6.342 Lecture 2 — February 9, 2009

Today:
• Standard spaces of sequences and functions
• Bases
• Linear operators
• Projection theorem and its applications

Homework #1: Due Wednesday, February 18

Recall: No lecture Wednesday, February 11
Monday, February 16 is Presidents Day (MIT holiday)
Lectures on Tuesday, Feb 17 and Wednesday, Feb 18

Readings:
• Chapters 0-1 of The World of Fourier and Wavelets
• Full book “release” delayed, so Chapter 2 will be posted in a few days

Last time

Definition 1.1 (Vector space). A vector space over the set of real numbers $\mathbb{R}$ (or the set of complex numbers $\mathbb{C}$) is a set of vectors, $V$, together with vector addition and scalar multiplication operations. For any $x, y, z$ in $V$ and $\alpha, \beta$ in $\mathbb{R}$ (or $\mathbb{C}$), these operations must satisfy the following properties:

(i) Commutativity: $x + y = y + x$.
(ii) Associativity: $(x + y) + z = x + (y + z)$ and $(\alpha \beta)x = \alpha(\beta x)$.
(iii) Distributivity: $\alpha(x + y) = \alpha x + \alpha y$ and $(\alpha + \beta)x = \alpha x + \beta x$.

Furthermore, the following hold:

(iv) Additive identity: There exists an element $0$ in $V$, such that $x + 0 = 0 + x = x$, for every $x$ in $V$.
(v) Additive inverse: For every $x$ in $V$, there exists a unique element $-x$ in $V$, such that $x + (-x) = (-x) + x = 0$.
(vi) Multiplicative identity: For every $x$ in $V$, $1 \cdot x = x$. 


**Last time**

**Definition 1.6 (Inner Product).** An inner product on a vector space $V$ over $\mathbb{R}$ (or $\mathbb{C}$) is a real-valued (or complex-valued) function $\langle \cdot, \cdot \rangle$ defined on $V \times V$ with the following properties for any $x, y, z \in V$ and $\alpha \in \mathbb{R}$ (or $\mathbb{C}$):

1. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.
2. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$.
3. $\langle x, y \rangle^* = \langle y, x \rangle$.
4. $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.

**Definition 1.7 (Norm).** Given a vector space $V$ over $\mathbb{R}$ (or $\mathbb{C}$), a norm $\| \cdot \|$ is a function mapping $V$ into $\mathbb{R}$ with the following properties:

1. $\|x\| \geq 0$ for any $x$ in $V$, and $\|x\| = 0$ if and only if $x = 0$.
2. $\|\alpha x\| = |\alpha| \|x\|$ for any $x$ in $V$ and $\alpha$ in $\mathbb{R}$ (or $\mathbb{C}$).
3. $\|x + y\| \leq \|x\| + \|y\|$ for any $x, y$ in $V$, with equality if and only if $y = \alpha x$.

*Distances Not Necessarily Induced by Norms*

A distance or metric $d : V \times V \to \mathbb{R}$ is a function with the following properties:

1. Nonnegativity: $d(x, y) \geq 0$ for every $x, y$ in $V$.
2. Symmetry: $d(x, y) = d(y, x)$ for every $x, y$ in $V$.
3. Triangle Inequality: $d(x, y) + d(y, z) \geq d(x, z)$ for every $x, y, z$ in $V$.
4. $d(x, x) = 0$ and $d(x, y) = 0$ implies $x = y$.

**Last time**

**Definition 1.11 (Convergence in normed vector spaces).** Let $x_1, x_2, \ldots$ be a sequence in a normed vector space $V$. The sequence is said to converge to $x \in V$ if $\lim_{n \to \infty} \|x - x_n\| = 0$. In other words: Given any $\epsilon > 0$, there exists $N$ such that

$$\|x - x_n\| < \epsilon \quad \text{for all } n \geq N.$$

**Definition 1.12 (Cauchy sequence).** The sequence $x_1, x_2, \ldots$ in a normed vector space is called a Cauchy sequence if: Given any $\epsilon > 0$, there exists $N$ such that

$$\|x_n - x_m\| < \epsilon \quad \text{for all } n, m \geq N.$$

**Definition 1.13 (Completeness, Banach space and Hilbert space).** In a normed vector space $V$, if every Cauchy sequence converges to a vector in $V$, then $V$ is said to be complete and is called a Banach space. A complete inner product space is called a Hilbert space.
Types of vector spaces and examples

Vector spaces

- Inner product spaces
- Banach spaces
- Normed vector spaces
- Hilbert spaces

Types of vector spaces:
- \( \mathbb{Q}^N \)
- \( \mathbb{R}^N \)
- \( \mathbb{C}^N \)
- \( C([a, b]) \)
- \( L^2(\mathbb{R}) \)
- \( L^1(\mathbb{R}) \)
- \( L^\infty(\mathbb{R}) \)
- \( \ell^1(\mathbb{Z}) \)
- \( \ell^2(\mathbb{Z}) \)
- \( \ell^\infty(\mathbb{Z}) \)

Sequence spaces

**\( \ell^p \) Spaces**

For sequences with domain \( \mathbb{Z} \), define the \( \ell^p \)-norm, for \( p \in [1, \infty) \), as

\[
\|x\|_p = \left( \sum_{k \in \mathbb{Z}} |x_k|^p \right)^{1/p}
\]

and the \( \ell^\infty \)-norm as

\[
\|x\|_\infty = \sup_{k \in \mathbb{Z}} |x_k|.
\]

As these norms are not necessarily finite; to use them we must define our vector space to include only those sequences for which the norm is well defined:

**Definition 1.9 (\( \ell^p(\mathbb{Z}) \)).** For any \( p \in [1, \infty] \), the space \( \ell^p(\mathbb{Z}) \) is the vector space of all complex-valued sequences \((x_i)_{i \in \mathbb{Z}}\) with finite \( \ell^p \)-norm.

Each of these is complete (hence a Banach space)
Function spaces

$L^p$ Spaces

For functions, define the $L^p$-norm, for $p \in [1, \infty)$, as

$$\|x\|_p = \left( \int_{t \in \mathbb{R}} |x(t)|^p \, dt \right)^{1/p},$$

and the $L^\infty$-norm as

$$\|x\|_\infty = \text{ess sup}_{t \in \mathbb{R}} |x(t)|.$$  \hspace{1cm} (1.27b)

As with the $L^p$-norms, finiteness of the norm is an issue, so we define vector spaces with the requirement of finite $L^p$-norm:

**Definition 1.10 ($L^p(\mathbb{R})$).** For any $p \in [1, \infty]$, the space $L^p(\mathbb{R})$ is the vector space of all complex-valued functions $x(t)$, $t \in \mathbb{R}$, with finite $L^p$-norm.

Completeness:
- Functions are measurable and integrals are Lebesgue integrals; or
- define spaces as completion of continuous functions under $L^p$ norm.

Function spaces

$C^p([a, b])$ Spaces

For any finite $a$ and $b$, the space $C([a, b])$ is defined as the inner product space of all complex-valued continuous functions on $[a, b]$, over the complex numbers, with pointwise addition and scalar multiplication. The inner product is

$$\langle x, y \rangle = \int_a^b x(t)y^*(t) \, dt,$$  \hspace{1cm} (1.29)

which induces the norm

$$\|x\| = \left( \int_a^b |x(t)|^2 \, dt \right)^{1/2}. \hspace{1cm} (1.30)$$

$C([a, b])$ is also called $C^0([a, b])$ because $C^p([a, b])$ is defined similarly but with the additional requirement that the functions have $p$ continuous derivatives.
Types of vector spaces and examples

Vector spaces

Normed vector spaces

Inner product spaces

Banach spaces

• $\mathbb{Q}^N$
• $\mathbb{R}^N$
• $\mathbb{C}^N$
• $C([a, b])$
• $L^2(\mathbb{R})$
• $C((a, b], \| \cdot \|_\infty)$
• $\ell^1(\mathbb{Z})$
• $L^\infty(\mathbb{R})$
• $L^1(\mathbb{R})$
• $\ell^0(\mathbb{Z})$
• $\ell^\infty(\mathbb{Z})$

Hilbert spaces

Defining bases

**Definition 1.2 (Subspace).** A subset $W$ of a vector space $V$ is a subspace if:

(i) For all $x$ and $y$ in $W$, $x + y$ is in $W$.

(ii) For all $x$ in $W$ and $\alpha$ in $\mathbb{R}$ (or $\mathbb{C}$), $\alpha x$ is in $W$.

**Definition 1.3 (Span).** The span of a set of vectors $S \subseteq V$ is a set of all finite linear combinations of vectors in $S$:

$$\text{span}(S) = \left\{ \sum_{k=1}^{N} \alpha_k s_k \mid \alpha_k \in \mathbb{R} \text{ (or } \mathbb{C}) \text{, } s_k \in S \text{ and } N \in \mathbb{N} \right\}.$$

**Definition 1.4 (Linear Independence).** The set $\{s_1, s_2, \ldots, s_N\} \subset V$ is called linearly independent when $\sum_{k=1}^{N} \alpha_k s_k = 0$ is true only if $\alpha_k = 0$ for all $k$. Otherwise, these vectors are linearly dependent. An infinite set of vectors is called linearly independent when every finite subset is linearly independent.

**Definition 1.5 (Basis).** The set $S \subset V$ is a basis for $V$ when $V = \text{span}(S)$ and $S$ is linearly independent. The number of elements in $S$ is the *dimension* of $V$ and may be infinite.
Interlude – Beware of Multiple “Definitions”

• It’s possible to own too much. A man with one watch knows what time it is; a man with two watches is never sure.
  — Lee Segall, creator of the original Dr. I.Q. tv/radio program

• There are several definitions of bases (Riesz basis, Schauder basis, Hamel basis, …). We won’t need subtle distinctions (for now).

Technicality: Separability

A space $X$ is called **separable** if it contains a countable dense subset.

Example: $\mathbb{R}$ is separable because of $\mathbb{Q} \subset \mathbb{R}$.

A Hilbert space is separable if and only if it contains a countable basis.

All Hilbert spaces we consider are separable.
Orthonormal bases

**Theorem 1.5.** Given an orthonormal set \( \{ \varphi_k \}_{k \in \mathcal{I}} \) in \( H \), the following are equivalent:

(i) The set of vectors \( \{ \varphi_k \}_{k \in \mathcal{I}} \) is an ONB for \( H \).

(ii) If \( \langle x, \varphi_k \rangle = 0 \) for all \( k \), then \( x = 0 \).

(iii) The \( \text{span}(\{ \varphi_k \}_{k \in \mathcal{I}}) \) is dense in \( H \), that is, every vector in \( H \) is a limit of a sequence of vectors in \( \text{span}(\{ \varphi_k \}_{k \in \mathcal{I}}) \).

(iv) For every \( x \) in \( H \), Parseval’s equality holds:

\[
\| x \|^2 = \sum_{k \in \mathcal{I}} | \langle x, \varphi_k \rangle |^2. \tag{1.44a}
\]

(v) For every \( x_1 \) and \( x_2 \) in \( H \), the generalized Parseval’s equality holds:

\[
\langle x_1, x_2 \rangle = \sum_{k \in \mathcal{I}} \langle x_1, \varphi_k \rangle \langle x_2, \varphi_k \rangle^*. \tag{1.44b}
\]

Biorthogonal bases

**Definition 1.2.** (Biorthogonal bases). Sets \( \{ \varphi_k \} \) and \( \{ \tilde{\varphi}_k \} \) in an Hilbert space \( H \) constitute a biorthogonal basis pair of \( H \) if:

(i) For all \( k, i \) in \( \mathbb{Z} \),

\[
\langle \varphi_k, \tilde{\varphi}_i \rangle = \delta_{k-i}. \tag{1.46a}
\]

(ii) There exist strictly positive constants \( A, B \), such that, for all \( x \) in \( H \),

\[
A \| x \|^2 \leq \sum_k | \langle x, \varphi_k \rangle |^2 \leq B \| x \|^2, \tag{1.46b}
\]

\[
\frac{1}{B} \| x \|^2 \leq \sum_k | \langle x, \tilde{\varphi}_k \rangle |^2 \leq \frac{1}{A} \| x \|^2. \tag{1.46c}
\]
A linear operator $A : V_1 \to V_2$ is a mapping between vector spaces over the same field $\mathbb{F}$ that satisfies

1. $A(x + y) = Ax + Ay$ for all $x, y \in V_1$; and
2. $A(\alpha x) = \alpha(Ax)$ for all $x \in V_1$ and $\alpha \in \mathbb{F}$

The (operator) norm of $A$ is $\|A\| = \sup_{x: \|x\| = 1} \|Ax\|

Examples for $\mathbb{R}^n \to \mathbb{R}^m$ or $\mathbb{C}^n \to \mathbb{C}^m$:

<table>
<thead>
<tr>
<th>Domain</th>
<th>$\ell_1$ norm of a column</th>
<th>$\ell_2$ norm of a column</th>
<th>$\ell_\infty$ max absolute entry of matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell_1$</td>
<td>Maximum</td>
<td>Maximum</td>
<td>Maximum absolute entry of matrix</td>
</tr>
<tr>
<td>$\ell_2$</td>
<td>NP-hard</td>
<td>Maximum singular value</td>
<td>Maximum $\ell_2$ norm of a row</td>
</tr>
<tr>
<td>$\ell_\infty$</td>
<td>NP-hard</td>
<td>NP-hard</td>
<td>Maximum $\ell_1$ norm of a row</td>
</tr>
</tbody>
</table>

$A$ is called **bounded** if $\|A\| < \infty$

A linear operator is continuous if and only if it is bounded

$A$ is called **invertible** if $\exists A^{-1} : V_2 \to V_1$ s.t.

1. $A^{-1}Ax = x$ for all $x \in V_1$; and
2. $AA^{-1}y = y$ for all $y \in V_2$

The adjoint operator $A^* : V_2 \to V_1$ is such that

$\langle Ax, y \rangle = \langle x, A^*y \rangle$ for all $x \in V_1$ and $y \in V_2$

Finite-dimensional cases: adjoint is Hermetian transpose
Projection operators

Definition 1.19 (Projection operator). An operator \( P \) is called idempotent if \( P^2 = P \). A linear operator that is idempotent is a projection operator.

Definition 1.20 (Orthogonal projection operator). A projection operator that is self-adjoint is an orthogonal projection operator.

Theorem 1.1 (Projection theorem [42]). Let \( W \) be a closed subspace of the Hilbert space \( H \). Corresponding to any vector \( x \in H \), there exists a unique vector \( w_0 \in W \) such that \( \| x - w_0 \| \leq \| x - w \| \) for all \( w \in W \). A necessary and sufficient condition for \( w_0 \) to be the unique minimizing vector is that \( x - w_0 \perp W \). The operator \( P_W \) mapping \( x \) into \( w_0 \) is an orthogonal projection operator.

\[ w_0 = P_W x \]

Matrix representations of vectors

\[ x = \sum_{k \in \mathbb{Z}} \alpha_k \varphi_k \]
\[ y = \sum_{k \in \mathbb{Z}} \beta_k \varphi_k \]
\[ \langle x, y \rangle = \langle \sum_{k \in \mathbb{Z}} \alpha_k \varphi_k, \sum_{i \in \mathbb{Z}} \beta_i \varphi_i \rangle = \sum_{k \in \mathbb{Z}} \sum_{i \in \mathbb{Z}} \alpha_k \beta_i^* \langle \varphi_k, \varphi_i \rangle = \beta^* G \alpha \]

\[ G = \begin{bmatrix} 
... & ... & ... \\
... & \langle \varphi_{-1}, \varphi_{-1} \rangle & \langle \varphi_0, \varphi_{-1} \rangle & \langle \varphi_1, \varphi_{-1} \rangle & ... \\
... & \langle \varphi_{-1}, \varphi_0 \rangle & \langle \varphi_0, \varphi_0 \rangle & \langle \varphi_1, \varphi_0 \rangle & ... \\
... & \langle \varphi_{-1}, \varphi_1 \rangle & \langle \varphi_0, \varphi_1 \rangle & \langle \varphi_1, \varphi_1 \rangle & ... \\
... & ... & ... & ... & ... 
\end{bmatrix} \]
Matrix representations of linear operators

\[
\beta_k = \left\langle A \left( \sum_{i \in \mathbb{Z}} \alpha_i \varphi_i \right), \tilde{\psi}_k \right\rangle \overset{(a)}{=} \left\langle \left( \sum_{i \in \mathbb{Z}} \alpha_i A \varphi_i \right), \tilde{\psi}_k \right\rangle \overset{(b)}{=} \sum_{i \in \mathbb{Z}} \alpha_i \left\langle A \varphi_i, \tilde{\psi}_k \right\rangle
\]

\[
\begin{bmatrix}
\vdots \\
\beta_{-1} \\
\beta_0 \\
\beta_1 \\
\vdots \\
\end{bmatrix}
= 
\begin{bmatrix}
\vdots \\
\left\langle A \varphi_{-1}, \tilde{\psi}_{-1} \right\rangle \\
\left\langle A \varphi_0, \tilde{\psi}_0 \right\rangle \\
\left\langle A \varphi_{-1}, \tilde{\psi}_0 \right\rangle \\
\vdots \\
\end{bmatrix}
= 
\begin{bmatrix}
\alpha_{-1} \\
\alpha_0 \\
\alpha_1 \\
\vdots \\
\end{bmatrix}
\]

Hilbert space of random variables

**Corollary 1.9 (Linear minimum MSE estimation).** Let \( X \) and \( Y^{(1)}, Y^{(2)}, \ldots, Y^{(n)} \) be complex random variables with finite second moments. Then there is a unique estimator of \( X \) from the \( Y^{(k)} \)s of the form

\[
\hat{X} = a_0 + a_1 Y^{(1)} + a_2 Y^{(2)} + \cdots + a_n Y^{(n)}
\]

such that mean-squared error \( \mathbb{E} \left[ (X - \hat{X})^2 \right] \) is minimized. This estimator is determined by the condition that

\[
\mathbb{E} \left[ (X - \hat{X})(b_0 + b_1 Y^{(1)} + b_2 Y^{(2)} + \cdots + b_n Y^{(n)})^* \right] = 0 \quad (1.58a)
\]

for all \( \{b_k\}_{k=0}^n \in \mathbb{C} \). More explicitly, the \( a_k \)s satisfy

\[
\begin{bmatrix}
1 \\
\mathbb{E} [Y^{(1)}] \\
\mathbb{E} [Y^{(2)}] \\
\vdots \\
\mathbb{E} [Y^{(n)}] \\
\end{bmatrix}
\begin{bmatrix}
\mathbb{E} [Y^{(1)*}] \\
\mathbb{E} [Y^{(1)*} Y^{(1)}] \\
\mathbb{E} [Y^{(1)*} Y^{(2)}] \\
\vdots \\
\mathbb{E} [Y^{(1)*} Y^{(n)}] \\
\end{bmatrix}
= 
\begin{bmatrix}
a_0 \\
a_1 \\
a_2 \\
\vdots \\
a_n \\
\end{bmatrix}
= 
\begin{bmatrix}
\mathbb{E} [X] \\
\mathbb{E} [Y^{(1)*} X] \\
\mathbb{E} [Y^{(2)*} X] \\
\vdots \\
\mathbb{E} [Y^{(n)*} X] \\
\end{bmatrix}
\]

(1.58b)