Chapter 6

Linear operators and adjoints

Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>6.1</td>
</tr>
<tr>
<td>Fundamentals</td>
<td>6.2</td>
</tr>
<tr>
<td>Spaces of bounded linear operators</td>
<td>6.3</td>
</tr>
<tr>
<td>Inverse operators</td>
<td>6.5</td>
</tr>
<tr>
<td>Linearity of inverses</td>
<td>6.5</td>
</tr>
<tr>
<td>Banach inverse theorem</td>
<td>6.5</td>
</tr>
<tr>
<td>Equivalence of spaces</td>
<td>6.5</td>
</tr>
<tr>
<td>Isomorphic spaces</td>
<td>6.6</td>
</tr>
<tr>
<td>Isometric spaces</td>
<td>6.6</td>
</tr>
<tr>
<td>Unitary equivalence</td>
<td>6.7</td>
</tr>
<tr>
<td>Adjoint in Hilbert spaces</td>
<td>6.11</td>
</tr>
<tr>
<td>Unitary operators</td>
<td>6.13</td>
</tr>
<tr>
<td>Relations between the four spaces</td>
<td>6.14</td>
</tr>
<tr>
<td>Duality relations for convex cones</td>
<td>6.15</td>
</tr>
<tr>
<td>Geometric interpretation of adjoints</td>
<td>6.15</td>
</tr>
<tr>
<td>Optimization in Hilbert spaces</td>
<td>6.16</td>
</tr>
<tr>
<td>The normal equations</td>
<td>6.16</td>
</tr>
<tr>
<td>The dual problem</td>
<td>6.17</td>
</tr>
<tr>
<td>Pseudo-inverse operators</td>
<td>6.18</td>
</tr>
<tr>
<td>Analysis of the DTFT</td>
<td>6.21</td>
</tr>
</tbody>
</table>

6.1

Introduction

The field of optimization uses linear operators and their adjoints extensively.

Example. differentiation, convolution, Fourier transform, Radon transform, among others.

Example. If $A$ is a $n \times m$ matrix, an example of a linear operator, then we know that $\|y - Ax\|_2$ is minimized when $x = \left[A' A\right]^{-1} A' y$.

We want to solve such problems for linear operators between more general spaces. To do so, we need to generalize “transpose” and “inverse.”
6.2
Fundamentals

We write \( T : D \to Y \) when \( T \) is a transformation from a set \( D \) in a vector space \( X \) to a vector space \( Y \).

The set \( D \) is called the domain of \( T \). The range of \( T \) is denoted
\[
R(T) = \{ y \in Y : y = T(x) \text{ for } x \in D \}.
\]

If \( S \subseteq D \), then the image of \( S \) is given by
\[
T(S) = \{ y \in Y : y = T(s) \text{ for } s \in S \}.
\]

If \( P \subseteq Y \), then the inverse image of \( P \) is given by
\[
T^{-1}(P) = \{ x \in D : T(x) \in P \}.
\]

Notation: for a linear operator \( A \), we often write \( Ax \) instead of \( A(x) \).

For linear operators, we can always just use \( D = X \), so we largely ignore \( D \) hereafter.

**Definition.** The nullspace of a linear operator \( A \) is
\[
N(A) = \{ x \in X : Ax = 0 \}.
\]

It is also called the kernel of \( A \), and denoted \( \ker(A) \).

**Exercise.** For a linear operator \( A \), the nullspace \( N(A) \) is a subspace of \( X \).

Furthermore, if \( A \) is continuous (in a normed space \( X \)), then \( N(A) \) is closed [3, p. 241].

**Exercise.** The range of a linear operator is a subspace of \( Y \).

**Proposition.** A linear operator on a normed space \( X \) (to a normed space \( Y \)) is continuous at every point \( X \) if it is continuous at a single point in \( X \).

**Proof.** Exercise. [3, p. 240].

**Luenberger does not mention that \( Y \) needs to be a normed space too.**

**Definition.** A transformation \( T \) from a normed space \( X \) to a normed space \( Y \) is called bounded iff there is a constant \( M \) such that \( \| T(x) \| \leq M \| x \| \), \( \forall x \in X \).

**Definition.** The smallest such \( M \) is called the norm of \( T \) and is denoted \( \| T \| \).

Formally:
\[
\| T \| \triangleq \inf \{ M \in \mathbb{R} : \| T(x) \| \leq M \| x \| , \forall x \in X \}.
\]

Consequently:
\[
\| T(x) \|_Y \leq \| T \| \| x \|_X , \quad \forall x \in X .
\]

**Fact.** For a linear operator \( A \), an equivalent expression (used widely!) for the operator norm is
\[
\| A \| = \sup_{\| x \| \leq 1} \| Ax \| .
\]

**Fact.** If \( X \) is the trivial vector space consisting only of the vector \( 0 \), then \( \| A \| = 0 \) for any linear operator \( A \).

**Fact.** If \( X \) is a nontrivial vector space, then for a linear operator \( A \) we have the following equivalent expressions:
\[
\| A \| = \sup_{x \neq 0} \frac{\| Ax \|}{\| x \|} = \sup_{\| x \| = 1} \| Ax \| .
\]
Example. Consider $X = (\mathbb{R}^n, \|\cdot\|_\infty)$ and $Y = \mathbb{R}$.

Clearly $A x = a_1 x_1 + \cdots + a_n x_n$ so $\|A x\|_Y = |A x| = |a_1 x_1 + \cdots + a_n x_n| \leq |a_1| |x_1| + \cdots + |a_n| |x_n| \leq (|a_1| + \cdots + |a_n|) \|x\|_\infty$.

In fact, if we choose $x$ such that $x_i = \text{sgn}(a_i)$, then $\|x\|_\infty = 1$ we get equality above. So we conclude $\|A\| = |a_1| + \cdots + |a_n|$.

Example. What if $X = (\mathbb{R}^n, \|\cdot\|_p)$?

\[\text{Proposition.}\] A linear operator is bounded iff it is continuous.

\[\text{Proof.}\] Exercise. [3, p. 240].

More facts related to linear operators.

- If $A : X \rightarrow Y$ is linear and $X$ is a finite-dimensional normed space, then $A$ is \textit{continuous} [3, p. 268].
- If $A : X \rightarrow Y$ is a transformation where $X$ and $Y$ are normed spaces, then $A$ is \textit{linear} and \textit{continuous} iff $A(\sum_{i=1}^{\infty} \alpha_i x_i) = \sum_{i=1}^{\infty} \alpha_i A(x_i)$ for all convergent series $\sum_{i=1}^{\infty} \alpha_i x_i$. [3, p. 237].

This is the \textit{superposition principle} as described in introductory signals and systems courses.

\section*{Spaces of bounded linear operators}

\textbf{Definition.} If $T_1$ and $T_2$ are both transformations with a common domain $X$ and a common range $Y$, over a common scalar field, then we define natural addition and scalar multiplication operations as follows:

\begin{align*}
(T_1 + T_2)(x) &= T_1(x) + T_2(x) \\
(\alpha T_1)(x) &= \alpha(T_1(x)).
\end{align*}

\[\text{Lemma.}\] With the preceding definitions, when $X$ and $Y$ are normed spaces the following space of \textit{operators} (!) is a \textit{vector space}:

\[B(X, Y) = \{\text{bounded linear transformations from } X \text{ to } Y\}.
\]

(The proof that this is a vector space is within the next proposition.)

This space is analogous to certain types of \textit{dual} spaces (see Ch. 5).

Not only is $B(X, Y)$ a vector space, it is a normed space when one uses the operator norm $\|A\|$ defined above.

\[\text{Proposition.}\] $(B(X, Y), \| \cdot \|)$ is a normed space when $X$ and $Y$ are normed spaces.

\[\text{Proof.}\] (sketch)

Claim 1. $B(X, Y)$ is a vector space.

Suppose $T_1, T_2 \in B(X, Y)$ and $\alpha \in \mathbb{F}$.

\[\|(\alpha T_1 + T_2)(x)\|_Y = \|\alpha T_1(x) + T_2(x)\|_Y \leq |\alpha| \|T_1(x)\|_Y + \|T_2(x)\|_Y \leq |\alpha| \|T_1\| \|x\|_X + \|T_2\| \|x\|_X = K \|x\|_X,\]

where $K \triangleq |\alpha| \|T_1\| + \|T_2\|$. So $\alpha T_1 + T_2$ is a bounded operator. Clearly $\alpha T_1 + T_2$ is a linear operator.

Claim 2. $\| \cdot \|$ is a norm on $B$.

The “hardest” part is verifying the triangle inequality:

\[\|T_1 + T_2\| = \sup_{\|x\|=1} \|(T_1 + T_2)x\|_Y \leq \sup_{\|x\|=1} \|T_1 x\|_Y + \sup_{\|x\|=1} \|T_2 x\|_Y = \|T_1\| + \|T_2\| \]

Are there other valid norms for $B(X, Y)$?

\[\text{Remark.}\] We did not really “need” linearity in this proposition. We could have shown that the space of bounded transformations from $X$ to $Y$ with $\| \cdot \|$ is a normed space.
Not only is \( B(\mathcal{X}, \mathcal{Y}) \) a normed space, but it is even **complete** if \( \mathcal{Y} \) is.

**Theorem.** If \( \mathcal{X} \) and \( \mathcal{Y} \) are normed spaces with \( \mathcal{Y} \) complete, then \((B(\mathcal{X}, \mathcal{Y}), \| \cdot \|)\) is complete.

**Proof.**

Suppose \( \{T_n\} \) is a Cauchy sequence (in \( B(\mathcal{X}, \mathcal{Y}) \)) of bounded linear operators, i.e., \( \|T_n - T_m\| \to 0 \) as \( n, m \to \infty \).

Claim 0. \( \forall x \in \mathcal{X} \), the sequence \( \{T_n(x)\} \) is Cauchy in \( \mathcal{Y} \).

\[
\|T_n(x) - T_m(x)\|_\mathcal{Y} = \|(T_n - T_m)(x)\|_\mathcal{Y} \leq \|T_n - T_m\| \|x\|_\mathcal{X} \to 0 \text{ as } n, m \to \infty.
\]

Since \( \mathcal{Y} \) is complete, for any \( x \in \mathcal{X} \) the sequence \( \{T_n(x)\} \) converges to some point in \( \mathcal{Y} \). (This is called **pointwise convergence**.) So we can define an operator \( T : \mathcal{X} \to \mathcal{Y} \) by \( T(x) \triangleq \lim_{n \to \infty} T_n(x) \).

To show \( B \) is complete, we must first show \( T \in B \), i.e., 1) \( T \) is linear, 2) \( T \) is bounded.

Then we show 3) \( T_n \to T \) (convergence w.r.t. the norm \( \| \cdot \| \)).

Claim 1. \( T \) is linear

\[
T(\alpha x + z) = \lim_{n \to \infty} T_n(\alpha x + z) = \lim_{n \to \infty} [\alpha T_n(x) + T_n(z)] = \alpha T(x) + T(z).
\]

(Recall that in a normed space, if \( u_n \to u \) and \( v_n \to v \), then \( \alpha u_n + v_n \to \alpha u + v \).)

Claim 2: \( T \) is bounded

Since \( \{T_n\} \) is Cauchy, it is bounded, so \( \exists K < \infty \) s.t. \( \|T_n\| \leq K \), \( \forall n \in \mathbb{N} \). Thus, by the continuity of norms, for any \( x \in \mathcal{X} \):

\[
\|T(x)\|_\mathcal{Y} = \lim_{n \to \infty} \|T_n(x)\|_\mathcal{Y} \leq \lim_{n \to \infty} \|T_n\| \|x\|_\mathcal{X} \leq K \|x\|_\mathcal{X}.
\]

Claim 3: \( T_n \to T \)

Since \( \{T_n\} \) is Cauchy,

\[
\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ s.t. } n, m > N \implies \|T_n - T_m\| \leq \varepsilon \implies \|T_n(x) - T_m(x)\|_\mathcal{Y} \leq \|T_n - T_m\| \|x\|_\mathcal{X}, \forall x \in \mathcal{X} \implies \lim_{m \to \infty} \|T_n(x) - T_m(x)\|_\mathcal{Y} \leq \varepsilon \|x\|_\mathcal{X}, \forall x \in \mathcal{X} \implies \|T_n(x) - T(x)\|_\mathcal{Y} \leq \varepsilon \|x\|_\mathcal{X}, \forall x \in \mathcal{X} \text{ by continuity of the norm} \implies \|T_n - T\| \leq \varepsilon.
\]

We have shown that every Cauchy sequence in \( B(\mathcal{X}, \mathcal{Y}) \) converges to some limit in \( B(\mathcal{X}, \mathcal{Y}) \), so \( B(\mathcal{X}, \mathcal{Y}) \) is complete. \( \square \)

**Corollary.** \((B(\mathcal{X}, \mathcal{Y}), \| \cdot \|)\) is a Banach space for any normed space \( \mathcal{X} \).

**Why?**

We write \( A \in B(\mathcal{X}, \mathcal{Y}) \) as shorthand for “\( A \) is a bounded linear operator from normed space \( \mathcal{X} \) to normed space \( \mathcal{Y} \).”

**Definition.** In general, if \( S : \mathcal{X} \to \mathcal{Y} \) and \( T : \mathcal{Y} \to \mathcal{Z} \), then we can define the **product operator** or **composition** as a transformation \( T S : \mathcal{X} \to \mathcal{Z} \) by \( (T S)(x) = T(S(x)) \).

**Proposition.** If \( S \in B(\mathcal{X}, \mathcal{Y}) \) and \( T \in B(\mathcal{Y}, \mathcal{Z}) \), then \( TS \in B(\mathcal{X}, \mathcal{Z}) \).

**Proof.** Linearity of the composition of linear operators is trivial to show.

To show that the composition is bounded:

\[
\|TSx\|_\mathcal{Z} \leq \|T\| \|Sx\|_\mathcal{Y} \leq \|T\| \|S\| \|x\|_\mathcal{X}.
\]

Does it follow that \( \|TS\| = \|T\|\|S\| \)?
Inverse operators

Definition. \( T : \mathcal{X} \to \mathcal{Y} \) is called one-to-one mapping of \( \mathcal{X} \) into \( \mathcal{Y} \) iff \( x_1, x_2 \in \mathcal{X} \) and \( x_1 \neq x_2 \implies T(x_1) \neq T(x_2). \)

Equivalently, \( T \) is one-to-one if the inverse image of any point \( y \in \mathcal{Y} \) is at most a single point in \( \mathcal{X} \), i.e., \( |T^{-1}\{y\}| \leq 1, \forall y \in \mathcal{Y}. \)

Definition. \( T : \mathcal{X} \to \mathcal{Y} \) is called onto iff \( T(\mathcal{X}) = \mathcal{Y} \). This is a stronger condition than into.

Fact. If \( T : \mathcal{X} \to \mathcal{Y} \) is one-to-one and onto \( \mathcal{Y} \), then \( T \) has an inverse denoted \( T^{-1} \) such that \( T(x) = y \iff T^{-1}(y) = x \).

Many optimization methods, e.g., Newton’s method, require inversion of the Hessian matrix (or operator) corresponding to a cost function.

Lemma. [3, p. 171]
If \( A : \mathcal{X} \to \mathcal{Y} \) is a linear operator between two vector spaces \( \mathcal{X} \) and \( \mathcal{Y} \), then \( A \) is one-to-one iff \( N(A) = \{0\} \).

Linearity of inverses

We first look at the algebraic aspects of inverse operators in vector spaces.

**Proposition.** If a linear operator \( A : \mathcal{X} \to \mathcal{Y} \) (for vector spaces \( \mathcal{X} \) and \( \mathcal{Y} \)) has an inverse, then that inverse \( A^{-1} \) is also linear.

**Proof.** Suppose \( A^{-1}(y_1) = x_1, A^{-1}(y_2) = x_2, A(x_1) = y_1, \) and \( A(x_2) = y_2. \) Then by the linearity of \( A \) we have \( A(\alpha x_1 + x_2) = \alpha A x_1 + Ax_2 = \alpha y_1 + y_2, \) so \( A^{-1}(\alpha y_1 + y_2) = \alpha x_1 + x_2 = \alpha A^{-1}(y_1) + A^{-1}(y_2). \)

6.4

Banach inverse theorem

Now we turn to the topological aspects, in normed spaces.

**Lemma.** (Baire) A Banach space \( \mathcal{X} \) is not the union of countably many nowhere dense sets in \( \mathcal{X}. \)

**Proof.** see text

**Theorem.** (Banach inverse theorem)
If \( A \) is a continuous linear operator from a Banach space \( \mathcal{X} \) onto a Banach space \( \mathcal{Y} \) for which the inverse operator \( A^{-1} \) exists, then \( A^{-1} \) is continuous.

**Proof.** see text

Combining with the earlier Proposition that linear operators are bounded iff they are continuous yields the following.

**Corollary.**
\( \mathcal{X}, \mathcal{Y} \) Banach and \( A \in B(\mathcal{X}, \mathcal{Y}) \) and \( A \) invertible \( \implies A^{-1} \in B(\mathcal{Y}, \mathcal{X}) \)

Equivalence of spaces (one way to use operators)

- in vector spaces
- in normed spaces
- in inner product spaces

<table>
<thead>
<tr>
<th>spaces</th>
<th>relation</th>
<th>operator</th>
<th>requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>vector normed</td>
<td>isomorphic</td>
<td>isomorphism</td>
<td>linear, onto, 1-1</td>
</tr>
<tr>
<td>normed topologically isomorphic</td>
<td>topological isomorphism</td>
<td>linear, onto, invertible, continuous</td>
<td></td>
</tr>
<tr>
<td>inner product unitarily equivalent</td>
<td>isometric isomorphism</td>
<td>linear, onto, norm preserving ( \implies 1-1, ) continuous</td>
<td></td>
</tr>
<tr>
<td></td>
<td>unitary</td>
<td>isometric isomorphism that preserves inner products</td>
<td></td>
</tr>
</tbody>
</table>
Isomorphic spaces

Definition. Vector spaces $X$ and $Y$ are called isomorphic (think: “same structure”) iff there exists a one-to-one, linear mapping $T$ of $X$ onto $Y$.

In such cases the mapping $T$ is called an isomorphism of $X$ onto $Y$.

Since an isomorphism $T$ is one-to-one and onto, $T$ is invertible, and by the “linear of inverses” proposition in 6.3, $T^{-1}$ is linear.

Example. Consider $X = \mathbb{R}^2$ and $Y = \{ f(t) = a + bt \text{ on } [0, 1] : a, b \in \mathbb{R} \}$.

An isomorphism is $f = T(x) \iff f(t) = a + bt$ where $x = (a, b)$, with inverse $T^{-1}(f) = (f(0), f(1) - f(0))$.

Exercise. Any real $n$-dimensional vector space is isomorphic to $\mathbb{R}^n$ (problem 2.14 p. 44) [3, p. 268].

However, they need not be topologically isomorphic [3, p. 270].

So far we have said nothing about norms. In normed spaces we can have a topological relationship too.

Definition. Normed spaces $X$ and $Y$ are called topologically isomorphic iff there exists a continuous linear transformation $T$ of $X$ onto $Y$ having a continuous inverse $T^{-1}$. The mapping $T$ is called a topological isomorphism.

Theorem. [3, p. 258] Two normed spaces are topologically isomorphic iff there exists a linear transformation $T$ with domain $X$ and range $Y$ and positive constants $m$ and $M$ s.t. $m \| x \|_X \leq \| T(x) \|_Y \leq M \| x \|_X$, $\forall x \in X$.

Example. In the previous example, consider $(X, \| \cdot \|_\infty)$ and $(Y, \| \cdot \|_2)$. Then for the same $T$ described in that example:

$\| T(x) \|_2^2 = \int_0^1 (a + bt)^2 dt = a^2 + ab + b^2/3 \leq a^2 + |a| |b| + b^2/3 \leq (1 + 1 + 1/3) \| x \|_\infty^2 = 7/3 \| x \|_\infty^2$.

Also $\| T(x) \|_Y^2 = a^2 + ab + b^2/3 = (a + b/2)^2 + b^2/12 = (a \sqrt{3}/2 + b/\sqrt{3})^2 + a^2/4 \geq \| x \|_\infty^2/12$.

So $X$ and $Y$ are topologically isomorphic for the given norms.

Exercise. $(C[0, 1], \| \cdot \|_\infty)$ and $(C[0, 1], \| \cdot \|_1)$ are not topologically isomorphic. Why? ??

Isometric spaces

Definition. Let $X$ and $Y$ be normed spaces. A mapping $T : X \to Y$ is called norm preserving iff $\| T(x) \|_Y = \| x \|_X$, $\forall x \in X$.

In particular, if $T$ is norm preserving, then $\| T \| = 1$. What about the converse? ??

Proposition. If $T$ is linear and norm preserving, then $T$ is one-to-one, i.e., $T(x) = T(z) \implies x = z$.

Proof. If $T(x) = y$ and $T(z) = y$, then by linearity $T(x - z) = 0$.

So since $T$ is norm preserving, $\| x - z \| = \| 0 \| = 0$, so $x = z$. $\square$

Definition. If $T : X \to Y$ is both linear and norm preserving, then $T$ is called a linear isometry.

If, in addition, $T$ is onto $Y$, then $X$ and $Y$ are called isometrically isomorphic, and $T$ is called an isometric isomorphism.

Remark. To illustrate why we require onto here, consider $T : E^n \to \ell_2$ defined by $T(x) = (x_1, \ldots, x_n, 0, 0, \ldots)$.

This $T$ is linear, one-to-one, and norm preserving, but not onto.

Exercise. Every normed space $X$ is isometrically isomorphic to a dense subset of a Banach space $X$. (problem 2.15 on p. 44)

Normed spaces that are isometrically isomorphic can, in some sense, be treated as being identical, i.e., they have identical properties. However, the specific isomorphism can be important sometimes.

Example. Consider $Y = \ell_p(\mathbb{N}) = \{ (a_1, a_2, \ldots) : \sum_{i=1}^\infty |a_i|^p < \infty \}$ and $X = \ell_p(\mathbb{Z}) = \{ \ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots : \sum_{i=-\infty}^\infty |a_i|^p < \infty \}$.

Define the mapping $T : X \to Y$ by $y = (b_1, b_2, \ldots) = T(x)$ if $b_i = a_{z(i)}$ where $z(i) = (-1)^i [i/2]$. Note that $z : \{ 1, 2, 3, \ldots \} \to 0, 1, -1, 2, -2, \ldots$.

This mapping $T$ is clearly an isometric isomorphism, so “$\ell_p(\mathbb{Z})$” and “$\ell_p(\mathbb{N})$” are isometrically isomorphic. Hence we only bother to work with $\ell_p = \ell_p(\mathbb{N})$ since we know all algebraic and topological results will carry over to double-sided sequences.
Unitary equivalence in inner product spaces

Definition. Two inner product spaces $\mathcal{X}$ and $\mathcal{Y}$ are **unitarily equivalent** iff there is an isomorphism $U : \mathcal{X} \rightarrow \mathcal{Y}$ that preserves inner products; $\langle Ux_1, Ux_2 \rangle_\mathcal{Y} = \langle x_1, x_2 \rangle_\mathcal{X}$, $\forall x_1, x_2 \in \mathcal{X}$. The mapping $U$ is called a **unitary operator**.

Fact. If $U$ is unitary, then $U$ is norm preserving since $\|Ux\| = \langle Ux, Ux \rangle = \langle x, x \rangle = \|x\|$. Clearly $\|U\| = 1$.

Furthermore, since $U$ is onto, $U$ is an **isometric isomorphism**. Interestingly, the converse is also true [3, p. 332].

**Theorem.** A mapping $U$ of $\mathcal{X}$ onto $\mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are inner product spaces, is an isometric isomorphism iff $U$ is a unitary operator.

Proof. ($\implies$) was just argued above.

($\impliedby$) Suppose $U$ is an isometric isomorphism.

Using the parallelogram law, the linearity of $U$, and the fact that $\|Ux\| = \|x\|$, we have:

$$4 \langle Ux, Uy \rangle = ||Ux + Uy||^2 - ||Ux - Uy||^2 + i ||Ux + iUy||^2 - i ||Ux - iUy||^2$$

$$= ||U(x + y)||^2 - ||U(x - y)||^2 + i ||U(x + iy)||^2 - i ||U(x - iy)||^2$$

$$= ||x + y||^2 - ||x - y||^2 + i ||x + iy||^2 - i ||x - iy||^2 = 4 \langle x, y \rangle .$$

Thus $U$ an isometric isomorphism $\implies$ $U$ unitary.

Remark. After defining adjoints we will show that $U^{-1} = U^*$ in Hilbert spaces.

**Exercise.** Any complex $n$-dimensional inner product space is unitarily equivalent to $\mathbb{C}^n$ [3, p. 332].

Every separable Hilbert space is unitarily equivalent with $\ell_2$ or some $\mathbb{C}^n$ [3, p. 339].

Example. Continue the previous $f(t) = a + bt$ example, but now use $\mathbb{R}^n$ and $\langle \cdot, \cdot \rangle_2$. If $g(t) = c + dt$ then

$$\langle T x_1, T x_2 \rangle = \langle f, g \rangle = \int_0^1 f(t)g(t) \, dt = \int_0^1 (a + bt)(c + dt) \, dt = ac + ad \frac{1}{2} + bc + bd \frac{1}{2} = [a \ b] \begin{bmatrix} 1/2 & 1/3 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = x'_1 G x_2 ,$$

so we define $U = T G^{-1/2}$, then

$$\langle U x_1, U x_2 \rangle = \langle T G^{-1/2} x_1, T G^{-1/2} x_2 \rangle = \langle G^{-1/2} x_1 \rangle ^* \langle G^{-1/2} x_1 \rangle = x'_1 x_1 = \langle x_1, x_2 \rangle .$$

Example. The Fourier transform $F(\nu) = \int_{-\infty}^{\infty} f(t) e^{-i\nu t} \, dt$ is a unitary mapping of $L_2[\mathbb{R}]$ onto itself.

Example. Soon we will analyze the **discrete-time Fourier transform** (DTFT) operator, defined by

$$G = F g \iff G(\omega) = \sum_{n=-\infty}^{\infty} g_n e^{i\omega n} .$$

We will show that $F \in B(\ell_2, L_2[-\pi, \pi])$ and $F$ is invertible. And Parseval’s relation from Fourier analysis is that

$$\left\langle \frac{1}{\sqrt{2\pi}} F g, \frac{1}{\sqrt{2\pi}} F h \right\rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\omega) H^*(\omega) \, d\omega = \sum_{n=-\infty}^{\infty} g_n h_n^* = \langle g, h \rangle .$$

So $\ell_2$ and $L_2[-\pi, \pi]$ are unitarily equivalent, and the unitary operator needed is simply $U = \frac{1}{\sqrt{2\pi}} F$. 
The following “extension theorem” is useful in proving that every separable Hilbert space is isometrically isomorphic to $\ell_2$.

**Theorem.** Let $\mathcal{X}$ be a normed space, $\mathcal{Y}$ a Banach space, and $M \subset \mathcal{X}$ and $N \subset \mathcal{Y}$ be subspaces. Suppose that

- $\overline{M} = \mathcal{X}$, and $\overline{N} = \mathcal{Y}$,
- $\tilde{T} : M \to N$ is a bounded linear operator.

Then there exists a unique linear operator $T : \mathcal{X} \to \mathcal{Y}$ such that $T \restriction M = \tilde{T}$. Moreover, $\| T \| = \| \tilde{T} \|$. If, in addition,

- $\mathcal{X}$ is a Banach space,
- $\tilde{T}^{-1} : N \to M$ exists and is bounded,

then $T$ is also onto $\mathcal{Y}$.

**Proof.** Note: $T \restriction M$ reads $T$ restricted to $M$.

$T \restriction M = \tilde{T}$ means $T(x) = \tilde{T}(x)$ for all $x \in M$.

Claim 1. If $\{ x_n \} \in M$ is Cauchy in $M$, then $\{ \tilde{T}(x_n) \}$ is Cauchy in $N$.

$$\| \tilde{T}(x_n) - \tilde{T}(x_m) \|_Y = \| T(x_n - x_m) \|_Y \leq \| \tilde{T} \| \| x_n - x_m \|_X \to 0 \text{ as } n, m \to \infty.$$  

Claim 2. For $\{ x_n \}, \{ \tilde{x}_n \} \in M$, suppose $x_n \to x \in \mathcal{X}$ and $\tilde{x}_n \to \tilde{x} \in \mathcal{Y}$.

Then $\{ \tilde{T}(x_n) \}$ and $\{ \tilde{T}(\tilde{x}_n) \}$ both converge to the same limit.

$$\{ \tilde{T}(x_n) \} \text{ and } \{ \tilde{T}(\tilde{x}_n) \} \text{ are both Cauchy in } N \text{ by Claim 1 since } x_n \text{ and } \tilde{x}_n \text{ both converge.}$$

Since $\mathcal{Y}$ is complete, $\exists y, \tilde{y} \in \mathcal{Y}$ s.t. $\tilde{T}(x_n) \to y$ and $\tilde{T}(\tilde{x}_n) \to \tilde{y}$.

But $\| y - \tilde{y} \|_Y = \lim_{n \to \infty} \| \tilde{T}(x_n) - \tilde{T}(\tilde{x}_n) \|_Y = \lim_{n \to \infty} \| \tilde{T}(x_n - \tilde{x}_n) \|_Y \leq \lim_{n \to \infty} \| \tilde{T} \| \| x_n - \tilde{x}_n \|_X = 0$, using norm continuity, linearity of $\tilde{T}$, and boundedness of $\tilde{T}$.

Now we define $T : \mathcal{X} \to \mathcal{Y}$ as follows.

For $x \in \mathcal{X} = \overline{M}$, $\exists \{ x_n \} \in M$ such that $x_n \to x$, so we define $T(x) = \lim_{n \to \infty} \tilde{T}(x_n)$.

By Claim 2, $T$ is well-defined, and moreover, if $x \in M$, then $T(x) = \tilde{T}(x)$, i.e., $T \restriction M = \tilde{T}$.

Claim 3. $\| T \| = \| \tilde{T} \|$.  

$\forall x \in \mathcal{X}, T(x) = \lim_{n \to \infty} \tilde{T}(x_n)$ where $x_n \in M$ and $x_n \to x$.

Thus $\| T(x) \|_Y = \lim_{n \to \infty} \| \tilde{T}(x_n) \|_Y \leq \lim_{n \to \infty} \| \tilde{T} \| \| x_n \|_X = \| \tilde{T} \| \| x \|_X$. Thus $\| T \| \leq \| \tilde{T} \|$.

However, $\forall x \in M, T(x) = \tilde{T}(x) \implies \| T \| \geq \| \tilde{T} \|$, since $T$ is defined on $\mathcal{X} \supseteq M$.

Thus $\| T \| = \| \tilde{T} \|$.

Claim 4. $T$ is linear, which is trivial to prove.

Claim 5. (uniqueness of $T$) If $L_1, L_2 \in B(\mathcal{X}, \mathcal{Y})$, then $L_1 \restriction M = L_2 \restriction M \implies L_1 = L_2$.

For $x \in \mathcal{X} = \overline{M}$, $\exists \{ x_n \} \in M$ s.t. $x_n \to x$. Thus, by the continuity of $L_1$ and $L_2$,

$L_1(x) = \lim_{n \to \infty} L_1(x_n) = \lim_{n \to \infty} L_2(x_n) = L_2(x)$, since $L_1(x_n) = L_2(x_n)$ by $L_1 \restriction M = L_2 \restriction M$.

Claim 6. If $\tilde{T}^{-1} : N \to M$ exists and is bounded, and if $\mathcal{X}$ is a Banach space, then $T$ is onto $\mathcal{Y}$.

Let $y \in \mathcal{Y}$. Since $\overline{N} = \mathcal{Y}$, $\exists \{ y_n \} \in N$ s.t. $y_n \to y$.

By the same reasoning as in Claim 1, $x_n = \tilde{T}^{-1}(y_n)$ is Cauchy in $\mathcal{X}$.

Since $\mathcal{X}$ is a Banach space, $\exists \tilde{x} \in \mathcal{X}$ s.t. $x_n \to \tilde{x}$.

Thus $T(x) = \lim_{n \to \infty} \tilde{T}(x_n) = \lim_{n \to \infty} \tilde{T}(\tilde{T}^{-1}(y_n)) = \lim_{n \to \infty} y_n = y$.

Thus, $\forall y \in \mathcal{Y}, \exists x \in \mathcal{X}$ s.t. $T(x) = y$.  

$\Box$
The following theorem is an important application of the previous theorem.

**Theorem.** Every separable Hilbert space \( \mathcal{H} \) is isometrically isomorphic to \( \ell_2 \).

**Proof.**
\( \mathcal{H} \) separable \( \implies \exists \) a countable orthonormal basis \( \{e_i\} \). (Homework.) Let \( M = \{e_i\}_{i=1}^{\infty} \).
Of course \( \ell_2 \) has a countable orthonormal basis \( \{\tilde{e}_i\} \), where \( \tilde{e}_{ij} = \delta_{i-j} \) (Kronecker). Let \( N = \{\tilde{e}_i\}_{i=1}^{\infty} \).

Define \( \tilde{T} : M \to N \) to be the linear operator for which \( \tilde{T}(e_i) = \tilde{e}_i \).

**(Exercise.** Think about why “the” here—consider \( M \).)

Then since any \( x \in M \) has form \( x = \sum_{i=1}^{n} c_i e_i \) for some \( n \in \mathbb{N} \):
\[
\left\| \tilde{T}(x) \right\|_2 = \left\| \tilde{T} \left( \sum_{i=1}^{n} c_i e_i \right) \right\|_2 = \left\| \sum_{i=1}^{n} c_i \tilde{T}(e_i) \right\|_2 = \left\| \sum_{i=1}^{n} c_i \tilde{e}_i \right\|_2 = \left\| \sum_{i=1}^{n} |c_i|^2 \right\|_{\mathcal{H}} = \left\| x \right\|_2,
\]
so \( \left\| \tilde{T} \right\| = 1 \) on \( M \).

Clearly \( \tilde{T}^{-1} \) exists on \( N \) and is given by \( \tilde{T}^{-1}(\tilde{e}_i) = e_i \), so \( \left\| \tilde{T} \right\| = 1 \) on \( N \) by a similar argument.

Since \( \ell_2 \) is a Banach space, we have established all the conditions of the preceding theorem, so there exists a unique linear operator \( T \) from \( \mathcal{H} \) to \( \ell_2 \) that is onto \( \ell_2 \) with \( \left\| T \right\| = \left\| \tilde{T} \right\| = 1 \).

Since \( T \) is bounded, it is continuous, so one can take limits as \( n \to \infty \) in (6-1) to show that \( \left\| T x \right\| = \left\| x \right\| \), so \( T \) is norm preserving.

Thus \( T \) is an isometric isomorphism for \( \mathcal{H} \) and \( \ell_2 \), so \( \mathcal{H} \) and \( \ell_2 \) are isometrically isomorphic.

**Corollary.** \( \mathcal{L}_2 \) is isometrically isomorphic to \( \ell_2 \).

A remarkable result. (Perhaps most of what is remarkable here is that \( \mathcal{L}_2 \) is separable.)

A practical consequence: usually one can search for counterexamples in \( \ell_2 \) (and its relatives) rather than \( \mathcal{L}_2 \).

**Example.** The (different) spaces of odd and even functions in \( \mathcal{L}_2[-\pi, \pi] \) are isometrically isomorphic.

**Example.** Elaborating. Let \( \mathcal{H} = \mathcal{L}_2[-\pi, \pi] \) and define
\[
e_{2k} = \cos(kt) / c_{2k} \quad e_{2k+1} = \sin(kt) / c_{2k+1}, \ k = 0, 1, 2, \ldots
\]
\[
c_{2k} = \int_{-\pi}^{\pi} \cos^2(kt) \, dt \quad c_{2k+1} = \int_{-\pi}^{\pi} \sin^2(kt) \, dt
\]
\[
\mathcal{X} = \{e_{2k}\} = \{f \in \mathcal{L}_2 : f \text{ even}\} \quad \mathcal{Y} = \{e_{2k+1}\} = \{f \in \mathcal{L}_2 : f \text{ odd}\}.
\]

Then \( \mathcal{H}, \mathcal{X}, \) and \( \mathcal{Y} \) are each Banach spaces that are isometrically isomorphic to each other! And each is isometrically isomorphic to \( \ell_2 \).

**Example.** \( \ell_2 \) is isometrically isomorphic to \( \mathcal{L}_2[-\pi, \pi] \). (Just use the DTFT.)

But an even stronger result holds than the above...

Every separable Hilbert space is unitarily equivalent with \( \ell_2 \) or some \( \mathbb{C}^n \) [3, p. 339].
Summary

**Insert 5.1-5.3 here!**
6.5
Adjoints in Hilbert spaces

When performing optimization in inner product spaces, often we need the “transpose” of a particular linear operator. The term transpose only applies to matrices. The more general concept for linear operators is called the adjoint.

Luenberger presents adjoints in terms of general normed spaces. In my experience, adjoints most frequently arise in inner product spaces, so in the interest of time and simplicity these notes focus on that case. The treatment here generalizes somewhat the treatment in Naylor [3, p. 352].

Recall the following fact from linear algebra. If $A : \mathbb{C}^n \to \mathbb{C}^m$ is a $m \times n$ matrix, then

$$\langle Ax, y \rangle_{\mathbb{C}^m} = y^T A x = \langle A^T y, x \rangle_{\mathbb{C}^n}.$$

This section generalizes the above relationship to general Hilbert spaces.

Let $A \in B(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces.

Let $y \in \mathcal{Y}$ be a fixed vector, and consider the following functional:

$$g_y : \mathcal{X} \to \mathbb{C}, \text{ where } g_y(x) \equiv \langle Ax, y \rangle_{\mathcal{Y}}.$$

- $g_y$ is clearly linear, since $A$ is linear and $\langle \cdot, y \rangle_{\mathcal{Y}}$ is linear.
- $\|g_y(x)\| \leq \|Ax\|_{\mathcal{Y}} \|y\|_{\mathcal{Y}} \leq \|A\| \|x\|_{\mathcal{X}} \|y\|_{\mathcal{Y}}$ by Cauchy-Schwarz and since $A$ is bounded. Thus $g_y$ is bounded, with $\|g_y\| \leq \|A\| \|y\|_{\mathcal{Y}}$.

In other words, $g_y \in \mathcal{X}^*$. 

**Definition.** By the Riesz representation theorem (here is where we use completeness), for each such $y \in \mathcal{Y}$ there exists a unique $z = z_y \in \mathcal{X}$ such that

$$g_y(x) = \langle x, z_y \rangle_{\mathcal{X}}, \quad \forall x \in \mathcal{X}.$$

So we can define legitimately a mapping $A^* : \mathcal{Y} \to \mathcal{X}$, called the adjoint of $A$, by the relationship $z_y = A^*(y)$.

The defining property of $A^*$ is then:

$$\langle Ax, y \rangle_{\mathcal{Y}} = \langle x, A^*(y) \rangle_{\mathcal{X}}, \quad \forall x \in \mathcal{X}, \ y \in \mathcal{Y}.$$

(At this point we should write $A^*(y)$ rather than $A^* y$ since we have not yet shown $A^*$ is linear, though we will soon.)

**Lemma.** $A^*$ is the only mapping of $\mathcal{Y}$ to $\mathcal{X}$ that satisfies the preceding equality.

**Proof.** For any $y \in \mathcal{Y}$, suppose $\forall x \in \mathcal{X}$ we have $\langle Ax, y \rangle_{\mathcal{Y}} = \langle x, T_1(y) \rangle_{\mathcal{X}} = \langle x, T_2(y) \rangle_{\mathcal{X}}$.

Then $0 = \langle x, T_1(y) - T_2(y) \rangle_{\mathcal{X}}$, so $T_1(y) = T_2(y), \forall y \in \mathcal{Y}$. \hfill $\square$

**Exercise.** Here are some simple facts about adjoints, all of which concur with those of Hermitian transpose in Euclidean space.

- $\langle \alpha x, y \rangle = \langle x, \alpha^* y \rangle$ so for $A : \mathcal{X} \to \mathcal{X}$ defined by $Ax = \alpha x$ we have $A^* y = \alpha^* y$. So “reuse” of the asterix is acceptable.
- $I^* = I$.
- $0_{\mathcal{Y} \to \mathcal{X}}^* = 0_{\mathcal{X} \to \mathcal{X}}$
- $A^{**} = A$ (see Thm below)
- $(ST)^* = T^* S^*$
- $(S + T)^* = S^* + T^*$
- $(\alpha A)^* = \alpha^* A^*$

Note: these last two properties are unrelated to the question of whether $A^*$ is a linear operator! 

**Example.** Consider $\mathcal{X} = L^2[0, 1], \mathcal{Y} = \mathbb{C}^2$, and $A : \mathcal{X} \to \mathcal{Y}$ is defined by

$$y = Ax \iff y_1 = [Ax]_1 = \int_0^1 t x(t) \, dt, \ y_2 = [Ax]_2 = \int_0^1 t^2 x(t) \, dt.$$

We can guess that the adjoint of $A$ is defined by $x = A^* y \iff x(t) = (A^* y)(t) = y_1 t + y_2 t^2$.

This is verified easily:

$$\langle Ax, y \rangle_{\mathcal{Y}} = y_1^* [Ax]_1 + y_2^* [Ax]_2 = y_1^* \int_0^1 t x(t) \, dt + y_2^* \int_0^1 t^2 x(t) \, dt = \int_0^1 x(t) \left[ y_1 t + y_2 t^2 \right] \, dt \equiv \langle x, A^* y \rangle_{\mathcal{X}}.$$
**Theorem.** Suppose $A \in B(X, Y)$ where $X$ and $Y$ are Hilbert spaces.
- The adjoint operator $A^*$ is linear and bounded, i.e., $A^* \in B(Y, X)$.
- $\|A^*\| = \|A\|
- (A^*)^* = A$

**Proof.**

Claim 1. $A^*$ is linear.

$\langle x, A^*(\alpha y + z) \rangle_X = \langle Ax, \alpha y + z \rangle_Y = \alpha^* \langle Ax, y \rangle_Y + \langle Ax, z \rangle_Y = \alpha^* \langle x, A^*(y) \rangle_X + \langle x, A^*(z) \rangle_X$

$= \langle x, \alpha A^*(y) + A^*(z) \rangle_X$, which holds for all $x \in X$ and $y \in Y$, so $A^*(\alpha y + z) = \alpha A^*(y) + A^*(z)$ by the usual Lemma once again.

Claim 2. $A^*$ is bounded and $\|A^*\| \leq \|A\|$.

$\|A^*y\|_X = \langle A^*y, A^*y \rangle_X = \langle AA^*y, y \rangle_Y \leq \|AA^*y\|_Y \|y\|_Y \leq \|A\| \|A^*y\|_X \|y\|_Y$

so $\|A^*y\|_X \leq \|A\| \|y\|_Y$ and thus $\|A^*\| \leq \|A\|$ and hence $A^* \in B(Y, X)$.

Claim 3. $A^{**} = A$.

Since we have shown $A^* \in B(Y, X)$, we can legitimately define the adjoint of $A^*$, denoted $A^{**}$, as the (bounded linear) operator that satisfies $\langle A^{**}y, x \rangle_Y = \langle y, A^*x \rangle_Y$, $\forall x \in X, y \in Y$.

Since $\langle y, Ax \rangle_Y = \langle y, A^*x \rangle_Y = \langle x, A^*y \rangle_X = \langle x, A^{**}y \rangle_X$, by the previous uniqueness arguments we see $A^{**} = A$.

Claim 4. $\|A^*\| = \|A\|$.

From Claim 2 with $A^*$: $\|A^{**}\| \leq \|A\|$ or equivalently: $\|A\| \leq \|A^*\|$.

**Corollary.** Under the same conditions as above: $\|A^*A\| = \|AA^*\| = \|A\|^2 = \|A^*\|^2$.

**Proof.** Recalling that $\|ST\| \leq \|S\| \|T\|$ we have by the preceding theorem:

$\|A^*A\| \leq \|A^*\| \|A\| = \|A\|^2 = \|A^*\|^2$,

and $\|Ax\|_Y^2 = \langle Ax, Ax \rangle_Y = \langle Ax, (A^*)^*Ax \rangle_X = \langle A^*Ax, x \rangle_X \leq \|A^*Ax\| \|x\| \leq \|A^*A\| \|x\|^2$

so $\|A\|^2 \leq \|A^*A\|$. Combining yields the equality $\|A\|^2 = \|A^*A\|$. The rest is obvious using $A^{**} = A$.

**Proposition.** If $A \in B(X, Y)$ is invertible, where $X, Y$ are Hilbert spaces, then $A^*$ has an inverse and $(A^*)^{-1} = (A^{-1})^*$.

**Proof.**

Claim 1. $A^* : Y \rightarrow X$ is one-to-one into $X$.

Consider $y_1, y_2 \in Y$ with $y_1 \neq y_2$, but suppose $A^*y_1 = A^*y_2$, so $A^*d = 0$ where $d = y_2 - y_1 \neq 0$.

Thus $\langle x, A^*d \rangle_X = \langle Ax, d \rangle_Y = 0$, $\forall x \in X$ and hence $\langle Ax, \rangle_Y = 0$, $\forall x \in X$.

Since $A$ is invertible, we can make the “change of variables” $z = Ax$ and hence $\langle z, d \rangle_Y = 0$, $\forall z \in X$.

But this implies $d = 0$, contradicting the supposition that $d \neq 0$. So $A^*$ is one-to-one.

Claim 2. $A^* : Y \rightarrow X$ is onto $X$.

By the Banach inverse theorem, since $A \in B(X, Y)$ and $A$ is invertible, $A^{-1} \in B(Y, X)$, so $A^{-1}$ has its own adjoint, $(A^{-1})^*$.

Pick any $z \in X$. Then for any $x \in X$

$\langle x, z \rangle = \langle A^{-1}Ax, z \rangle = \langle Ax, (A^{-1})^*z \rangle = \langle Ax, (A^{-1})z \rangle = \langle x, A^*(A^{-1})z \rangle$

Thus $z = A^*\left[\left(\frac{A^{-1}z}{\|z\|} \right)\right]$, showing that $z$ is in the range of $A^*$. Since $z \in X$ was arbitrary, $A^*$ is onto $X$.

Claim 3. $(A^*)^{-1} = (A^{-1})^*$.

Since $A^*$ is one-to-one and onto $X$, $A^*$ is invertible.

Furthermore, $z = A^*(A^{-1})z \implies (A^*)^{-1}z = (A^{-1})^*z$, $\forall z \in X$. Thus $(A^*)^{-1} = (A^{-1})^*$.

**Remark.** $AB = I$ and $BA = I \implies A, B$ invertible and $A^{-1} = B$.

**Remark.** $AB = I$ and $A$ invertible $\implies A^{-1} = B$. 
Definition. A ∈ B(ℋ, ℋ), where ℋ is a real Hilbert space, is called self adjoint if A* = A.

Exercise. An orthogonal projection P_M : ℋ → ℋ, where M is a closed subspace in a Hilbert space ℋ, is self adjoint. Case

Exercise. Conversely, if P ∈ B(ℋ, ℋ) and P^2 = P and P* = P, then P is an orthogonal projection operator. (L6.16)

Definition. A self-adjoint bounded linear operator A on a Hilbert space ℋ is positive semidefinite iff ⟨x, Ax⟩ ≥ 0, ∀x ∈ ℋ.

Remark. It is easily shown that ⟨x, Ax⟩ is real when A is self-adjoint.

Example. When M is a Chebyshev subspace in an inner product space, is P_M a self-adjoint operator?

Unitary operators

(Caution: the proof on [3, p. 358] is incomplete w.r.t. the “onto” aspects.)

We previously defined unitary operators; we now examine the adjoints of these.

Theorem. Suppose U ∈ B(X, Y) where X, Y are Hilbert spaces. Then the following are equivalent:
1. U is unitary, i.e., U is an isomorphism (linear, onto, and one-to-one) and ⟨Ux, Uz⟩ = ⟨x, z⟩, ∀x, z ∈ X,
2. U*U = I and UU* = I,
3. U is invertible with U^(-1) = U*.

Proof. (2 ⇒ 3) and (3 ⇒ 1) are obvious. (1 ⇒ 2) If U is unitary then for all x, z ∈ X: ⟨x, z⟩ = ⟨Ux, Uz⟩ = ⟨U*Ux, z⟩, so U*Ux = x, ∀x ∈ X so U*U = IU.

For any y ∈ Y, since U is onto there exists an x ∈ X s.t. Ux = y. Thus UU*y = UU*Ux = UIx x = Ux = y.

Since y ∈ Y was arbitrary, UU* = IY. □

Remark. A corollary is that if U is unitary, then so is U*.

Remark. To see why we need both U^*U = I and UU* = I above, consider U = [1 0], for which U^*U = I but U is not onto.

Example.

Consider X = Y = ℓ₂ and the (linear) discrete-time convolution operator A : ℓ₂ → ℓ₂ defined by

z = Ax ⇐⇒ z_n = ∑_{k=−∞}^{∞} h_{n−k}x_k, n ∈ ℤ,

where we assume that h ∈ ℓ₁, which is equivalent to BIBO stability. We showed previously that ||Ax||₂ ≤ ||h||₁ ||x||₂, so A is bounded with ||A|| ≤ ||h||₁, so A has an adjoint. (Later we will show ||A|| = ||H||ₙ∞ where H is the frequency response of h.)

Since A is bounded, it is legitimate to search for its adjoint:

⟨Ax, y⟩ = ∑_{n=−∞}^{∞} y_n^* ∑_{k=−∞}^{∞} x_k h_{n−k} = ∑_{k=−∞}^{∞} x_k ∑_{n=−∞}^{∞} y_n h_{n−k}^* = ∑_{k=−∞}^{∞} x_k [A*y]_k^* = ⟨x, A*y⟩,

where the adjoint is

[A*y]_k = ∑_{n=−∞}^{∞} h_{k−n}^* y_n \implies [A*y]_n = ∑_{k=−∞}^{∞} h_{k−n}^* y_k,

which is convolution with {h_{k−n}^*}.

When is A self adjoint? When h_l = h_{−l}^*, i.e., h is Hermitian symmetric.

When is A unitary? When h_n * h_{n}^* = δ[|n|], i.e., when |H(ω)|² = 1.
6.6 Relations between the four spaces

The following theorem relates the null spaces and range spaces of a linear operator and its adjoint.

Remark. Luenberger uses the notation \([R(A)]\) but this seems unnecessary since \(R(A)\) is a subspace.

**Theorem.** If \(A \in B(\mathcal{X}, \mathcal{Y})\) where \(\mathcal{X}, \mathcal{Y}\) are Hilbert spaces, then

\[
\begin{align*}
1. \ {R(A)}^\perp &= N(A^*), & 2. \ \overline{R(A)} &= {N(A^*)}^\perp, \\
3. \ {R(A^*)}^\perp &= N(A), & 4. \ R(A^*) &= {N(A)}^\perp.
\end{align*}
\]

**Proof.**

Claim 1. \(\{R(A)\}^\perp = N(A^*)\)

Pick \(z \in N(A^*)\) and any \(y \in R(A)\), so \(y = Ax\) for some \(x \in \mathcal{X}\).

Now \(\langle y, z \rangle_\mathcal{Y} = \langle Ax, z \rangle_\mathcal{Y} = \langle x, A^*z \rangle_\mathcal{X} = \langle x, 0 \rangle_\mathcal{X} = 0\).

Thus \(z \in N(A^*) \Rightarrow z \in \{R(A)\}^\perp\) since \(y \in R(A)\) was arbitrary. So \(N(A^*) \subseteq \{R(A)\}^\perp\).

Now pick \(y \in \{R(A)\}^\perp\). Then for all \(x \in \mathcal{X}, 0 = \langle Ax, y \rangle_\mathcal{Y} = \langle x, A^*y \rangle_\mathcal{X}\). So \(A^*y = 0\), i.e., \(y \in N(A^*)\).

Since \(y \in \{R(A)\}^\perp\) was arbitrary, \(\{R(A)\}^\perp \subseteq N(A^*)\). Combining: \(\{R(A)\}^\perp = N(A^*)\)

Claim 2. \(\overline{R(A)} = \{N(A^*)\}^\perp\)

Taking the orthogonal complement of part 1 yields \(\{R(A)\}^\perp = N(A^*)^\perp\).

Recall from proposition 3.4-1 that \(S^\perp = \overline{S}\) when \(S\) is a subset in a Hilbert space. Since \(R(A)\) is a subspace, \([R(A)] = R(A)\).

Parts 3 and 4 follow by applying 1 and 2 to \(A^*\) and using the fact that \(A^{**} = A\).

Example. (To illustrate why we need closure in \(\overline{R(A)}\) above. Modified from [4, p. 156].)

Consider the linear operator \(A : \ell_2 \to \ell_2\) defined by \(Ax = (x_1, x_2/\sqrt{2}, x_3/\sqrt{3}, \ldots)\). Clearly \(x \in \ell_2 \Rightarrow Ax \in \ell_2\) and \(\|Ax\| \leq \|x\|\), so \(A\) is bounded (and hence continuous). In fact \(\|A\| = 1\) (consider \(x = e_1\)).

Clearly \(R(A)\) includes all finitely nonzero sequences, so \(\overline{R(A)} = \ell_2\).

However, \(y = (1, 1/2, 1/3, \ldots) \notin R(A)\) (Why not?) yet \(y \in \ell_2\), so \(R(A)\) is not closed.

This problem never arises in finite-dimensional spaces!

Example. Consider \(A : \ell_2 \to \ell_2\) defined by the upsampling operator: \(Ax = (x_1, 0, x_2, 0, x_3, 0 \ldots)\).

It is easily shown that \(A\) is bounded and \(\|A\| = 1\).

One can verify that the adjoint operation is downsampling: \(A^*y = (y_1, y_3, y_5, \ldots)\).

Clearly \(R(A^*) = \ell_2\) and \(N(A) = \{0\} = \{R(A^*)\}^\perp\).

Furthermore, \(R(A) = \bigcup_{i=0}^{\infty} e_{2i+1}\) and \(N(A^*) = \bigcup_{i=1}^{\infty} e_{2i}\) and these two spaces are orthogonal complements of one another.
Example. Consider $A : \mathcal{L}_2[\mathbb{R}] \to \mathcal{L}_2[\mathbb{R}]$ (which is complete [3, p. 589]) defined by the shifted filtering operator:

$$y = Ax \iff y(t) = (Ax)(t) = \int_{-\infty}^{\infty} \text{sinc}^2(t - 3 - \tau) x(\tau) \, d\tau,$$

where $\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t}$ or unity if $t = 0$. The adjoint is

$$(A^*y)(t) = \int_{-\infty}^{\infty} \text{sinc}^2(t + 3 - \tau) y(\tau) \, d\tau.$$

The nullspace of $A^*$ consists of signals in $\mathcal{L}_2$ whose spectrum is zero over the frequencies $(-1/2, 1/2)$. The range of $A$ is all signals in $\mathcal{L}_2$ that are band-limited to that same range, so the orthogonal complement is the same as $N(A^*)$. (Picture).

Exercise. Why did I use $\text{sinc}^2(\cdot)$ rather than $\text{sinc}(\cdot)$? ??

6.7

Duality relations for convex cones

skip

6.8

Geometric interpretation of adjoints

(Presented in terms of general adjoints.)

skip

When $A \in B(\mathcal{X}, \mathcal{Y})$ with $\mathcal{X}$ and $\mathcal{Y}$ Hilbert spaces, consider the following hyperplanes:

$$V_1 = \left\{ x \in \mathcal{X} : \langle x, A^*y \rangle_{\mathcal{X}} = 1 \right\} \text{ for some } y \in \mathcal{Y}, \quad V_2 = \left\{ y \in \mathcal{Y} : \langle Ax_0, y \rangle_{\mathcal{Y}} = 1 \right\} \text{ for some } x_0 \in \mathcal{X}.$$
Optimization in Hilbert spaces

Consider the problem of “solving” \( y = Ax \), where \( y \in \mathcal{Y} \) is given, \( x \in \mathcal{X} \) is unknown, \( A \in B(\mathcal{X}, \mathcal{Y}) \) is given, and \( \mathcal{X} \) and \( \mathcal{Y} \) are Hilbert spaces. For any such \( y \), there are three possibilities for \( x \):

- a unique solution,
- no solution,
- multiple solutions.

If \( A \) is invertible, then there is a unique solution \( x = A^{-1}y \), which is the least interesting case.

6.9

The normal equations (No exact solutions, so we seek a minimum-norm, unconstrained approximation.)

We previously explored the normal equations in a setting where \( R(A) \) was finite dimensional. Now we have the tools to generalize.

The following theorem illustrates the fundamental role of adjoints in optimization.

**Theorem.** Let \( A \in B(\mathcal{X}, \mathcal{Y}) \) where \( \mathcal{X} \) and \( \mathcal{Y} \) are Hilbert spaces.

For a fixed \( y \in \mathcal{Y} \), a vector \( x \in \mathcal{X} \) minimizes \( \| y - Ax \|_\mathcal{Y} \) iff \( A^*Ax = A^*y \).

**Proof.** Consider the subspace \( M = R(A) \). Then the minimization problem is equivalent to \( \inf_{m \in M} \| y - m \| \).

By the pre-projection theorem, \( m_* \in M \) achieves the infimum iff \( y - m_* \perp M \), i.e., \( y - m_* \in M^\perp = [R(A)]^\perp = N(A^*) \), by a previous theorem. Thus, \( 0 = A^*(y - m_*) = A^*y - A^*Ax \), for some \( x \in \mathcal{X} \).

- There is no claim of existence here, since \( R(A) \) might not be closed.
- There is no claim of uniqueness of \( x \), here, since although \( m_* \) will be unique, there may be multiple solutions to \( m_* = Ax \).
- If a minimum distance solution \( x_* \) exists and \( A^*A \) is invertible, then the solution is unique and has the familiar form:

\[
x_* = (A^*A)^{-1}A^*y.
\]

Example. Find minimum-norm approximation to \( y \in \mathcal{H} \) of the form \( \hat{y} = \sum_{i=1}^n \alpha_i x_i \), where the \( x_i \)'s are linearly independent vectors in \( \mathcal{H} \).

We know how to solve this from Ch. 3, but the operator notation provides a concise expression.

Define the operator \( A \in B(\mathbb{C}^n, \mathcal{H}) \) by

\[
A\alpha \triangleq \sum_{i=1}^n \alpha_i x_i \text{ where } \alpha = (\alpha_1, \ldots, \alpha_n).
\]

Note that \( \| A\alpha \| = \sqrt{\alpha^*G\alpha} \leq \sqrt{\lambda_{\text{max}}(G)} \| \alpha \| \) where \( G = A^*A \) is the Gram matrix. Since the \( x_i \)'s are linearly independent, \( G \) is symmetric positive definite so its eigenvalues are real and positive. So \( A \) is bounded and \( \| A \| = \sqrt{\lambda_{\text{max}}(G)} \).

Our goal is to minimize \( \| y - A\alpha \| \) over \( \alpha \in \mathbb{C}^n \). By the preceding theorem, the optimal solution must satisfy \( A^*A\alpha = A^*y \).

What is \( A^*: \mathcal{H} \to \mathbb{C}^n \) here? Recall we need \( \langle A\alpha, y \rangle_\mathcal{H} = \langle \alpha, A^*y \rangle_{\mathbb{C}^n} \), \( \forall \alpha \in \mathbb{C}^n \), so

\[
\langle A\alpha, y \rangle_\mathcal{H} = \left\langle \sum_{i=1}^n \alpha_i x_i, y \right\rangle_\mathcal{H} = \sum_{i=1}^n \alpha_i \langle x_i, y \rangle_\mathcal{H} = \sum_{i=1}^n \alpha_i \langle y, x_i \rangle_\mathcal{H} = \sum_{i=1}^n \alpha_i [A^*y]_i^* = \langle \alpha, A^*y \rangle_{\mathbb{C}^n},
\]

where we see that \( [A^*y]_i = \langle y, x_i \rangle_\mathcal{H} \) and hence

\[
A^*y = (\langle y, x_1 \rangle_\mathcal{H}, \ldots, \langle y, x_n \rangle_\mathcal{H}).
\]

Thus one can easily show that \( A^*Ax = A^*y \) is equivalent to the usual normal equations.

So no computational effort has been saved, but the notation is more concise. Furthermore, the notation \( (A^*A)^{-1}A^*y \) is comfortably similar to the notation \( (A^TA)^{-1}A^Ty \) that we use for the least-squares solution of linear systems of equations in Euclidean space. So with everything defined appropriately, the generalization to arbitrary Hilbert spaces is very natural.
6.10

The dual problem (for minimum norm solutions)

If \( y = Ax \) has multiple solutions, then in some contexts it is reasonable to choose the solution that minimizes some type of norm. However, the appropriate norm is not necessarily the norm induced by the inner product.

Exercise. Generalize Luenberger’s treatment to weighted norms.

---

**Theorem.** Let \( X \) and \( Y \) be Hilbert spaces and \( A \in B(X,Y) \). Suppose \( y \in R(A) \) is given, i.e., \( y = Ax_0 \) for some \( x_0 \in X \). Assume that \( R(A^*) \) is closed in \( X \). **ERROR in Luenberger p. 161.** The unique vector \( x_* \in X \) having minimum norm and satisfying \( Ax = y \) is characterized by:

\[
x_* = \{ A^* z : AA^* z = y, \ z \in Y \}.
\]

**Proof.** Since \( x_0 \) is one solution to \( Ax = y \), the general solution has the form \( x = x_0 - m \), where \( m \in M \triangleq N(A) \).

In other words, we seek the minimum norm vector in the linear variety

\[
V = \{ x \in X : Ax = y \} = \{ x_0 - m : m \in M \}.
\]

Since \( A \) is continuous, \( N(A) \) is a closed subspace (homework). Thus \( V \) is a closed linear variety in a Hilbert space, and as such has a unique element \( x_* \) of minimum norm by the (generalized) projection theorem, and that element is characterized by the two conditions \( x_* = x_0 - m_* \perp M \), and \( x_* \in V \).

Since \( R(A^*) \) was assumed closed, by the previous “4-space” theorem we have \( M^\perp = \{ N(A) \}^\perp = R(A^*) = R(A^*) \).

Thus \( x_* \perp M \Rightarrow x_* \in M^\perp = R(A^*) \Rightarrow x_* = A^* z \) for some \( z \in Y \). (There may be more than one such \( z \).

Furthermore, \( x_* \in V \Rightarrow Ax_* = y \Rightarrow AA^* z = y \). (\( x_* \) is unique even if there are many such \( z \) values!)

If \( AA^* \) is invertible, then the minimum norm solution has the form:

\[
x_* = A^* (AA^*)^{-1} y.
\]

---

**Example.** \( X = \mathbb{R}^3 \), \( Y = \mathbb{R}^2 \), \( A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, y = (3,3) \).

There are multiple solutions including \( x_1 = (0,0,3) \) and \( x_2 = (3,3,0) \) and convex combinations thereof.

Here \( A^* = A^T \) so \( AA^* = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, [AA^*]^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \), \( z = [AA^*]^{-1} y = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \) and \( x_* = A^* z = (1,1,2) \).

Of course \( x_* \) has a smaller \( 2 \)-norm than the other solutions above.

However, \( x_1 \) is more “sparse” which can be important in some applications.

**Example.** \( X = \mathbb{R}^1 \), \( Y = \mathbb{R}^2 \), \( A = \begin{bmatrix} 1 \end{bmatrix}, y = (2,2) \). Then, using MATLAB’s \texttt{pinv} function, \( \hat{x} = 2 \).

In this case, \( AA^* = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \), so there are multiple solutions to \( AA^* z = y \), each of which leads to the same \( \hat{x} = A^* z \) however!

Example. The “downsampling by averaging” operator \( A : \ell_2 \to \ell_2 \) is defined by \( Ax = (x_1/2 + x_2/2, x_3/2 + x_4/2, \ldots) \).

One can show that this is bounded with \( \|A\| = 1/\sqrt{2} \), since \( \|1/2 1/2\| = 1/\sqrt{2} \).

The adjoint is \( A^* y = (y_1/2, y_1/2, y_2/2, y_2/2, \ldots) \), so \( AA^* z = (z_1/2, z_2/2, \ldots) = \frac{1}{2} z \Rightarrow AA^* = \frac{1}{2} I \). So \( z = 2y \).

Thus \( x_* = A^* z = 2A^* y = (y_1, y_1, y_2, y_2, \ldots) \), which is a sensible solution in this application.
6.11

Pseudo-inverse operators

This concept allows a more general treatment of finding “solutions” to $Ax = y$, regardless of how many such solutions.

**Definition.** Let $\mathcal{X}$ and $\mathcal{Y}$ be Hilbert spaces with $A \in B(\mathcal{X}, \mathcal{Y})$ and $R(A)$ closed in $\mathcal{Y}$. For any $y \in \mathcal{Y}$, define the following linear variety:

$$V_y = \left\{ x_1 \in \mathcal{X} : \|Ax_1 - y\|_Y = \min_{x \in \mathcal{X}} \|Ax - y\|_Y \right\}.$$  

Among all vectors $x_1 \in V_y$, let $x_0$ be the unique vector of minimum norm $\|\cdot\|_X$.

The **pseudo-inverse** $A^+$ of $A$ is the operator mapping each $y$ in $\mathcal{Y}$ into its corresponding $x_0$. So $A^+ : \mathcal{Y} \to \mathcal{X}$.

Note: closure of $R(A)$ usually arises from one of $\mathcal{X}$ or $\mathcal{Y}$ being finite dimensional.

This definition is legitimate since $\min_{x \in \mathcal{X}} \|y - Ax\| = \min_{m \in M = R(A)} \|y - m\|$ where $R(A)$ is assumed closed.

By the projection theorem there is a unique $\hat{y} \in M = R(A)$ of minimum distance to $y$.

However, the linear variety $V = \{ x \in \mathcal{X} : \hat{y} = Ax \}$ may nevertheless contain multiple points.

**What about uniqueness of the vector having minimum norm?**

The set $\{ x_1 \in \mathcal{X} : Ax_1 = \hat{y} \}$ is a linear variety, a translate of $N(A)$, which is closed. Why?

So by the Ch. 3 theorem on minimizing norms within a linear variety, $x_0$ is unique. Thus $A^+$ is well defined.

If $A$ is invertible, then $x_0 = A^{-1}y$ will of course be the minimizer, in which case we have $A^+ = A^{-1}$.

Often $A^+$ is many-to-one since many $y$ vectors will map to the same $x_0$.

**Geometric interpretation** (The above definition is algebraic.)

Since $N(A)$ is a closed subspace in the Hilbert space $\mathcal{X}$, by the theorem on orthogonal complements we have

$$\mathcal{X} = N(A) \oplus \{N(A)^\perp \}.$$  

Similarly, since we have assumed that $R(A)$ is closed (and it is a subspace):

$$\mathcal{Y} = R(A) \oplus \{R(A)^\perp \}.$$  

When restricted to the subspace $\{N(A)^\perp \}$, the operator $A$ is a mapping from $\{N(A)^\perp \}$ to $R(A)$ (of course).

Between these spaces, $A$ is **one-to-one**, due to the following Lemma.

Thus $A$ has a linear inverse on $R(A)$ that maps each point in $R(A)$ back into a point in $\{N(A)^\perp \}$. This inverse defines $A^+$ on $R(A)$. To define $A^+$ on all of $\mathcal{Y}$, define $A^+y = 0$ for $y \in \{R(A)^\perp \}$.

One way to write this is: $A^+ = \left[ A \mid \{N(A)^\perp \} \right]^{-1} P_{R(A)}$.

**Lemma.** Let $A : \mathcal{X} \to \mathcal{Y}$ be a linear operator on a Hilbert space $\mathcal{X}$.

If $S \subseteq \{N(A)^\perp \}$, then $A$ is one-to-one on $S$.

**Proof.** Suppose $As_1 = As_2$, where $s_1, s_2 \in S \subseteq \{N(A)^\perp \}$, which is a subspace, so $s = s_1 - s_2 \in \{N(A)^\perp \}$.

By the linearity of $A$, we have $As = 0$, so $s \in N(A)$. Thus $s = 0$ and hence $s_1 = s_2$. So $A$ is one-to-one on $S$. \[\square\]
Example. Consider $A = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$. Then

$$N(A) = \{(a, 0) : a \in \mathbb{R}\}, \quad \{N(A)\}^\perp = \{(0, b) : b \in \mathbb{R}\},$$

$$R(A) = \{(b, b) : b \in \mathbb{R}\}, \quad \{R(A)\}^\perp = \{(a, -a) : a \in \mathbb{R}\}.$$

Recall that

$$y = P_{[v]} x \iff y = \left\langle x, \frac{v}{\|v\|} \right\rangle \frac{v}{\|v\|},$$

so $P_{R(A)} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

And clearly $\left[A \mid \{N(A)\}\right]^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, so

$$A^+ = \left[A \mid \{N(A)\}\right]^{-1} P_{R(A)} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1/2 & 1/2 \end{bmatrix}.$$

Example.

$$(aI)^+ = \begin{cases} \frac{1}{a} I, & a \neq 0 \\ 0 I, & a = 0 \end{cases} = a^+ I.$$

Here are algebraic properties of a pseudoinverse.

**Proposition.** Let $A \in B(\mathcal{X}, \mathcal{Y})$ have closed range with pseudo-inverse $A^+ : \mathcal{Y} \to \mathcal{X}$. Then

- $A^+ \in B(\mathcal{Y}, \mathcal{X})$
- $(A^+)^+ = A$
- $A^+ A A^+ = A^+$
- $A A^+ A = A$
- $(A^*)^+ = (A^+)^*$
- $(A^+ A^*)^+ = A^+ A = A^* (A^+)^*$
- $A^+ = (A^* A)^{-1} A^*$ if $A^* A$ is invertible
- $A^+ = A^* (A^* A)^{-1}$ if $A A^*$ is invertible
- In finite-dimensional spaces, a “simple” formula is given in terms of the SVD of $A$.

**Proof.** (Exercise)

**L6.19:**

$$A^+ = \lim_{\varepsilon \to 0^+} [A^* A + \varepsilon I]^{-1} A^* = \lim_{\varepsilon \to 0^+} A^* [A A^* + \varepsilon I]^{-1},$$

where the limits represent convergence with respect to what norm? [??]
One property that is missing is $(CB)^+ \neq B^+ C^+$, unlike with adjoints and inverses.

However, $L\,6.21$ claims that if $B$ is onto and $C$ is one-to-one, then $(CB)^+ = B^+ C^+$.

Example.

$$C = [1\; 1], \quad C^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad C^+ = C^* (CC^*)^{-1} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad B^* = [1 \; 0], \quad B^+ = (B^* B)^{-1} B^* = [1 \; 0].$$

In this case, $CB = 1$, but $B^+ C^+ = 1/2$.

Example.  Something involving DTFT/filtering following DTFT analysis.

Example. Downsampling by averaging. (Handwritten notes.)

Example. (p.165)

From regularization design in tomography, form a LS approximation of the form: $f(\phi) \approx \sum_{k=0}^{3} \alpha_k \cos^2 \left( \phi - \frac{\pi}{4} k \right)$.

But those cos terms are linearly dependent!

More generally, if the $x_i$’s may be linearly dependent. how to we work with

$$\min_{\alpha \in \mathbb{C}^n} \left\| y - \sum_{i=1}^{n} \alpha_i x_i \right\|.$$

Define $A : \mathbb{C}^n \to \mathcal{H}$ by $\alpha = \sum_{i=1}^{n} \alpha_i x_i$, where $y, x_i \in \mathcal{H}$.

If minimizing $\alpha$ not unique, we could choose the one of minimum norm.

Here $R(A)$ is closed since it is a finite-dimensional subspace of $\mathcal{H}$.

So $A^+$ exists and $A^+ = (A^* A)^+ A^*$, where $G = A^* A$ is simply the $n \times n$ Gram matrix. So the minimum norm LS solution is

$$\hat{\alpha} = A^+ y = (A^* A)^+ A^* y.$$

Since $G$ is (Hermitian) symmetric nonnegative definite, it has an orthogonal eigenvector decomposition $G = QDQ'$, and one can show that $G^+ = QD^+ Q'$. 
Analysis of the DTFT

Given a discrete-time signal \( g(n) \), i.e., \( g : \mathbb{Z} \to \mathbb{C} \), the **discrete-time Fourier transform** or DTFT is “defined” in introductory signal processing books as follows:

\[
G = \mathcal{F}g \iff G(\omega) = \sum_{n=-\infty}^{\infty} g(n) e^{-i\omega n}.
\]  

(6-2)

This is an infinite series, so for a rigorous treatment we must find suitable normed spaces in which we can establish convergence. The natural family of norms is \( \ell_p \), for some \( 1 \leq p \leq \infty \). Why? What about the doubly infinite sum? ??

The logical meaning of the above definition is really

\[
G = \mathcal{F}g \iff G = \lim_{N \to \infty} \mathcal{F}_N g, \text{ where } G_N = \mathcal{F}_N g \iff G_N(\omega) = \sum_{n=-N}^{N} g(n) e^{-i\omega n}, \text{ where } N \in \mathbb{N}.
\]  

(6-3)

Alternatively, one might also try to show that \( \mathcal{F} = \lim_{N \to \infty} \mathcal{F}_N \), where the limit is with respect to the operator norm in \( B(\ell_p, \ell_r[-\pi, \pi]) \) for some \( p \) and \( r \), but this is in fact false! (See below.)

- Since \( G_N(\omega) \) is only a finite sum, clearly it is always well defined.
- Furthermore, being a finite sum of complex exponentials, \( G_N(\omega) \) is continuous in \( \omega \), and hence Lebesgue integrable on \([-\pi, \pi]\).

So we could write \( \mathcal{F}_N : \mathbb{R}^{2N+1} \to \mathcal{L}_1[-\pi, \pi] \) or perhaps more usefully: \( \mathcal{F}_N : \ell_p \to \mathcal{L}_r[-\pi, \pi] \) for any \( 1 \leq p, r \leq \infty \).

To elaborate, note that by Hölder’s inequality:

\[
|G_N(\omega)| = \left| \sum_{n=-N}^{N} g(n) e^{-i\omega n} \right| \leq \sum_{n=-N}^{N} |g(n)| = \sum_{n=-\infty}^{\infty} |g(n)| \ 1_{\{|n| \leq N\}} \leq \|g\|_p \ 1_{\{|n| \leq N\}} \|_q = \|g\|_p (2N + 1)^{1-1/p}.
\]

Thus

\[
\|\mathcal{F}_N g\|_r = \|G_N\|_r = \left( \int_{-\pi}^{\pi} |G_N(\omega)|^r \ d\omega \right)^{1/r} \leq (2\pi)^{1/r} \|g\|_p (2N + 1)^{1-1/p}.
\]  

(6-4)

Furthermore, for \( p = 1 \) the upper bound is achieved when \( g(n) = \delta[n] \), so \( \|\mathcal{F}_N\|_{1 \to r} = (2\pi)^{1/r} \).

- Thus \( \mathcal{F}_N \in B(\ell_p, \mathcal{L}_r[-\pi, \pi]) \) for any \( 1 \leq p, r \leq \infty \), and \( \|\mathcal{F}_N\|_{p \to r} \leq (2\pi)^{1/r}(2N + 1)^{1-1/p} \).
- **Remark.** \( f \in \mathcal{L}_\infty[a, b] \implies f \in \mathcal{L}_r[a, b] \) if \(-\infty < a < b < \infty \) and \( r \geq 1 \).

But to make (6-2) rigorous we must have normed spaces in which the limit in (6-3) exists.

Non-convergence of the operators

Note: treating \( \mathcal{F}_N : \ell_p \to \mathcal{L}_r[-\pi, \pi] \), by considering \( g_0(n) = \delta[n - (M + 1)] \) we have for \( N > M \):

\[
\|\mathcal{F}_N - \mathcal{F}_M\|_{p \to r} = \sup_{g : \|g\|_p \leq 1} \|(\mathcal{F}_N - \mathcal{F}_M)g\|_r \geq \|(\mathcal{F}_N - \mathcal{F}_M)g_0\|_r = \left| e^{-i\omega(M + 1)} - 0 \right| = (2\pi)^{1/r}.
\]

So \( \{\mathcal{F}_N\} \) is not Cauchy (and hence not convergent) in \( B(\ell_p, \mathcal{L}_r[-\pi, \pi]) \), no matter what \( p \) or \( r \) values one chooses. So we must analyze convergence of the spectra \( G_N(\omega) = \mathcal{F}_N g \), rather than convergence of the operators \( \mathcal{F}_N \) themselves.
So we can define the DTFT operator $F$. Thus, due to completeness, 

$${\{g(n)\}}_{N=1}^\infty$$

so the sequence of function $G_N$ converges pointwise to some limit, call it $G(\omega)$, where (6-2) is shorthand for that limit. Furthermore, when $g \in \ell_1$:

$$\|G_N - G_M\|_\infty = \sup_{|\omega| \leq \pi} |G_N(\omega) - G_M(\omega)| \to 0 \text{ as } N, M \to \infty.$$ 

So the sequence of function $\{G_N\}$ is Cauchy in $L_\infty[-\pi, \pi]$, which is complete, so $\{G_N\}$ converges to a limit $G \in L_\infty[-\pi, \pi]$. More generally, using (6-6):

$$\|G_N - G_M\|_r^r = \int_{-\pi}^{\pi} |G_N(\omega) - G_M(\omega)|^r d\omega \leq 2\pi \sum_{n>M} |g(n)| \to 0 \text{ as } N, M \to \infty,$$

so $\{G_N\}$ is Cauchy in $L_r[-\pi, \pi]$ for any $1 \leq r \leq \infty$.

Thus, due to completeness, $\{G_N\}$ converges to a limit $G \in L_r[-\pi, \pi]$. So we can define the DTFT operator $F : \ell_1 \to L_r[-\pi, \pi]$ by

$$Fg \triangleq \lim_{N \to \infty} F_N g.$$ 

**Proposition.** If $g \in \ell_1$, then $\{F_N g\}$ is Cauchy in $L_r[-\pi, \pi]$ for any $1 \leq r \leq \infty$.

**Proof.** Linearity of $F$ follows from linearity of $F_N$. For $g \in \ell_1$:

$$\|Fg\|_r = \left\| \lim_{N \to \infty} F_N g \right\|_r = \lim_{N \to \infty} \|F_N g\|_r \leq \lim_{N \to \infty} \|F_N\|_{1-r} \|g\|_1 = (2\pi)^{1/r} \|g\|_1,$$

using (6-4). Equality is achieved when $g(n) = \delta[n]$. □
Unfortunately, $\ell_1$ analysis is a bit restrictive; the class of signals is not as broad as we might like, and for least-squares problems we would rather work in $\ell_2$. This will allow us to apply Hilbert space methods.

However, if a signal is in $\ell_2$, it is not necessarily in $\ell_1$, so the above $\ell_1$ analysis does not apply to many signals in $\ell_2$. So we need a different approach.

**Proposition.** If $g \in \ell_2$, then $\{F_N g\}$ is Cauchy in $L_2[-\pi, \pi]$.

**Proof.** If $g \in \ell_2$, then using $G_N = F_N g$:

$$
\|G_N - G_M\|^2_2 = \int_{-\pi}^{\pi} \left| \sum_{n=-N}^{N} g(n) e^{-i \omega n} - \sum_{n=-M}^{M} g(n) e^{-i \omega n} \right|^2 d\omega = \int_{-\pi}^{\pi} \left| \sum_{n \in I(N,M)} g(n) e^{-i \omega n} \right|^2 d\omega
$$

$$
= \sum_{n \in I(N,M)} \sum_{m \in I(N,M)} g(n) g^*(m) \int_{-\pi}^{\pi} e^{-i \omega (n-m)} d\omega = \sum_{n \in I(N,M)} \sum_{m \in I(N,M)} g(n) g^*(m) 2\pi 1_{\{n=m\}}
$$

$$
= 2\pi \sum_{n \in I(N,M)} |g(n)|^2 \leq 2\pi \sum_{|n|>\min(N,M)} |g(n)|^2 \rightarrow 0 \quad \text{as} \quad N, M \rightarrow \infty,
$$

since $g \in \ell_2$. Thus $\{F_N g\}$ is Cauchy in $L_2[-\pi, \pi]$. □

Since $L_2[-\pi, \pi]$ is complete, $\{G_N\}$ is convergent (in the $L_2$ sense!) to some limit $G \in L_2[-\pi, \pi]$, and the expression in (6-2) is again a reasonable “shorthand” for that limit, whatever it may be, and now we can define $F : \ell_2 \rightarrow L_2[-\pi, \pi]$ via (6-7). In other words,

$$\|G_N - G_M\|^2_2 = \int_{-\pi}^{\pi} |G_N(\omega) - G(\omega)|^2 d\omega \rightarrow 0.$$

This is often called **mean square** convergence.

**Proposition.** $F \in B(\ell_2, L_2[-\pi, \pi])$ with $\|F\|_{2 \rightarrow 2} = \sqrt{2\pi}$.

**Proof.** Linearity of $F$ is easily shown.

Since $\{G_N\}$ is convergent it is bounded. In fact

$$
\|G_N\|^2_2 = \int_{-\pi}^{\pi} \left| \sum_{n=-N}^{N} g(n) e^{-i \omega n} \right|^2 d\omega = \sum_{n=-N}^{N} \sum_{m=-N}^{N} g(n) g^*(m) \int_{-\pi}^{\pi} e^{-i \omega (n-m)} d\omega = 2\pi \sum_{n=-N}^{N} |g(n)|^2 \leq 2\pi \|g\|^2_2,
$$

so $\|F g\|_2 \leq \sqrt{2\pi} \|g\|_2$ and hence $\|F\| \leq \sqrt{2\pi}$.

Furthermore, if we consider $g(n) = \delta[n]$, then $G(\omega) = 1$.

Thus $\|G\|^2_2 = \int_{-\pi}^{\pi} 1 d\omega = 2\pi$ which achieves the upper bound above. Hence $\|F\|_{2 \rightarrow 2} = \sqrt{2\pi}$. □
Adjoint

Since $F \in B(\ell_2, L_2[-\pi, \pi])$ and both $\ell_2$ and $L_2[-\pi, \pi]$ are Hilbert spaces, $F$ has an adjoint:

$$\langle Fg, S \rangle_{L_2[-\pi, \pi]} = \int_{-\pi}^{\pi} S^*(\omega) \left( \sum_n g(n) e^{-i\omega n} \right) d\omega = \sum_n g(n) \left( \int_{-\pi}^{\pi} S(\omega) e^{i\omega n} d\omega \right)^* = \langle g, F^* S \rangle_{\ell_2}$$

where $x = F^* S \iff x(n) = \int_{-\pi}^{\pi} S(\omega) e^{i\omega n} d\omega, \forall n \in \mathbb{Z}$.

So the adjoint is almost the same as the inverse DTFT (defined below).

Range

To apply the Hilbert space methods, we would like $R(F)$ to be closed in $L_2[-\pi, \pi]$. It suffices for $R(F)$ to be onto $L_2[-\pi, \pi]$, since of course $L_2[-\pi, \pi]$ itself is closed.

**Proposition.** The DTFT $F : \ell_2 \to L_2[-\pi, \pi]$ defined by (6-2) is onto $L_2[-\pi, \pi]$.

**Proof.** Let $e_k, k \in \mathbb{Z}$ denote the family of functions $e_k(\omega) = \frac{1}{\sqrt{2\pi}} e^{-i\omega k}$.

Recall that $\{e_k\}$ is a complete orthonormal basis for $L_2[-\pi, \pi]$ [4, p. 62].

Thus, by Parseval’s relation we have for any $G \in L_2[-\pi, \pi]$:

$$G = \sum_k \langle G, e_k \rangle e_k, \text{i.e., } G(\omega) = \sum_k \langle G, e_k \rangle \frac{1}{\sqrt{2\pi}} e^{-i\omega k},$$

where

$$\sum_k | \langle G, e_k \rangle |^2 = \|G\|^2_2.$$

Thus if we define $g(k) = \langle G, e_k \rangle / \sqrt{2\pi}$ then $g \in \ell_2$ and $G = Fg$, so $G \in R(F)$.

Since $G \in L_2[-\pi, \pi]$ was arbitrary, we conclude $R(F) = L_2[-\pi, \pi]$. \[\square\]

DTFT is (almost) unitary

One can show easily that

$$\langle Fu, Fv \rangle_{L_2[-\pi, \pi]} = 2\pi \langle u, v \rangle_2.$$

Thus (since $F$ is linear and invertible and hence an isomorphism), the normalized DTFT $U = 1/\sqrt{2\pi} F$ is unitary.

So $L_2[-\pi, \pi]$ and $\ell_2$ are unitarily equivalent.
Inverse DTFT

Define a “partial inverse DTFT” operator \( R_N : L_2[−\pi, π] \to \ell_2 \) by

\[
g_N = R_N G \iff g_N(n) = \frac{1}{2\pi} \int_{−\pi}^{\pi} G(\omega) e^{i\omega n} d\omega \quad 1_{\{|n| \leq N\}} = \frac{1}{\sqrt{2\pi}} \langle G, e_n \rangle 1_{\{|n| \leq N\}}.
\]

**Proposition.** If \( G \in L_2[−\pi, π] \), then \( \{R_N G\} \) is Cauchy in \( \ell_2 \).

**Proof.**

\[
\|R_N G - R_M G\| = \frac{1}{2\pi} \sum_{k \in I(N,M)} |\langle G, e_k \rangle|^2 \leq \frac{1}{2\pi} \sum_{|k| > \min(N,M)} |\langle G, e_k \rangle|^2 \to 0 \text{ as } N, M \to \infty.
\]

Since \( \ell_2 \) is complete, \( \{R_N G\} \) converges to some limit \( g \in \ell_2 \) and we define \( R G \) to be that limit: \( RG \equiv \lim_{N \to \infty} R_N G \).

**Proposition.** \( R \in B(L_2[−\pi, π], \ell_2) \) and \( \|R\| = 1/\sqrt{2\pi} \).

**Proof.**

\[
\|R_N G\|_2^2 = \sum_{|n| \leq N} \left| \frac{1}{2\pi} \int_{−\pi}^{\pi} G(\omega) e^{i\omega n} d\omega \right|^2 = \frac{1}{2\pi} \sum_{|k| \leq N} |\langle G, e_k \rangle|^2 \leq \frac{1}{2\pi} \sum_k |\langle G, e_k \rangle|^2 = \frac{1}{2\pi} \|G\|_2^2.
\]

So \( \|R_N\| \leq 1/\sqrt{2\pi} \) and \( R_N \in B(L_2[−\pi, π], \ell_2) \).

When \( G(\omega) = 1 \) we have \( \|R_N G\|_2^2 = \|\delta[n]\|_2^2 = 1 \) and \( \|G\|^2 = 2\pi \) so \( \|R_N\| = 1/\sqrt{2\pi} \).

Proof 2. \( \|RG\|_2 = \lim_{N \to \infty} \|R_N G\|_2 \leq \lim_{N \to \infty} \frac{1}{\sqrt{2\pi}} \|G\|_2 \). Consider \( G = 1 \) to show equality.

**Proposition.** \( RF = I_{\ell_2} \) and \( FR = I_{L_2} \), so \( F^{-1} = R \), where \( I_H \) denotes the identity operator for Hilbert space \( H \).

**Proof.** Exercise.

\(\)
Convolution revisited

Using time-domain analysis, we showed previously that if \( h \in \ell_1 \) and \( Ax = h * x \) then \( A \in B(\ell_p, \ell_p) \).

We have shown \( F \in B(\ell_2, \mathcal{L}_2[\pi, \pi]) \) and \( F^{-1} \in B(\mathcal{L}_2[\pi, \pi], \ell_2) \).

Consider the “band-limiting” linear operator \( D : \mathcal{L}_2[-\pi, \pi] \to \mathcal{L}_2[-\pi, \pi] \) defined by

\[
y = Dx \iff y(\omega) = \begin{cases} x(\omega), & |\omega| \leq \pi/2 \\ 0, & \text{otherwise} \end{cases}
\]

Clearly \( D \in B(\mathcal{L}_2[-\pi, \pi], \mathcal{L}_2[-\pi, \pi]) \) and in fact \( |D| = 1 \).

Now consider \( A \triangleq F^{-1}D F \). We previously showed in the analysis of the composition of operators that \( \|ST\| \leq \|S\|\|T\| \).

So \( A \in B(\ell_2, \ell_2) \) with \( \|A\| \leq \|F^{-1}\|\|D\|\|F\| \leq \frac{1}{2\pi} \cdot 1 \cdot \sqrt{2\pi} = 1 \).

But this \( A \) represents an ideal lowpass filter, i.e., convolution with \( h(n) = \frac{1}{2} \text{sinc} \left( \frac{n}{2} \right) \). But this \( h \notin \ell_1 \).

Evidently, the convolution operator, at least in \( \ell_2 \), has a looser requirement than \( h \in \ell_1 \).

In contrast, in \( \ell_\infty \), \( h \in \ell_1 \) is both necessary and sufficient for \( A_h \in B(\ell_\infty, \ell_\infty) \).

In \( \ell_2 \), a necessary and sufficient condition is that the frequency response be bounded.

**Proposition.** \( A_h \in B(\ell_2, \ell_2) \) (with \( \|A_h\| = \|H\|_\infty \) \( \iff \|Fh\|_\infty < \infty \), i.e., \( Fh \in L_\infty[-\pi, \pi] \).

**Proof.** (\( \Leftarrow \)) Suppose \( \|Fh\|_\infty \) is finite. Then using the convolution property of the DTFT and Parseval:

\[
\|h * x\|_2^2 = \frac{1}{2\pi} \|H X\|_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H(\omega)X(\omega)|^2 \, d\omega \leq \|H\|_\infty^2 \frac{1}{2\pi} \|X\|_2^2 = \|H\|_\infty^2 \|x\|_2^2,
\]

\( \|A_h\| \leq \|Fh\|_\infty = \|H\|_\infty \). The upper bounded is achieved when \( h(n) = \delta[n] \).

(\( \Rightarrow \)) If \( H \) is unbounded, then forall \( T \) there exists an interval over which \( H \geq T \). Choose \( x \) to be a signal whose spectrum is an indicator function on that interval, and then \( \|x * h\|_2 \geq T \|x\|_2 \), so \( A_h \) would be be unbounded. Take contrapositive. \( \square \)

Continuous-time case

Fourier transform

convolution

Young’s inequality