Chapter 6

Linear operators and adjoints

Contents

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 $\|\boldsymbol{y} - A\boldsymbol{x}\|_2$ is minimized when $\boldsymbol{x} = [A'A]^{-1}A'\boldsymbol{y}$.

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 $T : D \to \mathcal{Y}$ when T is a transformation from a set D in a vector space $\mathcal{X}$ to a vector space $\mathcal{Y}$.

The set D is called the **domain** of T. The **range** of T is denoted

$$R(T) = \{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{y} = T(\boldsymbol{x}) \text{ for } \boldsymbol{x} \in D\}.$$

If $S \subseteq D$, then the **image** of S is given by

$$T(S) = \{\boldsymbol{y} \in \mathcal{Y} : \boldsymbol{y} = T(\boldsymbol{s}) \text{ for } \boldsymbol{s} \in S\}.$$

If $P \subseteq \mathcal{Y}$, then the **inverse image** of P is given by

$$T^{-1}(P) = \{\boldsymbol{x} \in D : T(\boldsymbol{x}) \in P\}.$$

Notation: for a **linear operator** A, we often write $A\boldsymbol{x}$ instead of $A(\boldsymbol{x})$.

For linear operators, we can always just use $D = \mathcal{X}$, so we largely ignore D hereafter.

Definition. The **nullspace** of a linear operator A is $N(A) = \{\boldsymbol{x} \in \mathcal{X} : A\boldsymbol{x} = \boldsymbol{0}\}$.
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 $\mathcal{X}$.
Furthermore, if A is continuous (in a normed space $\mathcal{X}$), then $N(A)$ is **closed** [3, p. 241].

Exercise. The range of a linear operator is a **subspace** of $\mathcal{Y}$.

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

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

Luenberger does not mention that $\mathcal{Y}$ needs to be a normed space too.

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

Definition. The smallest such M is called the **norm** of T and is denoted $|\!|\!| T |\!|\!|$. Formally:

$$|\!|\!| T |\!|\!| \triangleq \inf \{M \in \mathbb{R} : \|T(\boldsymbol{x})\| \leq M \|\boldsymbol{x}\|, \ \forall \boldsymbol{x} \in \mathcal{X}\}.$$

Consequently:

$$\|T(\boldsymbol{x})\|_{\mathcal{Y}} \leq |\!|\!| T |\!|\!| \, \|\boldsymbol{x}\|_{\mathcal{X}}, \qquad \forall \boldsymbol{x} \in \mathcal{X}.$$

Fact. For a *linear* operator A, an equivalent expression (used widely!) for the operator norm is

$$|\!|\!| A |\!|\!| = \sup_{\|\boldsymbol{x}\| \leq 1} \|A\boldsymbol{x}\|.$$

Fact. If $\mathcal{X}$ is the trivial vector space consisting only of the vector **0**, then $|\!|\!| A |\!|\!| = 0$ for any linear operator A.

Fact. If $\mathcal{X}$ is a nontrivial vector space, then for a *linear* operator A we have the following equivalent expressions:

$$|\!|\!| A |\!|\!| = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \sup_{\|\boldsymbol{x}\| = 1} \|A\boldsymbol{x}\|.$$

Example. Consider $\mathcal{X} = (\mathbb{R}^n, \|\cdot\|_\infty)$ and $\mathcal{Y} = \mathbb{R}$.

Clearly $A\boldsymbol{x} = a_1 x_1 + \cdots + a_n x_n$ so $\|A\boldsymbol{x}\|_{\mathcal{Y}} = |A\boldsymbol{x}| = |a_1 x_1 + \cdots + a_n x_n| \leq |a_1||x_1| + \cdots + |a_n||x_n| \leq (|a_1| + \cdots + |a_n|) \|\boldsymbol{x}\|_\infty$.
In fact, if we choose $\boldsymbol{x}$ such that $x_i = \operatorname{sgn}(a_i)$, then $\|\boldsymbol{x}\|_\infty = 1$ we get equality above. So we conclude $|||A||| = |a_1| + \cdots + |a_n|$.

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

Proposition. A linear operator is bounded iff it is continuous.

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

More facts related to linear operators.

- If $A : \mathcal{X} \to \mathcal{Y}$ is linear and $\mathcal{X}$ is a finite-dimensional normed space, then A is **continuous** [3, p. 268].
- If $A : \mathcal{X} \to \mathcal{Y}$ is a transformation where $\mathcal{X}$ and $\mathcal{Y}$ are normed spaces, then A is **linear** and **continuous** iff $A(\sum_{i=1}^\infty \alpha_i \boldsymbol{x}_i) = \sum_{i=1}^\infty \alpha_i A(\boldsymbol{x}_i)$ for all convergent series $\sum_{i=1}^\infty \alpha_i \boldsymbol{x}_i$. [3, p. 237].
  This is the **superposition principle** as described in introductory signals and systems courses.

Spaces of bounded linear operators

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

$$\begin{aligned}(T_1 + T_2)(\boldsymbol{x}) &= T_1(\boldsymbol{x}) + T_2(\boldsymbol{x}) \\ (\alpha\, T_1)(\boldsymbol{x}) &= \alpha(T_1(\boldsymbol{x})).\end{aligned}$$

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

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

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

This space is analogous to certain types of **dual** spaces (see *Ch. 5*).

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

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

Proof. (sketch)

Claim 1. $B(\mathcal{X}, \mathcal{Y})$ is a vector space.
Suppose $T_1, T_2 \in B(\mathcal{X}, \mathcal{Y})$ and $\alpha \in \mathcal{F}$.

$\|(\alpha\, T_1 + T_2)(\boldsymbol{x})\|_{\mathcal{Y}} = \|\alpha\, T_1(\boldsymbol{x}) + T_2(\boldsymbol{x})\|_{\mathcal{Y}} \leq |\alpha|\, \|T_1(\boldsymbol{x})\|_{\mathcal{Y}} + \|T_2(\boldsymbol{x})\|_{\mathcal{Y}} \leq |\alpha| |||\, T_1 \,|||\, \|\boldsymbol{x}\|_{\mathcal{X}} + |||\, T_2 \,|||\, \|\boldsymbol{x}\|_{\mathcal{X}} = K \|\boldsymbol{x}\|_{\mathcal{X}}$, 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:

$$|||\, T_1 + T_2 \,||| = \sup_{\|\boldsymbol{x}\|=1} \|(T_1 + T_2)\boldsymbol{x}\|_{\mathcal{Y}} \leq \sup_{\|\boldsymbol{x}\|=1} \|T_1\, \boldsymbol{x}\|_{\mathcal{Y}} + \sup_{\|\boldsymbol{x}\|=1} \|T_2\, \boldsymbol{x}\|_{\mathcal{Y}} = |||\, T_1 \,||| + |||\, T_2 \,|||.$$

□

Are there other valid norms for $B(\mathcal{X}, \mathcal{Y})$? **??**

Remark. We did not really "need" linearity in this proposition. We could have shown that the space of bounded transformations from $\mathcal{X}$ to $\mathcal{Y}$ 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. $\forall \boldsymbol{x} \in \mathcal{X}$, the sequence $\{T_n(\boldsymbol{x})\}$ is Cauchy in $\mathcal{Y}$.

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

Since $\mathcal{Y}$ is complete, for any $\boldsymbol{x} \in \mathcal{X}$ the sequence $\{T_n(\boldsymbol{x})\}$ 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 $T(\boldsymbol{x}) \triangleq \lim_{n\to\infty} T_n(\boldsymbol{x})$.

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

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

(Recall that in a normed space, if $\boldsymbol{u}_n \to \boldsymbol{u}$ and $\boldsymbol{v}_n \to \boldsymbol{v}$, then $\alpha\boldsymbol{u}_n + \boldsymbol{v}_n \to \alpha\boldsymbol{u} + \boldsymbol{v}$.)

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 $\boldsymbol{x} \in \mathcal{X}$:

$$\|T(\boldsymbol{x})\|_{\mathcal{Y}} = \lim_{n\to\infty} \|T_n(\boldsymbol{x})\|_{\mathcal{Y}} \leq \lim_{n\to\infty} |\!|\!|T_n|\!|\!| \, \|\boldsymbol{x}\|_{\mathcal{X}} \leq K \, \|\boldsymbol{x}\|_{\mathcal{X}}\,.$$

Claim 3: $T_n \to T$
Since $\{T_n\}$ is Cauchy,

$$\begin{aligned}
\forall \varepsilon > 0,\ \exists N \in \mathbb{N} \text{ s.t. } n, m > N \quad &\Longrightarrow \quad |\!|\!|T_n - T_m|\!|\!| \leq \varepsilon \\
&\Longrightarrow \quad \|T_n(\boldsymbol{x}) - T_m(\boldsymbol{x})\|_{\mathcal{Y}} \leq \varepsilon \, \|\boldsymbol{x}\|_{\mathcal{X}}\,,\ \forall \boldsymbol{x} \in \mathcal{X} \\
&\Longrightarrow \quad \lim_{m\to\infty} \|T_n(\boldsymbol{x}) - T_m(\boldsymbol{x})\|_{\mathcal{Y}} \leq \varepsilon \, \|\boldsymbol{x}\|_{\mathcal{X}}\,,\ \forall \boldsymbol{x} \in \mathcal{X} \\
&\Longrightarrow \quad \|T_n(\boldsymbol{x}) - T(\boldsymbol{x})\|_{\mathcal{Y}} \leq \varepsilon \, \|\boldsymbol{x}\|_{\mathcal{X}}\,,\ \forall \boldsymbol{x} \in \mathcal{X} \text{ by continuity of the norm} \\
&\Longrightarrow \quad |\!|\!|T_n - T\,|\!|\!| \leq \varepsilon.
\end{aligned}$$

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. □

Corollary. $(B(\mathcal{X}, \mathbb{R}), |\!|\!|\cdot|\!|\!|)$ 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 $(T\,S)(\boldsymbol{x}) = T(S(\boldsymbol{x}))$.

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

Proof. Linearity of the composition of linear operators is trivial to show.
To show that the composition is bounded: $\|T\,S\boldsymbol{x}\|_{\mathcal{Z}} \leq |\!|\!|\,T\,|\!|\!| \, \|S\boldsymbol{x}\|_{\mathcal{Y}} \leq |\!|\!|\,T\,|\!|\!| |\!|\!|S|\!|\!| \, \|\boldsymbol{x}\|_{\mathcal{X}}\,.$ □

Does it follow that $|\!|\!|\,T\,S|\!|\!| = |\!|\!|\,T\,|\!|\!| |\!|\!|S|\!|\!|$? **??**

6.3

Inverse operators

Definition. $T : \mathcal{X} \to \mathcal{Y}$ is called **one-to-one** mapping of $\mathcal{X}$ into $\mathcal{Y}$ iff $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$ and $\boldsymbol{x}_1 \neq \boldsymbol{x}_2 \Longrightarrow T(\boldsymbol{x}_1) \neq T(\boldsymbol{x}_2)$.

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

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 $T(\boldsymbol{x}) = \boldsymbol{y}$ iff $T^{-1}(\boldsymbol{y}) = \boldsymbol{x}$.

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 $N(A) = \{\mathbf{0}\}$.

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 $A^{-1}(\boldsymbol{y}_1) = \boldsymbol{x}_1$, $A^{-1}(\boldsymbol{y}_2) = \boldsymbol{x}_2$, $A(\boldsymbol{x}_1) = \boldsymbol{y}_1$, and $A(\boldsymbol{x}_2) = \boldsymbol{y}_2$. Then by the linearity of A we have $A(\alpha\boldsymbol{x}_1 + \boldsymbol{x}_2) = \alpha A\boldsymbol{x}_1 + A\boldsymbol{x}_2 = \alpha\boldsymbol{y}_1 + \boldsymbol{y}_2$, so $A^{-1}(\alpha\boldsymbol{y}_1 + \boldsymbol{y}_2) = \alpha\boldsymbol{x}_1 + \boldsymbol{x}_2 = \alpha A^{-1}(\boldsymbol{y}_1) + A^{-1}(\boldsymbol{y}_2)$. □

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.

$$\mathcal{X}, \mathcal{Y} \text{ Banach and } A \in B(\mathcal{X}, \mathcal{Y}) \text{ and } A \text{ invertible} \Longrightarrow A^{-1} \in B(\mathcal{Y}, \mathcal{X})$$

Equivalence of spaces (one way to use operators)

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

spaces	relation	operator	requirements
vector	isomorphic	isomorphism	linear, onto, 1-1
normed	topologically isomorphic	topological isomorphism	linear, onto, invertible, continuous
normed	isometrically isomorphic	isometric isomorphism	linear, onto, norm preserving $\Longrightarrow$ 1-1, continuous
inner product	unitarily equivalent	unitary	isometric isomorphism that preserves inner products

---

Isomorphic spaces

Definition. Vector spaces $\mathcal{X}$ and $\mathcal{Y}$ are called **isomorphic** (think: "same structure") iff there exists a **one-to-one**, **linear** mapping T of $\mathcal{X}$ **onto** $\mathcal{Y}$.
In such cases the mapping T is called an **isomorphism** of $\mathcal{X}$ onto $\mathcal{Y}$.
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 $\mathcal{X} = \mathbb{R}^2$ and $\mathcal{Y} = \{f(t) = a + bt \text{ on } [0,1] : a, b \in \mathbb{R}\}$.
An isomorphism is $f = T(\boldsymbol{x}) \iff f(t) = a + bt$ where $\boldsymbol{x} = (a, b)$, 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 $\mathcal{X}$ and $\mathcal{Y}$ are called **topologically isomorphic** iff there exists a **continuous** linear transformation T of $\mathcal{X}$ **onto** $\mathcal{Y}$ 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 $\mathcal{X}$ and range $\mathcal{Y}$ and positive constants m and M s.t.* $m \|\boldsymbol{x}\|_{\mathcal{X}} \le \|T\boldsymbol{x}\|_{\mathcal{Y}} \le M \|\boldsymbol{x}\|_{\mathcal{X}}, \ \forall \boldsymbol{x} \in \mathcal{X}$.

Example. In the previous example, consider $(\mathcal{X}, \|\cdot\|_\infty)$ and $(\mathcal{Y}, \|\cdot\|_2)$. Then for the same T described in that example:
$\boldsymbol{x} = (a, b) \Longrightarrow \|T(\boldsymbol{x})\|_{\mathcal{Y}}^2 = \int_0^1 (a + bt)^2 \, dt = a^2 + ab + b^2/3 \le a^2 + |a|\,|b| + b^2/3 \le (1 + 1 + 1/3) \|\boldsymbol{x}\|_\infty^2 = 7/3 \|\boldsymbol{x}\|_\infty^2$.
Also $\|T(\boldsymbol{x})\|_{\mathcal{Y}}^2 = a^2 + ab + b^2/3 = (a + b/2)^2 + b^2/12 = (a\sqrt{3}/2 + b/\sqrt{3})^2 + a^2/4 \ge \|x\|_\infty^2 /12$.
So $\mathcal{X}$ and $\mathcal{Y}$ 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 $\mathcal{X}$ and $\mathcal{Y}$ be normed spaces. A mapping $T : \mathcal{X} \to \mathcal{Y}$ is called **norm preserving** iff $\|T(\boldsymbol{x})\|_{\mathcal{Y}} = \|\boldsymbol{x}\|_{\mathcal{X}}, \ \forall \boldsymbol{x} \in \mathcal{X}$.

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.*, $T(\boldsymbol{x}) = T(\boldsymbol{z}) \Longrightarrow \boldsymbol{x} = \boldsymbol{z}$.

Proof. If $T(\boldsymbol{x}) = \boldsymbol{y}$ and $T(\boldsymbol{z}) = \boldsymbol{y}$, then by linearity $T(\boldsymbol{x} - \boldsymbol{z}) = \boldsymbol{0}$.
So since T is norm preserving, $\|\boldsymbol{x} - \boldsymbol{z}\| = \|\boldsymbol{0}\| = 0$, so $\boldsymbol{x} = \boldsymbol{z}$. □

. . . .

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

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

Remark. To illustrate why we require **onto** here, consider $T : \mathbb{E}^n \to \ell_2$ defined by $T(\boldsymbol{x}) = (x_1, \ldots, x_n, 0, 0, \ldots)$.
This T is linear, one-to-one, and norm preserving, but not onto.

Exercise. Every normed space $\mathcal{X}$ is isometrically isomorphic to a dense subset of a Banach space $\hat{\mathcal{X}}$.
(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 $\mathcal{Y} = \ell_p(\mathbb{N}) = \{(a_1, a_2, \ldots) : \sum_{i=1}^\infty |a_i|^p < \infty\}$
and $\mathcal{X} = \ell_p(\mathbb{Z}) = \left\{(\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots) : \sum_{i=-\infty}^\infty |a_i|^p < \infty\right\}$. Define the mapping $T : \mathcal{X} \to \mathcal{Y}$ by $\boldsymbol{y} = (b_1, b_2, \ldots) = T(\boldsymbol{x})$ if $b_i = a_{z(i)}$ where $z(i) = (-1)^i \lfloor i/2 \rfloor$. 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** $U : \mathcal{X} \to \mathcal{Y}$ of $\mathcal{X}$ **onto** $\mathcal{Y}$ that preserves inner products: $\langle U\boldsymbol{x}_1, U\boldsymbol{x}_2\rangle_{\mathcal{Y}} = \langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle_{\mathcal{X}}, \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$. The mapping U is called a **unitary operator**.

Fact. If U is unitary, then U is norm preserving since $\|U\boldsymbol{x}\| = \langle U\boldsymbol{x}, U\boldsymbol{x}\rangle = \langle \boldsymbol{x}, \boldsymbol{x}\rangle = \|\boldsymbol{x}\|$. 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. ($\Longleftarrow$) was just argued above.
($\Longrightarrow$) Suppose U is an isometric isomorphism.
Using the parallelogram law, the linearity of U, and the fact that $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$, we have:

$$\begin{aligned} 4\langle U\boldsymbol{x}, U\boldsymbol{y}\rangle &= \|U\boldsymbol{x} + U\boldsymbol{y}\|^2 - \|U\boldsymbol{x} - U\boldsymbol{y}\|^2 + i\,\|U\boldsymbol{x} + iU\boldsymbol{y}\|^2 - i\,\|U\boldsymbol{x} - iU\boldsymbol{y}\|^2 \\ &= \|U(\boldsymbol{x} + \boldsymbol{y})\|^2 - \|U(\boldsymbol{x} - \boldsymbol{y})\|^2 + i\,\|U(\boldsymbol{x} + i\boldsymbol{y})\|^2 - i\,\|U(\boldsymbol{x} - i\boldsymbol{y})\|^2 \\ &= \|\boldsymbol{x} + \boldsymbol{y}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}\|^2 + i\,\|\boldsymbol{x} + i\boldsymbol{y}\|^2 - i\,\|\boldsymbol{x} - i\boldsymbol{y}\|^2 = 4\langle \boldsymbol{x}, \boldsymbol{y}\rangle\,. \end{aligned}$$

Thus U an isometric isomorphism $\Longrightarrow$ U unitary. □

Remark. After defining adjoints we will show that $U^{-1} = U^\star$ 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 ℓ_2 or some $\mathbb{C}^n$ [3, p. 339].

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

$$\langle T\,\boldsymbol{x}_1, T\,\boldsymbol{x}_2\rangle = \langle f, g\rangle = \int_0^1 f(t)g(t)\,\mathrm{d}t = \int_0^1 (a+bt)(c+dt)\,\mathrm{d}t = ac + \frac{ad}{2} + \frac{bc}{2} + \frac{bd}{3} = [a\ b]\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}\begin{bmatrix} c \\ d \end{bmatrix} = \boldsymbol{x}_1'\boldsymbol{G}\boldsymbol{x}_2,$$

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

$$\langle U\boldsymbol{x}_1, U\boldsymbol{x}_2\rangle = \langle T\,\boldsymbol{G}^{-1/2}\boldsymbol{x}_1, T\,\boldsymbol{G}^{-1/2}\boldsymbol{x}_2\rangle = (\boldsymbol{G}^{-1/2}\boldsymbol{x}_1)'\boldsymbol{G}(\boldsymbol{G}^{-1/2}\boldsymbol{x}_1) = \boldsymbol{x}_1'\boldsymbol{x}_1 = \langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle\,.$$

Example. The Fourier transform $F(\nu) = \int_{-\infty}^{\infty} f(t)\,\mathrm{e}^{-\imath 2\pi\nu t}\,\mathrm{d}t$ is a unitary mapping of $\mathcal{L}_2[\mathbb{R}]$ onto itself.

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

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

We will show that $\mathcal{F} \in B(\ell_2, \mathcal{L}_2[-\pi, \pi])$ and $\mathcal{F}$ is invertible. And Parseval's relation from Fourier analysis is that

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

So ℓ_2 and $\mathcal{L}_2[-\pi, \pi]$ are unitarily equivalent, and the unitary operator needed is simply $U = \frac{1}{\sqrt{2\pi}}\mathcal{F}$.

The following "extension theorem" is useful in proving that every separable Hilbert space is isometrically isomorphic to ℓ_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 $T\,|M = \tilde{T}$. 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: $T\,|M$ reads T restricted to M.

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

. . . .

Claim 1. If $\{\boldsymbol{x}_n\} \in M$ is Cauchy in M, then $\left\{\tilde{T}(\boldsymbol{x}_n)\right\}$ is Cauchy in N.

$\left\|\tilde{T}(\boldsymbol{x}_n) - \tilde{T}(\boldsymbol{x}_m)\right\|_{\mathcal{Y}} = \left\|\tilde{T}(\boldsymbol{x}_n - \boldsymbol{x}_m)\right\|_{\mathcal{Y}} \le \|\!|\, \tilde{T} \,|\!\| \, \|\boldsymbol{x}_n - \boldsymbol{x}_m\|_{\mathcal{X}} \to 0$ as $n, m \to \infty$.

. . . .

Claim 2. For $\{\boldsymbol{x}_n\}, \{\tilde{\boldsymbol{x}}_n\} \in M$, suppose $\boldsymbol{x}_n \to \boldsymbol{x} \in \mathcal{X}$ and $\tilde{\boldsymbol{x}}_n \to \boldsymbol{x}$.
Then $\left\{\tilde{T}(\boldsymbol{x}_n)\right\}$ and $\left\{\tilde{T}(\tilde{\boldsymbol{x}}_n)\right\}$ both converge and to the same limit.

$\left\{\tilde{T}(\boldsymbol{x}_n)\right\}$ and $\left\{\tilde{T}(\tilde{\boldsymbol{x}}_n)\right\}$ are both Cauchy in N by Claim 1 since $\boldsymbol{x}_n$ and $\tilde{\boldsymbol{x}}_n$ both converge.
Since $\mathcal{Y}$ is complete, $\exists \boldsymbol{y}, \tilde{\boldsymbol{y}} \in \mathcal{Y}$ s.t. $\tilde{T}(\boldsymbol{x}_n) \to \boldsymbol{y}$ and $\tilde{T}(\tilde{\boldsymbol{x}}_n) \to \tilde{\boldsymbol{y}}$.
But $\|\boldsymbol{y} - \tilde{\boldsymbol{y}}\|_{\mathcal{Y}} = \lim_{n\to\infty} \left\|\tilde{T}(\boldsymbol{x}_n) - \tilde{T}(\tilde{\boldsymbol{x}}_n)\right\|_{\mathcal{Y}} = \lim_{n\to\infty} \left\|\tilde{T}(\boldsymbol{x}_n - \tilde{\boldsymbol{x}}_n)\right\|_{\mathcal{Y}} \le \lim_{n\to\infty} \|\!|\, \tilde{T} \,|\!\| \, \|\boldsymbol{x}_n - \tilde{\boldsymbol{x}}_n\|_{\mathcal{X}}$
$= \|\!|\, \tilde{T} \,|\!\| \, \|\boldsymbol{x} - \boldsymbol{x}\|_{\mathcal{X}} = 0$, using norm continuity, linearity of $\tilde{T}$, and boundedness of $\tilde{T}$.

Now we define $T : \mathcal{X} \to \mathcal{Y}$ as follows.
For $\boldsymbol{x} \in \mathcal{X} = \overline{M}$, $\exists \{\boldsymbol{x}_n\} \in M$ such that $\boldsymbol{x}_n \to \boldsymbol{x}$, so we define $T(\boldsymbol{x}) \triangleq \lim_{n\to\infty} \tilde{T}(\boldsymbol{x}_n)$.
By Claim 2, T is well-defined, and moreover, if $\boldsymbol{x} \in M$, then $T(\boldsymbol{x}) = \tilde{T}(\boldsymbol{x})$, *i.e.*, $T\,|M = \tilde{T}$.

. . . .

Claim 3. $\|\!|\, T \,|\!\| = \|\!|\, \tilde{T} \,|\!\|$
$\forall \boldsymbol{x} \in \mathcal{X}$, $T(\boldsymbol{x}) = \lim_{n\to\infty} \tilde{T}(\boldsymbol{x}_n)$ where $\boldsymbol{x}_n \in M$ and $\boldsymbol{x}_n \to \boldsymbol{x}$.
Thus $\|T(\boldsymbol{x})\|_{\mathcal{Y}} = \lim_{n\to\infty} \left\|\tilde{T}(\boldsymbol{x}_n)\right\|_{\mathcal{Y}} \le \lim_{n\to\infty} \|\!|\, \tilde{T} \,|\!\| \, \|\boldsymbol{x}_n\|_{\mathcal{X}} = \|\!|\, \tilde{T} \,|\!\| \, \|\boldsymbol{x}\|_{\mathcal{X}}$. Thus $\|\!|\, T \,|\!\| \le \|\!|\, \tilde{T} \,|\!\|$.
However, $\forall \boldsymbol{x} \in M$, $T(\boldsymbol{x}) = \tilde{T}(\boldsymbol{x}) \implies \|\!|\, T \,|\!\| \ge \|\!|\, \tilde{T} \,|\!\|$, 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 $L_1|M = L_2|M \implies L_1 = L_2$.
For $\boldsymbol{x} \in \mathcal{X} = \overline{M}$, $\exists \{\boldsymbol{x}_n\} \in M$ s.t. $\boldsymbol{x}_n \to \boldsymbol{x}$. Thus, by the continuity of L_1 and L_2,
$L_1(\boldsymbol{x}) = \lim_{n\to\infty} L_1(\boldsymbol{x}_n) = \lim_{n\to\infty} L_2(\boldsymbol{x}_n) = L_2(\boldsymbol{x})$, since $L_1(\boldsymbol{x}_n) = L_2(\boldsymbol{x}_n)$ by $L_1|M = L_2|M$.

. . . .

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 $\boldsymbol{y} \in \mathcal{Y}$. Since $\overline{N} = \mathcal{Y}$, $\exists \{\boldsymbol{y}_n\} \in N$ s.t. $\boldsymbol{y}_n \to \boldsymbol{y}$.
By the same reasoning as in Claim 1, $\boldsymbol{x}_n = \tilde{T}^{-1}(\boldsymbol{y}_n)$ is Cauchy in $\mathcal{X}$.
Since $\mathcal{X}$ is a Banach space, $\exists \boldsymbol{x} \in \mathcal{X}$ s.t. $\boldsymbol{x}_n \to \boldsymbol{x}$.
Thus $T(\boldsymbol{x}) = \lim_{n\to\infty} \tilde{T}(\boldsymbol{x}_n) = \lim_{n\to\infty} \tilde{T}\left(\tilde{T}^{-1}(\boldsymbol{y}_n)\right) = \lim_{n\to\infty} \boldsymbol{y}_n = \boldsymbol{y}$.
Thus, $\forall \boldsymbol{y} \in \mathcal{Y}$, $\exists \boldsymbol{x} \in \mathcal{X}$ s.t. $T(\boldsymbol{x}) = \boldsymbol{y}$. □

The following theorem is an important application of the previous theorem.

Theorem. *Every separable Hilbert space $\mathcal{H}$ is isometrically isomorphic to ℓ_2.*

Proof.

$\mathcal{H}$ separable $\Longrightarrow \exists$ a countable orthonormal basis $\{\boldsymbol{e}_i\}$. (Homework.) Let $M = [\{\boldsymbol{e}_i\}_{i=1}^{\infty}]$.

Of course ℓ_2 has a countable orthonormal basis $\{\tilde{\boldsymbol{e}}_i\}$, where $\tilde{\boldsymbol{e}}_{ij} = \delta_{i-j}$ (Kronecker). Let $N = [\{\tilde{\boldsymbol{e}}_i\}_{i=1}^{\infty}]$.

Define $\tilde{T} : M \to N$ to be the linear operator for which $\tilde{T}(\boldsymbol{e}_i) = \tilde{\boldsymbol{e}}_i$.
(**Exercise.** Think about why "the" here—consider M.)

Then since any $\boldsymbol{x} \in M$ has form $\boldsymbol{x} = \sum_{i=1}^{n} c_i \boldsymbol{e}_i$ for some $n \in \mathbb{N}$:

$$\left\|\tilde{T}(\boldsymbol{x})\right\|_2 = \left\|\tilde{T}\left(\sum_{i=1}^{n} c_i \boldsymbol{e}_i\right)\right\|_2 = \left\|\sum_{i=1}^{n} c_i \tilde{T}(\boldsymbol{e}_i)\right\|_2 = \left\|\sum_{i=1}^{n} c_i \tilde{\boldsymbol{e}}_i\right\|_2 = \sqrt{\sum_{i=1}^{n} |c_i|^2} = \left\|\sum_{i=1}^{n} c_i \boldsymbol{e}_i\right\|_{\mathcal{H}} = \|\boldsymbol{x}\|, \tag{6-1}$$

so $\|\!| \tilde{T} \|\!| = 1$ on M.

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

Since ℓ_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 ℓ_2 that is onto ℓ_2 with $\|\!| T \|\!| = \|\!| \tilde{T} \|\!| = 1$.

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

Thus T is an isometric isomorphism for $\mathcal{H}$ and ℓ_2, so $\mathcal{H}$ and ℓ_2 are isometrically isomorphic. □

Corollary. $\mathcal{L}_2$ is isometrically isomorphic to ℓ_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 ℓ_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

$$\begin{aligned} \boldsymbol{e}_{2k} &= \cos(kt)/c_{2k} & \boldsymbol{e}_{2k+1} &= \sin(kt)/c_{2k+1},\ k = 0,1,2,\ldots \\ c_{2k} &= \int_{-\pi}^{\pi} \cos^2(kt)\,\mathrm{d}t & c_{2k+1} &= \int_{-\pi}^{\pi} \sin^2(kt)\,\mathrm{d}t \\ \mathcal{X} &= \overline{[\{\boldsymbol{e}_{2k}\}]} = \{f \in \mathcal{L}_2 : f \text{ even}\} & \mathcal{Y} &= \overline{[\{\boldsymbol{e}_{2k+1}\}]} = \{f \in \mathcal{L}_2 : f \text{ odd}\}. \end{aligned}$$

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 ℓ_2.

Example. ℓ_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 ℓ_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

$$\langle A\boldsymbol{x},\, \boldsymbol{y}\rangle_{\mathbb{C}^n} = \boldsymbol{y}' A\boldsymbol{x} = (A'\boldsymbol{y})'\boldsymbol{x} = \langle \boldsymbol{x},\, A'\boldsymbol{y}\rangle_{\mathbb{C}^m}.$$

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

$$g_{\boldsymbol{y}} : \mathcal{X} \to \mathbb{C}, \text{ where } g_{\boldsymbol{y}}(\boldsymbol{x}) \triangleq \langle A\boldsymbol{x},\, \boldsymbol{y}\rangle_{\mathcal{Y}}.$$

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

In other words, $g_{\boldsymbol{y}} \in \mathcal{X}^\star$.

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

$$g_{\boldsymbol{y}}(\boldsymbol{x}) = \langle \boldsymbol{x},\, \boldsymbol{z}_{\boldsymbol{y}}\rangle_{\mathcal{X}}, \qquad \forall \boldsymbol{x} \in \mathcal{X}.$$

So we can define legitimately a mapping $A^\star : \mathcal{Y} \to \mathcal{X}$, called the **adjoint** of A, by the relationship $\boldsymbol{z}_{\boldsymbol{y}} = A^\star(\boldsymbol{y})$.
The defining property of $A^\star$ is then:

$$\boxed{\langle A\boldsymbol{x},\, \boldsymbol{y}\rangle_{\mathcal{Y}} = \langle \boldsymbol{x},\, A^\star(\boldsymbol{y})\rangle_{\mathcal{X}},\ \forall \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{y} \in \mathcal{Y}.}$$

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

Lemma. $A^\star$ is the only mapping of $\mathcal{Y}$ to $\mathcal{X}$ that satisfies the preceding equality.
Proof. For any $\boldsymbol{y} \in \mathcal{Y}$, suppose $\forall \boldsymbol{x} \in \mathcal{X}$ we have $\langle A\boldsymbol{x},\, \boldsymbol{y}\rangle_{\mathcal{Y}} = \langle \boldsymbol{x},\, T_1(\boldsymbol{y})\rangle_{\mathcal{X}} = \langle \boldsymbol{x},\, T_2(\boldsymbol{y})\rangle_{\mathcal{X}}$.
Then $0 = \langle \boldsymbol{x},\, T_1(\boldsymbol{y}) - T_2(\boldsymbol{y})\rangle_{\mathcal{X}}$ so $T_1(\boldsymbol{y}) = T_2(\boldsymbol{y}),\ \forall \boldsymbol{y} \in \mathcal{Y}$. □

Exercise. Here are some simple facts about adjoints, all of which concur with those of Hermitian transpose in Euclidean space.
- $\langle \alpha\boldsymbol{x},\, \boldsymbol{y}\rangle = \langle \boldsymbol{x},\, \alpha^*\boldsymbol{y}\rangle$ so for $A : \mathcal{X} \to \mathcal{X}$ defined by $A\boldsymbol{x} = \alpha\boldsymbol{x}$ we have $A^\star\boldsymbol{y} = \alpha^*\boldsymbol{y}$. So "reuse" of the asterix is acceptable.
- $I^\star = I$, $\quad 0^\star_{\mathcal{X}\to\mathcal{Y}} = 0_{\mathcal{Y}\to\mathcal{X}}$
- $A^{\star\star} = A$ (see Thm below)
- $(ST)^\star = T^\star S^\star$
- $(S+T)^\star = S^\star + T^\star$
- $(\alpha A)^\star = \alpha^* A^\star$

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

<u>Example</u>. Consider $\mathcal{X} = \mathcal{L}_2[0,1]$, $\mathcal{Y} = \mathbb{C}^2$, and $A : \mathcal{X} \to \mathcal{Y}$ is defined by

$$\boldsymbol{y} = A\boldsymbol{x} \iff y_1 = [A\boldsymbol{x}]_1 = \int_0^1 t x(t)\,\mathrm{d}t,\ y_2 = [A\boldsymbol{x}]_2 = \int_0^1 t^2 x(t)\,\mathrm{d}t.$$

We can guess that the adjoint of A is defined by $\boldsymbol{x} = A^\star\boldsymbol{y} \iff x(t) = (A^\star\boldsymbol{y})(t) = y_1 t + y_2 t^2$.
This is verified easily:

$$\langle A\boldsymbol{x},\, \boldsymbol{y}\rangle_{\mathcal{Y}} = y_1^*[A\boldsymbol{x}]_1 + y_2^*[A\boldsymbol{x}]_2 = y_1^* \int_0^1 t x(t)\,\mathrm{d}t + y_2^* \int_0^1 t^2 x(t)\,\mathrm{d}t = \int_0^1 x(t) \Big[\underbrace{y_1 t + y_2 t^2}_{(A^\star\boldsymbol{y})(t)}\Big]^* \mathrm{d}t = \langle \boldsymbol{x},\, A^\star\boldsymbol{y}\rangle_{\mathcal{X}}.$$

Theorem. *Suppose $A \in B(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces.*
- *The adjoint operator $A^\star$ is linear and bounded,* i.e., $A^\star \in B(\mathcal{Y}, \mathcal{X})$.
- $\|A^\star\| = \|A\|$
- $(A^\star)^\star = A$

Proof.
Claim 1. $A^\star$ is linear.
$\langle \boldsymbol{x}, A^\star(\alpha \boldsymbol{y} + \boldsymbol{z}) \rangle_{\mathcal{X}} = \langle A\boldsymbol{x}, \alpha \boldsymbol{y} + \boldsymbol{z} \rangle_{\mathcal{Y}} = \alpha^* \langle A\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{Y}} + \langle A\boldsymbol{x}, \boldsymbol{z} \rangle_{\mathcal{Y}} = \alpha^* \langle \boldsymbol{x}, A^\star(\boldsymbol{y}) \rangle_{\mathcal{X}} + \langle \boldsymbol{x}, A^\star(\boldsymbol{z}) \rangle_{\mathcal{X}}$
$= \langle \boldsymbol{x}, \alpha A^\star(\boldsymbol{y}) + A^\star(\boldsymbol{z}) \rangle_{\mathcal{X}}$, which holds for all $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$, so $A^\star(\alpha \boldsymbol{y} + \boldsymbol{z}) = \alpha A^\star(\boldsymbol{y}) + A^\star(\boldsymbol{z})$ by the usual Lemma once again.

. . . .

Claim 2. $A^\star$ is bounded and $\|A^\star\| \le \|A\|$.

$$\|A^\star \boldsymbol{y}\|_{\mathcal{X}}^2 = \langle A^\star \boldsymbol{y}, A^\star \boldsymbol{y} \rangle_{\mathcal{X}} = \langle A A^\star \boldsymbol{y}, \boldsymbol{y} \rangle_{\mathcal{Y}} \le \|A A^\star \boldsymbol{y}\|_{\mathcal{Y}} \|\boldsymbol{y}\|_{\mathcal{Y}} \le \|A\| \, \|A^\star \boldsymbol{y}\|_{\mathcal{X}} \|\boldsymbol{y}\|_{\mathcal{Y}}$$

so $\|A^\star \boldsymbol{y}\|_{\mathcal{X}} \le \|A\| \, \|\boldsymbol{y}\|_{\mathcal{Y}}$ and thus $\|A^\star\| \le \|A\|$ and hence $A^\star \in B(\mathcal{Y}, \mathcal{X})$.

. . . .

Claim 3. $A^{\star\star} = A$.
Since we have shown $A^\star \in B(\mathcal{Y}, \mathcal{X})$, we can legitimately define the adjoint of $A^\star$, denoted $A^{\star\star}$, as the (bounded linear) operator that satisfies $\langle A^\star \boldsymbol{y}, \boldsymbol{x} \rangle_{\mathcal{X}} = \langle \boldsymbol{y}, A^{\star\star} \boldsymbol{x} \rangle_{\mathcal{Y}}, \ \forall \boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}$.
Since $\langle \boldsymbol{y}, A\boldsymbol{x} \rangle_{\mathcal{Y}} = \langle A\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{Y}}^* = \langle \boldsymbol{x}, A^\star \boldsymbol{y} \rangle_{\mathcal{X}}^* = \langle A^\star \boldsymbol{y}, \boldsymbol{x} \rangle_{\mathcal{X}} = \langle \boldsymbol{y}, A^{\star\star} \boldsymbol{x} \rangle_{\mathcal{Y}}$, by the previous uniqueness arguments we see $A^{\star\star} = A$.

. . . .

Claim 4. $\|A^\star\| = \|A\|$.
From Claim 2 with $A^\star$: $\|A^{\star\star}\| \le \|A^\star\|$ or equivalently: $\|A\| \le \|A^\star\|$. □

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

Proof. Recalling that $\|S\,T\| \le \|S\| \|T\|$ we have by the preceding theorem:

$$\|A^\star A\| \le \|A^\star\| \|A\| = \|A\|^2 = \|A^\star\|^2,$$

$$\text{and } \ \|A\boldsymbol{x}\|_{\mathcal{Y}}^2 = \langle A\boldsymbol{x}, A\boldsymbol{x} \rangle_{\mathcal{Y}} = \langle A\boldsymbol{x}, (A^\star)^\star \boldsymbol{x} \rangle_{\mathcal{Y}} = \langle A^\star A \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathcal{X}} \le \|A^\star A \boldsymbol{x}\| \, \|\boldsymbol{x}\| \le \|A^\star A\| \, \|\boldsymbol{x}\|^2$$

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

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

Proof.

Claim 1. $A^\star : \mathcal{Y} \to \mathcal{X}$ is one-to-one into $\mathcal{X}$.
Consider $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathcal{Y}$ with $\boldsymbol{y}_1 \neq \boldsymbol{y}_2$. but suppose $A^\star \boldsymbol{y}_1 = A^\star \boldsymbol{y}_2$, so $A^\star \boldsymbol{d} = \boldsymbol{0}$ where $\boldsymbol{d} = \boldsymbol{y}_2 - \boldsymbol{y}_1 \neq \boldsymbol{0}$.
Thus $\langle \boldsymbol{x}, A^\star \boldsymbol{d} \rangle_{\mathcal{X}} = 0, \ \forall \boldsymbol{x} \in \mathcal{X}$ and hence $\langle A\boldsymbol{x}, \boldsymbol{d} \rangle_{\mathcal{Y}} = 0, \ \forall \boldsymbol{x} \in \mathcal{X}$.
Since A is invertible, we can make the "change of variables" $\boldsymbol{z} = A\boldsymbol{x}$ and hence $\langle \boldsymbol{z}, \boldsymbol{d} \rangle_{\mathcal{Y}} = 0, \ \forall \boldsymbol{z} \in \mathcal{X}$.
But this implies $\boldsymbol{d} = \boldsymbol{0}$, contradicting the supposition that $\boldsymbol{d} \neq \boldsymbol{0}$. So $A^\star$ is one-to-one.

Claim 2. $A^\star : \mathcal{Y} \to \mathcal{X}$ is onto $\mathcal{X}$.
By the **Banach inverse theorem**, since $A \in B(\mathcal{X}, \mathcal{Y})$ and A is invertible, $A^{-1} \in B(\mathcal{Y}, \mathcal{X})$, so A^{-1} has its own adjoint, $(A^{-1})^\star$.
Pick any $\boldsymbol{z} \in \mathcal{X}$. Then for any $\boldsymbol{x} \in \mathcal{X}$

$$\langle \boldsymbol{x}, \boldsymbol{z} \rangle = \langle A^{-1} A \boldsymbol{x}, \boldsymbol{z} \rangle = \langle A\boldsymbol{x}, (A^{-1})^\star \boldsymbol{z} \rangle = \langle A\boldsymbol{x}, (A^{-1})^\star \boldsymbol{z} \rangle = \langle \boldsymbol{x}, A^\star (A^{-1})^\star \boldsymbol{z} \rangle .$$

Thus $\boldsymbol{z} = A^\star \Big[\underbrace{(A^{-1})^\star \boldsymbol{z}}_{\in \mathcal{Y}} \Big]$, showing that $\boldsymbol{z}$ is in the range of $A^\star$. Since $\boldsymbol{z} \in \mathcal{X}$ was arbitrary, $A^\star$ is onto $\mathcal{X}$.

Claim 3. $(A^\star)^{-1} = (A^{-1})^\star$.
Since $A^\star$ is one-to-one and onto $\mathcal{X}$, $A^\star$ is invertible.
Furthermore, $\boldsymbol{z} = A^\star (A^{-1})^\star \boldsymbol{z} \Longrightarrow (A^\star)^{-1} \boldsymbol{z} = (A^{-1})^\star \boldsymbol{z}, \ \forall \boldsymbol{z} \in \mathcal{X}$. Thus $(A^\star)^{-1} = (A^{-1})^\star$. □

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

Definition. $A \in B(\mathcal{H}, \mathcal{H})$, where $\mathcal{H}$ is a real Hilbert space, is called **self adjoint** if $A^\star = A$.

Exercise. An orthogonal projection $\mathrm{P}_M : \mathcal{H} \to \mathcal{H}$, where M is a closed subspace in a Hilbert space $\mathcal{H}$, is self adjoint. **??**

Exercise. Conversely, if $P \in B(\mathcal{H}, \mathcal{H})$ and $P^2 = P$ and $P^\star = P$, then P is an orthogonal projection operator. *(L6.16)*

Definition. A self-adjoint bounded linear operator A on a Hilbert space $\mathcal{H}$ is **positive semidefinite** iff $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle \geq 0, \ \forall \boldsymbol{x} \in \mathcal{H}$.

Remark. It is easily shown that $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle$ 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 $U \in B(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X}$,$\mathcal{Y}$ 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 $\langle U\boldsymbol{x}, U\boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle, \ \forall \boldsymbol{x}, \boldsymbol{z} \in \mathcal{X}$,*
2. *$U^\star U = I$ and $UU^\star = I$,*
3. *U is invertible with $U^{-1} = U^\star$.*

Proof. (2 $\Longrightarrow$ 3) and (3 $\Longrightarrow$ 1) are obvious.

(1 $\Longrightarrow$ 2) If U is unitary then for all $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{X}$: $\langle \boldsymbol{x}, \boldsymbol{z} \rangle_{\mathcal{X}} = \langle U\boldsymbol{x}, U\boldsymbol{z} \rangle = \langle U^\star U\boldsymbol{x}, \boldsymbol{z} \rangle$, so $U^\star U\boldsymbol{x} = \boldsymbol{x}, \ \forall \boldsymbol{x} \in \mathcal{X}$ so $U^\star U = I_{\mathcal{X}}$.
For any $\boldsymbol{y} \in \mathcal{Y}$, since U is onto there exists an $\boldsymbol{x} \in \mathcal{X}$ s.t. $U\boldsymbol{x} = \boldsymbol{y}$. Thus $UU^\star\boldsymbol{y} = UU^\star U\boldsymbol{x} = UI_{\mathcal{X}}\boldsymbol{x} = U\boldsymbol{x} = \boldsymbol{y}$.
Since $\boldsymbol{y} \in \mathcal{Y}$ was arbitrary, $UU^\star = I_{\mathcal{Y}}$. □

Remark. A corollary is that if U is unitary, then so is $U^\star$.

Remark. To see why we need *both* $U^\star U = I$ and $UU^\star = I$ above, consider $U = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, for which $U^\star U = I$ but U is not onto.

Example.

Consider $\mathcal{X} = \mathcal{Y} = \ell_2$ and the (linear) discrete-time convolution operator $A : \ell_2 \to \ell_2$ defined by

$$\boldsymbol{z} = A\boldsymbol{x} \iff z_n = \sum_{k=-\infty}^{\infty} h_{n-k} x_k, \quad n \in \mathbb{Z},$$

where we assume that $h \in \ell_1$, which is equivalent to BIBO stability. We showed previously that $\|A\boldsymbol{x}\|_2 \leq \|h\|_1 \|\boldsymbol{x}\|_2$, so A is bounded with $\|\!|A|\!\| \leq \|h\|_1$, so A has an adjoint. (Later we will show $\|\!|A|\!\| = \|H\|_\infty$ where H is the frequency response of $\boldsymbol{h}$.)

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

$$\langle A\boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{n=-\infty}^{\infty} y_n^* \left[\sum_{k=-\infty}^{\infty} x_k h_{n-k} \right] = \sum_{k=-\infty}^{\infty} x_k \left[\sum_{n=-\infty}^{\infty} y_n h_{n-k}^* \right]^* = \sum_{k=-\infty}^{\infty} x_k [A^\star \boldsymbol{y}]_k^* = \langle \boldsymbol{x}, A^\star \boldsymbol{y} \rangle,$$

where the adjoint is

$$[A^\star \boldsymbol{y}]_k = \sum_{n=-\infty}^{\infty} h_{n-k}^* y_n \Longrightarrow [A^\star \boldsymbol{y}]_n = \sum_{k=-\infty}^{\infty} h_{k-n}^* y_k,$$

which is convolution with $\left\{ h_{k-n}^* \right\}$.

When is A self adjoint? When $h_l = h_{-l}^*$, *i.e.*, h is **Hermitian symmetric**.
When is A unitary? When $h_n * h_{-n}^* = \delta[n]$, *i.e.*, when $|H(\omega)|^2 = 1$.

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*

$$1.\ \{R(A)\}^\perp = N(A^\star), \qquad 2.\ \overline{R(A)} = \{N(A^\star)\}^\perp,$$
$$3.\ \{R(A^\star)\}^\perp = N(A), \qquad 4.\ \overline{R(A^\star)} = \{N(A)\}^\perp.$$

Proof.

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

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

Now $\langle \boldsymbol{y}, \boldsymbol{z} \rangle_{\mathcal{Y}} = \langle A\boldsymbol{x}, \boldsymbol{z} \rangle_{\mathcal{Y}} = \langle \boldsymbol{x}, A^\star \boldsymbol{z} \rangle_{\mathcal{X}} = \langle \boldsymbol{x}, \boldsymbol{0} \rangle_{\mathcal{X}} = 0$.
Thus $\boldsymbol{z} \in N(A^\star) \Longrightarrow \boldsymbol{z} \in \{R(A)\}^\perp$ since $\boldsymbol{y} \in R(A)$ was arbitrary. So $N(A^\star) \subset \{R(A)\}^\perp$.

Now pick $\boldsymbol{y} \in \{R(A)\}^\perp$. Then for all $\boldsymbol{x} \in \mathcal{X}$, $0 = \langle A\boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{Y}} = \langle \boldsymbol{x}, A^\star \boldsymbol{y} \rangle_{\mathcal{X}}$. So $A^\star \boldsymbol{y} = 0$, *i.e.*, $\boldsymbol{y} \in N(A^\star)$.
Since $\boldsymbol{y} \in \{R(A)\}^\perp$ was arbitrary, $\{R(A)\}^\perp \subset N(A^\star)$. Combining: $\{R(A)\}^\perp = N(A^\star)$

. . . .

Claim 2. $\overline{R(A)} = \{N(A^\star)\}^\perp$,

Taking the orthogonal complement of part 1 yields $\{R(A)\}^{\perp\perp} = N(A^\star)^\perp$.
Recall from proposition 3.4-1 that $S^{\perp\perp} = \overline{[S]}$ 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 $A^\star$ and using the fact that $A^{\star\star} = A$. □

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 $A\boldsymbol{x} = (x_1, x_2/\sqrt{2}, x_3/\sqrt{3}, \ldots)$. Clearly $\boldsymbol{x} \in \ell_2 \Longrightarrow A\boldsymbol{x} \in \ell_2$ and $\|A\boldsymbol{x}\| \leq \|\boldsymbol{x}\|$, so A is bounded (and hence continuous). In fact $\|\!|A|\!\| = 1$ (consider $\boldsymbol{x} = \boldsymbol{e}_1$).
Clearly $R(A)$ includes all finitely nonzero sequences, so $\overline{R(A)} = \ell_2$.
However, $\boldsymbol{y} = (1, 1/2, 1/3, \ldots) \notin R(A)$ (Why not?) yet $\boldsymbol{y} \in \ell_2$, 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: $A\boldsymbol{x} = (x_1, 0, x_2, 0, x_3, 0 \ldots)$.
It is easily shown that A is bounded and $\|\!|A|\!\| = 1$.
One can verify that the adjoint operation is **downsampling**: $A^\star \boldsymbol{y} = (y_1, y_3, y_5, \ldots)$.
Clearly $R(A^\star) = \ell_2$ and $N(A) = \{\boldsymbol{0}\} = [R(A^\star)]^\perp$.
Furthermore, $R(A) = \overline{[\cup_{i=0}^\infty \boldsymbol{e}_{2i+1}]}$ and $N(A^\star) = \overline{[\cup_{i=1}^\infty \boldsymbol{e}_{2i}]}$ 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:

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

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

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

The nullspace of $A^\star$ 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 $N(A^\star)$. ***(Picture)*** .

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

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 **hyperplane**s:

$$V_1 = \left\{\boldsymbol{x} \in \mathcal{X} : \langle \boldsymbol{x},\, A^\star \boldsymbol{y} \rangle_{\mathcal{X}} = 1\right\} \text{ for some } \boldsymbol{y} \in \mathcal{Y}, \qquad V_2 = \left\{\boldsymbol{y} \in \mathcal{Y} : \langle A\boldsymbol{x}_0,\, \boldsymbol{y} \rangle_{\mathcal{Y}} = 1\right\} \text{ for some } \boldsymbol{x}_0 \in \mathcal{X}.$$

Optimization in Hilbert spaces

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

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

If A is invertible, then there is a unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{y}$, 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 $\boldsymbol{y} \in \mathcal{Y}$, a vector $\boldsymbol{x} \in \mathcal{X}$ minimizes $\|\boldsymbol{y} - A\boldsymbol{x}\|_{\mathcal{Y}}$ iff $A^\star A\boldsymbol{x} = A^\star \boldsymbol{y}$.

Proof. Consider the subspace $M = R(A)$. Then the minimization problem is equivalent to $\inf_{\boldsymbol{m}\in M} \|\boldsymbol{y} - \boldsymbol{m}\|$.
By the pre-projection theorem, $\boldsymbol{m}_\star \in M$ achieves the infimum iff $\boldsymbol{y} - \boldsymbol{m}_\star \perp M$, *i.e.*, $\boldsymbol{y} - \boldsymbol{m}_\star \in M^\perp = [R(A)]^\perp = N(A^\star)$, by a previous theorem. Thus, $\boldsymbol{0} = A^\star(\boldsymbol{y} - \boldsymbol{m}_\star) = A^\star\boldsymbol{y} - A^\star A\boldsymbol{x}$, for some $\boldsymbol{x} \in \mathcal{X}$. □

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

$$\boldsymbol{x}_\star = (A^\star A)^{-1} A^\star \boldsymbol{y}.$$

Example. Find minimum-norm approximation to $\boldsymbol{y} \in \mathcal{H}$ of the form $\hat{\boldsymbol{y}} = \sum_{i=1}^n \alpha_i \boldsymbol{x}_i$, where the $\boldsymbol{x}_i$'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

$$A\boldsymbol{\alpha} \triangleq \sum_{i=1}^n \alpha_i \boldsymbol{x}_i \text{ where } \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n).$$

Note that $\|A\boldsymbol{\alpha}\| = \sqrt{\boldsymbol{\alpha}'\boldsymbol{G}\boldsymbol{\alpha}} \le \sqrt{\lambda_{\max}(\boldsymbol{G})}\,\|\boldsymbol{\alpha}\|$ where $\boldsymbol{G} = A^\star A$ is the Gram matrix. Since the $\boldsymbol{x}_i$'s are linearly independent, $\boldsymbol{G}$ is symmetric positive definite so its eigenvalues are real and positive. So A is bounded and $|||A||| = \sqrt{\lambda_{\max}(\boldsymbol{G})}$.

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

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

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

where we see that $[A^\star\boldsymbol{y}]_i = \langle \boldsymbol{y},\, \boldsymbol{x}_i\rangle_{\mathcal{H}}$ and hence

$$A^\star\boldsymbol{y} = \left(\langle \boldsymbol{y},\, \boldsymbol{x}_1\rangle_{\mathcal{H}}, \ldots, \langle \boldsymbol{y},\, \boldsymbol{x}_n\rangle_{\mathcal{H}}\right).$$

Thus one can easily show that $A^\star A\boldsymbol{x} = A^\star\boldsymbol{y}$ is equivalent to the usual normal equations.

So no computational effort has been saved, but the notation is more concise. Furthermore, the notation $(A^\star A)^{-1}A^\star\boldsymbol{y}$ is comfortingly similar to the notation $(A'A)^{-1}A'\boldsymbol{y}$ 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 $\boldsymbol{y} = A\boldsymbol{x}$ 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 $\mathcal{X}$ and $\mathcal{Y}$ be Hilbert spaces and $A \in B(\mathcal{X}, \mathcal{Y})$. Suppose $\boldsymbol{y} \in R(A)$ is given,* i.e., *$\boldsymbol{y} = A\boldsymbol{x}_0$ for some $\boldsymbol{x}_0 \in \mathcal{X}$. Assume that $R(A^\star)$ is closed in $\mathcal{X}$.* ***ERROR in Luenberger p. 161.***
> *The* unique *vector $\boldsymbol{x}_\star \in \mathcal{X}$ having minimum norm and satisfying $A\boldsymbol{x} = \boldsymbol{y}$ is characterized by:*
>
> $$\boldsymbol{x}_\star = \{A^\star \boldsymbol{z} : AA^\star \boldsymbol{z} = \boldsymbol{y},\ \boldsymbol{z} \in \mathcal{Y}\}.$$

Proof. Since $\boldsymbol{x}_0$ is one solution to $A\boldsymbol{x} = \boldsymbol{y}$, the general solution has the form $\boldsymbol{x} = \boldsymbol{x}_0 - \boldsymbol{m}$, where $\boldsymbol{m} \in M \triangleq N(A)$.
In other words, we seek the minimum norm vector in the linear variety

$$V = \{\boldsymbol{x} \in \mathcal{X} : A\boldsymbol{x} = \boldsymbol{y}\} = \{\boldsymbol{x}_0 - \boldsymbol{m} : \boldsymbol{m} \in M\}.$$

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 $\boldsymbol{x}_\star$ of minimum norm by the (generalized) projection theorem, and that element is characterized by the two conditions $\boldsymbol{x}_\star = \boldsymbol{x}_0 - \boldsymbol{m}_\star \perp M$, and $\boldsymbol{x}_\star \in V$.

Since $R(A^\star)$ was assumed closed, by the previous "4-space" theorem we have $M^\perp = \{N(A)\}^\perp = \overline{R(A^\star)} = R(A^\star)$.
Thus $\boldsymbol{x}_\star \perp M \Longrightarrow \boldsymbol{x}_\star \in M^\perp = R(A^\star) \Longrightarrow \boldsymbol{x}_\star = A^\star \boldsymbol{z}$ for some $\boldsymbol{z} \in \mathcal{Y}$. (There may be more than one such $\boldsymbol{z}$.)
Furthermore, $\boldsymbol{x}_\star \in V \Longrightarrow A\boldsymbol{x}_\star = \boldsymbol{y} \Longrightarrow AA^\star \boldsymbol{z} = \boldsymbol{y}$. ($\boldsymbol{x}_\star$ is unique even if there are many such $\boldsymbol{z}$ values!) □

. . . .

If $AA^\star$ is invertible, then the minimum norm solution has the form:

$$\boldsymbol{x}_\star = A^\star (AA^\star)^{-1} \boldsymbol{y}.$$

. . . .

<u>Example</u>. $\mathcal{X} = \mathbb{R}^3$, $\mathcal{Y} = \mathbb{R}^2$, $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$, $\boldsymbol{y} = (3, 3)$.
There are multiple solutions including $\boldsymbol{x}_1 = (0, 0, 3)$ and $\boldsymbol{x}_2 = (3, 3, 0)$ and convex combinations thereof.
Here $A^\star = A^T$ so $AA^\star = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $[AA^\star]^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$, $\boldsymbol{z} = [AA^\star]^{-1}\boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{x}_\star = A^\star \boldsymbol{z} = (1, 1, 2)$.
Of course $\boldsymbol{x}_\star$ has a smaller 2-norm than the other solutions above.
However, $\boldsymbol{x}_1$ is more "sparse" which can be important in some applications.

<u>Example</u>. $\mathcal{X} = \mathbb{R}^1$, $\mathcal{Y} = \mathbb{R}^2$, $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\boldsymbol{y} = (2, 2)$. Then, using MATLAB's `pinv` function, $\hat{\boldsymbol{x}} = 2$.
In this case, $AA^\star = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, so there are multiple solutions to $AA^\star \boldsymbol{z} = \boldsymbol{y}$, each of which leads to the same $\hat{\boldsymbol{x}} = A^\star \boldsymbol{z}$ however!

. . . .

<u>Example</u>. The "downsampling by averaging" operator $A : \ell_2 \to \ell_2$ is defined by $A\boldsymbol{x} = (x_1/2 + x_2/2, x_3/2 + x_4/2, \ldots)$.
One can show that this is bounded with $|||A||| = 1/\sqrt{2}$, since $|||[1/2\ 1/2]||| = 1/\sqrt{2}$.
The adjoint is $A^\star \boldsymbol{y} = (y_1/2, y_1/2, y_2/2, y_2/2, \ldots)$, so $AA^\star \boldsymbol{z} = (z_1/2, z_2/2, \ldots) = \frac{1}{2}\boldsymbol{z} \Longrightarrow AA^\star = \frac{1}{2}I$. So $\boldsymbol{z} = 2\boldsymbol{y}$.
Thus $\boldsymbol{x}_\star = A^\star \boldsymbol{z} = 2A^\star \boldsymbol{y} = (y_1, y_1, y_2, y_2, \ldots)$, which is a sensible solution in this application.

6.11

Pseudo-inverse operators

This concept allows a more general treatment of finding "solutions" to $A\boldsymbol{x} = \boldsymbol{y}$, 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 $\boldsymbol{y} \in \mathcal{Y}$, define the following linear variety:

$$V_{\boldsymbol{y}} = \left\{ \boldsymbol{x}_1 \in \mathcal{X} : \|A\boldsymbol{x}_1 - \boldsymbol{y}\|_{\mathcal{Y}} = \min_{\boldsymbol{x} \in \mathcal{X}} \|A\boldsymbol{x} - \boldsymbol{y}\|_{\mathcal{Y}} \right\}.$$

Among all vectors $\boldsymbol{x}_1 \in V_{\boldsymbol{y}}$, let $\boldsymbol{x}_0$ be the unique vector of minimum norm $\|\cdot\|_{\mathcal{X}}$.
The **pseudo-inverse** A^+ of A is the operator mapping each $\boldsymbol{y}$ in $\mathcal{Y}$ into its corresponding $\boldsymbol{x}_0$. So $A^+ : \mathcal{Y} \to \mathcal{X}$.

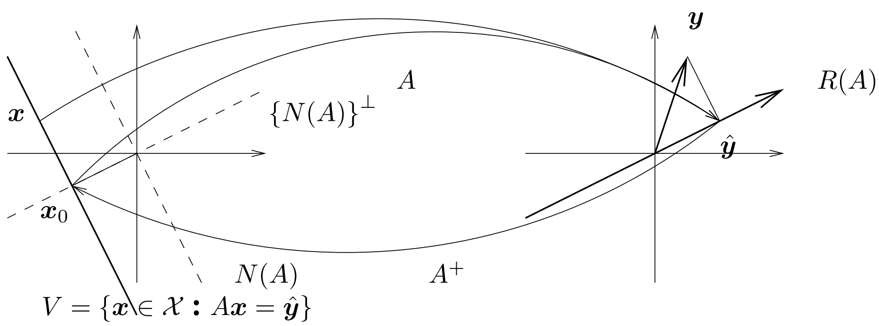
Note: closure of $R(A)$ usually arises from one of $\mathcal{X}$ or $\mathcal{Y}$ being finite dimensional.

This definition is legitimate since $\min_{\boldsymbol{x} \in \mathcal{X}} \|\boldsymbol{y} - A\boldsymbol{x}\| = \min_{\boldsymbol{m} \in M = R(A)} \|\boldsymbol{y} - \boldsymbol{m}\|$ where $R(A)$ is assumed closed.
By the projection theorem there is a unique $\hat{\boldsymbol{y}} \in M = R(A)$ of minimum distance to $\boldsymbol{y}$.
However, the linear variety $V = \{\boldsymbol{x} \in \mathcal{X} : \hat{\boldsymbol{y}} = A\boldsymbol{x}\}$ may nevertheless contain multiple points.

What about uniqueness of the vector having minimum norm?

The set $\{\boldsymbol{x}_1 \in \mathcal{X} : A\boldsymbol{x}_1 = \hat{\boldsymbol{y}}\}$ 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, $\boldsymbol{x}_0$ is unique. Thus A^+ is well defined.

If A is invertible, then $\boldsymbol{x}_0 = A^{-1}\boldsymbol{y}$ will of course be the minimizer, in which case we have $A^+ = A^{-1}$.



Often A^+ is many-to-one since many $\boldsymbol{y}$ vectors will map to the same $\boldsymbol{x}_0$.

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

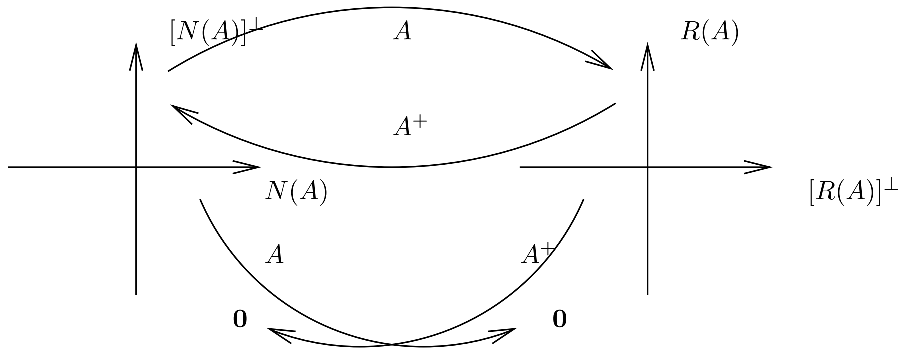
$$\mathcal{X} = N(A) \oplus \{N(A)\}^\perp.$$

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 $A^+\boldsymbol{y} = \boldsymbol{0}$ for $\boldsymbol{y} \in \{R(A)\}^\perp$.
One way to write this is: $A^+ = \left[A|\,\{N(A)\}^\perp\right]^{-1} \mathrm{P}_{R(A)}$.

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 $A\boldsymbol{s}_1 = A\boldsymbol{s}_2$. where $\boldsymbol{s}_1, \boldsymbol{s}_2 \in S \subseteq \{N(A)\}^\perp$, which is a subspace, so $\boldsymbol{s} = \boldsymbol{s}_1 - \boldsymbol{s}_2 \in \{N(A)\}^\perp$.
By the linearity of A, we have $A\boldsymbol{s} = \boldsymbol{0}$, so $\boldsymbol{s} \in N(A)$. Thus $\boldsymbol{s} = \boldsymbol{0}$ and hence $\boldsymbol{s}_1 = \boldsymbol{s}_2$. So A is one-to-one on S. □



Example. Consider $A = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$. Then

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

Recall that

$$\boldsymbol{y} = \mathrm{P}_{[\boldsymbol{v}]}\, \boldsymbol{x} \iff \boldsymbol{y} = \left\langle \boldsymbol{x}, \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} \right\rangle \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|},$$

so $\mathrm{P}_{R(A)} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} [1\ 1] = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

And clearly $\left[A \mid \{N(A)\}^\perp\right]^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, so $A^+ = \left[A \mid \{N(A)\}^\perp\right]^{-1} \mathrm{P}_{R(A)} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1/2 & 1/2 \end{bmatrix}$.

Example.

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

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

Proof. (Exercise)

L6.19:

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

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.

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

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 $\boldsymbol{x}_i$'s may be linearly dependent. how to we work with

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

Define $A : \mathbb{C}^n \to \mathcal{H}$ by $\boldsymbol{\alpha} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i$, where $\boldsymbol{y}, \boldsymbol{x}_i \in \mathcal{H}$.
If minimizing $\boldsymbol{\alpha}$ 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 $A^+ = (A^\star A)^+ A^\star$, where $\boldsymbol{G} = A^\star A$ is simply the $n \times n$ Gram matrix. So the minimum norm LS solution is

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

Since $\boldsymbol{G}$ is (Hermitian) symmetric nonnegative definite, it has an orthogonal eigenvector decomposition $\boldsymbol{G} = \boldsymbol{Q}D\boldsymbol{Q}'$, and one can show that $\boldsymbol{G}^+ = \boldsymbol{Q}D^+\boldsymbol{Q}'$.

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:

$$G = \mathcal{F}g \iff G(\omega) = \sum_{n=-\infty}^{\infty} g(n)\, \mathrm{e}^{-\imath\omega n}. \tag{6-2}$$

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 ℓ_p, for some $1 \le p \le \infty$. Why? What about the doubly infinite sum? **??**

The logical meaning of the above definition is really

$$G = \mathcal{F}g \iff G = \lim_{N\to\infty} \mathcal{F}_N g, \text{ where } G_N = \mathcal{F}_N g \iff G_N(\omega) \triangleq \sum_{n=-N}^{N} g(n)\, \mathrm{e}^{-\imath\omega n}, \text{ where } N \in \mathbb{N}. \tag{6-3}$$

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 $B(\ell_p, \mathcal{L}_r[-\pi,\pi])$ 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 ω, 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 $1 \le p, r \le \infty$.
  To elaborate, note that by Hölder's inequality:

$$|G_N(\omega)| = \left|\sum_{n=-N}^{N} g(n)\, \mathrm{e}^{-\imath\omega n}\right| \le \sum_{n=-N}^{N} |g(n)| = \sum_{n=-\infty}^{\infty} |g(n)\; 1_{\{|n|\le N\}}| \le \|g\|_p \left\| 1_{\{|n|\le N\}} \right\|_q = \|g\|_p\, (2N+1)^{1-1/p}.$$

  Thus

$$\|\mathcal{F}_N g\|_r = \|G_N\|_r = \left(\int_{-\pi}^{\pi} |G_N(\omega)|^r \,\mathrm{d}\omega\right)^{1/r} \le (2\pi)^{1/r}\, \|g\|_p\, (2N+1)^{1-1/p}. \tag{6-4}$$

  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 $1 \le p, r \le \infty$, and $|||\mathcal{F}_N|||_{p\to r} \le (2\pi)^{1/r}(2N+1)^{1-1/p}$.
- *Remark.* $f \in \mathcal{L}_\infty[a,b] \Longrightarrow f \in \mathcal{L}_r[a,b]$ if $-\infty < a < b < \infty$ and $r \ge 1$.

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 $N > M$:

$$|||\mathcal{F}_N - \mathcal{F}_M|||_{p\to r} = \sup_{g\,:\,\|g\|_p\le 1} \|(\mathcal{F}_N - \mathcal{F}_M)g\|_r \ge \|(\mathcal{F}_N - \mathcal{F}_M)g_0\|_r = \left\|\mathrm{e}^{-\imath\omega(M-1)} - 0\right\|_r = (2\pi)^{1/r}.$$

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.

ℓ_1 analysis

Proposition. If $g \in \ell_1$, then $\{\mathcal{F}_N g\}$ is Cauchy in $\mathcal{L}_r[-\pi, \pi]$ for any $1 \leq r \leq \infty$.

Proof.

If $g \in \ell_1$, then defining

$$I(N, M) \triangleq \{n \in \mathbb{Z} : \min(N, M) < |n| \leq \max(N, M)\} \tag{6-5}$$

and $G_N = \mathcal{F}_N g$ we have

$$\begin{aligned} |G_N(\omega) - G_M(\omega)| &= \left| \sum_{n=-N}^{N} g(n)\, \mathrm{e}^{-\imath\omega n} - \sum_{n=-M}^{M} g(n)\, \mathrm{e}^{-\imath\omega n} \right| = \left| \sum_{n \in I(N,M)} g(n)\, \mathrm{e}^{-\imath\omega n} \right| \\ &\leq \sum_{n \in I(N,M)} |g(n)| \leq \sum_{|n| > \min(N,M)} |g(n)| \to 0 \text{ as } N, M \to \infty. \end{aligned} \tag{6-6}$$

So for each $\omega \in \mathbb{R}$, the sequence $\{G_N(\omega)\}_{N=1}^{\infty}$ is Cauchy in $\mathbb{R}$, and hence convergent by the completeness of $\mathbb{R}$, provided $g \in \ell_1$. Thus for each ω, $\{G_N(\omega)\}$ **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 $\mathcal{L}_\infty[-\pi, \pi]$, which is complete, so $\{G_N\}$ converges to a limit $G \in \mathcal{L}_\infty[-\pi, \pi]$.

More generally, using (6-6):

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

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

....

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

$$\mathcal{F} g \triangleq \lim_{N \to \infty} \mathcal{F}_N g. \tag{6-7}$$

Proposition. $\mathcal{F} \in B(\ell_1, \mathcal{L}_r[-\pi, \pi])$ for any $1 \leq r \leq \infty$ with $|||\mathcal{F}|||_{1 \to r} = (2\pi)^{1/r}$.

Proof. Linearity of $\mathcal{F}$ follows from linearity of $\mathcal{F}_N$. For $g \in \ell_1$:

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

using (6-4). Equality is achieved when $g(n) = \delta[n]$. □

ℓ_2 analysis

Unfortunately, ℓ_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 ℓ_2. This will allow us to apply Hilbert space methods.

However, if a signal is in ℓ_2, it is not necessarily in ℓ_1, so the above ℓ_1 analysis *does not apply* to many signals in ℓ_2. So we need a different approach.

Proposition. If $g \in \ell_2$, then $\{\mathcal{F}_N g\}$ is Cauchy in $\mathcal{L}_2[-\pi, \pi]$.

Proof. If $g \in \ell_2$,then using $G_N = \mathcal{F}_N g$:

$$\begin{aligned}\|G_N - G_M\|_2^2 &= \int_{-\pi}^{\pi} \left| \sum_{n=-N}^{N} g(n)\,\mathrm{e}^{-\imath\omega n} - \sum_{n=-M}^{M} g(n)\,\mathrm{e}^{-\imath\omega n} \right|^2 \mathrm{d}\omega = \int_{-\pi}^{\pi} \left| \sum_{n\in I(N,M)} g(n)\,\mathrm{e}^{-\imath\omega n} \right|^2 \mathrm{d}\omega \\ &= \sum_{n\in I(N,M)} \sum_{m\in I(N,M)} g(n)g^*(m) \int_{-\pi}^{\pi} \mathrm{e}^{-\imath\omega(n-m)}\,\mathrm{d}\omega = \sum_{n\in I(N,M)} \sum_{m\in I(N,M)} g(n)g^*(m) 2\pi\, 1_{\{n=m\}} \\ &= 2\pi \sum_{n\in I(N,M)} |g(n)|^2 \leq 2\pi \sum_{|n|>\min(N,M)} |g(n)|^2 \to 0 \text{ as } N, M \to \infty,\end{aligned}$$

since $g \in \ell_2$. Thus $\{\mathcal{F}_N g\}$ is Cauchy in $\mathcal{L}_2[-\pi, \pi]$. □

Since $\mathcal{L}_2[-\pi, \pi]$ is complete, $\{G_N\}$ is convergent (*in the $\mathcal{L}_2$ sense!*) to some limit $G \in \mathcal{L}_2[-\pi, \pi]$, and the expression in (6-2) is again a reasonable "shorthand" for that limit, whatever it may be, and now we can define $\mathcal{F} : \ell_2 \to \mathcal{L}_2[-\pi, \pi]$ via (6-7). In other words,

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

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

Proposition. $\mathcal{F} \in B(\ell_2, \mathcal{L}_2[-\pi, \pi])$ with $|||\mathcal{F}|||_{2\to 2} = \sqrt{2\pi}$.

Proof. Linearity of $\mathcal{F}$ is easily shown.
Since $\{G_N\}$ is convergent it is bounded. In fact

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

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

Furthermore, if we consider $g(n) = \delta[n]$, then $G(\omega) = 1$.
Thus $\|G\|_2^2 = \int_{-\pi}^{\pi} 1 \,\mathrm{d}\omega = 2\pi$ which achieves the upper bound above. Hence $|||\mathcal{F}|||_{2\to 2} = \sqrt{2\pi}$. □

Adjoint

Since $\mathcal{F} \in B(\ell_2, \mathcal{L}_2[-\pi, \pi])$ and both ℓ_2 and $\mathcal{L}_2[-\pi, \pi]$ are Hilbert spaces, $\mathcal{F}$ has an adjoint:

$$\langle \mathcal{F} g,\, S \rangle_{\mathcal{L}_2[-\pi,\pi]} = \int_{-\pi}^{\pi} S^*(\omega) \left(\sum_n g(n)\, \mathrm{e}^{-\imath\omega n} \right) \mathrm{d}\omega = \sum_n g(n) \left(\int_{-\pi}^{\pi} S(\omega)\, \mathrm{e}^{\imath\omega n}\, \mathrm{d}\omega \right)^* = \langle g,\, \mathcal{F}^\star S \rangle_{\ell_2}$$

where $x = \mathcal{F}^\star S \iff x(n) = \int_{-\pi}^{\pi} S(\omega)\, \mathrm{e}^{\imath\omega n}\, \mathrm{d}\omega,\ \forall n \in \mathbb{Z}$.
So the adjoint is almost the same as the inverse DTFT (defined below).

And of course we know that $\mathcal{F}^\star \in B(\mathcal{L}_2[-\pi, \pi], \ell_2)$.

Range

To apply the Hilbert space methods, we would like $R(\mathcal{F})$ to be closed in $\mathcal{L}_2[-\pi, \pi]$. It suffices for $R(\mathcal{F})$ to be onto $\mathcal{L}_2[-\pi, \pi]$, since of course $\mathcal{L}_2[-\pi, \pi]$ itself is closed.

Proposition. The DTFT $\mathcal{F} : \ell_2 \to \mathcal{L}_2[-\pi, \pi]$ defined by (6-2) is **onto** $\mathcal{L}_2[-\pi, \pi]$.

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

Recall that $\{\boldsymbol{e}_k\}$ is a **complete orthonormal basis** for $\mathcal{L}_2[-\pi, \pi]$ [4, p. 62].

Thus, by **Parseval's relation** we have for any $G \in \mathcal{L}_2[-\pi, \pi]$:

$$G = \sum_k \langle G,\, \boldsymbol{e}_k \rangle\, \boldsymbol{e}_k, \textit{i.e.}, G(\omega) = \sum_k \langle G,\, \boldsymbol{e}_k \rangle \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\imath\omega k},$$

where

$$\sum_k |\langle G,\, \boldsymbol{e}_k \rangle|^2 = \|G\|_2^2 .$$

Thus if we define $g(k) = \langle G,\, \boldsymbol{e}_k \rangle / \sqrt{2\pi}$ then $g \in \ell_2$ and $G = \mathcal{F} g$, so $G \in R(\mathcal{F})$.
Since $G \in \mathcal{L}_2[-\pi, \pi]$ was arbitrary, we conclude $R(\mathcal{F}) = \mathcal{L}_2[-\pi, \pi]$. □

DTFT is (almost) unitary

One can show easily that

$$\langle \mathcal{F} \boldsymbol{u},\, \mathcal{F} \boldsymbol{v} \rangle_{\mathcal{L}_2[-\pi,\pi]} = 2\pi \langle \boldsymbol{u},\, \boldsymbol{v} \rangle_2 .$$

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

So $\mathcal{L}_2[-\pi, \pi]$ and ℓ_2 are unitarily equivalent.

Inverse DTFT

Define a "partial inverse DTFT" operator $\mathcal{R}_N : \mathcal{L}_2[-\pi, \pi] \to \ell_2$ by

$$g_N = \mathcal{R}_N G \iff g_N(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\omega)\, e^{\imath \omega n} \, d\omega \;\; \mathbf{1}_{\{|n| \le N\}} = \frac{1}{\sqrt{2\pi}} \langle G,\, \boldsymbol{e}_n \rangle \;\; \mathbf{1}_{\{|n| \le N\}}.$$

Proposition. If $G \in \mathcal{L}_2[-\pi, \pi]$, then $\{\mathcal{R}_N G\}$ is Cauchy in ℓ_2.

Proof.

$$\|\mathcal{R}_N G - \mathcal{R}_M G\| = \frac{1}{2\pi} \sum_{k \in I(N,M)} |\langle G,\, \boldsymbol{e}_k \rangle|^2 \le \frac{1}{2\pi} \sum_{|k| > \min(N,M)} |\langle G,\, \boldsymbol{e}_k \rangle|^2 \to 0 \text{ as } N, M \to \infty.$$

□

Since ℓ_2 is complete, $\{\mathcal{R}_N G\}$ converges to some limit $g \in \ell_2$ and we define $\mathcal{R}G$ to be that limit: $\mathcal{R}G \triangleq \lim_{N\to\infty} \mathcal{R}_N G$.

Proposition. $\mathcal{R} \in B(\mathcal{L}_2[-\pi, \pi], \ell_2)$ and $|||\mathcal{R}||| = 1/\sqrt{2\pi}$.

Proof.

$$\|\mathcal{R}_N G\|_2^2 = \sum_{|n| \le N} \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\omega)\, e^{\imath \omega n} \, d\omega \right|^2 = \frac{1}{2\pi} \sum_{|k| \le N} |\langle G,\, \boldsymbol{e}_k \rangle|^2 \le \frac{1}{2\pi} \sum_k |\langle G,\, \boldsymbol{e}_k \rangle|^2 = \frac{1}{2\pi} \|G\|_2^2 .$$

So $|||\mathcal{R}_N||| \le 1/\sqrt{2\pi}$ and $\mathcal{R}_N \in B(\mathcal{L}_2[-\pi, \pi], \ell_2)$.
When $G(\omega) = 1$ we have $\|\mathcal{R}_N G\|^2 = \|\delta[n]\|^2 = 1$ and $\|G\|^2 = 2\pi$ so $|||\mathcal{R}_N||| = 1/\sqrt{2\pi}$.

Proof 2. $\|\mathcal{R}G\|_2 = \lim_{N\to\infty} \|\mathcal{R}_N G\|_2 \le \lim_{N\to\infty} \frac{1}{\sqrt{2\pi}} \|G\|_2$. Consider $G = 1$ to show equality. □

Proposition. $\mathcal{R}\mathcal{F} = I_{\ell_2}$ and $\mathcal{F}\mathcal{R} = I_{\mathcal{L}_2}$, so $\mathcal{F}^{-1} = \mathcal{R}$, where $I_{\mathcal{H}}$ denotes the identity operator for Hilbert space $\mathcal{H}$.

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 $\mathcal{F} \in B(\ell_2, \mathcal{L}_2[-\pi, \pi])$ and $\mathcal{F}^{-1} \in B(\mathcal{L}_2[-\pi, \pi], \ell_2)$.
Consider the "band-limiting" linear operator $D : \mathcal{L}_2[-\pi, \pi] \to \mathcal{L}_2[-\pi, \pi]$ defined by

$$y = Dx \iff y(\omega) = \begin{cases} x(\omega), & |\omega| \le \pi/2 \\ 0, & \text{otherwise.} \end{cases}$$

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

Now consider $A \triangleq \mathcal{F}^{-1} D \mathcal{F}$. We previously showed in the analysis of the composition of operators that $|||ST||| \le |||S||| |||T|||$.
So $A \in B(\ell_2, \ell_2)$ with $|||A||| \le |||\mathcal{F}^{-1}||| |||D||| |||\mathcal{F}||| \le \frac{1}{\sqrt{2\pi}} \cdot 1 \cdot \sqrt{2\pi} = 1$.
But this A represents an ideal lowpass filter, *i.e.*, convolution with $h(n) = \frac{1}{2} \operatorname{sinc}\left(\frac{n}{2}\right)$. But this $h \notin \ell_1$.

Evidently, the convolution operator, at least in ℓ_2, has a looser requirement than $h \in \ell_1$.

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

In ℓ_2, a necessary and sufficient condition is that the frequency response be bounded.

Proposition. $A_h \in B(\ell_2, \ell_2)$ (with $|||A_h||| = \|H\|_\infty$) $\iff$ $\|\mathcal{F}h\|_\infty < \infty$, *i.e.*, $\mathcal{F}h \in \mathcal{L}_\infty[-\pi, \pi]$.

Proof. ($\Longleftarrow$) Suppose $\|\mathcal{F}h\|_\infty$ is finite. Then using the convolution property of the DTFT and Parseval:

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

$|||A_h||| \le \|\mathcal{F}h\|_\infty = \|H\|_\infty$. The upper bounded is achieved when $h(n) = \delta[n]$.

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

Continuous-time case

Fourier transform

convolution

Young's inequality

1. P. Enflo. A counterexample to the approximation problem in Banach spaces. *Acta Math*, 130:309–17, 1973.
2. I. J. Maddox. *Elements of functional analysis*. Cambridge, 2 edition, 1988.
3. A. W. Naylor and G. R. Sell. *Linear operator theory in engineering and science*. Springer-Verlag, New York, 2 edition, 1982.
4. D. G. Luenberger. *Optimization by vector space methods*. Wiley, New York, 1969.
5. J. Schauder. Zur theorie stetiger abbildungen in funktionenrumen. *Math. Zeitsch.*, 26:47–65, 1927.
6. L. Grafakos. *Classical and modern Fourier analysis*. Pearson, NJ, 2004.
7. P. P. Vaidyanathan. Generalizations of the sampling theorem: Seven decades after Nyquist. *IEEE Tr. Circ. Sys. I, Fundamental theory and applications*, 48(9):1094–109, September 2001.
8. A. M. Ostrowski. *Solution of equations in Euclidian and Banach spaces*. Academic, 3 edition, 1973.
9. R. R. Meyer. Sufficient conditions for the convergence of monotonic mathematical programming algorithms. *J. Comput. System. Sci.*, 12(1):108–21, 1976.
10. M. Rosenlicht. *Introduction to analysis*. Dover, New York, 1985.
11. A. R. De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. *IEEE Tr. Med. Imag.*, 12(2):328–33, June 1993.
12. A. R. De Pierro. On the convergence of the iterative image space reconstruction algorithm for volume ECT. *IEEE Tr. Med. Imag.*, 6(2):174–175, June 1987.
13. A. R. De Pierro. Unified approach to regularized maximum likelihood estimation in computed tomography. In *Proc. SPIE 3171, Comp. Exper. and Num. Meth. for Solving Ill-Posed Inv. Imaging Problems: Med. and Nonmed. Appl.*, pages 218–23, 1997.
14. J. A. Fessler. Grouped coordinate descent algorithms for robust edge-preserving image restoration. In *Proc. SPIE 3170, Im. Recon. and Restor. II*, pages 184–94, 1997.
15. A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Tr. Med. Imag.*, 14(1):132–137, March 1995.
16. J. A. Fessler and A. O. Hero. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *IEEE Tr. Im. Proc.*, 4(10):1417–29, October 1995.
17. M. W. Jacobson and J. A. Fessler. Properties of MM algorithms on convex feasible sets. *SIAM J. Optim.*, 2003. Submitted. # 061996.
18. P. L. Combettes and H. J. Trussell. Method of successive projections for finding a common point of sets in metric spaces. *J. Optim. Theory Appl.*, 67(3):487–507, December 1990.
19. F. Deutsch. The convexity of Chebyshev sets in Hilbert space. In A. Yanushauskas Th. M. Rassias, H. M. Srivastava, editor, *Topics in polynomials of one and several variables and their applications*, pages 143–50. World Sci. Publishing, River Edge, NJ, 1993.
20. M. Jiang. On Johnson's example of a nonconvex Chebyshev set. *J. Approx. Theory*, 74(2):152–8, August 1993.
21. V. S. Balaganskii and L. P. Vlasov. The problem of convexity of Chebyshev sets. *Russian Mathematical Surveys*, 51(6):1127–90, November 1996.
22. V. Kanellopoulos. On the convexity of the weakly compact Chebyshev sets in Banach spaces. *Israel Journal of Mathematics*, 117:61–9, 2000.
23. A. R. Alimov. On the structure of the complements of Chebyshev sets. *Functional Analysis and Its Applications*, 35(3):176–82, July 2001.
24. Y. Bresler, S. Basu, and C. Couvreur. Hilbert spaces and least squares methods for signal processing, 2000. Draft.
25. M. Vetterli and J. Kovacevic. *Wavelets and subband coding*. Prentice-Hall, New York, 1995.
26. D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part I—Theory. *IEEE Tr. Med. Imag.*, 1(2):81–94, October 1982.
27. M. Unser and T. Blu. Generalized smoothing splines and the optimal discretization of the wiener filter. *IEEE Tr. Sig. Proc.*, 2004. in press.