

Chapter 6

Linear Transformations and Operators

6.1 The Algebra of Linear Transformations

Theorem 6.1.1. *Let V and W be vector spaces over the field F . Let T and U be two linear transformations from V into W . The function $(T + U)$ defined pointwise by*

$$(T + U)(\underline{v}) = T\underline{v} + U\underline{v}$$

is a linear transformation from V into W . Furthermore, if $s \in F$, the function (sT) defined by

$$(sT)(\underline{v}) = s(T\underline{v})$$

is also a linear transformation from V into W . The set of all linear transformation from V into W , together with the addition and scalar multiplication defined above, is a vector space over the field F .

Proof. Suppose that T and U are linear transformation from V into W . For $(T + U)$ defined above, we have

$$\begin{aligned}(T + U)(s\underline{v} + \underline{w}) &= T(s\underline{v} + \underline{w}) + U(s\underline{v} + \underline{w}) \\ &= s(T\underline{v}) + T\underline{w} + s(U\underline{v}) + U\underline{w} \\ &= s(T\underline{v} + U\underline{v}) + (T\underline{w} + U\underline{w}) \\ &= s(T + U)\underline{v} + (T + U)\underline{w},\end{aligned}$$

which shows that $(T + U)$ is a linear transformation. Similarly, we have

$$\begin{aligned}
 (rT)(s\underline{v} + \underline{w}) &= r(T(s\underline{v} + \underline{w})) \\
 &= r(s(T\underline{v}) + (T\underline{w})) \\
 &= rs(T\underline{v}) + r(T\underline{w}) \\
 &= s(r(T\underline{v})) + rT(\underline{w}) \\
 &= s((rT)\underline{v}) + (rT)\underline{w}
 \end{aligned}$$

which shows that (rT) is a linear transformation.

To verify that the set of linear transformations from V into W together with the operations defined above is a vector space, one must directly check the conditions of Definition 3.3.1. These are straightforward to verify, and we leave this exercise to the reader. \square

We denote the space of linear transformations from V into W by $L(V, W)$. Note that $L(V, W)$ is defined only when V and W are vector spaces over the same field.

Fact 6.1.2. *Let V be an n -dimensional vector space over the field F , and let W be an m -dimensional vector space over F . Then the space $L(V, W)$ is finite-dimensional and has dimension mn .*

Theorem 6.1.3. *Let V , W , and Z be vector spaces over a field F . Let $T \in L(V, W)$ and $U \in L(W, Z)$. Then the composed function UT defined by $(UT)(\underline{v}) = U(T(\underline{v}))$ is a linear transformation from V into Z .*

Proof. Let $\underline{v}_1, \underline{v}_2 \in V$ and $s \in F$. Then, we have

$$\begin{aligned}
 (UT)(s\underline{v}_1 + \underline{v}_2) &= U(T(s\underline{v}_1 + \underline{v}_2)) \\
 &= U(sT\underline{v}_1 + T\underline{v}_2) \\
 &= sU(T\underline{v}_1) + U(T\underline{v}_2) \\
 &= s(UT)(\underline{v}_1) + (UT)(\underline{v}_2),
 \end{aligned}$$

as desired. \square

Definition 6.1.4. *If V is a vector space over the field F , a **linear operator** on V is a linear transformation from V into V .*

Definition 6.1.5. An **algebra** over a field F is a vector space V over F that has a bilinear **vector product** “ \cdot ”: $V \times V \rightarrow V$ satisfying $(s\underline{u}) \cdot (t\underline{v}) = (st)(\underline{u} \cdot \underline{v})$ and

$$(s\underline{u} + \underline{v}) \cdot (t\underline{w} + \underline{x}) = st(\underline{u} \cdot \underline{w}) + s(\underline{u} \cdot \underline{x}) + t(\underline{v} \cdot \underline{w}) + (\underline{v} \cdot \underline{x}),$$

for all $s, t \in F$ and $\underline{u}, \underline{v}, \underline{w}, \underline{x} \in V$. If V is a Banach space and the norm of the vector product satisfies $\|\underline{u} \cdot \underline{v}\| \leq \|\underline{u}\| \|\underline{v}\|$, then it is called a **Banach algebra**.

Example 6.1.6. The set $L(V, V)$ of linear operators on V forms an algebra when the vector product is defined by functional composition $UT(\underline{v}) = U(T(\underline{v}))$. If V is a Banach space and $L(V, V)$ is equipped with the induced operator norm, then it forms a Banach algebra.

Definition 6.1.7. A linear transformation T from V into W is called **invertible** if there exists a function U from W to V such that UT is the identity function on V and TU is the identity function on W . If T is invertible, the function U is unique and is denoted by T^{-1} . Furthermore, T is invertible if and only if

1. T is one-to-one: $T\underline{v}_1 = T\underline{v}_2 \implies \underline{v}_1 = \underline{v}_2$
2. T is onto: the range of T is W .

Example 6.1.8. Consider the vector space V of semi-infinite real sequences \mathbb{R}^ω where $\underline{v} = (v_1, v_2, v_3, \dots) \in V$ with $v_n \in \mathbb{R}$ for $n \in \mathbb{N}$. Let $L: V \rightarrow V$ be the left-shift linear transformation defined by

$$L\underline{v} = (v_2, v_3, v_4, \dots)$$

and $R: V \rightarrow V$ be the right-shift linear transformation defined by

$$R\underline{v} = (0, v_1, v_2, \dots).$$

Notice that L is onto but not one-to-one and R is one-to-one but not onto. Therefore, neither transformation is invertible.

Example 6.1.9. Consider the normed vector space V of semi-infinite real sequences \mathbb{R}^ω with the standard Schauder basis $\{\underline{e}_1, \underline{e}_2, \dots\}$. Let $T: V \rightarrow V$ be the linear transformation that satisfies $T\underline{e}_i = i^{-1}\underline{e}_i$ for $i = 1, 2, \dots$. Let the linear transformation $U: V \rightarrow V$ satisfy $U\underline{e}_i = i\underline{e}_i$ for $i = 1, 2, \dots$. It is easy to verify that $U = T^{-1}$ and $UT = TU = I$.

This example should actually bother you somewhat. Since T reduces vector components arbitrarily, its inverse must enlarge them arbitrarily. Clearly, this is not a desirable property. Later, we will introduce a norm for linear transforms which quantifies this problem.

Theorem 6.1.10. *Let V and W be vector spaces over the field F and let T be a linear transformation from V into W . If T is invertible, then the inverse function T^{-1} is a linear transformation from W onto V .*

Proof. Let \underline{w}_1 and \underline{w}_2 be vectors in W and let $s \in F$. Define $\underline{v}_j = T^{-1}\underline{w}_j$, for $j = 1, 2$. Since T is a linear transformation, we have

$$T(s\underline{v}_1 + \underline{v}_2) = sT(\underline{v}_1) + T(\underline{v}_2) = s\underline{w}_1 + \underline{w}_2.$$

That is, $s\underline{v}_1 + \underline{v}_2$ is the unique vector in V that maps to $s\underline{w}_1 + \underline{w}_2$ under T . It follows that

$$T^{-1}(s\underline{w}_1 + \underline{w}_2) = s\underline{v}_1 + \underline{v}_2 = s(T^{-1}\underline{w}_1) + T^{-1}\underline{w}_2$$

and T^{-1} is a linear transformation. □

A **homomorphism** is a mapping between algebraic structures which preserves all relevant structure. An **isomorphism** is a homomorphism which is also invertible. For vector spaces, the relevant structure is given by vector addition and scalar multiplication. Since a linear transformation preserves both of these operation, it is also a *vector space homomorphism*. Likewise, an invertible linear transformation is a *vector space isomorphism*.

6.2 The Dual Space

Definition 6.2.1. *Let V be a vector space. The collection of all linear functionals on V , denoted $L(V, F)$, forms a vector space. We also denote this space by V^* and call it the **dual space** of V .*

The following theorem shows that, if V is finite dimensional, then

$$\dim V^* = \dim V.$$

In this case, one actually finds that V is isomorphic to V^* . Therefore, the two spaces can be identified with each other so that $V = V^*$ for finite dimensional V .

Theorem 6.2.2. *Let V be a finite-dimensional vector space over the field F , and let $\mathcal{B} = \underline{v}_1, \dots, \underline{v}_n$ be a basis for V . There is a unique dual basis $\mathcal{B}^* = f_1, \dots, f_n$ for V^* such that $f_j(\underline{v}_i) = \delta_{ij}$. For each linear functional on V , we have*

$$f = \sum_{i=1}^n f(\underline{v}_i) f_i$$

and for each vector \underline{v} in V , we have

$$\underline{v} = \sum_{i=1}^n f_i(\underline{v}) \underline{v}_i.$$

Proof. Let $\mathcal{B} = \underline{v}_1, \dots, \underline{v}_n$ be a basis for V . According to Theorem 3.4.7, there is a unique linear functional f_i on V such that

$$f_i(\underline{v}_j) = \delta_{ij}.$$

Thus, we obtain from \mathcal{B} a set of n distinct linear functionals f_1, \dots, f_n on V . These functionals are linearly independent; suppose that

$$f = \sum_{i=1}^n s_i f_i,$$

then

$$f(\underline{v}_j) = \sum_{i=1}^n s_i f_i(\underline{v}_j) = \sum_{i=1}^n s_i \delta_{ij} = s_j.$$

In particular, if f is the zero functional, $f(\underline{v}_j) = 0$ for $j = 1, \dots, n$ and hence the scalars $\{s_j\}$ must all equal 0. It follows that the functionals f_1, \dots, f_n are linearly independent. Since $\dim V^* = n$, we conclude that $\mathcal{B}^* = f_1, \dots, f_n$ forms a basis for V^* , the **dual basis** of \mathcal{B} .

Next, we want to show that there is a unique basis which is dual to \mathcal{B} . If f is a linear functional on V , then f is some linear combination of f_1, \dots, f_n with

$$f = \sum_{i=1}^n s_i f_i.$$

Furthermore, by construction, we must have $s_j = f(\underline{v}_j)$ for $j = 1, \dots, n$. Similarly, if

$$\underline{v} = \sum_{i=1}^n t_i \underline{v}_i.$$

is a vector in V , then

$$f_j(\underline{v}) = \sum_{i=1}^n t_i f_j(\underline{v}_i) = \sum_{i=1}^n t_i \delta_{ij} = t_j.$$

That is, the unique expression for \underline{v} as a linear combination of $\underline{v}_1, \dots, \underline{v}_n$ is

$$\underline{v} = \sum_{i=1}^n f_i(\underline{v}) \underline{v}_i.$$

□

One important use of the dual space is to define the transpose of a linear transform in a way that generalizes to infinite dimensional vector spaces. Let V, W be vector spaces over F and $T: V \rightarrow W$ be a linear transform. If $g \in W^*$ is a linear functional on W (i.e., $g: W \rightarrow F$), then $g(T\underline{v}) \in F$ is a linear functional on V . The **transpose** of T is the mapping $U: W^* \rightarrow V^*$ defined by $f(\underline{v}) = g(T\underline{v}) \in F$ for all $g \in W^*$. If V, W are finite-dimensional, then one can identify $V = V^*$ and $W = W^*$ via isomorphism and recover the standard transpose mapping $U: W \rightarrow V$ implied by the matrix transpose.

The details of this definition are not used in the remainder of these notes, but can be useful in understanding the subtleties of infinite dimensional spaces. For infinite dimensional Hilbert spaces, we will see later that the definition again simplifies because one identifies $V = V^*$ via isomorphism. The interesting case that does not simplify is that of linear transforms between infinite dimensional Banach spaces.

6.3 Operator Norms

For any vector space of linear transforms, one can define a norm to get a normed vector space of linear transforms (e.g., consider the Frobenius norm of a matrix). In contrast, an operator norm is defined for linear transforms between normed spaces and it is induced by the vector norms of the underlying spaces. Intuitively, the induced operator norm is the largest factor by which a linear transform can increase the length of a vector. This defines a simple “worst-case” expansion for any linear transform.

Definition 6.3.1. Let V and W be two normed vector spaces and let $T: V \rightarrow W$ be a linear transformation. The induced **operator norm** of T is defined to

$$\|T\| = \sup_{\underline{v} \in V - \{0\}} \frac{\|T\underline{v}\|}{\|\underline{v}\|} = \sup_{\underline{v} \in V, \|\underline{v}\|=1} \|T\underline{v}\|.$$

A common question about the operator norm is, “How do I know the two expressions give the same result?”. To see this, we can write

$$\sup_{\underline{v} \in V - \{0\}} \frac{\|T\underline{v}\|}{\|\underline{v}\|} = \sup_{\underline{v} \in V - \{0\}} \left\| T \frac{\underline{v}}{\|\underline{v}\|} \right\| = \sup_{\underline{u} \in V, \|\underline{u}\|=1} \|T\underline{u}\|.$$

Previously, we have seen that the set $L(V, W)$ of linear transformations from V into W , with the standard addition and scalar multiplication, satisfies the conditions required to be a vector space. Now, we have a norm for that vector space. Interested readers should verify that the above definition satisfies the first two standard conditions required by a norm. To verify the triangle inequality, we can write

$$\begin{aligned} \|T + U\| &= \sup_{\underline{v} \in V, \|\underline{v}\|=1} \|(T + U)\underline{v}\| \\ &\leq \sup_{\underline{v} \in V, \|\underline{v}\|=1} (\|T\underline{v}\| + \|U\underline{v}\|) \\ &\leq \sup_{\underline{v} \in V, \|\underline{v}\|=1} \|T\underline{v}\| + \sup_{\underline{v} \in V, \|\underline{v}\|=1} \|U\underline{v}\| \\ &= \|T\| + \|U\|. \end{aligned}$$

The induced operator norm also has another property that follows naturally from its definition. Notice that

$$\|T\| = \sup_{\underline{v} \in V - \{0\}} \frac{\|T\underline{v}\|}{\|\underline{v}\|} \geq \frac{\|T\underline{u}\|}{\|\underline{u}\|}$$

for all non-zero $\underline{u} \in V$. Checking the special case of $\underline{u} = 0$ separately, one can show the **induced operator-norm inequality** $\|T\underline{u}\| \leq \|T\| \|\underline{u}\|$ for all $\underline{u} \in V$.

For the space $L(V, V)$ of linear operators on V , a norm is called **submultiplicative** if $\|TU\| \leq \|T\| \|U\|$ for all $T, U \in L(V, V)$. The induced operator-norm inequality shows that all induced operator norms are submultiplicative because

$$\|UT\underline{v}\| \leq \|U\| \|T\underline{v}\| \leq \|U\| \|T\| \|\underline{v}\|.$$

This also defines a submultiplicative norm for the algebra of linear operators on V .

6.3.1 Bounded Transformations

Definition 6.3.2. A linear transformation is called **bounded** if its induced operator norm is finite.

Theorem 6.3.3. A linear transformation $T: V \rightarrow W$ is bounded if and only if it is continuous.

Proof. Suppose that T is bounded; that is, there exists M such that $\|T\underline{v}\| \leq M \|\underline{v}\|$ for all $\underline{v} \in V$. Let $\underline{v}_1, \underline{v}_2, \dots$ be a convergent sequence in V , then

$$\|T\underline{v}_i - T\underline{v}_j\| = \|T(\underline{v}_i - \underline{v}_j)\| \leq M \|\underline{v}_i - \underline{v}_j\|.$$

This implies that $T\underline{v}_1, T\underline{v}_2, \dots$ is a convergent sequence in W , and T is continuous.

Conversely, assume T is continuous and notice that $T\underline{0} = \underline{0}$. Therefore, for any $\epsilon > 0$, there is a $\delta > 0$ such that $\|T\underline{v}\| < \epsilon$ for all $\|\underline{v}\| < \delta$. Since the norm of $\underline{u} = \frac{\delta \underline{v}}{2\|\underline{v}\|}$ is equal to $\delta/2$, we get

$$\|T\underline{v}\| = \left\| T \frac{\delta \underline{v}}{2\|\underline{v}\|} \right\| \frac{2\|\underline{v}\|}{\delta} < \frac{2\epsilon}{\delta} \|\underline{v}\|.$$

The value $M = \frac{2\epsilon}{\delta}$ serves as an upper bound on $\|T\|$. □

Then, by showing that linear transformations over finite-dimensional spaces are continuous, one concludes that they are also bounded. This is accomplished in the following theorem.

Theorem 6.3.4. Let V and W be normed vector spaces and let $T: V \rightarrow W$ be a linear transformation. If V is finite dimensional, then T is continuous and bounded.

Lemma 6.3.5. Let V be a finite-dimensional normed vector space, and let

$$\mathcal{B} = \underline{v}_1, \dots, \underline{v}_n$$

be a basis for V . Then, for $\underline{v} \in V$, each coefficient s_i in the expansion

$$\underline{v} = s_1 \underline{v}_1 + \dots + s_n \underline{v}_n$$

is a continuous linear function of \underline{v} . Being continuous, it is also bounded, so there exists a constant M such that $|s_i| \leq M \|\underline{v}\|$.

Proof of Lemma. The linearity property is straightforward, its proof is omitted. It will suffice to show that there is an $m > 0$ such that

$$m|s_i| \leq m(|s_1| + \cdots + |s_n|) \leq \|\underline{v}\|, \quad (6.1)$$

since (6.1) implies that $|s_i| \leq m^{-1} \|\underline{v}\|$. We first show that this holds for coefficients $\{s_1, \dots, s_n\}$ satisfying the condition $|s_1| + \cdots + |s_n| = 1$. Let

$$S = \left\{ (s_1, \dots, s_n) \mid \sum_{i=1}^n |s_i| = 1 \right\}.$$

This set is closed and bounded; it is therefore compact. Define the function $f: S \rightarrow \mathbb{R}$ by

$$f(s_1, \dots, s_n) = \|s_1 \underline{v}_1 + \cdots + s_n \underline{v}_n\|.$$

It can be shown that f is continuous, and it is clear that $f > 0$ over S . Let

$$m = \min_{(s_1, \dots, s_n) \in S} f(s_1, \dots, s_n).$$

Since f is continuous and S is compact, this minimum exists and is attained by some point $(s'_1, \dots, s'_n) \in S$. Note that $m > 0$ for otherwise $\underline{v}_1, \dots, \underline{v}_n$ are linearly dependent, contradicting the fact that \mathcal{B} is a basis. Thus m so defined satisfies (6.1).

For general sets of coefficients $\{s_i\}$, let $c = |s_1| + \cdots + |s_n|$. If $c = 0$, the result is trivial. If $c > 0$, then write

$$\begin{aligned} \|s_1 \underline{v}_1 + \cdots + s_n \underline{v}_n\| &= c \left\| \frac{s_1}{c} \underline{v}_1 + \cdots + \frac{s_n}{c} \underline{v}_n \right\| \\ &= cf \left(\frac{s_1}{c}, \dots, \frac{s_n}{c} \right) \\ &\geq cm = m(|s_1| + \cdots + |s_n|). \end{aligned}$$

This is the desired result. □

We are now ready to prove the theorem.

Proof of Theorem. Let $\mathcal{B} = \underline{v}_1, \dots, \underline{v}_n$ be a basis for V . Let $\underline{v} \in V$ be expressed in terms of this basis as

$$\underline{v} = s_1 \underline{v}_1 + \cdots + s_n \underline{v}_n.$$

Let $C = \max_{1 \leq i \leq n} \|T\underline{v}_i\|$. Then,

$$\begin{aligned} \|T\underline{v}\| &= \|T(s_1\underline{v}_1 + \cdots + s_n\underline{v}_n)\| \\ &\leq |s_1| \|T\underline{v}_1\| + \cdots + |s_n| \|T\underline{v}_n\| \\ &\leq C(|s_1| + \cdots + |s_n|). \end{aligned}$$

By the previous lemma, this implies that there exists an M such that $|s_1| + \cdots + |s_n| \leq M \|\underline{v}\|$, so that

$$\|T\underline{v}\| \leq CM \|\underline{v}\|.$$

□

6.3.2 The Neumann Expansion

Theorem 6.3.6. *Let $\|\cdot\|$ be a submultiplicative operator norm and $T: V \rightarrow V$ be a linear operator with $\|T\| < 1$. Then, $(I - T)^{-1}$ exists and*

$$(I - T)^{-1} = \sum_{i=0}^{\infty} T^i.$$

Proof. First, we observe that the sequence

$$A_n = \sum_{i=0}^{n-1} T^i.$$

is Cauchy. This follows from the fact that, for $m < n$, we have

$$\|A_n - A_m\| = \left\| \sum_{i=m}^{n-1} T^i \right\| \leq \sum_{i=m}^{n-1} \|T\|^i = \frac{\|T\|^m - \|T\|^n}{1 - \|T\|} \leq \frac{\|T\|^m}{1 - \|T\|}.$$

Since this goes to zero as $m \rightarrow \infty$, we see that the limit $\lim_{n \rightarrow \infty} A_n$ exists.

Next, we observe that

$$(I - T)(I + T + T^2 + \cdots + T^{n-1}) = I - T^n.$$

Since $\|T\| < 1$, we have $\lim_{k \rightarrow \infty} T^k = 0$ because $\|T^k\| \leq \|T\|^k \rightarrow 0$. Taking the limit $n \rightarrow \infty$ of both sides gives

$$(I - T) \sum_{i=0}^{\infty} T^i = \lim_{n \rightarrow \infty} (I - T^n) = I.$$

Likewise, reversing the order multiplication results in the same result. This shows that $\sum_{i=0}^{\infty} T^i$ must be the inverse of $I - T$. □

If one only needs to show that $I - T$ is non-singular, then proof by contradiction is somewhat simpler. Suppose $I - T$ is singular, then there exists a non-zero vector \underline{v} such that $(I - T)\underline{v} = \underline{0}$. But, this implies that $\|\underline{v}\| = \|T\underline{v}\| \leq \|T\| \|\underline{v}\|$. Since $\|\underline{v}\| \neq 0$, this gives the contradiction $\|T\| \geq 1$ and implies that $I - T$ is non-singular.

6.3.3 Matrix Norms

$$\|A\|_{\infty} = \max_{\|\underline{v}\|_{\infty}=1} \|A\underline{v}\|_{\infty} = \max_i \sum_j |a_{ij}|$$

$$\|A\|_1 = \max_{\|\underline{v}\|_1=1} \|A\underline{v}\|_1 = \max_j \sum_i |a_{ij}|$$

The 2-norm of a matrix can be found by solving

$$\max_{\underline{v}^H \underline{v}=1} \|A\underline{v}\|_2^2 = \underline{v}^H A^H A \underline{v}.$$

Using the Lagrange multiplier technique, one seeks to minimize

$$J = \underline{v}^H A^H A \underline{v} - \lambda \underline{v}^H \underline{v}.$$

Taking the gradient with respect to \underline{v} and equating the result to zero, we get

$$A^H A \underline{v} = \lambda \underline{v}.$$

The corresponding \underline{v} must be an eigenvector of the matrix $A^H A$. Left multiplying this equation by \underline{v}^H and using the fact that $\underline{v}^H \underline{v} = 1$, we obtain

$$\underline{v}^H A^H A \underline{v} = \lambda \underline{v}^H \underline{v} = \lambda.$$

Since we are maximizing the left hand side of this equation, λ must be the largest eigenvalue of $A^H A$. For an $n \times n$ matrix B with eigenvalues $\lambda_1, \dots, \lambda_n$, the **spectral radius** $\rho(B)$ is defined by

$$\rho(B) = \max_i |\lambda_i|.$$

The spectral radius of B is the smallest radius of a circle centered at the origin that contains all the eigenvalues of B . It follows that

$$\|A\|_2 = \sqrt{\rho(A^H A)}.$$

When A is Hermitian, $\|A\|_2 = \rho(A)$. The 2-norm is also called the **spectral norm**.

The **Frobenius norm** is given by

$$\|A\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}}.$$

This norm is also called the **Euclidean norm**. Note that $\|A\|_F^2 = \text{tr}(A^H A)$.

6.4 Linear Functionals on Hilbert Spaces

Let V be an inner-product space, and let \underline{v} be some fixed vector in V . Define the function $f_{\underline{v}}$ from V into F by

$$f_{\underline{v}}(\underline{w}) = \langle \underline{w}, \underline{v} \rangle.$$

Clearly, $f_{\underline{v}}$ is a linear functional on V . If V is a Hilbert space, then every continuous linear functional on V arises in this way from some vector \underline{v} . This result is known as the **Riesz representation theorem**.

Lemma 6.4.1. *If $\langle \underline{v}, \underline{w} \rangle = \langle \underline{u}, \underline{w} \rangle$ for all $\underline{w} \in V$, then $\underline{v} = \underline{u}$.*

Proof. Then, $\langle \underline{v} - \underline{u}, \underline{w} \rangle = 0$ for all $\underline{w} \in V$. Therefore, $\langle \underline{v} - \underline{u}, \underline{v} - \underline{u} \rangle = 0$ and this implies $\underline{v} - \underline{u} = \underline{0}$. \square

Theorem 6.4.2 (Riesz). *Let V be a Hilbert space and f be a continuous linear functional on V . Then, there exists a unique vector $\underline{v} \in V$ such that $f(\underline{w}) = \langle \underline{w}, \underline{v} \rangle$ for all $\underline{w} \in V$.*

Proof. While the result holds in any Hilbert space, this proof assumes V is separable for simplicity. Therefore, we let $\underline{v}_1, \underline{v}_2, \dots$ be a countable orthonormal basis for V . We wish to find a candidate vector \underline{v} for the inner product.

First, we note that f is bounded because it is continuous and, as such, there exists M such that $|f(\underline{x})| \leq M\|\underline{x}\|$ for all $\underline{x} \in V$. Let $\underline{x}_n = \sum_{i=1}^n \overline{f(\underline{v}_i)} \underline{v}_i$. For any n , we have

$$\begin{aligned} M\|\underline{x}_n\| &\geq |f(\underline{x}_n)| = \left| \sum_{i=1}^n \overline{f(\underline{v}_i)} f(\underline{v}_i) \right| = \sum_{i=1}^n |f(\underline{v}_i)|^2 = \sum_{i=1}^n f(\underline{v}_i) \overline{f(\underline{v}_i)} \\ &= \sum_{i=1}^n \left\langle \overline{f(\underline{v}_i)} \underline{v}_i, \overline{f(\underline{v}_i)} \underline{v}_i \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \left\langle \overline{f(\underline{v}_j)} \underline{v}_j, \overline{f(\underline{v}_i)} \underline{v}_i \right\rangle \\ &= \left\langle \sum_{j=1}^n \overline{f(\underline{v}_j)} \underline{v}_j, \sum_{i=1}^n \overline{f(\underline{v}_i)} \underline{v}_i \right\rangle = \langle \underline{x}_n, \underline{x}_n \rangle = \|\underline{x}_n\|^2. \end{aligned}$$

This implies that $\|\underline{x}_n\| \leq M$ for all n . Hence, $\lim_{n \rightarrow \infty} \sum_{i=1}^n |f(\underline{v}_i)|^2$ is bounded and the vector

$$\underline{v} = \sum_{i=1}^{\infty} \overline{f(\underline{v}_i)} \underline{v}_i,$$

is in V because it is the limit point of a Cauchy sequence. Let $f_{\underline{v}}$ be the functional defined by

$$f_{\underline{v}}(\underline{w}) = \langle \underline{w}, \underline{v} \rangle.$$

By the Cauchy-Schwarz, we can verify that

$$\|f_{\underline{v}}\| \triangleq \sup_{\underline{u} \in V - \{0\}} \frac{f_{\underline{v}}(\underline{u})}{\|\underline{u}\|} = \|\underline{v}\|.$$

Since f is continuous, it follows that $\|f\| < \infty$ and $\|\underline{v}\| < \infty$. Then,

$$f_{\underline{v}}(\underline{v}_j) = \left\langle \underline{v}_j, \sum_{i=1}^{\infty} \overline{f(\underline{v}_i)} \underline{v}_i \right\rangle = f(\underline{v}_j).$$

Since this is true for each \underline{v}_j , it follows that $f = f_{\underline{v}}$. Now, consider any $\underline{v}' \in V$ such that $\langle \underline{w}, \underline{v} \rangle = \langle \underline{w}, \underline{v}' \rangle$ for all $\underline{w} \in W$. Applying Lemma 6.4.1 shows that $\underline{v} = \underline{v}'$ and we conclude that \underline{v} is unique. \square

An important consequence of this theorem is that the continuous dual space V^* of a Hilbert space V is isometrically isomorphic to the original space V . Let $R: V^* \rightarrow V$ be the implied Riesz mapping from continuous linear functionals on V (i.e., V^*) to elements of V . Then, $f(\underline{v}) = \langle \underline{v}, R(f) \rangle$ for all $f \in V^*$. The isomorphism can be shown by verifying that $R(sf_1 + f_2) = \bar{s}R(f_1) + R(f_2)$ and one finds that the mapping R is conjugate linear. The mapping is isometric because $\|f\| = \|R(f)\|$. Based on this isomorphism, one can treat a Hilbert space as self-dual and assume without confusion that $V = V^*$.

Theorem 6.4.3. *Let V and W be Hilbert spaces, and assume $T: V \rightarrow W$ is a continuous linear transformation. Then, the **adjoint** is the unique linear transformation T^* on W such that*

$$\langle T\underline{v}, \underline{w} \rangle = \langle \underline{v}, T^*\underline{w} \rangle$$

for all vectors $\underline{v} \in V$, $\underline{w} \in W$.

Proof. Let \underline{w} be any vector in W . Then $f(\underline{v}) = \langle T\underline{v}, \underline{w} \rangle$ is a continuous linear functional on V . It follows from the Riesz representation theorem (Theorem 6.4.2) that there exists a unique vector $\underline{v}' \in V$ such that $f(\underline{v}) = \langle T\underline{v}, \underline{w} \rangle = \langle \underline{v}, \underline{v}' \rangle$. Of course, the vector \underline{v}' depends on the choice of \underline{w} . So, we define the adjoint mapping $T^*: W \rightarrow V$ to give the required \underline{v}' for each \underline{w} . In other words,

$$\underline{v}' = T^*\underline{w}.$$

Next, we must verify that T^* is a linear transformation. Let $\underline{w}_1, \underline{w}_2$ be in W and s be a scalar. For all $\underline{v} \in V$,

$$\begin{aligned} \langle \underline{v}, T^*(s\underline{w}_1 + \underline{w}_2) \rangle &= \langle T\underline{v}, (s\underline{w}_1 + \underline{w}_2) \rangle \\ &= s\langle T\underline{v}, \underline{w}_1 \rangle + \langle T\underline{v}, \underline{w}_2 \rangle \\ &= s\langle \underline{v}, T^*\underline{w}_1 \rangle + \langle \underline{v}, T^*\underline{w}_2 \rangle \\ &= \langle \underline{v}, sT^*\underline{w}_1 \rangle + \langle \underline{v}, T^*\underline{w}_2 \rangle \\ &= \langle \underline{v}, sT^*\underline{w}_1 + T^*\underline{w}_2 \rangle. \end{aligned}$$

Since this holds for all $\underline{v} \in V$, we gather from Lemma 6.4.1 that $T^*(s\underline{w}_1 + \underline{w}_2) = sT^*\underline{w}_1 + T^*\underline{w}_2$. Therefore, T^* is linear. The uniqueness of T^* is inherited from Theorem 6.4.2 because, for each $\underline{w} \in W$, the vector $T^*\underline{w}$ is determined uniquely as the vector \underline{v}' such that $\langle T\underline{v}, \underline{w} \rangle = \langle \underline{v}, \underline{v}' \rangle$ for all $\underline{v} \in V$. \square

Theorem 6.4.4. *Let V be a finite-dimensional inner-product space and let*

$$\mathcal{B} = \underline{v}_1, \dots, \underline{v}_n$$

be an orthonormal basis for V . Let T be a linear operator on V and let A be the matrix representation of T in the ordered basis \mathcal{B} . Then $A_{kj} = \langle T\underline{v}_j, \underline{v}_k \rangle$.

Proof. Since \mathcal{B} is an orthonormal basis, we have

$$\underline{v} = \sum_{k=1}^n \langle \underline{v}, \underline{v}_k \rangle \underline{v}_k.$$

The matrix A is defined by

$$T\underline{v}_j = \sum_{k=1}^n A_{kj} \underline{v}_k$$

and since

$$T\underline{v}_j = \sum_{k=1}^n \langle T\underline{v}_j, \underline{v}_k \rangle \underline{v}_k,$$

we conclude that $A_{kj} = \langle T\underline{v}_j, \underline{v}_k \rangle$. \square

Corollary 6.4.5. *Let V be a finite-dimensional inner-product space, and let T be a linear operator on V . In any orthonormal basis for V , the matrix for T^* is the conjugate transpose of the matrix of T .*

Proof. Let $\mathcal{B} = \underline{v}_1, \dots, \underline{v}_n$ be an orthonormal basis for V , let $A = [T]_{\mathcal{B}}$ and $B = [T^*]_{\mathcal{B}}$. According to the previous theorem,

$$\begin{aligned} A_{kj} &= \langle T\underline{v}_j, \underline{v}_k \rangle \\ B_{kj} &= \langle T^*\underline{v}_j, \underline{v}_k \rangle \end{aligned}$$

By the definition of T^* , we then have

$$B_{kj} = \langle T^*\underline{v}_j, \underline{v}_k \rangle = \overline{\langle \underline{v}_k, T^*\underline{v}_j \rangle} = \overline{\langle T\underline{v}_k, \underline{v}_j \rangle} = \overline{A_{jk}}.$$

□

We note here that every linear operator on a finite-dimensional inner-product space V has an adjoint on V . However, in the infinite-dimensional case this is not necessarily true. In any case, there exists at most one such operator T^* .

6.5 Fundamental Subspaces

There are four fundamental subspaces of a linear transformation $T: V \rightarrow W$ when V and W are Hilbert spaces. We have already encountered two such spaces: The range of T and the nullspace of T . Recall that the range of a linear transformation T is the set of all vectors $\underline{w} \in W$ such that $\underline{w} = T\underline{v}$ for some $\underline{v} \in V$. The nullspace of T consists of all vectors $\underline{v} \in V$ such that $T\underline{v} = \underline{0}$.

The other two fundamental subspaces of T are the **range of the adjoint** T^* , denoted R_{T^*} and the **nullspace of the adjoint** T^* , denoted N_{T^*} . The various subspaces of the transformation $T: V \rightarrow W$ can be summarized as follows,

$$\begin{aligned} R_T &\subseteq W \\ N_T &\subseteq V \\ R_{T^*} &\subseteq V \\ N_{T^*} &\subseteq W. \end{aligned}$$

Theorem 6.5.1. *Let V and W be Hilbert spaces and $T: V \rightarrow W$ be a bounded linear transformation from V to W such that R_T and R_{T^*} are both closed. Then,*

1. *the range R_T is the orthogonal complement of N_{T^*} , i.e., $[R_T]^\perp = N_{T^*}$;*
2. *the nullspace N_T is the orthogonal complement of R_{T^*} , i.e., $[R_{T^*}]^\perp = N_T$.*

Complementing these equalities, we get

$$\begin{aligned}\overline{R_T} &= R_T = [N_{T^*}]^\perp \\ \overline{R_{T^*}} &= R_{T^*} = [N_T]^\perp.\end{aligned}$$

Proof. Let $\underline{w} \in R_T$, then there exists $\underline{v} \in V$ such that $T\underline{v} = \underline{w}$. Assume that $\underline{n} \in N_{T^*}$, then

$$\langle \underline{w}, \underline{n} \rangle = \langle T\underline{v}, \underline{n} \rangle = \langle \underline{v}, T^*\underline{n} \rangle = 0.$$

That is, \underline{w} and \underline{n} are orthogonal vectors. It follows that $N_{T^*} \subseteq [R_T]^\perp$. Now, let $\underline{w} \in [R_T]^\perp$. Then, for every $\underline{v} \in V$, we have

$$\langle T\underline{v}, \underline{w} \rangle = 0.$$

This implies that $\langle \underline{v}, T^*\underline{w} \rangle = 0$, by the definition of the adjoint. Since this is true for every $\underline{v} \in V$, we get $T^*\underline{w} = \underline{0}$, so $\underline{w} \in N_{T^*}$. Then $[R_T]^\perp \subseteq N_{T^*}$, which combined with our previous result yields $[R_T]^\perp = N_{T^*}$. Using a similar argument, one can show that $[R_{T^*}]^\perp = N_T$. \square

6.6 Pseudoinverses

Theorem 6.6.1. *Let V and W be Hilbert spaces and T be a bounded linear transformation from V to W where R_T is closed. The equation $T\underline{v} = \underline{w}$ has a solution if and only if $\langle \underline{w}, \underline{u} \rangle = 0$ for every vector $\underline{u} \in N_{T^*}$, i.e.,*

$$\underline{w} \in R_T \Leftrightarrow \underline{w} \perp N_{T^*}.$$

In matrix notation, $A\underline{v} = \underline{w}$ has a solution if and only if $\underline{u}^H \underline{w} = 0$ for every vector \underline{u} such that $A^H \underline{u} = \underline{0}$.

Proof. Assume that $T\underline{v} = \underline{w}$, and let $\underline{u} \in N_{T^*}$. Since T is bounded, the adjoint T^* exists and

$$\langle \underline{w}, \underline{u} \rangle = \langle T\underline{v}, \underline{u} \rangle = \langle \underline{v}, T^*\underline{u} \rangle = \langle \underline{v}, \underline{0} \rangle = 0.$$

To prove the reverse implication, suppose that $\langle \underline{w}, \underline{u} \rangle = 0$ when $\underline{u} \in N_{T^*}$ and $T\underline{v} = \underline{w}$ has no solution. Since $\underline{w} \notin R_T$ and R_T is closed, it follows that

$$\underline{w}_o = \underline{w} - P_{R_T}\underline{w} = \underline{w} - \underline{w}_r \neq \underline{0}.$$

But

$$\langle \underline{w}, \underline{w}_o \rangle = \langle \underline{w}_r + \underline{w}_o, \underline{w}_o \rangle = \langle \underline{w}_o, \underline{w}_o \rangle > 0,$$

which contradicts the assumption that $\langle \underline{w}, \underline{u} \rangle = 0$ when $\underline{u} \in N_{T^*}$. We must conclude that $T\underline{v} = \underline{w}$ has a solution. \square

Fact 6.6.2. *The solution to $T\underline{v} = \underline{w}$ (if it exists) is unique if and only if the only solution to $T\underline{v} = \underline{0}$ is $\underline{v} = \underline{0}$. That is, if $N_T = \{\underline{0}\}$.*

6.6.1 Least Squares

Let $T: V \rightarrow W$ be a bounded linear transformation. If the equation $T\underline{v} = \underline{w}$ has no solution, then we can find a vector \underline{v} that minimizes

$$\|T\underline{v} - \underline{w}\|^2.$$

Theorem 6.6.3. *The vector $\underline{v} \in V$ minimizes $\|T\underline{v} - \underline{w}\|$ if and only if*

$$T^*T\underline{v} = T^*\underline{w}.$$

Proof. Minimizing $\|\underline{w} - T\underline{v}\|$ is equivalent to minimizing $\|\underline{w} - \hat{\underline{w}}\|$, where $\hat{\underline{w}} = T\underline{v} \in R_T$. By the projection theorem, we must have

$$\underline{w} - \hat{\underline{w}} \in [R_T]^\perp.$$

But this is equivalent to

$$\underline{w} - \hat{\underline{w}} \in N_{T^*}.$$

That is, $T^*(\underline{w} - \hat{\underline{w}}) = \underline{0}$, or equivalently $T^*\underline{w} = T^*\hat{\underline{w}}$. Conversely, if $T^*T\underline{v} = T^*\underline{w}$, then

$$T^*(T\underline{v} - \underline{w}) = \underline{0},$$

so that $T\underline{v} - \underline{w} \in N_{T^*}$. Hence, the error is orthogonal to the subspace R_T and has minimal length by the projection theorem. \square

Corollary 6.6.4. *If A is a matrix such that $A^H A$ is invertible, then the least-squares solution to $A\underline{v} = \underline{w}$ is*

$$\underline{v} = (A^H A)^{-1} A^H \underline{w}.$$

The matrix $(A^H A)^{-1} A^H$ is the left inverse of A and is an example of a Moore-Penrose **pseudoinverse**.

Theorem 6.6.5. *Suppose the vector $\hat{\underline{v}} \in V$ minimizes $\|\underline{v}\|$ over all $\underline{v} \in V$ satisfying $T\underline{v} = \underline{w}$. Then, $\hat{\underline{v}} \in [N_T]^\perp$ and, if R_{T^*} is closed, $\hat{\underline{v}} = T^* \underline{u}$ for some $\underline{u} \in W$.*

Proof. Suppose $\hat{\underline{v}} \notin [N_T]^\perp$, then the orthogonal decomposition $V = [N_T]^\perp + N_T$ shows that the projection of $\hat{\underline{v}}$ onto $[N_T]^\perp$ has smaller norm but still satisfies $T\hat{\underline{v}} = \underline{w}$. This gives a contradiction and shows that $\hat{\underline{v}} \in [N_T]^\perp$. If R_{T^*} is closed, then $R_{T^*} = [N_T]^\perp$ and $\hat{\underline{v}} = T^* \underline{u}$ for some $\underline{u} \in W$. \square

Corollary 6.6.6. *If A is a matrix such that AA^H is invertible, then the minimum-norm solution to $A\underline{v} = \underline{w}$ is*

$$\underline{v} = A^H (AA^H)^{-1} \underline{w}.$$

Proof. The theorem shows that $\underline{v} = A^H \underline{u}$ and $A\underline{v} = AA^H \underline{u} = \underline{w}$. Since AA^H is invertible, this gives $\underline{u} = (AA^H)^{-1} \underline{w}$ and computing \underline{v} gives the desired result. \square

The matrix $A^H (AA^H)^{-1}$ is the right inverse of A and is another example of a Moore-Penrose **pseudoinverse**.

Definition 6.6.7. *Let $T: V \rightarrow W$ be a bounded linear transformation, where V and W are Hilbert spaces, and R_T is closed. For each $\underline{w} \in W$, there is a unique vector $\hat{\underline{v}}$ of minimum norm in the set of vectors that minimize $\|T\underline{v} - \underline{w}\|$. The **pseudoinverse** T^\dagger is the transformation mapping each $\underline{w} \in W$ to its unique $\hat{\underline{v}}$.*