VECTOR SPACES

1. Introduction

Let V be a *set* on which two operations are defined: *addition* $(+)$ and *scalar multiplication*, *i.e.,* for $u, v \in V$ and $\lambda, \mu \in \mathbb{R}$ the *sum* of u and v is $u + v$ and the *scalar product* of u by λ is λu.
If the following properties are satisfied by all elements $u, v, w \in V$ and all scalars λ, μ:

(1) $u, v \in V \Rightarrow u + v \in V$
(2) $u + v = v + u$
(3) $u + (v + w) = (u + v) + w$ and so $= u + v + w$
(4) there is an element denoted 0 such that $0 + u = u + 0 = u$ for all $u \in V$
(5) for all $u \in V$ there is an element $-u \in V$ such that $u + (-u) = -u + u = 0$
(6) $\lambda \in \mathbb{R}, u \in V \Rightarrow \lambda u \in V$
(7) $\lambda(u + v) = \lambda u + \lambda v$
(8) $(\lambda + \mu)u = \lambda u + \mu u$
(9) $\lambda(\mu u) = (\lambda\mu)u$ and so $= \lambda\mu u$
(10) $1u = u$

then V is a vector space; the elements of V are vectors and 0 is the zero vector.

Examples.

- $V = \mathbb{R}^n$ (column vectors with n elements). Addition is standard vector addition.
- $V = \mathbb{R}^{m \times n}$ the set of $m \times n$ matrices. Addition is standard matrix addition.
- $V =$ the set of real-valued functions defined on the real line $\mathbb{R}$. If $f \equiv f(x)$ and $g \equiv g(x)$ are in V the sum of f and g is $f + g$ defined by $(f + g)(x) = f(x) + g(x)$ and the scalar product is λf defined by $(\lambda f)(x) = \lambda[f(x)]$. The zero vector 0 is the constant zero-valued function defined by $0(x) = 0$ for all x.
- Let $V = \{0\}$. Let addition be defined by $0 + 0 = 0$ and scalar multiplication by $\lambda 0 = 0$ for all λ. Then all 10 properties are satisfied and V is a vector space called the *zero vector space*.
- Let V be the set of solutions of a second order linear homogeneous differential equation:

$$y'' + a(x)y' + b(x)y = 0$$

on an interval $x_0 < x < x_1$ then

$$u \in V \Rightarrow u'' + a(x)u' + b(x)u = 0$$

$$v \in V \Rightarrow v'' + a(x)v' + b(x)v = 0$$

Then, defining $(u + v)(x) = u(x) + v(x)$

$$(u + v)'' + a(x)(u + v)' + b(x)(u + v) = 0 \Rightarrow u + v \in V$$

and for any $\lambda \in \mathbb{R}$

$$(\lambda u)'' + a(x)(\lambda u)' + b(x)(\lambda u) = \lambda[u'' + a(x)u' + b(x)u] = 0 \Rightarrow \lambda u \in V$$

and it may be verified that all properties are satisfied.

2. Subspaces.

If W is a subset of V ($W \subset V$), then W is a subspace of V if it is itself a vector space under the addition and scalar multiplication defined on V.

REMARK 2.1. To check that W is a vector space we only need to check properties 1,4,5,6. The other six properties are inherited from V. In fact we can go further:

THEOREM 2.2. *If W is a non-empty subset of V, a vector space, then W is a subspace of V if and only if*

(a): $u, v \in W \Rightarrow u + v \in W$

(b): $\lambda \in \mathbb{R}, u \in W \Rightarrow \lambda u \in W$

i.e., W is closed under addition and scalar multiplication.

PROOF. If W is a subspace, then all 10 properties hold; thus (a) and (b) hold. Conversely, if (a) and (b) hold, which are equivalent to properties 1 and 6, we have to check that properties 4 and 5 also hold. Let $u \in W$. From (b) $\lambda u \in W$ for all λ. Thus $0u = 0 \in W$, which proves Property 4 (existence of a zero element) and also $(-1)u = -u \in W$ which proves property 5. □

Examples of subspaces.

- The set of all 3×3 diagonal matrices, *i.e.,* of the form
$$\begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix}$$
is a subspace of $\mathbb{R}^{3\times 3}$.
- The set of all polynomials of degree $\leq n$, *i.e.,* $\{p_n(x) : p_n(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0\}$, where $a_0, a_1, \ldots, a_n \in \mathbb{R}$ is a subspace of the vector space of all real functions. We call this space P_n.
- Other subspaces of the space of all real functions is $C^0(-\infty, \infty)$, the space of functions which are continuous everywhere and $C^0[a, b]$ the space of functions which are continuous in the interval $[a, b]$.
- The family of subspaces $C^n(-\infty, \infty)$ and $C^n[a, b]$, $n = 0, 1, 2, \ldots$, the spaces of functions which are continuously differentiable of order n. If $m > n$ then $C^m[a, b] \subset C^n[a, b]$ and for all n $C^n(-\infty, \infty) \subset C^n[a, b]$. The limit of this sequence of subspaces is $C^\infty(-\infty, \infty)$. This is a subspace of all the other spaces. For example, if $f(x) = \sin x$, then $f \in C^n[a, b]$ for any n and any $[a, b]$ and hence $f \in C^\infty(-\infty, \infty)$.

3. Linear combinations.

A vector u is a *linear combination* of the vectors $v_1, v_2, \ldots, v_m$ if
$$u = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_m v_m = \sum_{i=1}^{m} \lambda_i v_i$$
for some scalars $\lambda_1, \ldots, \lambda_m$.

Example. To show that $u = \begin{pmatrix} -1 \\ 2 \\ 7 \end{pmatrix} \in \mathbb{R}^3$ is a linear combination of $v = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $w = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$, let $u = \lambda v + \mu w$; then

$$\begin{aligned} -1 &= \lambda - \mu \\ 2 &= 2\lambda \quad \Rightarrow \lambda = 1, \mu = 2 \\ 7 &= 3\lambda + 2\mu \end{aligned}$$

However, if we try to write $y = \begin{pmatrix} 1 \\ 2 \\ 8 \end{pmatrix}$ as a linear combination of v and w then

$$\begin{aligned} 1 &= \lambda - \mu \\ 2 &= 2\lambda \quad \Rightarrow \lambda = 1, \mu = 0, \mu = 2.5 \\ 8 &= 3\lambda + 2\mu \end{aligned}$$

The system of equations is inconsistent and thus y is not a linear combination of v and w.

DEFINITION 3.1. If $v_1, v_2, \ldots, v_r \in V$, where V is a vector space, and if every vector in V can be expressed as a linear combination of $v_1, v_2, \ldots, v_r$, then these vectors are said to *span* V.

Examples. $i = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $j = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $k = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ span $\mathbb{R}^3$, since any vector in $\mathbb{R}^3$ of the form $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ can be written as $ai + bj + ck$.

The polynomials $1, x, x^2, \ldots, x^n$ span P_n since any polynomial $p \in P_n$ can be written as $p = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$ which is a linear combination of $1, x, \ldots, x^n$.

Do the vectors $v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ $v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 1 \\ 4 \\ 4 \end{pmatrix}$ span $\mathbb{R}^3$? *i.e.*, can any vector $u = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ be expressed as

$$u = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3$$

If so, then

$$\begin{array}{ccccccc} \lambda_1 & + & \lambda_2 & + & \lambda_3 & = & a \\ 2\lambda_1 & & & + & 4\lambda_3 & = & b \\ 3\lambda_1 & + & 2\lambda_2 & + & 4\lambda_3 & = & c \end{array}$$

This has a unique solution only if the coefficient matrix

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

is invertible. But A^{-1} does not exist (since, *e.g.* $\det A = 0$), So v_1, v_2, v_3 do not span $\mathbb{R}^3$.

THEOREM 3.2. *If $v_1, v_2, \ldots, v_r$ are vectors in a vector space V, then*

(a): *the set W of all linear combinations of $v_1, v_2, \ldots, v_r$ is a subspace of V;*

(b): *W is the smallest subspace containing $v_1, v_2, \ldots, v_r$.*

PROOF. **(a)**

$$u, v \in W \Rightarrow u = \sum_{i=1}^{r} \lambda_i v_i \quad \text{and} \quad v = \sum_{i=1}^{r} \mu_i v_i$$

for some scalars $\{\lambda_i, \mu_i\}$. Then

$$u + v = \sum_{i=1}^{r} (\lambda_i + \mu_i) v_i$$

and

$$ku = \sum_{i=1}^{r}(k\lambda_i)v_i$$

for any scalar k. Thus $u + v$ and $ku \in W$.

(b) Each of $v_1, v_2, \ldots, v_r \in W$, since, *e.g.*,

$$v_1 = 1v_1 + 0v_2 + 0v_3 + \ldots + 0v_r$$

$$v_2 = 0v_1 + 1v_2 + 0v_3 + \ldots + 0v_r$$

Let W' be another subspace containing $v_1, v_2, \ldots, v_r$. Because W' is closed under addition and scalar multiplication, it contains all linear combinations of $v_1, v_2, \ldots, v_r$. Thus, W is a subspace of W'. □

Definition 3.3. The vector space spanned by a set of vectors $S = \{v_1, v_2, \ldots, v_r\}$ is called span(S) or lin(S).

4. Linear independence

If $S = \{v_1, v_2, \ldots, v_r\}$ is a set of vectors, then the equation

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_r v_r = 0$$

has *at least* one solution: $\lambda_i = 0 \; \forall i$. If there is *no other* solution, then S is a *linearly independent* set. Otherwise S is *linearly dependent*.

Remark 4.1. If S is linearly dependent, then at least one of the scalars $\{\lambda_i\}$ is non-zero. Assume $\lambda_k \neq 0$; then

$$v_k = -\frac{\lambda_1}{\lambda_k}v_1 - \frac{\lambda_2}{\lambda_k}v_2 - \ldots - \frac{\lambda_r}{\lambda_k}v_r$$

i.e., v_k is a linear combination of the other vectors in the set.

Example. Are the vectors $\begin{pmatrix}1\\1\\1\end{pmatrix}\begin{pmatrix}1\\0\\1\end{pmatrix}\begin{pmatrix}3\\2\\3\end{pmatrix}$ linearly independent?

$$\lambda_1\begin{pmatrix}1\\1\\1\end{pmatrix} + \lambda_2\begin{pmatrix}1\\0\\1\end{pmatrix} + \lambda_3\begin{pmatrix}3\\2\\3\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}$$

Therefore

$$\begin{array}{ccccccc}\lambda_1 & + & \lambda_2 & + & 3\lambda_3 & = & 0\\ \lambda_1 & + & & & 2\lambda_3 & = & 0\\ \lambda_1 & + & \lambda_2 & + & 3\lambda_3 & = & 0\end{array}$$

$\Rightarrow \lambda_3$ is arbitrary $= k$ (say)

$$\lambda_1 = -2k \qquad \lambda_2 = 2k - 3k = -k$$

Thus, there is an infinite number of non-trivial solutions and thus the set is linearly dependent.

For example,

$$\begin{pmatrix}3\\2\\3\end{pmatrix} = 2\begin{pmatrix}1\\1\\1\end{pmatrix} + 1\begin{pmatrix}1\\0\\1\end{pmatrix}$$

Example. Consider $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$$\lambda_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

or

$$\lambda_1 + \lambda_2 = 0$$

$$\lambda_1 = 0$$

Thus, $\lambda_1 = \lambda_2 = 0$ and the set is thus linearly independent.

Example. The polynomials

$$p_1 = 1 - 2x \qquad p_2 = 2 + 3x + x^2 \qquad p_3 = 7x + x^2$$

are linearly dependent since $p_3 = p_2 - 2p_1$.

Remark 4.2. 2 vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ are linearly dependent if and only if they lie on the same line through the origin and 3 vectors in $\mathbb{R}^3$ are linearly dependent if and only if they lie on the same plane through the origin.

Theorem 4.3. *Let $S = \{v_1, v_2, \ldots, v_r\}$ be vectors in $\mathbb{R}^n$. If $r > n$ then S is linearly dependent.*

Proof. Let $v_1 = \begin{pmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{n1} \end{pmatrix}$ $v_1 = \begin{pmatrix} v_{12} \\ v_{22} \\ \vdots \\ v_{n2} \end{pmatrix}$ $\ldots$

Then $\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_r v_r = 0 \Rightarrow$

$$\begin{array}{ccccccccc} v_{11}\lambda_1 & + & v_{12}\lambda_2 & + & \ldots & + & v_{1r}\lambda_r & = & 0 \\ v_{21}\lambda_1 & + & v_{22}\lambda_2 & + & \ldots & + & v_{2r}\lambda_r & = & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ v_{n1}\lambda_1 & + & v_{n2}\lambda_2 & + & \ldots & + & v_{nr}\lambda_r & = & 0 \end{array}$$

This is a system of n equations in r unknowns. Since $r > n$ the system is *underdetermined*. Thus, there is an infinity of (non-zero) solutions and the set S is linearly dependent. □

5. Bases and dimension

Definition 5.1. A finite set of vectors $S = \{v_1, v_2, \ldots, v_r\}$ in a vector space V is called a *basis* for V if S is linearly independent and S spans V.

Example. Let $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ $e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ $\ldots e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$ be a set of vectors in $\mathbb{R}^n$. The set $S = \{e_1, e_2, \ldots, e_n\}$ is linearly independent and spans $\mathbb{R}^n$, so it is a basis for $\mathbb{R}^n$ called the *standard basis* for $\mathbb{R}^n$

Example. The set $S = \{1, x, x^2, \ldots, x^n\}$ was shown to span P_n. To show that S is also linearly independent let

$$\lambda_0 + \lambda_1 x + \ldots + \lambda_{n-1} x^{n-1} + \lambda_n x^n \equiv 0$$

Noting that a polynomial $p_n \in P_n$ has *at most* n distinct zeroes, $p_n = 0$ for all $x \Rightarrow p_n$ is identically zero, *i.e.*, $\lambda_0 = \lambda_1 = \ldots = \lambda_{n-1} = \lambda_n = 0$. Thus S is linearly independent and so forms a basis for P_n called the *standard basis* for P_n.

Example. If $S = \{v_1, v_2, \ldots, v_r\}$ is a linearly independent set in a vector space V, then, since S spans span (S) (by definition), S is a basis for the subspace span (S).

DEFINITION 5.2. A non-zero vector space V is called *finite-dimensional* if it contains a finite set S which forms a basis for V. Otherwise it is *infinite-dimensional.*

The zero vector space is also defined to be finite dimensional.

THEOREM 5.3. *If $S = \{v_1, v_2, \ldots, v_n\}$ is a basis for a vector space V, then every set with more than n vectors is linearly dependent.*

PROOF. Let $W = \{w_1, w_2, \ldots, w_m\} \subset V$ with $m > n$. Then, for $i = 1, \ldots, m$

$$w_i = \lambda_{i1}v_1 + \lambda_{i2}v_2 + \ldots + \lambda_{in}v_n$$

for some scalars $\{\lambda_{ij}\}$ (say).

To show W is linearly dependent, find $k_1, \ldots, k_m$, not all zero, such that

$$k_1w_1 + k_2w_2 + \ldots + k_mw_m = 0$$

That is

$$\begin{aligned} & k_1(\lambda_{11}v_1 + \lambda_{12}v_2 + \ldots + \lambda_{1n}v_n) \\ + \quad & k_2(\lambda_{21}v_1 + \lambda_{22}v_2 + \ldots + \lambda_{2n}v_n) + \ldots \\ + \quad & k_m(\lambda_{m1}v_1 + \lambda_{m2}v_2 + \ldots + \lambda_{mn}v_n) = 0 \end{aligned}$$

Since S is linearly independent, we have

$$\begin{aligned} \lambda_{11}k_1 + \lambda_{21}k_2 + \ldots + \lambda_{m1}k_m &= 0 \\ \lambda_{12}k_2 + \lambda_{22}k_2 + \ldots + \lambda_{m2}k_m &= 0 \\ \vdots \quad \vdots \quad \vdots \\ \lambda_{1n}k_1 + \lambda_{2n}k_2 + \ldots + \lambda_{mn}k_m &= 0 \end{aligned}$$

Again, the number of equations n is less than the number of unknowns m, so there is an infinity of non-zero solutions.

Thus W is linearly independent. □

COROLLARY 5.4. *Any two bases for a finite dimensional vector space have the same number of vectors.*

PROOF. Exercise (or see Anton). □

DEFINITION 5.5. The dimension of a vector space V is the number of vectors in a basis for V. (The zero vector space has dimension 0).

Examples. $\mathbb{R}^n$ has dimension n. P_n has dimension $n + 1$.

THEOREM 5.6. **(a):** *If $S = \{v_1, v_2, \ldots, v_n\}$ is a set of n linearly independent vectors in an n-dimensional space V, then S is a basis for V.*

(b): *If S is a set of n vectors which spans an n-dimensional space V, then S is a basis for V.*

(c): *If $S = \{v_1, v_2, \ldots, v_r\}$ is a linearly independent set in V with $r < n$, then there exist vectors $v_{r+1}, \ldots, v_n$ such that $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is a basis for V.*

PROOF. **(a):** Let w be any vector in V. The set $\{v_1, v_2, \ldots, v_n, w\}$ is linearly *dependent*. Thus, setting

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n + \mu w = 0$$

for scalars $\lambda_1, \ldots, \lambda_n, \mu$.

Any non-zero solution $\Rightarrow \mu \neq 0$. Thus

$$w = -\frac{\lambda_1}{\mu} v_1 - \frac{\lambda_2}{\mu} v_2 - \ldots - \frac{\lambda_n}{\mu} v_n$$

That is, $\{v_1, v_2, \ldots, v_n\}$ spans V.

(b): Suppose $S = \{v_1, v_2, \ldots, v_n\}$ is linearly dependent.

Then for some i

$$v_i = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_{i-1} v_{i-1} + \lambda_{i+1} v_{i+1} + \ldots + \lambda_n v_n$$

Any $w \in V$ can be expressed as

$$\begin{aligned} w &= \mu_1 v_1 + \mu_2 v_2 + \ldots + \mu_i v_i + \ldots + \mu_n v_n \\ &= (\mu_1 + \mu_i \lambda_1) v_1 + (\mu_2 + \mu_i \lambda_2) v_2 + \ldots + (\mu_n + \mu_i \lambda_n) v_n \end{aligned}$$

Thus $S' = \{v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n\}$ spans V.

But S' has only $n-1$ vectors, which contradicts the assumption that V is n-dimensional.

Thus S is linearly independent and hence a basis for V.

(c): Exercise.

□

6. Matrices

DEFINITION 6.1. Let

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

The rows of A: $r_1 = (a_{11}, a_{12}, \ldots, a_{1n})$, …, $r_m = (a_{m1}, a_{m2}, \ldots, a_{mn})$ are called the row vectors of A.

The columns of A: $c_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} \ldots c_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$ are called the column vectors of A.

The row vectors span a subspace of $\mathbb{R}^n$ called the *row space* of A or *row* A.

The column vectors span a subspace of $\mathbb{R}^m$ called the *column space* of A or *col* A.

Recall that *elementary row operations* are:

(a): multiplication of a row by a non-zero constant;

(b): interchange of two rows;

(c): addition of a multiple of one row to another.

THEOREM 6.2. *If B is obtained from A by elementary row operations, row A =row B.*

PROOF. **(i):** If operation is interchange of rows, then the rows of B are the same as those of A; thus *row* A = *row* B.

(ii): If operation is scalar multiplication or scalar multiplication and addition, then the row vectors of B are linear combinations of the row vectors of A; thus they are in *row* A. Thus all linear combinations of the rows of B are in *row* A. Thus, *row* A contains *row* B. Analogously, *row* B contains *row* A. Thus *row* A = *row* B.

□

6.1. Echelon form. Recall that a matrix can be reduced to echelon form:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & a_{rr} & \dots & a_{rn} \\ 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix}$$

by elementary row operations.

Thus, if a matrix A is reduced to echelon form, its row space is unchanged. But the non-zero row vectors of the echelon form are linearly independent. Thus they form a basis for the row space of A.

Example. Find a basis for the space spanned by

$$v_1 = \begin{pmatrix} 1 \\ -2 \\ 0 \\ 3 \end{pmatrix} \qquad v_2 = \begin{pmatrix} 0 \\ 2 \\ 5 \\ 2 \end{pmatrix} \qquad v_3 = \begin{pmatrix} 2 \\ 1 \\ 0 \\ 4 \end{pmatrix}$$

This space is the row space of the matrix

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

Gauss elimination:

$$\text{row } 3 = \text{row } 3 \text{ -2 row } 1 \qquad \begin{pmatrix} 1 & -2 & 0 & 3 \\ 0 & 2 & 5 & 2 \\ 0 & 5 & 0 & -2 \end{pmatrix}$$

$$\text{row } 3\text{=row } 3\text{-2.5 row } 2 \qquad \begin{pmatrix} 1 & -2 & 0 & 3 \\ 0 & 2 & 5 & 2 \\ 0 & 0 & -12.5 & -7 \end{pmatrix}$$

Thus

$$\begin{pmatrix} 1 \\ -2 \\ 0 \\ 3 \end{pmatrix} \qquad \begin{pmatrix} 0 \\ 2 \\ 5 \\ 2 \end{pmatrix} \qquad \begin{pmatrix} 0 \\ 0 \\ -12.5 \\ -7 \end{pmatrix}$$

form a basis for span (v_1, v_2, v_3).

REMARK 6.3. col A = row A^T

THEOREM 6.4. *dim row A = dim col A.*

PROOF. see Anton ☐

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

To find row A:

$$\text{row 2=row 2 - 3 row 1} \qquad \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 2 & 4 & -2 \end{pmatrix}$$

$$\text{row 3=row 3 - 2 row 2} \qquad \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus, dim row $A = 2$.

To find col A: col A = row A^T.

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

$$\begin{array}{r} \text{row 2 = row 2 - row 1} \\ \text{row 4 = row 4 - row 1} \end{array} \qquad \begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 4 \\ 0 & -1 & -2 \end{pmatrix}$$

$$\begin{array}{r} \text{row 3 = row 3 - 2 row 2} \\ \text{row 4 = row 4 + row 2} \end{array} \qquad \begin{pmatrix} 1 & 3 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Thus dim col $A = 2 =$ dim row A.

DEFINITION 6.5. rank A = dim row A = dim col A.

THEOREM 6.6. *The following are all equivalent for any $n \times n$ matrix A:*

- *A is invertible, i.e., A^{-1} exists*
- $Ax = 0 \Rightarrow x = 0$
- *A is row-equivalent to I_n*
- *$Ax = b$ is consistent for all $b \in \mathbb{R}^n$*
- *det $A \neq 0$*
- *rank $A = n$*
- *the row vectors of A are linearly independent*
- *the column vectors of A are linearly independent*

7. Inner products and norms

7.1. Inner product spaces. An *inner product* on a vector space V is a mapping of a pair of vectors $u, v \in V$ onto a real number denoted $< u, v >$ so that

(1) $< u, v >=< v, u >$
(2) $< u + v, w >=< u, w > + < v, w >$
(3) $< \lambda u, v >= \lambda < u, v >$
(4) $< u, u >\geq 0$ and $< u, u >= 0 \iff u = 0$

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

Example. $u, v \in \mathbb{R}^n$ $< u, v >= u^T v$

$$u = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad v = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$$< u, v >= u^T v = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = (1)(1) + (2)(0) + (3)(-1) = -2$$

This is an inner product since

(1) $u^T v = v^T u$
(2) $(u + v)^T w = u^T w + v^T w$
(3) $(\lambda u)^T v = \lambda u^T v \qquad \lambda \in \mathbb{R}$
(4) $u^T u = u_1^2 + u_2^2 + \ldots + u_n^2 \geq 0$ and $= 0 \iff u = 0$

This is the standard inner product on $\mathbb{R}^n$.

Example. $u, v \in \mathbb{R}^3$ $< u, v >= u_1 v_1 + 2u_2 v_2 + 3u_3 v_3$

(1) $< u, v >=< v, u >$
(2)

$$\begin{aligned} < u + v, w > \quad &= (u_1 + v_1)w_1 + 2(u_2 + v_2)w_2 + 3(u_3 + v_3)w_3 \\ &= u_1 w_1 + 2u_2 w_2 + 3u_3 w_3 + v_1 w_1 + 2v_2 w_2 + 3v_3 w_3 \\ &=< u, w > + < v, w > \end{aligned}$$

(3) automatic
(4) $< u, u >= u_1^2 + 2u_2^2 + 3u_3^2 \geq 0$ and $= 0 \iff u = 0$

Thus, more than one inner product can be associated with any vector space.

Example. $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$

$< A, B >= a_{11} b_{11} + a_{12} b_{12} + a_{21} b_{21} + a_{22} b_{22}$

Example. $p = a_0 + a_1 x + a_2 x^2 \quad q = b_0 + b_1 x + b_2 x^2$

$< p, q >= a_0 b_0 + a_1 b_1 + a_2 b_2$ is an inner product on the space P_2

Example. $u, v \in C^0[a, b]$

$< u, v >= \int_a^b u(x) v(x)\, dx$

This clearly satisfies the first three properties and also

$< u, v >= \int_a^b u^2(x)\, dx \geq 0$ and $= 0 \iff u \equiv 0$

This is the standard inner product on $C^0[a, b]$

Remark 7.1. For $u, v \in \mathbb{R}^3$, $< u, v >= u^T v = u_1 v_1 + u_2 v_2 + u_3 v_3 = u.v$

i.e., the dot product of the vectors. Thus, here $< u,v >= |u|\,|v|\cos\theta$ where θ is the angle between u and v. Thus $< u,v >^2 = |u|^2|v|^2\cos^2\theta$. Here, $|u| = \sqrt{u_1^2+u_2^2+u_3^2} = \sqrt{< u,u >}$. Therefore, $< u,v >^2 = < u,u >< v,v > \cos^2\theta$. and, since $\cos^2\theta \le 1$, we have $< u,v >^2 \le < u,u >< v,v >$.
Will this inequality hold for all inner products?

THEOREM 7.2. (**Cauchy-Schwarz inequality**) *For any two vectors in an inner product space* $< u,v >^2 \le < u,u >< v,v >$

PROOF. If $u = 0$ then inequality clearly holds. So assume $u \neq 0$.
Let $a = < u,u >$, $b = 2 < u,v >$, $c = < v,v >$
For any $t \in \mathbb{R}$, we have

$$\begin{aligned} 0 &\le < tu+v, tu+v > \\ &= < u,u > t^2 + 2 < u,v > t + < v,v > \\ &= at^2 + bt + c \end{aligned}$$

Since the quadratic $at^2+bt+c \ge 0$ for all t, it either has no real roots $(b^2-4ac<0)$ or 1 repeated real root $(b^2-4ac=0)$.
In either case, $b^2-4ac \le 0$ or $4 < u,v >^2 \le 4 < u,u >< v,v >$ as required. □

Example. $u,v \in C^0[a,b]$. Then Cauchy-Schwarz $\Rightarrow$

$$\left[\int_a^b u(x)v(x)\,dx\right]^2 \le \left[\int_a^b u^2(x)\,dx\right]\left[\int_a^b v^2(x)\,dx\right]$$

7.2. Normed vector spaces. A *norm* is a mapping of a vector v onto $\mathbb{R}$, i.e. an association with v of a real number $\|v\|$ called the norm of v which satisfies the properties:

(1) $\|v\| \ge 0$ and $\|v\| = 0 \iff v = 0$
(2) $\|\lambda v\| = |\lambda|\,\|v\| \quad \lambda \in \mathbb{R}$
(3) $\|u+v\| \le \|u\| + \|v\|$ the Triangle Inequality

Example. $\|v\| = < v,v >^{1/2}$ for some inner product $< .,. >$
Properties 1 and 2 follow immediately.

$$\begin{aligned} \text{Property 3: } \|u+v\|^2 &= < u+v, u+v > \\ &= < u,u > + 2 < u,v > + < v,v > \\ &\le < u,u > + 2 < u,u >^{1/2} < v,v >^{1/2} + < v,v > \\ &\quad \text{by Cauchy Schwarz} \\ &= \|u\|^2 + 2\|u\|\,\|v\| + \|v\|^2 \\ &= (\|u\| + \|v\|)^2 \end{aligned}$$

as required. This norm is called the norm *induced* from the inner product.

REMARK 7.3. Cauchy Schwarz $\Rightarrow < u,v >^2 \le \|u\|^2\|v\|^2$ and so

$$\frac{< u,v >^2}{\|u\|^2\|v\|^2} \le 1$$

or

$$-1 \le \frac{< u,v >}{\|u\|\,\|v\|} \le 1$$

Thus, there is a unique $0 \leq \theta \leq \pi$ such that

$$\cos\theta = \frac{< u, v >}{\|u\| \, \|v\|}$$

which we call the *angle between* u and v.

Example. $A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ $B = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$
Using the inner product defined earlier, $< A, B >= 2$ and the induced norm $\|A\| = \sqrt{1+0+4+1} = \sqrt{6}$ and $\|B\| = \sqrt{0+4+1+0} = \sqrt{5}$. Thus, $\cos\theta = \frac{2}{\sqrt{30}}$ and $\theta = \cos^{-1}\frac{2}{\sqrt{30}}$.

7.3. Orthogonality.

REMARK 7.4. If $< u, v >= 0$, then $\theta = \pi/2$ and u and v are said to be *orthogonal* to each other.

REMARK 7.5. Orthogonality depends on the choice of inner product.

Example. $< u, v >= \int_{-1}^{1} u(x)v(x)\,dx$
If $u = x$ and $v = x^2$, then $< u, v >= \int_{-1}^{1} x\,x^2\,dx = \left[\frac{x^4}{4}\right]_{-1}^{1} = 0.$

DEFINITION 7.6. A set of vectors $\{v_1, v_2, \ldots, v_n\}$ in an inner product space is orthogonal if

$$< v_i, v_j >= 0 \quad \text{for} \quad i \neq j$$

for all $i, j = 1, \ldots, n$. If also $\|v_i\| = 1$ for all i then the set is *orthonormal*, in which case,

$$< v_i, v_j >= \delta_{ij} \quad \text{the Kronecker delta}$$

where $\delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases}$

Example. $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $v_2 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$ $v_3 = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$
The set $\{v_1, v_2, v_3\}$ is orthonormal.

REMARK 7.7. If v is a non-zero vector in a normed vector space, then the vector $u = v/\|v\|$ has norm 1, since

$$\left\| \frac{v}{\|v\|} \right\| = \frac{1}{\|v\|}\|v\| = 1$$

This process is called *normalising* v.

THEOREM 7.8. *If $S = \{v_1, \ldots, v_n\}$ is an orthonormal basis for an inner product space V and $u \in V$, then*

$$u =< u, v_1 > v_1 + \ldots + < u, v_n > v_n \quad = \quad \sum_{i=1}^{n} < u, v_i > v_i$$

PROOF. S is a basis. Therefore, for some scalars λ_i $u = \sum_{i=1} \lambda_i v_i$ But, for any $v_j \in S$ $< u, v_j >= \sum_{i=1}^{n} \lambda_i < v_i, v_j >= \sum_{i=1}^{n} \lambda_i \delta_{ij} = \lambda_j$ as required. □

Example. $v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $v_2 = \begin{pmatrix} -4/5 \\ 0 \\ 3/5 \end{pmatrix}$ $v_3 = \begin{pmatrix} 3/5 \\ 0 \\ 4/5 \end{pmatrix}$ form an orthonormal basis for $\mathbb{R}^3$.

Express $u = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ as a linear combination of v_1, v_2 and v_3.

$< u, v_1 >= 2$ $< u, v_2 >= -4/5 + 9/5 = 1$ $< u, v_3 >= 3/5 + 12/5 = 3$

Thus, $u = 2v_1 + v_2 + 3v_3$.

THEOREM 7.9. *If $S = \{v_1, v_2, \ldots, v_n\}$ is an orthogonal set of non-zero vectors in an inner product space, then S is linearly independent.*

PROOF. Let $\sum_{i=1}^{n} \lambda_i v_i = 0$

Then, for all $v_j \in S$, $\sum_{i=1}^{n} \lambda_i < v_i, v_j >= 0$

that is $\lambda_j < v_j, v_j >= 0$

$v_j \neq 0 \Rightarrow < v_j, v_j >\neq 0 \Rightarrow \lambda_j = 0$

Thus, for all j $\lambda_j = 0$ and so S is linearly independent. □

Example. $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ $v_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ $v_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$

form an orthogonal set and therefore are linearly independent.

Thus they form a basis for $\mathbb{R}^3$.

7.4. Orthogonal projections.

THEOREM 7.10. *Let V be an inner product space and $S = \{v_1, v_2, \ldots, v_r\}$ an orthonormal set in V. Let $W = span\{v_1, v_2, \ldots, v_r\}$. Then, all vectors $u \in V$ can be expressed in the form $u = w_1 + w_2$ where $w_1 \in W$ and $w_2 \perp W$.*

PROOF. Let $w_1 = \sum_{i=1}^{r} < u, v_i > v_i$ and $w_2 = u - w_1$

Clearly $u = w_1 + w_2$ and $w_1 \in W$.

To show $w_2 \perp W$: for any $v_j \in S$ we have

$$\begin{aligned} < w_2, v_j > &= < u, v_j > - \sum_{i=1}^{r} < u, v_i >< v_i, v_j > \\ &= < u, v_j > - < u, v_j >= 0 \end{aligned}$$

and so, for any $v \in W$ $v = \sum_{i=1}^{r} < v, v_i > v_i$

and thus $< v, w_2 >= \sum_{i=1}^{r} < v, v_i >< v_i, w_2 >= 0$.

That is w_2 is orthogonal to all vectors in W and so $w_2 \perp W$ □

DEFINITION 7.11. w_1 is called the **orthogonal projection of** u **onto** W or $\text{proj}_W u$. $w_2 = u - \text{proj}_W u$ is the component of u orthogonal to W.

Example. Let $W = \text{span}\{v_1, v_2\}$

where $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 0 \\ 4/5 \\ -3/5 \end{pmatrix}$

(which are orthonormal with respect to the standard inner product)

If $u = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ then

$$\begin{aligned} \text{proj}_W u &= < u, v_1 > v_1 + < u, v_2 > v_2 \\ &= 1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (8/5 - 9/5) \begin{pmatrix} 0 \\ 4/5 \\ -3/4 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -4/25 \\ 3/25 \end{pmatrix} = \begin{pmatrix} 1 \\ -4/25 \\ 3/25 \end{pmatrix} \end{aligned}$$

and $u - \text{proj}_W u = \begin{pmatrix} 0 \\ 54/25 \\ 72/25 \end{pmatrix}$ is orthogonal to v_1 and v_2 and hence to W.

THEOREM 7.12. **(Gram-Schmidt)** *Every non-zero finite dimensional inner product space has an orthonormal basis.*

PROOF. Construction of orthonormal basis.

Let V be an n-dimensional inner product space with a basis $S = \{v_1, v_2, \dots, v_n\}$.

To produce an orthonormal basis:

(1) $u_1 = v_1/\|v_1\|$

(2) Construct a vector u_2 of norm 1 (a unit vector) orthogonal to u_1.

First, find the component of v_2 orthogonal to $W_1 = \text{span}\{u_1\}$.

This is $v_2 - \text{proj}_{W_1} v_2 = v_2 - < v_2, u_1 > u_1$

Normalise it to obtain

$$u_2 = \frac{v_2 - < v_2, u_1 > u_1}{\|v_2 - < v_2, u_1 > u_1\|}$$

(3)

$$u_3 = \frac{v_3 - < v_3, u_1 > u_1 - < v_3, u_2 > u_2}{\|v_3 - < v_3, u_1 > u_1 - < v_3, u_2 > u_2\|}$$

and so on until u_n is obtained. □

Example. Construct an orthonormal basis for $\mathbb{R}^3$ starting form the basis $v_1 = (1, 1, 1)$ $v_2 = (-1, 1, 0)$ $v_3 = (1, 2, 1)$

(1) $u_1 = v_1/\|v_1\| = \frac{1}{\sqrt{3}}(1, 1, 1) = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$

(2)

$$\begin{aligned} v_2 - \text{proj}_{W_1} v_2 &= v_2 - < v_2, u_1 > u_1 \\ &= (-1, 1, 0) - 0u_1 = (-1, 1, 0) \end{aligned}$$

Thus $u_2 = \frac{1}{\sqrt{2}}(-1, 1, 0) = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$

(3)

$$\begin{aligned} v_3 - \text{proj}_{W_2} v_3 &= v_3 - < v_3, u_1 > u_1 - < v_3, u_2 > u_2 \\ &= (1,2,1) - \frac{4}{\sqrt{3}}\frac{1}{\sqrt{3}}(1,1,1) - \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}(-1,1,0) \\ &= (1,2,1) - \left(\frac{4}{3},\frac{4}{3},\frac{4}{3}\right) - \left(-\frac{1}{2},\frac{1}{2},0\right) \\ &= \left(\frac{1}{6},\frac{1}{6},-\frac{1}{3}\right) = \frac{1}{6}(1,1,-2) \end{aligned}$$

Thus $\|v_3 - \text{proj}_{W_2} v_3\| = \frac{1}{6}\sqrt{6} = \frac{1}{\sqrt{6}}$

and so $u_3 = \frac{1}{\sqrt{6}}(1,1,-2)$

Example. Construct orthonormal basis for P_2 with respect to the inner product

$$< p, q >= \int_{-1}^{1} p(x)q(x)\, dx$$

We start from the standard basis $v_1 = 1$ $v_2 = x$ $v_3 = x^2$

(1) $u_1 = v_1/\|v_1\|$

$$\|v_1\|^2 = \int_{-1}^{1} 1^2\, dx = [x]_{-1}^{1} = 2 \qquad \Rightarrow \|v_1\| = \sqrt{2}$$

Thus $u_1 = 1/\sqrt{2}$

(2) $v_2 - \text{proj}_{W_1} v_2 \quad = \quad v_2 - < v_2, u_1 > u_1$

$$< v_2, u_1 >= \frac{1}{\sqrt{2}}\int_{-1}^{1} x\, dx = \frac{1}{\sqrt{2}}\left[\frac{x^2}{2}\right]_{-1}^{1} = 0$$

Thus $u_2 = v_2/\|v_2\|$

$$\|v_2\|^2 = \int_{-1}^{1} x^2\, dx = \frac{1}{3}[x^3]_{-1}^{1} = 2/3$$

Thus $\|v_2\| = \sqrt{\frac{2}{3}}$ and so $u_2 = \sqrt{\frac{3}{2}}x$

(3) $v_3 - \text{proj}_{W_2} v_3 \quad = \quad v_3 - < v_3, u_1 > u_1 - < v_3, u_2 > u_2$

$$< v_3, u_1 >= \frac{1}{\sqrt{2}}\int_{-1}^{1} x^2\, dx = \frac{1}{3\sqrt{2}}\left[x^3\right]_{-1}^{1} = \sqrt{2}/3$$

$$< v_3, u_2 >= \sqrt{\frac{3}{2}}\int_{-1}^{1} x^2\, x\, dx = \sqrt{\frac{3}{2}}\,\frac{1}{4}\left[x^4\right]_{-1}^{1} = 0$$

Thus $v_3 - \text{proj}_{W_2} v_3 \quad = \quad x^2 - \frac{\sqrt{2}}{3}\frac{1}{\sqrt{2}} = x^2 - 1/3$

Thus $u_3 = \frac{x^2 - 1/3}{\|x^2 - 1/3\|}$

$$\begin{aligned} \|x^2 - 1/3\|^2 &= \int_{-1}^{1} (x^2 - 1/3)^2\, dx \\ &= \int_{-1}^{1} (x^4 - 2/3\, x^2 + 1/9)\, dx \\ &= \left[(x^5/5 - 2/9\, x^3 + x/9)\right]_{-1}^{1} \\ &= 2/5 - 4/9 + 2/9 \quad = \quad 8/45 \end{aligned}$$

Thus $\|x^2 - 1/3\| = \sqrt{\frac{8}{45}} = \frac{2}{3}\sqrt{\frac{2}{5}}$

and so $u_3 = \frac{\sqrt{5}}{2\sqrt{2}}(3x^2 - 1)$

THEOREM 7.13. **Projection Theorem.** *If W is a finite dimensional subspace of an inner product space V, then every $u \in V$ can be expressed uniquely as*

$$u = w_1 + w_2$$

with $w_1 \in W$ and $w_2 \perp W$

PROOF. By Gram-Schmidt W has an orthonormal basis $\{v_1, v_2, \ldots, v_r\}$
and $w_1 = \text{proj}_W u$ $w_2 = u - \text{proj}_W u$ satisfy the requirements.
To prove uniqueness, assume $u = z_1 + z_2$ with $z_1 \in W$ and $z_2 \perp W$.
Then subtracting:

$$\begin{aligned} u &= w_1 + w_2 \\ u &= z_1 + z_2 \\ \Rightarrow \quad 0 &= w_1 - z_1 + w_2 - z_2 \\ \text{or} \quad w_1 - z_1 &= z_2 - w_2 \end{aligned}$$

Thus $z_2 - w_2 \perp W$. But $w_1 - z_1 \in W \quad \Rightarrow \quad z_2 - w_2 \in W$
Therefore $< z_2 - w_2, z_2 - w_2 >= 0$ which implies $z_2 = w_2$ and $z_1 = w_1$
Thus the projection is unique. □

THEOREM 7.14. **Best Approximation Theorem.** *If W is a finite dimensional subspace of an inner product space V and $u \in V$, then $proj_W u$ is the best approximation to u in W. i.e., for all $w \in W$, $\|u - proj_W u\| \leq \|u - w\|$*

PROOF. For any $w \in W$ we have

$$\begin{array}{ccccc} u - w & = & u - \text{proj}_W u & + & \text{proj}_W u - w \\ & & \uparrow & & \uparrow \\ & & \perp W & & \in W \end{array}$$

Thus $u - \text{proj}_W u$ is orthogonal to $\text{proj}_W u - w$

REMARK 7.15.

$$\begin{aligned} u \perp v \quad &\Rightarrow \quad < u + v, u + v >=< u, u > +2 < u, v > + < v, v > \\ &\Rightarrow \quad \|u + v\|^2 = \|u\|^2 + \|v\|^2 \end{aligned}$$

(this is just Pythagoras' theorem).

Thus $\|u - w\|^2 = \|u - \text{proj}_W u\|^2 + \|\text{proj}_W u - w\|^2$ and so $\|u - \text{proj}_W u\|^2 \leq \|u - w\|^2$ as required □

8. Application to approximation problems

Find the best possible approximation over an interval $[a, b]$ to some function f using only approximations from a specified subspace of $C^0[a, b]$. 'Best possible' here means 'smallest error'.
We could define the error as $\int_a^b |f(x) - g(x)| \, dx$ where g is the approximating function. The preferred measure though is the *mean square error* defined by

$$\int_a^b [f(x) - g(x)]^2 \, dx$$

Note that if $< f, g >= \int_a^b f(x)g(x)\,dx$ then

$$\int_a^b [f(x) - g(x)]^2 \; dx =< f - g, f - g >= \|f - g\|^2$$

Thus the mean square error is minimised if $\|f - g\|$ is minimised.

Bu, from the best approximation theorem, this happens if $g = \text{proj}_W f$

In this case, g is called the *least squares approximation* to f from the space W.

Example. Fourier Series

Let T_n be the space of trigonometric polynomials of degree $\leq n$.

t is a trigonometric ploynomial of degree n if $t(x) = c_0 + \sum_{k=1}^{n} (c_k \cos kx + d_k \sin kx)$ and c_n and d_n are not both zero.

The functions $1, \quad \cos x, \quad \cos 2x, \ldots, \cos nx, \quad \sin x, \quad \sin 2x, \ldots, \sin nx$ form a basis for T_n and so T_n is $2n+1$-dimensional.

Consider the least squares approximation of a continuous function f on the interval $[-\pi, \pi]$ by $g \in T_n$.

We know that $g = \text{proj}_{T_n} f$ but to compute it we must have an orthonormal basis $g_0, g_1 \ldots g_{2n}$ for T_n so that

$$\text{proj}_{T_n} f \quad = \quad \sum_{k=0}^{2n} < f, g_k > g_k$$

We obtain this basis by using the Gram-Schmidt process with the inner product $< u, v >= \int_{-\pi}^{\pi} u(x)v(x)\,dx$

Thus, $g_0 = 1/\|1\|$. $\|1\|^2 = \int_{-\pi}^{\pi} 1\,dx = 2\pi$ and so $g_0 = 1/\sqrt{2\pi}$

$$g_1 = \frac{\cos x - < \cos x, g_0 > g_0}{\| \cos x - < \cos x, g_0 > g_0\|}$$

$$< \cos x, g_0 >= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \cos x\,dx = 0$$

Thus $g_1 = \cos x / \| \cos x\|$

$$\begin{aligned} \| \cos x\|^2 &= \int_{-\pi}^{\pi} \cos^2 x\,dx \\ &= \frac{1}{2} \int_{-\pi}^{\pi} (1 + \cos 2x) \; dx \\ &= \frac{1}{2} \left[x + \frac{1}{2} \sin 2x \right]_{-\pi}^{\pi} \; dx \\ &= \pi \end{aligned}$$

Thus $g_1 = \dfrac{1}{\sqrt{\pi}} \cos x$

Similarly, we can obtain

$$g_k = \begin{cases} \frac{1}{\sqrt{\pi}} \cos kx, & k = 1, 2, \ldots, n \\ \frac{1}{\sqrt{\pi}} \sin(k - n)x, & k = n + 1, n + 2, \ldots, 2n \end{cases}$$

Thus

$$\begin{aligned} \text{proj}_{T_n} f = \frac{1}{\sqrt{2\pi}} < f, g_0 > \quad &+ \quad \frac{}{\sqrt{\pi}} \sum_{k=1}^{n} (< f, g_k > \cos kx \\ &\qquad + < f, g_{n+k} > \sin kx) \end{aligned}$$

or if we set

$$\begin{aligned} a_0 &= \frac{1}{\sqrt{2\pi}} < f, g_0 > \\ a_k &= \frac{1}{\sqrt{\pi}} < f, g_k > \quad k = 1, \ldots, n \\ b_k &= \frac{1}{\sqrt{\pi}} < f, g_{n+k} > \quad k = 1, \ldots, n \end{aligned}$$

$$\mathrm{proj}_{T_n} f = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

where

$$a_0 = \frac{1}{\sqrt{2\pi}} < f, g_0 >= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\, dx$$

$$a_k = \frac{1}{\sqrt{\pi}} < f, g_k >= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx\, dx$$

$$b_k = \frac{1}{\sqrt{\pi}} < f, g_{n+k} >= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx\, dx$$

that is $\mathrm{proj}_{T_n} f$ is the nth partial sum of the Fourier series for f. As $n \to \infty$ the mean square error $\to 0$

Example. Find the least squares approximation of e^x over the interval $[0, 1]$ by a straight line.

A straight line takes the form $g(x) = a + bx$. $g \in P_1$. We start with the standard basis $v_1 = 1$, $v_2 = x$ and produce first an orthonormal basis u_1, u_2 for P_1.

$u_1 = v_1/\|v_1\|$ $\|v_1\| = \|1\| = \int_0^1 1\, dx = [x]_0^1 = 1$

Thus $u_1 = v_1 = 1$

$$u_2 = \frac{v_2 - < v_2, u_1 > u_1}{\|v_2 - < v_2, u_1 > u_1\|}$$

$< v_2, u_1 >= \int_0^1 x\, dx = \left[\frac{x^2}{2}\right]_0^1 = 1/2$

Thus $v_2 - < v_2, u_1 > u_1 = x - 1/2$

$$\begin{aligned} \|x - 1/2\|^2 &= \int_0^1 (x - 1/2)^2\, dx \\ &= \int_0^1 \left(x^2 - x + 1/4\right)\, dx \\ &= \left[\frac{x^3}{3} - \frac{x^2}{2} + \frac{x}{4}\right]_0^1 \\ &= 1/3 - 1/2 + 1/4 = 1/12 \end{aligned}$$

Thus $\|x - 1/2\| = 1/\sqrt{12} = 1/2\sqrt{3}$ and

$$u_2 = \sqrt{3}(2x - 1)$$

The best approximation to e^x is $\mathrm{proj}_{P_1} e^x =< e^x, u_1 > u_1 + < e^x, u_2 > u_2$

$$< e^x, u_1 >= \int_0^1 e^x\, dx = [e^x]_0^1 = e - 1$$

$$\begin{aligned} < e^x, u_2 > &= \sqrt{3} \int_0^1 (2x - 1) e^x\, dx \\ \text{by parts :} & \quad p = 2x - 1 \qquad dq = e^x\, dx \\ & \quad dp = 2\, dx \qquad q = e^x \\ \Rightarrow < e^x, u_2 > &= \sqrt{3} \left\{ [(2x - 1)e^x]_0^1 - 2 \int_0^1 e^x\, dx \right\} \\ &= \sqrt{3} \{e + 1 - 2(e - 1)\} = \sqrt{3}(3 - e) \end{aligned}$$

and so

$$\text{proj}_{P_1} e^x = e - 1 + 3(3-e)(2x-1) = 4e - 10 + (18 - 6e)x$$

Example. Let f be a function defined on $[0,3]$ and let f have the following values at the points $x_i, \quad i = 1,\dots,4$

i	x_i	$f(x_i)$
1	0	1
2	1	1/2
3	2	1/3
4	3	1/4

Find the least squares approximation to the function f in both P_1 and P_2 with respect to the inner product defined by

$$< p, q >= \sum_{i=1}^{4} p(x_i)q(x_i)$$

where $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ and $x_4 = 3$
We first construct orthonormal bases for P_1 and P_2 starting from the standard basis $v_1 = 1$, $v_2 = x$ for P_1 and the standard basis v_1, v_2 and $v_3 = x^2$ for P_2

$$u_1 = v_1/\|v_1\| = 1/\|1\|$$

$$\|1\|^2 = \sum_{i=1}^{4} 1^2 = 4 \Rightarrow \|1\| = 2 \quad \Rightarrow \quad u_1 = 1/2$$

$$u_2 = \frac{v_2 - < v_2, u_1 > u_1}{\|v_2 - < v_2, u_1 > u_1\|}$$

$$< v_2, u_1 >= \frac{1}{2}\sum_{i=1}^{4} x_i = \frac{1}{2}(0+1+2+3) = 3$$

Thus

$$u_2 = \frac{x - 3/2}{\|x - 3/2\|}$$

$$\begin{aligned} \|x - 3/2\|^2 &= \sum_{i=1}^{4}(x_i - 3/2)^2 \\ &= (0-3/2)^2 + (1-3/2)^2 + (2-3/2)^2 + (3-3/2)^2 \\ &= 9/4 + 1/4 + 1/4 + 9/4 \\ &= 5 \end{aligned}$$

Thus $u_2 = \dfrac{1}{\sqrt{5}}(x - 3/2)$

$$u_3 = \frac{v_3 - < v_3, u_1 > u_1 - < v_3, u_2 > u_2}{\|v_3 - < v_3, u_1 > u_1 - < v_3, u_2 > u_2\|}$$

$$< v_3, u_1 >= \frac{1}{2}\sum_{i=1}^{4} x_i^2 = \frac{1}{2}(0+1+4+9) = 7$$

$$\begin{aligned} < v_3, u_2 > &= \frac{1}{\sqrt{5}} \sum_{i=1}^{4} x_i^2 (x_i - 3/2) \\ &= \frac{1}{\sqrt{5}} \left(0 + 1(1-3/2) + 4(2-3/2) + 9(3-3/2))\right) \\ &= \frac{1}{\sqrt{5}} \left(0 - 1/2 + 2 + 27/2))\right) \\ &= \frac{1}{\sqrt{5}} 15 \quad = \quad 3\sqrt{5} \end{aligned}$$

Thus

$$\begin{aligned} u_3 &= \frac{x^2 - 7/2 - 3(x - 3/2)}{\|x^2 - 7/2 - 3(x - 3/2)\|} \\ &= \frac{x^2 - 3x + 1}{\|x^2 - 3x + 1\|} \end{aligned}$$

$$\begin{aligned} \|x^2 - 3x + 1\|^2 &= \sum_{i=1}^{4} (x_i^2 - 3x_i + 1)^2 \\ &= 1 + (1 - 3 + 1)^2 + (4 - 6 + 1)^2 + (9 - 9 + 1)^2 \\ &= 1 + 1 + 1 + 1 \quad = \quad 4 \end{aligned}$$

Thus $\|x^2 - 3x + 1\| = 2$ and so $u_3 = \frac{1}{2}(x^2 - 3x + 1)$

The least squares approximation to f in P_1 is then

$$\text{proj}_{P_1} f = < f, u_1 > u_1 + < f, u_2 > u_2$$

$$\begin{aligned} < f, u_1 > &= \frac{1}{2} \sum_{i=1}^{4} f(x_i) \\ &= \frac{1}{2} (1 + 1/2 + 1/3 + 1/4) \\ &= \frac{12 + 6 + 4 + 3}{24} = \frac{25}{24} \end{aligned}$$

$$\begin{aligned} < f, u_2 > &= \frac{1}{\sqrt{5}} \sum_{i=1}^{4} (x_i - 3/2) f(x_i) \\ &= \frac{1}{\sqrt{5}} ((-3/2)(1) + (1 - 3/2)(1/2) + (2 - 3/2)(1/3) + \\ (3 - 3/2)(1/4)) \\ &= \frac{1}{\sqrt{5}} (-3/2 - 1/4 + 1/6 + 3/8) \\ &= \frac{1}{\sqrt{5}} \frac{-36 - 6 + 4 + 9}{24} \\ &= -\frac{29}{24\sqrt{5}} \end{aligned}$$

Thus

$$\begin{aligned} \text{proj}_{P_1} f &= \frac{25}{48} - \frac{29}{120}(x - 3/2) \\ &= \frac{1}{120} (19 - 29x) \end{aligned}$$

The mean square error is $\|f - \text{proj}_{P_1} f\|^2$

The best approximation to f in P_2 is

$$\text{proj}_{P_2} f = < f, u_1 > u_1 + < f, u_2 > u_2 + < f, u_3 > u_3$$

$$
\begin{aligned}
< f, u_3 > &= \frac{1}{2}\sum_{i=1}^{4}(x_i^2 - 3x_i + 1)f(x_i) \\
&= \frac{1}{2}\left(1 + (1-3+1)(1/2) + (4-6+1)(1/3) + \right. \\
&\quad \left. (9-9+1)(1/4)\right) \\
&= \frac{1}{2}\left(1 - 1/2 - 1/3 + 1/4\right) \\
&= \frac{12-6-4+3}{24} \\
&= \frac{5}{24}
\end{aligned}
$$

Thus

$$< f, u_3 > u_3 = \frac{5}{48}\left(x^2 - 3x + 1\right)$$

and so

$$
\begin{aligned}
\mathrm{proj}_{P_2} f &= \frac{1}{120}(19 - 29x) + \frac{5}{48}\left(x^2 - 3x + 1\right) \\
&= \frac{1}{240}\left(63 - 133x + 25x^2\right)
\end{aligned}
$$

9. Eigenvalues and eigenvectors

If $A \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n$ then x is an eigenvector of A if $x \neq 0$ and $Ax = \lambda x$ for some scalar λ, which is called an eigenvalue of A. x is the eigenvector corresponding to the eigenvalue λ.

$$
\begin{aligned}
Ax &= \lambda x \\
\Rightarrow (A - \lambda I)x &= 0 \\
\text{and } x \neq 0 \Rightarrow \det(A - \lambda I) &= 0
\end{aligned}
$$

The equation $\det(A - \lambda I) = 0$ or equivalently $\det(\lambda I - A) = 0$ is called the characteristic equation of A. $\det(A - \lambda I)$ is a polynomial of degree n called the characteristic polynomial of A and so there are n roots of this equation, not necessarily all distinct, and some or all of which may be complex.

Example. $A = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$ The characteristic equation is

$$\begin{vmatrix} 1-\lambda & 2 \\ -1 & 4-\lambda \end{vmatrix} = 0$$

which is a quadratic equation

$$(1-\lambda)(4-\lambda) - (-1)(2) = 0$$

Thus

$$
\begin{aligned}
4 - 5\lambda + \lambda^2 + 2 &= 0 \\
\text{or } \lambda^2 - 5\lambda + 6 &= 0
\end{aligned}
$$

This has 2 real roots $\lambda_1 = 2$ and $\lambda_2 = 3$ which are the eigenvalues of A.

To find the corresponding eigenvectors we substitute each of these values in turn into the equation $(A - \lambda I)x = 0$:

$$
\begin{aligned}
(A - 2I)x &= 0 \\
\Rightarrow \begin{pmatrix} -1 & 2 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\end{aligned}
$$

Note that we have an underdetermined system of equations

$$-x_1 + 2x_2 = 0$$

Thus if we set x_2 to be an arbitrary constant k then $x_1 = 2k$ and an eigenvector corresponding to $\lambda_1 = 2$ takes the form

$$k\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

In fact this is a vector space called the eigenspace of A corresponding to $\lambda_1 = 2$, which in this case is one-dimensional.

Similarly

$$\begin{aligned} (A-3I)x &= 0 \\ \Rightarrow \begin{pmatrix} -2 & 2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned}$$

Again we have an underdetermined system of equations

$$-x_1 + x_2 = 0$$

and so setting x_2 to be an arbitrary constant k then $x_1 = k$ and an eigenvector corresponding to $\lambda_1 = 3$ takes the form

$$k\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

We can normalise these eigenvectors, using the Euclidean norm, to obtain the unit eigenvector $u_1 = \dfrac{1}{\sqrt{3}}\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ corresponding to λ_1 and the unit eigenvector $u_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ corresponding to λ_2

Remark 9.1. The eigenspace of an $n \times n$ matrix corresponding to a particular eigenvalue λ is a subspace of $\mathbb{R}^n$ since, if x and y are eigenvectors and μ is a scalar, then

$$A(x+y) = Ax + Ay = \lambda x + \lambda y = \lambda(x+y)$$

that is, the space is closed under addition, and

$$A(\mu x) = \mu Ax = \mu\lambda x = \lambda(\mu x)$$

that is, the space is closed under scalar multiplication.

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

Characteristic equation: $\begin{vmatrix} 1-\lambda & -2 & 0 \\ -3 & 2-\lambda & 0 \\ 0 & 0 & 4-\lambda \end{vmatrix} = 0$

or $(4-\lambda)[(1-\lambda)(2-\lambda) - (-2)(-3)] = 0$

Thus $(4-\lambda)(\lambda^2 - 3\lambda - 4) = 0$ or $(\lambda-4)^2(\lambda+1) = 0$ and there are two eigenvalues $\lambda_1 = -1$ and $\lambda_2 = 4$.

Eigenspace corresponding to $\lambda_1 = -1$:

$$\begin{pmatrix} 2 & -2 & 0 \\ -3 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

Thus $x_3 = 0$ and $2x_1 - 2x_2 = 0$.

Thus a basis for the eigenspace is $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ and an orthonormal basis is $u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

Eigenspace corresponding to $\lambda_2 = 4$:

$$\begin{pmatrix} -3 & -2 & 0 \\ -3 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0$$

Thus both x_2 and x_3 are arbitrary and $3x_1 + 2x_2 = 0$ An element of the eigenspace thus takes the form $\begin{pmatrix} -2k \\ 3k \\ \ell \end{pmatrix} = k \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix} + \ell \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

and a basis is $\begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

These are orthogonal, so an orthonormal basis is

$$u_1 = \frac{1}{\sqrt{13}} \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix} \quad \text{and} \quad u_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Thus since $\lambda_2 = 4$ occurs twice as an eigenvalue of A the eigenspace is two-dimensional.

9.1. Diagonalisation.

DEFINITION 9.2. A square matrix is said to be diagonal if all off-diagonal elements of the matrix are zero.

DEFINITION 9.3. A square matrix A is said to be *diagonalisable* if there exists an invertible matrix P such that $P^{-1}AP$ is diagonal.

THEOREM 9.4. *If A is an $n \times n$ matrix, then the following are equivalent:*

(a): *A is diagonalisable*

(b): *A has n linearly independent eigenvectors*

PROOF. **(b) $\Rightarrow$ (a)**

If A has n linearly independent eigenvectors $p_1, p_2, \ldots, p_n$, then if P is the matrix with columns $p_1, \ldots, p_n$ the columns of AP are

$$Ap_1, Ap_2, \ldots, Ap_n \quad = \quad \lambda_1 p_1, \lambda_2 p_2, \ldots, \lambda_n p_n$$

Therefore $AP = PD$ where $D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$

Since the columns of P are linearly independent, P has rank n and thus is invertible.

Thus $P^{-1}AP = D$

(a) $\Rightarrow$ (b)

If A is diagonalisable then there is an invertible matrix P such that $P^{-1}AP = D$ where $D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & \dots & d_n \end{pmatrix}$

Therefore $AP = PD$ which implies that $Ap_1 = d_1p_1$, $Ap_2 = d_2p_2$, …, $Ap_n = d_np_n$.

Since P is invertible, its columns are linearly independent and therefore non-zero.

Thus $p_1, p_2, \dots, p_n$ are linearly independent eigenvectors of A corresponding to the eigenvelues $d_1, d_2, \dots, d_n$ □

Remark 9.5. This theorem implies that, if A does not have n linearly independent eigenvectors, it is not diagonalisable.

Example. $A = \begin{pmatrix} 1 & -2 & 0 \\ -3 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$ has eigenvectors $p_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ corresponding to $\lambda_1 = -1$ and eigenvectors $p_2 = \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix}$ and $p_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ corresponding to $\lambda_2 = \lambda_3 = 4$

Thus we know that $P^{-1}AP = D$ where $P = \begin{pmatrix} 1 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $D = \mathrm{diag}(-1, 4, 4) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$

As a verification, to calculate P^{-1} use Gauss-Jordan elimination:

$$\left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

Subtract row 1 from row 2:

$$\left(\begin{array}{ccc|ccc} 1 & -2 & 0 & 1 & 0 & 0 \\ 0 & 5 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

Divide row 2 by 5 and then add twice this row to row 1:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 3/5 & 2/5 & 0 \\ 0 & 1 & 0 & -1/5 & 1/5 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

Thus $P^{-1} = \frac{1}{5}\begin{pmatrix} 3 & 2 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix}$

We have then

$$AP = \begin{pmatrix} 1 & -2 & 0 \\ -3 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}\begin{pmatrix} 1 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -8 & 0 \\ -1 & 12 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

and thus

$$P^{-1}AP = \frac{1}{5}\begin{pmatrix} 3 & 2 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix}\begin{pmatrix} -1 & -5 & 0 \\ -1 & 12 & 0 \\ 0 & 0 & 4 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

Note that there is no preferred order for the columns of P. A permutation of the columns just leads to a permutation of the diagonals of D

Application Calculate A^4 where A is as in the example above.

Note that $P^{-1}AP = D \Rightarrow AP = PD \Rightarrow A = PDP^{-1}$

Thus $A^4 = PDP^{-1}PDP^{-1}PDP^{-1}PDP^{-1} = PD^4P^{-1}$

$$\begin{aligned} D^4P^{-1} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 256 & 0 \\ 0 & 0 & 256 \end{pmatrix}\frac{1}{5}\begin{pmatrix} 3 & 2 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix} \\ &= \frac{1}{5}\begin{pmatrix} 3 & 2 & 0 \\ -256 & 256 & 0 \\ 0 & 0 & 1280 \end{pmatrix} \end{aligned}$$

and so

$$\begin{aligned} A^4 = PD^4P^{-1} &= \begin{pmatrix} 1 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}\frac{1}{5}\begin{pmatrix} 3 & 2 & 0 \\ -256 & 256 & 0 \\ 0 & 0 & 1280 \end{pmatrix} \\ &= \frac{1}{5}\begin{pmatrix} 515 & -510 & 0 \\ -765 & 770 & 0 \\ 0 & 0 & 1280 \end{pmatrix} \\ &= \begin{pmatrix} 103 & -102 & 0 \\ -153 & 154 & 0 \\ 0 & 0 & 256 \end{pmatrix} \end{aligned}$$

Of course, for any k, $A^k = PD^kP^{-1}$.

REMARK 9.6. A special case is when A has n distinct eigenvalues. In this case the corresponding eigenvectors are linearly independent and thus A is diagonalisable. (Proof in Anton)

9.2. Orthogonal diagonalisation.

Definition A square matrix A with the property $A^{-1} = A^T$ is said to be an orthogonal matrix.

Definition A square matrix A is *orthogonally diagonalisable* if there is an orthogonal matrix P such that $P^{-1}AP = P^TAP$ is diagonal. The matrix P is said to orthogonally diagonalise A.

Note If A is orthogonally diagonalisable then $P^{-1}AP$ is diagonal where P is orthogonal;
that is $P^TP = I$ or if p_i for $i = 1, 2, \ldots, n$ are the columns of P then $p_i^T p_j = \delta_{ij}$
that is, the columns of P which are eigenvectors of A are orthonormal and so A has n orthonormal eigenvectors.
Similarly if A has n orthonormal eigenvectors $p_1, p_2 \ldots, p_n$ then the matrix P whose columns are these eigenvectors diagonalises A orthogonally.
Thus A is orthogonally diagonalisable if and only if A has n orthonormal eigenvectors.

THEOREM 9.7. *If A is an $n \times n$ matrix, then A is orthogonally diagonalisable if and only if A is symmetric.*

PROOF. We only prove that if A is orthogonally diagonalisable then it must be symmetric.

$$\begin{aligned} D &= P^{-1}AP \\ \Rightarrow A &= PDP^{-1} = PDP^T \\ \text{Therefore } A^T &= (PDP^T)^T = PD^TP^T = PDP^T = A \end{aligned}$$

Thus A is symmetric □

THEOREM 9.8. *If A is symmetric, then eigenvectors corresponding to different eigenvalues are orthogonal.*

PROOF. Let λ and μ be two different eigenvalues of A with corresponding eigenvectors x and y
Thus

$$Ax = \lambda x \tag{9.1}$$

$$Ay = \mu y \tag{9.2}$$

Multiplying (2) on the left by $x^T \Rightarrow x^TAy = \mu x^Ty$
Also (1) $\Rightarrow (Ax)^T = (\lambda x)^T$ or $x^TA = \lambda x^T$
Multiplying on the right by $y \Rightarrow x^TAy = \lambda x^Ty$
Thus $\lambda x^Ty = \mu x^Ty$ or $(\lambda - \mu)x^Ty = 0$
Since $\lambda - \mu \neq 0$ $x^Ty = 0$; that is, x and y are orthogonal. □

Thus, to orthogonally diagonalise a symmetric matrix A, the following procedure is used:

(1) Find a basis for each eigenspace of A.
(2) Using the Gram-Schmidt process, obtain an orthonormal basis for each eigenspace.
(3) Use these basis vectors as the columns of the matrix P.

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

The characteristic equation is $\begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix} = 0$

This gives the quadratic equation

$$\begin{aligned} (2-\lambda)[((2-\lambda)^2 - 1] - [2-\lambda-1] + [1-(2-\lambda)] &= 0 \\ \Rightarrow (2-\lambda)[\lambda^2 - 4\lambda + 3 + 2(\lambda - 1) &= 0 \\ \Rightarrow (2-\lambda)(\lambda-3)(\lambda-1) + 2(\lambda-1) &= 0 \\ \Rightarrow (\lambda-1)[-\lambda^2 + 5\lambda - 6 + 2] &= 0 \\ \Rightarrow -(\lambda-1)^2(\lambda-4) &= 0 \end{aligned}$$

Thus the eigenvalues are $\lambda = 1$ and $\lambda = 4$.
To find the eigenspace corresponding to $\lambda = 1$:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Thus x_2 and x_3 are arbitrary and $x_1 = -x_2 - x_3$

A basis for the eigenspace is thus $v_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ and $v_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$

To obtain an orthonormal basis use Gram-Schmidt:

$u_1 = v_1/\|v_1\|$ $\|v_1\| = \sqrt{2}$ and so $u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$

$$u_2 = \frac{v_2 - < v_2, u_1 > u_1}{\|v_2 - < v_2, u_1 > u_1\|}$$

$$< v_2, u_1 >= \frac{1}{\sqrt{2}}$$

$$v_2 - < v_2, u_1 > u_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1/2 \\ -1/2 \\ 1 \end{pmatrix}$$

Thus $u_2 = \frac{(-1,-1,2)^T}{\|(-1,-1,2)^T\|} = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$

To find the eigenspace corresponding to $\lambda = 4$:

$$\begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Thus

$$\begin{pmatrix} -2 & 1 & 1 \\ 0 & -3/2 & 3/2 \\ 0 & 3/2 & -3/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and so

$$\begin{pmatrix} -2 & 1 & 1 \\ 0 & -3/2 & 3/2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Thus x_3 is arbitrary $x_2 = x_3$ and $x_1 = (x_2 + x_3)/2 = x_3$.

Therefore a basis for the eigenspace is $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and an orthonormal basis is $v_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Using u_1, u_2 and u_3 as column vectors, we have

$$P = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

One can then verify that $P^T P = I$ and $P^T AP = D$ where $D = \text{diag}(1, 1, 4)$