# Contents



# 1 Hilbert Space

## 1.1 Definition

We will focus primarily on the Hilbert space whose elements are random vectors with mean zero and finite variance. For random variable $Z$, it is equipped with an underlying probability space $(\mathscr{Z}, \mathcal{A}, P)$, where $\mathscr{Z}$ is the sample space, $\mathcal{A}$ is the corresponding $\sigma$-algebra, and $P$ is the probability measure.
Consider the space consisting of $q$-dimensional mean-zero random functions of $Z$,

$$h : \mathscr{Z} \to \mathbb{R}^q$$

where $h(Z)$ is measurable and also satisfies

$$\text{(i) } E\{h(Z)\} = 0$$
$$\text{(ii) } E\left\{h^T(Z)h(Z)\right\} < \infty$$

In the same way that we consider points in Euclidean space as vectors from the origin, here we will consider the q-dimensional random functions as points in a space. Therefore, we define the origin of this space as

$$h(Z) = 0^{q\times 1}$$

A Hilbert space, denoted by $\mathcal{H}$, is a complete normed *linear vector space* equipped with an *inner product*. As well as being a linear space, a Hilbert space also allows us to consider distance between elements and angles and orthogonality between vectors in the space. This is accomplished by defining an inner product.

**Definition 1.1.** ***(Inner product)***
*Corresponding to each pair of elements $h_1$, $h_2$ belonging to a linear vector space $\mathcal{H}$, an inner product, defined by $\langle h_1, h_2\rangle$, is a function that maps to the real line. That is, $\langle h_1, h_2\rangle$ is a scalar that satisfies*

1. $\langle h_1, h_2\rangle = \langle h_2, h_1\rangle$
2. $\langle h_1 + h_2, h_3\rangle = \langle h_1, h_3\rangle + \langle h_2, h_3\rangle$, *with* $h_1, h_2, h_3 \in \mathcal{H}$
3. $\langle \lambda h_1, h_2\rangle = \lambda \langle h_1, h_2\rangle$
4. $\langle h_1, h_1\rangle \geq 0$ *with equality if and only if* $h_1 = 0$

**Definition 1.2.** ***(Inner product for $q$-dimensional measurable random functions)***

$$\langle h_1, h_2\rangle = E\left(h_1^T h_2\right)$$

*Definitions that follows*

1. *equivalence:* $h_1 \equiv h_2$ *if* $h_1 = h_2$ *a.e.*
2. *norm/length:* $\|h\| = \langle h, h\rangle^{1/2}$
3. *orthogonality:* $h_1 \perp h_2$ *if* $\langle h_1, h_2\rangle = 0$

**Definition 1.3.** ***(linear subspace)***
*A space $\mathcal{U} \subset \mathcal{H}$ is a linear subspace if $u_1, u_2 \in \mathcal{U}$ implies $au_1 + bu_2 \in \mathcal{U}$, $\forall a, b \in R$. A linear subspace must contain the origin. This is clear by letting the scalars be $a = b = 0$*

**Theorem 1.4.** ***(Cauchy-Schwartz inequality)***
*$\forall\, h_1, h_2 \in \mathcal{H}$*

$$|\langle h_1, h_2 \rangle|^2 \le \|h_1\|^2 \, \|h_2\|^2$$

*with equality holds if and only if $h_1 = ch_2$ for some $c \in R$*

## 1.2 Projection theory

**Theorem 1.5.** ***(projection of $h$ onto $\mathcal{U}$)***
*Let $\mathcal{H}$ be a Hilbert space and $\mathcal{U}$ a linear subspace that is closed (i.e., contains all its limit points). Corresponding to any $h \in \mathcal{H}$, , there exists a unique $u_0 \in \mathcal{U}$ that is closest (in terms of the distance defined above) to $h$; that is*

$$\|h - u_0\| \le \|h - u\|, \quad \forall u \in \mathcal{U},$$

*in addition, $(h - u_0) \perp \mathcal{U}$, that is*

$$\langle h - u_0, u \rangle = 0, \quad \forall u \in \mathcal{U}.$$

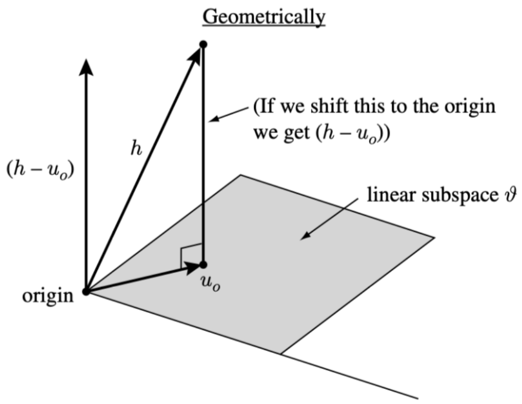


Figure 1: Projection onto a linear subspace

**Theorem 1.6.** ***(Pythagorean Theorem)***
*if $h_1, h_2 \in \mathcal{H}$ and $h_1 \perp h_2$ then*

$$\|h_1 + h_2\|^2 = \|h_1\|^2 + \|h_2\|^2$$

**Example 1.** ***(One-Dimensional Random Functions)***
*Consider the Hilbert space $\mathcal{H}$ of one-dimensional random functions, $h(Z)$, with mean zero and finite variance. Let $u_1(Z), \dots, u_k(Z) \in \mathcal{H}$, then $\mathcal{U}$ is the linear subspace spanned by $\{u_1, \cdots, u_k\}$.*

$$\mathcal{U} = \left\{a^T u; \quad \textit{for} \quad a \in \mathbb{R}^k\right\}, \quad u^{k\times 1} = \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix}$$

*$u$ is the basis.*
*To derive the projection of $\forall h \in \mathcal{H}$, let's go through the following steps*

$$\begin{aligned}
&\left\langle h - a_0^T u, a^T u \right\rangle = 0, \quad \forall a = (a_1, \dots, a_k)^T \in \mathbb{R}^k \\
&E\left(hu^T\right) - a_0^T E\left(uu^T\right) = 0^{(1\times k)} \\
&a_0^T = E\left(hu^T\right)\left\{E\left(uu^T\right)\right\}^{-1} \\
&u_0 = a_0^T u = E\left(hu^T\right)\left\{E\left(uu^T\right)\right\}^{-1} u.
\end{aligned}$$

*The norm-squared of this projection is equal to*

$$E\left(hu^T\right)\left\{E\left(uu^T\right)\right\}^{-1} E(uh^T)$$

*since*

$$\left\|a_0^\top \mu\right\|^2 = E\left[E\left(hu^T\right)\left\{E\left(uu^T\right)\right\}^{-1} uu^T \left\{E\left(uu^T\right)\right\}^{-1} E(uh^T)\right]$$
$$= E\left(hu^T\right)\left\{E\left(uu^T\right)\right\}^{-1} E(uh^T)$$

# 2 The Geometry of Influence Functions

**Problem setup**: $Z_1, \ldots, Z_n$ are i.i.d random vectors with the density belongs to $\{p_Z(z;\theta), \theta \epsilon \Omega\}$. $\theta = \left(\beta^T, \eta^T\right)^T$, where $\beta^{q\times 1}$ is the parameterof interest and $\eta$ is the nuisance parameter, may be finite- or infinite-dimensional. An estimator $\hat{\beta}_n$ of $\beta$ is a $q$-dimensional measurable random function of $Z_1, \ldots, Z_n$. Most reasonable estimators for $\beta$ are *asymptotically linear*; that is, there exists a random vector (i.e., a $q$-dimensional measurable random function) $\varphi^{q\times 1}(Z)$ s.t. $E\{\varphi(Z)\} = 0^{q\times 1}$,

$$n^{1/2}\left(\hat{\beta}_n - \beta_0\right) = n^{-1/2}\sum_{i=1}^{n} \varphi\left(Z_i\right) + o_p(1).$$

It follows that by the CLT

$$n^{-1/2}\sum_{i=1}^{n} \varphi\left(Z_i\right) \xrightarrow{\mathcal{D}} N\left(0^{q\times 1}, E\left(\varphi\varphi^T\right)\right),$$

and

$$n^{1/2}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{\mathcal{D}} N\left(0, E\left(\varphi\varphi^T\right)\right).$$

**Remark 1.** *The function $\varphi(Z)$ is defined with respect to the true distribution $p\left(z, \theta_0\right)$. Consequently, we sometimes may write $\varphi(Z, \theta)$. Therefore,*

$$E_{\theta_0}\left\{\varphi\left(Z, \theta_0\right)\right\}$$

The class of influence functions for estimators belongs to the Hilbert space of all mean-zero $q$-dimensional random functions with finite variance.

**Theorem 2.1.** ***(Uniqueness of influence function)***
*An asymptotically linear estimator has a unique (a.s.) influence function.*

**Example 2.** *For $Z_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, \cdot, n$, $\hat{\mu}_n = n^{-1}\sum_{i=1}^n Z_i$, and $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \left(Z_i - \hat{\mu}_n\right)^2$.*

$$n^{1/2}\left(\hat{\mu}_n - \mu_0\right) = n^{-1/2}\sum_{i=1}^{n}\left(Z_i - \mu_0\right)$$
$$n^{1/2}\left(\hat{\sigma}_n^2 - \sigma_0^2\right) = n^{-1/2}\sum\left\{\left(Z_i - \mu_0\right)^2 - \sigma_0^2\right\} + n^{1/2}\left(\hat{\mu}_n - \mu_0\right)^2.$$

**Remark 2.** ***(Asymptotically efficient)***
*We know that the variance of any unbiased estimator must be greater than or equal to the Cràmer-Rao lower bound; see, for example, Casella and Berger (2002, Section 7.3). When considering asymptotic theory, where we let the sample size $n$ go to infinity, most reasonable estimators are asymptotically unbiased. Thus, we might expect the asymptotic variance of such asymptotically unbiased estimators also to be greater than the Cràmer-Rao lower bound. This indeed is the case for the most part, and estimators whose asymptotic variance equals the Cràmer-Rao lower bound are referred to as asymptotically efficient.*

## 2.1 Super-Efficiency

**Example 3.** ***(Hodges's example)***
*XXX*

**Definition 2.2.** ***(Regular)***
*Consider a local data generating process (LDGP), where, for each $n$, the data are distributed according to $\theta_n$, where $n^{1/2}\left(\theta_n - \theta^*\right) \to c$, for some $c \in R$. That is*

$$Z_{1n}, Z_{2n}, \ldots, Z_{nn} \sim_{\text{i.i.d}} p\left(z, \theta_n\right),$$

*where $\theta_n = \left(\beta_n^T, \eta_n^T\right)^T$, $\quad \theta^* = \left({\beta^*}^T, {\eta^*}^T\right)^T$. An estimator $\hat{\beta}_n(Z_{1n}, \ldots, Z_{nn})$, is said to be regular if, for each $\theta^*$, $n^{1/2}\left(\hat{\beta}_n - \beta_n\right)$ has a limiting distribution that does not depend on the LDGP.*

**Definition 2.3.** ***(Score vector)***
*For* $Z \sim p_Z(z, \theta), \theta = \left(\beta^T, \eta^T\right)^T$,

$$
\begin{aligned}
S_\theta(z, \theta_0) &= \left.\frac{\partial \log p_Z(z,\theta)}{\partial \theta}\right|_{\theta=\theta_0} \\
S_\theta(Z, \theta_0) &= \left\{S_\beta^T(Z, \theta_0), S_\eta^T(Z, \theta_0)\right\}^T \\
&\text{where} \\
S_\beta(z, \theta_0) &= \left.\frac{\partial \log p_Z(z,\theta)}{\partial \beta}\right|_{\theta=\theta_0}^{q\times 1} \\
S_\eta(z, \theta_0) &= \left.\frac{\partial \log p_Z(z,\theta)}{\partial \eta}\right|_{\theta=\theta_0}^{r\times 1}.
\end{aligned}
$$

**Theorem 2.4.** *Let the parameter of interest* $\beta(\theta)$ *be a* $q$*-dimensional function of the* $p$*-dimensional parameter* $\theta$*,* $q < p$*, such that* $\Gamma^{q\times p}(\theta) = \partial\beta(\theta)/\partial\theta^T$*, the* $q \times p$*-dimensional matrix of partial derivatives, exists, has rank* $q$*, and is continuous in* $\theta$ *in a neighborhood of the truth* $\theta_0$*. Also let* $\hat{\beta}_n$ *nbe an asymptotically linear estimator with influence function* $\varphi(Z)$ *such that* $E_\theta\left(\varphi^T\varphi\right)$ *exists and is continuous in* $\theta$ *in a neighborhood of* $\theta_0$*. Then if* $\hat{\beta}_n$ *is regular, this will imply that*

$$
E\left\{\varphi(Z) S_\theta^T(Z, \theta_0)\right\} = \Gamma(\theta_0).
$$

*In the special case where* $\theta$ *can be partitioned as* $\left(\beta^T, \eta^T\right)^T$*, we have*

$$
\begin{aligned}
E\left\{\varphi(Z) S_\beta^T(Z, \theta_0)\right\} &= I^{q\times q} \\
E\left\{\varphi(Z) S_\eta^T(Z, \theta_0)\right\} &= 0^{q\times r},
\end{aligned}
$$

*where* $I^{q\times q}$ *is the* $q \times q$ *identity matrix and* $0^{q\times r}$ *is the* $q \times r$ *matrix of zeros.*

## 2.2 Geometry of Influence Functions for Parametric Models

Consider the Hilbert space $\mathcal{H}$ of all $q$-dimensional measurable functions of $Z$ with mean zero and finite variance equipped with the inner product $\langle h_1, h_2\rangle = E\left(h_1^T h_2\right)$.

**Definition 2.5.** ***(Tangent space)***
*We first note that the score vector* $S_\theta(Z, \theta_0)$*, under suitable regularity conditions, has mean zero (*$E\left\{S_\theta(Z, \theta_0)\right\} = 0^{p\times 1}$*). We can define the finite-dimensional linear subspace* $\mathscr{T} \subset \mathcal{H}$ *spanned by the* $p$*-dimensional score vector* $S_\theta(Z, \theta_0)$ *as the set of all q-dimensional mean-zero random vectors consisting of*

$$
B^{q\times p} S_\theta(Z, \theta_0), \quad \forall B^{q\times p}.
$$

*The linear subspace* $\mathscr{T}$ *is referred to as the tangent space.*

**Definition 2.6.** ***(Nuisance tangent space)***
*In the case where* $\theta = \left(\beta^T, \eta^T\right)^T$*, consider the linear subspace spanned by the nuisance score vector* $S_\eta(Z, \theta_0)$*,*

$$
B^{q\times r} S_\eta(Z, \theta_0), \quad \forall B^{q\times r}.
$$

*This space is referred to as the nuisance tangent space and will be denoted by* $\Lambda$*.*

Note that Theorem 2.4 states that the $q$-dimensional influence function $\varphi_{\hat{\beta}_n}(Z)$ for $\hat{\beta}_n$ is orthogonal to the nuisance tangent space $\Lambda$.

## 2.3 Constructing Estimators

Influence functions of RAL estimators for $\beta$ must satisfy conditions of Theorem 2.4, , a natural question is whether the converse is true; that is, for any element of the Hilbert space satisfying conditions of Theorem 2.4, does there exist an RAL estimator for $\beta$ with that influence function? Let $\varphi(Z)$ be a $q$-dimensional measurable function with zero mean and finite variance that satisfies conditions of Theorem 2.4. Define

$$
m(Z, \beta, \eta) = \varphi(Z) - E_{\beta,\eta}\{\varphi(Z)\}.
$$

Assume that we can find a root-$n$ consistent estimator for the nuisance parameter $\hat{\eta}_n$ ($n^{1/2}(\hat{\eta}_n - \eta_0)$ is bounded in probability). In many cases the estimator $\hat{\eta}_n$ will be $\beta$-dependent ($\hat{\eta}_n(\beta)$). For example, we might use the MLE for $\eta$, or the restricted MLE for $\eta$, fixing the value of $\beta$. We argue that the solution to the equation

$$
\sum_{i=1}^{n} m\left\{Z_i, \beta, \hat{\eta}_n(\beta)\right\} = 0
$$

which we denote by $\hat{\beta}_n$ will be an asymptotically linear estimator with influence function $\varphi(Z)$.

*Proof.* By construction, we have

$$E_{\beta_0,\eta}\left\{m\left(Z,\beta_0,\eta\right)\right\}=0,$$

or equivalently,

$$\int m\left(z,\beta_0,\eta\right)p\left(z,\beta_0,\eta\right)d\nu(z)=0.$$

Consequently,

$$\begin{aligned}&\frac{\partial}{\partial\eta^T}\bigg|_{\eta=\eta_0}\int m\left(z,\beta_0,\eta\right)p\left(z,\beta_0,\eta\right)d\nu(z)\\&=\int\frac{\partial m\left(z,\beta_0,\eta_0\right)}{\partial\eta^T}p\left(z,\beta_0,\eta_0\right)d\nu(z)+\int m\left(z,\beta_0,\eta_0\right)S_\eta^T\left(z,\beta_0,\eta_0\right)p\left(z,\beta_0,\eta_0\right)d\nu(z)=0.\end{aligned}$$

By definition, $E\left\{\varphi(Z)S_\eta^T\left(Z,\theta_0\right)\right\}=0$. Consequently,

$$E\left\{\frac{\partial}{\partial\eta^T}m\left(Z,\beta_0,\eta_0\right)\right\}=0.$$

Similarly, we can show that

$$E\left\{\frac{\partial}{\partial\beta^T}m\left(Z,\beta_0,\eta_0\right)\right\}=-I^{q\times q}.$$

A standard expansion yields,

$$\begin{aligned}0&=\sum_{i=1}^n m\left\{Z_i,\hat\beta_n,\hat\eta_n\left(\hat\beta_n\right)\right\}\\&=\sum_{i=1}^n m\left\{Z_i,\beta_0,\hat\eta_n\left(\hat\beta_n\right)\right\}\\&+\left[\sum_{i=1}^n\frac{\partial m}{\partial\beta^T}\left\{Z_i,\beta_n^*,\underbrace{\hat\eta_n\left(\hat\beta_n\right)}_{\text{Notice that this term is held fixed}}\right\}\right]\left(\hat\beta_n-\beta_0\right),\quad\beta_n^*\in(\hat\beta_n,\ \beta_0).\end{aligned}$$

Therefore,

$$\begin{aligned}&n^{1/2}\left(\hat\beta_n-\beta_0\right)\\&=-\underbrace{\left[n^{-1}\sum_{i=1}^n\frac{\partial}{\partial\beta^T}m\left\{Z_i,\beta_n^*,\hat\eta_n\left(\hat\beta_n\right)\right\}\right]^{-1}}_{\Downarrow}\left[n^{-1/2}\sum_{i=1}^n m\left\{Z_i,\beta_0,\hat\eta_n\left(\hat\beta_n\right)\right\}\right]\\&\left[E\left\{\frac{\partial}{\partial\beta^T}m\left(Z,\beta_0,\eta_0\right)\right\}\right]^{-1}=-I^{q\times q}\end{aligned}$$

For the second term,

$$\begin{aligned}&n^{-1/2}\sum_{i=1}^n m\left\{Z_i,\beta_0,\hat\eta_n\left(\hat\beta_n\right)\right\}\\&=n^{-1/2}\sum_{i=1}^n m\left(Z_i,\beta_0,\eta_0\right)+\underbrace{\left\{n^{-1}\sum_{i=1}^n\frac{\partial m\left(Z_i,\beta_0,\eta_n^*\right)}{\partial\eta^T}\right\}}_{\Rightarrow_p E\left\{\frac{\partial}{\partial\eta^T}m(Z,\beta_0,\eta_0)\right\}=0}\underbrace{\left[n^{1/2}\left\{\hat\eta_n\left(\hat\beta_n\right)-\eta_0\right\}\right]}_{\text{bounded in probability}}.\end{aligned}$$

Therefore,

$$\begin{aligned}n^{1/2}\left(\hat\beta_n-\beta_0\right)&=n^{-1/2}\sum_{i=1}^n m\left(Z_i,\beta_0,\eta_0\right)+o_p(1)\\&=n^{-1/2}\sum_{i=1}^n\varphi\left(Z_i\right)+o_p(1)\end{aligned}$$

□