Notes on Hilbert Space

The Projection Theorem and Some of Its Consequences

Basic Results

Let X denote a vector space over the field of real scalars $\mathbf{R}$. An *inner product* on X is a function $\langle \cdot | \cdot \rangle : X \times X \to \mathbf{R}$ satisfying, for every $x, y, z \in X$ and $l \in \mathbf{R}$,

1. $\langle x | y \rangle = \langle y | x \rangle$, symmetry
2. $\langle x + y | z \rangle = \langle x | z \rangle + \langle y | z \rangle$, linearity
3. $\langle l\, x | y \rangle = l \langle x | y \rangle$, linearity
4. $\langle x | x \rangle \geq 0$ and $\langle x | x \rangle = 0$ if and only if $x = 0$, positivity.

The space X with an inner product is called a *pre-Hilbert space*. We assume henceforth that X is a pre-Hilbert space. For each $x \in X$, let $\|x\| = \langle x | x \rangle^{1/2}$. We will show below (in Proposition 2) that this gives a norm on X. A pre-Hilbert space that is complete in this norm is called a Hilbert space. So you have some idea that this is not an uninteresting abstraction, you should note that all Euclidean spaces $\mathbf{R}^n$ for finite n are Hilbert spaces with the standard Euclidean norm $\|x\|$ = $\left(\sum_{i=1}^n x_1^2\right)^{1/2}$ and inner product $\langle x | y \rangle = x \cdot y = \sum_{i=1}^n x_i y_i$, the standard dot product. We will introduce and different inner product below. The space of random variables with finite second moment is a Hilbert space. It is this space that is a standard model for contingent claims (securities, payoffs) in finance. First we need

Lemma 1. (*The Cauchy-Schwartz Inequality*) For each $x, y \in X$, $\left|\langle x|y\rangle\right| \le \|x\|\|y\|$. Equality holds if and only if $x = l\, y$ or $y = 0$.

Proof. If $y = 0$, the inequality holds trivially since by 1 and 3, $\langle y|x\rangle = \langle 0y|x\rangle = 0\langle y|x\rangle = 0$. Therefore assume that $y \ne 0$. For all scalars l, we have

$$0 \le \langle x - l\, y | x - l\, y\rangle \; = \; \langle x|x\rangle - l\, \langle x|y\rangle - l\, \langle y|x\rangle + |l|^2 \langle y|y\rangle . \qquad (1)$$

There can be equality in (1) if and only if $x = l\, y$. In particular, for $l = \langle x|y\rangle / \langle y|y\rangle$, we have from (1) that

$$0 \le \langle x|x\rangle - 2\left|\langle x|y\rangle\right|^2 / \langle y|y\rangle + \left|\langle x|y\rangle\right|^2 / \langle y|y\rangle = \langle x|x\rangle - \left|\langle x|y\rangle\right|^2 / \langle y|y\rangle ,$$

or

$$\left|\langle x|y\rangle\right|^2 \le \langle x|x\rangle \langle y|y\rangle .$$

Taking square roots completes the proof.[]

Proposition 2. On a pre-Hilbert space X, the function $\|x\| = \langle x|x\rangle^{1/2}$ is a norm on X.

Proof. Clearly by 1 and 3, $\|a\,x\| = |a|\|x\|$, $a \in \mathbf{R}$, and by 4, $\|x\| > 0, x \ne 0$. The only requirement for a norm that remains is to prove the triangle inequality. For $x, y \in X$, we have that

$$\begin{aligned} \|x+y\|^2 &= \langle x+y|x+y\rangle = \langle x|x\rangle + 2\langle x|y\rangle + \langle y|y\rangle \\ &\le \|x\|^2 + 2\left|\langle x|y\rangle\right| + \|y\|^2 \\ &\le \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \end{aligned}$$

where the last inequality follows from the Cauchy-Schwartz inequality. []

There are various properties of the inner product that are useful. Two easy ones you should be able to show (ysbats) are the following.

Lemma 3. For a given $x \in X$, $\langle x | y \rangle = 0$, for every $y \in X$, implies that $x = 0$.

Lemma 4. (*The Parallelogram Law*) In a pre-Hilbert space, $\|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right)$.

A sequence $\{x_n\} \subset X$, converges to $x \in X$ if $\lim_{n\to\infty} \|x_n - x\| = 0$. In this case we abbreviate by writing $x_n \to x$.[1]

Lemma 5. (*Continuity of the Inner Product*) Suppose that $x_n \to x$ and $y_n \to y$ in X. Then $\langle x_n | y_n \rangle \to \langle x | y \rangle$.

Proof. Since the sequence $\{x_n\}$ converges, it is norm bounded, say $\|x_n\| \le M$. Now

$$\left|\langle x_n | y_n \rangle - \langle x | y \rangle\right| = \left|\langle x_n | y_n \rangle - \langle x_n | y \rangle + \langle x_n | y \rangle - \langle x | y \rangle\right| \le \left|\langle x_n | y_n - y \rangle\right| + \left|\langle x_n - x | y \rangle\right|,$$

the inequality following from the triangle inequality for absolute value. By the Cauchy-Schwartz inequality, we obtain

$$\left|\langle x_n | y_n \rangle - \langle x | y \rangle\right| \le \|x_n\|\|y_n - y\| + \|x_n - x\|\|y\| \le M\|y_n - y\| + \|x_n - x\|\|y\| \to 0. \text{ []}$$

In a pre-Hilbert space, two vectors x and y are said to be *orthogonal* (perpendicular) if $\langle x | y \rangle = 0$. We symbolize this by writing $x \perp y$. A vector x is said to be orthogonal to a set S, written $x \perp S$, if $x \perp s$ for every $s \in S$. Using the first few lines of the proof of Proposition 2, ysbats that the Pythagorean Theorem holds in pre-Hilbert spaces.

Lemma 6. (*Pythagorean Theorem*) If $x \perp y$, then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.

The Projection Theorem

The next result is one of the most important optimization results in functional analysis.

Theorem 7. (*Projection Theorem*) Let X be a pre-Hilbert space, let M be a subspace of X, and let x be an arbitrary vector in X. If there is a vector $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$, then m_0 is unique. A necessary and sufficient condition that $m_0 \in M$ be the unique vector that minimizes $\|x - m\|$ over $m \in M$ is that $x - m_0 \perp M$. If M is complete (Cauchy sequences in M converge in M), then a unique vector that minimizes $\|x - m\|$ over $m \in M$ exists.

Proof. Let $x \in X$ be arbitrary. We show first that if $m_0 \in M$ minimizes $\|x - m\|$ over $m \in M$, then $x - m_0$ is orthogonal to M. Suppose not. Then there exists $m \in M$ such that $\langle x - m_0 | m \rangle = d \neq 0$. Without loss of generality we may assume that $\|m\|$ = 1. Let $m_1 = m_0 + d\,m$. Then $m_1 \in M$, and

$$\|x - m_1\|^2 = \|x - m_0 - d\,m\|^2 = \|x - m_0\|^2 - \langle x - m_0 | d\,m \rangle - \langle d\,m | x - m_0 \rangle + |d|^2$$

$$= \|x - m_0\|^2 - |d|^2 < \|x - m_0\|^2 .$$

Thus if $x - m_0$ is not orthogonal to M, then m is not a minimizing vector.

Next we show that if $x - m_0$ is orthogonal to M, then it is a unique minimizing vector. If $x - m_0$ is orthogonal to M, then for any $m \in M$, the Pythagorean theorem (Lemma 6) gives

---

[1] Any convergent sequence is norm bounded in the following sense. Suppose that $x_n \to x$. Then there exists and integer N such that for all $n \geq N$, $\|x - x_n\| \leq 1$. We then have $\|x_n\| = \|x - (x - x_n)\| \leq \|x\| + \|x - x_n\| \leq \|x\| + 1$, for all $n \geq N$. Of course for $n < N$, $\|x_n\| \leq \max_{1 \leq m \leq N} \|x_m\|$.

$\|x-m\|^2 = \|x-m_0+m_0-m\|^2 = \|x-m_0\|^2 + \|m_0-m\|^2$. Thus $\|x-m\|^2 > \|x-m_0\|^2$, for any $m \neq m_0$. Taking square roots gives the result.

Finally, we prove that a minimizing m_0 exists when M is complete. If $x \in M$, then $m_0 = x$ and there is nothing to prove. Therefore assume that $x \notin M$ and let $d = \inf\{\|x-m\| : m \in M\}$. Let $\{m_n\}$ be a sequence of vectors in M such that $\|x-m_n\| \to d$. Now, by the parallelogram law (Lemma 4),

$$\|(m_k-x)+(x-m_n)\|^2 + \|(m_k-x)-(x-m_n)\|^2 = 2\|m_k-x\|^2 + 2\|x-m_n\|^2. \qquad (2)$$

The first term on the left is $\|m_k-m_n\|^2$, while the second term is

$$\|2(x-(m_k+m_n)/2)\|^2 = 4\|x-(m_k+m_n)/2\|^2.$$

Substituting in (2) and rearranging, we get that

$$\|m_k-m_n\|^2 = 2\|m_k-x\|^2 + 2\|x-m_n\|^2 - 4\|x-(m_k+m_n)/2\|^2.$$

For all $k, n, (m_k+m_n)/2$ is in M, and $\|x-(m_k+m_n)/2\| \geq d$. Therefore,

$$\|m_k-m_n\|^2 \leq 2\|m_k-x\|^2 + 2\|x-m_n\|^2 - 4d^2.$$

Since $\|m_k-x\|^2 \to d^2$ and $\|x-m_n\|^2 \to d^2$, we conclude that $\|m_k-m_n\| \to 0$ as $k, n \to \infty$. Thus $\{m_n\}$ is Cauchy and hence it converges to some $m_0 \in M$. By continuity of the inner product (Lemma 5) and hence the norm, we have that $d = \lim_{n\to\infty}\|x-m_n\| = \|x-m_0\|$. []

Theorem 7 does two things. First it characterizes the unique vector in the subspace M, if it exists, that minimizes the distance of x to M as the one m_0 such that $x-m_0$ is orthogonal to M. That vector m_0 is therefore called the *orthogonal projection* of x onto M.

The second thing that Theorem 7 does is give us sufficient conditions for the existence of the orthogonal projection, namely that the subspace M be complete. Of course a complete subspace is closed in the norm topology, but the converse need not be true. It is true in an important class of cases, namely when the pre-Hilbert space is itself complete. A pre-Hilbert space that is complete in the norm of Proposition 2 is called a *Hilbert space*. A Hilbert space is then a Banach space (normed vector space that is complete) with an inner product that induces the norm. In a Banach space, a subspace is complete if and only if it is closed. As indicated above, all Euclidean spaces $\mathbf{R}^n$ for finite n are Hilbert spaces. The space of random variables with finite second moment is a Hilbert space (see below).

The development above was taken from David G. Luenberger, *Optimization by Vector Space Methods* , John Wiley & Sons, 1969. The projection theorem has numerous applications in optimization theory and statistics. See Luenberger for a range of these applications. It is used extensively in asset pricing theory, in particular, in John H. Cochrane, *Asset Pricing*, Princeton University Press, 2001, referred to hereafter as (C), and in Darrell Duffie, *Dynamic Asset Pricing Theory*, 3rd ed., Princeton University Press, 2001, referred to hereafter as (D).

Representation of Linear Functionals

One purely mathematical application that has some import for asset pricing has to do with the representation of continuous linear functionals on a Hilbert space. Ysbats that asset price functionals must be (positive and) linear or there will be arbitrage opportunities. See Theorems page 66, 71 in (C), lemma of 1F in (D), lemma of 2C in (D), lemma of 10B in (D).

Theorem 8. (*Riesz Representation Theorem*). Let X be a Hilbert space and let $F : X \to \mathbf{R}$ be a continuous linear functional. Then there is a unique $p \in X$ such that $F(x) = \langle p \mid x \rangle$, for each $x \in X$.

Proof.[2] Given X and F as hypothesized, let $J = \{x \in X : F(x) = 0\}$, the kernel of F. Since F is continuous, (ysbats) J is a closed subspace of X. If $J = X$, then F is identically zero and the representation $F(x) = \langle p | x \rangle$, with $p = 0$ (the zero vector in X) works. Its uniqueness follows from property 4 of the inner product. If $J \neq X$, then (ysbats) there exists an element $z_1 \in X$ with $F(z_1) = 1$. By the Projection Theorem, there exists $z_2 \in J$ such that $\langle z_1 - z_2 | x \rangle = 0$, for all $x \in J$. Let $z = z_1 - z_2$ and let $p = z / \langle z | z \rangle$. Note that $F(z) = F(z_1) - F(z_2) = F(z_1) = 1$ since $z_2 \in J$. Note also that $p \perp J$ and that for any x, $x - F(x)z \in J$. Then for any x,

$$0 = \langle p | x - F(\;)z \rangle = \langle p | x \rangle - \frac{\langle z | F(x)z \rangle}{\langle z | z \rangle} = \langle p | x \rangle - F(x).$$

This choice of p thus works to represent F. It is unique because if there is another $\hat{p}$ that represents F, then $\langle p - \hat{p} | x \rangle = 0$ for all $x \in X$. By Lemma 3 above, $p - \hat{p} = 0$.[]

In the application of the Riesz Representation Theorem to asset pricing, the vector p representing a pricing functional is sometimes referred to as a discount factor, a state-price density, state-price deflator, etc. There may be many discount factors and much of asset pricing theory is devoted to providing characterizations of these discount factors. An essential tool in those characterizations is the orthogonal decomposition of a Hilbert space. See in particular section 5.3 of (C).

Orthogonal Compliments

[2] The proof here is an elaboration of Exercise 1.17 in (D) referred to above in the text.

Given a subset S of a pre-Hilbert space, the set of all vectors orthogonal to S is called the orthogonal complement of S and is denoted $S^{\perp}$. Ysbats that $S^{\perp}$ is a closed subspace (using Lemma 5 above). We say that a vector space X is the direct sum of two subspaces M and N if every vector $x \in X$ has a unique representation of the form $x = m + n$, where $m \in M$ and $n \in N$. When X is the direct sum of M and N, we write $X = M \oplus N$. The following is another result of the Projection Theorem.

Theorem 9. If S is a closed subspace of a Hilbert space H, then $H = S \oplus S^{\perp}$ and $S = S^{\perp\perp}$.

Proof. Let $x \in H$. By the Projection Theorem, there is a unique vector $s \in S$ such that $\|x-s\| \quad \|x-y\|$ for all $y \in S$ and $\hat{s} \equiv x - s \in S^{\perp}$. Thus $x = s + \hat{s}$ with $s \in S$ and $\hat{s} \in S^{\perp}$. This representation is unique, again by the Projection Theorem. This establishes that $H = S \oplus S^{\perp}$.

To show that $S = S^{\perp\perp}$, note first that if $x \in S$, then $x \perp y$ for all $y \in S^{\perp}$. Therefore $x \in S^{\perp\perp}$, and hence $S \subseteq S^{\perp\perp}$. To go the other way, let $x \in S^{\perp\perp}$. By the first part of this result, there exists $s \in S$ and $\hat{s} \in S^{\perp}$ with $x = s + \hat{s}$. Since both $x \in S^{\perp\perp}$ and $s \in S \subseteq S^{\perp\perp}$, we have $\hat{s} = x - s \in S^{\perp\perp}$. But also $\hat{s} \in S^{\perp}$, implying that $\langle \hat{s} | \hat{s} \rangle = 0$, and hence by property 4 of the inner product $\hat{s} = 0$. Thus $x = s \in S$, proving $S^{\perp\perp} \subseteq S$.[]

Many applications of interest in the theory of valuation involve the space of random variables with finite variances. We will casually refer to such spaces as payoff spaces. Fix a probability space (Ω, Φ, P) as the underlying model of uncertainty and let $L^2 = L^2(\Omega, \Phi, P)$ be the class of real valued (Φ-measurable) random variables X such that the expectation of the square of X is finite, i. e., $E(X^2) < \infty$, where E denotes expectation with respect to the probability measure P. The space L^2 with the obvious definitions of addition and scalar multiplication (with real scalars) is a real vector space. For each $X,Y \in L^2$, we have $E\left(|X|\right) < \infty$, $Var\left(X\right) = E\left(X^2\right) - E\left(X\right)^2 < \infty$, and $E\left(|XY|\right) \le E\left(X^2\right)^{1/2} E\left(Y^2\right)^{1/2} < \infty$, with the last (weak) inequalty following the Cauchy-Schwartz inequality for square integrable functions. It follows that $Cov\left(X,Y\right) = E\left(XY\right) - E\left(X\right)E\left(Y\right)$ exists and is finite.

This square integrable structure allows us to conclude that L^2 (or at least equivalence classes of L^2 functions under the equivalence relation = almost surely) is a Hilbert space with inner product $\left\langle X \middle| Y \right\rangle \equiv E\left(XY\right)$ and norm $\|X\| = \left\langle X \mid X \right\rangle^{1/2} = E\left(X^2\right)^{1/2}$. The completeness of this space in this norm is proved in most books on probability theory. See Theorem 6.6.2 in the text by Resnick, *A Probability Path*, or Chapter 6, especially section 6.10 of the Williams text *Probability with Martingales*. To test your understanding of these ideas, use the Pythagorean theorem to show that $Var\left(X+Y\right) = Var\left(X\right) + Var\left(Y\right)$, provided $Cov\left(X,Y\right) = 0$.

Special cases of great interest in asset pricing are when Ω is a finite set of states, say $\Omega = \{1,\ldots,S\}$, S a finite integer. In this case Φ can be taken to be the set of all subsets of Ω, and the measure P can be specified by any vector $p = \left(p_1,\ldots,p_S\right)$ >> 0 with $P\left(A\right) = \sum_{s\in A} p_s$, as in Section 1.F in (D). Any random variable or payoff X defined on Ω can be identified with an element of the Euclidean space $\mathbf{R}^S$ with $X = \left(X\left(1\right),\ldots,X\left(S\right)\right)$, where $X\left(s\right)$ is the payoff of X in the s^{th} state. We know that $\mathbf{R}^S$ is a Hilbert space in the Euclidean norm and inner product. For most applications in asset pricing, however, it is more convenient to endow $\mathbf{R}^S$ with the machinery of $L^2\left(\Omega,\Phi,P\right)$. This is done in Section 1.F of (D) and less formally by (C) throughout when the space of payoffs is $\mathbf{R}^S$.

For many applications in asset pricing, it is necessary to have a state space that is not finite. For example, for normally distributed or log normally distributed returns the underlying state space has to be at least as large as the set $\mathbf{R}$ of real numbers. When Ω is not finite, then the structure of $L^2(\Omega, \Phi, P)$ is still very valuable but this vector space is no longer a finite dimensional vector space. It is infinite dimensional and occasionally this infinite dimensional character must be handled with care because the mathematics of infinite dimensional vector spaces is more complicated than that of finite dimensional spaces.

Much of what is done in asset pricing, however, is done in terms of a finite number of assets or payoffs. In this case we would be dealing with the span of a finite number of elements of $L^2(\Omega, \Phi, P)$, a finite dimensional subspace. Finite dimensional subspaces of $L^2(\Omega, \Phi, P)$ are in most respects just like the spaces we deal with when Ω is finite.

Another application of the Projection Theorem in the context of $L^2(\Omega, \Phi, P)$ is the existence of conditional expectations of square integrable random variables. Of course such things exist, and you need to be proficient in the use of their properties as described in section 2 of the paper L. C. G. Rogers, "Stochastic Calculus and Markov Methods." But the existence argument using the Projection Theorem allows an interpretation of the conditional expectation in terms of its optimality properties. This argument is given at the end of these notes, following the more technical existence stuff, which can be skipped.

Conditional Expectation: An Application

Let Σ be a sub-σ-algebra of Φ, and let X denote an integrable random variable on (Ω, Φ, P). As in section 2 of the paper L. C. G. Rogers, "Stochastic Calculus and Markov Methods," and in Appendix C of D. Duffie, *Dynamic Asset Pricing Theory*, there exists a random variable $E\left[X\,|\Sigma\right]$, called the conditional expectation of X given Σ, that has the following defining properties:

(i) $E\left[X\,|\Sigma\right]$ is Σ-measurable and integrable,

(ii) $\int_A XdP = \int_A E\left[X\,|\Sigma\right]dP$, for all $A \in \Sigma$.

We want to apply the Projection Theorem to give the existence of this conditional expectation for square integrable random variables. We can then extend this existence result using denseness properties of a particular subset of L^2.

Therefore, assume first that X is square integrable, i. e., $X \in L^2\left(\Omega, \Phi, P\right)$. Let $L^2\left(\Omega, \Sigma, P\right)$ denote the space of square integrable Σ-measurable random variables. As noted above, (the spaces of equivalence classes of random variables) are Hilbert spaces and therefore $Y \in L^2\left(\Omega, \Sigma, P\right)$ is a complete subspace of $L^2\left(\Omega, \Phi, P\right)$. Hence by Theorem 7, there exists an element $Y \in L^2\left(\Omega, \Sigma, P\right)$ such that

(a) $\|X - Y\|^2 = E\left(\left(X - Y\right)^2\right) = \inf\left\{\|x - Z\|^2 : Z \in L^2\left(\Omega, \Sigma, P\right)\right\}$

(b) $\langle X - Y \,|\, Z\rangle = 0,$ all $Z \in L^2\left(\Omega, \Sigma, P\right)$

Since $Y \in L^2\left(\Omega, \Sigma, P\right)$, Y is Σ-measurable and square integrable and hence satisfies condition (i) above. Now if $A \in \Sigma$, then $Z = I_A$ (the indicator function of the set A) belongs to $L^2\left(\Omega, \Sigma, P\right)$ and (b) states that

$$0 = E\left(\left(X - Y\right)Z\right) = \int_A \left(X - Y\right)dP = \int_A XdP - \int_A YdP$$

But this is just condition (ii) above. Thus Y is $E\left[X\,|\Sigma\right]$.

The existence of $E\left[X|\Sigma\right]$ for X such that $E\left(X^2\right)<\infty$ given by the Projection Theorem allows us an interesting interpretation that is not immediately available via other existence arguments. This interpretation is that $E\left[X|\Sigma\right]$ is the least-squares-best Σ-measurable predictor of X: among all predictors V of X which can be computed based on information in Σ, $E\left[X|\Sigma\right]$ minimizes the squared deviation $E\left(\left(X-Y\right)^2\right)$. This interpretation is the property (a) above, which is equivalent to (b) by the Projection Theorem.

See Williams, *Probability with Martingales*, Chapter 9 for motivation of the defining properties (i) and (ii) above for conditional expectation and for the extension of the existence argument to larger sets of payoffs (beyond L^2).