Math Camp 2010

Lecture 4: Linear Algebra

Xiao Yu Wang

MIT

Aug 2010

Game Plan

- Vector Spaces
- Linear Transformations and Matrices
- Determinant
- Diagonalization
- Inner Product Spaces
- Definiteness

Definition (Real Vector Space)

A triple $(V, +, \cdot)$ consisting of a set V, addition $+$:

$$V \times V \rightarrow V$$

$$(x, y) \rightarrow x + y$$

and multiplication $\cdot$:

$$\mathbb{R} \times V \rightarrow V$$

$$(\lambda, x) \rightarrow \lambda \cdot x$$

is called a **Real Vector Space** if the following 8 conditions hold:

1. $(x+y)+z = x+(y+z)$ for all $x, y, z \in V$ (assoc of add)
2. $x+y = y+x$ for all $x, y \in V$ (commut of add)
3. There is an element $0 \in V$ (the zero vector) s.t. $x+0 = x \ \forall x \in V$
4. For each element $x \in V$, there is an element $y \in V$ such that $x+y=0$ (additive inverse)
5. $1 \cdot x = x$ for every $x \in V$ (identity)
6. $\lambda \cdot (\mu \cdot x) = (\lambda\mu) \cdot x$ for all $\lambda, \mu \in \mathbb{R}, x \in V$ (assoc mult)
7. $\lambda \cdot (x+y) = \lambda \cdot x + \lambda \cdot y$ for all $\lambda \in \mathbb{R}, x, y \in V$ (distr)
8. $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$ for all $\lambda, \mu \in \mathbb{R}, x \in V$ (distr)

Definition (Subspace)

Let V be a real vector space. A subset $U \subseteq V$ is called a **subspace** of V if U is nonempty and for any $x, y \in U$ and $\lambda \in \mathbb{R}$, $x + y \in U$ and $\lambda \cdot x \in U$

That is: A subspace is a subset that is closed under addition and multiplication.

The smallest subspace of any vector space is the **Null Space** $\{0\}$.

Any intersection of subspaces of a vector space V is a subspace of V.

Example

Finite Euclidean Space:

$\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R} = \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}, \forall i\}$

Note that by $x = (x_1, \ldots, x_n)$ we denote a vector in $\mathbb{R}^n$. One of the conventions is to call that vector a row vector and employ the notation x^T for a column vector.

Definition

Let $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ be two vectors. Define the sum by:

$$x + y = (x_1, ..., x_n) + (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n)$$

Definition

Let $x = (x_1, ..., x_n)$ be a vector and λ a real number. Define multiplication by:

$$\lambda \cdot x = \lambda(x_1, ..., x_n) = (\lambda x_1, ..., \lambda x_n)$$

One can verify that all 8 conditions from the definition of the vector space hold true for $\mathbb{R}^n$ so it is a vector space.

Example

A more complicated vector space is the set of all functions from a given set E to $\mathbb{R}$. E can be anything. Define $f+g$ such that $(f+g)(x) = f(x) + g(x)$ and λf such that $(\lambda f)(x) = \lambda f(x)$.
Subspaces are given by any property that is conserved by addition and multiplication; for example, the set of continuous functions from E into $\mathbb{R}$ denoted by $C(E)$, or the set of n-times continuously differentiable functions $C^n(E)$ are vector spaces.
Be careful and distinguish between $0 \in V$ and $0 \in \mathbb{R}$; we use the same notation for convenience.

Example

The set of all polynomials with coefficients from F is denoted by P(F).

Example

The set of all polynomials up to degree n with coefficients from F is denoted by $P_n(F)$.

Example

The set of all matrices of dimension $m \times n$ with coefficients from F is denoted by $M_{m \times n}(F)$.

Definition (Linear Combination)

Let V be a real vector space . A vector $v \in V$ is called a **linear combination** of vectors if there exist a finite number of vectors $v_1, \ldots, v_n \in V$ and scalars $a_1, \ldots, a_n \in \mathbb{R}$ such that $v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$. If the scalars are nonnegative and sum to 1 then we may also use the term convex combination.

Definition (Span)

Let S be a nonempty subset of a vector space V. The **span** of S, denoted span(S), is the set consisting of all linear combinations of the vectors in S. For convenience, we define span$(\emptyset) = \{0\}$.

Theorem

The span of any subset S of a vector space V is a subspace of V. Moreover, any subspace of V that contains S must also contain the span of S.

Definition (Generates (or spans))

A subset S of a vector space V **generates** (or **spans**) V if $\text{span}(S) = V$.

Example

The vectors $(1,1,0)$, $(1,0,1)$, and $(0,1,1)$ span $\mathbb{R}^3$ since an arbitrary vector $(a_1, a_2, a_3) \in \mathbb{R}^3$ is a linear combination of the three given vectors. In fact, the scalars r, s, t for which

$$r(1,1,0) + s(1,0,1) + t(0,1,1) = (a_1, a_2, a_3)$$

are $r = \frac{1}{2}(a_1 + a_2 - a_3), s = \frac{1}{2}(a_1 - a_2 + a_3), t = \frac{1}{2}(-a_1 + a_2 + a_3)$
Note that $(1,0,0)$, $(0,1,0)$, and $(0,0,1)$ also span $\mathbb{R}^3$.

Definition (Linearly Dependent)

A subset S of a vector space V is called **linearly dependent** if there exists a finite number of distinct vectors $u_1, u_2, \ldots, u_n$ in S and scalars $a_1, a_2, \ldots, a_n$, not all zero, such that

$$a_1 u_1 + a_2 u_2 + \cdots + a_n u_n = 0$$

In this case, we also say that the vectors of S are linearly dependent.

Example

Let $S = \{(1,3,-4,2),(2,2,-4,0),(1,-3,2,-4),(-1,0,1,0)\} \in \mathbb{R}^4$. S is linearly dependent if we can find scalars a_1, a_2, a_3, and a_4, not all zero, such that

$$a_1(1,3,-4,2) + a_2(2,2,-4,0) + a_3(1,-3,2,-4) + a_4(-1,0,1,0) = 0$$

One such solution is $(a_1, a_2, a_3, a_4) = (4,-3,2,0)$. So these vectors are linearly dependent (alt, S is linearly dependent).

Definition (Linearly Independent)

A subset S of a vector space that is not linearly dependent is called **linearly independent**. As before, we say the vectors of S are linearly independent.

Properties

1. The empty set is linearly independent
2. A set consisting of a single nonzero vector is linearly independent
3. A set is linearly independent iff the only representations of 0 as linear combinations of its vectors are trivial representations (i.e., coefficients all zero)

Definition (Basis)

A **basis** β for a vector space V is a linearly independent subset of V that generates V.

Examples

- $\varnothing$ is a basis for $\{0\}$
- The standard basis of $\mathbb{R}^n$ is $\{e_i : e_i = (0, ..., 1, ..., 0), i = 1, ..., n\}$
- The standard basis of $\mathrm{P}_n(\mathrm{F})$ is the set $\{1, x, x^2, ..., x^n\}$
- A basis of $\mathrm{P}(\mathrm{F})$ is the set $\{1, x, x^2, ...\}$

Theorem

Let V be a vector space and $\beta = \{u_1, ..., u_n\}$ be a subset of V. Then β is a basis for V iff each $v \in V$ can be **uniquely** *expressed as a linear combination of vectors of β.*

Proof

$\Rightarrow$ Let β be a basis for V. If $v \in V$, then $v \in$span(β). Thus v is a linear combination of the vectors of β. Suppose that

$$v = \sum_{i=1}^{n} a_i u_i = \sum_{i=1}^{n} b_i u_i$$

are two such representations of v. Subtracting these two equations yields

$$0 = \sum_{i=1}^{n} (a_i - b_i) u_i$$

Since β is linearly independent, $a_i - b_i = 0 \forall i$, so v is uniquely expressible as a linear combination of the vectors of β.

$\Leftarrow$ Exercise.$\blacksquare$

Theorem (Replacement Theorem)

Let V be a vector space that is generated by a set G ($\text{span}(G) = V$) containing exactly n vectors, and let L be a linearly independent subset of V containing exactly m vectors. Then $m \leq n$ and there exists a subset H of G containing exactly $n - m$ vectors such that $L \cup H$ generates V.

Corollary

Let V be a vector space having a finite basis. Then every basis for V contains the same number of vectors. (If you added a vector, it would break linear independence; if you took a vector away, it would break span.)

Definition (Dimension)

A vector space is called **finite-dimensional** if it has a basis consisting of a finite number of vectors. The unique number of vectors in each basis for V is called the **dimension** of V and is denoted by $\dim(V)$. A vector space that is not finite-dimensional is called **infinite-dimensional**.

Examples

- The vector space $\{0\}$ has dimension 0
- The vector space $\mathbb{R}^n$ has dimension n. (Recall the standard basis $(1, 0.., 0), (0, 1, 0, ..., 0), ...)$
- $C^0([0, 1])$ has infinite dimension. (Cts functions on $[0, 1]$.) Intuition: Weierstrass. (Allow polynomial to have infinite terms (countable).)

Corollary

Let V be a vector space with dimension n.
a) Any finite generating set for V contains at least n vectors, and a generating set for V that contains exactly n vectors is a basis for V.
b) Any linearly independent subset of V that contains exactly n vectors is a basis for V.
c) Every linearly independent subset of V can be extended to a basis for V

Proof of (c)

Let β be a basis for V. If L is a linearly independent subset of V containing m vectors, then the replacement theorem asserts that there is a subset H of β containing exactly $n - m$ vectors such that $L \cup H$ generates V. Now $L \cup H$ contains at most n vectors; therefore (a) implies that $L \cup H$ contains exactly n vectors and that $L \cup H$ is a basis for V.■

Example

Do the polynomials $x^3 - 2x^2 + 1, 4x^2 - x + 3$, and $3x - 2$ generate $P_3(\mathbb{R})$? (Hint: think of standard basis.)

Example

Is $\{(1, 4, -6), (1, 5, 8), (2, 1, 1), (0, 1, 0)\}$ a linearly independent subset of $\mathbb{R}^3$? (Hint: think of standard basis.)

- Linear Transformations
- Operations on Matrices
- Rank of a Matrix

Definition (Linear Transformation)

Let V and W be real vector spaces. We call a function $T : V \to W$ a **linear transformation** from V to W if, for all $x, y \in V$ and $c \in \mathbb{R}$, we have

a) $T(x + y) = T(x) + T(y)$ and

b) $T(cx) = cT(x)$

A Matrix

An $m \times n$ **matrix** A with entries from $\mathbb{R}$ is a rectangular array of the form

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

where each entry a_{ij} $(1 \leq i \leq m, 1 \leq j \leq n) \in \mathbb{R}$. We call the entries a_{ii} the **diagonal** entries of the matrix. An $m \times 1$ matrix is an element of $\mathbb{R}^m$ and is called a **column vector**. A $1 \times n$ matrix is an element of $\mathbb{R}^n$ and is called a **row vector**.

Matrix representation of a linear transformation

The matrix A is associated with the linear mapping from $x \in \mathbb{R}^n$ to $y \in \mathbb{R}^m$ that satisfies

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

We write $y = Ax$.

Example

Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation defined by $T(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2)$. Then the matrix associated with this mapping is

$$M_T = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}$$

since $M_T a = (a_1 + 3a_2, 0, 2a_1 - 4a_2)$.

The set of all $m \times n$ matrices with real entries is a vector space, which we denote by $\mathrm{M}_{m\times n}(\mathbb{R})$, with the following operations of **matrix addition** and **scalar multiplication:** For $A, B \in \mathrm{M}_{m\times n}(\mathbb{R})$ and $c \in \mathbb{R}$,

$$(A+B)_{ij} = A_{ij} + B_{ij}$$

$$(cA)_{ij} = cA_{ij}$$

for $1 \leq i \leq m, 1 \leq j \leq n$.

Properties of Matrices

- $A + B = B + A$
- $(A + B) + C = A + (B + C)$
- $(\lambda + \mu)A = \lambda A + \mu A$
- $\lambda(A + B) = \lambda A + \lambda B$
- $\lambda(\mu A) = (\lambda\mu)A$

The **null matrix** is the matrix that has all elements equal to 0 and is denoted by 0.

Gaussian Elimination

Definition ((Row) Echelon Form)

A matrix is in **(row) echelon form** if
a) All rows that contain only zeros are grouped at the bottom of the matrix
b) For each row that does not contain only zeros, the first nonzero entry of the row (the **pivot**) appears strictly to the right of the pivot of each row that appears above it.

Definition

A matrix is in reduced row echelon form if in addition, every leading coefficient is 1 and is the only nonzero entry in its column.

Example

This matrix is in reduced row echelon form:

$$\begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 2 \\ 0 & \mathbf{1} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 & -5 \\ 0 & 0 & 0 & \mathbf{1} & 3 \end{bmatrix}$$

Every matrix can be transformed to a matrix in row echelon form using (a finite number of) elementary row operations, which are defined as:

a) Add a multiple of one row to another

b) Interchange two rows

c) Multiply a row by a nonzero constant

This is Gaussian elimination.

The **matrix product** of A and B is defined when A is $m \times n$ and B is $n \times p$. The product is then defined by:

$$(AB)_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

and AB is an $m \times p$ matrix. This definition is motivated by the fact that if A and B represent linear mappings then AB represents the composition of the mappings $A \circ B$.

Example

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1*1+2*2+3*3=15 \\ 4*1+2*2+3*3=17 \end{bmatrix}$$

Properties

- (AB)C =A(BC) (assoc)
- A(B+C)=AB+AC (distr)
- (A+B)C=AC+BC (distr)

Properties

Note that if AB exists, BA need not be defined and even if it exists, it is not usually true that $AB = BA$.
However, commutativity holds for symmetric matrices.

Definition

(Symmetric matrix) An $n \times m$ matrix A is symmetric iff $a_{ij} = a_{ji}$ for $i \in \{1, ..., n\}$ and $j \in \{1, ..., m\}$.

Several other types of matrices are important:

Definition

(Diagonal matrix) A diagonal matrix is a square matrix in which all elements off the main diagonal are zero. That is, $d_{ij} = 0$ for all $i \neq j$, $i, j \in \{1, ..., n\}$.

For example: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ the **identity matrix** is diagonal.

Definition

(Triangular matrix) An $m \times n$ matrix is triangular if either all the elements above the main diagonal and the main diagonal are nonzero, and all other elements are zero, or all the elements below the main diagonal and the main diagonal are zero. The first matrix is called "upper triangular", and the second, "lower triangular".

For example: $\begin{bmatrix} 1 & 2 & 5 \\ 0 & 4 & 3 \\ 0 & 0 & 1 \end{bmatrix}$ is upper triangular.

Definition

(Permutation matrix) A permutation matrix is a square $n \times n$ binary matrix that has exactly one entry 1 in each row and each column and 0's elsewhere. Each such matrix represents a specific permutation of n elements and, when used to multiply another matrix, can produce that permutation in the rows or columns of the other matrix.

For ex: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}$

Definition

(Doubly stochastic matrix) A doubly stochastic matrix is a square matrix of nonnegative real numbers, each of whose rows and columns sums to 1.

The **transpose** of an $m \times n$ matrix A is denoted by A^T or A'. If A has elements a_{ij}, then A' has elements a_{ji}.
Note that if the matrix is symmetric, then it equals its transpose.
Properties

- $(A')' = A$
- $(A+B)' = A' + B'$
- $(AB)' = B'A'$

We call A **non-singular (invertible)** if there exists a matrix A^{-1} such that

$$AA^{-1} = A^{-1}A = I \text{ where } I = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

is the **identity matrix.**

Note that these are square matrices. Nonsquare matrices are not invertible, though they may have "left" or "right" inverses (not important).

Noninvertible square matrices are called **singular**.

Properties

- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A')^{-1} = (A^{-1})'$

The mapping that corresponds to the inverse of a matrix A corresponds to the inverse of the mapping associated with A.

Algorithm for Computing Matrix Inverse

- Augment matrix A with identity matrix
- Apply elementary row operations to this augmented matrix to transform A to I
- The matrix in the place of the original identity matrix is now A^{-1}

Example

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

Perform elementary operations: multiply the second row by -1 and add it to the first row.

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Thus

$$A^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Trick for inverting 2×2 matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

1. Calculate $(ad - bc) = D$
2. Then the inverse is: $\frac{1}{D}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Definition (Column/Row Rank)

The **column rank** of an $m \times n$ matrix A is the largest number of linearly independent column vectors A contains. The **row rank** is the largest number of linearly independent row vectors A contains.

It can be shown that the row rank and column rank are equal. The row and column rank of a matrix is often referred to simply as the **rank of the matrix**. We denote the rank of A by $rk(A)$.

Definition (Column Space)

The **column space** of A is the set of vectors generated by the columns of A, denoted by $CS(A) = \{y : y = Ax, x \in \mathbb{R}^n\}$. The dimension of $CS(A) = rk(A)$.

Definition (Null Space)

The **null space** of A, $N(A) = \{x \in \mathbb{R}^n : Ax = 0\}$.

Theorem

If A is an $m \times n$ matrix, then $rk(A) + dim N(A) = n$.

Theorem

Let A be an $n \times n$ matrix. Then the following are equivalent
a) A is nonsingular (invertible)
b) $N(A) = \{0\}$
c) $rk(A) = n$

Determining Rank and Bases for Row, Column, and Null Spaces

Let $A \in \mathrm{M}_{m\times n}(\mathbb{R})$ and U be a matrix in row echelon form obtained from A by row operations. Then

- $RS(A) = RS(U)$ and $N(A) = N(U)$
- The nonzero rows of U form a basis for $RS(U)$
- The columns of U that contain the pivots form a basis for the $CS(U)$
- Whenever certain columns of U form a basis for $CS(U)$, the corresponding column vectors of A form a basis for the $CS(A)$

Example

Find a basis for the row space, column space, and null space, and $rk(A)$, where

$$A = \begin{pmatrix} 2 & -1 & 3 \\ 4 & 2 & 1 \\ 2 & 3 & -2 \\ 8 & 0 & 7 \end{pmatrix}$$

Row echelon form gives

$$U = \begin{pmatrix} 2 & -1 & 3 \\ 0 & 4 & -5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

A basis for $RS(A)$ is $\{(2,-1,3),(0,4,-5)\}$.

A basis for the $CS(A)$ is $\left\{\begin{pmatrix}2\\4\\2\\8\end{pmatrix},\begin{pmatrix}-1\\2\\3\\0\end{pmatrix}\right\}$. And clearly $rk(A)=2$.

To find a basis for the $N(A) = N(U)$ we can examine the vectors, x, such that $Ux = 0$. This means that $x_2 = \frac{5}{4}s$ and $x_1 = \frac{1}{2}x_2 - \frac{3}{2}s$ where $x_3 = s$. Thus

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\frac{5}{4}s - \frac{3}{2}s \\ \frac{5}{4}s \\ s \end{pmatrix} = s\begin{pmatrix} -\frac{7}{8} \\ \frac{5}{4} \\ 1 \end{pmatrix}$$

So, $\left\{ \begin{pmatrix} -\frac{7}{8} \\ \frac{5}{4} \\ 1 \end{pmatrix} \right\}$ is a basis for $N(A)$.

Definition (Determinant)

Let $A \in \mathrm{M}_{n\times n}(\mathbb{R})$. If $n = 1$, so that $A = (A_{11})$, we define $\det(A) = A_{11}$. For $n \geq 2$, we define $\det(A)$ recursively as

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij})$$

for any $i \in \{1, ..., n\}$ where $\tilde{A}_{ij}$ denotes the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j. The scalar $\det(A)$ is called the **determinant** of A and is also denoted by $|A|$. The scalar

$$c_{ij} = (-1)^{i+j} \det(\tilde{A}_{ij})$$

is called the **cofactor** of the entry of A in row i, column j.

Example

Find the determinant of matrix A:

$$A = \begin{bmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & 6 \end{bmatrix}$$

Example

Then: $\det(A) = 1\det\begin{bmatrix} -5 & 2 \\ 4 & 6 \end{bmatrix} - 3\det\begin{bmatrix} -3 & 2 \\ -4 & 6 \end{bmatrix} - 3\det\begin{bmatrix} -3 & -5 \\ -4 & 4 \end{bmatrix} =$ $22 - 3(26) - 3(-32) = 40.$

Theorem

For any $A, B \in \mathrm{M}_{n\times n}(\mathbb{R})$, $det(AB) = det(A) \cdot det(B)$.

Corollary

A matrix $A \in \mathrm{M}_{n\times n}(\mathbb{R})$ *is invertible iff* $det(A) \neq 0$. *Furthermore, if A is invertible, then* $det(A^{-1}) = \frac{1}{det(A)}$.

Theorem

For any $A \in \mathrm{M}_{n\times n}(\mathbb{R})$, $det(A') = det(A)$.

Definition

(Kronecker product) If A is an $n \times m$ matrix and B is a $p \times q$ matrix, the Kronecker product of A and B is denoted by $A \otimes B$, where this equals:

$$\begin{bmatrix} a_{11}B & a_{1m}B \\ & \\ a_{n1}B & a_{nm}B \end{bmatrix}$$

In what follows, $A \in \mathrm{M}_{n\times n}(\mathbb{R})$.

Definition (Eigenvalue/Eigenvector)

If $Ax = \lambda x$ and $x \in \mathbb{R}^n \backslash \{0\}$, we say that λ is an **eigenvalue** of A and x is an **eigenvector** associated with λ.

Theorem

The eigenvalues are the roots of the characteristic polynomial of A, defined as $P_A(\lambda) = det(A - \lambda I)$. There are at most n eigenvalues.

Theorem

If $x_1, ..., x_k$ are eigenvectors corresponding to $\lambda_1, ..., \lambda_k$ distinct eigenvalues of A, then $\{x_1, ..., x_k\}$ is linearly independent.

Definition (Similar)

A is said to be **similar** to B if there exists an invertible matrix such that

$$A = PBP^{-1}$$

Definition (Diagonalizable)

A is **diagonalizable** if A is similar to a diagonal matrix, i.e. there exists an invertible matrix P and a diagonal matrix Λ such that $A = P\Lambda P^{-1}$

Theorem

A is diagonalizable iff A has n linearly independent eigenvectors.

Theorem

If A has n real, distinct eigenvalues, then A is diagonalizable.

Definition (Splits)

A polynomial $f(t)$ **splits over** $\mathbb{R}$ if there are scalars $c, a_1, ..., a_n$ (not necessarily distinct) in $\mathbb{R}$ such that

$$f(t) = c(t - a_1)(t - a_2)\cdots(t - a_n)$$

Theorem

The characteristic polynomial of any diagonalizable matrix splits.

Definition (Multiplicity)

Let λ be an eigenvalue of a matrix with characteristic polynomial $f(t)$. The **multiplicity** of λ is the largest possible integer k for which $(t-\lambda)^k$ is a factor of $f(t)$.

Theorem

A is diagonalizable iff
i) The characteristic polynomial of A splits
ii) For each eigenvalue λ of A, the multiplicity of λ equals $n - rk(A - \lambda I)$.

Example

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

The characteristic polynomial of A is $\det(A - \lambda I) = (1-\lambda)^2(2-\lambda)$, which splits and has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$ with multiplicities 2 and 1, respectively.

Example

Note that

$$3 - rk(A - \lambda_1 I) = 3 - rk \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} = 3 - 1 = 2$$

so, from the last theorem, A is diagonalizable. We can also find the basis composed of eigenvectors for $\mathbb{R}^3$. To do so, we find the null space of $(A - \lambda_i I)$ for $i = 1, 2$.

Example

For $\lambda_1 = 1$, the corresponding nullspace is

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \right\}$$

Example

This is the solution space for the system $x_2 + x_3 = 0$ and has a basis of

$$\gamma_1 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \right\}$$

Example

For $\lambda_2 = 2$, the corresponding nullspace is

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} -1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \right\}$$

Example

This is the solution space for the system $x_1 = x_3$ and has a basis of

$$\gamma_2 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Example

Let

$$\gamma = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Then γ is an ordered basis for $\mathbb{R}^3$ consisting of eigenvectors of A.

Example

One can verify that $A = P\Lambda P^{-1}$ where

$$P = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, P^{-1} = \begin{pmatrix} 1 & -1 & -1 \\ 0 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, and \Lambda = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Note that the columns of P are the eigenvectors corresponding to the eigenvalues, which are on the diagonal of Λ.

Trace

Definition (Trace)

The **trace** of A, denoted $tr(A)$, is the sum of the diagonal entries of A.

Properties:

1. $tr(A) + tr(B) = tr(A + B)$, for square matrices.
2. $tr(cA) = ctr(A)$, $c \in \mathbb{R}$, for square matrices.
3. If A is an $m \times n$ matrix and B is an $n \times m$ matrix, then $tr(AB) = tr(BA)$.

This yields the following useful theorem:

Theorem

Let A be an $n \times n$ matrix that is similar to an upper triangular matrix and has the distinct eigenvalues $\lambda_1, ..., \lambda_k$ with corresponding multiplicities $m_1, ..., m_k$. Then

a) $det(A) = (\lambda_1)^{m_1}(\lambda_2)^{m_2} \cdots (\lambda_k)^{m_k}$

b) $tr(A) = \sum_{i=1}^{k} m_i\lambda_i$

- Inner Products and Norms
- Orthonormal Basis and Projections
- Symmetric Matrices

Definition (Inner Product)

Let V be a real vector space. An **inner product** on V is a function that assigns, to every ordered pair of vectors $x, y \in V$, a real scalar, denoted $\langle x, y \rangle$, such that for all $x, y, z \in V$ and all real scalars c, the following hold

a) $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$
b) $\langle cx, y \rangle = c\langle x, y \rangle$
c) $\langle x, y \rangle = \langle y, x \rangle$
d) $\langle x, x \rangle > 0$ if $x \neq 0$

Example

When $V = \mathbb{R}^n$, the standard inner product is given by:

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x_i y_i$$

In terms of matrix notations: x and y are n-dimensional vectors, so they can be written as $n \times 1$ matrices and then the inner product is equivalent to the matrix multiplication:

$$\langle x, y \rangle = x' y$$

Example

Definition (Norm)

Let V be a vector space endowed with an inner product. For $x \in V$ we define the **norm** of x by:

$$\|x\| = \sqrt{\langle x, x \rangle}$$

Example

When $V = \mathbb{R}^n$, then $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$ is the Euclidean definition of length.

Let V be a vector space endowed with an inner product. Then for all $x, y \in V$ and $c \in \mathbb{R}$, the following are true:

- $\|cx\| = |c| \cdot \|x\|$
- $\|x\| = 0$ iff $x = 0$. In any case, $\|x\| \geq 0$.
- (Cauchy-Shwarz Inequality) $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$
- (Triangle Inequality) $\|x + y\| \leq \|x\| + \|y\|$

Definition (Orthogonal/Orthonormal)

Let V be a vector space endowed with an inner product. Vectors $x, y \in V$ are **orthogonal** if $\langle x, y \rangle = 0$. A susbset S of V is **orthogonal** if any two distinct vectors in S are orthogonal. A vector $x \in V$ is a **unit vector** if $\|x\| = 1$. Finally, a subset S of V is **orthonormal** if S is orthogonal and consists entirely of unit vectors.

We say that an $n \times n$ matrix A is orthonormal iff its column vectors form an orthonormal basis of $\mathbb{R}^n$. It should be clear that A is orthonormal iff $A'A = I$ (which implies that $A' = A^{-1}$ for orthonormal matrices; $A'A$ is a diagonal matrix for A orthogonal).
Example: standard basis of $\mathbb{R}^n$.

Theorem

Any subspace S of $\mathbb{R}^n$ has an orthonormal basis.

Definition (Orthogonal Complement)

Let S be a nonempty subset of an inner product space V. We define $S^\perp$ to be the set of all vectors in V that are orthogonal to every vector in S; that is, $S^\perp = \{x \in V : \langle x, y \rangle = 0 \text{ for all } y \in S\}$. The set $S^\perp$ is called the **orthogonal complement** of S.

It should be clear that $S^\perp$ is a subspace of V for any subset S of V.

Examples

- $\{0\}^\perp = V$ and $V^\perp = \{0\}$ for any inner product space V
- If $V = \mathbb{R}^3$ and $S = \{e_3\}$, then $S^\perp$ equals the $xy-$plane

Theorem

Let W be a finite-dimensional subspace of an inner product space V and let $y \in V$. Then there exists unique vectors $u \in W$ and $z \in W^{\perp}$ such that $y = u + z$. Furthermore, if $\{v_1, ..., v_k\}$ is an orthonormal basis for W, then

$$u = \sum_{i=1}^{k} \langle y, v_i \rangle v_i$$

Corollary

In the notation of the preceding theorem, the vector u is the unique vector in W that is "closest" to y; that is, for any $x \in W$, $\|y - x\| \geq \|y - u\|$, and this inequality is an equality iff $x = u$.

The vector u in the corollary is called the **orthogonal projection** of y on W.

Using all of this, it is straightforward to find the ONB of W, but we never do this in econ so I won't go through the "formula", but it's basically normalizing stuff.

Proof

We have $y = u + z$, where $z \in W^{\perp}$. Let $x \in W$. Then $u - x$ is orthogonal to z, so we have

$$\begin{aligned} \|y - x\|^2 &= \|u + z - x\|^2 = \|(u - x) + z\|^2 = \|u - x\|^2 + \|z\|^2 \\ &\geq \|z\|^2 = \|y - u\|^2 \end{aligned}$$

where the third equality can be shown to be true by applying the definition of a norm and using the linearity of the inner product. Now suppose that $\|y - x\| = \|y - u\|$. Then the inequality above becomes an equality, and therefore $\|u - x\| = 0$, and hence $u = x$. The proof of the converse is obvious.■

Application: Ordinary Least Squares Regression (OLS)

Consider a vector of explained data $y \in \mathbb{R}^n$ and vectors $x_1, ..., x_m \in \mathbb{R}^n$ of explanatory variables. Here, n is the number of observations; usually $n >> m$. In OLS we look for a linear combination of explanatory data that is closest to the explained data. It is given by a projection of y on the subspace of $\mathbb{R}^n$ spanned by the explanatory variables. The coefficients on the explanatory variables in the projected vector are called the regression parameters.

Diagonalization and Symmetric Matrices

Theorem

Let A be an $n \times n$ symmetric matrix. Then A is diagonalizable on an orthonormal basis of eigenvectors and all of the eigenvalues are real. Hence, A can be expressed as $A = Q\Lambda Q'$ where Q is an orthonormal matrix and Λ is a diagonal matrix.

Definition (Idempotent Matrix)

A $n \times n$ matrix A is said to be **idempotent** if $MM = M$.

Theorem

A symmetric idempotent matrix has eigenvalues equal to either 0 or 1.

Proof

Let λ be an eigenvalue of a matrix A. Then $Ax = \lambda x$. Therefore, $\lambda x = Ax = AAx = A\lambda x = \lambda Ax = \lambda^2 x$. Since x is nonzero and λ is real, it follows that λ can only be 0 or 1.■

Let A be an $n \times n$ matrix and $x \in \mathbb{R}^n$. Then

$$x'Ax = \sum_{i,j} a_{ij} x_i x_j$$

is a quadratic form. Note that $x'Ax$ is a scalar. Therefore it follows that $x'Ax = x'A'x$ such that we can write $x'Ax = \frac{1}{2}x'(A + A')x$ where $\frac{A+A'}{2}$ is symmetric. Thus, assume, without loss of generality, that A is symmetric. Then A is

- **positive (negative) definite** if $x'Ax > (<)0 \qquad \forall x \neq 0.$
- **positive (negative) semidefinite** if $x'Ax \geq (\leq)0 \qquad \forall x$

Theorem

For any matrix A, $A'A$ is positive semidefinite.

Proof

Let $y = Ax$. Then $x'A'Ax = y'y = \sum_{i=1}^{n} y_i^2 \geq 0.\blacksquare$

Theorem

A positive definite matrix is nonsingular.

Proof

Assume that A is positive definite and singular. Then there exists an $x \neq 0$ such that $Ax = 0$. But then $x'Ax = 0$ which contadicts that A is positive definite.$\blacksquare$

Theorem

If A is $m \times n$ matrix with $m \geq n$ and $rk(A) = n$, then $A'A$ is positive definite.

Proof

Let $y = Ax$ and observe that Ax is zero only if $x = 0$ because A is of full column rank n. Then for $x \neq 0$, $y \neq 0$, so $x'A'Ax = y'y > 0$.■

Theorem

Any symmetric matrix with strictly positive eigenvalues is positive definite. Any symmetric matrix with nonnegative eigenvalues is positive semidefinite.

In particular, any symmetric idempotent matrix is positive semidefinite.

A Determinantal Test of Definiteness

For any $T \times S$ matrix M, we denote by ${}_tM$ the $t \times S$ submatrix of M where only the first $t \leq T$ rows are retained. Similarly, we let M_s be the $T \times s$ submatrix of M where only the first $s \leq S$ columns are retained. Finally, we let ${}_tM_s$ be the $t \times s$ submatrix of M where only the first $t \leq T$ rows and $s \leq S$ columns are retained. Also, if M is an $N \times N$ matrix, then for any permutation π of the indices $\{1, ..., N\}$ we denote by M^{π} the matrix in which rows and columns are correspondingly permuted.

Theorem

Let M be an $N \times N$ symmetric matrix.
i) Then M is negative definite iff $(-1)^r |{}_rM_r| > 0$ for every $r = 1, ..., N$.
ii) Then M is negative semidefinite iff $(-1)^r |{}_rM_r^{\pi}| \geq 0$ for every $r = 1, ..., N$ and for every permutation π of the indices.

14.381 Statistical Method in Economics
Fall 2013