Lectures 16,17,18

6 Inner Product Spaces

6.1 Basic Definition

Parallelogram law, the ability to measure angle between two vectors and in particular, the concept of perpendicularity make the euclidean space quite a special type of a vector space. Essentially all these are consequences of the dot product. Thus, it makes sense to look for operations which share the basic properties of the dot product. In this section we shall briefly discuss this.

Definition 6.1 *Let V be a vector space. By an* **inner product** *on V we mean a binary operation, which associates a scalar say $\langle \mathbf{u}, \mathbf{v}\rangle$ for each pair of vectors $\mathbf{u}, \mathbf{v})$ in V, satisfying the following properties for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and α, β any scalar. (Let "$-$" denote the complex conjugate of a complex number.)*

(1) $\langle \mathbf{u}, \mathbf{v}\rangle = \overline{\langle \mathbf{v}, \mathbf{u}\rangle}$ *(Hermitian property or conjugate symmetry);*
(2) $\langle \mathbf{u}, \alpha\mathbf{v} + \beta\mathbf{w}\rangle = \alpha\langle \mathbf{u}, \mathbf{v}\rangle + \beta\langle \mathbf{u}, \mathbf{w}\rangle$ *(sesquilinearity);*
(3) $\langle \mathbf{v}, \mathbf{v}\rangle > 0$ *if* $\mathbf{v} \neq 0$ *(positivity).*

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

Remark 6.1
(i) Observe that we have not mentioned whether V is a real vector space or a complex vector space. The above definition includes both the cases. The only difference is that if $\mathbb{K} = \mathbb{R}$ then the conjugation is just the identity. Thus for real vector spaces, (1) will becomes 'symmetric property' since for a real number c, we have $\bar{c} = c$.
(ii) Combining (1) and (2) we obtain
(2') $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w}\rangle = \bar{\alpha}\langle \mathbf{u}, \mathbf{w}\rangle + \bar{\beta}\langle \mathbf{v}, \mathbf{w}\rangle$. *In the special case when $\mathbb{K} = \mathbb{R}$, (2) and (2') together are called 'bilinearity'.*

Example 6.1
(i) The dot product in $\mathbb{R}^n$ is an inner product. This is often called the standard inner product.
(ii) Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)^t$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^t \in \mathbb{C}^n$. Define $\langle \mathbf{x}, \mathbf{y}\rangle = \sum_{i=1}^n \overline{x_i} y_i = \mathbf{x}^*\mathbf{y}$. This is an inner product on the complex vector space $\mathbb{C}^n$.
(iii) Let $V = C[0, 1]$. For $f, g \in V$, put

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

It is easy to see that $\langle f, g\rangle$ is an inner product.

6.2 Norm Associated to an Inner Product

Definition 6.2
Let V be an inner product space. For any $\mathbf{v} \in V$, the norm of $\mathbf{v}$, denoted by $\|\mathbf{v}\|$, is the positive square root of $\langle \mathbf{v}, \mathbf{v} \rangle$: $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.

For standard inner product in $\mathbb{R}^n$, $\|\mathbf{v}\|$ is the usual length of the vector $\mathbf{v}$.

Proposition 6.1 *Let V be an inner product space. Let $\mathbf{u}, \mathbf{v} \in V$ and c be a scalar. Then*
(i) $\|c\mathbf{u}\| = |c|\|\mathbf{u}\|$;
(ii) $\|\mathbf{u}\| > 0$ *for* $\mathbf{u} \neq 0$;
(iii) $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$;
(iv) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$, *equality holds in this triangle inequality iff $\mathbf{u}, \mathbf{v}$ are linearly dependent over $\mathbb{R}$.*

Proof: (i) and (ii) follow from definitions 6.1,6.2.
(iii) This is called the **Cauchy-Schwarz inequality**. If either $\mathbf{u}$ or $\mathbf{v}$ is zero, it is clear. So let $\mathbf{u}$ be nonzero. Consider the vector

$$\mathbf{w} = \mathbf{v} - a\mathbf{u}.$$

where a is a scalar to be specified later. Then

$$\begin{aligned} 0 \leq \langle \mathbf{w}, \mathbf{w} \rangle &= \langle \mathbf{v} - a\mathbf{u}, \mathbf{v} - a\mathbf{u} \rangle \\ &= \langle \mathbf{v}, \mathbf{v} \rangle - \langle \mathbf{v}, a\mathbf{u} \rangle - \langle a\mathbf{u}, \mathbf{v} \rangle + \langle a\mathbf{u}, a\mathbf{u} \rangle \\ &= \langle \mathbf{v}, \mathbf{v} \rangle - a\langle \mathbf{v}, \mathbf{u} \rangle - \overline{a}\langle \mathbf{u}, \mathbf{v} \rangle + a\overline{a}\langle \mathbf{u}, \mathbf{u} \rangle \end{aligned}$$

Substitute $a = \langle \mathbf{u}, \mathbf{v} \rangle / \langle \mathbf{u}, \mathbf{u} \rangle$ and multiply by $\langle \mathbf{u}, \mathbf{u} \rangle$, we get $|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle$. This proves (iii).
(iv) Using (iii), we get $\Re(\langle \mathbf{u}, \mathbf{v} \rangle) \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$. Therefore,

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \|\mathbf{u}\|^2 + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 + 2\Re\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2. \end{aligned}$$

Thus $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

Now equality holds in (iv) iff $\Re\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$ and equality holds in (iii). This is the same as saying that $\langle \mathbf{u}, \mathbf{v} \rangle$ is real and $\mathbf{w} = 0$. This is the same as saying that $\mathbf{v} = a\mathbf{u}$ and $\langle \mathbf{v}, \mathbf{u} \rangle = a\langle \mathbf{u}, \mathbf{u} \rangle$ is real. Which is the same as saying that $\mathbf{v}$ is a real multiple of $\mathbf{u}$. ♠

6.3 Distance and Angle

Definition 6.3 *In any inner product space V, the distance between $\mathbf{u}$ and $\mathbf{v}$ is defined as*

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Definition 6.4 *Let V be a real inner product space. The angle between vectors $\mathbf{u}, \mathbf{v}$ is defined to be the angle $\theta, 0 \leq \theta \leq \pi$ so that*

$$\cos\theta = \frac{\langle u,\, v\rangle}{\|u\|\|v\|}\,.$$

Proposition 6.2
(i) $d(\mathbf{u}, \mathbf{v}) \geq 0$ *and* $d(\mathbf{u}, \mathbf{v}) = 0$ *iff* $\mathbf{u} = \mathbf{v}$;
(ii) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$;
(iii) $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$.

6.4 Gram-Schmidt Orthonormalization

The standard basis $E = \{e_1, e_2, \ldots, e_n\}$ of $\mathbb{R}^n$ has two properties :
(i) each e_i has length one
(ii) any two elements of E are mutually perpendicular.

When vectors are expressed in terms of the standard basis, calculations become easy due to the above properties. Now, let V be any inner product space of dimension d. We wish to construct an analogue of E for V.

Definition 6.5 *Two vectors $\mathbf{u}, \mathbf{v}$ of an inner product space V are called* orthogonal to each other *or* perpendicular to each other *if $\langle \mathbf{u},\, \mathbf{v}\rangle = 0$. We express this symbolically by $\mathbf{u} \perp \mathbf{v}$. A set S consisting of non zero elements of V is called* orthogonal *if $\langle \mathbf{u},\, \mathbf{v}\rangle = 0$ for every pair of elements of S. Further if every element of V is of unit norm then S is called an* orthonormal *set. Further, if S is a basis for V then it is called an* orthonormal basis.

One of the biggest advantage of an orthogonal set is that Pythagorus theorem and its general form are valid, which makes many computations easy.

Theorem 6.1 Pythagorus Theorem: *Given $\mathbf{w}, \mathbf{u} \in V$ such that $\mathbf{w} \perp \mathbf{u}$ we have*

$$\|\mathbf{w} + \mathbf{u}\|^2 = \|\mathbf{w}\|^2 + \|\mathbf{u}\|^2. \tag{47}$$

Proof: We have,

$$\|\mathbf{w} + \mathbf{u}\|^2 = \langle \mathbf{w} + \mathbf{u},\, \mathbf{w} + \mathbf{u}\rangle = \langle \mathbf{w},\, \mathbf{w}\rangle + \langle \mathbf{u},\, \mathbf{w}\rangle + \langle \mathbf{w},\, \mathbf{u}\rangle + \langle \mathbf{u},\, \mathbf{u}\rangle = \|\mathbf{w}\|^2 + \|\mathbf{u}\|^2$$

since $\langle \mathbf{u},\, \mathbf{w}\rangle = \langle \mathbf{w},\, \mathbf{u}\rangle = 0$. ♠

Proposition 6.3 *Let $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an orthogonal set. Then S is linearly independent.*

Proof: Suppose $c_1, c_2, \ldots, c_n$ are scalars with

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_n\mathbf{u}_n = 0.$$

Take inner product with $\mathbf{u}_i$ on both sides to get

$$c_i\langle \mathbf{u}_i,\, \mathbf{u}_i\rangle = 0.$$

Since $\mathbf{u}_i \neq 0$, we get $c_i = 0$. Thus U is linearly independent. ♠

Definition 6.6 *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in an inner product space* V *and* $\mathbf{v} \neq 0$ *then the vector*

$$\langle \mathbf{v},\, \mathbf{u} \rangle \frac{\mathbf{v}}{\|\mathbf{v}\|^2}$$

is called the **orthogonal projection** *of* $\mathbf{u}$ *along* $\mathbf{v}$.

Remark 6.2 *The order in which you take* $\mathbf{u}$ *and* $\mathbf{v}$ *matters here when you are working over complex numbers because of the inner product is linear in the second slot and conjugate linear in the first slot. Over the real numbers, however, this is not a problem.*

Theorem 6.2 (Gram-Schmidt Orthonormalization Process)*: Let* V *be any inner product space. Given a linearly independent subset* $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ *there exists an orthonormal set of vectors* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *such that* $L(\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}) = L(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$ *for all* $1 \leq k \leq n$. *In particular, if* V *is finite dimensional, then it has an orthonormal basis.*

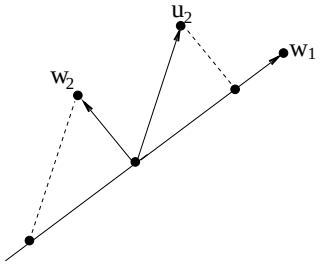
Proof: The construction of the orthonormal set is algorithmic and of course, inductive. First we construct an intermediate orthogonal set $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ with the same property for each k. Then we simply 'normalize' each element, viz, take

$$\mathbf{v}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}.$$

That will give us an orthonormal set as required. The last part of the theorem follows if we begin with $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ as any basis for V.

Take $\mathbf{w}_1 := \mathbf{u}_1$. So, the construction is over for $k = 1$. To construct $\mathbf{w}_2$, subtract from $\mathbf{u}_2$, its orthogonal projection along $\mathbf{w}_1$ Thus

$$\mathbf{w}_2 = \mathbf{u}_2 - \langle \mathbf{w}_1,\, \mathbf{u}_2 \rangle \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|^2}.$$



Check that $\langle \mathbf{w}_2,\, \mathbf{w}_1 \rangle = 0$. Thus $\{\mathbf{w}_1, \mathbf{w}_2\}$ is an orthogonal set. Check also that $L(\{\mathbf{u}_1, \mathbf{u}_2\}) = L(\{\mathbf{w}_1, \mathbf{w}_2\})$. Now suppose we have constructed $\mathbf{w}_i$ for $i \leq k < n$ as required. Put

$$\mathbf{w}_{k+1} = \mathbf{u}_{k+1} - \sum_{j=1}^{k} \langle \mathbf{w}_j,\, \mathbf{u}_{k+1} \rangle \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|^2}.$$

Now check that $\mathbf{w}_{k+1}$ is orthogonal to $\mathbf{w}_j, j \leq k$ and $L(\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{k+1}\}) = L(\{\mathbf{u}_1, \ldots, \mathbf{u}_k\})$. By induction, this completes the proof. ♠

Example 6.2 *Let $V = P_3[-1,-1]$ denote the real vector space of polynomials of degree at most 3 defined on $[-1,1]$ together with the zero polynomial. V is an inner product space under the inner product*

$$\langle f,\, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt.$$

To find an orthonormal basis, we begin with the basis $\{1, x, x^2, x^3\}$. Set $\mathbf{v}_1 = \mathbf{w}_1 = 1$. Then

$$\begin{aligned}
\mathbf{w}_2 &= x - \langle x,\, 1\rangle \frac{1}{\|1\|^2} \\
&= x - \frac{1}{2}\int_{-1}^{1} t dt = x, \\
\mathbf{w}_3 &= x^2 - \langle x^2,\, 1\rangle \frac{1}{2} - \langle x^2,\, x\rangle \frac{x}{(2/3)} \\
&= x^2 - \frac{1}{2}\int_{-1}^{1} t^2 dt - \frac{3}{2}x\int_{-1}^{1} t^3 dt \\
&= x^2 - \frac{1}{3}, \\
\mathbf{w}_4 &= x^3 - \langle x^3,\, 1\rangle \frac{1}{2} - \langle x^3,\, x\rangle \frac{x}{(2/3)} - \langle x^3,\, x^2 - \frac{1}{3}\rangle \frac{x^2 - \frac{1}{3}}{(2/5)} \\
&= x^3 - \frac{3}{5}x.
\end{aligned}$$

Thus $\{1, x, x^2 - \frac{1}{3}, x^3 - \frac{3}{5}x\}$ is an orthogonal basis. We divide these by respective norms to get an orthonormal basis.

$$\left\{ \frac{1}{\sqrt{2}},\ \ x\sqrt{\frac{3}{2}},\ \ (x^2 - \frac{1}{3})\frac{3\sqrt{5}}{2\sqrt{2}},\ \ (x^3 - \frac{3}{5}x)\frac{5\sqrt{7}}{2\sqrt{2}} \right\}.$$

You will meet these polynomials while studying differential equations.

6.5 Orthogonal Complement

Definition 6.7 *Let V be an inner product space. Given $\mathbf{v}, \mathbf{w} \in V$ we say $\mathbf{v}$ is perpendicular to $\mathbf{W}$ and write $\mathbf{v} \perp \mathbf{w}$ if $\langle \mathbf{v},\, \mathbf{w}\rangle = 0$. Given a subspace W of V, we define*

$$W^{\perp} := \{\mathbf{v} \in V \ :\ \mathbf{v} \perp \mathbf{w},\ \forall\ \mathbf{w} \in W\}$$

and call it the orthogonal complement *of W.*

Proposition 6.4 *Let W be a subspace of an inner product space. Then*
(i) $W^{\perp}$ is a subspace of V;
(ii) $W \cap W^{\perp} = (0)$;
(iii) $W \subset (W^{\perp})^{\perp}$;

Proof: All the proofs are easy.

Proposition 6.5 *Let W be a finite dimensional subspace of a vector space. Then every element $\mathbf{v} \in V$ is expressible as a unique sum*

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp}$$

where $\mathbf{w} \in W$ and $\mathbf{w}^{\perp} \in W^{\perp}$. Further $\mathbf{w}$ is characterised by the property that it is the near most point in W to $\mathbf{v}$.

Proof: Fix an orthonormal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ for W. Define $\mathbf{w} = \sum_{i=1}^{k} \langle \mathbf{w}_i, \mathbf{v} \rangle \mathbf{w}_i$. Take $\mathbf{w}^\perp = \mathbf{v} - \mathbf{w}$ and verify that $\mathbf{w}^\perp \in W^\perp$. This proves the existence of such an expression. Now suppose, we have $\mathbf{v} = \mathbf{u} + \mathbf{u}^\perp$ is another similar expression. Then it follows that $\mathbf{w} - \mathbf{u} = \mathbf{w}^\perp - \mathbf{u}^\perp \in W \cap W^\perp = (0)$. Therefore $\mathbf{w} = \mathbf{u}$ and $\mathbf{w}^\perp = \mathbf{u}^\perp$. This proves the uniqueness. Finally, let $\mathbf{u} \in W$ be any vector. Then $\mathbf{v} - \mathbf{u} = (\mathbf{v} - \mathbf{w}) + (\mathbf{w} - \mathbf{u})$ and since $\mathbf{w} - \mathbf{u} \in W$ and $\mathbf{v} - \mathbf{w} \in W^\perp$, we can apply Pythagorus theorem to conclude that $\|\mathbf{v} - \mathbf{u}\|^2 \geq \|\mathbf{v} - \mathbf{w}\|^2$ and equality holds iff $\mathbf{u} = \mathbf{w}$. This proves that $\mathbf{w}$ is the (only) nearest point to $\mathbf{v}$ in W. ♠

Remark 6.3
(i) The element $\mathbf{w}$ *so obtained is called the* orthogonal projection *of* $\mathbf{v}$ *on* W. *One can easily check that the assignment* $\mathbf{v} \mapsto \mathbf{w}$ *itself defines a linear map* $P_W : V \longrightarrow W$. *(Use the formula* $P_W(\mathbf{v}) = \sum_i \langle \mathbf{w}_i, \mathbf{v} \rangle \mathbf{w}_i$.*) Also observe that* P_W *is identity on* W *and hence* $P_W^2 = P_W$.
(ii) Observe that if V *itself is finite dimensional, then it follows that the above proposition is applicable to all subspaces in* V. *In this case it easily follows that* $\dim W + \dim W^\perp = \dim V$ *(see exercise at the end of section Linear Dependence.)*

6.6 Least square problem

As an application of orthogonal projection, let us consider the following problem. Suppose we are given a system of linear equations and a likely candidate for the solution. If this candidate is actually a solution well and good. Otherwise we would like to get a solution which is nearmost to the candidate. This problem is a typical example of finding 'best approximations'. (It can also be discussed using Lagrange multipliers.)

Thus if the system is a homgeneous set of m linear equations involving n variables, we get a linear transformation $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ associated to it and the space of solutions is nothing but the subspace $W = \mathcal{N}(f)$, the null space. Given a point $\mathbf{v} \in \mathbb{R}^n$ the nearmost point in W to $\mathbf{v}$ is nothing but the orthogonal projection of $\mathbf{v}$ on W. In case the system is inhomogeneous, we still consider the linear map f defined by the homogeneous part and put $W = \mathcal{N}(f)$. Then we must first of all have one correct solution $\mathbf{p}$ of the inhomogeneous system. We then take $\mathbf{v} - \mathbf{p}$, take its projection on W as above and then add back $\mathbf{p}$ to it to get the nearmost solution to $\mathbf{v}$.

Problem (Method of Least Squares) Given a linear system of m equations in n variables $A\mathbf{x} = b$, and an n-vector $\mathbf{v}$, find the solution $\mathbf{y}$ such that $\|\mathbf{y} - \mathbf{v}\|$ is minimal.

Answer: Step 1. Apply GJEM to find a basis for the null-space W of A and a particular solution $\mathbf{p}$ of the system.
Step 2. Apply Gram-Schmidt process to the basis of W to get an orthonomal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ for W.
Step 3. Let $P_W(\mathbf{v} - \mathbf{p}) = \sum_{j=1}^{k} \langle (\mathbf{v} - \mathbf{p}), \mathbf{v}_j \rangle \mathbf{v}_j$ denote the orthogonal projection of $\mathbf{v} - \mathbf{p}$ on the space W. Take $\mathbf{y} = P_w(\mathbf{v} - \mathbf{p}) + \mathbf{p}$. (Verify that this is the required solution.)

Example 6.3 *Consider a system of three linear equations in 5 variables* $x_1, \ldots, x_5$ *given by*

$$\begin{array}{rrrrcl} & 2x_3 & -2x_4 & +x_5 & = & 2 \\ 2x_2 & -8x_3 & +14x_4 & -5x_5 & = & 2 \\ x_2 & +3x_3 & & +x_5 & = & 8 \end{array}$$

(a) Write down a basis for the space of solution of the associated homogeneous system.
(b) Given the vector $\mathbf{v} = (1, 6, 2, 1, 1)^t$, *find the solution to the system which is nearmost to* $\mathbf{v}$.

Solution: *The matrix form of the system is*

$$\left[\begin{array}{ccccc|c} 0 & 0 & 2 & -2 & 1 & 2 \\ 0 & 2 & -8 & 14 & -5 & 2 \\ 0 & 1 & 3 & 0 & 1 & 8 \end{array}\right]$$

Applying GJEM, we see that the Gauss-Jordan form of this system is

$$\left[\begin{array}{ccccc|c} 0 & 1 & 0 & 3 & -1/2 & 5 \\ 0 & 0 & 1 & -1 & 1/2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Therefore the general solution is of the form

$$\begin{bmatrix} x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 & 0 & -3 & 1/2 \\ 1 & 0 & 1 & -1/2 \end{bmatrix} \begin{bmatrix} 1 \\ x_1 \\ x_4 \\ x_5 \end{bmatrix}.$$

A particular solution of the system is obtained by taking $x_1 = x_4 = x_5 = 0$ *and hence* $x_2 = 5, x_3 = 1$, *i.e.,* $\mathbf{p} = (0, 5, 1, 0, 0)$. *Also the space* W *of solutions of the homogeneous system is given by*

$$\begin{bmatrix} x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 & 0 & -3 & 1/2 \\ 1 & 0 & 1 & -1/2 \end{bmatrix} \begin{bmatrix} 0 \\ x_1 \\ x_4 \\ x_5 \end{bmatrix}.$$

Therefore a basis for this space W *(which is the null-space of the un-augemented matrix) is*

$$\{(1, 0, 0, 0, 0)^t, (0, -3, 1, 1, 0)^t, (0, 1/2, -1/2, 0, 1)^t\}$$

So we consider the orthogonal projection of $\mathbf{v} - \mathbf{p} = (1, 1, 1, 1, 1)$ *on the space* W *which is equal to*

$\mathbf{w} = p_W(\mathbf{v} - \mathbf{p}) = (1, 0, 0, 0, 0)^t - (0, -3, 1, 1, 0)^t(\frac{1}{11}) + (0, 1/2, -1/2, 0, 1)^t(\frac{2}{3})$
$= (1, 2/33, -8/33, 1/11, 2/3)^t$.

Now the solution to the original system which is nearmost to the given vector $\mathbf{v}$ *is obtained by adding* $\mathbf{p}$ *to* $\mathbf{w}$, *viz.*

Answer: $(1, 167/33, 25/33, 1/11, 2/3)^t$.

7 Eigenvalues and Eigenvectors

7.1 Introduction

The simplest of matrices are the diagonal ones. Thus a linear map will be also easy to handle if its associated matrix is a diagonal matrix. Then again we have seen that the matrix associated depends upon the choice of the bases to some extent. This naturally leads us to the problem of investigating the existence and construction of a suitable basis with respect to which the matrix associated to a given linear transformation is diagonal.

Definition 7.1 *A $n \times n$ matrix A is called* diagonalizable *if there exists an invertible $n \times n$ matrix M such that $M^{-1}AM$ is a diagonal matrix. A linear map $f : V \longrightarrow V$ is called diagonalizable if the matrix associated to f with respect to some basis is diagonal.*

Remark 7.1
(i) Clearly, f is diagonalizable iff the matrix associated to f with respect to some basis (any basis) is diagonalizable.
(ii) Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis. The matrix M_f of a linear transformation f w.r.t. this basis is diagonal iff $f(\mathbf{v}_i) = \lambda_i \mathbf{v}_i,\ 1 \leq i \leq n$ for some scalars λ_i. Naturally a subquestion here is: does there exist such a basis for a given linear transformation?

Definition 7.2 *Given a linear map $f : V \longrightarrow V$ we say $\mathbf{v} \in V$ is an* eigenvector *for f if $\mathbf{v} \neq 0$ and $f(\mathbf{v}) = \lambda \mathbf{v}$ for some $\lambda \in \mathbb{K}$. In that case λ is called as* eigenvalue *of f. For a square matrix A we say λ is an eigenvalue if there exists a non zero column vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda \mathbf{v}$. Of course $\mathbf{v}$ is then called the eigenvector of A corresponding to λ.*

Remark 7.2
(i) It is easy to see that eigenvalues and eigenvectors of a linear transformation are same as those of the associated matrix.
(ii) Even if a linear map is not diagonalizable, the existence of eigenvectors and eigenvalues itself throws some light on the nature of the linear map. Thus the study of eigenvalues becomes extremely important. They arise naturally in the study of differential equations. Here we shall use them to address the problem of diagonalization and then see some geometric applications of diagonalization itself.

7.2 Characteristic Polynomial

Proposition 7.1
(1) Eigenvalues of a square matrix A are solutions of the equation

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

(2)The null space of $A - \lambda I$ is equal to the eigenspace

$$E_A(\lambda) := \{\mathbf{v} \ : \ A\mathbf{v} = \lambda \mathbf{v}\} = \mathcal{N}(A - \lambda I).$$

Proof: (1) If $\mathbf{v}$ is an eigenvector of A then $\mathbf{v} \neq 0$ and $A\mathbf{v} = \lambda \mathbf{v}$ for some scalar λ. Hence $(A - \lambda I)\mathbf{v} = 0$. Thus the nullity of $A - \lambda I$ is positive. Hence $\text{rank}(A - \lambda I)$ is less than n. Hence $\det(A - \lambda I) = 0$.
(2) $E_A(\lambda) = \{\mathbf{v} \in V \ : \ A\mathbf{v} = \lambda \mathbf{v}\} = \{\mathbf{v} \in V \ : \ (A - \lambda I)\mathbf{v} = 0\} = \mathcal{N}(A - \lambda I)$. ♠

Definition 7.3 *For any square matrix A, the polynomial $\chi_A(\lambda) = \det(A - \lambda I)$ in λ is called the* **characteristic polynomial** *of A.*

Example 7.1

(1) $A = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$. *To find the eigenvalues of A, we solve the equation*

$$\det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & 2 \\ 0 & 3-\lambda \end{bmatrix} = (1-\lambda)(3-\lambda) = 0.$$

Hence the eigenvalues of A are 1 and 3. Let us calculate the eigenspaces $E(1)$ and $E(3)$. By definition

$$E(1) = \{\mathbf{v} \mid (A - I)\mathbf{v} = 0\} \text{ and } E(3) = \{\mathbf{v} \mid (A - 3I)\mathbf{v} = 0\}.$$

$A - I = \begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix}$. *Hence* $(x, y)^t \in E(1)$ *iff* $\begin{bmatrix} 0 & 2 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2y \\ 2y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. *Hence* $E(1) = L\{(1, 0)\}$.

$A - 3I = \begin{bmatrix} 1-3 & 2 \\ 0 & 3-3 \end{bmatrix} = \begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix}$. *Suppose* $\begin{bmatrix} -2 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Then $\begin{bmatrix} -2x + 2y \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. *This is possible iff $x = y$. Thus* $E(3) = L(\{(1, 1)\})$.

(2) Let $A = \begin{bmatrix} 3 & 0 & 0 \\ -2 & 4 & 2 \\ -2 & 1 & 5 \end{bmatrix}$. *Then* $\det(A - \lambda I) = (3 - \lambda)^2(6 - \lambda)$.

Hence eigenvalues of A are 3 and 6. The eigenvalue $\lambda = 3$ is a double root of the characteristic polynomial of A. We say that $\lambda = 3$ has **algebraic multiplicity** *2. Let us find the eigenspaces $E(3)$ and $E(6)$.*

$\underline{\lambda = 3}$: $A - 3I = \begin{bmatrix} 0 & 0 & 0 \\ -2 & 1 & 2 \\ -2 & 1 & 2 \end{bmatrix}$. *Hence* $rank(A - 3I) = 1$. *Thus* $nullity(A - 3I) = 2$. *By solving the system $(A - 3I)\mathbf{v} = 0$, we find that*

$$\mathcal{N}(A - 3I) = E_A(3) = L(\{(1, 0, 1), (1, 2, 0)\}).$$

The dimension of $E_A(\lambda)$ is called the **geometric multiplicity** *of λ. Hence geometric multiplicity of $\lambda = 3$ is 2.*

$\underline{\lambda = 6}$: $A - 6I = \begin{bmatrix} -3 & 0 & 0 \\ -2 & -2 & 2 \\ -2 & 1 & -1 \end{bmatrix}$. *Hence* $rank(A - 6I) = 2$. *Thus* $\dim E_A(6) = 1$. *(It can be shown that $\{(0, 1, 1)\}$ is a basis of $E_A(6)$.) Thus both the algebraic and geometric multiplicities of the eigenvalue 6 are equal to 1.*

(3) $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. *Then* $\det(A - \lambda I) = (1 - \lambda)^2$. *Thus $\lambda = 1$ has algebraic multiplicity 2.*

$A - I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. *Hence nullity* $(A - I) = 1$ *and* $E_A(1) = L\{e_1\}$. *In this case the geometric multiplicity is less than the algebraic multiplicity of the eigenvalue* 1.

Remark 7.3
(i) Observe that $\chi_A(\lambda) = \chi_{M^{-1}AM}(\lambda)$. *Thus the characteristic polynomial is an invariant of similarity. Thus the characteristic polynomial of any linear map* $f : V \longrightarrow V$ *is also defined (where* V *is finite dimensional) by choosing some basis for* V, *and then taking the characteristic polynomial of the associated matrix* $\mathcal{M}(f)$ *of* f. *This definition does not depend upon the choice of the basis.*
(ii) If we expand $\det(A - \lambda I)$ *we see that there is a term*

$$(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda).$$

This is the only term which contributes to λ^n *and* λ^{n-1}. *It follows that the degree of the characteristic polynomial is exactly equal to* n, *the size of the matrix; moreover, the coefficient of the top degree term is equal to* $(-1)^n$. *Thus in general, it has* n *complex roots, some of which may be repeated, some of them real, and so on. All these patterns are going to influence the geometry of the linear map.*
(iii) If A *is a real matrix then of course* $\chi_A(\lambda)$ *is a real polynomial. That however, does not allow us to conclude that it has real roots. So while discussing eigenvalues we should consider even a real matrix as a complex matrix and keep in mind the associated linear map* $\mathbb{C}^n \longrightarrow \mathbb{C}^n$. *The problem of existence of real eigenvalues and real eigenvectors will be discussed soon.*
(iv) Next, the above observation also shows that the coefficient of λ^{n-1} *is equal to*

$$(-1)^{n-1}(a_{11} + \cdots + a_{nn}) = (-1)^{n-1} tr\, A.$$

Lemma 7.1 Suppose A is a real matrix with a real eigenvalue λ. Then there exists a real column vector $\mathbf{v} \neq 0$ such that $A\mathbf{v} = \lambda\mathbf{v}$.

Proof: Start with $A\mathbf{w} = \lambda\mathbf{w}$ where $\mathbf{w}$ is a non zero column vector with complex entries. Write $\mathbf{w} = \mathbf{v} + \imath\mathbf{v}'$ where both $\mathbf{v}, \mathbf{v}'$ are real vectors. We then have

$$A\mathbf{v} + \imath A\mathbf{v}' = \lambda(\mathbf{v} + \imath\mathbf{v}')$$

Compare the real and imaginary parts. Since $\mathbf{w} \neq 0$, at least one of the two $\mathbf{v}, \mathbf{v}'$ must be a non zero vector and we are done. ♠

Proposition 7.2 *Let* A *be an* $n \times n$ *matrix with eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_n$. *Then*
(i) $tr\,(A) = \lambda_1 + \lambda_2 + \ldots + \lambda_n$.
(ii) $\det A = \lambda_1\lambda_2 \ldots \lambda_n$.

Proof: The characteristic polynomial of A is

$$\det(A - \lambda I) = \det \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{bmatrix}$$

$$(-1)^n\lambda^n + (-1)^{n-1}\lambda^{n-1}(a_{11} + \ldots + a_{nn}) + \ldots \tag{48}$$

Put $\lambda = 0$ to get $\det A =$ constant term of $\det(A - \lambda I)$.
Since $\lambda_1, \lambda_2, \ldots, \lambda_n$ are roots of $\det(A - \lambda I) = 0$ we have

$$\det(A - \lambda I) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n). \tag{49}$$

$$\tag{50}$$

$$(-1)^n[\lambda^n - (\lambda_1 + \lambda_2 + \ldots + \lambda_n)\lambda^{n-1} + \ldots + (-1)^n\lambda_1\lambda_2\ldots\lambda_n]. \tag{51}$$

Comparing (49) and 51), we get, the constant term of $\det(A - \lambda I)$ is equal to $\lambda_1\lambda_2\ldots\lambda_n = \det A$ and $tr(A) = a_{11} + a_{22} + \ldots + a_{nn} = \lambda_1 + \lambda_2 + \ldots + \lambda_n$. ♠

Proposition 7.3 *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be eigenvectors of a matrix A associated to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent.*

Proof: Apply induction on k. It is clear for $k = 1$. Suppose $k \geq 2$ and $c_1\mathbf{v}_1 + \ldots + c_k\mathbf{v}_k = 0$ for some scalars $c_1, c_2, \ldots, c_k$. Hence $c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 + \ldots + c_kA\mathbf{v}_k = 0$
Hence

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \ldots + c_k\lambda_k\mathbf{v}_k = 0$$

Hence

$$\begin{aligned}&\lambda_1(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_k\mathbf{v}_k) - (\lambda_1c_1\mathbf{v}_1 + \lambda_2c_2\mathbf{v}_2 + \ldots + \lambda_kc_k\mathbf{v}_k)\\ &= (\lambda_1 - \lambda_2)c_2\mathbf{v}_2 + (\lambda_1 - \lambda_3)c_3\mathbf{v}_3 + \ldots + (\lambda_1 - \lambda_k)c_k\mathbf{v}_k = 0\end{aligned}$$

By induction, $\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_k$ are linearly independent. Hence $(\lambda_1 - \lambda_j)c_j = 0$ for $j = 2, 3, \ldots, k$. Since $\lambda_1 \neq \lambda_j$ for $j = 2, 3, \ldots, k$, $c_j = 0$ for $j = 2, 3, \ldots, k$. Hence c_1 is also zero. Thus $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent. ♠

Proposition 7.4 *Suppose A is an $n \times n$ matrix. Let A have n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let C be the matrix whose column vectors are respectively $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ where $\mathbf{v}_i$ is an eigenvector for λ_i for $i = 1, 2, \ldots, n$. Then*

$$C^{-1}AC = D(\lambda_1, \ldots, \lambda_n) = D$$

the diagonal matrix.

Proof: It is enough to prove $AC = CD$. For $i = 1, 2, \ldots, n$: let $C^i(= \mathbf{v}_i)$ denote the i^{th} column of C etc.. Then

$$(AC)^i = AC^i = A\mathbf{v}_i = \lambda_i\mathbf{v}_i.$$

Similarly,

$$(CD)^i = CD^i = \lambda_i\mathbf{v}_i.$$

Hence $AC = CD$ as required.] ♠

7.3 Relation Between Algebraic and Geometric Multiplicities

Recall that

Definition 7.4 *The* algebraic multiplicity $a_A(\mu)$ *of an eigenvalue* μ *of a matrix* A *is defined to be the multiplicity* k *of the root* μ *of the polynomial* $\chi_A(\lambda)$. *This means that* $(\lambda-\mu)^k$ *divides* $\chi_A(\lambda)$ *whereas* $(\lambda-\mu)^{k+1}$ *does not.*

Definition 7.5 *The* geometric multiplicity *of an eigenvalue* μ *of* A *is defined to be the dimension of the eigenspace* $E_A(\lambda)$;

$$g_A(\lambda) := dim\, E_A(\lambda).$$

Proposition 7.5 *Both algebraic multiplicity and the geometric multiplicities are invariant of similarity.*

Proof: We have already seen that for any invertible matrix C, $\chi_A(\lambda) = \chi_{C^{-1}AC}(\lambda)$. Thus the invariance of algebraic multiplicity is clear. On the other hand check that $E_{C^{-1}AC}(\lambda) = C(E_A(\lambda))$. Therefore, $dim\,(E_{C^{-1}AC}(\lambda)) = dim\, C(E_A\lambda)) = dim\,(E_A(\lambda))$, the last equality being the consequence of invertibility of C.

♠

We have observed in a few examples that the geometric multiplicity of an eigenvalue is at most its algebraic multiplicity. This is true in general.

Proposition 7.6 *Let* A *be an* $n\times n$ *matrix. Then the geometric multiplicity of an eigenvalue* μ *of* A *is less than or equal to the algebraic multiplicity of* μ.

Proof: Put $a_A(\mu) = k$. Then $(\lambda-\mu)^k$ divides $\det(A-\lambda I)$ but $(\lambda-\mu)^{k+1}$ does not. Let $g_A(\mu) = g$, be the geometric multiplicity of μ. Then $E_A(\mu)$ has a basis consisting of g eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_g$. We can extend this basis of $E_A(\mu)$ to a basis of $\mathbb{C}^n$, say $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_g, \ldots, \mathbf{v}_n\}$. Let B be the matrix such that $B^j = \mathbf{v}_j$. Then B is an invertible matrix and

$$B^{-1}AB = \left[\begin{array}{c|c} \mu I_g & X \\ \hline 0 & Y \end{array}\right]$$

where X is a $g\times(n-g)$ matrix and Y is an $(n-g)\times(n-g)$ matrix. Therefore,

$$\begin{aligned} \det(A-\lambda I) = \det[B^{-1}(A-\lambda I)B] &= \det(B^{-1}AB-\lambda I) \\ &= (\det(\mu-\lambda)I_g)(\det(C-\lambda I_{n-g}) \\ &= (\mu-\lambda)^g \det(Y-\lambda I_{n-g}). \end{aligned}$$

Thus $g \le k$. ♠

Remark 7.4 *We will now be able to say something about the diagonalizability of a given matrix* A. *Assuming that there exists* B *such that* $B^{-1}AB = D(\lambda_1, \ldots, \lambda_n)$, *as seen in the previous proposition, it follows that* $AB = BD\ldots$ *etc..* $AB^i = \lambda B^i$ *where* B^i *denotes the* i^{th} *column vector of* B. *Thus we need not hunt for* B *anywhere but look for eigenvectors of* A. *Of course* B^i *are linearly independent, since* B *is invertible. Now the problem turns to*

the question whether we have n linearly independent eigenvectors of A so that they can be chosen for the columns of B. The previous proposition took care of one such case, viz., when the eigenvalues are distinct. In general, this condition is not forced on us. Observe that the geometric multiplicity and algebraic multiplicity of an eigenvalue co-incide for a diagonal matrix. Since these concepts are similarity invariants, it is necessary that the same is true for any matrix which is diagonalizable. This turns out to be sufficient also. The following theorem gives the correct condition for diagonalization.

Theorem 7.1 *A $n \times n$ matrix A is diagonalizable if and only if for each eigenvalue μ of A we have the algebraic and geometric multiplicities are equal: $a_A(\mu) = g_A(\mu)$.*

Proof: We have already seen the necessity of the condition. To prove the converse, suppose that the two multiplicities coincide for each eigenvalue. Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_k$ are all the eigenvalues of A with algebraic multiplicities $n_1, n_2, \ldots, n_k$. Let

$$\begin{aligned} B_1 = \{\mathbf{v}_{11}, \mathbf{v}_{12}, \ldots, \mathbf{v}_{1n_1}\} &= \text{a basis of } E(\lambda_1), \\ B_2 = \{\mathbf{v}_{21}, \mathbf{v}_{22}, \ldots, \mathbf{v}_{2n_2}\} &= \text{a basis of } E(\lambda_2), \\ &\vdots \\ B_k = \{\mathbf{v}_{k1}, \mathbf{v}_{k2}, \ldots, \mathbf{v}_{kn_k}\} &= \text{a basis of } E(\lambda_k). \end{aligned}$$

Use induction on k to show that $B = B_1 \cup B_2 \cup \ldots \cup B_k$ is a linearly independent set. (The proof is exactly similar to the proof of proposition (7.3). Denote the matrix with columns as elements of the basis B also by B itself. Then, check that $B^{-1}AB$ is a diagonal matrix. Hence A is diagonalizable. ♠