Chapter 2
Technical preliminaries: Some basic notions of Hilbert Spaces

Outline

Recommended readings: Tsiatis, ch 2, and Luenberger, ch 3

Inner product

Let V be a vector space and let x and y denote vectors in V.

An (real) inner product is any function that assigns to any pair of vectors x and y, a scalar denoted by $\langle x, y\rangle$ which satisfies (for any a, b real numbers),

$\langle x, y\rangle = \langle y, x\rangle$,	(commutative)
$\langle ax + bz, y\rangle = a\langle x, y\rangle + b\langle z, y\rangle$,	(linearity)
$\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$.	(positive definitiveness)

A vector space together with an inner product is called an inner product space or pre-Hilbert space.

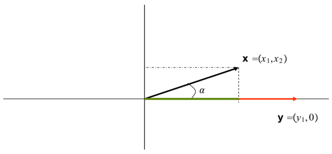
Norm of a vector is $\|x\|^2 = \langle x, x\rangle$, and $\|x\|$ is called the length of x.

Example 2.1. In the n-dimensional Euclidean space

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \ldots + x_ny_n,$$
$$\|x\|^2 = x_1^2 + x_2^2 + \ldots + x_n^2.$$

Cauchy-Schwartz inequality $\langle x, y\rangle^2 \leq \|x\|^2\|y\|^2$.

In the 2-dimensional space,



$$\begin{aligned}
\cos(\alpha) &= \frac{x_1}{\|x\|}\,\mathrm{sg}\,(y_1) \\
&= \frac{x_1}{\|x\|}\frac{y_1}{|y_1|} \\
&= \frac{x_1}{\|x\|}\frac{y_1}{\|y\|} = \frac{x_1 y_1 + x_2 y_2}{\|x\|\|y\|} \\
&= \frac{\langle x, y\rangle}{\|x\|\|y\|}.
\end{aligned}$$

In an arbitrary vector space, we define the cosine of the “angle” between two vectors as

$$\cos(\alpha) = \frac{\langle x, y\rangle}{\|x\|\|y\|}.$$

The right hand side has absolute value ≤ 1 because of the Cauchy-Schwartz inequality.

In the two dimensional Euclidean space, two vectors are orthogonal if they form a straight angle, i.e. if the cosine of their angle is zero, or equivalently, if their inner product is zero.

In an arbitrary inner product space we say that two vectors x and y are orthogonal iff

$$\langle x, y\rangle = 0,$$

and we write $x \perp y$.

Let $x_1, x_2, \ldots$ be a sequence of vectors in a space V, the sequence is said to converge in V if there exists a vector x in V such that

$$\lim_{n\to\infty} \|x_n - x\| = 0,$$

in which case x is called the limit of the sequence.

A subspace M of a pre-Hilbert space V is closed iff the limit of every converging sequence in M is also in M. That is, if $x_1, x_2, \ldots$ is a sequence of vectors in M which converges to x, then x is in M.

A Cauchy sequence is any sequence of vectors $x_1, x_2, \ldots$ which satisfies that for all $\varepsilon > 0$ there exists n_0 such that if $n \geq n_0$ and $m \geq n_0$, then

$$\|x_n - x_m\| \leq \varepsilon.$$

A space V is complete if for every Cauchy sequence $x_1, x_2, \ldots$ whose elements are in V, there exists a vector x in V such that x is the limit of $x_1, x_2, \ldots$, i.e.

$$\lim_{n\to\infty} \|x_n - x\| = 0.$$

A Hilbert space is a complete linear inner product space.

Any Hilbert space V is closed. This follows from the following argument: Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is a sequence of vectors in $\mathbf{V}$ such that $\lim_{n\to\infty} \mathbf{x}_n = \mathbf{x}$. Then, since every converging sequence is a Cauchy sequence, we have that $\mathbf{x}_1, \mathbf{x}_2, \ldots$ is a Cauchy sequence. Finally, since $\mathbf{V}$ is a complete space, then the limit of the Cauchy sequence, i. e. $\mathbf{x}$ must be an element of $\mathbf{V}$.

Another result is that a finite dimensional subspace of a Hilbert space is always closed.

For our purposes we will be interested in the spaces

$$\mathcal{L}_2(\theta) \equiv \left\{ b(\cdot) \text{ real valued: } \int b(x)^2 f(x;\theta)dx < \infty \right\},$$

and

$$\mathcal{L}_2^0(\theta) \equiv \left\{ b(\cdot) \text{ real valued: } \int b(x)^2 f(x;\theta)dx < \infty, \int b(x) f(x;\theta)dx = 0 \right\},$$

where x is a scalar or a vector.

The spaces $\mathcal{L}_2(\theta)$ and $\mathcal{L}_2^0(\theta)$ are Hilbert spaces with inner product given by

$$\langle b_1(X), b_2(X) \rangle_\theta = E_\theta \left\{ b_1(X) b_2(X) \right\}.$$

Also, in $\mathcal{L}_2{}^0(\theta)$

$$\|b(X)\|_\theta^2 = \mathrm{var}_\theta\{b(X)\}.$$

Here and throughout E_θ and var_θ stand for expectation and variance under F_θ.

Some notational remarks

If B denotes an arbitrary set of a Hilbert space H, then $\bar{B}$ denotes the closure of the set B and $[B]$ denotes the linear span of the set B, i. e.

$$\bar{B} = \left\{ v \in H : \exists v_1, v_2, \ldots \in B \text{ such that } \lim_{n\to\infty} v_n = v \right\},$$

and

$$[B] = \left\{ v \in H : \exists v_i \in B \text{ and } a_i \in \mathbb{R}, i = 1, \ldots, k \text{ such that } v = \sum_{i=1}^k a_i v_i \right\}.$$

Thus, in particular, $\overline{[B]}$ denotes the closed linear span of B.

Theorem 2.1 (Pythagorean theorem). Let $v_1, v_2, \ldots, v_k$ be mutually orthogonal vectors in Ω, then

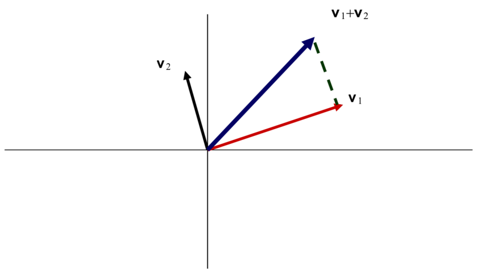
$$\left\| \sum_{i=1}^k v_i \right\|^2 = \sum_{i=1}^k \|v_i\|^2.$$

Proof.

$$\begin{aligned} \left\| \sum_{i=1}^k v_i \right\|^2 &= \left\langle \sum_{i=1}^k v_i, \sum_{j=1}^k v_j \right\rangle \\ &= \sum_{i=1}^k \sum_{j=1}^k \langle v_i, v_j \rangle \\ &= \sum_{i=1}^k \langle v_i, v_i \rangle. \end{aligned}$$

□

In the 2-dimensional space,



$$\|v_1\|^2 + \|v_2\|^2 = \|v_1 + v_2\|^2.$$

Lemma 2.1 (The Projection lemma 1). Let V be a Hilbert space and let M be a closed linear subspace. Then

- Corresponding to any vector x in V, there exists a unique vector m_0 in M such that

$$\|x - m_0\| \leq \|x - m\| \text{ for all } m \text{ in } M. \tag{1}$$

- m_0 satisfies (1) if and only if

$$x - m_0 \perp m \text{ for all } m \text{ in } M.$$

The Projection lemma 1

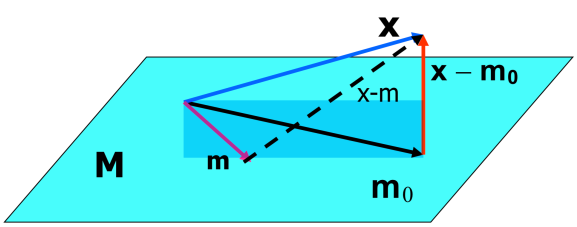


Figure 1: The Projection lemma

Lemma 2.2 (The Projection lemma 2). Suppose that M is a subspace (not necessarily closed) of an inner product space V (not necessarily complete) and let x be a vector in V. Then,

- If there exists m_0 in M such that

$$\|x - m_0\| \leq \|x - m\| \text{ for all } m \text{ in } M, \tag{2}$$

  then m_0 is unique.

- m_0 satisfies (2) if and only if

$$x - m_0 \perp m \quad \text{for all } m \text{ in } M.$$

Notational remarks

- The vector m_0 of the Projection Lemmas, when it exists, is called the projection of the vector x into the space M. Throughout, we will denote it with $\Pi[x \mid M]$.

- If a vector v satisfies

$$V \perp m \text{ for all } m \in M,$$

then we say that m is orthogonal to M and we write

$$V \perp M.$$

Notational remarks

- The collection of all vectors that are orthogonal to a set M is denoted with $M^{\perp}$.

- If M_1 and M_2 are any subsets, then

$$M_1 + M_2 \equiv \{v_1 + v_2 : v_1 \in M_1 \text{ and } v_2 \in M_2\}.$$

- If M_1 and M_2 are orthogonal sets (i.e. any vector of M_1 is orthogonal to any vector of M_2), then we write

$$M_1 \oplus M_2 \text{ instead of } M_1 + M_2.$$

- $M^{\perp}$ is always a closed linear space (i.e. even if M is neither closed nor linear).

- If $v \perp M$ then $\Pi[v \mid M] = 0$.

- If M is a closed linear space of a Hilbert space V then for any v in V there exist unique vectors m in M and $m^{\perp}$ in $M^{\perp}$ such that

$$v = m + m^{\perp}.$$

- Suppose that $M_a, a \in \mathcal{A}$ is a, not necessarily countable, collection of closed linear spaces. Let

$$M \equiv \overline{\left[\bigcup_{a \in \mathcal{A}} M_a\right]}.$$

Suppose that

$$\Pi[v \mid M_a] = 0 \text{ for all } a \in \mathcal{A},$$

then

$$\Pi[v \mid M] = 0.$$

- If M_1 and M_2 are orthogonal closed linear spaces, then

$$\Pi[v \mid M_1 \oplus M_2] = \Pi[v \mid M_1] + \Pi[v \mid M_2].$$

- A very important implication of the Pythagorean Theorem is that projecting a vector "contracts"it or leaves it the same, i.e.

$$\|\Pi[x \mid M]\|^2 \le \|x\|^2,$$
$$\|\Pi[x \mid M]\|^2 = \|x\|^2 \text{ iff } x \in M.$$

Proof.

$$\begin{aligned}\|x\|^2 &= \|\Pi[x \mid M]\|^2 + \|x - \Pi[x \mid M]\|^2 \\ &\ge \|\Pi[x \mid M]\|^2.\end{aligned}$$

Further,

$$\|\Pi[x \mid M]\|^2 = \|x\|^2 \text{ iff } \|x - \Pi[x \mid M]\|^2 = 0,$$

or equivalently iff $x \in M$. □

Suppose that M is a subspace of dimension p, spanned by the vectors $v_1, \dots, v_p$. Then, since $\Pi[x \mid M]$ is in M, it must be that

$$\Pi[x \mid M] = \sum_{i=1}^{p} a_i v_i.$$

We will now calculate the values of $a_1, \dots, a_p$. Since $x - \Pi[x \mid M] \perp v_j$ for $j = 1, \dots, p$, we have

$$\left\langle x - \sum_{i=1}^{p} a_i v_i, v_j \right\rangle = 0 \text{ for } j = 1, \dots, p,$$

or equivalently

$$\sum_{i=1}^{p} a_i \langle v_i, v_j \rangle = \langle x, v_j \rangle \text{ for } j = 1, \dots, p.$$

In matrix form, this is

$$\underbrace{\begin{bmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle & \cdots & \langle v_1, v_p \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle & \cdots & \langle v_2, v_p \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_p, v_1 \rangle & \langle v_p, v_2 \rangle & \cdots & \langle v_p, v_p \rangle \end{bmatrix}}_{\text{This is called the Gram Matrix}} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \langle x, v_1 \rangle \\ \langle x, v_2 \rangle \\ \vdots \\ \langle x, v_p \rangle \end{bmatrix}.$$

This system of linear equations is called the NORMAL EQUATIONS.

We will often apply the previous result in the following setting.

The Hilbert space V will be $\mathcal{L}_2^0(\theta)$ and the vectors $v_1, \dots, v_p$ will be the scores with respect to the components of a parameter vector θ indexing a parametric model. That is,

$$v_j = S_{\theta_j}(\theta); \text{ typically, } S_{\theta_j}(\theta) = \frac{\partial \ln f(X;\theta)}{\partial \theta_j} \quad j = 1, \dots, p,$$

and the subspace M is the space generated by the components of the score vector, throughout denoted by $\Lambda(\theta)$, and called the TANGENT SPACE FOR THE MODEL AT F_θ, i. e.

$$\Lambda(\theta) = \left\{ a^T S_\theta(\theta) : a \in \mathbb{R}^p \right\},$$

where

$$S_\theta(\theta) = \left(S_{\theta_1}(\theta), S_{\theta_2}(\theta), \dots, S_{\theta_p}(\theta) \right)^T.$$

The Gram Matrix in this case is

$$
\begin{aligned}
& \begin{bmatrix} \langle v_1, v_1\rangle & \langle v_1, v_2\rangle & \cdots & \langle v_1, v_p\rangle \\ \langle v_2, v_1\rangle & \langle v_2, v_1\rangle & \cdots & \langle v_2, v_p\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_p, v_1\rangle & \langle v_p, v_2\rangle & \cdots & \langle v_p, v_p\rangle \end{bmatrix} \\
= & \begin{bmatrix} E_\theta\left[S_{\theta_1}(\theta)S_{\theta_1}(\theta)\right] & E_\theta\left[S_{\theta_1}(\theta)S_{\theta_2}(\theta)\right] & \cdots & E_\theta\left[S_{\theta_1}(\theta)S_{\theta_p}(\theta)\right] \\ E_\theta\left[S_{\theta_2}(\theta)S_{\theta_1}(\theta)\right] & E_\theta\left[S_{\theta_2}(\theta)S_{\theta_2}(\theta)\right] & \cdots & E_\theta\left[S_{\theta_2}(\theta)S_{\theta_p}(\theta)\right] \\ \vdots & \vdots & \ddots & \vdots \\ E_\theta\left[S_{\theta_p}(\theta)S_{\theta_1}(\theta)\right] & E_\theta\left[S_{\theta_p}(\theta)S_{\theta_2}(\theta)\right] & \cdots & E_\theta\left[S_{\theta_p}(\theta)S_{\theta_p}(\theta)\right] \end{bmatrix} \\
= & E_\theta\left[S_\theta(\theta)S_\theta(\theta)^T\right] \\
= & I(\theta).
\end{aligned}
$$

Then, if the information matrix $I(\theta)$ is non-singular, the projection of any real valued function $b(X)$ of X in $\mathcal{L}_2^0(\theta)$ into the tangent space is

$$\Pi_\theta[b(X) \mid \Lambda(\theta)] = a^T S_\theta(\theta),$$

where

$$
\begin{aligned}
a^T &= \left(E_\theta\left[b(X)S_{\theta_1}(\theta)\right], E_\theta\left[b(X)S_{\theta_2}(\theta)\right], \ldots, E_\theta\left[b(X)S_{\theta_p}(\theta)\right]\right) I(\theta)^{-1} \\
&= E_\theta\left[b(X)S_\theta(\theta)^T\right] I(\theta)^{-1}.
\end{aligned}
$$

Conclusion: if $\Lambda(\theta)$ is the tangent space in a parametric model with score vector $S_\theta(\theta)$, then

$$\Pi_\theta[b(X) \mid \Lambda(\theta)] = E_\theta\left[b(X)S_\theta(\theta)^T\right] I(\theta)^{-1} S_\theta(\theta).$$

Remembering that $\|\Pi[x \mid M]\|^2 \leq \|x\|^2$ and that in $\mathcal{L}_2^0(\theta)$,

$$\|b(X)\|^2 = \mathrm{var}_\theta\{b(X)\},$$

and noticing that

$$
\begin{aligned}
& \mathrm{var}_\theta\left\{E_\theta\left[b(X)S_\theta(\theta)^T\right] I(\theta)^{-1} S_\theta(\theta)\right\} \\
= & E_\theta\left[b(X)S_\theta(\theta)^T\right] I(\theta)^{-1} E_\theta\left[b(X)^T S_\theta(\theta)\right],
\end{aligned}
$$

we conclude that for any $b(X) \in \mathcal{L}_2^0(\theta)$,

$$\mathrm{var}_\theta\{b(X)\} \geq E_\theta\left[b(X)S_\theta(\theta)^T\right] I(\theta)^{-1} E_\theta\left[S_\theta(\theta)b(X)\right].$$

We can also deduce the following useful result. Suppose that $b_1(X), \ldots, b_k(X)$ are k vectors in $\mathcal{L}_2^0(\theta)$. Define

$$\underline{b}(X) = \left(b_1(X), \ldots, b_k(X)\right)^T,$$

and in a slight abuse of notation define,

$$\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)] = \begin{bmatrix} \Pi_\theta\left[b_1(X) \mid \Lambda(\theta)\right] \\ \Pi_\theta\left[b_2(X) \mid \Lambda(\theta)\right] \\ \vdots \\ \Pi_\theta\left[b_k(X) \mid \Lambda(\theta)\right] \end{bmatrix},$$

and we have that

$$\underline{b}(X) - \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)] = \begin{bmatrix} b_1(X) - \Pi_\theta\left[b_1(X) \mid \Lambda(\theta)\right] \\ b_2(X) - \Pi_\theta\left[b_2(X) \mid \Lambda(\theta)\right] \\ \vdots \\ b_p(X) - \Pi_\theta\left[b_k(X) \mid \Lambda(\theta)\right] \end{bmatrix}.$$

So,

$$E\left\{\left(\underline{b}(X) - \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right)\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]^T\right\} = \underbrace{0}_{k\times k},$$

and

$$\begin{aligned}\underbrace{\mathrm{var}_\theta(\underline{b}(X))}_{k\times k} &= \mathrm{var}_\theta\left\{\left(\underline{b}(X) - \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right) + \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right\} \\ &= \mathrm{var}_\theta\left\{\left(\underline{b}(X) - \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right)\right\} + \mathrm{var}_\theta\left\{\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right\} \\ &+ E\left\{\left(\underline{b}(X) - \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right)\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]^T\right\} \\ &+ E\left\{\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\left(\underline{b}(X) - \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right)^T\right\} \\ &= \mathrm{var}_\theta\left\{\left(\underline{b}(X) - \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right)\right\} + \mathrm{var}_\theta\left\{\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right\}.\end{aligned}$$

Therefore,

$$\mathrm{var}_\theta(\underline{b}(X)) - \mathrm{var}_\theta\left\{\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right\} = \mathrm{var}_\theta\left\{\left(\underline{b}(X) - \Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right)\right\}.$$

Hence, since any variance-covariance matrix is positive semidefinite, we obtain that

$$\mathrm{var}_\theta(\underline{b}(X)) \geq \mathrm{var}_\theta\left\{\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right\},$$

and throughout, if A and B are squared conformable matrices then, $A \geq B$ designates that $A - B$ is positive semidefinite, i.e. for all conformable vectors v,

$$v^T(A-B)v \geq 0.$$

Finally, remembering that

$$\begin{aligned}\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)] &= \begin{bmatrix} \Pi_\theta\left[b_1(X) \mid \Lambda(\theta)\right] \\ \Pi_\theta\left[b_2(X) \mid \Lambda(\theta)\right] \\ \vdots \\ \Pi_\theta\left[b_k(X) \mid \Lambda(\theta)\right] \end{bmatrix} \\ &= \begin{bmatrix} E_\theta\left[b_1(X)S_\theta(\theta)^T\right]I(\theta)^{-1}S_\theta(\theta) \\ E_\theta\left[b_2(X)S_\theta(\theta)^T\right]I(\theta)^{-1}S_\theta(\theta) \\ \vdots \\ E_\theta\left[b_k(X)S_\theta(\theta)^T\right]I(\theta)^{-1}S_\theta(\theta) \end{bmatrix} \\ &= E_\theta\left[\underline{b}(X)S_\theta(\theta)^T\right]I(\theta)^{-1}S_\theta(\theta),\end{aligned}$$

we have that

$$\mathrm{var}_\theta\left\{\Pi_\theta[\underline{b}(X) \mid \Lambda(\theta)]\right\} = E_\theta\left[\underline{b}(X)S_\theta(\theta)^T\right] I(\theta)^{-1} E_\theta\left[S_\theta(\theta)\underline{b}(X)^T\right],$$

and we conclude that

$$\mathrm{var}_\theta\{\underline{b}(X)\} \geq E_\theta\left[\underline{b}(X)S_\theta(\theta)^T\right] I(\theta)^{-1} E_\theta\left[S_\theta(\theta)\underline{b}(X)^T\right].$$