Lecture 11: Orthogonality in Hilbert Spaces and Adjoint Operators

We introduce the idea of a coordinate system in a Hilbert space $\mathcal{H}$.

### 3.5.2.1 Orthonormal Basis in Hilbert Space

An **orthonormal set** in $\mathcal{H}$ is a set $\Psi = \{\psi_1, \psi_2, \ldots\}$ such that $\|\psi_i\| = 1$, $\forall i$, and $\psi_i \perp \psi_j$, $\forall i \neq j$.

The set $\Psi$ is a complete orthonormal set or **orthonormal basis** if $\mathcal{H} = \mathcal{L}(\Psi)$ (smallest subspace containing $\Psi$ is the space itself.)

This is equivalent to saying that if $x \in \mathcal{H}$, then there exists a sequence of real numbers $\{\alpha_i\}$ such that

$$\|x - x_n\| \to 0,$$

where $x_n = \sum_{i=1}^{n} \alpha_i \psi_i$.

**Note:**

By the Schwarz Inequality,

$$\left| \langle \psi_j, x - x_n \rangle \right| \leq \|\psi_j\| \|x - x_n\| = \|x - x_n\|.$$

But, for $n > j$, $\langle \psi_j, x - x_n \rangle = \alpha_j$, so that

$$\left| \langle \psi_j, x - x_n \rangle \right| = \left| \langle \psi_j, x \rangle - \alpha_j \right| \leq \|x - x_n\|,$$

which implies $\alpha_j = \langle \psi_j, x \rangle$ as $n \to \infty$. Thus, each $x \in \mathcal{H}$ has the representation

$$x = \sum_{i=1}^{\infty} \alpha_i \psi_i = \sum_{i=1}^{\infty} \langle x, \psi_i \rangle \psi_i.$$

This series expansion in orthogonal vectors is a generalization of the Fourier series.

**Proposition:**

If $\Psi$ is an orthonormal basis of $\mathcal{H}$ and given $x, y \in \mathcal{H}$, then

$$\langle x, y \rangle = \sum_{i=1}^{\infty} \langle x, \psi_i \rangle \langle y, \psi_i \rangle.$$
In particular,

$$\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, \psi_i \rangle|^2.$$ (Parseval's equality)

Proof:

Second part first. Suppose $x_n \to x \in \mathcal{H}$. Then, by the continuity of the inner product, $\|x_n\| \to \|x\|$ and

$$\|x_n\|^2 = \left\langle \sum_{i=1}^{n} \alpha_i \psi_i, \sum_{j=1}^{n} \alpha_j \psi_j \right\rangle = \sum_{j=1}^{n} |\alpha_j|^2$$

$$\Rightarrow \sum_{j=1}^{n} |\alpha_j|^2 = \sum_{j=1}^{\infty} |\langle x, \psi_j \rangle|^2 \to \|x\|^2$$

For the second part, assume the inner product is real-valued, and consider the identity:

$$\langle x, y \rangle = \frac{1}{4} \left( \|x + y\|^2 - \|x - y\|^2 \right),$$

so that

$$\langle x, y \rangle = \frac{1}{4} \left( \sum_{j=1}^{\infty} \langle x + y, \psi_j \rangle^2 - \sum_{j=1}^{\infty} \langle x - y, \psi_j \rangle^2 \right)$$

$$= \frac{1}{4} \left( \sum_{j=1}^{\infty} \left( \langle x, \psi_j \rangle + \langle y, \psi_j \rangle \right)^2 - \sum_{j=1}^{\infty} \left( \langle x, \psi_j \rangle - \langle y, \psi_j \rangle \right)^2 \right)$$

$$= \frac{1}{4} \left( \sum_{j=1}^{\infty} \left( \langle x, \psi_j \rangle + \langle y, \psi_j \rangle \right)^2 - \left( \langle x, \psi_j \rangle - \langle y, \psi_j \rangle \right)^2 \right)$$

$$= \frac{1}{4} \sum_{j=1}^{\infty} \left( 2 \langle x, \psi_j \rangle \langle y, \psi_j \rangle + 2 \langle x, \psi_j \rangle \langle y, \psi_j \rangle \right)$$

$$= \sum_{j=1}^{\infty} \langle x, \psi_j \rangle \langle y, \psi_j \rangle$$

3.5.2.2 Dense Subset of a Metric Space

Let $(\mathcal{X}, d(\bullet, \bullet))$ be a metric space. Then a subset $\mathcal{A} \subseteq \mathcal{X}$ is dense if any vector in $\mathcal{X}$ can be approximated arbitrarily closely by vectors in $\mathcal{A}$, i.e., if for any $\varepsilon > 0$, there exists $a \in \mathcal{A}$ such that $d(a, x) < \varepsilon$.

$\mathcal{X}$ is called separable if it has a countable dense subset $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots\}$.

Notes:
(a) If $\mathcal{B}$ is any subset of a Hilbert space $\mathcal{H}$ then any $x \in \mathcal{H}$ has the decomposition

$$x = y + z, \ y \in \mathcal{L}(\mathcal{B}), \ z \perp \mathcal{L}(\mathcal{B}),$$

and for any $b \in \mathcal{B}$, $d(b, x) \geq \|z\|$. It follows that:

$$B' := \left\{ x \in \mathcal{H} : x = \sum_{i=1}^{n} \alpha_i b_i, \ \text{for some} \ \alpha_i, b_i \in \mathbb{R}, \ b_i \in \mathcal{B} \right\}$$

is dense iff $x \perp b, \ \forall b \in B' \Rightarrow x = 0$. Note that $B'$ is not countable, but could be made countable by using rational coefficients $\alpha_i$.

(b) Any countable set in $\mathcal{H}$ can be orthogonalized using the Gram-Schmidt procedure to produce a sequence $\{x_n\}$ such that $\|x_n\| = 1, \ x_i \perp x_j, \ \forall i \neq j$

Examples:

The basic example of a Hilbert space possessing an orthonormal basis is $L^2_2[0,1]$.

(a) The Haar functions $f_0, f_{k/2^n}, \ k = 1, 3, \ldots, 2^n - 1, \ n = 1, 2, 3, \ldots$ defined over $t \in [0,1]$ by:

$$f_0(t) = \begin{cases} 1 & \text{for} \ t = 0 \\ \frac{2^{(n-1)/2}}{2^n} & \text{for} \ (k-1)2^{-n} \leq t < k2^{-n} \\ -\frac{2^{(n-1)/2}}{2^n} & \text{for} \ k2^{-n} \leq t < (k+1)2^{-n} \\ 0 & \text{otherwise} \end{cases}$$

form an orthonormal basis of $L^2_2[0,1]$. The Haar functions are actually wavelets that can be used for multiresolution analysis.
(b) The trigonometric functions \( \varphi_0, \varphi_1, \varphi_2, \ldots, \psi_1, \psi_2, \ldots \) defined by:

\[
\varphi_0(t) = 1, \quad \varphi_k(t) = \sqrt{2} \sin 2\pi k t, \quad \psi_k(t) = \sqrt{2} \cos 2\pi k t
\]

form an orthonormal basis of \( L^2_2[0,1] \). This basis is used in Fourier series.

### 3.5.3 Adjoint Operators

#### 3.5.3.1 Definition of the Adjoint

Let \((U, \langle \cdot, \cdot \rangle_U)\), \((V, \langle \cdot, \cdot \rangle_V)\) be inner product spaces, and let \( A : U \rightarrow V, \ u \mapsto Au = v \) be a linear transformation. Then \( A^* : V \rightarrow U \) is the adjoint of \( A \) if

\[
\langle Au, v \rangle_V = \langle u, A^* v \rangle_U, \ \forall u \in U, \ \forall v \in V.
\]

**Examples:**

(a) If \( U, V \) are finite-dimensional, then the inner product coincides with the scalar product, from which it is easy to show that \( A^* \) always exists, and is given by the matrix \( \overline{A^T} = A^* \):

\[
\langle Au, v \rangle_V = (Au)^T v = \overline{u}^T \overline{A^T} v = \langle u, \overline{A^T} v \rangle_U = \langle u, A^* v \rangle_U, \ \forall u \in U, \ \forall v \in V
\]

\[\Rightarrow A^* = \overline{A^T} = A^*\]

In the finite-dimensional case, \( A^* \) is unique and \( (A^*)^* = A \), but in the infinite-dimensional case, \( (A^*)^* \) and \( A \) do not always coincide.

(b) Let \( L : l_2[t_0, t_{k-1}] \rightarrow \mathbb{R}^n, \ L u = \sum_{i=0}^{k-1} \Phi(t_k, t_{i+1}) B(t_i) u(t_i) \)

This is the controllability operator, or zero-state response of the discrete-time linear time-varying system \( x(t_{j+1}) = A(t_j) x(t_j) + B(t_j) u(t_j) \). We regard the domain and codomain of \( L \) as being equipped with inner products which are simply the real scalar products.

\[
\langle Lu, z \rangle_{\mathbb{R}^n} = \left( \sum_{i=0}^{k-1} \Phi(t_k, t_{i+1}) B(t_i) u(t_i) \right)^T z = \sum_{i=0}^{k-1} (u(t_i))^T B(t_i)^T \Phi(t_k, t_{i+1})^T z
\]

\[
= \langle u(\bullet), B(\bullet)^T \Phi(t_k, \bullet+1)^T z \rangle_{l_2[t_0, t_{k-1}]} = \langle u(\bullet), L^c z \rangle_{l_2[t_0, t_{k-1}]} \quad \forall u \in l_2[t_0, t_{k-1}], \forall z \in \mathbb{R}^n
\]

\[\Rightarrow L^c = B(\bullet)^T \Phi(t_k, \bullet+1)^T\]

Here, \( L^c \) is a function in \( l_2[t_0, t_{k-1}] \).
3.5.3.2 Structure of the Adjoint

Let \( \mathcal{A} : \mathcal{U} \to \mathcal{V} \) be a linear transformation. From Chapter 2,

\[
\mathcal{U} = \mathcal{D}(\mathcal{A}) = \mathcal{N}(\mathcal{A}) \oplus \mathcal{N}(\mathcal{A})^c = \mathcal{N}(\mathcal{A}) \oplus \mathcal{D}(\mathcal{A})/\mathcal{N}(\mathcal{A})
\]

\[
\mathcal{V} = \mathcal{C}(\mathcal{A}) = \mathcal{R}(\mathcal{A}) \oplus \mathcal{R}(\mathcal{A})^c
\]

If \( \mathcal{U}, \mathcal{V} \) are Hilbert spaces, then we can use the orthogonal complements:

\[
\mathcal{U} = \mathcal{D}(\mathcal{A}) = \mathcal{N}(\mathcal{A}) \oplus \mathcal{N}(\mathcal{A})^\perp
\]

\[
\mathcal{V} = \mathcal{C}(\mathcal{A}) = \mathcal{R}(\mathcal{A}) \oplus \mathcal{R}(\mathcal{A})^\perp
\]

where \( \mathcal{N}(\mathcal{A})^\perp \), \( \mathcal{R}(\mathcal{A})^\perp \) are unique. Now the claim is that

\[
\mathcal{N}(\mathcal{A})^\perp = \mathcal{R}(\mathcal{A}^*)
\]

\[
\mathcal{R}(\mathcal{A})^\perp = \mathcal{N}(\mathcal{A}^*)
\]

where \( \mathcal{A}^* : \mathcal{V} \to \mathcal{U} \) is the adjoint of \( \mathcal{A} \).
Theorem:

Let $\mathcal{A} : \mathcal{U} \rightarrow \mathcal{V}$ be a linear transformation, where $\mathcal{U}, \mathcal{V}$ are finite-dimensional inner product spaces (e.g., Hilbert spaces). Then,

(a) $\mathcal{U} = \mathcal{N}\{\mathcal{A}\}^\perp \oplus \mathcal{R}\{\mathcal{A}^*\}$

(b) $\mathcal{V} = \mathcal{R}\{\mathcal{A}\}^\perp \oplus \mathcal{N}\{\mathcal{A}^*\}$

Proof:

(a) $x \in \mathcal{R}\{\mathcal{A}^*\}^\perp \iff \langle x, \mathcal{A}^* y \rangle_{\mathcal{U}} = 0, \ \forall y \in \mathcal{V}$

$\iff \langle \mathcal{A}x, y \rangle_{\mathcal{V}} = 0, \ \forall y \in \mathcal{V}$

$\iff \mathcal{A}x = 0$

$\iff x \in \mathcal{N}\{\mathcal{A}\}$

Hence, $\mathcal{N}\{\mathcal{A}\} = \mathcal{R}\{\mathcal{A}^*\}^\perp$ and $\mathcal{U} = \mathcal{N}\{\mathcal{A}\} \oplus \mathcal{R}\{\mathcal{A}^*\}$.

(b) $y \in \mathcal{R}\{\mathcal{A}\}^\perp \iff \langle y, \mathcal{A}x \rangle_{\mathcal{V}} = 0, \ \forall x \in \mathcal{U}$

$\iff \langle \mathcal{A}^* y, x \rangle_{\mathcal{U}} = 0, \ \forall x \in \mathcal{U}$

$\iff \mathcal{A}^* y = 0$

$\iff y \in \mathcal{N}\{\mathcal{A}^*\}$

Hence, $\mathcal{N}\{\mathcal{A}^*\} = \mathcal{R}\{\mathcal{A}\}^\perp$ and $\mathcal{V} = \mathcal{N}\{\mathcal{A}^*\} \oplus \mathcal{R}\{\mathcal{A}\}$.

3.5.3.3 Solving $\mathcal{A}u = v$ Using the Adjoint

Consider the linear equation:

$\mathcal{A}u = v,$

where $\mathcal{A} : \mathcal{U} \rightarrow \mathcal{V}$ is a linear transformation, where $\mathcal{U}, \mathcal{V}$ are finite-dimensional inner product spaces of dimension $n$ and $m$, respectively. Operator $\mathcal{A}$ has a matrix representation $\mathcal{A}$ under the chosen bases for $\mathcal{U}$ and $\mathcal{V}$, so we can write $\mathcal{A}u = v$.

We seek test for $\mathcal{A}$ to be onto, and if it is, an explicit formula for solutions $u$. 
Theorem:

\[ A : U \rightarrow V \text{ is onto } V \]

(a) \( \Leftrightarrow \) columns of \( A^* \) are linearly independent,

(b) \( \Leftrightarrow \) columns of \( A^* \) are linearly independent,

(c) \( \Leftrightarrow (AA^*)^{-1} \) exists

Proof:

(a) \( A \) is onto:

\[ \Leftrightarrow \mathcal{R}\{A\} = V \]
\[ \Leftrightarrow \dim\{\mathcal{R}\{A\}\} = \dim\{V\} = \# \text{ independent rows of } A \]
\[ = \# \text{ independent columns of } A^* \]
\[ \Leftrightarrow \text{all columns of } A^* \text{ are linearly independent} \]

(b) \( AA^* : V \rightarrow V \) and matrix \( AA^* \) is square. Thus,

\[ (AA^*)^{-1} \text{ exists} \Leftrightarrow \mathcal{N}\{AA^*\} = 0 \]
\[ \Leftrightarrow \mathcal{N}\{A^*\} = 0 \text{ (shown below)} \]
\[ \Leftrightarrow A \text{ is onto (by the structure theory)} \]

Now we prove the claim \( \mathcal{N}\{AA^*\} = 0 \Leftrightarrow \mathcal{N}\{A^*\} = 0 : \)

\[ \mathcal{N}\{AA^*\} := \{ y \in V : AA^*y = \theta \} = \{ y \in V : A^*y = \theta \} = \mathcal{N}\{A^*\} \]

Now, we can state the result providing a solution to \( Au = v \) when \( A \) is onto.

Theorem:

If \( A : U \rightarrow V \) is onto \( V \), then \( u_1 = A^*(AA^*)^{-1}v \) is one solution of \( Au = v \).
Proof:

If $A$ is onto, then $(AA^*)^{-1}$ exists by the above theorem. Now, $u_i = A^* (AA^*)^{-1} v$ is a solution to $Au = v$ as $Au_i = AA^* (AA^*)^{-1} v = v$.

With this solution at hand, recall that all solutions to $Au = v$ lie in the coset $u_i + N\{A\}$. Note that in many cases, the solution $u_i = A^* (AA^*)^{-1} v$ also holds in infinite-dimensional Hilbert spaces.

### 3.5.3.4 Moore-Penrose Pseudoinverse

If $A$ is onto, then its Moore-Penrose pseudoinverse is defined as $A^\odot := A^* (AA^*)^{-1}$.

**Notes:**

1. $A^\odot$ is a right inverse, i.e., $AA^\odot = AA^* (AA^*)^{-1} = I$.

2. $A^\odot A = A^* (AA^*)^{-1} A \neq I$, unless $A$ is one-to-one.

3. If $A^{-1}$ exists, then $A^{-1} = A^\odot$.

### 3.5.3.5 Quadratic Optimality of the Pseudoinverse

Let $(\mathcal{U}, \|\cdot\|)$ be a normed linear space with $\mathcal{U}$ finite-dimensional and $\|\cdot\|$ the inner-product-induced Euclidean norm. Then, we will show that the Moore-Penrose pseudoinverse solution $u_i = A^\odot v$ minimizes $\|u_i\|$.

However, if another norm on $\mathcal{U}$ is used, say the 1-norm $\|u\|_1 = \sum_{i=1}^{n} |u_i|$, then this norm is not minimized.

The optimality of $u_i = A^\odot v$ holds true in a more general setting with $A : \mathcal{U} \to \mathcal{V}$ a bounded linear transformation, and $\mathcal{U}, \mathcal{V}$ Hilbert spaces (can be infinite-dimensional) provided an adjoint $A^*$ can be found. Here is the result establishing optimality of the pseudoinverse solution in Euclidean space.
Theorem:

The solution \( u_1 = A^* (AA^*)^{-1} v \) is the unique minimum-norm solution of \( Au = v \) in Euclidean space.

Proof:

If \( u_2 \) is any other solution, then \( u_2 - u_1 \in \mathcal{N}\{A\} \). Also, \( u_1 \in \mathcal{R}\{A^*\} \) since \( A^* \) is the last transformation in the expression \( A^* (AA^*)^{-1} \). Now, \( \mathcal{N}\{A\} \perp \mathcal{R}\{A^*\} \) by structure, it must be true that \( u_2 - u_1 \perp u_1 \). The conclusion that \( \|u_2\| \geq \|u_1\| \) follows from the Pythagorean theorem: Let \( \Delta u := u_2 - u_1 \). Then,

\[
\|u_2\|^2 = \|u_1 + \Delta u\|^2 = \langle u_1 + \Delta u, u_1 + \Delta u \rangle \\
= \langle u_1, u_1 \rangle + \langle \Delta u, u_1 \rangle + \langle u_1, \Delta u \rangle + \langle \Delta u, \Delta u \rangle \\
= \langle u_1, u_1 \rangle + \langle \Delta u, \Delta u \rangle = \|u_1\|^2 + \|\Delta u\|^2 \geq \|u_1\|^2
\]

Thus, \( u_1 \in \mathcal{U} \) is smaller in norm than any other solution \( u_2 \in \mathcal{U} \). Vector \( u_1 \) is the unique smallest solution because the last inequality is strict unless \( \Delta u = 0 \). 

\[\blacksquare\]