MATH 2121 LINEAR ALGEBRA
NOTES

Systems of linear equations

Definition We refer to $a_1x_1+a_2x_2+\cdots+a_nx_n=b$ as a **linear equation** in variables $x_1, x_2, \cdots x_n$.

A **system of linear equations** or **linear system** is a list of linear equations.

Definition A **solution** of a linear system in variables $x_1, x_2, \cdots, x_n$ is a list of n numbers with the property that if we set $x_1=s_1$, $x_2=s_2, \cdots x_n=s_n$ in our equations, we get all true statements.

A solution $(s_1, s_2, \cdots s_n)$ is **nonzero** if at least one number $s_i \neq 0$.

Two linear systems are **equivalent** if they have the same set of solutions.

Define **homogeneous system** as linear system in the form of $a_1x_1+a_2x_2+\cdots+a_nx_n=0$.

A linear system is **consistent** if it has at least one solution, and **inconsistent** if it has zero solutions.

Theorem A linear system has either 0, 1, or infinitely many solutions.

Matrices

A **matrix** is a rectangular array of numbers.

$$A=\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

Consider the linear system

$$\begin{cases} x_1-2x_2+x_3=0 \\ \quad 2x_2-8x_3=8 \\ 5x_1 \quad -5x_3=10 \end{cases}$$

Define the **coefficient matrix** to be $\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$

Define the **augmented matrix** to be $\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$.

Solving linear systems and row operations

Definition Peform the following **(elementary) row operations**

1. Replacement: $r_i+r_j\times c$
2. Scaling: $r_i\times c$
3. Interchange: $r_i \leftrightarrow r_j$

Two matrices are **row equivalent** if one can be transformed to the other by a sequence of row operations. Row operation is reversible.

Theorem If the augmented matrices of two linear systems are row equivalent, then the system are equivalent (have same solution)

Row reduction to echelon form (Gaussian elimination)

A row (column) is **nonzero** if not every entry in the row (column) is 0.

Definition A matrix is in **echelon form** if

1. If a row is nonzero, then every row above it is also nonzero
2. The leading entry in a nonzero row is to the right of the leading entry of any earlier row.
3. If a row is nonzero, then every entry below its leading entry in the same column is zero.

Definition A matrix is in **reduced echelon form** (RREF) if

1. The matrix is in echelon form.
2. Each nonzero row has leading entry 1.
3. The leading 1 in each nonzero row is the only nonzero number in its column.

Theorem Each matrix A is row equivalent to exactly one matrix RREF(A) in reduced echelon form.

Proposition In any echelon form E of a matrix A, the locations of the leading entries are the same.

A **pivot position** in a matrix A is the location containing a leading 1 in the RREF for A. A **pivot column** in a matrix A is a column containing a pivot position.

Solutions of linear systems

A variable x_i is **basic** if i is a pivot column, and **free** otherwise

- The system has **0 solutions** if the **last column is a pivot column**.
- The system has only **1 solution** if there are **no free variables** and the **last column is not a pivot.**
- Otherwise, the system has **infinitely many solutions.**

Vectors

Add two vectors of the same size:

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} u_1+v_1 \\ u_2+v_2 \\ \vdots \\ u_3+v_3 \end{bmatrix}$$

Note that $u+v=v+u$.

If v is a vector and $c\in\mathbb{R}$ is a **scalar**, define **scalar multiple**

$$cv = c\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}$$

The **zero vector** $\mathbb{R}^n$ is the vector $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. $0+v=v+0$.

Definition Suppose $v_1,\cdots,v_p\in\mathbb{R}^n$, $c_1,\cdots,c_p\in\mathbb{R}$

the vector $y=c_1v_1+c_2v_2+\cdots+c_pv_p$ is called a **linear combination** of $v_1,v_2,\cdots,v_p$. $c_1,c_2,\cdots c_p$ are called **coefficients.**

Definition The **span** of a list of vectors $v_1,v_2,\cdots,v_p\in\mathbb{R}^n$ is the set of all vectors $y\in\mathbb{R}^n$ that are linear combinations of $v_1,v_2,\cdots,v_p$

$\mathbb{R}\text{-span}\{v_1,v_2,\cdots,v_p\}$ or $\text{span}\{v_1,v_2,\cdots v_p\}$

Proposition If $v_1,v_2,\cdots v_p\in\mathbb{R}^n$, then **$y\in\mathbb{R}^n$ belongs to $\mathbb{R}\text{-span}\{v_1,v_2,\cdots v_p\}$** if and only if **$[v_1\ v_2\ \cdots v_p\ y]$** is the augmented matrix of a consistent linear system.

Multiplying matrices and vectors

Definition If A is a matrix with column $a_1,a_2,\cdots,a_m\in\mathbb{R}^m$ and $v\in\mathbb{R}^n$.

the **matrix-vector product** Av is given by

$$Av=[a_1\ a_2\ \cdots\ a_n]\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1a_1+v_2a_2+\cdots+v_na_n\in\mathbb{R}^m$$

A transforms vector in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$.

The transformation is linear:

$A(u+v)=Au+Av$, $A(cv)=c(Av)$.

Matrix equations

If A is an $m\times n$ matrix with columns $a_1,a_2,\cdots,a_n\in\mathbb{R}^m$ and

$$x=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},\ b=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

We call $Ax=b$ a **matrix equation**

Proposition $\underset{m\times 1}{Ax=b}$ has the same solution as both the vector equation $a_1x_1+a_2x_2+\cdots+a_nx_n=b$ and the linear system whose augmented matrix is $[a_1\ a_2\ \cdots\ a_n\ b]$.

Proposition The matrix equation **$Ax=b$ has a solution** if and only if **b is a linear combination of the columns of A**, that is $b\in\mathbb{R}\text{-span}\{a_1,a_2,\cdots a_n\}$.

Theorem $A: m\times n$. The following statements are equivalent

1. $\forall b\in\mathbb{R}^m$, $Ax=b$ has a solution
2. $\forall b\in\mathbb{R}^m$ is a linear combination of the columns of A.
3. The span of the columns of A is the set $\mathbb{R}^m$.
4. **A has a pivot position in every row.** $\rightarrow \begin{bmatrix} \ddots & \vdots \\ & b_n \end{bmatrix}$ if $b_n\neq 0$, then must ensure pivot postion in last row.
   $(\Rightarrow n\geq m)$

Linear independence

$v_1, v_2, \cdots, v_p \in \mathbb{R}^n$. Those vectors are linearly indepedent if the only solution to the vector function $v_1x_1 + v_2x_2 + \cdots + \cdots v_px_p = 0$ is given by $x_1 = x_2 = \cdots = x_p = 0$. Otherwise they're linearly dependent, if there're numbers $c_1, c_2, \cdots, c_p \in \mathbb{R}$, at least one of which is nonzero, such that $c_1v_1 + c_2v_2 + \cdots + c_pv_p = 0$.

Theorem The columns of A are linearly independent if and only if A has a pivot position in every column.

Theorem Suppose $v_1, v_2, \cdots v_p \in \mathbb{R}^n$. If $p > n$, then these vectors are linearly dependent.

If $y = c_1v_1 + c_2v_2 + \cdots + c_pv_p$ for $c_i \in \mathbb{R}$ is any linear combination of the vectors $v_1, \cdots v_p$, then $\mathbb{R}\text{-span}\{v_1, \cdots v_p\} = \mathbb{R}\text{-span}\{v_1, \cdots v_p, y\}$.

since $a_1v_1 + \cdots + a_pv_p + by = (a_1 + bc_1)v_1 + \cdots + (a_p + bc_p)v_p \in \mathbb{R}\text{-span}\{v_1, \cdots, v_p\}$.

Definition Consider the p sets given by

$$\{0\} \subseteq \mathbb{R}\text{-span}\{v_1\} \subseteq \mathbb{R}\text{-span}\{v_1, v_2\} \subseteq \cdots \subseteq \mathbb{R}\text{-span}\{v_1, \cdots, v_p\}.$$

The vectors $v_1, \cdots, v_p$ are linearly independent if these sets are all distinct

A linear dependence among $v_1, v_2, \cdots v_p \Leftrightarrow 0 = c_1v_1 + c_2v_2 + \cdots + c_pv_p$ for some $c_1, \cdots c_p \in \mathbb{R}$ that are not all zeros.

Proposition The vectors $v_1, v_2, \cdots v_p \in \mathbb{R}^n$ are linearly independent if and only if no linear dependence among it.

Vectors $v_1, v_2, \cdots, v_p \in \mathbb{R}$ are linearly dependent if and only if some vector v_i is a linear combination of other vectors.

How to determine if $v_1, v_2, \cdots v_p \in \mathbb{R}^n$ are linearly independent:

$$n \times p \text{ matrix } A = [v_1 \cdots v_p] \Rightarrow \text{RREF} \Rightarrow \begin{cases} \text{every column is a pivot} \Rightarrow \text{in-} \\ \text{Some columns not pivot} \Rightarrow \text{de-} \end{cases}$$

e.g. $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \\ 16 \end{bmatrix}$ are linearly dependent since

$$\begin{bmatrix} 1 & 2 & 5 \\ 0 & 3 & 9 \\ -1 & 5 & 16 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 5 \\ 0 & 1 & 3 \\ 0 & 1 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

A list of vectors in $\mathbb{R}^n$ is linearly dependent if it includes the zero vector. A single vector v is linearly independent if and only if $v \neq 0$.

Linear Transformations

$f: X \to Y$. X is called domain, Y is called codomain

The image of an input x in X under f is the output $f(x)$. The range of f is the set $\text{range}(f) = \{f(x) : x \in X\}$

Definition Let $f: \mathbb{R}^n \to \mathbb{R}^m$ be a function whose domain and codomain are sets of vectors. The function f is a linear transformation (linear function) if both the properties hold:

(1) $f(u+v) = f(u) + f(v)$ for all vectors, $u, v \in \mathbb{R}^n$

(2) $f(cv) = cf(v)$ for all vectors $v \in \mathbb{R}^n$ and scalars $c \in \mathbb{R}$.

e.g. $A: m \times n$. $T: \mathbb{R}^n \to \mathbb{R}^m$, $T(v) = Av$ for $v \in \mathbb{R}$ is a linear transformation

Proposition If $f: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation then

(3) $f(0) = 0$;

(4) $f(u-v) = f(u) - f(v)$ for $u, v \in \mathbb{R}^n$

(5) $f(au+bv) = af(u) + bf(v)$ for all $a, b \in \mathbb{R}$, $u, v \in \mathbb{R}^n$.

Theorem Suppose $T: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then there's a unique $m \times n$ matrix A such that $\underset{m\times 1}{T(v)} = \underset{m\times n\ n\times 1}{Av}$ for all $v \in \mathbb{R}^n$.

Define $e_1, \cdots, e_n \in \mathbb{R}^n$.

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \cdots e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

Define $\underset{m\times 1}{a_i} = \underset{n\times 1}{T(e_i)}$, $A = [a_1 \ a_2 \cdots a_n]$. $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{R}^n$.

$$T(w) = T(w_1e_1 + w_2e_2 + \cdots + w_ne_n) = \sum_{i=1}^{n} w_i \underset{m\times 1}{T(e_i)} = \sum_{i=1}^{n} w_i \underset{m\times 1}{a_i} = Aw$$

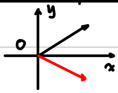
A is called the standard matrix of linear transformation T.

This could be used to evaluate A.

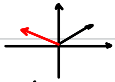
$$A = [T(e_1) \ T(e_2) \ \cdots \ T(e_n)].$$

Geometric interpretations of $\mathbb{R}^n \to \mathbb{R}^n$

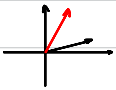
$\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}a\\-b\end{bmatrix}$



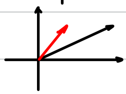
$\begin{bmatrix}-1&0\\0&1\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}-a\\b\end{bmatrix}$



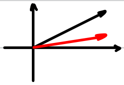
$\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}b\\a\end{bmatrix}$



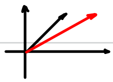
$\begin{bmatrix}k&0\\0&1\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}ka\\b\end{bmatrix}$ $(0<k<1)$



$\begin{bmatrix}1&0\\0&k\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}a\\kb\end{bmatrix}$



$\begin{bmatrix}1&k\\0&1\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}=\begin{bmatrix}a+bk\\b\end{bmatrix}$



One-to-one and onto functions

Definition A function $f: X \to Y$ is one-to-one (or injective) if $f(a)=f(b)$ implies $a=b$.

e.g. $T:\mathbb{R}^3 \to \mathbb{R}^2$, $T(v)=Av$. $A=\begin{bmatrix}1&2&5\\0&5&3\end{bmatrix}$.

A's columns are linearly dependent. Thus $\exists\, v \neq 0 \in \mathbb{R}^3$, s.t.

$[a_1\ a_2\ a_3]\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}=0 \Rightarrow T(v)=Av=0$. However, $T(0)=0$. Hence T is not one-to-one.

Theorem If $T:\mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then the following means the same thing:

(a) T is one-to-one;

(b) The only solution to $T(x)=0$ is $x=0 \in \mathbb{R}^n$.

(c) The columns of the standard matrix A of T are linearly independent. (A has pivot in every column).

Definition A function $f: X \to Y$ is onto (or surjective) if $\mathrm{range}(f)=\{f(x): x \in X\}=Y$, i.e. f is onto if its range is equal to its codomain. ($\exists\, y \in Y$, s.t. $\forall x \in X$, $f(x) \neq y$, then f is not onto).

Theorem Suppose $T:\mathbb{R}^n \to \mathbb{R}^m$ is the linear transformation with standard matrix A. The following statements are equivalent:

(a) T is onto

(b) Matrix equation $Ax=b$, has a solutions for $\forall b \in \mathbb{R}^m$. ($A$ has a pivot in every row).

(c) The span of columns of A is $\mathbb{R}^m$.

e.g. $T:\mathbb{R}^2 \to \mathbb{R}^3$, $T\left(\begin{bmatrix}v_1\\v_2\end{bmatrix}\right)=\begin{bmatrix}3v_1+v_2\\5v_1+7v_2\\v_1+3v_2\end{bmatrix}$.

$a_1=T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)=\begin{bmatrix}3\\5\\1\end{bmatrix}$, $a_2=T(e_2)=\begin{bmatrix}1\\7\\3\end{bmatrix}$

$A=\begin{bmatrix}3&1\\5&7\\1&3\end{bmatrix}\sim\begin{bmatrix}1&3\\0&1\\0&1\end{bmatrix}\sim\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}=\mathrm{RREF}(A)$.

Columns of A are linearly independent $\Rightarrow$ one-to-one

No pivot in every row $\Rightarrow$ not onto. $A: m\times n$

Corollary A linear transformation $T:\mathbb{R}^n \to \mathbb{R}^m$ is one-to-one only if $n \le m$, and onto only if $n \ge m$.

one-to-one $\Rightarrow n \le m$
onto $\Rightarrow n \ge m$

Operations on linear transformations

Sums and scalar multiples. Suppose $T:\mathbb{R}^n \to \mathbb{R}^m$, $U:\mathbb{R}^n \to \mathbb{R}^m$.

$T+U:\mathbb{R}^n \to \mathbb{R}^m$, $(T+U)(v)=T(v)+U(v)$, for $v \in \mathbb{R}^n$.

$c \in \mathbb{R}$, $cT:\mathbb{R}^n \to \mathbb{R}^m$, $(cT)(v)=cT(v)$, for $v \in \mathbb{R}^n$.

Both $T+U$ and cT are linear transformation.

Suppose T and U have standard matrices

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

Proposition The standard matrices of $T+U$ and cT are

$$A+B=\underbrace{\begin{bmatrix}a_{11}+b_{11}&a_{12}+b_{12}&\cdots&a_{1n}+b_{1n}\\a_{21}+b_{21}&a_{22}+b_{22}&\cdots&a_{2n}+b_{2n}\\\vdots&\vdots&&\vdots\\a_{m1}+b_{m1}&a_{m2}+b_{m2}&\cdots&a_{mn}+b_{mn}\end{bmatrix}}_{m\times n},\ cA=\underbrace{\begin{bmatrix}ca_{11}&ca_{12}&\cdots&ca_{1n}\\ca_{21}&ca_{22}&\cdots&ca_{2n}\\\vdots&\vdots&&\vdots\\ca_{m1}&ca_{m2}&\cdots&ca_{mn}\end{bmatrix}}_{k\times m}$$

Composition Suppose $U:\mathbb{R}^n \to \mathbb{R}^m$ and $T:\mathbb{R}^m \to \mathbb{R}^k$ are functions.

Assume the codomain of U is equal to the domain of T.

The composition $T\circ U$ is the function $\mathbb{R}^n \to \mathbb{R}^k$ given by

$$(T\circ U)(v)=T(U(v)).$$

If T and U are linear, then $T\circ U$ is linear.

$(T\circ U)(v)=Cv$. $T(v)=Av$. $U(w)=Bw$. Define the matrix product

$$AB=C.$$

Theorem Suppose B has columns $b_1, b_2, \cdots b_n \in \mathbb{R}^m$ so that $B=[b_1\ b_2 \cdots b_n]$.

Then $AB=[Ab_1\ Ab_2 \cdots Ab_n]$

An elementary matrix is a matrix E obtained by doing one row operation to I.
If B is obtained from A by doing one row operation then $B=EA$ for some E. Thus $RREF(A)=E_k\cdots E_1A$.

① cr_i $E=\begin{bmatrix}1&&&\\&1&&\\&&c&\\&&&\ddots\\&&&&1\end{bmatrix}$ ② $r_i \leftrightarrow r_j$ $E=\begin{bmatrix}0&1&&\\1&0&&\\&&\ddots&\\&&&1\end{bmatrix}$ ③ $r_i+r_j\times k$ $E=\begin{bmatrix}1&&&\\&1&a&\\&&1&\\&&&\ddots\\&&&&1\end{bmatrix}$

A position (i,j) in a matrix is diagonal if $i=j$.

Write $I_n=\begin{bmatrix}1&&\\&1&\\&&\ddots\\&&&1\end{bmatrix}$. I_n is the standard matrix of the identity map

$id_{\mathbb{R}^n}:\mathbb{R}^n\to\mathbb{R}^n$. $id_{\mathbb{R}^n}(v)=v$ for all $v\in\mathbb{R}^n$.

Proposition If A is an $m\times n$ matrix then $I_mA=AI_n=A$.

Proposition Assume A is $m\times n$, B is $n\times \ell$, C is $\ell\times k$.

Then $A(BC)=(AB)C$.

Note: 1. The product AB is defined only if the number of columns of A is the number of rows of B.

2. Even if AB, BA are both defined, $AB\neq BA$

3. $AB=AC \not\Rightarrow B=C$; $AC=BC \not\Rightarrow A=B$.

4. It can happen that $AB=0$, if $A\neq 0$, $B\neq 0$.

If A and B are both square matrix of the same size and $AB=BA$, then we say that A and B commute.

Matrix transpose

The transpose of an $m\times n$ matrix A is the $n\times m$ matrix A^T whose columns are the rows of A.

If $A=\begin{bmatrix}a&b&c\\d&e&f\end{bmatrix}$, $A^T=\begin{bmatrix}a&d\\b&e\\c&f\end{bmatrix}$.

Some basic properties:

1. $(A^T)^T=A$.

2. If $c\in\mathbb{R}$ and A, B have the same size, then $(A+B)^T=A^T+B^T$.

$(cA)^T=cA^T$.

3. If A is a $k\times m$ matrix and B is an $m\times n$ matrix, then $(AB)^T=B^TA^T$

Inverses

Definition The function f is invertible or bijective if f is both onto and one-to-one.

Proposition f is invertible if and only if for each b there's exacty one $a\in X$ with $f(a)=b$.

The identity function on a set X is the function $id_X: X\to X$ with $id_X(a)=a$ for all $a\in X$.

$id_{\mathbb{R}^n}:\mathbb{R}^n\to\mathbb{R}^n$'s standard function is the $n\times n$ identity matrix

$I_n=\begin{bmatrix}1&&\\&1&\\&&\ddots\\&&&1\end{bmatrix}$. $AI_n=A$, $I_nA=A$.

Proposition The function $f: X\to Y$ is invertible if and only if there's a function $f^{-1}: Y\to X$ such that $f\circ f^{-1}=id_Y$ and $f^{-1}\circ f=id_X$

Proposition If $T:\mathbb{R}^n\to\mathbb{R}^m$ is linear and invertible, then $n=m$ and T^{-1} is linear.

Definition Let A be an $n\times n$ matrix and define $T:\mathbb{R}^n\to\mathbb{R}^n$ by $T(x)=Ax$.

The matrix A is invertible is the function T is invertible. Its inverse is the unique matrix A^{-1} such that $T^{-1}(x)=A^{-1}x$.

Proposition Let A be an $n\times n$ matrix. The following mean the same thing:

(1) A is invertible

(2) There's an $n\times n$ matrix A^{-1} such that $AA^{-1}=A^{-1}A=I_n$.

(3) For each $b\in\mathbb{R}^n$, $Ax=b$ has a unique solution

(4) $RREF(A)=I_n$

A synonym for an invertible matrix is a non-singular matrix

Theorem Let $A=\begin{bmatrix}a&b\\c&d\end{bmatrix}$.

(1) A is invertible if $|ad-bc|\neq 0$.

(2) If $ad-bc\neq 0$, then $A^{-1}=\frac{1}{ad-bc}\begin{bmatrix}d&-b\\-c&a\end{bmatrix}$.

Theorem Let A and B be $n\times n$ matrices.

1. If A is invertible then $(A^{-1})^{-1}=A$.

2. If A and B are both intertible then AB is invertible with $(AB)^{-1}=B^{-1}A^{-1}$.

3. If A is invertible then A^T is invertible with $(A^T)^{-1}=(A^{-1})^T$

Process to compute A^{-1}: $RREF([A\ I_n])=[I_n\ A^{-1}]$.

Stronger characterization of invertible matrices

Theorem When A is an $n\times n$ matrix, the following are equivalent

(a) A is invertible (b) The columns of A are linearly independent (c) The span of the columns of A is $\mathbb{R}^n$.

Corollary A and B are both $n\times n$ matrices. If $AB=I_n$ then $BA=I_n$

Subspaces of $\mathbb{R}^n$

Let n be a positive integer. $0=\begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix}\in\mathbb{R}^n$.

Definition Let H be a subset of $\mathbb{R}^n$. The subset H is a **subspace** if

1. $0\in H$ 2. $\forall v,u\in H,\ u+v\in H$ 3. $\forall v\in H,\ c\in\mathbb{R},\ cv\in H$

Common examples:

1. $\mathbb{R}^n$ is a subspace of itself;
2. $\{0\}$ is a subspace of $\mathbb{R}^n$. ϕ is not a subspace of $\mathbb{R}^n$.
3. The span of a set of vectors in $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$. Conversely, any subspace of $\mathbb{R}^n$ is the span of a finite set of vectors.

Definition The **column space** of an $n\times n$ matrix A is the subspace

$$\mathrm{Col}\,A\subseteq\mathbb{R}^m.$$

given by the span of the columns of A.

Remark If $T:\mathbb{R}^n\to\mathbb{R}^m$, $T(x)=Ax$ then $\mathrm{Col}\,A=\mathrm{range}(T)$

A vector $b\in\mathbb{R}^m$ belongs to $\mathrm{Col}\,A$ iff $Ax=b$ has a solution.

$\Rightarrow \mathrm{Col}\,A=\mathbb{R}^m$ iff $Ax=b$ has a solution for each $b\in\mathbb{R}^m$.

Definition The **null space** of an $m\times n$ matrix A is the subspace

$$\mathrm{Nul}\ A\subseteq\mathbb{R}^n.$$

given by the set of vectors $v\in\mathbb{R}^n$ with $Av=0$.

Remark If $T:\mathbb{R}^n\to\mathbb{R}^m$, $T(x)=Ax$, the $\mathrm{Nul}\,A=\{x\in\mathbb{R}^n:\ T(x)=0\}$

Definition Let H be a subspace of $\mathbb{R}^n$. A **basis** for H is a set of vectors $\{v_1,v_2,\cdots,v_k\}\subseteq H$ that are linearly independent and have span equal to H.

ϕ is considered to be a basis for $\{0\}$.

If V,W are subspaces of $\mathbb{R}^n$,
then $V\cup W$ is not necessarily a subspace.
but $V\cap W$ is a subspace.

If A and B have the same number of rows,
then $\mathrm{Col}[A,B]=\mathrm{Col}\,A+\mathrm{Col}B$
If A and B have the same number of columns,
then $\mathrm{Nul}\begin{bmatrix}A\\B\end{bmatrix}=\mathrm{Nul}\,A\cap\mathrm{Nul}\,B$.

The set $\underbrace{\{\vec{e}_1,\vec{e}_2,\cdots\vec{e}_n\}}^{\subseteq\mathbb{R}^n}$ is a basis for $\mathbb{R}^n$, is called the **standard basis** of $\mathbb{R}^n$

Theorem Every subspace H of $\mathbb{R}^n$ has a basis of size at most n.

e.g. $A=\begin{bmatrix}-3&6&-1&1&-7\\1&-2&2&3&-1\\2&4&5&8&-4\end{bmatrix}$. Find a basis for $\mathrm{Nul}\,A$.

$$[A,0]\sim\begin{bmatrix}1&-2&0&-1&3&0\\0&0&1&2&-2&0\\0&0&0&0&0&0\end{bmatrix},\quad\begin{cases}x_1=2x_2+x_4-3x_5\\x_3=-2x_4+2x_5.\end{cases}$$

$$x=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix}=\begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix}x_2+\begin{bmatrix}1\\0\\-2\\1\\0\end{bmatrix}x_4+\begin{bmatrix}-3\\0\\2\\0\\1\end{bmatrix}x_5$$

the basis

The size of $\mathrm{Nul}\,A$ is equal to the number of free variables in the linear system $Ax=0$.

Proposition Let A be any matrix. The pivot columns form a basis for $\mathrm{Col}\,A$.

Coordinate System

Suppose H is a subspace of $\mathbb{R}^n$. Let $b_1,b_2,\cdots,b_k$ be a basis of H.

Theorem Let $v\in H$. There're unique coefficient $C_1,C_2,\cdots C_k\in\mathbb{R}$ s.t.

$$C_1b_1+C_2b_2+\cdots+C_kb_k=v$$

$B=(b_1,b_2,\cdots b_k)$ consist of the basis vectors in some fixed order.

Given $v\in H$, $[v]_B=\begin{bmatrix}C_1\\C_2\\\vdots\\C_k\end{bmatrix}\in\mathbb{R}^k$ as the unique vector with $C_1b_1+\cdots C_kb_k=v$

Equivalently, $[v]_B$ is the unique solution to $[b_1\ b_2\cdots b_k]x=v$.

$[v]_B$ is called the **coordinate vector of v in the basis B**, or just **v in the basis B.**

e.g. Let $b_1=\begin{bmatrix}3\\6\\2\end{bmatrix}$, $b_2=\begin{bmatrix}-1\\0\\1\end{bmatrix}$, $v=\begin{bmatrix}3\\12\\7\end{bmatrix}$. $B=(b_1,b_2)$.

$$x=\begin{bmatrix}x_1\\x_2\end{bmatrix}\in\mathbb{R}^2\ \text{s.t.}\ \begin{bmatrix}3&-1\\6&0\\2&1\end{bmatrix}x=\begin{bmatrix}3\\12\\7\end{bmatrix}$$

$$\begin{bmatrix}3&-1&3\\6&0&12\\2&1&7\end{bmatrix}\sim\begin{bmatrix}1&0&2\\0&1&3\\0&0&0\end{bmatrix},\ [v]_B=\begin{bmatrix}2\\3\end{bmatrix}$$

Theorem H is a subspace of $\mathbb{R}^n$. Then all bases of H have the same number of elements.

Dimension

Definition The **dimension** of a subspace H is the number of vectors in any basis of H, denoted by $\dim H$.

If $H=\{0\}$, then $\dim H=0$.

Corollary The dimension of $\mathrm{Nul}\,A$ is the number of free variables in the linear system $Ax=0$.

Corollary The dimension of $\mathrm{Col}\,A$ is the number of pivot columns in A.

Definition The **rank** of a matrix A is $\mathrm{rank}\,A=\dim \mathrm{Col}\,A$

Theorem (Rank-nullity theorem) If A is a matrix with n columns, then $\mathrm{rank}\,A+\dim \mathrm{Nul}\,A=n$.

Theorem (Basis theorem) If H is a subspace of $\mathbb{R}^n$ with $\dim H=p$ then

1. Any set of p linearly independent vectors in H is a basis for H
2. Any set of p vectors in H whose span is H is the basis for H.

Corollary If H is an n-dimensional subspace of $\mathbb{R}^n$, then $H=\mathbb{R}^n$.

Corollary If $U, V\subseteq\mathbb{R}^n$ are subspaces with $U\subseteq V$ but $U\neq V$, then $\dim U<\dim V\leq n$.

Corollary Let A be an $n\times n$ matrix. The following are equivalent:

(a) A is invertible;

(b) The columns of A form a basis for $\mathbb{R}^n$.

(c) $\mathrm{rank}\,A=\dim\mathrm{Col}\,A=n$.

(d) $\dim\mathrm{Nul}\,A=0$.

Determinants

Theorem Let n be any positive integer. There exists a unique function $\det:\{n\times n \text{ matrices}\}\to\mathbb{R}$, called **determinants**, with the following properties.

(1) $\det I_n=1$.

(2) $a_1, a_2, \cdots, a_n\in\mathbb{R}^n$, $1\leq i\leq j\leq n$, then

$$\det[a_1\cdots a_i\cdots a_j\cdots a_n]=-\det[a_1\cdots a_j\cdots a_i\cdots a_n]$$

(3) $a_2, a_3, \cdots a_n\in\mathbb{R}^n$. If $u, v\in\mathbb{R}^n$, then

$$\det[u+v\ a_2\ a_3\cdots a_n]=\det[u\ a_2\ a_3\cdots a_n]+\det[v\ a_2\cdots a_n]$$

and if $c\in\mathbb{R}$, then

$$\det[cv\ a_2\ a_3\cdots a_n]=c\cdot\det[v\ a_2\ a_3\cdots a_n]$$

Lemma If A has two equal columns then $\det A=0$.

Corollary If the columns of A are not linearly independent, then $\det A=0$.

Corollary If A is not invertible then $\det A=0$.

Lemma Suppose $1\leq i\leq n$ and $a_i=pu+qv$ for some $p, q\in\mathbb{R}$ and $u, v\in\mathbb{R}^n$. Define $B=[a_1\cdots a_{i-1}\ u\ a_{i+1}\cdots a_n]$, $C=[a_1\cdots a_{i-1}\ v\ a_{i+1}\cdots a_n]$. Then $\det A=p\det B+q\det C$.

$\det\begin{bmatrix} a & b \\ c & d\end{bmatrix}=ad-bc$.

Permutation matrices

Rearranging the columns of I.

A **permutation matrix** is an $n\times n$ matrix whose entries are all 0 or 1, and which has exactly one non-zero entry in each row and in each column, denoted by S_n.

Let R_n be the set of $n\times n$ matrices whose entries are all 0 or 1, and which have exactly one nonzero entry in each column

(Size of S_n) $=A_n^n=n!$, (size of R_n) $=n^n$.

Lemma If $X\in R_n$ but $X\notin S_n$ then $\det X=0$

Given $X\in S_n$, define $\mathrm{inv}(X)$ to be the number of 2×2 submatrices of X equal to $\begin{bmatrix} 0 & 1 \\ 1 & 0\end{bmatrix}$

Lemma If $X\in S_n$, then $\det X=(-1)^{\mathrm{inv}(X)}$.

A formular for det A

Given a matrix $X \in R_n$ and an arbitrary $n\times n$ matrix A, define

prod(X,A) = the product of the entries of A in the non-zero positions of X.

e.g. $$\mathrm{prod}\left(\begin{bmatrix} 0&0&1\\ 1&0&0\\ 0&1&0\end{bmatrix}, \begin{bmatrix} a&b&c\\ d&e&f\\ g&h&i\end{bmatrix}\right) = cdh.$$

Theorem Suppose A is an $n\times n$ matrix. Then

$$\det A = \sum_{X\in S_n} \mathrm{prod}(X,A)(-1)^{\mathrm{inv}(X)}.$$

$$\det\begin{bmatrix} a&b&c\\ d&e&f\\ g&h&i\end{bmatrix} = a(ei-hf)-b(di-gf)+c(dh-eg).$$

More properties of the determinant.

Lemma If $X\in S_n$, then $X^T\in S_n$ and $\mathrm{inv}(X)=\mathrm{inv}(X^T)$

Corollary If A is any square matrix then **$\det(A)=\det(A^T)$**

Corollary If A is a square matrix with **two equal rows** then $\det A=0$.

Lemma Let A and B be $n\times n$ matrices with $\det A\neq 0$. Then

$$\det(AB)=(\det A)(\det B)=\det(BA)$$

Determinants of triangular and invertible matrices

An $n\times n$ matrix A is **upper-triangular** if it looks like $\begin{bmatrix} *&*&*&*\\ 0&*&*&*\\ 0&0&*&*\\ 0&0&0&*\end{bmatrix}$

An $n\times n$ matrix A is **lower-triangular** if it looks like $\begin{bmatrix} *&0&0&0\\ *&*&0&0\\ *&*&*&0\\ *&*&*&*\end{bmatrix}$.

The transpose of an upper-triangular matrix is lower triangular and vice versa. $\begin{bmatrix} *&0&0&0\\ 0&*&0&0\\ 0&0&*&0\\ 0&0&0&*\end{bmatrix}$

A matrix is **diagonal** if it's both upper- and lower-triangular.

Proposition If A is a triangular matrix then $\det A$ is the **product of the diagonal entries** of A.

Lemma If A is an $n\times n$ matrix then $\det A$ is a **nonzero multiple of $\det(\mathrm{RREF}(A))$**.

Theorem An $n\times n$ matrix A is an **invertible** if and only if **$\det A\neq 0$**.

Lemma A and B are $n\times n$ matrices. If A or B is not invertible then AB is not invertible.

If X,Y are square matrices then $XY=I \Rightarrow X^{-1}=Y$ and $Y^{-1}=X$ and $YX=I$.

Computing determinants

Start by denom $=1$. Row reduce A to an echelon form E.

(a) Switch two rows, **multiply denom by -1**.

(b) Multiply $\lambda\neq 0$ to a row, **multiply denom by λ**.

(c) Add multiple of one row to anoth row, **do nothing to λ**.

Then **$\det A = \dfrac{\det E}{\mathrm{denom}}$**.

Theorem If A is the $n\times n$ matrix with entry a_{ij} row i and j, then

(1) **$\det A = a_{11}\det A^{(1,1)} - a_{12}\det A^{(1,2)} - (-1)^n a_{1n}\det A^{(1,n)}$**

(2) $\det A = a_{11}\det A^{(1,1)} - a_{21}\det A^{(2,1)} - (-1)^n a_{n1}\det A^{(n,1)}$.

Interpreting the determinant geometrically

Proposition If A is an $n\times n$ matrix then $|\det A|$ is the **volume** of the n-dimensional parallelogram

$$P(A)=\{Av : v\in\mathbb{R}^n \text{ with } 0\le v_i\le 1 \text{ for all } i=1,2,\cdots n\}.$$

Corollary Suppose $T:\mathbb{R}^n\to\mathbb{R}^n$ is a linear transformation with standard matrix A. If S is any region in $\mathbb{R}^n$ with finite volume then **the volume of $T(S)$ is the volumn of S times $|\det A|$**.

Vector spaces

Definition A **vector space** is a nonempty set V with two operations called **vector addition** and **scalar multiplication** satisfying several conditions. We refer to the elements of V as **vectors**.

The vector addition operation for V: $u,v\in V \Rightarrow u+v\in V$.

(a) **$u+v=v+u$**; (b) **$(u+v)+w=u+(v+w)$**

(c) There exists a unique **zero vector** $0\in V$ with the property that **$0+v=v$** for all $v\in V$.

The scalar multiplication operation for V: $c\in\mathbb{R}, v\in V \Rightarrow cv\in V$.

(a) If $c=-1$, then **$v+(-1)v=0$**. (b) **$c(u+v)=cu+cv$**.

e.g. $\mathbb{R}$ is a vector space

$\mathbb{R}^n$ and its subspace are vector spaces

$\mathbb{R}^{m\times n}$ = (set of $m\times n$ matrices) is a vector space e.g. $\mathrm{Fun}(\{1,2,3,\dots 19\}, \mathbb{R}) \cong \mathbb{R}^{19}$

19 rows

(c) $(c+d)v = cv + dv$, for $c, d \in \mathbb{R}$; (d) $c(dv) = (cd)v$ for $c, d \in \mathbb{R}$

(e) If $c=1$ then $1v = v$.

Proposition Let X be a set and let V be a vector space. Then the set $\mathrm{Fun}(X, V)$ of all functions $f: X \to V$ is a vector space once we define

$f + g =$ (the function that maps $x \mapsto f(x) + g(x)$ for $x \in X$)

$cf =$ (the function that maps $x \mapsto cf(x)$ for $x \in X$)

$0 =$ (the function that maps $x \mapsto 0$ for $x \in X$)

for $f, g \in \mathrm{Fun}(X, V)$ and $c \in \mathbb{R}$.

Definition A subset $H \subseteq V$ is a **subspace** if $0 \in H$ and if $u+v \in H$ and $cv \in H$ for all $u, v \in H$ and $c \in \mathbb{R}$.

A function $f: U \to V$ is **linear** if $f(u+v) = f(u) + f(v)$ and $f(cv) = cf(v)$ for all $u, v \in U$ and $c \in \mathbb{R}$.

Proposition If U, V, W are vector spaces and $f: V \to W$ and $g: U \to V$ are linear functions then $f \circ g: U \to W$ is also linear, if we define $f \circ g(x) = f(g(x))$ for $x \in U$.

Definition A **linear combination** of a finite list of vectors $v_1, \dots, v_k \in V$ is a vector of the form $c_1 v_1 + c_2 v_2 + \dots + c_k v_k$ for some $c_1, c_2, \dots, c_k \in \mathbb{R}$.

A linear combination of an infinite set of vectors is a linear combination of some finite subset of vectors.

Definition The **span** of a set of vectors is the set of all linear combinations that can be formed from the vectors.

The span of a set of vectors in V is a subspace of V.

Example The subspace of polynomials P in $\mathrm{Fun}(\mathbb{R}, \mathbb{R})$ is the span of the set of functions $1, x, x^2, x^3 \dots$ The infinite sum $e^x = 1 + x + \frac{1}{2}x + \cdots$ does not belong to the subspace.

not a linear combination $\quad \dim(P) = \infty$

Definition A finite list of vectors $v_1, v_2, \dots, v_k \in V$ is **linearly independent** if it is impossible to express $0 = c_1 v_1 + c_2 v_2 + \dots + c_k v_k$ except when $c_1 = c_2 = \dots = c_k = 0$. An infinite list of vectors is defined to be **linearly independent** if every finite subset of the vectors is linearly independent.

Definition A **basis** of a vector space V is a subset of linearly independent vectors whose span is V. $b_1, b_2, b_3, \dots$ is a basis for V is the same as saying that for each $v \in V$, there are unique coefficients $x_1, x_2, x_3 \dots \in \mathbb{R}$, all but finitely many of which are 0, such that $v = x_1 b_1 + x_2 b_2 + \cdots$

Theorem Let V be a vector space

1. V has at least one basis
2. Every basis of V has the same number of elements.
3. If A is a subset of linearly independent vectors in V then V has a basis B with $A \subseteq B$.
4. If C is a subset of vectors in V whose span is V then V has a basis B with $B \subseteq C$.

Definition The **dimension** of a vector space V is the number $\dim V$ of elements in any of its bases.

e.g. If X is a finite set then $\dim \mathrm{Fun}(X, \mathbb{R}) = |X|$, where $|X|$ is the size of X.

Definition Suppose U and V are vector spaces and $f: U \to V$ is a linear function. Define $\mathrm{range}(f) = \{f(x) : x \in U\} \subseteq V$ and $\mathrm{kernel}(f) = \{x \in U : f(x) = 0\} \subseteq U$.

Theorem (Rank-Nullity Theorem) If $\dim U < \infty$ then $\dim \mathrm{range}(f) + \dim \mathrm{kernel}(f) = \dim U$.

Eigenvectors and eigenvalues

Suppose A is a square $n\times n$ matrix.

Definition An **eigenvector** of A is a nonzero vector $v \in \mathbb{R}^n$ such that $Av=\lambda v$ for a number $\lambda \in \mathbb{R}$. λ is called the **eigenvalue** of A for the eigenvector v. (λ could be 0)

Proposition A number $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $A-\lambda I$ is not invertible.

Proof. $Ax=\lambda x \Rightarrow (A-\lambda I)x=0$ has a nonzero solution $\Rightarrow \mathrm{Nul}(A-\lambda I) \neq \{0\}$. $\Rightarrow A-\lambda I$ is not invertible.

Corollary A number $\lambda \in \mathbb{R}$ is an eigenvalue of A if and only if $\det(A-\lambda I)=0$.

We call the set of all $v \in \mathbb{R}^n$ with $Av=\lambda v$ the **eigenspace** of A for λ. We also call this the **λ-eigenspace** of A. This is just the null space of $A-\lambda I$. A number is an eigenvalue of A if and only if the corresponding eigenspace is nonzero.

Proposition The following statements are equivalent:

1. There exists an eigenvector $v \in \mathbb{R}^n$ for A with eigenvalue λ.
2. The matrix $A-\lambda I$ is not invertible.
3. $\det(A-\lambda I)=0$
4. The λ-eigenspace for A contains a non-zero vector.

The **characteristic polynomial** of a square matrix A is $\det(A-xI)$.

Theorem The eigenvalues of a triangular square matrix A are its diagonal entries. If these numbers are $d_1, d_2, \cdots, d_n$ then the characteristic polynomial of A is $(d_1-x)(d_2-x)\cdots(d_n-x)$.

Theorem Suppose $\lambda_1, \lambda_2, \cdots, \lambda_r$ are **distinct** eigenvalues for A, meaning $\lambda_i \neq \lambda_j$ for $i\neq j$. Let $v_1, v_2, \cdots, v_r \in \mathbb{R}^n$ be the corresponding eigenvectors, so that $Av_i=\lambda v_i$ for $i=1,2,\cdots,r$.

The the vectors are linearly independent.

The eigenvalues of A are precisely the solutions to the equation $\det(A-xI)=0$, which is the **characteristic equation** for A.

e.g. The matrix

$$A=\begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

has characteristic polynomial of $\det(A-xI)=(5-x)^2(3-x)(1-x)$.

Say 5 is an eigenvalue of A with **algebraic multiplicity** 2. 1 and 3 have **algebraic multiplicity** 1.

e.g. Consider the matrix $A=\begin{bmatrix} 1 & 5 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

The characteristic polynomial is $(1-x)(2-x)(3-x)$. Its eigenvalues are 1, 2, 3, each has algebraic multiplicity 1.

1-eigenspace. $A-I=\begin{bmatrix} 0 & 5 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \sim \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \mathrm{RREF}(A-I)$.

$$\Rightarrow x=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}x_1.$$

2-eigenspace $A-2I=\begin{bmatrix} -1 & 5 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -5 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \mathrm{RREF}(A-2I)$.

$$x=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} 5x_2 \\ x_2 \\ 0 \end{bmatrix}=\begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}x_2.$$

3-eigenspace $A-3I=\begin{bmatrix} -2 & 5 & 4 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \mathrm{RREF}(A-3I)$.

$$x=\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} 2x_3 \\ 0 \\ x_3 \end{bmatrix}=\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}x_3.$$

$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$ are linearly independent.

Lemma For any integer $n\geq 0$, we have $A^n=(PDP^{-1})^n=PD^nP^{-1}$.

Lemma For any integer $n\geq 0$, we have

$$D^n=\begin{bmatrix} 1^n & & \\ & 2^n & \\ & & 3^n \end{bmatrix}=\begin{bmatrix} 1 & & \\ & 2^n & \\ & & 3^n \end{bmatrix}$$

Similar matrices

Definition Two $n\times n$ matrices X and Y are **similar** if there exists an invertible $n\times n$ matrix P with $X = PYP^{-1}$.

In this case it also holds that $Y = P^{-1}PYP^{-1}P = P^{-1}XP$. We say X is similar to Y and Y is similar to X.

Definition A squre matrix X is **diagonalizable** if X is similar to a diagonal matrix.

Proposition An $n\times n$ matrix is always similar to itself.

Proposition Suppose A, B, C are $n\times n$ matrices. Assume A and B are similar and B and C are also similar. Then A and C are similar.

Theorem If A and B are similar $n\times n$ matrices the A and B have the same characteristic polynomial and so have the same eigenvalues.

Theorem An $n\times n$ matrix A is diagonalizable if and only if $\mathbb{R}^n$ has a basis $v_1, v_2, \cdots, v_n$ whose elements are all eigenvectors of A.

If λ_i is the eigenvalue such that $Av_i = \lambda_i v_i$, then $A = PDP^{-1}$ for $P = [v_1\ v_2 \cdots v_n]$ and $D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$.

Theorem If A is an $n\times n$ matrix with n distinct eigenvalues then A is diagonalizable.

A is diagonalizable if $A = PDP^{-1}$ for some $D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$. In this case:

- $\lambda_1, \lambda_2 \cdots \lambda_n$ are eigenvalues of A since A and D are similar.
- If $P = [v_1\ v_2 \cdots v_n]$ then $Av_i = \lambda_i v_i$ for each $i = 1, 2, \cdots, n$.

  $Pe_i = v_i$, $P^{-1}v_i = e_i$, $De_i = \lambda_i e_i \Rightarrow Av_i = PDP^{-1}v_i = PDe_i = P\lambda_i e_i = \lambda_i v_i$.
- The columns of P are a basis for $\mathbb{R}^n$ of eigenvectors of A.

  These columns are eigenvectors and they are linearly independent as P is invertible.

Note Not all diagonalizable $n\times n$ matrices have n distinct eigenvalues, however. Not all $n\times n$ matrices are diagonalizable. Not all $n\times n$ matrices have n distinct eigenvalues.

Diagonalization and Fibonacci numbers

The sequence f_n of Fibonacci numbers

$$f_0 = 0,\ f_1 = 1,\ f_2 = 1,\ f_3 = 2,\ f_4 = 3,\ f_5 = 5 \cdots$$

For $n \geq 2$, $f_n = f_{n-2} + f_{n-1}$.

Define $a_n = f_{2n} = f_{2n-2} + f_{2n-1} = a_{n-1} + b_{n-1}$.

$b_n = f_{2n+1} = f_{2n-1} + f_{2n} = b_{n-1} + a_n = a_{n-1} + 2b_{n-1}$.

$$\Rightarrow \begin{bmatrix} a_n \\ b_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} a_{n-1} \\ b_{n-1} \end{bmatrix} \ (n>0)$$

$$\begin{bmatrix} a_n \\ b_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} a_{n-1} \\ b_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^2\begin{bmatrix} a_{n-2} \\ b_{n-2} \end{bmatrix} = \cdots = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^n\begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}^n\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix},\ \det(A - xI) = \begin{vmatrix} 1-x & 1 \\ 1 & 2-x \end{vmatrix} = (x-2)(x-1) - 1 = x^2 - 3x + 1 = 0.$$

$\Rightarrow \alpha = \frac{3+\sqrt{5}}{2}$, $\beta = \frac{3-\sqrt{5}}{2}$. These are the eigenvalues of A.

$$A - \alpha I = \begin{bmatrix} 1-\alpha & 1 \\ 1 & 2-\alpha \end{bmatrix} \sim \begin{bmatrix} 1 & 2-\alpha \\ 1-\alpha & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 2-\alpha \\ 0 & 1-(2-\alpha)(1-\alpha) \end{bmatrix} = \begin{bmatrix} 1 & 2-\alpha \\ 0 & 0 \end{bmatrix}$$

$\Rightarrow x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \alpha-2 \\ 1 \end{bmatrix} x_2$, $x_2 \in \mathbb{R}$. $\Rightarrow v = \begin{bmatrix} \alpha-2 \\ 1 \end{bmatrix}$ is an eigenvector.

Similarly we have $w = \begin{bmatrix} \beta-2 \\ 1 \end{bmatrix}$ is also an eigenvector.

Thus $P = [v\ w] = \begin{bmatrix} \alpha-2 & \beta-2 \\ 1 & 1 \end{bmatrix}$, $P^{-1} = \frac{1}{\alpha-\beta}\begin{bmatrix} 1 & 2-\beta \\ -1 & \alpha-2 \end{bmatrix}$, $D = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}$

$$\Rightarrow \begin{bmatrix} a_n \\ b_n \end{bmatrix} = A^n\begin{bmatrix} 0 \\ 1 \end{bmatrix} = PD^nP^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{\alpha-\beta}\begin{bmatrix} \alpha-2 & \beta-2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \alpha^n & 0 \\ 0 & \beta^n \end{bmatrix}\begin{bmatrix} 1 & 2-\beta \\ -1 & \alpha-2 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\Rightarrow f_{2n} = a_n = \frac{1}{\sqrt{5}}\left((\alpha-1)^{2n} - (\beta-1)^{2n}\right),\ f_{2n+1} = \frac{1}{\sqrt{5}}\left((\alpha-1)^{2n+1} - (\beta-1)^{2n+1}\right).$$

$$\Rightarrow f_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

Diagonalizing matrices whose eigenvalues are not distinct

The **(algebraic) multiplicity** of eigenvalue λ is the largest integer $m \geq 1$ such that we can write the characteristic polynomial of A as the product $\det(A - xI) = (\lambda - x)^m p(x)$ for some polynomial $p(x)$.

Theorem Let A be an $n\times n$ matrix. Suppose A has distinct eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p$ where $p \leq n$. The following properties then holds:

(a) For each $i = 1, 2, \cdots, p$, it holds that $\dim \mathrm{Nul}(A - \lambda_i I)$ is at most the multiplicity of λ_i;

(b) A is diagonalizable if and only if the sum of the dimensions of the eigenspaces of A is n, i.e. $\sum_{i=1}^{p} \dim \mathrm{Nul}(A - \lambda_i I) = n$.

(c) Suppose A is diagonalizable and B_i is a basis for the λ_i-eigenspace. Then the union $B_1 \cup B_2 \cup \cdots \cup B_p$ is a basis for $\mathbb{R}^n$ consisting of eigenvectors of A. If the elements of this union are the vectors $v_1, v_2, \cdots, v_n$ then the matrix $P=[v_1\ v_2 \cdots v_n]$ is invertible and the matrix $D=P^{-1}AP$ is diagonal and $A=PDP^{-1}$.

e.g. $A=\begin{bmatrix} 5&0&0&0\\0&5&0&0\\1&4&-3&0\\-1&-2&0&3\end{bmatrix} \Rightarrow \det(A-xI)=(5-x)^2(-3-x)^2$. Thus the eigenvalues of A are 5 and -3, each with multiplicity 2

$$A-5I=\begin{bmatrix}0&0&0&0\\0&0&0&0\\1&4&-8&0\\-1&-2&0&-8\end{bmatrix}\sim\begin{bmatrix}1&0&8&16\\0&1&-4&-4\\0&0&0&0\\0&0&0&0\end{bmatrix}=RREF(A-5I)$$

$\Rightarrow x=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}-8\\4\\1\\0\end{bmatrix}x_3+\begin{bmatrix}-16\\4\\0\\1\end{bmatrix}x_4$. 5-eigenspace has dimension 2.

$$A+3I=\begin{bmatrix}8&0&0&0\\0&8&0&0\\1&4&0&0\\-1&-2&0&0\end{bmatrix}\sim\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}=RREF(A+3I)$$

$\Rightarrow x=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}0\\0\\1\\0\end{bmatrix}x_3+\begin{bmatrix}0\\0\\0\\1\end{bmatrix}x_4$. -3-eigenspace has dimension 2

$\Rightarrow A$ is diagonalizable. $P=\begin{bmatrix}-8&-16&0&0\\4&4&0&0\\1&0&1&0\\0&1&0&1\end{bmatrix}$, $D=\begin{bmatrix}5&&&\\&5&&\\&&-3&\\&&&-3\end{bmatrix}$.

Proposition Let A be an $n\times n$ upper-triangular matrix:

$$A=\begin{bmatrix}\lambda&*&\cdots&*\\&\lambda&\ddots&*\\&&\ddots&\vdots\\&&&\lambda\end{bmatrix}$$

If A is diagonalizable then A is equal to the diagonal matrix λI. This means that if A is not diagonal then A is not diagonalizable.

Theorem Let A be an $n\times n$ matrix. Suppose $\lambda_1, \lambda_2, \cdots \lambda_p$ are distinct eigenvalues of A. Let $d_i=\dim Nul(A-\lambda_i I)$ for $i=1,2,\cdots p$.

By the definition of an eigenvalue, $\forall i$, $1\le d_i\le n$.

1. We always have $d_1+d_2+\cdots+d_p\le n$.

2. The matrix A is diagonalizable if and only if $d_1+d_2+\cdots+d_p=n$

3. Suppose A is diagonalizable. Let $D_i=\lambda_i I_{d_i}$ and define D as the $n\times n$ diagonal matrix

$$D=\begin{bmatrix}D_1&&&\\&D_2&&\\&&\ddots&\\&&&D_p\end{bmatrix}$$

Choose n vectors $v_1, v_2, \cdots, v_n\in\mathbb{R}^n$ such that the first d_1 vectors are a basis for $Nul(A-\lambda_1 I)$, the next d_2 vectors are a basis for $Nul(A-\lambda_2 I)$, $\cdots$, the last d_p vectors are basis for $Nul(A-\lambda_p I)$. Then $A=PDP^{-1}$ for $P=[v_1\ v_2 \cdots v_n]$.

<u>Complex numbers</u>

Let $i=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$, $I_2=\begin{bmatrix}1&0\\0&1\end{bmatrix}$.

Suppose $a, b\in\mathbb{R}$. To simplify, write I_2 as 1, aI_2+bi as $a+bi$.

Consider $a=a+0i$, $bi=0+bi$, $0=0+0i$. Then

$$a+bi=a\begin{bmatrix}1&0\\0&1\end{bmatrix}+b\begin{bmatrix}0&-1\\1&0\end{bmatrix}=\begin{bmatrix}a&-b\\b&a\end{bmatrix}.$$

Define $\mathbb{C}=\{a+bi : a,b\in\mathbb{R}\}=\left\{\begin{bmatrix}a&-b\\b&a\end{bmatrix} : a,b\in\mathbb{R}\right\}$. This is called the set of complex number. Each element of $\mathbb{C}$ is a 2×2 matrix, called a complex number.

Fact $(a+bi)+(c+di)=\begin{bmatrix}a&-b\\b&a\end{bmatrix}+\begin{bmatrix}c&-d\\d&c\end{bmatrix}=\begin{bmatrix}a+c&-(b+d)\\b+d&a+c\end{bmatrix}$

$=(a+c)+(b+d)i\in\mathbb{C}$.

Clearly $(a+bi)+(c+di)=(c+di)+(a+bi)=(a+c)+(b+d)i$.

Fact $(a+bi)-(c+di)=\begin{bmatrix}a&-b\\b&a\end{bmatrix}-\begin{bmatrix}c&-d\\d&c\end{bmatrix}=\begin{bmatrix}a-c&-b+d\\b-d&a-c\end{bmatrix}$

$=(a-c)+(b-d)i\in\mathbb{C}$.

Fact $(a+bi)(c+di)=\begin{bmatrix}a&-b\\b&a\end{bmatrix}\begin{bmatrix}c&-d\\d&c\end{bmatrix}=\begin{bmatrix}ac-bd&-(bc+ad)\\bc+ad&ac-bd\end{bmatrix}$

$=(ac-bd)+(ad+bc)i. \in\mathbb{C}$.

$(a+bi)(c+di)=(c+di)(a+bi)=(ac-bd)+(ad+bc)i$.

Fact. If $a, b, x \in \mathbb{R}$, then

$$(a+bi)x = x(a+bi) = \begin{bmatrix} ax & -bx \\ bx & ax \end{bmatrix} = (ax)+(bx)i \in \mathbb{C}.$$

Fact. If $a, b, x \in \mathbb{R}$ and $x \neq 0$. Then $\frac{a+bi}{x} = \frac{a}{x} + \frac{b}{x} i \in \mathbb{C}$.

A complex number $a+bi$ is **nonzero** if $a \neq 0$ or $b \neq 0$.

$$\det(a+bi) = \det \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = a^2+b^2.$$

Fact. If $a, b, c, d \in \mathbb{R}$ and $c+di \neq 0$ then define

$$\frac{a+bi}{c+di} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} c & -d \\ d & c \end{bmatrix}^{-1} = \frac{ac+bd}{c^2+d^2} + \frac{ad-bc}{c^2+d^2} i \in \mathbb{C}$$

More generally, if $c+di \neq 0$, $\frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{c^2+d^2}$.

The **complex conjugate** of $c+di$ is defined to be

$$\overline{c+di} = (c+di)^T = c-di \in \mathbb{C}.$$

When $c+di \neq 0$, $(c+di)^{-1} = \begin{bmatrix} c & -d \\ d & c \end{bmatrix}^{-1} = \frac{1}{c^2+d^2} \begin{bmatrix} c & d \\ -d & c \end{bmatrix} = \frac{1}{c^2+d^2} \cdot \overline{c+di}$.

Since $x, y \in \mathbb{C}$ satisfy $xy = yx$ and $(xy)^T = y^T x^T$, it follows that

$$\overline{xy} = \bar{y} \cdot \bar{x} = \bar{x} \cdot \bar{y}.$$

Also, $(a+bi)+c = c+(a+bi) = (a+bi)+(c+0i) = (a+c)+bi$.

$$i^2+1 = (0+i)(0+i)+(1+0i) = 0+0i = 0 \Rightarrow i^2 = -1$$

Theorem Define the exponential function $\mathbb{C} \to \mathbb{C}$ by the convergent power series: $e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots$

Then $e^{i\pi}+1 = 0$.

Theorem (Fundamental theorem of algebra) $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is a polynomial with degree n. There are n (not necessarily distinct) complex numbers $r_1, r_2, \cdots, r_n \in \mathbb{C}$ such that $p(x) = a_n(x-r_1)(x-r_2)\cdots(x-r_n)$. $r_1, r_2, \cdots, r_n$ are called roots of $p(x)$.

Complex eigenvalues

The characteristic equation of an $n \times n$ matrix is a degree n polynomial with real coefficients. $\det(A - xI)$ has exactly n roots but some of them may be complex number.

Define $\mathbb{C}^n$ to be the set of vectors $v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ with n rows and entries $v_1, v_2, \cdots, v_n \in \mathbb{C}$. We have $\mathbb{R}^n \subset \mathbb{C}^n$.

Definition Let A be an $n \times n$ matrix with entries in $\mathbb{R}$ or $\mathbb{C}$.

The following statements are equivalent:

(1) λ is an **(complex) eigenvalue** of A;

(2) $Av = \lambda v$ for some nonzero vector $v \in \mathbb{C}^n$.

(3) $\det(A - \lambda I) = 0$.

Fact A is an $n \times n$ matrix, then A has n (not necessarily real or distinct) eigenvalues $\lambda \in \mathbb{C}$, counting repeated eigenvalues with their respective multiplicities.

Define $\bar{A}$ to be the matrix given by replacing all entries of A by their complex conjugates.

Proposition Suppose A is an $n \times n$ matrix with real entries, then $\bar{A} = A$. If $Av = \lambda v$, then $A\bar{v} = \bar{\lambda}\bar{v}$ $(v \neq 0, \bar{v} \neq 0)$.

Some final properties of eigenvalues of eigenvectors

Lemma Suppose a polynomial

$$(\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

for some complex number $\lambda_1, \lambda_2, \cdots, \lambda_n, a_0, a_1, \cdots, a_n \in \mathbb{C}$.

Then $a_n = (-1)^n$, $a_{n-1} = (-1)^{n-1}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$, $a_0 = \lambda_1 \lambda_2 \cdots \lambda_n$

Define $tr(A)$ to be the sum of the diagonal entries of A. Call $tr(A)$ the **trace** of A. $tr(A^T) = tr(A)$

Proposition If A, B are $n \times n$ matrices then $tr(A+B) = tr(A) + tr(B)$.

Theorem Suppose $\det(A - \lambda I) = (\lambda_1 - x)(\lambda_2 - x)\cdots(\lambda_n - x)$.

Then $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$ and $tr(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Corollary Suppose A is 2×2. Let $p = \det A$, $q = trA$. Then A has distinct eigenvalue if and only if $q^2 \neq 4p$.

Proposition If A is a square matrix then A and A^T have the same eigenvalues.

Proposition Let A be a square matrix. Then A is invertible if and only if 0 is not its eigenvalue. Assume A is invertible, then A and A^{-1} have the same eigenvectors, but v is an eigenvector of A with eigenvalue λ if and only if v is an eigenvector of A^{-1} with eigenvalue $1/\lambda$.

$$Av=\lambda v \Rightarrow A^{-1}v=\lambda^{-1}v.$$

Corollary If A is invertible and diagonalizable then A^{-1} is diagonalizable.

Corollary. If A is diagonalizable then A^T is diagonalizable.

Inner products and orthogonality

Definition The inner product or dot product of two vectors

$$u=\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix} \text{ and } v=\begin{bmatrix}v_1\\v_2\\\vdots\\v_n\end{bmatrix} \text{ in } \mathbb{R}^n \text{ is the scalar}$$

$$u\cdot v=u_1v_1+u_2v_2+\cdots+u_nv_n=u^Tv=v^Tu=v\cdot u.$$

Definition The length of a vector $v\in\mathbb{R}^n$ is

$$\|v\|=\sqrt{v\cdot v}=\sqrt{v_1^2+v_2^2+\cdots+v_n^2}.$$

Properties: Let $u,v,w\in\mathbb{R}^n$ and $c\in\mathbb{R}$,

(a) $u\cdot v=v\cdot u$, $(u+v)\cdot w=u\cdot w+v\cdot w$, $(cv)\cdot w=c(v\cdot w)$, $\|cv\|=|c|\|v\|$

(b) $v\cdot v=v_1^2+v_2^2+\cdots+v_n^2\geq 0$ and $\|v\|\geq 0$.

(c) $v\cdot v=0$ if and only if $\|v\|=0$ if and only if $v=0\in\mathbb{R}^n$.

(d) $u\cdot v=\|u\|\|v\|\cos\theta$, θ is the angle between u and v.

The distance between two vectors $u,v\in\mathbb{R}^n$ is the length of the difference $\|u-v\|$. A unit vector is a vector $u\in\mathbb{R}^n$ with $\|u\|=1$.

If $v\in\mathbb{R}^n$ is any nonzero vector, then the unit vector in the direction of v is $u=\frac{1}{\|v\|}v\in\mathbb{R}^n$.

Definition Two vectors $u,v\in\mathbb{R}^n$ are orthogonal if $u\cdot v=0$.

If $u,v\in\mathbb{R}^n$ are orthogonal, $u=\begin{bmatrix}a\\b\end{bmatrix}\neq 0$, then v is a scalar multiple $\begin{bmatrix}-b\\a\end{bmatrix}$.

Orthogonal complements

Let $V\subseteq\mathbb{R}^n$ be a subspace. The orthogonal complement of V is $V^\perp=\{w\in\mathbb{R}^n : v\cdot w=0 \text{ for all } v\in V\}$, pronounce "V perp".

Proposition If $V\subseteq\mathbb{R}^n$ is a subspace then its orthogonal complement $V^\perp\subseteq\mathbb{R}^n$ is also a subspace.

Theorem Suppose A is an $m\times n$ matrix. Then $(\mathrm{Col}\,A)^\perp=\mathrm{Nul}(A^T)$.

Lemma Let $V\subseteq\mathbb{R}^n$ be a subspace. If $w\in V\cap V^\perp$ then $w=0$.

Proposition Let $V\subseteq\mathbb{R}^n$ be a subspace. If $S\subseteq V$ and $T\subseteq V^\perp$ are two sets of linearly independent vectors, then $S\cup T$ is also linearly independent.

Corollary If $V\subseteq\mathbb{R}^n$ is a subspace then $\dim V^\perp\leq n-\dim V$.

Orthogonal bases and orthogonal projections

Proposition (Generalized Pythagorean theorem) Two vectors $v,w\in\mathbb{R}^n$ are orthogonal if and only if $\|u+v\|^2=\|u\|^2+\|v\|^2$.

A collection of vectors $u_1,u_2,\cdots,u_p\in\mathbb{R}^n$ is orthogonal if $u_i\cdot u_j=0$ whenever $1\leq i<j\leq p$. In particular, an orthogonal basis of $\mathbb{R}^n$ is a basis in which two vectors are orthogonal. For example, the standard basis $e_1,e_2,\cdots,e_n$ is an orthogonal basis for $\mathbb{R}^n$.

Theorem Suppose the vectors $u_1,u_2,\cdots,u_p\in\mathbb{R}^n$ are orthogonal and all nonzero. Then $u_1,u_2,\cdots,u_p$ are linearly independent.

Corollary Any set of nonzero, orthogonal vectors is an orthogonal basis for the subspace they span.

Any set of n nonzero, orthogonal vectors in $\mathbb{R}^n$ is an orthogonal basis for $\mathbb{R}^n$.

Proposition Suppose $u_1, u_2, \cdots, u_p$ is an orthogonal basis for a subspace $V \subseteq \mathbb{R}^n$. Let $y \in V$. Then we can write $y = c_1u_1 + \cdots + c_pu_p$ where $c_i = \frac{y \cdot u_i}{u_i \cdot u_i} = \frac{y \cdot u_i}{\|u_i\|^2}$.

Definition The orthogonal projection of y onto u is the vector $\hat{y} = \frac{y \cdot u}{u \cdot u} u$, which is a scalar multiple of u, and can be zero. The component of y orthogonal to u is the vector $z = y - \hat{y} = y - \frac{y \cdot u}{u \cdot u} u$. Then $z \cdot u = 0$

Observation The vectors $\hat{y}$ and z do not change if u is replaced by a nonzero scalar multiple: $\frac{y \cdot cu}{(cu) \cdot (cu)} cu = \frac{y \cdot u}{u \cdot u} = \hat{y}$.

Let $L = \mathbb{R}\text{-span}\{u\}$. Then $\hat{y}$ and z may also be called the orthogonal projection of y onto L, the component of y orthogonal to L, write $proj_L(y) = \hat{y} \in L$.

Proposition The only vector $\hat{y} \in L$ with $y - \hat{y} \in L^\perp$ is the orthogonal projection $\hat{y} = proj_L(y)$

e.g. If $y = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $L = \mathbb{R}\text{-span}\left\{\begin{bmatrix} 4 \\ 2 \end{bmatrix}\right\}$, then

$$proj_L(y) = \frac{\begin{bmatrix} 7 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix}}{\begin{bmatrix} 4 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix}} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \frac{28+12}{16+4} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}.$$

Orthonormal Vectors

A set of vectors $u_1, u_2, \cdots, u_p$ is orthonormal if the vectors are orthogonal and each vector is a unit vector. In other words, if $u_i \cdot u_j = 0$ when $i \neq j$ and $u_i \cdot u_j = 1$ if $i = j$. An orthonormal basis of a subspace is a basis that is orthonormal.

A square matrix with orthonormal columns is called an orthogonal matrix. The standard basis $e_1, e_2, \cdots, e_n$ is an orthonormal basis for $\mathbb{R}^n$.

Theorem Let U be an $n \times n$ matrix. The columns of U are orthonormal vectors if and only if $U^TU = I_n$. If U is square then its columns are orthonormal if and only if $U^T = U^{-1}$. (A matrix U is orthogonal if and only if U is square and $U^T = U^{-1}$.

Corollary If U is an orthogonal matrix then $\det(U) = \pm 1$.

Theorem Let U be an $m \times n$ matrix with orthonormal columns. Suppose $x, y \in \mathbb{R}^n$, then

1. $\|Ux\| = \|x\|$
2. $(Ux) \cdot (Uy) = x \cdot y$
3. $(Ux) \cdot (Uy) = 0$ if and only if $x \cdot y = 0$.

Orthogonal projections onto subspace

Theorem Let $W \subseteq \mathbb{R}^n$ be any subspace. Let $y \in \mathbb{R}^n$. Then there are unique vectors $\hat{y} \in W$ and $z \in W^\perp$ such that $y = \hat{y} + z$.

If $u_1, u_2, \cdots, u_p$ is an orthogonal basis for W then $\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p$ and $z = y - \hat{y}$.

Definition The vector $\hat{y}$ is called the orthogonal projection of y onto W, write $proj_W(y) = \hat{y}$.

Corollary If $W \subseteq \mathbb{R}^n$ is any subspace then $\dim W^\perp = n - \dim W$

Fact If $y \in W$ then $proj_W(y) = y$. If $y \in W^\perp$ then $proj_W(y) = 0$.

Proposition If $v \in W$ and $y \in \mathbb{R}^n$ and $v \neq proj_W(y)$ then $\|y - proj_W(y)\| < \|y - v\|$, that is, the projection $proj_W(y)$ is the vector in W that is closest to y.

Fact Suppose $u_1, u_2, \cdots, u_p$ is an orthonormal basis of W. Then $proj_W(y) = (y \cdot u_1)u_1 + (y \cdot u_2)u_2 + \cdots + (y \cdot u_p)u_p$.

Define $U = [u_1\ u_2 \cdots u_p]$, then $proj_W(y) = UU^Ty$

The Gram-Schmidt process

The Gram-Schmidt process is an algorithm that takes an arbitrary basis for some subspace of $\mathbb{R}^n$ as input, and produces an orthogonal basis of the same subspace as output.

Theorem Let $W \subseteq \mathbb{R}^n$ be a nonzero subspace. Then W has an orthogonal basis.

Gram-Schmidt process Suppose $x_1, x_2, \cdots, x_p$ is any basis for W. Then an orthogonal basis is given by $v_1, v_2, \cdots, v_p$

$v_1 = x_1$

$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$

$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2.$

$v_4 = x_4 - \frac{x_4 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_4 \cdot v_2}{v_2 \cdot v_2} v_2 - \frac{x_4 \cdot v_3}{v_3 \cdot v_3} v_3.$

$\vdots$

$v_p = x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - \cdots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}.$

Let $W_i = \mathbb{R}\text{-span}\{v_1, v_2, \cdots, v_i\}$ for each $i = 1, 2, \cdots, p$. Then $v_1, v_2, \cdots, v_i$ is an orthogonal basis for W_i and $v_{i+1} = x_{i+1} - \text{proj}_{W_i}(x_{i+1})$

Remark To find an orthonormal basis for a subspace W, first find an orthogonal basis $v_1, v_2, \cdots, v_p$. Then replace each vector v_i by $u_i = \frac{1}{\|v_i\|} v_i$.

Least-squares problems

Define the **distance** between vectors $u, v \in \mathbb{R}^n$ to be

$$\|u-v\| = \sqrt{(u-v)\cdot(u-v)} = \sqrt{(u_1-v_1)^2 + (u_2-v_2)^2 + \cdots + (u_n-v_n)^2}.$$

The distance function $\|\cdot\|$ is called the **Euclidean distance** or **L^2-distance**.

Definition If A is an $m \times n$ matrix and $b \in \mathbb{R}^m$, then a **least-squares solution** to the linear system $Ax = b$ is a vector $s \in \mathbb{R}^n$ such that $\|b - As\| \leq \|b - Ax\|$ for all $x \in \mathbb{R}^n$.

There is always at least one least-squares solution.

Lemma The least-squares solutions to $Ax = b$ are precisely those $s \in \mathbb{R}^n$ such that $As = \text{proj}_{\text{Col}A} b$.

Theorem The set of least-squares solutions to $Ax = b$ is the set of exact solutions to the linear system $A^T A x = A^T b$. This new linear system is always consistent so its set of solutions is nonempty.

Theorem Let A be an $m \times n$ matrix. The following are equivalent

(a) $Ax = b$ has a unique least-squares solution for each $b \in \mathbb{R}^m$.

(b) The columns of A are linearly independent.

(c) $A^T A$ is invertible

When these properties hold, the solution is $s = (A^T A)^{-1} A^T b$.

Suppose we have n data: $(a_1, b_1), (a_2, b_2), \cdots, (a_n, b_n)$. We want to find $\beta_0, \beta_1 \in \mathbb{R}$ such that $y = \beta_0 + \beta_1 x$ is the **line of best fit** for the data. If the points are all on the same line, then $\forall i \in \mathbb{R}$, $b_i = \beta_0 + \beta_1 a_i$, meaning that $x = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ is an exact solution for $Ax = b$ where

$$A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

If the points are not on the same line, then we need to find a least-squares solution to $Ax = b$, which is $A^T A x = A^T b$.

$\Rightarrow x = (A^T A)^{-1} A^T b$.

Symmetric matrix

A matrix is **symmetric** if $A^T = A$. This happens if A is square and $A_{ji} = A_{ij}$ for all i, j.

Proposition If A is symmetric and $k \in \mathbb{N}^*$, then A^k is also symmetric.

Proposition If A is an invertible symmetric matrix then A^{-1} is also symmetric.

Theorem Suppose A is symmetric. Then any two eigenvectors from different eigenspaces of A are orthogonal.

A matrix P is **orthogonal** if P is invertible and $P^{-1} = P^T$.

Definition A matrix A is *orthogonally diagonalizable* if there is an orthogonal matrix and a diagonal matrix D such that $A = PDP^{-1} = PDP^T$.

Lemma If A is orthogonally diagonalizable then A is symmetric

Lemma All complex eigenvalues of an $n \times n$ symmetric matrix A with real entries belong to $\mathbb{R}$.

Lemma An $n \times n$ matrix with all real eigenvalues can be written as $A = URU^T$, where U is an $n \times n$ orthogonal matrix and R is an $n \times n$ upper-triangular matrix. This is called a *Schur factorization*.

Theorem A square matrix is orthogonally diagonalizable if and only if it is symmetric.

Corollary If $A = UDU^T$ where $U = [u_1 \ u_2 \cdots u_n]$ has orthonormal columns and

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

is diagonal, then $A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T$

Each product $u_i u_i^T$ is an $n \times n$ matrix of rank 1. This is called a *spectral decomposition* of A.

Singular value decomposition

A is a $m \times n$ matrix. Then A^TA is a symmetric $n \times n$ matrix, since $(A^TA)^T = A^T(A^T)^T = A^TA$. Then A^TA has all real eigenvalues.

Lemma All eigenvalues of A^TA are nonnegative real numbers. If λ is an eigenvalue of A^TA and $v \in \mathbb{R}^n$ is a unit vector with $A^TAv = \lambda v$, then $\lambda = \|Av\|^2$.

Definition Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of A^TA arranged in decreasing order. Define $\sigma_i := \sqrt{\lambda_i}$. We call the numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ the *singular value* of A.

Theorem Suppose $v_1, v_2, \cdots, v_n$ is an orthonormal basis of $\mathbb{R}^n$ composed of eigenvectors of A^TA, arranged so that if $\lambda_i \in \mathbb{R}$ is the eigenvalue of v_i then $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Assume A has r nonzero singular values. Then $Av_1, Av_2, \cdots Av_r$ is an orthogonal basis for the column space of A and $\operatorname{rank} A = r$.

Corollary The rank of a matrix is the same as its number of nonzero singular values.

Theorem (Existence of SVDs) Let A be an $m \times n$ matrix with rank r. Suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are the nonzero singular values of A. Then we can write

$$A = U\Sigma V^T$$

where U is some $m \times m$ orthogonal matrix, V is some orthogonal matrix, Σ is the $m \times n$ matrix $\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$ where $D = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$.

Definition A factorization $A = U\Sigma V^T$ is a *singular value decomposition (SVD)* of A. U and V are not uniquely determined by A, but Σ is. The columns of U are called *left singular vectors* of A. The columns of V are called *right singular vectors* of A.

Method to find the SVD of A:

1. Find an orthogonal diagonalization of A^TA. Let the eigenvalues be the decreasing list. Then $v_1, v_2, \cdots, v_n$ is a list of orthonormal eigenvectors. Then $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$
2. For each $i = 1, 2, \cdots, r$, define $u_i := \frac{Av_i}{\|Av_i\|} = \frac{1}{\sigma_i} Av_i$. Choose $u_{r+1}, u_{r+2}, \cdots, u_m \in \mathbb{R}^n$ such that $u_1, u_2, \cdots, u_m$ is an orthonormal

basis for $\mathbb{R}^m$.

3. Let $U=[u_1\ u_2 \cdots u_n]$, $V=[v_1\ v_2 \cdots v_n]$, $\Sigma=\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$

Then $A=U\Sigma V^T$.

Definition A **persudo-inverse** of an $m\times n$ matrix A is an $n\times m$ matrix A^+ such that $AA^+A=A$, $A^+AA^+=A^+$. If A is a square, invertible matrix, then $A^+=A^{-1}$.

Theorem Every matrix has a (unique) pseudo-inverse. If $A=U\Sigma V^T$, Σ^+ is formed by transposing Σ and replacing all its nonzero entries by their reciprocals, then $A^+=V\Sigma^+U^T$.

SVDs for symmetric matrices

Suppose $A=A^T$ is an $n\times n$ symmetric matrix. The eigenvalues of A are $\lambda_1, \lambda_2, \cdots, \lambda_n \in \mathbb{R}$. Suppose the eigenvalues are ordered such that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \geq 0$. Let $D=\operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$.

Then there exists U such that $A=UDU^T$. Define $\varepsilon_i=1$ if $\lambda_i \geq 0$ and $\varepsilon_i=-1$ if $\lambda_i<0$. Let

$$E=\begin{bmatrix} \varepsilon_1 & & & \\ & \varepsilon_2 & & \\ & & \ddots & \\ & & & \varepsilon_n \end{bmatrix}$$

Proposition A singular value decomposition for the symmetric matrix $A=A^T=UDU^T$ is $A=U\Sigma V^T$, where $\Sigma=DE$ and $V=UE$. The singular values of A are $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. In general, the singular values of any symmetric matrix are just the absolute values of its eigenvalues.

SVDs for 2×2 matrices

Proposition Every 2×2 orthogonal matrix takes the form

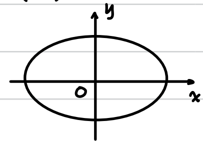
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos(\theta-\frac{\pi}{2}) & -\sin(\theta-\frac{\pi}{2}) \\ \sin(\theta-\frac{\pi}{2}) & \cos(\theta-\frac{\pi}{2}) \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Suppose U is a 2×2 orthogonal matrix. The mapping $v \mapsto Uv$ for $v\in\mathbb{R}^2$.

① $\det U=1$. $U=\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, v is rotated counterclockwise

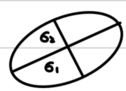
② $\det U=-1$, $U=\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$, v is reflected across $y=x$ and then rotated counterclockwise.

The **unit disc** Disc is the set of vectors $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$ with $v_1^2+v_2^2 \leq 1$. Fix $r_1, r_2 \geq 0$. Consider the set $\mathcal{E}$ of vectors $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2$ with $\left(\frac{v_1}{r_1}\right)^2+\left(\frac{v_2}{r_2}\right)^2 \leq 1$. Call $\mathcal{E}$ a **solid standard ellipse**.



Proposition. It holds that $\mathcal{E}=\left\{\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix} v : v\in \text{Disc}\right\}$.

Proposition Suppose U is some orthognal 2×2 matrix and $\Sigma=\begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix}$. Then the set of vectors $\{U\Sigma v \in \mathbb{R}^2 : v \in \text{Disc}\}$ is an **ellipse** whose radii have length r_1 and r_2.



Proposition A is a 2×2 matrix. Then $\{Av : v\in\mathbb{R}\}$ is an ellipse

The lengths of the radii are the singular values of A.