Projection Properties

1 Vector space generalities

Our vector space in this course will be the L_2-space associated with some probability on a sample space. However, for some of the concepts it may be clearer to (briefly) discuss a more general setting.

The vector space concept

A vector space is a set, together with two operations, one called vector addition and one called multiplication by a number. We often refer to a number as a *scalar*. In our work we use real numbers, and so they are our scalars. In other situations complex numbers might be used instead of real numbers, and then the complex numbers would be the scalars. The vector space operations must satisfy the following basic properties.

(i) Vector addition must be associative and commutative.

(ii) Multiplication by a scalar must be associative, in the obvious sense that $s(tv) = (st)v$ for any numbers s, t and any vector v.

(iii) Multiplication by a scalar must be distributive over vector addition.

(iv) There must exist a vector (denoted here by $\bar{0}$), such that $\bar{0} + a = a$ for all vectors a. We must have $0a = \bar{0}$ for every vector a.

(v) We must have $1v = v$ for every vector v.

The main vector space example

If A is any set and V is the set of all functions from A to $\mathbb{R}$ then V is a vector space, where the two vector space operations are *pointwise* addition of functions and *pointwise* multiplication of a function by a real number.

Each vector in the example space V is a function from A to $\mathbb{R}$. The axioms defining a vector space hold for our example space V by a simple argument, using the fact that the real numbers satisfy the same properties.

The space $\mathbb{R}^n$ of *coordinate vectors* is a special case of the main vector space example, because we can think of an n-tuple $(x_1, \ldots, x_n)$ as a function from $\{1, \ldots, n\}$ to $\mathbb{R}$.

Subspaces

If V is a vector space, and if G is any subset of V which is **closed under vector addition** and **closed under multiplication by a scalar**, then G is said to be a subspace of V. The *key point* here is that any subspace of vector space is also a vector space in its own right, using the same vector operations.

Because of this key point, the main vector space example gives rise to many other examples, which are all subspaces of the main example.

The norm concept

Let V be a vector space. A *norm* on V is a function φ from V to the scalars with the following properties.

(i) φ is *positive*: $\varphi(v) \geq 0$ for all vectors v, and if $\varphi(v) = 0$ then $v = 0$.

(ii) The *triangle inequality holds*: $\varphi(v + w) \leq \varphi(v) + \varphi(w)$ for all vectors v, w.

(iii) φ is *homogeneous*: $\varphi(cv) = |c|\,\varphi(v)$ for every vector v and every scalar c.

The value $\varphi(v)$ is called the norm of v. We often write $\|v\|$ for the norm of v.

The scalar product concept

Let V be a vector space. A *scalar product* on V is a function λ from $V \times V$ to the scalars with the following properties.

(i) λ is *bilinear*: $\lambda(u, av + bw) = a\lambda(u, v) + b\lambda(u, w)$ for all vectors u, v, w and all scalars a, b; $\lambda(au + bw, v) = a\lambda(u, v) + b\lambda(w, v)$ for all vectors u, v, w and all scalars a, b.

(ii) λ is *symmetric*: $\lambda(u, v) = \lambda(v, u)$ for all vectors u, v.

(iii) φ is *positive*: $\lambda(u, u) \geq 0$ for all vectors u, and if $\lambda(u, u) = 0$ then $u = 0$.

We often write $\langle u, v\rangle$ for the scalar product of u and v.

A scalar product space

If we have a scalar product on a vector space K, we will refer to K as a *scalar product space*, and we will write the scalar product of two vectors u and v as $\langle u, v\rangle$ unless otherwise stated.

It is an important fact that whenever we have a scalar product space, we can define an *associated norm* by $\|v\| = \sqrt{\langle v, v\rangle}$. The proof is given on page 98 of the text for the case of L_2 of a finite set Ω, but the same argument works in all cases.

It is also important to know that whenever we have a scalar product space, then Cauchy's Inequality holds:

$$|\langle u, v\rangle| \leq \|u\|\,\|v\|\,.$$

Again, the proof is given on page 98 of the text for the case of L_2 of a finite set Ω, but the same argument works in all cases.

The theorem of Pythagoras

Consider vectors in an scalar product space. We write $u \perp v$ if $\langle u, v\rangle = 0$ and say that u and v are perpendicular. (We also say, equivalently, that u and v are orthogonal.) For any perpendicular vectors u and v, we have

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 .$$

This is the essential content of the theorem of Pythagoras. It follows immediately by expanding $\langle u+v, u+v\rangle$. This is a good example of the geometric power of the scalar product concept. Problem 4.1.4 in the text, the *parallelogram identity*, is a related example.

2 Scalar products and projections

In honor of our textbook, from now on we'll tend to write vectors with uppercase letters, so they *look like* random variables. But the facts we prove will hold for vectors in any scalar product space.

Lemma 1 (Uniqueness of projections) *Let K be a scalar product space. Let $\mathcal{L}$ be any subspace of K. Let $X \in K$. There is at most one element Z in $\mathcal{L}$ such that $X - Z \perp \mathcal{L}$.*

Proof Suppose that Z in $\mathcal{L}$ with $X - Z \perp \mathcal{L}$, and also H in $\mathcal{L}$ with $X - H \perp \mathcal{L}$. We must show that $Z = H$.

We have

$$\langle X - Z, Y\rangle = 0 \text{ for every } Y \in \mathcal{L}$$

and

$$\langle X - H, Y\rangle = 0 \text{ for every } Y \in \mathcal{L}.$$

Subtracting,

$$\langle H - Z, Y\rangle = 0 \text{ for every } Y \in \mathcal{L}.$$

But $H - Z \in \mathcal{L}$, so we have $\langle H - Z, H - Z\rangle = 0$, and so $H - Z = 0$.

This proves the lemma.

Lemma 1 contains the **uniqueness part** of **Theorem 4.1.9**.

Definition 1 (The projection of a vector on a subspace) *Let K be a scalar product space. Let $\mathcal{L}$ be a subspace of K. Let $X \in K$. Then $\Pi_{\mathcal{L}} X$ is that vector Z, if it exists, such that $Z \in \mathcal{L}$ and $X - Z \perp \mathcal{L}$.*

Our definition of projection is the same as that given in the **text in the middle of page 99**, except that we already proved uniqueness.

Lemma 2 (Projections give the same scalar products) *Let* K *be a scalar product space. Let* $\mathcal{L}$ *be a subspace of* K*. Let* $X \in K$*. Let* Z *in* $\mathcal{L}$*. The following statements are equivalent.*

(i) $\Pi_{\mathcal{L}} X$ *exists and* $Z = \Pi_{\mathcal{L}} X$*.*

(ii) $\langle Z, Y \rangle = \langle X, Y \rangle$ *for every* $Y \in \mathcal{L}$*.*

Proof Suppose that (i) holds. Then by definition we have $X - Z \perp \mathcal{L}$. That is, $\langle X - Z, Y \rangle = 0$ for every $Y \in \mathcal{L}$. But then $\langle X, Y \rangle = \langle Z, Y \rangle$ for every $Y \in \mathcal{L}$, so (ii) holds.

Suppose that (ii) holds. Then $\mathcal{Z} \in \mathcal{L}$ and $\langle Z - X, Y \rangle = 0$, so that by definition $\Pi_{\mathcal{L}} X$ exists and $Z = \Pi_{\mathcal{L}} X$. Thus (i) holds.

This proves the lemma.

Lemma 2 implies **part (i) of Theorem 4.1.22** in the text. Lemma 2 is also relevant to **Remark 4.2.8** in the text.

Lemma 3 (Projection is linear) *Let* K *be a scalar product space. Let* $\mathcal{L}$ *be a subspace of* K*. Let* $X, Y \in K$ *and let* $a, b \in \mathbb{R}$*. Suppose that* $\Pi_{\mathcal{L}}\, X$ *exists and* $\Pi_{\mathcal{L}}\, Y$ *exists. Then* $\Pi_{\mathcal{L}}\, (aX + bY)$ *exists, and*

$$\Pi_{\mathcal{L}}\, (aX + bY) = a\Pi_{\mathcal{L}}\, X + b\Pi_{\mathcal{L}}\, Y. \tag{1}$$

Proof Let $G = \Pi_{\mathcal{L}}\, X$ and let $H = \Pi_{\mathcal{L}}\, Y$. Let $Z = aG + bH$. By definition, $G, H \in \mathcal{L}$, $X - G \perp \mathcal{L}$ and $Y - H \perp \mathcal{L}$.

Hence for every $U \in \mathcal{L}$ we have from Lemma 2 that

$$\begin{aligned} \langle Z, U \rangle &= \langle aG + bH, U \rangle \\ &= a\langle G, U \rangle + b\langle H, U \rangle \\ &= a\langle X, U \rangle + b\langle Y, U \rangle \\ &= \langle aX + bY, U \rangle. \end{aligned}$$

Using Lemma 2 one more time we conclude that $\Pi_{\mathcal{L}}\, (aX + bY)$ exists, and $\Pi_{\mathcal{L}}\, (aX + bY) = Z$. But then substituting for G and H in the definition of Z shows that (1) holds, so the lemma is proved.

Lemma 3 implies **Problem 4.1.10 in the text**.

Lemma 4 (Projections telescope) *Let* K *be a scalar product space. Let* $\mathcal{L}_1$ *and* $\mathcal{L}_2$ *be subspaces of* K*. with* $\mathcal{L}_1 \subset \mathcal{L}_2$*.*

Let $X \in K$*, such that* $\Pi_{\mathcal{L}_2}\, X$ *exists. Then* $\Pi_{\mathcal{L}_1}\, \Pi_X\, \mathcal{L}_2$ *exists if and only if* $\Pi_{\mathcal{L}_1}\, X$ *exists, and if these projections exist then*

$$\Pi_{\mathcal{L}_1}\, \Pi_{\mathcal{L}_2}\, X = \Pi_{\mathcal{L}_1}\, X. \tag{2}$$

Proof Let $G = \Pi_{\mathcal{L}_2} X$.

Assume that $\Pi_{\mathcal{L}_1} \Pi_X \mathcal{L}_2$ exists. Let $Z = \Pi_{\mathcal{L}_1} \Pi_X \mathcal{L}_2$.

Using Lemma 2 twice,

$$\langle Z, Y\rangle = \langle G, Y\rangle = \langle X, Y\rangle$$

for every $Y \in \mathcal{L}_1$. But then using Lemma 2 again, $\Pi_{\mathcal{L}_1} X = Z$, so that (2) holds.

Now assume that $\Pi_{\mathcal{L}_1} X$ exists. Let $Z = \Pi_{\mathcal{L}_1} X$. Using Lemma 2 twice,

$$\langle Z, Y\rangle = \langle X, Y\rangle = \langle G, Y\rangle$$

for every $Y \in \mathcal{L}_1$. But then using Lemma 2 again, $\Pi_{\mathcal{L}_1} G = Z$, so that (2) holds again.

This proves the lemma.

Lemma 4 implies **part (ii) of Theorem 4.1.22**.

Lemma 5 (A projection is a best approximation) *Let K be a scalar product space. Let $\mathcal{L}$ be a subspace of K. Let $X \in K$. Let Z in $\mathcal{L}$. The following statements are equivalent.*

(i) $X - Z \perp \mathcal{L}$.

(ii) $\|X - Z\| \leq \|X - Y\|$ *for every* $Y \in \mathcal{L}$.

Remark Note that the condition in part (ii) says that Z is the point in $\mathcal{L}$ which is at a *minimum distance* from X.

Proof of Lemma 5 Suppose that (i) holds. Let $Y \in \mathcal{L}$. Using Pythagoras and the fact that $Z - Y \in \mathcal{L}$, we have

$$\begin{aligned} \|X - Y\|^2 &= \|X - Z + Z - Y\|^2 \\ &= \|X - Z\|^2 + \|Z - Y\|^2 \\ &\geq \|X - Z\|^2 . \end{aligned}$$

Thus (ii) holds. (This argument is given on **page 102 of the text**, to prove equation (10) on that page directly.)

Now suppose that (ii) holds. Let $Y \in \mathcal{L}$. Then for every real number t, $Z + tY \in \mathcal{L}$, so we have

$$\|X - (Z + tY)\|^2 \geq \|X - Z\|^2 .$$

Hence the polynomial $\|X - (Z + tY)\|^2$ obtains its *minimum* when $t = 0$, and therefore its derivative must be zero at that point. But the derivative at $t = 0$ is $-2\langle X - Z, Y\rangle$, so we have shown that $X - Z \perp \mathcal{L}$. Thus (i) holds. (This is the argument used in the text, at the bottom of page 101, as a step in the **existence part** of the proof of **Theorem 4.1.9**. We haven't finished the proof of existence of projections, however.)

This completes the proof of the lemma.

Lemma 6 (Projections on complete subspaces exist) *Let K be a scalar product space. Let $\mathcal{L}$ be a subspace of K. Suppose that $\mathcal{L}$ is complete, meaning that whenever Y_n is a sequence of elements in $\mathcal{L}$ such that*

$$\lim_{m,n\to\infty} \|Y_m - Y_n\| = 0, \tag{3}$$

the sequence Y_n converges to a limit Z in $\mathcal{L}$.

Let $X \in K$. Then $\Pi_{\mathcal{L}} X$ exists.

Proof Let

$$c = \inf \left\{ \|X - Y\| : \ Y \in \mathcal{L} \right\}.$$

Let Y_n be a sequence of elements in $\mathcal{L}$ such that $\|X - Y_n\|$ becomes as small as possible, that is, such that

$$\lim_{n\to\infty} \|X - Y_n\| = c.$$

We claim that the sequence Y_n is Cauchy.

To see that, we study the *parallelogram identity*, proved in **Problem 4.1.4** in the text. Dividing the original inequality by 4, we can write the parallelogram identity as

$$\left\| \frac{V+W}{2} \right\|^2 + \left\| \frac{V-W}{2} \right\|^2 = \frac{\|V\|^2 + \|W\|^2}{2} \tag{4}$$

for any vectors V, W. Equivalently,

$$\left\| \frac{V-W}{2} \right\|^2 = \frac{\|V\|^2 + \|W\|^2}{2} - \left\| \frac{V+W}{2} \right\| \tag{5}$$

for any vectors V, W. (Thus the *length of the average* of two vectors must be significantly *shorter* than the average of their lengths, *unless* the vectors are *close to each other*.)

Using (5) with $V = X - Y_m$ and $W = X - Y_n$, we see that

$$\begin{aligned} 0 \le \left\| \frac{Y_m - Y_m}{2} \right\|^2 = \frac{\|X - Y_m\|^2 + \|X - Y_n\|^2}{2} &- \left\| X - \frac{Y_m + Y_n}{2} \right\| \\ \le \frac{\|X - Y_m\|^2 + \|X - Y_n\|^2}{2} &- c. \end{aligned} \tag{6}$$

Since

$$\lim_{m,n\to\infty} \frac{\|X - Y_m\|^2 + \|X - Y_n\|^2}{2} - c = 0, \tag{7}$$

we see by the “Pinching Theorem” of calculus that

$$\lim_{m,n\to\infty} \left\| \frac{Y_m - Y_m}{2} \right\|^2 = 0. \tag{8}$$

That is, the sequence Y_n is Cauchy.

By assumption, there is an element $Z \in \mathcal{L}$ such that $\|Y_n - Z\| \to 0$ as $n \to \infty$. Since

$$c \leq \|X - Z\| \leq \|X - Y_n\| + \|Y_n - Z\|,$$

letting $n \to \infty$ we see that

$$\|X - Z\| = c. \tag{9}$$

Hence condition (ii) of Lemma 5 holds, and so by Lemma 5 we know that $\Pi_{\mathcal{L}} X$ exists.

This proves the lemma.

Lemma 6 completes the proof of **Theorem 4.1.9** of the text. The argument given here to prove Lemma 6 appears on page 110 of the text as part of the proof of Theorem 4.1.9.