Notes on Inner Product Spaces

Let V be a vector space with an inner product $\langle\ ,\ \rangle$ (see the textbook for the definition). These notes give some of the basic facts and properties in an order a bit different from the one in the textbook.

Distance Function. In an inner product space V, we define the distance between two vectors $\mathbf{v}$ and $\mathbf{w}$ to be

$$d(\mathbf{v}, \mathbf{w}) \;=\; \sqrt{\|\mathbf{v} - \mathbf{w}\|^2}.$$

This makes V into a *metric space*, i.e. V is a set with a distance function $V \times V \to \mathbb{R}$ that satisfies, for all $\mathbf{v}, \mathbf{w}$ and $\mathbf{u}$ in V,

1. $d(\mathbf{v}, \mathbf{v}) \geq 0$ with equality if and only if $\mathbf{v} = 0$.
2. $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$.
3. (Triangle inequality) $d(\mathbf{v}, \mathbf{w}) \leq d(\mathbf{v}, \mathbf{u}) + d(\mathbf{u}, \mathbf{w})$

The distance function allows us to talk about converging sequences of vectors in V; more on this below.

Orthonormal Bases

Definition. Let V be a vector space with an inner product. A set an *orthogonal set* if $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = 0$ for all $i \neq j$, and is $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an *orthonormal set* if

$$\langle \mathbf{v}_i, \mathbf{v}_j\rangle \;=\; \delta_{ij} \;=\; \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

The concise equation $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{ij}$ means that each $\mathbf{v}_i$ is a unit vector and is orthogonal to all the other $\mathbf{v}$'s.

Theorem 0.1. *Orthonormal sets are linearly independent.*

Proof. Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \dots\}$ is an orthonormal set and that $\sum_i x^i \mathbf{v}_i = 0$. Then for each j, $0 = \langle \mathbf{v}_j, 0\rangle = \langle \mathbf{v}_j, \sum x^i \mathbf{v}_i\rangle = \sum x^i \langle \mathbf{v}_j, \mathbf{v}_i\rangle = x^j$. Thus the set is linearly independent. $\square$

Note that if $\dim V = n$, any orthogonal set has at most n vectors, and if it has n vectors then it is a basis.

Fourier Coefficients. Orthonormal bases are especially convenient because the coordinates of a vector are found by simply taking inner products.

Theorem. *Let V be a finite-dimensional inner product space with an orthonormal basis* $\{\mathbf{e}_1, \mathbf{e}_2, \dots\}$. *Then for each* $\mathbf{v} \in V$

$$\mathbf{v} = \sum_{\text{all } n} x^i\, \mathbf{e}_i \qquad \text{where} \quad x^i = \langle \mathbf{e}_i, \mathbf{v}, \rangle. \tag{0.1}$$

Furthermore, if $\mathbf{v} = \sum_{\text{all } n} x^i\, \mathbf{e}_i$ and $\mathbf{w} = \sum_{\text{all } n} y^i\, \mathbf{e}_i$, then

$$\langle \mathbf{v}, \mathbf{w}\rangle = \sum_i \bar{x}^i y^i. \tag{0.2}$$

Proof. Because $\{\mathbf{e}_1, \mathbf{e}_2, \dots\}$ is a basis, $\mathbf{v}$ has a unique expansion $\mathbf{v} = \sum x^i \, \mathbf{e}_i$. But taking the inner product with $\mathbf{e}_j$ gives $\langle \mathbf{e}_j, \mathbf{v} \rangle = \langle \mathbf{e}_j, \sum x^i \mathbf{e}_i \rangle = \sum x^i \langle \mathbf{e}_j, \mathbf{e}_i \rangle = x^j$. The last statement follows similarly:

$$\langle \mathbf{v}, \mathbf{w} \rangle \;=\; \Big\langle \sum_i x^i \mathbf{e}_i,\; \sum_j y^j \mathbf{e}_j \Big\rangle \;=\; \sum_i \sum_j \bar{x}^i y^j \, \langle \mathbf{e}_i, \mathbf{e}_j \rangle \;=\; \sum_i \sum_j \bar{x}^i y^j \, \delta_{ij} \;=\; \sum_i \bar{x}^i y^i.$$

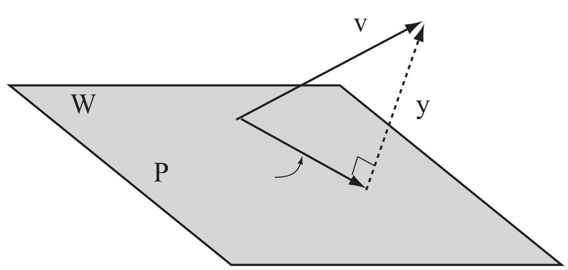
□

The numbers $x^i = \langle \mathbf{e}_i, \mathbf{v} \rangle$ are often called the "Fourier coefficients of $\mathbf{v}$" for reasons that will become clear soon. Note that according to (0.2) the inner product of any two vectors is the (complex) dot product of their Fournier coefficients.

Projections.

Definition. Let V be an inner product space (possibly infinite-dimensional). Given a vector $\mathbf{v} \in V$ and a subspace $W \subset V$ with an orthonormal basis $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$, the *orthogonal projection of* $\mathbf{v}$ *onto* W is the vector

$$\mathrm{Proj}_W(\mathbf{v}) \;=\; \sum \langle \mathbf{w}_j, \mathbf{v} \rangle \; \mathbf{w}_j.$$



Proposition. *Proj*$_W(\mathbf{v})$ *is the vector in* W *closest to* $\mathbf{v}$.

Proof. Choose an orthonormal basis $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ of W (see Gram-Schmidt below) and consider the function

$$D(x^1, \cdots x^k) \;=\; \Big\| v - \sum x^i \mathbf{e}_i \Big\|,$$

which is the distance from $\mathbf{v}$ to an arbitrary vector $\mathbf{w} = \sum x^i \mathbf{e}_i$ in W. Expanding by the bilinear property and completing the square, we have

$$\begin{aligned}
\Big\| v - \sum x^i \mathbf{e}_i \Big\|^2 \;&=\; \|\mathbf{v}\|^2 - 2 \sum x^i \langle \mathbf{e}_i, \mathbf{v} \rangle + \sum_i \sum_j x^i x^j \langle \mathbf{e}_i, \mathbf{e}_j \rangle \\
&=\; \sum_i \big| x^i - \langle \mathbf{e}_i, \mathbf{v} \rangle \big|^2 + \|\mathbf{v}\|^2 - \sum_i |\langle \mathbf{e}_i, \mathbf{v} \rangle|^2 \qquad (0.3) \\
&=\; \sum_i \big| x^i - \langle \mathbf{e}_i, \mathbf{v} \rangle \big|^2 + C
\end{aligned}$$

where C is a constant independent of the numbers x^i. Thus $D(\mathbf{x}) \geq C$ with equality if and only if $x^i = \langle \mathbf{v}, \mathbf{e}_i \rangle$ for all i.

□

Corollary (Bessel's inequality). *If* $\mathbf{v} = \sum x^i \mathbf{e}_i$ *then* $\sum_i |x^i|^2 \leq \|\mathbf{v}\|^2$.

Proof. Take $x^i = \langle \mathbf{e}_i, \mathbf{v} rangle$ in (0.3), noting that the left-hand side is non-negative. □

Note that (0.2) shows that Bessel's inequality is actually an equality when V is finite-dimensional. But the infinite-dimensional case is especially interesting – see below.

The Gram-Schmidt Process. Given a basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of an inner product space, we can form an orthonormal basis $\{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ by these steps:

1. Normalize $\mathbf{u}_1$ by setting $\mathbf{e}_1 = \dfrac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$.

2a. Replace $\mathbf{u}_2$ by the component of $\mathbf{u}_2$ orthogonal to $\mathbf{e}_1$ by setting $\mathbf{v}_2 = \mathbf{u}_2 - \langle \mathbf{e}_1, \mathbf{u}_2 \rangle\ \mathbf{e}_1$.

2b. Normalize $\mathbf{v}_2$ by setting $\mathbf{e}_2 = \dfrac{\mathbf{u}_2}{\|\mathbf{u}_2\|}$.

3a. Replace $\mathbf{u}_3$ by the component of $\mathbf{u}_3$ orthogonal to $W_2 = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ by setting

$$\mathbf{v}_3 = \mathbf{u}_3 - \langle \mathbf{e}_1, \mathbf{u}_3 \rangle\ \mathbf{e}_1 - \langle \mathbf{e}_2, \mathbf{u}_3 \rangle\ \mathbf{e}_2.$$

3b. Normalize $\mathbf{v}_3$, etc.

Thus at the kth step we calculate

$$\mathbf{v}_k \;=\; \mathbf{u}_k - \Big(\langle \mathbf{e}_1, \mathbf{u}_k \rangle\ \mathbf{e}_1 + \langle \mathbf{e}_2, \mathbf{u}_k \rangle\ \mathbf{e}_2 + \cdots + \langle \mathbf{e}_{k-1}, \mathbf{u}_k \rangle\ \mathbf{e}_{k-1} \Big)$$

and then set $\mathbf{e}_k = \dfrac{\mathbf{v}_k}{\|\mathbf{v}_k\|}$.

Tips: When applying the Gram-Schmidt process with specific vectors, it is helpful to:

(i) Pull out common factors. For example, write $\sqrt{2}(1,1,1)$ instead of $(\sqrt{2}, \sqrt{2}, \sqrt{2})$.

(ii) Erase the common factor before normalizing. For example, the normalization of $(\sqrt{2}, \sqrt{2}, \sqrt{2})$ is the same as the normalization of $(1,1,1)$, which is much easier to compute.