ME(EE) 550

Foundations of Engineering Systems Analysis

Chapter Four: Inner Product Spaces

The concept of normed vector spaces, presented in Chapter 3, synergistically combines the topological structure of metric spaces and the algebraic structure of vector spaces. An inner product spaces are a special case of a normed space, where every inner product induces a unique norm; and the resulting geometry is largely similar to the familiar Euclidean spaces. Many familiar notions (e.g., root mean square value, standard deviation, least squares estimation, and orthogonal functions) in engineering analysis can be explained from the perspectives of inner product spaces. For example, the inner product of two vectors is a generalization of the familiar dot product in vector calculus. This chapter should be read along with Chapter 5 Parts A and B of Naylor & Sell. Specifically, both solved examples and exercises in Naylor & Sell are very useful.

1 Basic Concepts

Definition 1.1. (Inner Product) Let $(V, \oplus)$ be a vector space over a (complete) field $\mathbb{F}$, where we choose $\mathbb{F}$ to be $\mathbb{R}$ or $\mathbb{C}$. Then, a function $\langle \bullet, \bullet \rangle : V \times V \to \mathbb{F}$ is defined to be an inner product if the following conditions hold $\forall x, y, z \in V \ \ \forall \alpha \in \mathbb{F}$:

- *Additivity*: $\langle (x \oplus y), z \rangle = \langle x, z \rangle + \langle y, z \rangle$.
- *Homogeneity*: $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$.
- *Hermiticity*: $\langle y, x \rangle = \overline{\langle x, y \rangle}$.
- *Positive Definiteness*: $\langle x, x \rangle > 0$ if $x \neq \mathbf{0}_V$.

A vector space with an inner product is called an inner product space.

Example 1.1. Examples of Inner Product Spaces:

1. Let $V = \mathbb{F}^n$ and $x, y \in V$. Then, $\sum_{k=1}^{n} \bar{y}_k x_k$ is an inner product, where the vector $x \equiv [x_1, x_2, \cdots, x_n]$.
2. Let $V = \mathbb{F}^{n\times m}$ and $A, B \in V$. Then, $trace(B^H A)$ is an inner product.
3. Let $V = \ell_2(\mathbb{F})$ and $x, y \in V$. Then, $\sum_{k=1}^{\infty} \bar{y}_k \; x_k$ is an inner product.
4. Let $V = L_2(\mu)$ and $x, y \in V$. Then, $\int d\mu(t) \; \bar{y}(t) \; x(t)$ is an inner product.

Remark 1.1. Let V be a vector space over a (complete) field $\mathbb{F}$ and $x \in V$. Then $\langle x, x\rangle \in \mathbb{R}$, because it follows from the homogeneity property in Definition 1.1 that $\langle x, x\rangle = \overline{\langle x, x\rangle}$. Furthermore, it follows from the additivity and positive definiteness properties in Definition 1.1 that $\langle x, x\rangle \in [0, \infty) \; \forall x \in V$.

Remark 1.2. The following results are derived from Definition 1.1.

- $\langle y, \alpha x\rangle = \bar{\alpha} \; \langle y, x\rangle$.
- $\langle x, (y \oplus z)\rangle = \langle x, y\rangle + \langle x, z\rangle$.
- If $\langle x, x\rangle = 0$, then then $x = \mathbf{0}_V$.
- If $\langle x, y\rangle = 0 \; \forall y \in V$, then $x = \mathbf{0}_V$.

Lemma 1.1. *(Cauchy-Schwarz Inequality) Let V be a vector space over a (complete) field $\mathbb{F}$. Then, the following inequality holds: $|\langle x, y\rangle| \leq \|x\|\|y\| \quad \forall x, y \in V$, where $\|x\| \triangleq \sqrt{\langle x, x\rangle}$. The equality holds if and only if either $x = \alpha y$ or $y = \alpha x$ for some $\alpha \in \mathbb{F}$.*

Proof. Let $x, y \in V$ and $\alpha = \frac{\langle x,y\rangle}{\langle y,y\rangle}$. Then, it follows that

$$0 \leq \langle (x - \alpha y), (x - \alpha y)\rangle = \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2}$$

which completes the proof. □

Corollary 1.1. *The equality part in Lemma 1.1 holds if and only if x and y are collinear, i.e., if either $x = \alpha y$ or $y = \alpha x$ for some $\alpha \in \mathbb{F}$.*

Proof. The proof follows by substituting x or y in the inequality in Lemma 1.1. □

Remark 1.3. Note that $\|x \oplus y\| = \|x\| + \|y\|$ or $\|x \oplus y\| = |\|x\| - \|y\||$ if and only if x and y are collinear.

Lemma 1.2. *Let V be a vector space and $x \in V$. Then $\|x\| \triangleq \sqrt{\langle x, x\rangle}$ is a valid norm on the vector space V.*

Proof. It directly follows from Definition 1.1 that $\|x\|$ satisfies the strict positivity and homogeneity properties of a norm. What remains to be shown is the triangular inequality property. By choosing arbitrary vectors $x, y \in V$ and applying the Cauchy-Schwarz inequality in Lemma 1.1, it follows that

$$\|x \oplus y\|^2 = \langle (x \oplus y), (x \oplus y)\rangle \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = \big(\|x\| + \|y\|\big)^2$$

□

Lemma 1.3. *(Continuity of inner products) Let $(V, \langle \bullet, \bullet\rangle)$ be an inner product space over a (complete) field $\mathbb{F}$. Let $\{x^k\}$ and $\{y^k\}$ be two Cauchy sequences of vectors in V, converging to $x \in V$ and $y \in V$, respectively. Then, the sequence of inner products $\{\langle x^k, y^k\rangle\}$ converge to the inner product $\langle x, y\rangle$.*

Proof. Subtracting and adding a term, using triangle inequality on scalars, and using the Cauchy-Schwarz inequality and the continuity property of a norm, it follows that

$$\begin{aligned}
|\langle x^k, y^k\rangle - \langle x, y\rangle| \qquad &= |\langle x^k, y^k\rangle - \langle x^k, y\rangle + \langle x^k, y\rangle - \langle x, y\rangle| \\
&\leq |\langle x^k, (y^k \oplus (-y))\rangle| + |\langle (x^k \oplus (-x)), y\rangle| \\
&\leq \|x^k\|\|(y^k \oplus (-y))\| + \|(x^k \oplus (-x))\|\|y^k\| \\
&\to 0
\end{aligned}$$

because $\|x^k\| < \infty$, $\|y^k\| < \infty$, $\|(y^k \oplus (-y))\| \to 0$, and $\|(x^k \oplus (-x))\| \to 0$ as $k \to \infty$. □

Remark 1.4. Every inner product space is a normed space but every normed space is not an inner product space. For example, the normed space ℓ_2 is an inner product space but the normed space ℓ_p for $p \neq 2$ is not an inner product space. Given a normed space, how do we find out whether it qualifies as an inner product space? This issue is addressed in the following theorem.

Theorem 1.1. *(Parallelogram Equality) Let $(V, \langle \bullet, \bullet\rangle)$ be an inner product space. Let $\| \bullet \|$ be the norm induced by the inner product $\langle \bullet, \bullet\rangle$. Then, for all $x, y \in V$, the following equality, known as the parallelogram equality holds:*

$$\|x \oplus y\|^2 + \|x \oplus (-y)\|^2 = 2\|x\|^2 + 2\|y\|^2$$

Proof. The proof follows by substituting $\langle x, x\rangle$ for $\|x\|^2$ and performing algebraic operations. $\square$

It is also a fact that the converse of Theorem 1.1 is true. That is, if V is a normed vector space and satisfies the parallelogram equality, then there is a unique inner product defined on the V, which generates the norm. In other words, V is an inner product space if and only if the norm satisfies the parallelogram equality. The issue is how to find the inner product from the knowledge of the norm. This issue is addressed in the following theorem.

Theorem 1.2. *(Polarization Identity) Let $(V, \|\bullet\|)$ be a normed vector space that satisfies the parallelogram equality. Then, the (unique) inner product $\langle \bullet, \bullet\rangle$ on V that satisfies the the relation $\langle x, x\rangle = \|x\|^2 \quad \forall x \in V$ is given as follows:*

1. *If the scalar field $\mathbb{F}$ over which the vector space is constructed is the real field $\mathbb{R}$, then the inner product is constructed from the given norm as follows:*
$$\langle x, y\rangle = \frac{1}{4}\Big(\|x \oplus y\|^2 - \|x \oplus (-y)\|^2\Big)$$
2. *If the scalar field $\mathbb{F}$ over which the vector space is constructed is the complex field $\mathbb{C}$, then the inner product is constructed from the given norm as follows:*
$$\langle x, y\rangle = \frac{1}{4}\Big(\|x \oplus y\|^2 - \|x \oplus (-y)\|^2 + i\|x \oplus iy\|^2 - i\|x \oplus i(-y)\|^2\Big)$$
where $i \triangleq \sqrt{-1}$.

Proof. The proof follows by substituting $\langle x, x\rangle$ for $\|x\|^2$ and performing algebraic operations. $\square$

Definition 1.2. A complete inner product space is called a Hilbert space.

Remark 1.5. Every inner product space $(V, \langle \bullet, \bullet\rangle$ induces a unique norm in the vector space V defined over the field $\mathbb{F}$. Since a norm is a valid metric in a vector space, every inner product space is a metric space. Therefore, the topology of an inner product space is the one generated by the metric: $d(x, y) \triangleq \|x \oplus (-y)\| = \sqrt{\langle (x \oplus (-y)), (x \oplus (-y))\rangle}$; for example, this is the usual topology in the Euclidean space $\mathbb{R}^n$ or the unitary space $\mathbb{C}^n$. It is also seen in Chapter 1 that every metric space has an (essentially) unique completion; therefore, every inner product space has an (essentially) unique completion that leads to the construction of a unique Hilbert space from the given inner product space. In the next lemma, the inner product $\langle \bullet, \bullet\rangle$ is viewed as a (continuous) mapping of the product space $(V, \langle \bullet, \bullet\rangle) \times (V, \langle \bullet, \bullet\rangle)$ into the scalar field $\mathbb{F}$.

1.1 Orthogonality in Inner Product Spaces

The Hilbert space is a close cousin of the Euclidean space in the sense that the geometrical notions in these spaces are similar. The rationale for this similarity is largely due to the concept of orthogonality that is admissible in inner product spaces.

Definition 1.3. (Orthogonality) Let $(V, \langle \bullet, \bullet \rangle)$ be an inner product space. Then, two vectors $x, y \in V$ are said to orthogonal (to each other) if $\langle x, y \rangle = 0$. Orthogonality of x and y is denoted as $x \perp y$. Two subsets (not necessarily subspaces) $A, B \subseteq V$ are said to be orthogonal (to each other) if $\langle x, y \rangle = 0$ for all $x \in A$ and all $y \in B$; this is denoted as $A \perp B$.

Lemma 1.4. *(Pythagorean Theorem) Let $(V, \langle \bullet, \bullet \rangle)$ be an inner product space and let $x, y \in V$. If $x \perp y$, then $\|x \oplus y\|^2 = \|x\|^2 + \|y\|^2$.*

Proof. $\|x \oplus y\|^2 = \langle (x \oplus y), (x \oplus y) \rangle = \|x\|^2 + \langle x, y \rangle + \langle y, x \rangle + \|y\|^2 = \|x\|^2 + \|y\|^2$ because $\langle x, y \rangle = \langle y, x \rangle = 0$. □

Definition 1.4. (Direct Sum) Let V be a vector space and let U and W be two subspaces of V. Then, a subspace Y is the direct sum of U and W, denoted as $Y = U \bigoplus W$, if $\forall\, y \in Y\ \exists\, u \in U$ and $w \in W$ such that there is a *unique* representation $y = u \oplus w$; and U is called the algebraic complement of W (alternatively, W is called the algebraic complement of U) in Y.

Definition 1.5. (Orthogonal Complement) Let $(V, \langle \bullet, \bullet \rangle)$ be an inner product space and let $\emptyset \subset M \subseteq V$. (Note that M may or may not be a vector space.) Then, the orthogonal complement of M in $(V, \langle \bullet, \bullet \rangle)$ is defined as

$$M^\perp = \big\{ x \in V :\ x \perp y \ \ \forall\, y \in M \big\}$$

i.e., $M^\perp$ is made of all vectors in V, which are orthogonal to each vector in M. An $x \in M^\perp$ is denoted as $x \perp M$.

Remark 1.6. If $M = \{\mathbf{0}_V\}$, then $M^\perp = V$ and if $M = V$, then $M^\perp == \{\mathbf{0}_V\}$. Also note that

$$M \bigcap M^\perp = \begin{cases} \{\mathbf{0}_V\} & \text{if } \mathbf{0}_V \in M \\ \emptyset & \text{if } \mathbf{0}_V \notin M \end{cases}$$

Theorem 1.3. *(Closure Complement) Let $(V, \langle \bullet, \bullet \rangle)$ be an inner product space and let $\emptyset \subset M \subseteq V$. Then, $M^{\perp}$ is a closed subspace of $(V, \langle \bullet, \bullet \rangle)$.*

Proof. It follows from Definition 1.5 that $M^{\perp}$ satisfies all conditions for a vector space because any linear combination of vectors orthogonal to M is also orthogonal to M. Therefore, $M^{\perp}$ is a subspace of V. What remains to show that $M^{\perp}$ is closed in $(V, \langle \bullet, \bullet \rangle)$.

Let $\{x^k\}$ be a convergent sequence in $M^{\perp}$ and let $x^k \to \tilde{x}$. Since $\langle x^k, y \rangle = 0 \,\forall\, x^k$ for any $y \in M$, it follows from continuity of inner products (see Lemma 1.3) that $\langle \tilde{x}, y \rangle = \lim_{k\to\infty} \langle x^k, y \rangle = 0$. Therefore, the limit $\tilde{x} \in M^{\perp}$ and so $M^{\perp}$ is a closed subspace of $(V, \langle \bullet, \bullet \rangle)$. □

Theorem 1.4. *: Let $(V, \langle \bullet, \bullet \rangle)$ be an inner product space and let $\emptyset \subset S \subseteq V$ and $\emptyset \subset T \subseteq V$. Then, the following statements are valid.*

1. *If $S \subseteq T$, then $T^{\perp} \subseteq S^{\perp}$ and $S^{\perp\perp} \subseteq T^{\perp\perp}$.*
2. *$S \subseteq S^{\perp\perp}$ and $S^{\perp} = S^{\perp\perp\perp}$.*
3. *If $x \in S \bigcap S^{\perp}$, then $x = \mathbf{0}_V$. Hence, it follows taht*

$$S \bigcap S^{\perp} = \begin{cases} \{\mathbf{0}_V\} & \text{if } \mathbf{0}_V \in S \\ \emptyset & \text{if } \mathbf{0}_V \notin S \end{cases}$$

4. *$\{\mathbf{0}_V\}^{\perp} = V$ and $V^{\perp} = \{\mathbf{0}_V\}$.*
5. *If S is a dense subset of V, then $S^{\perp} = \{\mathbf{0}_V\}$.*

Note: These results do not need completion of the inner product space $(V, \langle \bullet, \bullet \rangle)$.

Proof. The proof is as follows.

1. Let $x \in S \subseteq T$ and $y \in T^{\perp}$, where both choices of x and y are arbitrary. Therefore, $x \perp y$ and hence $y \in S^{\perp}$. So, $T^{\perp} \subseteq S^{\perp}$. Applying this procedure once more on the deived condition: $T^{\perp} \subseteq S^{\perp}$, it follows that $S^{\perp\perp} \subseteq T^{\perp\perp}$.
2. Let $x \in S$, where the choice of x is arbitrary. So $x \perp S^{\perp}$, which implies that $x \in S^{\perp\perp}$. Hence, $S \subseteq S^{\perp\perp}$. Then, by (1) it follows that $S^{\perp\perp\perp} \subseteq S^{\perp}$ and also $S^{\perp} \subseteq (S^{\perp})^{\perp\perp} = S^{\perp\perp\perp}$. Hence, $S^{\perp} = S^{\perp\perp\perp}$.
3. If $x \in S \bigcap S^{\perp}$, then $x \perp x$ implying thar $\langle x, x \rangle = 0$. That is, $\|x\| = 0$ or $x = \mathbf{0}_V$. The remaining part follows in the same way from Definition 1.5.
4. The proof follows directly from Definition 1.5.

5. If $x \in S^{\perp}$, then $\forall\ y \in S$ it follows from Pythagorean theorem that

$$\|x \oplus (-y)\| = \|x\|^2 + \|y\|^2 \geq \|x\|^2$$

Since S is dense in V, y can be chosen such that $\|x \oplus (-y)\|$ becomes arbitrarily small. Hence, $\|x\| = 0$. (Note that if $S^{\perp} = \{\mathbf{0}_V\}$, then S is dense in V only if $(V, \langle \bullet, \bullet \rangle)$ is a Hilbert space.)

□

Theorem 1.5. *Let $(H, \langle \bullet, \bullet \rangle)$ be a Hilbert space and let G be a subspace of H. Then, the following statements are valid.*

1. *$\overline{G} = G^{\perp\perp}$, where $\overline{G}$ is the closure of G.*
2. *If G is closed in $(H, \langle \bullet, \bullet \rangle)$, then $G^{\perp\perp} = G$.*
3. *If $G^{\perp} = \{\mathbf{0}_V\}$ if and only if G is dense in H.*
4. *If G is closed in $(H, \langle \bullet, \bullet \rangle)$ and if $G^{\perp} = \{\mathbf{0}_V\}$, then $G = H$.*

Note: These results do need completion of the inner product space $(H, \langle \bullet, \bullet \rangle)$.

Proof. The proof is as follows.

1. It is known that $G \subseteq \overline{G}$. It follows from Theorem 1.4 that $G \subseteq G^{\perp\perp}$, It also follows from Theorem 1.3 that $\overline{G} \subseteq G^{\perp\perp}$. Now, if $\overline{G} \neq G^{\perp\perp}$, then there is a nonzero vector $z \in G^{\perp\perp}$ such that $z \perp \overline{G}$.since $G \subseteq \overline{G}$, it follows that $z \perp G$. Therefore, $z \in G \bigcap G^{\perp}$ $z = \mathbf{0}_V$. This is a contradiction. Hence, $\overline{G} = G^{\perp\perp}$.
2. The proof follows directly from part (1) above.
3. The "if" part follows from Theorem 1.4 part (5). Now let $G^{\perp} = \{\mathbf{0}_V\}$. Then, $G^{\perp\perp} = V$ and it follows from part (1) that $\overline{G} = V$. Hence, G is dense in V.
4. This part follows from part (3) above.

□

Theorem 1.6. *Let $(H, \langle \bullet, \bullet \rangle)$ be a Hilbert space and let F and G be two closed subspaces of H. If $F \perp G$, the direct sum $F \bigoplus G$ is a closed subspace of H.*

Note: This result does need completion of the inner product space $(H, \langle \bullet, \bullet \rangle)$

Proof. Let $\{z_k\}$ be a convergent sequence in $F \bigoplus G$ with $z_k \to z$. we want to show that $z \in F \bigoplus G$. That is, $z_k = x_k \oplus y_k$, where $x_k \in F$ and $y_k \in G$. Since $F \perp G$, it follows from Pythagorean Theorem that

$$\|z_k - z_\ell\|^2 = \|x_k - x_\ell\|^2 + \|y_k - y_\ell\|^2$$

hence, $\{x_k\}$ and $\{y_k\}$ are Cauchy sequences in F and G , respectively. It follows from continuity of addition that $z = x \oplus y$. Therefore, $F \bigoplus G$ is closed in V. □

Theorem 1.7. *(Projection Theorem: Version 1) Let $(H, \langle \bullet, \bullet \rangle)$ be a Hilbert space and let G be a closed subspace of H. Then,*

1. *$H = G \bigoplus G^\perp$.*
2. *Each $x \in H$ can be uniquely expressed as $x = y \oplus z$, where $y \in G$ and $z \in G^\perp$, and $\|x\|^2 = \|y\|^2 + \|z\|^2$.*

Note: This result does need completion of the inner product space $(H, \langle \bullet, \bullet \rangle)$] and closedness of G in H.

Proof. It follows from Theorem 1.6 that $G \bigoplus G^\perp$ is a closed subspace of H. Since $G \subseteq G \bigoplus G^\perp$ and $G^\perp \subseteq G \bigoplus G^\perp$, it follows from Theorem 1.4 and Theorem 1.5 that $H = G \bigoplus G^\perp$.

See Naylor & Sell: Proof of Theorem 5.15.6 in Page 298 for more details. □

1.2 Orthogonal Projection

Definition 1.6. (Orthogonal Projection) A projection P on an inner product space is called an orthogonal projection if the range space and null space of P are orthogonal, i.e., $\mathcal{R}(P) \perp \mathcal{N}(P)$.

Remark 1.7. It follows from Definition 1.6 that if P is an orthogonal projection, then so is $I - P$.

Theorem 1.8. *(Continuity of Orthogonal Projection) An orthogonal projection in an inner product space is continuous.*

Proof. See Naylor & Sell: Proof of Theorem 5.16.2 in Page 300. □

Theorem 1.9. *Let P be an orthogonal projection on an inner product space V. Then,*

1. *$\mathcal{R}(P)$ and $\mathcal{N}(P)$ are closed subspaces of V.*

2. *$\mathcal{N}(P) = \mathcal{R}(P)^{\perp}$ and $\mathcal{R}(P) = \mathcal{N}(P)^{\perp}$.*

3. *Every vector $x \in V$ can be uniquely expressed as $x = r \oplus n$, where $r \in \mathcal{R}(P)$ and $n \in \mathcal{N}(P)$.*

4. *$\|x\|^2 = \|r\|^2 + \|n\|^2$.*

Proof. See Naylor & Sell: Proof of Theorem 5.16.3 in Page 301. □

Theorem 1.10. *(Projection Theorem: Version 2) Let $(H, \langle \bullet, \bullet \rangle)$ be a Hilbert space and let G be a closed subspace of H. Then,*

1. *$\mathcal{R}(P)$ and $\mathcal{N}(P)$ are closed subspaces of V.*

2. *$\mathcal{N}(P) = \mathcal{R}(P)^{\perp}$ and $\mathcal{R}(P) = \mathcal{N}(P)^{\perp}$.*

3. *Every vector $x \in V$ can be uniquely expressed as $x = r \oplus n$, where $r \in \mathcal{R}(P)$ and $n \in \mathcal{N}(P)$.*

4. *$\|x\|^2 = \|r\|^2 + \|n\|^2$.*

Proof. See Naylor & Sell: Proof of Theorem 5.16.2 in Page 300. □

1.3 Optimization in a Hilbert Space

Let us digress slightly from the main theme of this chapter and focus our attention on optimization in a Hilbert space, where the projection theorem characterizes the optimal solution. We present two slightly different versions of the theorem – one dealing with an arbitrary inner product space and the other providing a stronger solution in a Hilbert space. In this context, the problem is stated as follows.

Let V be an inner product space and let W be a subspace of V. Given an arbitrary vector $y \in V$, the task is to identify a vector $\hat{x} \in W$ closest to y in the sense that the error $\|y - \hat{x}\|$ is minimized. If $y \in W$, the the obvious solution is $\hat{x} = y$. Therefore, we deal with the situation $y \notin W$, where the following questions must be answered for a complete solution of the optimization problem.

(i) (Existence) Does there always exist a vector $\hat{x} \in W$ that minimizes the error $\|(y - \hat{x})\|$, or there is no $\hat{x} \in W$ that is as least as good as all others?

(ii) (Uniqueness) Is the solution unique?

(iii) (Characterization) How the solution is characterized if it exists?

Theorem 1.11. *(Uniqueness of the minimizing vector) Let V be an inner product space and let W be a subspace of V. Given an arbitrary vector $y \in V$, if there exists a vector $\hat{x} \in W$ such that $\|(y-\hat{x})\| \leq \|(y-x)\|$ $\forall y \in V$, then $\hat{x} \in W$ is unique. A necessary and sufficient condition that $\hat{x}$ be a unique minimizing vector in W is that the error vector $(y-\hat{x})$ is orthogonal to W.*

Proof. First we apply a contradiction to show that if $\hat{x} \in W$ is a minimizing vector, then the error vector $(y-\hat{x})$ is orthogonal to W. Let us assume that there exists an $x \in W \setminus \{\mathbf{0}_H\}$ that is not orthogonal to $(y-\hat{x})$. It is noted that $\|x\| > 0$ and let $\varepsilon \triangleq \frac{\langle (y-\hat{x}),x\rangle}{\|x\|}$; obviously, $\varepsilon \neq 0$. Let $\tilde{x} \triangleq \hat{x} + \varepsilon x$. Then,

$$\|y-\tilde{x}\|^2 = \|y-\hat{x}-\varepsilon x\|^2 = \|y-\hat{x}\|^2 - \langle (y-\hat{x}), \varepsilon x\rangle - \langle \varepsilon x, (y-\hat{x})\rangle + |\varepsilon|^2\|x\|^2$$

$$\Rightarrow \ \|y-\tilde{x}\|^2 = \|y-\hat{x}\|^2 - |\varepsilon|^2|x\|^2 < \|y-\hat{x}\|^2$$

Thus, if $(y-\hat{x})$ is not orthogonal to W, then $\hat{x}$ is not a minimizing vector.

Next we show that if $(y-\hat{x})$ is orthogonal to W, then $\hat{x}$ is a unique minimizing vector. For any $x \in W$, the Pythagorean Theorem yields

$$\|y-x\|^2 = \|(y-\hat{x}) + (\hat{x}-x)\|^2 = \|y-\hat{x}\|^2 + \|\hat{x}-x\|^2$$

Thus, $\|y-x\| > \|y-\hat{x}\|$ $\forall x \neq \hat{x}$. □

Next we address the conditions for existence of the minimizing vector by strengthening the above theorem on uniqueness. Thus, we will satisfy the criteria for both existence and uniqueness of the minimizing vector.

Theorem 1.12. *(Existence and uniqueness of the minimizing vector) Let H be a Hilbert space and G be a closed subspace of H. Then, corresponding to any vector $y \in H$, there exists a unique vector $\hat{x} \in G$ such that $\|y-\hat{x}\| \leq \|y-x\|$ $\forall x \in G$. Furthermore, a necessary and sufficient condition that $\hat{x} \in G$ be the unique minimizing vector is that $(y-\hat{x})$ be orthogonal to G.*

Proof. The uniqueness and orthogonality have been established in Theorem 1.11. What remains to be establish is existence of the minimizing vector $\hat{x} \in G$.

If $y \in G$, then trivially $\hat{x} = y$; let $y \notin G$ and $\delta \triangleq \inf_{x \in G} \|y-x\|$. In order to identify an $\hat{x} \in G$ such that $\|y-\hat{x}\| = \delta$, we proceed as follows.

Let $\{x^k\}$ be a sequence of vectors in G such that $\lim_{k\to\infty} \|y-x^k\| = \delta$. Then, it follows by using the parallelogram law that

$$\|(x^k-y)+(y-x^\ell)\|^2 + \|(x^k-y)+(y-x^\ell)\|^2 = 2\big(\|x^k-y\|^2 + \|y-x^\ell\|^2\big)$$

A rearrangement of the above expression yields

$$\|(x^k - x^\ell)\|^2 = 2\|x^k - y\|^2 + 2\|y - x^\ell\|^2 - 4\left\|y - \frac{(x^k + x^\ell)}{2}\right\|^2$$

Since the vector $\frac{(x^k+x^\ell)}{2} \in G \; \forall k, \ell \in \mathbb{N}$, it follows that $\left\|y - \frac{(x^k+x^\ell)}{2}\right\| \geq \delta$. Therefore,

$$\|(x^k - x^\ell)\|^2 \leq 2\|x^k - y\|^2 + 2\|y - x^\ell\|^2 - 4\delta^2$$

As $\|y - x^k\|^2 \to \delta^2$ and $\|y - x^\ell\|^2 \to \delta^2$, it follows that $\lim_{k.\ell \to \infty} \|x^k - x^\ell\|^2 = 0$. Therefore, $\{x^k\}$ is a Cauchy sequence in G that is a closed (and hence a complete) subspace of the Hilbert space H. Therefore, $\{x^k\}$ converges to the limit point in G. Then, by continuity of the norm, it follows that the limit is $\hat{x} \in G$. □

Remark 1.8. Although the proof of the above theorem does not make any explicit reference to the inner product, it does make use of the parallelogram law that, in turn, makes use of the inner product for its proof.

2 Orthonormal Sets and Bases

We have presented the notion of Hamel basis in Chapter 02 and Schauder basis in Chapter 03, which involve only algebraic notions. Now we incorporate both algebraic and topological notions in the formulation of a basis.

Definition 2.1. (Orthogonal Set) Let $\mathcal{I}$ be an arbitrary nonempty index set (i.e., finite or countable or uncountable). A set $\{x^\alpha : \alpha \in \mathcal{I}\}$ in an inner product space is called orthogonal if $\forall \alpha \neq \beta$, $x^\alpha \perp x^\beta$, i.e., $\langle x^\alpha, x^\beta \rangle = 0$. In addition, if $\langle x^\alpha, x^\alpha \rangle = 1 \; \forall \alpha \in \mathcal{I}$, i.e., $\langle x^\alpha, x^\beta \rangle = \delta_{\alpha\beta} \;\; \alpha, \beta \in \mathcal{I}$, then the set $\{x^\alpha\}$ is called orthonormal. Note that

$$\delta_{\alpha\beta} \triangleq \begin{cases} 1 & \text{if } \alpha = \beta \\ 0 & \text{if } \alpha \neq \beta \end{cases}$$

is called the Kronecker delta. Also note that any orthogonal set of non-zero vectors can be converted into an orthonormal set by replacing x^α with $\frac{x^\alpha}{\|x^\alpha\|} \; \forall \alpha$.

Lemma 2.1. *Every orthonormal set of vectors in an inner product space is linearly independent.*

Proof. Let $S = \{x^\alpha\}$ be an orthonormal set in an inner product space V. It suffices to show that, for every subset $\{x^k \in S : k \in \{1, 2, \cdots, n\}$ and $n \in \mathbb{N}$, the only solution of the algebraic equation $\sum_{k=1}^{n} \alpha_k x^k = \mathbf{0}_V$ is $\alpha_k = 0$ for all $k \in \{1, 2, \cdots, n\}$. Now, taking inner product on both sides with x^j for all $j \in \{1, 2, \cdots, n\}$,

$$\Big\langle \sum_{k=1}^{n} \alpha_k x^k, x^j \Big\rangle = \langle \mathbf{0}_V, x^j \rangle = 0 \ \Rightarrow \alpha_j = 0 \ \forall j.$$

□

Definition 2.2. (Complete Orthonormal Set) Let $\mathcal{I}$ be an arbitrary nonempty index set. An orthonormal set $B = \{x_\alpha\}$ in an inner product space V is called complete (or maximal) if the following condition holds

$$\big(x \perp x_\alpha \ \forall \alpha \in \mathcal{I}\big) \Rightarrow x = \mathbf{0}_V$$

In other words, no unit vector $x \in V$ exists such that $Big(B \bigcup \{x\}$ is an orthonormal set. A complete orthonormal set in a Hilbert space V is called an orthonormal basis for V.

Definition 2.3. (Orthonormal Basis) Let $\mathcal{I}$ be an arbitrary nonempty index set. An orthonormal set $B = \{x_\alpha\}$ in an inner product space V is called complete (or maximal) if the following condition holds

$$\big(x \perp x_\alpha \ \forall \alpha \in \mathcal{I}\big) \Rightarrow \big(x = \mathbf{0}_V\big)$$

Equivalently, no unit vector $x \in V$ exists such that $B \bigcup \{x\}$ is an orthonormal set.

Remark 2.1. Every orthonormal set of vectors in in a Hilbert space can be extended to form an orthonormal basis. Furthermore, two orthonormal bases of a Hilbert space must have the same cardinality.

2.1 The Fourier Series Theory

This subsection presents the fundamental properties of orthogonal bases by combining the topological and algebraic structures of a Hilbert space.

Lemma 2.2. *(Bessel Inequality) Let $\{x^k\}$ be an orthnormal set in an inner product space V. Then, for every $x \in V$, the following condition holds;*

$$\sum_{k \in \mathbb{N}} \mid \langle x, x^k \rangle \mid^2 \leq \|x\|^2$$

Proof. Let $\{x^1, \cdots, x^n\}$ be a finite subset from $\{x^k\}$. Then, by making use of the orthonarmality $\langle x^j, x^k\rangle = \delta_{jk}$, it follows that

$$0 \le \Big\| x - \Big(\sum_{j=1}^{n} \langle x, x^j\rangle x^j\Big)\Big\|^2$$

$$= \Big\langle \Big(x - \sum_{j=1}^{n}\langle x, x^j\rangle x^j\Big), \Big(x - \sum_{k=1}^{n}\langle x, x^k\rangle x^j\Big)\Big\rangle$$

$$= \langle x, x\rangle - \sum_{j=1}^{n}\langle x, x^j\rangle\langle x^j, x\rangle - \sum_{k=1}^{n}\langle x, x^k\rangle\langle x^k, x\rangle + \sum_{j=1}^{n}\sum_{k=1}^{n}\langle x, x^j\rangle\langle x^j, x\rangle\langle x, x^k\rangle\langle x^k, x\rangle$$

$$= \|x\|^2 - \sum_{k=1}^{n} |\ \langle x, x^k\rangle\ |^2$$

Since the above inequality is true for all $n \in \mathbb{N}$, it must also hold true for a (countable) infinite sum. Hence, the proof is established. □

Theorem 2.1. *(Fourier Series Theorem) Let $\{x^k\}$ be an orthonormal set in a Hilbert space H. Then, the following statements are equivalent.*

(a) *(Orthonormal basis) $\{x^k\}$ be an orthonormal basis of H.*

(b) *(Fourier series expansion) Every $x \in H$ can be expanded as $\sum_k \langle x, x^k\rangle x^k$.*

(c) *(Parseval equality) $\forall x, y \in H$, the following relation holds; $\langle x, y\rangle = \sum_k \langle x, x^k\rangle\ \overline{\langle y, x^k\rangle}$.*

(d) *(Norm decomposition) The norm of every $x \in H$ can be decomposed as:*

$$\|x\|^2 = \sum_k |\langle x, x^k\rangle|^2 < \infty.$$

(e) *(Dense subspace of H) Let V be a subspace of H such that $\{x^k\} \subset V$. Then, V is dense in H, i.e. closure(V)=H.*

Proof. Following Naylor & Sell (pp. 307-312), the proof is completed by showing the equivalence of individual parts in the following order: $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (a)$ and $(b) \Rightarrow (e) \Rightarrow (b)$. □

Corollary 2.1. *Let V be a closed subspace of a Hilbert space H, where $V \neq \{\mathbf{0}_H\}$; and let $\{x^k\}$ be an orthonormal basis of V. Then, the orthogonal projection P of H onto V is given by $Px = \sum_k \langle x, x^k\rangle x^k \quad \forall x \in H$.*

Proof. The proof follows from Theorem 2.1. □

2.1.1 Gram-Schmidt Procedure

Let V be an inner product space and let $\{y^k\}$ be a linearly independent set with countable cardinality in V. The objective here is to construct an orthonormal sequence from the By making use of Bessel inequality in Lemma 2.2. Now we present the following theorem.

Theorem 2.2. *(Gram-Schmidt Orthonormalization) Let $\{x^k\}$ be an at most countable (i.e., finite or countably infinite) set of linearly independent vectors in an inner product space V. Then, there exists an orthonormal sequence $\{e^k\} \subset V$ such that, for every $n \in \mathbb{N}$,*

$$Span[e^1, \cdots, e^n] = Span[x^1, \cdots, x^n].$$

Proof. First let us normalize the first vector x^1 as $e^1 \triangleq \frac{x^1}{\|x^1\|}$ that obviously spans the same space as x^1 does. Next we construct $z^2 \triangleq x^2 - \langle x^2, e^1\rangle e^1$. Also note that $z^2 \neq \mathbf{0}_V$ because x^2 and x^1 are linearly independent and $Span[e^1, e^2] = Span[x^1, x^2]$. Now, let $e^2 \triangleq \frac{z^2}{\|z^2\|}$, which assures $\langle e^2, e^1\rangle = 0$ and $\langle e^2, e^2\rangle = 1$.

Proceeding in this way, the remaining e^k's are generated by induction as follows. Given $z^k \triangleq x^k - \sum_{j=1} k - 1\langle x^k, e^j\rangle e^j$ and $e^k \triangleq \frac{z^k}{\|z^k\|}$ for all $k \in \{3, \cdots, n\}$. Since $\langle e^j, e^k\rangle = \delta_{jk}\ \forall j, k \in \{1, \cdots, n\}$, we have $\langle z^{n+1}, e^k\rangle = 0$, i.e., z^{n+1} is orthogonal to $Span[e^1, \cdots, e^n] = Span[x^1, \cdots, x^n]$. Next, define $e^{n+1} \triangleq \frac{z^{n+1}}{\|z^{n+1}\|}$. The proof by induction is thus completed. □

Example 2.1. (Orthogonal polynomials) Let $T = (-1, 1)$ and let us consider the (linearly independent) sequence of vectors $\{t^k : k \in \mathbb{N} \cup \{0\}$ and $t \in T\}$ on the inner product space $L_2(T)$ over the real field $\mathbb{R}$, where the inner product is defined as $\langle x, y\rangle \triangleq \int_T dt\ x(t)y(t)\ \forall x, y \in L_2(T)$. Then, Gram-Schmidt orthonormalization of the sequence $\{t^k\}$ generates a set of orthonormal polynomials, called Legendre polynomials as seen below.

Noting that $\langle 1, 1\rangle = \int_{-1} 1\ dt = 2 \Rightarrow \|1\| = \sqrt{2}$, we have $e^1 = \frac{t^0}{\|t^0\|} = \frac{1}{\sqrt{2}}$.

$z^2 \triangleq t - \langle t, e^1\rangle e^1 = t - \frac{1}{2}\int_{-1} 1\ t\ dt = t \Rightarrow e^2 = \frac{z^2}{\|z^2\|} = \sqrt{\frac{3}{2}}\ t$.

$z^3 \triangleq t^2 - \langle t^2, \frac{1}{\sqrt{2}}\rangle \frac{1}{\sqrt{2}} - \langle t^2, \frac{3}{\sqrt{2}t}\rangle \frac{3}{\sqrt{2}t} = t^2 - \frac{1}{3} \Rightarrow e^3 = \frac{z^3}{\|z^3\|} = \sqrt{\frac{45}{8}}\ (t^2 - \frac{1}{3})$.

$\vdots$

There are many other families of orthonormal polynomials that can by gram-Schmidt orthonormalization of the (linearly independent) sequence of vectors $\{t^k : k \in \mathbb{N} \cup \{0\}$ and $t \in T\}$ on different inner product spaces. A few examples follow.

- Laugure Polynomials on the space where the inner product is defined as: $\langle x, y\rangle \triangleq \int_T dt \ \exp(-t) \ x(t) \ y(t)$, where $T = (0, \infty)$.
- Hermite Polynomials on the space where the inner product is defined as: $\langle x, y\rangle \triangleq \int_T dt \ \exp(-t^2) \ x(t) \ y(t)$, where $T = (-\infty, \infty)$.
- Chebyshev Polynomials on the space where the inner product is defined as: $\langle x, y\rangle \triangleq \int_T dt \ \sqrt{(1-t^2)} \ x(t) \ y(t)$, where $T = (-1, 1)$.
- Chebyshev Polynomials on the space where the inner product is defined as: $\langle x, y\rangle \triangleq \int_T dt \ \sqrt{(1-t^2)} \ x(t) \ y(t)$, where $T = (-1, 1)$.

2.1.2 Fourier Transform

Now we introduce the concept of Fourier transform of (absolute) integrable functions, i.e., the functions belonging to the $L_1(\mu)$ space, where μ is the Lebesgue measure on the real line $\mathbb{R}$. We note that, in the space $L_p(\mu)$ for $p \in [1, \infty]$, Lebesgue measure μ is defined on any measurable subset of the real line $\mathbb{R}$. In other words, we consider functions $f \in L_1(\mu)$, where $\int_{\mathbb{R}} d\mu(t)|f(t)| < \infty$, for Fourier transform.

Definition 2.4. (Fourier transform) For $f \in L_1(\mu)$, the Fourier transform of f is:

$$\hat{f}(\xi) \triangleq \int_{\mathbb{R}} d\mu(t) \ \exp(-i2\pi\xi t) f(t) \ \ \forall \xi \in \mathbb{R} \qquad \text{Fourier Analysis Formula}$$

and for $\hat{f} \in L_1(\mu)$, i.e., $\int_{\mathbb{R}} d\mu(\xi)|\hat{f}(\xi)| < \infty$, inverse Fourier transform of $\hat{f}$ is:

$$f(t) \triangleq \int_{\mathbb{R}} d\mu(\xi) \ \exp(i2\pi\xi t)\hat{f}(\xi) \ \ \forall t \in \mathbb{R} \qquad \text{Fourier Synthesis Formula}$$

Remark 2.2. The Fourier transform converts time translation into a phase shift and vice versa, and convolution into products as seen below. To see this, let $f \in L_1(\mu)$ and $\alpha, \lambda \in \mathbb{R}$. Then,

1. If $g(t) = f(t)\exp(i2\pi\alpha t)$, then $\hat{g}(\xi) = \hat{f}(\xi - \alpha)$ and if $g(t) = f(t - \alpha)$, then $\hat{g}(\xi) = \hat{f}(\xi)\exp(-i2\pi\alpha\xi)$.
2. If $g \in L_1(\mu)$ and $h = f \star g$, then $\hat{h}(\xi) = \hat{f}(\xi)\hat{g}(\xi)$. Similarly, if $\hat{g} \in L_1(\mu)$ and $\hat{h} = \hat{f} \star \hat{g}$, then $h(t) = f(t)g(t)$.
3. If $g(t) = \overline{f(-t)}$, then $\hat{g}(\xi) = \overline{\hat{f}(\xi)}$.
4. If $\lambda \neq 0$ and $g(t) = f(\frac{t}{\lambda})$, then $\hat{g}(\xi) = |\lambda|\hat{f}(\lambda\xi)$.

5. Let $n \in \mathbb{N}$. If f is a C^{∞}-continuous function, then Fourier transform of $\frac{\partial^n f(t)}{\partial t^n}$ is $(i2\pi\xi)^n \hat{f}(\xi)$; similarly, Fourier transform of $(-i2\pi t)^n f(t)$ is $\frac{\partial^n \hat{f}(\xi)}{\partial \xi^n}$, provided that the Fourier transforms exist.

6. Fourier transform of $\int_{-\infty}^{t} d\mu(\tau) f(\tau)$ is $\frac{\hat{f}(\xi)}{i2\pi\xi}$ provided that $\hat{f}(0) = 0$.

7. Let $n \in \mathbb{N} \cup \{0\}$. Then, $(-i2\pi)^n \int_{-\infty}^{\infty} d\mu(\tau)\tau^n f(\tau) = \left.\frac{\partial^n \hat{f}}{\partial \xi^n}\right|_{\xi=0}$, provided that the integral and the derivatives exist.

Also note that if $f \in L_1(\mu)$ and $\hat{f}(\xi) = 0 \ \forall \xi \in \mathbb{R}$, then $f(t) = 0$ almost everywhere (μ-a.e.) on $\mathbb{R}$.

Next we introduce the notion of Plancherel Theorem. Since the Lebesgue measure μ of $\mathbb{R}$ is not finite (it is σ-finite), we cannot claim $L_2(\mu)$ to be a subspace of $L_1(\mu)$. Fourier transform in Definition 2.4 is not directly applicable to every $f \in L_2(\mu)$. However, the definition does apply if $f \in L_1(\mu) \bigcap L_2(\mu)$ and it turns out that $\hat{f} \in L_2(\mu)$. In fact, $\|\hat{f}\|_{L_2} = \|f\|_{L_2}$. This isometry of $L_1(\mu) \bigcap L_2(\mu)$ into $L_2(\mu)$ extends to an isometry of $L_2(\mu)$ onto $L_2(\mu)$, and this defines an extension of the Fourier transform, which is often called as the Plancherel transform and is applicable to every $f \in L_2(\mu)$. The resulting L_2-theory has more symmetry than that in the L_1-theory. In other words, f and $\hat{f}$ play the same role in $L_2(\mu)$.

Theorem 2.3. *(Plancherel Theorem) Each function $f \in L_2(\mu)$ can be associated to a function $\hat{f} \in L_2(\mu)$ such that the following properties hold:*

1. *If $f \in L_1(\mu) \bigcap L_2(\mu)$, then $\hat{f}$ is the the Fourier transform of f, i.e., $\hat{f}(\xi) = \int_{\mathbb{R}} dt \ \exp(-i2\pi\xi t) f(t) \quad \forall \xi \in \mathbb{R}$.*

2. *$\|\hat{f}\|_{L_2} = \|f\|_{L_2}$ for every $f \in L_2(\mu)$.*

3. *The mapping $f \mapsto \hat{f}$ is a Hilbert space isomorphism of $L_2(\mu)$ onto $L_2(\mu)$.*

4. *The following symmetric relation exists between f and $\hat{f}$: If $\varphi_A(\xi) \triangleq \int_{-A}^{A} d\mu(t) \exp(-i2\pi\xi t) f(t)$ and $\psi_A(t) \triangleq \int_{-A}^{A} d\mu(\xi) \exp(i2\pi\xi t) \hat{f}(\xi)$, then $\|\varphi_A - \hat{f}\|_{L_2} \to 0$ and $\|\psi_A - f\|_{L_2} \to 0$ as $A \to \infty$.*

Proof. See Real and Complex Analysis by Walter Rudin (pp.187-189). □

2.1.3 Hilbert Dimension and Separable Hilbert Spaces

Let us classify Hilbert spaces (over the same field) by cardinality of the orthonormal basis set. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two orthonormal bases of a Hilbert space H.

Since $\mathcal{B}_1$ and $\mathcal{B}_2$ must have the same cardinal numbers, this cardinal number is called the *Hilbert dimension* of H. For finite-dimension spaces H, the Hilbert dimension of H is the same as the dimension related to a Hamel base. However, for infinite-dimensional Hilbert spaces, the Hibert dimension could be countable or uncountable. An example of countable Hilbert dimension is the Hilbert space of periodic functions $L_2[-\pi, \pi]$.

Example 2.2. The Hilbert space ℓ_2 is isometrically isomorphic to the Hilbert space $L_2[-\pi, \pi]$. That is, there exists a linear bijective mapping $f : \ell_2 \to L_2[-\pi, \pi]$ such that the inner products satisfy the following equality

$$\langle x, \tilde{x} \rangle_{\ell_2} = \langle f(x), f(\tilde{x}) \rangle_{L_2[-\pi,\pi]} \quad \forall x, \tilde{x} \in \ell_2$$

To see this, we proceed as follows. Let $\{x^k\}$ be an orthonormal basis for $L_2[-\pi, \pi]$ and let $f : \ell_2 \to L_2[-\pi, \pi]$ be defined as: $f(a) = \sum_{k \in \mathbb{N}} a_k x^k$

We cite below an example of a Hilbert space having uncountable Hilbert dimension.

Example 2.3. (Uncountable Hilbert dimension) Let us consider a trigonometric polynomial $p : \mathbb{R} \to \mathbb{C}$ such that $\forall t \in \mathbb{R} \quad p(t) \triangleq \sum_{k=1}^{m} c_k \exp(i r_k t)$ for some $m \in \mathbb{N}$, where $r_1, \cdots, r_m \in \mathbb{R}$ and $c_1, \cdots, c_m \in \mathbb{C}$. Let $\tilde{p} : \mathbb{R} \to \mathbb{C}$ be another trigonometric polynomial defined as $\forall t \in \mathbb{R} \quad \tilde{p}(t) \triangleq \sum_{k=1}^{\tilde{m}} \tilde{c}_k \exp(i \tilde{r}_k t)$ for some $\tilde{m} \in \mathbb{N}$, where $\tilde{r}_1, \cdots, \tilde{r}_{\tilde{m}} \in \mathbb{R}$ and $\tilde{c}_1, \cdots, \tilde{c}_{\tilde{m}} \in \mathbb{C}$. Then, it follows that

$$p(t)\ \overline{\tilde{p}}(t) = \sum_{j=1}^{m} \sum_{k=1}^{\tilde{m}} c_j \overline{\tilde{c}}_k \exp\Big(i(r_j - \tilde{r}_k)t\Big)$$

We define an inner product in the space trigonometric polynomials as

$$\langle p, \tilde{p} \rangle \triangleq \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} dt\ p(t)\ \overline{\tilde{p}}(t) = \sum_{j=1}^{m} \sum_{k=1}^{\tilde{m}} c_j \overline{\tilde{c}}_k$$

because

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} dt\ \exp\big(irt\big) = \begin{cases} 1 & \text{if } r = 0 \\ 0 & \text{if } r \in \mathbb{R} \setminus \{0\} \end{cases}$$

The completion of the inner product space of trigonometric polynomials is a Hilbert space called the space of almost periodic functions. Let us call this Hilbert space as H_{trigP}. For all $r \in \mathbb{R}$ let us consider the function $u_r(t) \triangleq \exp(irt),\ t \in \mathbb{R}$.

Then, $\{u_r : r \in \mathbb{R}\}$ is an uncountable orthonormal set in H_{trigP}. Since the set of all trigonometric polynomials is dense in H_{trigP}, it follows that if $p \in H$ and $\langle p, u_r\rangle = 0 \ \forall r \in \mathbb{R}$, then $\|p\|^2 = \langle p, p\rangle = 0 \Rightarrow p = \mathbf{0}_H$. Hence, $\{u_r : r \in \mathbb{R}\}$ is an uncountable orthonormal basis for H_{trigP}.

If $\{z_n : \ n \in \mathbb{N}\}$ is a countable subset of H_{trigP}, then the open ball $B_{\frac{1}{\sqrt{2}}}\big(z_n\big)$ contain at most one u_r, because $\|u_r - u_s\| = \sqrt{2}$ for $\forall\ r \neq s$. Hence, the collection of all open balls $B_{\frac{1}{\sqrt{2}}}\big(z_n\big)$ can cover only a countable number of u_r's. Since the set $\{u_r : r \in \mathbb{R}\}$ has uncountable cardinality, no countable set can be dense in H_{trigP}. Hence, H_{trigP} is a non-separable Hilbert space.

Theorem 2.4. *(Separable Hilbert Spaces) Let H be a Hilbert space. Then, H is a separable Hilbert space (i.e., H has a countable subset that dense in H) if and only if every orthonormal basis of H is countable.*

Proof. To prove the only if part, let the Hilbert space H be separable, $U \subseteq H$ be countable and dense in H, and M be an orthonormal basis of H. Then, for every $x, y \in M$,

$$\|x - y\| = \begin{cases} 0 & \text{if } x = y \\ \sqrt{2} & \text{if } x \neq y \end{cases}$$

Then, for $\forall x, y \in U$ and $x \neq y$, the open neighborhoods $N_{\frac{1}{\sqrt{2}}}(x)$ and $N_{\frac{1}{\sqrt{2}}}(y)$ in H are disjoint. Therefore, every $N_{\frac{1}{\sqrt{2}}}(x)$ in H can have at most one element of M. Since U is countable, M must be countable; else U cannot be dense in H.

To prove the if part, let $\{e^k\}$ be a countable orthonormal basis of H and let A be a set of all (finite) linear combinations $\gamma_1^n e^n + \ \cdots \ + \gamma_n^n e^n$ for all $n \in \mathbb{N}$, where $\gamma_j^n \triangleq \alpha_j^n + i\beta_j^n$ and $\alpha_j^n, \beta_j^n \in \mathbb{Q}$, the set of rational numbers. Note that the set A is countable and if H is defined over $\mathbb{R}$, then set all $\beta_j^n = 0$. We will prove that A is dense in H by showing

$$\forall x \in H \ \ \forall \varepsilon > 0 \ \ \exists\ \tilde{x} \in A \ \ \text{such that} \ \ \|x - \tilde{x}\| < \varepsilon$$

Given any $x \in H$ and $\varepsilon > 0$, let Y_n be the space spanned by a selected subset of vectors $\{e^1, \cdots e^n\}$ from the orthonormal basis $\{e^k\}$ such that $\|x - y\| < \frac{\varepsilon}{2} \ \forall y \in Y_n$. This is possible because $\{e^k\}$ is a complete ortonormal set in H. It follows, by orthogonal projection of x onto Y_n, that

$$y = \sum_{k=1}^{n} \langle x, e^k\rangle e^k \ \Rightarrow \ \Big\|x - \sum_{k=1}^{n} \langle x, e^k\rangle e^k\Big\| < \frac{\varepsilon}{2}$$

Let $\tilde{x} = \sum_{k=1}^{n} \gamma_k^n e^k$ be the orthogonal decomposition of $\tilde{x} \in A$. Then,

$$\|x - \tilde{x}\| = \left\| x - \sum_{k=1}^{n} \gamma_k^n e^k \right\|$$

$$\leq \left\| x - \sum_{k=1}^{n} \langle x, e^k \rangle e^k \right\| + \left\| \sum_{k=1}^{n} \langle x, e^k \rangle e^k - \sum_{k=1}^{n} \gamma_k^n e^k \right\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

which proves that A is dense in H; and since A is countable, H is separable. □

Example 2.4. The spaces ℓ_2 and $L_2([-\pi, \pi])$, defined on the same field $\mathbb{F}$, have the same Hilbert dimension. It is also true that ℓ_2 and $L_2([-\pi, \pi])$ are isometrically isomorphic as explained below.

Let $\{x^k\}$ be an orthonormal basis for the Hilbert space $L_2([-\pi, \pi])$ and let $f : \ell_2 \to L_2([-\pi, \pi])$ be defined as follows:

$$f(\mathbf{a}) \triangleq \sum_{k=1}^{\infty} a_k x^k, \text{ where } \mathbf{a} \triangleq \{a_1 \ a_2 \ \cdots\} \in \ell_2.$$

Now we show that f is linear and bijective. Linearity is established as follows.

$$f(\alpha \mathbf{a} + \mathbf{b}) = \sum_{k=1}^{\infty} \big(\alpha a + b\big)_k \ x^k \ = \ \alpha \sum_{k=1}^{\infty} a_k x^k \ + \sum_{k=1}^{\infty} b_k x^k \ = \ \alpha f(\mathbf{a}) + f(\mathbf{b})$$

To establish injectivity of f, let $f(\mathbf{a}) = f(\mathbf{b})$. Since f and $\{x^k\}$ is orthonormal, it follows that

$$0 = \|f(\mathbf{a}) - f(\mathbf{b})\|^2 = \left\| \sum_{k=1}^{\infty} (a-b)_k \ x^k \right\|^2 = \sum_{k=1}^{\infty} |(a-b)_k|^2 \ \|x^k\|^2 = \sum_{k=1}^{\infty} |(a-b)_k|^2$$

which implies that $\mathbf{a} = \mathbf{b}$.

To establish surjectivity of f and isometry of the spaces ℓ_2 and $L_2([-\pi, \pi])$, let $x \in L_2([-\pi, \pi])$. Then, it follows from Fourier series theorem that $x = \sum_{k=1}^{\infty} \langle x, x^k \rangle$. Setting $a_k = \langle x, x^k \rangle$, it follows from orthonormality of $\{x^k\}$ that

$$\|x\|^2 = \left\| \sum_{k=1}^{\infty} a_k \ x^k \right\|^2 = \sum_{k=1}^{\infty} |a_k|^2 \ \|x^k\|^2 = \sum_{k=1}^{\infty} |a_k|^2 = \|\mathbf{a}\|^2$$

which implies that, for each $x \in L_2([-\pi, \pi])$, there exists an $\mathbf{a} \in \ell_2$ such that $f(\mathbf{a}) = x$ and $\|\mathbf{a}\|_{\ell_2} = \|x\|_{L_2([-\pi,\pi])}$.

3 Applications of Hilbert Spaces

Hilbert spaces have numerous applications in Science and Engineering. In modern Physics (that started in early 120's), the mathematics of Quantum Mechanics is built upon the concept of Hilbert spaces over the complex field $\mathbb{C}$. While there is abundance of literature in mathematics of Quantum Mechanics, Chapter 11 of *Introductory Functional Analysis with Applications* by E. Kreyszig provides a brief explanation of the underlying concepts. In this section, we provide a few examples of applications of Hilbert spaces in Signal Processing and Quantum Mechanics.

3.1 Shannon Sampling: Bandlimited Signals

Sampling of continuous-time signal is a key step to discrete-time signal processing. In engineering practice, sampling is accomplished by analog-to-digital (A/D) conversion of continuously varying signals that are generated by various sensors. It is apparent that the sampling frequency should be selected depending on the bandwidth of the signal and the sensor that is used to measure the signal. For example, if the sampling is performed infrequently, the vital information content of the signal might be lost. On the other hand, a very high sampling frequency may generate excessively redundant data that would require additional memory and computation time. We present the analytical relationship between the signal bandwidth and the smallest sampling frequency for no loss of information.

Theorem 3.1. *(The Shannon Sampling Theorem) Let a (real-valued) signal* $f \in L_2(\mathbb{R})$ *be band-limited by* Ω*, i.e.,* $\hat{f}(\xi) = 0 \; \forall \; \xi \notin [-\Omega, \Omega]$*. Then,* $f(t)$ *can be perfectly reconstructed from its samples* $f(t_k)$*, collected at time instants* $t_k \triangleq \frac{k}{2\Omega}$*,* $k \in \mathbb{Z}$*, by the following interpolation formula:*

$$f(t) = \sum_{k \in \mathbb{Z}} \frac{\sin\big(2\pi\Omega \; (t - t_k)\big)}{2\pi\Omega \; (t - t_k)} \; f(t_k)$$

Proof. By Plancherel theorem, the L_2-norm of Fourier transform is expressed as:

$$\begin{aligned} \| \hat{f} \|^2_{L_2(-\Omega,\Omega)} &= \int_{-\Omega}^{\Omega} d\xi \;\; | \hat{f}(\xi) |^2 \\ &= \int_{-\infty}^{\infty} d\xi \;\; | \hat{f}(\xi) |^2 = \| \hat{f} \|^2_{L_2(\mathbb{R})} \\ &= \| f \|^2_{L_2(\mathbb{R})} \end{aligned}$$

Expansion of $\hat{f} \in L_2(-\Omega, \Omega)$ in a Fourier series yields

$$\hat{f}(\xi) = \frac{1}{2\Omega} \sum_{k \in \mathbb{Z}} c_k \; \exp\big(-i \; 2\pi\xi t_k \big)$$

where the Fourier coefficients c_k are identically equal to the samples of the signal $f(t)$ taken the time instants $t_k \triangleq \frac{k}{2\Omega}$, i.e.,

$$c_k = \int_{-\Omega}^{\Omega} d\xi \; \exp\big(i \; 2\pi\xi t_k \big) \hat{f}(\xi) = f(t_k)$$

It follows by equality in the L_2-sense and by exchange (due to uniform convergence of the summand) of the integral and the infinite sum that

$$\begin{aligned} f(t) &= \int_{\mathbb{R}} d\xi \; \exp\big(i \; 2\pi\xi t_k \big) \hat{f}(\xi) \\ &= \int_{-\Omega}^{\Omega} d\xi \; \exp\big(i \; 2\pi\xi t_k \big) \hat{f}(\xi) \\ &= \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} d\xi \; \exp\big(i \; 2\pi\xi t_k \big) \\ &\quad \times \sum_{k \in \mathbb{Z}} f(t_k) \; \exp\big(-i \; 2\pi\xi t_k \big) \\ &= \frac{1}{2\Omega} \sum_{k \in \mathbb{Z}} \int_{-\Omega}^{\Omega} d\xi \; \exp\big(i \; 2\pi\xi (t - t_k\big) f(t_k) \end{aligned}$$

Since $\int_{-\Omega}^{\Omega} d\xi \; \exp\big(i \; 2\pi\xi (t - t_k\big) = \frac{\sin(2\pi\Omega \; (t - t_k))}{2\pi\Omega \; (t - t_k)}$, it follows that

$$f(t) = \sum_{k \in \mathbb{Z}} \frac{\sin\big(2\pi\Omega \; (t - t_k) \big)}{2\pi\Omega \; (t - t_k)} \; f(t_k)$$

in the L_2-sense. As band-limited signals are analytic in the entire complex plane, it suffices that $f(\bullet)$ is continuous. Hence, the above equality holds pointwise everywhere in the domain of $f(\bullet)$. $\square$

3.2 Fourier Transform in Quantum Mechanics

Let x be the position of a particle and the quantum wave function $\psi(x)$ be a complex-valued function of x, which belongs to a Hilbert space over the complex field $\mathbb{C}$. The state of a classical particle in one dimension is specified by both position x and momentum p while, in quantum physics, the wave function depends on only one of the two variables. The information on momentum p is contained in the wave function $\psi(x)$. A convenient way to extract the information on p as a Fourier transform of $\psi(x)$ is to construct the *momentum wave function*

$$\hat{\psi}(p) \triangleq \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dx\ \exp\Big(\frac{-ipx}{\hbar}\Big)\psi(x) = \frac{1}{\sqrt{h}} \int_{-\infty}^{\infty} dx\ \exp\Big(\frac{-i2\pi px}{h}\Big)\psi(x)$$

where the Planck constant $\hbar = 1.054572 \times 10^{-34}$ Joule-sec and the original Plank constant $h = 2\pi\hbar$.

The above Fourier transform can be inverted to generate the information on position x from $\hat{\psi}(p)$ as

$$\psi(x) \triangleq \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dp\ \exp\Big(\frac{+ipx}{\hbar}\Big)\hat{\psi}(p) = \frac{1}{\sqrt{h}} \int_{-\infty}^{\infty} dp\ \exp\Big(\frac{+i2\pi px}{h}\Big)\hat{\psi}(p)$$

Therefore, the position waveform $\psi(x)$ and the momentum waveform $\hat{\psi}(p)$ contain the precisely the same information about a quantum state. By Plancherel Theorem, the inner product is obtained in either position domain or momentum domain as

$$\Big(\langle\varphi, \psi\rangle = \langle\hat{\varphi}, \hat{\psi}\rangle\Big) \Leftrightarrow \left(\int_{-\infty}^{\infty} dx\ \overline{\psi}(x)\varphi(x) = \int_{-\infty}^{\infty} dp\ \overline{\hat{\psi}}(p)\hat{\varphi}(p)\right)$$

In essence, the information about the momentum of a quantum particle can be obtained from the momentum wave function in the same way that information about its position can be obtained from the position wave function as explained below.

The position wave function has a compact support, i.e., $\psi(x)$ vanishes outside in a finite interval $x_1 \leq x \leq x_2$; similarly, momentum wave function has a compact support, i.e., $\hat{\psi}(p)$ vanishes outside a finite interval $p_1 \leq p \leq p_2$. Therefore, it is safe to say that the quantum particle does not lie outside a position interval and a momentum interval. However, we may not be able to specify a deterministic value of the position and momentum within these intervals. This issue is further examined in the next section from the perspectives of both Quantum Mechanics and Signal Processing.

3.3 Heisenberg Uncertainty Principle and Signal Analysis

Let us start with the statement that the use of "uncertainty" in the so-called *Uncertainty Principle* is perhaps a misnomer, because there is nothing uncertain about the uncertainty principle. In the discipline of signal analysis, it is a well-known mathematical fact that

A narrow-band waveform yields a wide spectrum and a wide-band waveform yields a narrow spectrum.

It is possible that both time and frequency can be narrow, but the probability densities of time and frequency localization cannot be made simultaneously arbitrarily narrow, i.e., their variances cannot be made simultaneously arbitrarily small.

In both physics and signal analysis, the uncertainty principle applies to a pair of variables whose associated operators do not commute.

In the physics literature,

Position Operator: $\chi \equiv (x)$

Momentum Operator: $p \equiv \big(-i\hbar \frac{\partial}{\partial x}\big)$ or $\big(-i\frac{h}{2\pi} \frac{\partial}{\partial x}\big)$,

The commutator of χ and p on an arbitrary (differentiable) function $f(x)$ is defined as:

$$[\chi, p] \triangleq \chi p - p\chi \quad \Rightarrow \quad [\chi, p]f(x) = x\Big(-i\hbar\frac{\partial}{\partial x}\Big)f(x) + i\hbar\frac{\partial}{\partial x}\Big(xf(x)\Big)$$
$$= -i\hbar\Big(x\frac{\partial f}{\partial x} - x\frac{\partial f}{\partial x} - f(x)\Big) = i\hbar f(x)$$
$$\Rightarrow [\chi, p] = i\hbar \;\text{ or }\; [\chi, p] = i\frac{h}{2\pi}$$

Similarly, in the signal analysis literature,

Time Operator: $\Im \equiv (t) \triangleq \big(\frac{i}{2\pi}\frac{d}{d\xi}\big)$

Frequency Operator: $\Xi \equiv (\xi) \triangleq \big(\frac{-i}{2\pi}\frac{d}{dt}\big)$

The commutator of $\Im$ and Ξ on an arbitrary (differentiable) signal $s(t)$ is defined as:

$$[\Im,\ \Xi] \triangleq \Im\ \Xi - \Xi\ \Im \ \Rightarrow\ [\Im,\ \Xi]\ s(t) = (t)\Big(\frac{-i}{2\pi}\frac{d}{dt}\Big)s(t) - \Big(\frac{-i}{2\pi}\Big)\frac{d}{dt}\Big(ts(t)\Big)$$
$$= \frac{-i}{2\pi}\Big(t\frac{ds}{dt} - t\frac{ds}{dt} - s(t)\Big) = \frac{i}{2\pi}s(t)$$
$$\Rightarrow [\Im,\ \Xi] = \frac{i}{2\pi}$$

Alternatively,

$$[\Im,\ \Xi] \triangleq \Im\ \Xi - \Xi\ \Im \ \Rightarrow\ [\Im,\ \Xi]\ \hat{s}(\xi) = \Big(\frac{i}{2\pi}\frac{d}{d\xi}\Big)(\xi)\hat{s}(\xi)) - (\xi)\Big(\frac{i}{2\pi}\frac{d}{d\xi}\hat{s}(\xi)\Big)$$
$$= \frac{i}{2\pi}\Big(\hat{s}(\xi) + \xi\frac{d\hat{s}}{d\xi} - \xi\frac{d\hat{s}}{d\xi}\Big) = \frac{i}{2\pi}\xi\frac{d\hat{s}}{d\xi}$$
$$\Rightarrow [\Im,\ \Xi] = \frac{i}{2\pi}$$

3.4 Time-frequency Localization

Let us consider a unit energy signal $s(t) \in L_2(\mathbb{R})$, i.e., $\int_{-\infty}^{\infty} dt \mid s(t) \mid^2 = 1$ and $\int_{-\infty}^{\infty} d\xi \mid \hat{s}(\xi) \mid^2 = 1$. In line with the concepts in Quantum Mechanics, $\mid s(t) \mid^2$ and $\mid \hat{s}(\xi) \mid^2$ are the probability densities of time and frequency localization, respectively. By appropriate time translation and frequency modulation, it follows that

$$\int_{-\infty}^{\infty} dt \mid s(t) \mid^2 t = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} d\xi \mid \hat{s}(\xi) \mid^2 \xi = 0$$

without loss of generality. Let us define the so-called variances of time localization and frequency localization as:

$$\sigma_t^2 \triangleq \int_{-\infty}^{\infty} dt \mid s(t) \mid^2 t^2 \quad \text{and} \quad \sigma_\xi^2 \triangleq \int_{-\infty}^{\infty} d\xi \mid \hat{s}(\xi) \mid^2 \xi^2$$

Theorem 3.2. *(Heisenberg Uncertainty for Time-frequency Localization) If the signal $s(t) \in L_2(\mathbb{R})$ decays to zero faster than $|t|^{-\frac{1}{2}}$ as $t \to \pm\infty$, then $\sigma_t \sigma_\xi \geq \frac{1}{4\pi}$ and the equality holds for Gaussian signals $s(t) = \sqrt{\frac{\alpha}{\pi}} \exp(-\alpha t^2)$.*

Proof. By Cauchy-Schwarz inequality

$$\Big| \int_{-\infty}^{\infty} dt \left(t\ s(t) \frac{ds}{dt} \right) \Big|^2 \leq \int_{-\infty}^{\infty} dt \left| t\ s(t) \right|^2 \int_{-\infty}^{\infty} dt \left| \frac{ds}{dt} \right|^2.$$

The first integral on the right is equal to σ_t^2.Taking Fourier transform of $\frac{ds}{dt} = i2\pi\xi\hat{s}(\xi)$ and by using Plancherel Theorem, i.e., $\int_{-\infty}^{\infty} dt \left| s(t) \right|^2 = \int_{-\infty}^{\infty} d\xi \left| \hat{s}(\xi) \right|^2$, the second integral on the right is expressed as:

$$\int_{-\infty}^{\infty} dt \left| \frac{ds}{dt} \right|^2 = \int_{-\infty}^{\infty} d\xi \left| i2\pi\xi\hat{s}(\xi) \right|^2 = (2\pi)^2 \sigma_\xi^2$$

Therefore, $\Big| \int_{-\infty}^{\infty} dt \left(t\ s(t) \frac{ds}{dt} \right) \Big| \leq 2\pi\sigma_t\sigma_\xi$. Now, integration by parts yields

$$\int_{-\infty}^{\infty} dt \left(t\ s(t) \frac{ds}{dt} \right) = \frac{1}{2} \int_{-\infty}^{\infty} dt \Big(t\ \frac{d\ |s(t)|^2}{dt} \Big)$$

$$= \frac{1}{2} t\ |s(t)|^2 \Big|_{-\infty}^{\infty} - \frac{1}{2} \int_{-\infty}^{\infty} dt\ |s(t)|^2 = 0 - \frac{1}{2} = -\frac{1}{2} \quad \Rightarrow \sigma_t \sigma_\xi \geq \frac{1}{4\pi}$$

The Cauchy-Schwarz inequality becomes an equality if $ts(t)$ and $\frac{ds}{dt}$ are collinear. For Gaussian signals,

$$\frac{ds}{dt} = \frac{d}{dt} \Big(\sqrt{\frac{\alpha}{\pi}} \exp(-\alpha t^2) \Big)$$

$$= -2\alpha t \sqrt{\frac{\alpha}{\pi}} \exp(-\alpha t^2) = -2\alpha t s(t)$$

Hence, $\sigma_t \sigma_\xi = \frac{1}{4\pi}$ for $s(t) = \sqrt{\frac{\alpha}{\pi}} \exp(-\alpha t^2)$. □