \underline{0})$. For instance, $\bar{d}(2\underline{e}_1, \underline{0}) = 1 < 2\bar{d}(\underline{e}_1, \underline{0})$. Thus, the $f(\underline{v}) = \bar{d}(\underline{v}, \underline{0})$ is not a norm.

3.5: Complete Normed Spaces

Definition

A vector $\underline{v} \in V$ is called normalized if $\|\underline{v}\| = 1$. Any non-zero $\underline{v}$ can be normalized:

$$\underline{u} = \frac{\underline{v}}{\|\underline{v}\|}$$

has norm $\|\underline{u}\| = 1$. A normalized vector is called a unit vector.

Definition

A complete normed vector space is called a Banach space.

Example

Vector spaces $\mathbb{R}^n$ (or $\mathbb{C}^n$) with any well-defined norm are Banach spaces.

Example

The vector space of all continuous functions from $[a, b]$ to $\mathbb{R}$ is a Banach space under the supremum norm

$$\|f(t)\| = \sup_{t \in [a,b]} f(t).$$

3.5: Schauder Basis

Definition

A Banach space V has a Schauder basis, $\underline{v}_1, \underline{v}_2, \ldots$, if every $\underline{v} \in V$ can be written uniquely as

$$\underline{v} = \sum_{i=1}^{\infty} s_i \underline{v}_i,$$

where convergence is determined by the norm topology.

Example

Let $V = \mathbb{R}^\infty$ be the vector space of semi-infinite real sequences. The standard Schauder basis is the countably infinite extension $\{\underline{e}_1, \underline{e}_2, \ldots\}$ of the standard basis.

3.5: Convergence of Sums

Banach space convergence via the induced metric $d(\underline{v},\underline{w}) = \|\underline{v}-\underline{w}\|$.

Lemma

If $\sum_{i=1}^{\infty} \|\underline{v}_i\| = a < \infty$, then $\underline{u}_n = \sum_{i=1}^{n} \underline{v}_i$ satisfies $\underline{u}_n \to \underline{u}$ with $\|\underline{u}\| \leq a$.

Proof.

- Let $a_n = \sum_{i=1}^{n} \|\underline{v}_i\|$ and observe that, for $n > m$,

$$|a_n - a_m| = \left| \sum_{i=1}^{n} \|\underline{v}_i\| - \sum_{i=1}^{m} \|\underline{v}_i\| \right| = \sum_{i=m+1}^{n} \|\underline{v}_i\|$$

$$\|\underline{u}_n - \underline{u}_m\| = \left\| \sum_{i=1}^{n} \underline{v}_i - \sum_{i=1}^{m} \underline{v}_i \right\| = \left\| \sum_{i=m+1}^{n} \underline{v}_i \right\| \leq \sum_{i=m+1}^{n} \|\underline{v}_i\|$$

- Since $\sum_{i=1}^{\infty} \|\underline{v}_i\|$ converges in $\mathbb{R}$, a_n must be a Cauchy sequence
- Since $\|\underline{u}_n - \underline{u}_m\| \leq |a_n - a_m|$, $\underline{u}_n$ is also a Cauchy sequence
- The norm bound follows from the triangle inequality □

3.5: Open and Closed Subspaces

Definition

A closed subspace of a Banach space is a subspace that is a closed set in the topology generated by the norm.

Theorem

All finite dimensional subspaces of a Banach space are closed.

Example

Let $W = \{\underline{w}_1, \underline{w}_2, \ldots\}$ be a linearly independent sequence of normalized vectors in a Banach space. The span of W only includes finite linear combinations. However, a sequence of finite linear combinations, like

$$\underline{u}_n = \sum_{i=1}^{n} \frac{1}{i^2} \underline{w}_i,$$

converges to $\lim_{n\to\infty} \underline{u}_n$ if it exists. Thus, the span of W is not closed.

Show convergence on whiteboard.

6.1/6.3: Vector Space of Linear Transforms and Norms

Definition

Let $L(V, W)$ denote the vector space of all linear transforms from V into W, where V and W are vector spaces over a field F.

An operator norm is a norm on a vector space of linear transforms.

Definition (Induced Operator Norm)

Let V and W be two normed vector spaces and let $T \colon V \to W$ be a linear transformation. The induced operator norm of T is defined to

$$\|T\|_{\mathrm{op}} = \sup_{\underline{v} \in V - \{\underline{0}\}} \frac{\|T\underline{v}\|_W}{\|\underline{v}\|_V} = \sup_{\underline{v} \in V, \|\underline{v}\|_V = 1} \|T\underline{v}\|_W .$$

A common question about the operator norm is, "How do I know the two expressions give the same result?". To see this, we can write

$$\sup_{\underline{v} \in V - \{\underline{0}\}} \frac{\|T\underline{v}\|_W}{\|\underline{v}\|_V} = \sup_{\underline{v} \in V - \{\underline{0}\}} \left\| T \frac{\underline{v}}{\|\underline{v}\|_V} \right\|_W = \sup_{\underline{u} \in V, \|\underline{u}\|_V = 1} \|T\underline{u}\|_W .$$

6.3: Operator Norms

The induced operator norm has another property:

$$\|T\| = \sup_{\underline{v} \in V - \{\underline{0}\}} \frac{\|T\underline{v}\|}{\|\underline{v}\|} \geq \frac{\|T\underline{u}\|}{\|\underline{u}\|},$$

for non-zero $\underline{u} \in V$. This implies that $\|T\underline{u}\| \leq \|T\|\|\underline{u}\|$ for all non-zero $\underline{u} \in V$. By checking $\underline{u} = \underline{0}$ separately, one can see it holds for all $\underline{u} \in V$.

Definition

For the space $L(V, V)$ of linear operators on V, a norm is called submultiplicative if $\|TU\| \leq \|T\|\|U\|$ for all $T, U \in L(V, V)$.

Using the above inequality, induced operator norms are submultiplicative:

$$\|UT\underline{v}\| \leq \|U\| \, \|T\underline{v}\| \leq \|U\| \, \|T\| \, \|\underline{v}\| .$$

6.3.3: Matrix Norms

A norm on a vector space of matrices is called a matrix norm.

Definition

For $A \in F^{m\times n}$, the matrix norm induced by the ℓ^p vector norm $\|\cdot\|_p$, is:

$$\|A\|_p \triangleq \max_{\|\underline{v}\|_p=1} \|A\underline{v}\|_p .$$

In special cases, there are exact formulae:

$$\|A\|_\infty = \max_i \sum_j |a_{ij}|$$

$$\|A\|_1 = \max_j \sum_i |a_{ij}|$$

$$\|A\|_2 = \sqrt{\rho(A^H A)},$$

where the $\rho(B)$ is the maximum absolute value of all eigenvalues.

Examples on whiteboard.

6.3.2: Neumann Expansion

Theorem

Let $\|\cdot\|$ be a submultiplicative operator norm and $T\colon V \to V$ be a linear operator with $\|T\| < 1$. Then, $(I-T)^{-1}$ exists and

$$(I-T)^{-1} = \sum_{i=0}^{\infty} T^i.$$

Proof.

- Observe that $\sum_{i=0}^{\infty} \|T^i\| \leq \sum_{i=0}^{\infty} \|T\|^i = \frac{1}{1-\|T\|}$ because $\|T\| < 1$
- Recall an infinite sum $\sum_{i=0}^{\infty} T^i$ converges if $\sum_{i=0}^{\infty} \|T^i\|$ converges
- Next, observe that $(I-T)\left(I+T+T^2+\cdots+T^{n-1}\right) = I - T^n$
- $T^n \to 0$ because $\|T^n\| \leq \|T\|^n \to 0$ since $\|T\| < 1$
- Thus, $\sum_{i=0}^{\infty} T^i$ is a right inverse for $I-T$
- Same argument works on the left, so we're done. $\square$

3.6: Inner-Product Spaces

Definition

Let F be the field of real numbers or the field of complex numbers, and assume V is a vector space over F. An inner product on V is a function which assigns to each ordered pair of vectors $\underline{v}, \underline{w} \in V$ a scalar $\langle \underline{v} | \underline{w} \rangle \in F$ in such a way that for all $\underline{u}, \underline{v}, \underline{w} \in V$ and any scalar $s \in F$

1. $\langle \underline{u} + \underline{v} | \underline{w} \rangle = \langle \underline{u} | \underline{w} \rangle + \langle \underline{v} | \underline{w} \rangle$
2. $\langle s\underline{v} | \underline{w} \rangle = s \langle \underline{v} | \underline{w} \rangle$
3. $\langle \underline{v} | \underline{w} \rangle = \overline{\langle \underline{w} | \underline{v} \rangle}$, where the overbar denotes complex conjugation;
4. $\langle \underline{v} | \underline{v} \rangle \geq 0$ with equality iff $\underline{v} = \underline{0}$.

Note that these conditions imply that:

$$\langle s\underline{v} + \underline{w} | \underline{u} \rangle = s \langle \underline{v} | \underline{u} \rangle + \langle \underline{w} | \underline{u} \rangle$$
$$\langle \underline{u} | s\underline{v} + \underline{w} \rangle = \bar{s} \langle \underline{u} | \underline{v} \rangle + \langle \underline{u} | \underline{w} \rangle$$

3.6: Example Inner Products

Example (Standard Inner Product on F^n)

Consider the inner product on F^n defined by

$$\langle \underline{v} | \underline{w} \rangle = \langle (v_1, \ldots, v_n) | (w_1, \ldots, w_n) \rangle \triangleq \sum_{j=1}^{n} v_j \overline{w}_j.$$

For column vectors, it follows that $\langle \underline{v} | \underline{w} \rangle = \underline{w}^H \underline{v}$

Example (Standard Inner Product on a Function Space)

Let V be the vector space of all continuous complex-valued functions on the unit interval $[0, 1]$. Then, the following defines an inner product

$$\langle f | g \rangle = \int_0^1 f(t) \overline{g(t)} \, dt$$

Example (Inner Product on Space of Random Variables)

Let W be a set of real-valued random variables with finite 2nd moments. Then, $V = \mathrm{span}(W)$ is a vector space over $\mathbb{R}$ with inner product

$$\langle X | Y \rangle = \mathrm{E}[XY]$$

3.6: Properties of the Inner Product (1)

Theorem

Let V be a finite-dimensional space with ordered basis $\mathcal{B} = \underline{w}_1, \ldots, \underline{w}_n$. Then, any inner product on V is determined by the values

$$g_{ij} = \langle \underline{w}_j | \underline{w}_i \rangle .$$

Proof.

If $\underline{u} = \sum_j s_j \underline{w}_j$ and $\underline{v} = \sum_i t_i \underline{w}_i$, then

$$\begin{aligned} \langle \underline{u} | \underline{v} \rangle &= \left\langle \sum_j s_j \underline{w}_j \middle| \underline{v} \right\rangle = \sum_j s_j \langle \underline{w}_j | \underline{v} \rangle \\ &= \sum_j s_j \left\langle \underline{w}_j \middle| \sum_i t_i \underline{w}_i \right\rangle = \sum_j \sum_i s_j \bar{t}_i \langle \underline{w}_j | \underline{w}_i \rangle \\ &= \sum_j \sum_i \bar{t}_i g_{ij} s_j = [\underline{v}]_{\mathcal{B}}^H G [\underline{u}]_{\mathcal{B}} \end{aligned}$$

where $[\underline{u}]_{\mathcal{B}} = (s_1, \ldots, s_n)$ and $[\underline{v}]_{\mathcal{B}} = (t_1, \ldots, t_n)$ are the coordinate matrices of $\underline{u}$, $\underline{v}$ in the ordered basis $\mathcal{B}$. The matrix G is called the weight matrix of the inner product in the ordered basis $\mathcal{B}$. □

3.6: Properties of the Inner Product (2)

- From the property $\langle \underline{v} | \underline{w} \rangle = \overline{\langle \underline{w} | \underline{v} \rangle}$, it follows that
  - the weight matrix G of an inner product is Hermitian: $G = G^H$.
- Using $\langle \underline{v} | \underline{v} \rangle \geq 0$, we see that $\langle \underline{v} | \underline{v} \rangle = \underline{w}^H G \underline{w} > 0$ for all $\underline{w} \neq \underline{0}$
  - A Hermitian matrix satisfying this is called positive definite
- If G is an $n \times n$ matrix that is Hermitian and positive definite, then:
  - G is the weight matrix (in ordered basis $\mathcal{B}$) of the inner product

$$\langle \underline{u} | \underline{v} \rangle_G = [\underline{v}]_{\mathcal{B}}^H G [\underline{u}]_{\mathcal{B}} .$$

Definition (Orthogonal)

Let $\underline{v}$ and $\underline{w}$ be vectors in inner-product space V. Then, $\underline{v}$ is orthogonal to $\underline{w}$ (denoted $\underline{v} \perp \underline{w}$) iff $\langle \underline{v} | \underline{w} \rangle = 0$. Since this implies $\langle \underline{w} | \underline{v} \rangle = 0$, $\underline{w}$ is also orthogonal to $\underline{v}$, we simply say that $\underline{v}$ and $\underline{w}$ are orthogonal.

3.6.1: Induced Norm

Definition (Induced Norm)

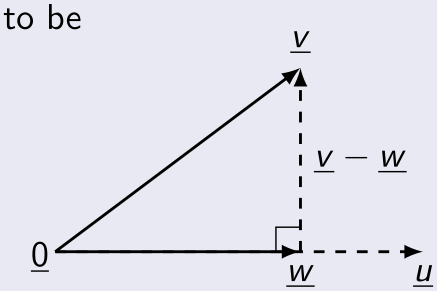
Let V be an inner-product space with inner product $\langle \cdot | \cdot \rangle$. This inner product naturally defines the induced norm

$$\|\underline{v}\| = \langle \underline{v} | \underline{v} \rangle^{\frac{1}{2}}.$$

Definition (Projection)

Let $\underline{u}, \underline{v}$ be vectors in an inner-product space V with inner product $\langle \cdot | \cdot \rangle$. The projection of $\underline{v}$ onto $\underline{u}$ is defined to be

$$\underline{w} = \frac{\langle \underline{v} | \underline{u} \rangle}{\|\underline{u}\|^2} \underline{u}$$



3.6.1: Properties of the Induced Norm

Theorem

If V is an inner-product space over F and $\|\underline{v}\| \triangleq \sqrt{\langle \underline{v} | \underline{v} \rangle}$, then for any $\underline{v}, \underline{w} \in V$ and any $s \in F$, it follows that

1. $\|s\underline{v}\| = |s| \, \|\underline{v}\|$
2. $\|\underline{v}\| > 0$ *for* $\underline{v} \neq \underline{0}$
3. $|\langle \underline{v} | \underline{w} \rangle| \leq \|\underline{v}\| \, \|\underline{w}\|$ *with equality iff* $\underline{v} = \underline{0}$, $\underline{w} = \underline{0}$, *or* $\underline{v} = s\underline{w}$
4. $\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\|$ *with equality iff* $\underline{v} = \underline{0}$, $\underline{w} = \underline{0}$, *or* $\underline{v} = s\underline{w}$.

Sketch of Proof.

The first two follow immediately from definitions. The third inequality, $|\langle \underline{v} | \underline{w} \rangle| \leq \|\underline{v}\| \, \|\underline{w}\|$, is called the Cauchy-Schwarz inequality. The fourth inequality is the triangle inequality for the induced norm and can be shown using the Cauchy-Schwarz inequality. □

Proof of Cauchy-Schwarz on whiteboard.

3.7: Sets of Orthogonal Vectors

Definition

Let V be an inner-product space and U, W be subspaces. Then, the subspace U is an orthogonal to the subspace W (denoted $U \perp W$) if:

$$\underline{u} \perp \underline{w} \text{ for all } \underline{u} \in U \text{ and } \underline{w} \in W.$$

Definition

A subset $W \subset V$ of vectors is an orthogonal set if all distinct pairs in W are orthogonal. A orthogonal set is orthonormal if all vectors normalized.

Example

For $\mathbb{R}^n$ with standard inner product, the standard basis is an orthonormal.

Example

Let V be the vector space (over $\mathbb{C}$) of continuous functions $f : [0, 1] \to \mathbb{C}$ with the standard inner product. Let $f_n(x) = \sqrt{2} \cos 2\pi nx$ and $g_n(x) = \sqrt{2} \sin 2\pi nx$. Then, $\{1, f_1, g_1, f_2, g_2, \ldots\}$ is a countably infinite orthonormal set and a Schauder basis for the closure of V.

3.7: Properties of Orthogonal Sets

Lemma

Let V be an inner-product space and $W \subset V$ be an orthogonal set of non-zero vectors. Let $\underline{v} = s_1 \underline{w}_1 + \cdots + s_n \underline{w}_n$ be a linear combination of distinct vectors in W. Then,

$$s_i = \frac{\langle \underline{v} | \underline{w}_i \rangle}{\|\underline{w}_i\|^2}$$

Proof.

The inner product $\langle \underline{v} | \underline{w}_i \rangle$ is given by

$$\langle \underline{v} | \underline{w}_i \rangle = \left\langle \textstyle\sum_j s_j \underline{w}_j | \underline{w}_i \right\rangle = \textstyle\sum_j s_j \left\langle \underline{w}_j | \underline{w}_i \right\rangle = s_i \left\langle \underline{w}_i | \underline{w}_i \right\rangle.$$

Dividing both sides by $\|\underline{w}_i\|^2 = \langle \underline{w}_i | \underline{w}_i \rangle > 0$, gives the stated result. $\square$

Theorem

An orthogonal set of non-zero vectors is linearly independent.

Proof by contradiction on whiteboard.

3.7: Gram-Schmidt Orthogonalization (1)

Gram-Schmidt Process

Let V be an inner-product space and assume $\underline{v}_1, \ldots, \underline{v}_n$ are linearly independent vectors in V. Then, an orthogonal set of vectors $\underline{w}_1, \ldots, \underline{w}_n$ with the same span is produced by Gram-Schmidt process:

1. Let $\underline{w}_1 = \underline{v}_1$
2. For $m = 1, \ldots, n-1$, define
$$\underline{w}_{m+1} = \underline{v}_{m+1} - \sum_{i=1}^{m} \frac{\left\langle \underline{v}_{m+1} | \underline{w}_i \right\rangle}{\left\|\underline{w}_i\right\|^2} \underline{w}_i.$$

- Vector $\underline{w}_{m+1} \neq 0$. Otherwise, $\underline{v}_{m+1}$ is a linear combination of $\underline{w}_1, \ldots, \underline{w}_m$ and hence a linear combination of $\underline{v}_1, \ldots, \underline{v}_m$
- Vectors $\underline{w}_{m+1}$ and $\underline{w}_j$ are orthogonal for $j = 1, \ldots, m$:
$$\begin{aligned}\left\langle \underline{w}_{m+1} | \underline{w}_j \right\rangle &= \left\langle \underline{v}_{m+1} | \underline{w}_j \right\rangle - \sum_{i=1}^{m} \frac{\left\langle \underline{v}_{m+1} | \underline{w}_i \right\rangle}{\left\|\underline{w}_i\right\|^2} \left\langle \underline{w}_i | \underline{w}_j \right\rangle \\ &= \left\langle \underline{v}_{m+1} | \underline{w}_j \right\rangle - \frac{\left\langle \underline{v}_{m+1} | \underline{w}_j \right\rangle}{\left\|\underline{w}_j\right\|^2} \left\langle \underline{w}_j | \underline{w}_j \right\rangle = 0\end{aligned}$$

3.7: Gram-Schmidt Orthogonalization (2)

Example

Let $V = \mathbb{R}^3$ be standard vector space equipped with the standard inner product and define
$$\begin{aligned}\underline{v}_1 &= (2,2,1) \\ \underline{v}_2 &= (3,6,0) \\ \underline{v}_3 &= (6,3,9)\end{aligned}$$

Applying the Gram-Schmidt process to $\underline{v}_1, \underline{v}_2, \underline{v}_3$ results in:

$$\underline{w}_1 = (2,2,1)$$
$$\begin{aligned}\underline{w}_2 &= (3,6,0) - \frac{\langle (3,6,0) | (2,2,1) \rangle}{9}(2,2,1) \\ &= (3,6,0) - 2(2,2,1) = (-1,2,-2)\end{aligned}$$
$$\begin{aligned}\underline{w}_3 &= \underline{v}_3 - \frac{\langle (6,3,9) | (2,2,1) \rangle}{9}(2,2,1) - \frac{\langle (6,3,9) | (-1,2,-2) \rangle}{9}(-1,2,-2) \\ &= (6,3,9) - 3(2,2,1) + 2(-1,2,-2) = (-2,1,2)\end{aligned}$$

It is easily verified that $\{\underline{w}_1, \underline{w}_2, \underline{w}_3\}$ is an orthogonal set of vectors.

3.7.1: Hilbert Spaces

Definition

A complete inner-product space is called a Hilbert space.

Example

Consider the Banach space ℓ^2 of infinite real/complex sequences with Euclidean norm $\|\underline{v}\| = \left(\sum_{i=1}^{\infty} |v_i|^2\right)^{1/2} < \infty$. The set ℓ^2 is a Hilbert space under the standard inner product because it induces the Euclidean norm.

Theorem

If Hilbert space V has a countable dense subset, then there is a linear transform $T : V \to \ell^2$ such that $\langle \underline{u} | \underline{v} \rangle_V = \langle T\underline{u} | T\underline{v} \rangle_{\ell^2}$ for all $\underline{u}, \underline{v} \in V$.

3.8: Linear Functionals and the Riesz Theorem

Definition

Let V be a vector space over a field F. A linear transformation f from V into the scalar field F is called a linear functional on V.

Example

Thus, $f : V \to F$ is a function on V such that

$$f(s\underline{v}_1 + \underline{v}_2) = sf(\underline{v}_1) + f(\underline{v}_2)$$

for all $\underline{v}_1, \underline{v}_2 \in V$ and $s \in F$.

Theorem (Riesz)

Let V be a Hilbert space and f be a continuous linear functional on V. Then, there exists a unique vector $\underline{v} \in V$ such that $f(\underline{w}) = \langle \underline{w} | \underline{v} \rangle$ for all $\underline{w} \in V$.

Derivatives in Banach Spaces

The foundation of engineering is the ability to use math and physics to design and optimize complex systems.

Computers have now made this possible on an unprecedented scale.

In vector analysis, derivatives are usually introduced using Banach spaces:

- For a function $f\colon X \to Y$, the definition of the derivative requires a linear structure (to define differences) and a topology (to define convergence) on both X and Y
- If $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, then the derivative is a linear transform from X to Y represented by the Jacobian matrix $f'(\underline{x}) \in \mathbb{R}^{m\times n}$
- Thus, we generally assume $f\colon X \to Y$ be a mapping from the Banach space $(X, \|\cdot\|_X)$ to the Banach space $(Y, \|\cdot\|_Y)$
- For directional derivatives, one only needs the linear structure on X

Directional Derivatives

Definition (Directional Derivative)

Let $f\colon X \to Y$ map vector space X to a Banach space $(Y, \|\cdot\|)$. Then, if it exists, the Gâteaux differential of f at $\underline{x}$ in direction $\underline{h}$ is given by

$$\delta f(\underline{x}; \underline{h}) \triangleq \lim_{t\to 0} \frac{f(\underline{x} + t\underline{h}) - f(\underline{x})}{t}.$$

Example

Consider $X = Y = \mathbb{R}^2$ and $f(\underline{x}) = (x_1 x_2, x_1 + x_2^2)$. For $\underline{x} = (1,1)$, $\underline{h} = (1,2)$:

$$\delta f(\underline{x}, \underline{h}) = \frac{d}{dt}\big((1+t)(1+2t), (1+t) + (1+2t)^2\big)\Big|_{t=0} = (3,5).$$

Lemma

Let $Y = \mathbb{R}$ and suppose that $\delta f(\underline{x}; \underline{h})$ exists and is negative for some f, $\underline{x}$, and $\underline{h}$. Then, there exists $t_0 > 0$ such that, for all $t \in (0, t_0)$, one has

$$f(\underline{x} + t\underline{h}) < f(\underline{x}).$$

Proof on whiteboard.

What is Meant by Differentiable?

Definition

Let $f: X \to Y$ be a mapping from a vector space X to a Banach space $(Y, \|\cdot\|)$. Then, f is Gâteaux differentiable at $\underline{x}$ if the Gâteaux differential $\delta f(\underline{x}; \underline{h})$ exists for all $\underline{h} \in X$ and is a continuous linear function of $\underline{h}$.

Definition (Differentiable)

Let $f: X \to Y$ be a mapping from a Banach space $(X, \|\cdot\|_X)$ to a Banach space $(Y, \|\cdot\|_Y)$. Then, f is Fréchet differentiable at $\underline{x}$ if there is a continuous linear transformation $T: X \to Y$ satisfying

$$\lim_{\underline{h} \to \underline{0}} \frac{\|f(\underline{x}+\underline{h}) - f(\underline{x}) - T(\underline{h})\|_Y}{\|\underline{h}\|_X} = 0,$$

where the limit is with respect to the implied Banach space mapping $X \to \mathbb{R}$. In this case, the Fréchet derivative $f'(\underline{x})$ equals T.

Gradient Descent

- Gradient descent adds the gradient $\nabla f(\underline{x})$ to an element of X
- But, for a Banach space, the derivative is a linear functional mapping X to $\mathbb{R}$!
- How can one add a linear mapping to X?
- In Hilbert space, the Riesz representation theorem states every linear functional is represented by the inner product with a fixed vector
- Thus, the gradient $\nabla f(\underline{x}) \in X$ is defined as the representative vector

Definition (Gradient Descent)

Let $f: X \to Y$ be a mapping from a Hilbert space X to the standard Banach space of real numbers. Starting from $\underline{x}_1 \in X$, gradient descent defines the sequence

$$\underline{x}_{n+1} = \underline{x}_n - \delta_n \nabla f(\underline{x}_n),$$

where δ_n is the step size and the gradient $\nabla f(\underline{x}) \in X$ satisfies

$$\langle \underline{h} | \nabla f(\underline{x}) \rangle = f'(\underline{x})(\underline{h}) \text{ for all } \underline{h} \in X.$$