Chapter 3

Linear Algebra

3.1 Fields

Consider a set F of objects and two operations on the elements of F, addition and multiplication. For every pair of elements $s, t \in F$ then $(s + t) \in F$. For every pair of elements $s, t \in F$ then $st \in F$. Suppose that these two operations satisfy

1. addition is commutative: $s + t = t + s \ \forall s, t \in F$

2. addition is associative: $r + (s + t) = (r + s) + t \ \forall r, s, t \in F$

3. to each $s \in F$ there exists a unique element $(-s) \in F$ such that $s + (-s) = 0$

4. multiplication is commutative: $st = ts \ \forall s, t \in F$

5. multiplication is associative: $r(st) = (rs)t \ \forall r, s, t \in F$

6. there is a unique non-zero element $1 \in F$ such that $s1 = s \ \forall s \in F$

7. to each $s \in F - 0$ there exists a unique element $s^{-1} \in F$ such that $ss^{-1} = 1$

8. multiplication distributes over addition: $r(s + t) = rs + rt \ \forall r, s, t \in F$.

Then, the set F together with these two operations is a **field**.

Example 3.1.1. *The real numbers with the usual operations of addition and multiplication form a field. The complex numbers with these two operations also form a field.*

Example 3.1.2. *The set of integers with addition and multiplication is not a field.*

Problem 3.1.3. *Is the set of rational numbers a subfield of the real numbers?*

Example 3.1.4. *Is the set of all real numbers of the form $s + t\sqrt{2}$, where s and t are rational, a subfield of the complex numbers?*

The set $F = \left\{ s + t\sqrt{2} : s, t \in \mathbb{Q} \right\}$ together with the standard addition and multiplication is a field. Let $s, t, u, v \in \mathbb{Q}$,

$$s + t\sqrt{2} + u + v\sqrt{2} = (s + u) + (t + v)\sqrt{2} \in F$$

$$\left(s + t\sqrt{2} \right)\left(u + v\sqrt{2} \right) = (su + 2tv) + (sv + tu)\sqrt{2} \in F$$

$$\left(s + t\sqrt{2} \right)^{-1} = \frac{s - t\sqrt{2}}{s^2 + 2t^2} = \frac{s}{s^2 + 2t^2} - \frac{t}{s^2 + 2t^2}\sqrt{2} \in F$$

Again, the remaining properties are straightforward to prove. The field $s + t\sqrt{2}$, where s and t are rational, is a subfield of the complex numbers.

3.2 Matrices

Let F be a field and consider the problem of finding n scalars $x_1, \ldots, x_n$ which satisfy the conditions

$$
\begin{array}{ccccccccc}
a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & y_1 \\
a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & y_2 \\
\vdots & & \vdots & & & & \vdots & & \vdots \\
a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & y_m
\end{array}
\tag{3.1}
$$

where $\{y_i : 1 \leq i \leq n\} \subset F$ and $\{a_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\} \subset F$. These conditions form a *system of m linear equations in n unknowns*. A shorthand notation for (3.1) is the matrix equation

$$A\underline{x} = \underline{y},$$

where $\underline{x} = (x_1, \ldots, x_n)^T$, $\underline{y} = (y_1, \ldots, y_m)^T$, and A is the matrix given by

$$
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}.
$$

We also use $[A]_{i,j}$ to denote the entry of A in the i-th row and j-th column (i.e., a_{ij}) and $[\underline{x}_i]$ to denote the i-th entry in $\underline{x}$ (i.e., x_i).

Definition 3.2.1. *Let A be an $m \times n$ matrix over F and let B be an $n \times p$ matrix over F. The **matrix product** AB is the $m \times p$ matrix C whose i, j entry is*

$$c_{ij} = \sum_{r=1}^{n} a_{ir} b_{rj}. \tag{3.2}$$

Remark 3.2.2. *Consider (3.2) when j is fixed and i is eliminated by grouping the elements of C and A into column vectors $\underline{c}_1, \ldots, \underline{c}_p$ and $\underline{a}_1, \ldots, \underline{a}_n$. For this case, (3.2) shows that the j-th column of C is a linear combination of the columns of A,*

$$\underline{c}_j = \sum_{r=1}^{n} \underline{a}_r b_{rj},$$

Similarly, one can fix i and eliminate the index j by grouping the elements of C and B into row vectors $\underline{c}_1, \ldots, \underline{c}_m$ and $\underline{b}_1, \ldots, \underline{b}_n$. Then, (3.2) shows that the i-th row of C is a linear combination of the rows of B,

$$\underline{c}_i = \sum_{r=1}^{n} a_{ir} \underline{b}_r.$$

Definition 3.2.3. *Consider an $m \times n$ matrix A with elements $a_{ij} \in F$. The **transpose** of A is the $n \times m$ matrix $B = A^T$ with elements defined by $b_{ij} = a_{ji}$.*

Definition 3.2.4. *Consider a complex $m \times n$ matrix A with elements $a_{ij} \in \mathbb{C}$. Its **Hermitian transpose** $B = A^H$ is the $n \times m$ matrix with elements defined $b_{ij} = \overline{a_{ji}}$, where $\bar{a}$ denotes the complex conjugate of a.*

Problem 3.2.5. *For matrices $A \in \mathbb{C}^{m \times p}$ and $B \in \mathbb{C}^{p \times n}$, show $(AB)^H = B^H A^H$.*

Definition 3.2.6. *An $m \times n$ matrix A over F is in **row echelon form** if*

1. *all rows containing only zeros, if they exist, are the bottom of the matrix, and*

2. *For non-zero rows, the leading coefficient (i.e., the first non-zero element from the left) is strictly to the right of the leading coefficient of the row above it.*

*These two conditions imply that entries below the leading coefficient in a column are zero. A matrix is **column echelon form** if its transpose is in row echelon form.*

Definition 3.2.7. *An $m \times n$ matrix A over F is in **reduced row echelon form** if it is in row echelon form and*

1. *every leading coefficient is 1, and*

2. *every leading coefficient is the only non-zero element in its column.*

Definition 3.2.8. *Let A be an $n \times n$ matrix over F. An $n \times n$ matrix B is called the **inverse** of A if*

$$AB = BA = I.$$

*In this case, A is called **invertible** and its inverse is denoted by A^{-1}.*

Problem 3.2.9. *For a matrix $A \in \mathbb{C}^{n \times n}$, show that $(A^H)^{-1} = (A^{-1})^H$ if A^{-1} exists.*

Definition 3.2.10. *An **elementary row operation** on an $m \times n$ matrix consists of*

1. *multiplying a row by a non-zero scalar,*

2. *swapping two rows, or*

3. *adding a non-zero scalar multiple of one row to another row.*

*An **elementary column operation** is the same but applied to the columns.*

Lemma 3.2.11. *For any $m \times n$ matrix A over F, there is an invertible $m \times m$ matrix P over F such that $R = PA$ is in reduced row echelon form.*

Sketch of Proof. This follows from the fact that elementary row operations (i.e., Gaussian elimination) can be used to reduce any matrix to reduced row echelon form. To construct the P matrix, one applies Gaussian elimination to the augmented matrix $A' = [A \ I]$. This results in an augmented matrix $R' = [R \ P]$ in reduced row echelon form. It follows that R is also in reduced row echelon form. Since elementary row operations can be implemented by (invertible) matrix multiplies on the left side, one also finds that $R' = PA'$, $R = PA$, and P is invertible.   □

Lemma 3.2.12. *Let A be an $m \times n$ matrix over F with $m < n$. Then, there exists a length-n column vector $\underline{x} \neq \underline{0}$ (over F) such that $A\underline{x} = \underline{0}$.*

Proof. First, we use row reduction to compute the reduced row echelon form $R = PA$ of A, where P is invertible. Then, we observe that the columns of R containing leading elements can be combined in a linear combination to cancel any other column of R. This allows us to construct a vector $\underline{x}$ satisfying $R\underline{x} = \underline{0}$ and thus $A\underline{x} = P^{-1}R\underline{x} = \underline{0}$.   □

3.3 Vector Spaces

Definition 3.3.1. *A **vector space** consists of the following,*

1. *a field F of scalars*

2. *a set V of objects, called vectors*

3. *an operation called vector addition, which associates with each pair of vectors $\underline{v}, \underline{w} \in V$ a vector $\underline{v} + \underline{w} \in V$ such that*

    (a) addition is commutative: $\underline{v} + \underline{w} = \underline{w} + \underline{v}$

    (b) addition is associative: $\underline{u} + (\underline{v} + \underline{w}) = (\underline{u} + \underline{v}) + \underline{w}$

    (c) there is a unique vector $\underline{0} \in V$ such that $\underline{v} + \underline{0} = \underline{v}$, $\forall \underline{v} \in V$

    (d) to each $\underline{v} \in V$ there is a unique vector $-\underline{v} \in V$ such that $\underline{v} + (-\underline{v}) = \underline{0}$

4. *an operation called scalar multiplication, which associates with each $s \in F$ and $\underline{v} \in V$ a vector $s\underline{v} \in V$ such that*

    (a) $1\underline{v} = \underline{v}$, $\forall \underline{v} \in V$

    (b) $(s_1 s_2)\underline{v} = s_1(s_2\underline{v})$

    (c) $s(\underline{v} + \underline{w}) = s\underline{v} + s\underline{w}$

    (d) $(s_1 + s_2)\underline{v} = s_1\underline{v} + s_2\underline{v}$.

Example 3.3.2. *Let F be a field, and let V be the set of all n-tuples $\underline{v} = (v_1, \ldots, v_n)$ of scalar $v_i \in F$. If $\underline{w} = (w_1, \ldots, w_n)$ with $w_i \in F$, the sum of $\underline{v}$ and $\underline{w}$ is defined by*

$$\underline{v} + \underline{w} = (v_1 + w_1, \ldots, v_n + w_n).$$

The product of a scalar $s \in F$ and vector $\underline{v}$ is defined by

$$s\underline{v} = (sv_1, \ldots, sv_n).$$

The set of n-tuples, denoted by F^n, with the vector addition and scalar product defined above forms a vector space. This is the standard vector space for F^n.

Example 3.3.3. *Let X be a non-empty set and let Y be a vector space over F. Consider the set V of all functions from X into Y. The sum of two vectors $f, g \in V$ is the function from X into Y defined by*

$$(f + g)(x) = f(x) + g(x) \quad \forall x \in X,$$

where the RHS uses vector addition from Y. The product of scalar $s \in F$ and the function $f \in V$ is the function sf defined by

$$(sf)(x) = sf(x) \, \forall x \in X,$$

where the RHS uses scalar multiplication from Y. This is the standard vector space of functions from a set X to a vector space Y.

Definition 3.3.4. *A vector $\underline{w} \in V$ is said to be a **linear combination** of the vectors $\underline{v}_1, \ldots, \underline{v}_n \in V$ provided that there exist scalars $s_1, \ldots, s_n \in F$ such that*

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

3.3.1   Subspaces

Definition 3.3.5. *Let V be a vector space over F. A **subspace** of V is a subset $W \subset V$ which is itself a vector space over F.*

Fact 3.3.6. *A non-empty subset $W \subset V$ is a subspace of V if and only if for every pair $\underline{w}_1, \underline{w}_2 \in W$ and every scalar $s \in F$ the vector $s\underline{w}_1 + \underline{w}_2$ is again in W.*

If V is a vector space then the intersection of any collection of subspaces of V is a subspace of V.

Example 3.3.7. *Let A be an $m \times n$ matrix over F. The set of all $n \times 1$ column vectors V such that*

$$\underline{v} \in V \implies A\underline{v} = \underline{0}$$

is a subspace of $F^{n \times 1}$.

Definition 3.3.8. *Let U be a set (or list) of vectors in V. The **span** of U, denoted $\mathrm{span}(U)$, is defined to be the set of all finite linear combinations of vectors in U.*

The subspace spanned by U can be defined equivalently as the intersection of all subspaces of V that contain U. To see this, we note that the intersection of all subspaces containing U is a subspace containing U because the intersection of subspaces is also a subspace. The intersection cannot be larger than U, however, because U is a subspace containing U.

Definition 3.3.9. *Let V be a vector space and U, W be subspaces. If U, W are disjoint (i.e., $U \cap W = \{\underline{0}\}$), their **direct sum** $U \oplus W$ is defined by*

$$U \oplus W \triangleq \{\underline{u} + \underline{w} | \underline{u} \in U, \underline{w} \in W\}.$$

An important property of a direct sum is that any vector $\underline{v} \in U \oplus W$ has a unique decomposition $\underline{v} = \underline{u} + \underline{w}$ where $\underline{u} \in U$ and $\underline{w} \in W$.

3.3.2 Bases and Dimensions

The dimension of a vector space is defined using the concept of a basis.

Definition 3.3.10. *Let V be a vector space over F. A list of vectors $\underline{u}_1, \ldots, \underline{u}_n \in V$ is called **linearly dependent** if there are scalars $s_1, \ldots, s_n \in F$, not all of which are 0, such that*

$$\sum_{i=1}^{n} s_i \underline{u}_i = 0.$$

*A list that is not linearly dependent is called **linearly independent**. Similarly, a subset $U \subset V$ is called linearly dependent if there is a finite list $\underline{u}_1, \ldots, \underline{u}_n \in U$ of distinct vectors that is linearly dependent. Otherwise, it is called linearly independent.*

A few important consequences follow immediately from this definition. Any subset of a linearly independent set is also linearly independent. Any set which contains the $\underline{0}$ vector is linearly dependent. A set $U \subset V$ is linearly independent if and only if each finite subset of U is linearly independent.

Definition 3.3.11. *Let V be a vector space over F. Let $\mathcal{B} = \{\underline{v}_\alpha | \alpha \in A\}$ be a subset of linearly independent vectors from V such that every $\underline{v} \in V$ has a unique decomposition as a finite linear combination of vectors from $\mathcal{B}$. Then, the set $\mathcal{B}$ is a **Hamel basis** for V. The space V is **finite-dimensional** if it has a finite basis.*

Theorem 3.3.12. *Every vector space has a Hamel basis.*

Proof. Let X be the set of linearly independent subsets of V. Furthermore, for $x, y \in X$ consider the strict partial order defined by proper inclusion. By the maximum principle, if x is an element of X, then there exists a maximal simply ordered subset Z of X containing x. This element is a Hamel basis for V. $\qquad\square$

Example 3.3.13. *Let F be a field and let $U \subset F^n$ be the subset consisting of the vectors $\underline{e}_1, \ldots, \underline{e}_n$ defined by*

$$
\begin{aligned}
\underline{e}_1 &= (1, 0, \ldots, 0) \\
\underline{e}_2 &= (0, 1, \ldots, 0) \\
\vdots &= \qquad \vdots \\
\underline{e}_n &= (0, 0, \ldots, 1).
\end{aligned}
$$

For any $\underline{v} = (v_1, \ldots, v_n) \in F^n$, we have

$$
\underline{v} = \sum_{i=1}^{n} v_i \underline{e}_i. \tag{3.3}
$$

*Thus, the collection $U = \{\underline{e}_1, \ldots, \underline{e}_n\}$ spans F^n. Since $\underline{v} = \underline{0}$ in (3.3) if and only if $v_1 = \cdots = v_n = 0$, U is linearly independent. Accordingly, the set U is a basis for $F^{n \times 1}$. This basis is termed the **standard basis** of F^n.*

Example 3.3.14. *Let A be an invertible matrix over F. Then, the columns of A, denoted by $\underline{a}_1, \ldots, \underline{a}_n$, form a basis for the space of column vectors $F^{n \times 1}$. This can be seen as follows. If $\underline{v} = (v_1, \ldots, v_n)^T$ is a column vector, then*

$$
A\underline{v} = \sum_{i=1}^{n} v_i \underline{a}_i.
$$

Since A is invertible,

$$
A\underline{v} = \underline{0} \implies I\underline{v} = A^{-1}\underline{0} \implies \underline{v} = \underline{0}.
$$

Thus, $\{\underline{a}_1, \ldots, \underline{a}_n\}$ is a linearly independent set. Next, for any column vector $\underline{w} \in F^n$, let $\underline{v} = A^{-1}\underline{w}$. It follows that $\underline{w} = A\underline{v}$ and, as a consequence, $\{\underline{a}_1, \ldots, \underline{a}_n\}$ is a basis for F^n.

Theorem 3.3.15. *Let V be a finite-dimensional vector space that is spanned by a finite set of vectors $W = \{\underline{w}_1, \ldots, \underline{w}_n\}$. If $U = \{\underline{u}_1, \ldots, \underline{u}_m\} \subset V$ is a linearly independent set of vectors, then $m \leq n$.*

Proof. Suppose that $U = \{\underline{u}_1, \ldots, \underline{u}_m\} \subset V$ is linearly independent and $m > n$. Since W spans V, there exists scalars a_{ij} such that

$$\underline{u}_j = \sum_{i=1}^{n} a_{ij}\underline{w}_i.$$

For any m scalars $s_1, \ldots, s_m$ we have

$$\sum_{j=1}^{m} s_j\underline{u}_j = \sum_{j=1}^{m} s_j \sum_{i=1}^{n} a_{ij}\underline{w}_i = \sum_{j=1}^{m}\sum_{i=1}^{n}(a_{ij}s_j)\underline{w}_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} a_{ij}s_j\right)\underline{w}_i.$$

Collecting the a_{ij} coefficients into an n by m matrix A shows that

$$\begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix} = A \begin{bmatrix} s_1 \\ \vdots \\ s_m \end{bmatrix}.$$

Since $A \in F^{n \times m}$ with $n < m$, Lemma 3.2.12 implies there are scalars $s_1, \ldots, s_n$, not all 0, such that $t_1 = t_2 = \cdots = t_m = 0$. For these scalars, $\sum_{j=1}^{m} s_j\underline{u}_j = \underline{0}$. Thus, the set U is linearly dependent. and the contradiction implies $m \geq n$. $\quad\square$

Now, suppose that V is a finite-dimensional vector space with bases $U = \{\underline{u}_1, \ldots, \underline{u}_n\}$ and $W = \{\underline{w}_1, \ldots, \underline{w}_m\}$ where $m \neq n$. Then, without loss of generality, we can assume $m > n$ and apply Theorem 3.3.15 to see that W must be linearly dependent. Since a basis must be linearly independent, this gives a contradiction and implies that $m = n$. Hence, if V is a finite-dimensional vector space, then any two bases of V have the same number of elements. Therefore, the dimension of a finite-dimensional vector space is uniquely defined. Thus, our intuition about dimension from $\mathbb{R}^n$ does not break down for other vector spaces and fields.

Definition 3.3.16. *The **dimension** of a finite-dimensional vector space is defined as the number of elements in any basis for V. We denote the dimension of a finite-dimensional vector space V by $\dim(V)$.*

The zero subspace of a vector space V is the subspace spanned by the vector $\underline{0}$. Since the set $\{\underline{0}\}$ is linearly dependent and not a basis, we assign a dimension 0 to the zero subspace. Alternatively, it can be argue that the empty set $\emptyset$ spans $\{\underline{0}\}$ because the intersection of all the subspaces containing the empty set is $\{0\}$. Though this is only a minor point.

Theorem 3.3.17. *Let A be an $n \times n$ matrix over F and suppose that the columns of A, denoted by $\underline{a}_1, \ldots, \underline{a}_n$, form a linearly independent set of vectors in $F^{n\times 1}$. Then A is invertible.*

Proof. Suppose that W is the subspace of $F^{n\times 1}$ spanned by $\underline{a}_1, \ldots, \underline{a}_n$. Since $\underline{a}_1, \ldots, \underline{a}_n$ are linearly independent, $\dim(W) = n = \dim(F^{n\times 1})$. It follows that $W = V$ and, as a consequence, there exist scalars $b_{ij} \in F$ such that

$$\underline{e}_j = \sum_{i=1}^{n} b_{ij}\underline{a}_i, \quad 1 \leq j \leq n$$

where $\{\underline{e}_1, \ldots, \underline{e}_n\}$ is the standard basis for $F^{n\times 1}$. Then, for the matrix B with entries b_{ij}, we have

$$AB = I.$$

Note also that if the rows of A form a linearly independent set of vectors in $F^{1\times n}$ then A is invertible. $\qquad\square$

3.3.3 Coordinate System

Let $\{\underline{v}_1, \ldots, \underline{v}_n\}$ be a basis for the n-dimensional vector space V and recall that every vector $\underline{w} \in V$ can be expressed uniquely as

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

While standard vector and matrix notation requires that the basis elements be ordered, a set is an unordered collection of objects. Ordering this set (e.g., $\underline{v}_1, \ldots, \underline{v}_n$) allows the first element in the coordinate vector to be associated with the first vector in our basis and so on.

Definition 3.3.18. *If V is a finite-dimensional vector space, an **ordered basis** for V is a finite list of vectors that is linearly independent and spans V.*

In particular, if the sequence $\underline{v}_1, \ldots, \underline{v}_n$ is an ordered basis for V, then the set $\{\underline{v}_1, \ldots, \underline{v}_n\}$ is a basis for V. The ordered basis $\mathcal{B}$, denoted by $(\underline{v}_1, \ldots, \underline{v}_n)$, defines the set and a specific ordering of the vectors. Based on this ordered basis, a vector $\underline{v} \in V$ can be unambiguously represented as an n-tuple $(s_1, \ldots, s_n) \in F^n$ such that

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

Definition 3.3.19. *For a finite-dimensional vector space V with ordered basis $\mathcal{B} = (\underline{v}_1, \ldots, \underline{v}_n)$, the **coordinate vector** of $\underline{v} \in V$ is denoted by $[\underline{v}]_\mathcal{B}$ and equals the unique vector $\underline{s} = F^n$ such that*

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

The dependence of the coordinate vector $[\underline{v}]_\mathcal{B}$ on the basis is explicitly specified using the subscript. This can be particularly useful when multiple coordinates systems are involved.

Example 3.3.20. *The canonical example of an ordered basis is the standard basis for F^n introduced in Section 3.3.2. Note that the standard basis contains a natural ordering: $\underline{e}_1, \ldots, \underline{e}_n$. Vectors in F^n can therefore be unambiguously expressed as n-tuples.*

Problem 3.3.21. *Suppose that $\mathcal{A} = \underline{v}_1, \ldots, \underline{v}_n$ is an ordered basis for V. Let P be an $n \times n$ invertible matrix. Show that there exists an ordered basis $\mathcal{B} = \underline{w}_1, \ldots, \underline{w}_n$ for V such that*

$$[\underline{u}]_\mathcal{A} = P [\underline{u}]_\mathcal{B}$$
$$[\underline{u}]_\mathcal{B} = P^{-1} [\underline{u}]_\mathcal{A}$$

for every $\underline{u} \in V$.

S 3.3.21. Consider the ordered basis $\mathcal{A} = \underline{v}_1, \ldots, \underline{v}_n$ and let $Q = P^{-1}$. For all $\underline{u} \in V$, we have $\underline{u} = \sum_{i=1}^{n} s_i \underline{v}_i$, where

$$[\underline{u}]_\mathcal{A} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix}.$$

If we define

$$\underline{w}_i = \sum_{k=1}^{n} p_{ki}\underline{v}_k \qquad \text{and} \qquad t_i = \sum_{j=1}^{n} q_{ij}s_j,$$

then we find that

$$\sum_{i=1}^{n} t_i\underline{w}_i = \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}s_j\underline{w}_i = \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}s_j \sum_{k=1}^{n} p_{ki}\underline{v}_k$$

$$= \sum_{j=1}^{n} s_j \sum_{k=1}^{n} \underline{v}_k \sum_{i=1}^{n} p_{ki}q_{ij} = \sum_{j=1}^{n} s_j \sum_{k=1}^{n} \underline{v}_k\delta_{jk}$$

$$= \sum_{j=1}^{n} s_j\underline{v}_j = \underline{u}.$$

This shows that $\mathcal{B} = \underline{w}_1, \ldots, \underline{w}_n$ is an ordered basis for V and

$$[\underline{u}]_{\mathcal{B}} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}.$$

The definition of t_i also shows that $[\underline{u}]_{\mathcal{B}} = P^{-1}[\underline{u}]_{\mathcal{A}}$ and therefore $[\underline{u}]_{\mathcal{A}} = P[\underline{u}]_{\mathcal{B}}$.

3.4   Linear Transformations

3.4.1   Definitions

Definition 3.4.1. *Let V and W be vector spaces over a field F. A **linear transform** from V to W is a function T from V into W such that*

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

for all $\underline{v}_1$ and $\underline{v}_2$ in V and all scalars s in F.

Definition 3.4.2. *Let $L(V,W)$ denote the **set of all linear transforms** from V into W, where V and W are vector spaces over a field F.*

Example 3.4.3. *Let A be a fixed $m \times n$ matrix over F. The function T defined by $T(\underline{v}) = A\underline{v}$ is a linear transformation from $F^{n \times 1}$ into $F^{m \times 1}$.*

Example 3.4.4. *Let $P \in F^{m \times m}$ and $Q \in F^{n \times n}$ be fixed matrices. Define the function T from $F^{m \times n}$ into itself by $T(A) = PAQ$. Then T is a linear transformation from $F^{m \times n}$ into $F^{m \times n}$. In particular,*

$$T(sA + B) = P(sA + B)Q$$
$$= sPAQ + PBQ$$
$$= sT(A) + T(B).$$

Example 3.4.5. *Let V be the space of continuous functions from $\mathbb{R}$ to $\mathbb{R}$, and define T by*

$$(Tf)(x) = \int_0^x f(t)dt.$$

Then T is a linear transformation from V into V. The function Tf is continuous and differentiable.

It is important to note that if T is a linear transformation from V to W, then $T(\underline{0}) = \underline{0}$. This is essential since

$$T(\underline{0}) = T(\underline{0} + \underline{0}) = T(\underline{0}) + T(\underline{0}).$$

Definition 3.4.6. *A linear transformation $T\colon V \to W$ is **singular** if there is a non-zero vector $\underline{v} \in V$ such that $T\underline{v} = \underline{0}$. Otherwise, it is called **non-singular**.*

3.4.2 Properties

Theorem 3.4.7. *Let V, W be vector spaces over F and $\mathcal{B} = \{\underline{v}_\alpha | \alpha \in A\}$ be a Hamel basis for V. For each mapping $G\colon \mathcal{B} \to W$, there is a unique linear transformation $T\colon V \to W$ such that $T\underline{v}_\alpha = G(\underline{v}_\alpha)$ for all $\alpha \in A$.*

Proof. Since $\mathcal{B}$ is a Hamel basis for V, every vector $\underline{w} \in V$ has a unique expansion

$$\underline{w} = \sum_{\alpha \in A} s_\alpha(\underline{w}) \underline{v}_\alpha,$$

where $s_\alpha(\underline{w})$ is the unique α coefficient for $\underline{w}$ and $s_\alpha(\underline{w}) \neq 0$ only for a finite subset of A. Using the unique expansion and vector space properties, one can show that

$$s_\alpha(t\underline{w}_1 + \underline{w}_2) = ts_\alpha(\underline{w}_1) + s_\alpha(\underline{w}_2).$$

Next, we define the mapping $T\colon V \to W$ in terms of $s_\alpha(\cdot)$ and $G(\cdot)$ with

$$T\underline{w} = \sum_{\alpha \in A} s_\alpha(\underline{w})\, G(\underline{v}_\alpha).$$

Using the linearity of $s_\alpha(\cdot)$, it is easy to verify that T is a linear transform.

To show that T is unique, we let $U\colon V \to W$ be any other linear mapping satisfying $U\underline{v}_\alpha = G(\underline{v}_\alpha)$ for all $\alpha \in A$. In this case, the linearity of U guarantees that

$$U\underline{w} = U\left(\sum_{\alpha \in A} s_\alpha(\underline{w})\, \underline{v}_\alpha\right) = \sum_{\alpha \in A} s_\alpha(\underline{w})\, U(\underline{v}_\alpha) = \sum_{\alpha \in A} s_\alpha(\underline{w})\, G(\underline{v}_\alpha).$$

From this, we see that $U\underline{w} = T\underline{w}$ for all $\underline{w} \in V$ and therefore that $U = T$. $\square$

This theorem illuminates a very important structural element of linear transformations: they are uniquely defined by where they map an set of basis vectors for their domain.

Definition 3.4.8. *If T is a linear transformation from V into W, the **range** of T is the set of all vectors $\underline{w} \in W$ such that $\underline{w} = T\underline{v}$ for some $\underline{v} \in V$. We denote the range of T by*

$$\mathcal{R}(T) \triangleq \{\underline{w} \in W \,|\, \exists \underline{v} \in V \text{ s.t. } T\underline{v} = \underline{w}\} = \{T\underline{v} \,|\, \underline{v} \in V\}.$$

The set $\mathcal{R}(T)$ is a subspace of W. Let $\underline{w}_1, \underline{w}_2 \in \mathcal{R}(T)$ and let s be a scalar. By definition, there exist vectors $\underline{v}_1$ and $\underline{v}_2$ in V such that $T\underline{v}_1 = \underline{w}_1$ and $T\underline{v}_2 = \underline{w}_2$. Since T is a linear transformation, we have

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

which shows that $s\underline{w}_1 + \underline{w}_2$ is also in $\mathcal{R}(T)$.

Definition 3.4.9. *If T is a linear transformation from V into W, the **nullspace** of T is the set of all vectors $\underline{v} \in V$ such that $T\underline{v} = \underline{0}$. We denote the nullspace of T by*

$$\mathcal{N}(T) \triangleq \{\underline{v} \in V \,|\, T\underline{v} = \underline{0}\}.$$

It can easily be verified that $\mathcal{N}(T)$ is a subspace of V.

$$T\left(\underline{0}\right) = \underline{0} \implies \underline{0} \in \mathcal{N}(T).$$

Furthermore, if $T\underline{v}_1 = T\underline{v}_2 = 0$ then

$$T\left(s\underline{v}_1 + \underline{v}_2\right) = sT\left(\underline{v}_1\right) + \left(\underline{v}_2\right) = s\underline{0} + \underline{0} = \underline{0},$$

so that $s\underline{v}_1 + \underline{v}_2 \in \mathcal{N}(T)$.

Definition 3.4.10. *Let V and W be vector spaces over a field F, and let T be a linear transformation from V into W. If V is finite-dimensional, the **rank** of T is the dimension of the range of T and the **nullity** of T is the dimension of the nullspace of T.*

Theorem 3.4.11. *Let V and W be vector spaces over the field F and let T be a linear transformation from V into W. If V is finite-dimensional, then*

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = \mathrm{dim}(V)$$

Proof. Let $\underline{v}_1, \ldots, \underline{v}_k$ be a basis for $\mathcal{N}(T)$, the nullspace of T. There are vectors $\underline{v}_{k+1}, \ldots, \underline{v}_n \in V$ such that $\underline{v}_1, \ldots, \underline{v}_n$ is a basis for V. We want to show that $T\underline{v}_{k+1}, \ldots, T\underline{v}_n$ is a basis for the range of T. The vectors $T\underline{v}_1, \ldots, T\underline{v}_n$ certainly span $\mathcal{R}(T)$ and, since $T\underline{v}_j = \underline{0}$ for $j = 1, \ldots, k$, it follows that $T\underline{v}_{k+1}, \ldots, \underline{v}_n$ span $\mathcal{R}(T)$. Suppose that there exist scalars $s_{k+1}, \ldots, s_n$ such that

$$\sum_{j=k+1}^{n} s_j T\underline{v}_j = \underline{0}.$$

This implies that

$$T\left(\sum_{j=k+1}^{n} s_j \underline{v}_j\right) = \underline{0}.$$

and accordingly the vector $\underline{v} = \sum_{j=k+1}^{n} s_j \underline{v}_j$ is in the nullspace of T. Since $\underline{v}_1, \ldots, \underline{v}_k$ form a basis for $\mathcal{N}(T)$, there must be a linear combination such that

$$\underline{v} = \sum_{j=1}^{k} t_j \underline{v}_j.$$

But then,

$$\sum_{j=1}^{k} t_j \underline{v}_j - \sum_{j=k+1}^{n} s_j \underline{v}_j = \underline{0}.$$

Since the vectors $\underline{v}_1, \ldots, \underline{v}_n$ are linearly independent, this implies that

$$t_1 = \cdots = t_k = s_{k+1} = \ldots s_n = 0.$$

That is, the set $T\underline{v}_{k+1}, \ldots, T\underline{v}_n$ is linearly independent in W and therefore forms a basis for $\mathcal{R}(T)$. In turn, this implies that $n = \operatorname{rank}(T) + \operatorname{nullity}(T)$. $\square$

Theorem 3.4.12. *If A is an $m \times n$ matrix with entries in the field F, then*

$$\operatorname{row\,rank}(A) \triangleq \dim(\mathcal{R}(A^T)) = \dim(\mathcal{R}(A)) \triangleq \operatorname{rank}(A).$$

Proof. Let $R = PA$ be the reduced row echelon form of A, where P is invertible. Let r be the number of non-zero rows in R and observe that $\operatorname{row\,rank}(A) = r$ because the rows of R form a basis for the row space of A. Next, we write $A = P^{-1}R$ and observe that each column of A is a linear combination of the first r columns in P^{-1}. Thus, the column space of A is spanned by r vectors and $\operatorname{rank}(A) \leq \operatorname{row\,rank}(A)$. Applying this to both A and A^T gives

$$\operatorname{rank}(A) \leq \operatorname{row\,rank}(A) = \operatorname{rank}(A^T) \leq \operatorname{row\,rank}(A^T) = \operatorname{rank}(A),$$

which completes the proof. $\square$

When $F = \mathbb{C}$, the space $\mathcal{R}(A^H)$ has many nice properties and can also be called the row space of A. Regardless, it holds that $\operatorname{rank}(A) = \operatorname{rank}(A^T) = \operatorname{rank}(A^H)$.

3.5   Norms

Let V be a vector space over the real numbers or the complex numbers.

Definition 3.5.1. *A **norm** on vector space V is a real-valued function $\|\cdot\| : V \to \mathbb{R}$ that satisfies the following properties.*

1. *$\|\underline{v}\| \geq 0 \quad \forall \underline{v} \in V$; equality holds if and only if $\underline{v} = \underline{0}$*

2. *$\|s\underline{v}\| = |s| \, \|\underline{v}\| \quad \forall \underline{v} \in V, s \in F$*

3. *$\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\| \quad \forall \underline{v}, \underline{w} \in V.$*

The concept of a norm is closely related to that of a metric. For instance, a metric can be defined from any norm. Let $\|\underline{v}\|$ be a norm on vector space V, then

$$d(\underline{v}, \underline{w}) = \|\underline{v} - \underline{w}\|$$

is the metric induced by the norm.

Normed vector spaces are very useful because they have all the properties of a vector space and all the benefits of a topology generated by the norm. Therefore, one can discuss limits and convergence in a meaningful way.

Example 3.5.2. *Consider vectors in $\mathbb{R}^n$ with the euclidean metric*

$$d\left(\underline{v}, \underline{w}\right) = \sqrt{(v_1 - w_1)^2 + \cdots + (v_n - w_n)^2}.$$

Recall that the standard bounded metric introduced in Problem 2.1.5 is given by

$$\bar{d}\left(\underline{v}, \underline{w}\right) = \min\left\{d\left(\underline{v}, \underline{w}\right), 1\right\}.$$

Define the function $f \colon \mathbb{R}^n \to \mathbb{R}$ by $f\left(\underline{v}\right) = \bar{d}\left(\underline{v}, \underline{0}\right)$. Is the function f a norm?

By the properties of a metric, we have

1. $\bar{d}\left(\underline{v}, \underline{0}\right) \geq 0 \quad \forall \underline{v} \in V$; equality holds if and only if $\underline{v} = \underline{0}$

2. $\bar{d}\left(\underline{v}, \underline{0}\right) + \bar{d}\left(\underline{w}, \underline{0}\right) = \bar{d}\left(\underline{v}, \underline{0}\right) + \bar{d}\left(\underline{0}, \underline{w}\right) \geq \bar{d}\left(\underline{v}, \underline{w}\right) \quad \forall \underline{v}, \underline{w} \in V.$

However, $\bar{d}\left(s\underline{v}, \underline{0}\right)$ is not necessarily equal to $s\bar{d}\left(\underline{v}, \underline{0}\right)$. For instance, $\bar{d}\left(2\underline{e}_1, \underline{0}\right) = 1 < 2\bar{d}\left(\underline{e}_1, \underline{0}\right)$. Thus, the function $f \colon \mathbb{R}^n \to \mathbb{R}$ defined by

$$f\left(\underline{v}\right) = \bar{d}\left(\underline{v}, \underline{0}\right).$$

is not a norm.

Example 3.5.3. *The following functions are examples of norms for $\mathbb{R}^n$ and $\mathbb{C}^n$:*

1. the l^1 norm: $\|\underline{v}\|_1 = \sum_{i=1}^{n} |v_i|$

2. the l^p norm: $\|\underline{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}, \quad p \in (1, \infty)$

3. the l^∞ norm: $\|\underline{v}\|_\infty = \max_{1,\ldots,n}\{|v_i|\}.$

Example 3.5.4. *Similarly, norms can be defined for the vector space of functions from $[a, b]$ to $\mathbb{R}$ (or $\mathbb{C}$) with*

1. the L^1 norm: $\|f(t)\|_1 = \int_a^b |f(t)| dt$

2. the L^p norm: $\|f(t)\|_p = \left(\int_a^b |f(t)|^p dt\right)^{\frac{1}{p}}, \quad p \in (1, \infty)$

3. the L^∞ norm: $\|f(t)\|_\infty = \text{esssup}_{[a,b]}\{|f(t)|\}.$

In this example, the integral notation refers to the **Lebesgue integral** (rather than the **Riemann integral**).

What is the Lebesgue Integral?

Many important spaces include functions that are not Riemann integrable. The Lebesgue integral is defined using measure theory and is often used in advanced probability courses. Since there are many non-zero Lebesgue-integrable functions whose integral is zero, this definition has a subtlety. The norm of a function is zero if and only if it is zero **almost everywhere** (abbreviated a.e.). Therefore, two functions are *equal almost everywhere* if the norm of their difference is zero. Strictly speaking, a vector space of "functions" with the L^p norm actually has elements that are equivalence classes of functions defined by *equality almost everywhere*.

Consider the set of all functions $f \colon X \to \mathbb{R}$ from X to the real numbers. The normed vector space $L^p(X)$ (with $1 \le p < \infty$) is the subset where the Lebesgue integral

$$\|f\|_{L^p} \triangleq \left(\int_X |f(x)|^p \, dx \right)^{1/p}$$

exists and is finite. Of course, this definition begs the question, "What is the Lebesgue integral?". For this course, we will be satisfied with the following definition:

Definition 3.5.5. *The **Lebesgue integral** is a generalization of the Riemann integral that applies to wider class of functions. The values of these two integrals coincide on the set of Riemann integrable functions. Loosely speaking, one can construct any non-negative function $f \in L^p(X)$ by considering sequences $f_1, f_2, \ldots$ of "simple" functions that are constructed by rounding f down to values in an increasing sequence of finite sets that contain 0. By construction, the sequence of functions is non-decreasing (i.e., $f_{n+1}(x) \ge f_n(x)$ for all $x \in X$) and, therefore, it converges pointwise to a limit function $f(x)$. Moreover, the Lebesgue integral of each simple function is easy to define. Thus, this sequence of simple functions gives rise to a non-decreasing sequence of Lebesgue integrals and one can define the Lebesgue integral of $f(x)$ to be the limit of this sequence. In fact, the non-negative functions in $L^p(X)$ are in one-to-one correspondence with the limits of non-decreasing sequences of simple functions that satisfy $\|f_n\|_{L^p} \to M < \infty$.*

Definition 3.5.6. *The **Lebesgue measure** of a set is equal to the Lebesgue integral of its indicator function when both quantites exist. In particular, the set is measurable iff the Lebesgue integral of its indicator function exists.*

Definition 3.5.7. *A vector $\underline{v} \in V$ is said to be **normalized** if $\|\underline{v}\| = 1$. Any vector can be normalized, except the zero vector:*

$$\underline{u} = \frac{\underline{v}}{\|\underline{v}\|} \tag{3.4}$$

*has norm $\|\underline{u}\| = 1$. A normalized vector is also referred to as a **unit vector**.*

Definition 3.5.8. *A complete normed vector space is called a **Banach space**.*

Banach spaces are the standard setting for many problems because completeness is a powerful tool for solving problems.

Example 3.5.9. *The vector spaces $\mathbb{R}^n$ (or $\mathbb{C}^n$) with any well-defined norm are Banach spaces.*

Example 3.5.10. *The vector space of all continuous functions from $[a, b]$ to $\mathbb{R}$ is a Banach space under the supremum norm*

$$\|f(t)\| = \sup_{t \in [a,b]} f(t).$$

Definition 3.5.11. *A Banach space V has a **Schauder basis**, $\underline{v}_1, \underline{v}_2, \ldots$, if every $\underline{v} \in V$ can be written uniquely as*

$$\underline{v} = \sum_{i \in \mathbb{N}} s_i \underline{v}_i.$$

Example 3.5.12. *Let $V = \mathbb{R}^\omega$ be the vector space of semi-infinite real sequences. The **standard Schauder basis** is the countably infinite extension $\{\underline{e}_1, \underline{e}_2, \ldots\}$ of the standard basis.*

Definition 3.5.13. *A **closed subspace** of a Banach space is a subspace that is a closed set in the topology generated by the norm.*

Theorem 3.5.14. *All finite dimensional subspaces of a Banach space are closed.*

Proof. This proof requires material from later in the notes, but is given here for completeness. Let $\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n$ be a basis for a finite dimensional subspace W of a Banach space V over F. Let $U = F^n$ be the standard Banach space, which is closed by definition, and consider the mapping $f : U \to W$ defined by

$$f(\underline{s}) = \sum_{i=1}^{n} s_i \underline{w}_i.$$

It is easy to verify that this linear mapping is non-singular and onto. Therefore, it has a linear inverse mapping $g = f^{-1}$ that must be continuous (i.e., bounded) because U, W are finite dimensional. Since g is continuous, we find that $W = g^{-1}(U) = f(U)$ is closed because U is closed. $\qquad\square$

Example 3.5.15. *Let $V = L^p([a, b])$, for $1 \le p < \infty$, be the set of real Lebesgue-integrable functions on $[a, b]$. We say that $f \in V$ is continuous if the equivalence class generated by equality almost everywhere contains a continuous function. It is easy to verify that the subset $W \subset V$ of continuous functions is a subspace. It is not closed, however, because sequences in W may converge to discontinuous functions. In fact, the set of continuous functions is dense in $L^p([a, b])$ for $p \in [1, \infty)$.*

Example 3.5.16. *By definition, the span of an infinite set of vectors only includes finite linear combinations. However, a sequence of finite linear combinations, like*

$$\underline{v}_n = \sum_{i=1}^n \frac{1}{i^2} \underline{e}_i,$$

can converge to an infinite linear combination. Thus, the span of any infinite set of linearly independent vectors is not closed.

3.6   Inner Products

Definition 3.6.1. *Let F be the field of real numbers or the field of complex numbers, and assume V is a vector space over F. An **inner product** on V is a function which assigns to each ordered pair of vectors $\underline{v}, \underline{w} \in V$ a scalar $\langle \underline{v}|\underline{w} \rangle \in F$ in such a way that for all $\underline{u}, \underline{v}, \underline{w} \in V$ and any scalar $s \in F$*

1. $\langle \underline{u} + \underline{v}|\underline{w} \rangle = \langle \underline{u}|\underline{w} \rangle + \langle \underline{v}|\underline{w} \rangle$

2. $\langle s\underline{v}|\underline{w} \rangle = s \langle \underline{v}|\underline{w} \rangle$

3. $\langle \underline{v}|\underline{w} \rangle = \overline{\langle \underline{w}|\underline{v} \rangle}$, *where the overbar denotes complex conjugation;*

4. $\langle \underline{v}|\underline{v} \rangle \ge 0$ *with equality iff $\underline{v} = \underline{0}$.*

Note that the conditions of Definition 3.6.1 imply that

$$\langle \underline{u}|s\underline{v} + \underline{w} \rangle = \overline{s} \langle \underline{u}|\underline{v} \rangle + \langle \underline{u}|\underline{w} \rangle .$$

Definition 3.6.2. *A real or complex vector space equipped with an inner product is called an* **inner-product space**.

Example 3.6.3. *Consider the inner product on F^n defined by*

$$\langle \underline{v} | \underline{w} \rangle = \langle (v_1, \ldots, v_n) | (w_1, \ldots, w_n) \rangle = \sum_{j=1}^{n} v_j \overline{w}_j.$$

This inner product is called the **standard inner product**. *When $F = \mathbb{R}$, the standard inner product can also be written as*

$$\langle \underline{v} | \underline{w} \rangle = \sum_{j=1}^{n} v_j w_j.$$

In this context it is often called the dot product, denoted by $\underline{v} \cdot \underline{w}$. In either case, it can also be written in terms of the Hermitian transpose as $\langle \underline{v} | \underline{w} \rangle = \underline{w}^H \underline{v}$.

Problem 3.6.4. *For $\underline{v} = (v_1, v_2)$ and $\underline{w} = (w_1, w_2)$ in $\mathbb{R}^2$, show that*

$$\langle \underline{v} | \underline{w} \rangle = v_1 w_1 - v_2 w_1 - v_1 w_2 + 4 v_2 w_2$$

is an inner product.

S 3.6.4. For all $\underline{u}, \underline{v}, \underline{w} \in V$ and all scalars s

$$\begin{aligned}
\langle \underline{u} + \underline{v} | \underline{w} \rangle &= (u_1 + v_1)w_1 - (u_2 + v_2)w_1 - (u_1 + v_1)w_2 + 4(u_2 + v_2)w_2 \\
&= u_1 w_1 - u_2 w_1 - u_1 w_2 + 4 u_2 w_2 + v_1 w_1 - v_2 w_1 - v_1 w_2 + 4 v_2 w_2 \\
&= \langle \underline{u} | \underline{w} \rangle + \langle \underline{v} | \underline{w} \rangle.
\end{aligned}$$

Also, we have

$$\langle s\underline{v} | \underline{w} \rangle = s v_1 w_1 - s v_2 w_1 - s v_1 w_2 + 4 s v_2 w_2 = s \langle \underline{v} | \underline{w} \rangle.$$

Since $V = \mathbb{R}^2$, we have $\langle \underline{v} | \underline{w} \rangle = \overline{\langle \underline{w} | \underline{v} \rangle}$. Furthermore,

$$\langle \underline{v} | \underline{v} \rangle = v_1^2 - 2 v_1 v_2 + 4 v_2^2 = (v_1 - v_2)^2 + 3 v_2^2 \geq 0 \quad \text{with equality iff } \underline{v} = \underline{0}.$$

That is, $\langle \underline{v} | \underline{v} \rangle$ is an inner product.

Example 3.6.5. *Let V be the vector space of all continuous complex-valued functions on the unit interval $[0, 1]$. Then*

$$\langle f | g \rangle = \int_0^1 f(t) \overline{g(t)} dt$$

is an inner product.

Example 3.6.6. *Let V and W be two vector spaces over F and suppose that $\langle \cdot | \cdot \rangle_W$ is an inner product on W. If T is a non-singular linear transformation from V into W, then the equation*

$$\langle \underline{v}_1, \underline{v}_2 \rangle_V = \langle T\underline{v}_1 | T\underline{v}_2 \rangle_W$$

defines an inner product on V.

Example 3.6.7. *Let $V = F^{m \times n}$ be the space of $m \times n$ matrices over F and define the inner product for matrices $A, B \in V$ to be*

$$\langle A|B \rangle \triangleq \mathrm{tr}\left(B^H A\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \overline{b}_{j,i} a_{j,i}.$$

This also equals $\mathrm{tr}\left(AB^H\right)$ and both are identical to writing the entries of the matrices as length-mn vectors and then applying the standard inner product.

Theorem 3.6.8. *Let V be a finite-dimensional space, and suppose that*

$$\mathcal{B} = \underline{w}_1, \ldots, \underline{w}_n$$

is an ordered basis for V. Any inner product on V is determined by the values

$$g_{ij} = \left\langle \underline{w}_j | \underline{w}_i \right\rangle$$

that it takes on pairs of vectors in $\mathcal{B}$.

Proof. If $\underline{u} = \sum_j s_j \underline{w}_j$ and $\underline{v} = \sum_i t_i \underline{w}_i$, then

$$\langle \underline{u} | \underline{v} \rangle = \left\langle \sum_j s_j \underline{w}_j \middle| \underline{v} \right\rangle = \sum_j s_j \left\langle \underline{w}_j | \underline{v} \right\rangle$$

$$= \sum_j s_j \left\langle \underline{w}_j \middle| \sum_i t_i \underline{w}_i \right\rangle = \sum_j \sum_i s_j \overline{t}_i \left\langle \underline{w}_j | \underline{w}_i \right\rangle$$

$$= \sum_j \sum_i \overline{t}_i g_{ij} s_j = [\underline{v}]_{\mathcal{B}}^H \, G \, [\underline{u}]_{\mathcal{B}}$$

where $[\underline{u}]_{\mathcal{B}}$ and $[\underline{v}]_{\mathcal{B}}$ are the coordinate matrices of $\underline{u}, \underline{v}$ in the ordered basis $\mathcal{B}$. The matrix G is called the weight matrix *of the inner product in the ordered basis $\mathcal{B}$.* $\quad \square$

It is easily verified that G is a Hermitian matrix, i.e., $G = G^H$. Furthermore, G must satisfy the additional condition

$$\underline{w}^H G \underline{w} > 0, \quad \forall \underline{w} \neq \underline{0} \tag{3.5}$$

so that the induced norm is non-negative and zero only for the zero vector. A Hermitian matrix that satisfies this condition is called positive definite and this also implies that G is invertible.

Conversely if G is an $n \times n$ Hermitian matrix over F which satisfies (3.5), then G is the matrix in the ordered basis $\mathcal{B}$ of an inner product on V. This inner product is given by

$$\langle \underline{u} | \underline{v} \rangle_G = [\underline{v}]_{\mathcal{B}}^H \, G \, [\underline{u}]_{\mathcal{B}} \,.$$

Problem 3.6.9. *Let V be a vector space over F. Show that the sum of two inner products on V is an inner product on V. Show that a positive multiple of an inner product is also an inner product.*

Example 3.6.10. *Consider any set W of real-valued random variables, defined on a common probability space, that have finite 2nd moments. It turns out that $V = \operatorname{span}(W)$ is a vector space over $\mathbb{R}$. In fact, one can define the inner product*

$$\langle X | Y \rangle = \mathrm{E} \left[XY \right],$$

for any $X, Y \in V$. Using the induced norm, this inner product provides the topology of mean-square convergence and two random variables $X, Y \in V$ are considered equal if $\| X - Y \|^2 = E \left[|X - Y|^2 \right] = 0$ (or equivalently $\Pr(X \neq Y) = 0$).

In terms of abstract mathematics, the introduction of an inner product allows one to introduce the key concept of orthogonality.

Definition 3.6.11. *Let $\underline{v}$ and $\underline{w}$ be vectors in an inner-product space V. Then $\underline{v}$ is **orthogonal** to $\underline{w}$ (denoted $\underline{v} \perp \underline{w}$) if $\langle \underline{v} | \underline{w} \rangle = 0$. Since this relation is reflexive and $\underline{w}$ is also orthogonal to $\underline{v}$, we simply say that $\underline{v}$ and $\underline{w}$ are orthogonal.*

3.6.1 Induced Norms

A finite-dimensional real inner-product space is often referred to as a **Euclidean space**. A complex inner-product space is sometimes called a unitary space.
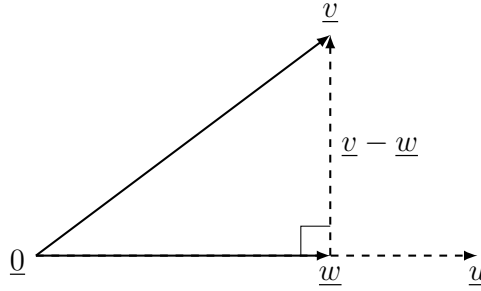
Figure 3.1: The projection of $\underline{v}$ onto $\underline{u}$ is given by $\underline{w}$ and $\underline{v} - \underline{w}$ is orthogonal to $\underline{u}$.

Definition 3.6.12. *Let V be an inner-product space with inner product $\langle \cdot | \cdot \rangle$. This inner product can be used to define a norm, called the **induced norm**,*

$$\|\underline{v}\| = \langle \underline{v} | \underline{v} \rangle^{\frac{1}{2}}$$

for every $\underline{v} \in V$.

Definition 3.6.13. *Let $\underline{u}, \underline{v}$ be vectors in an inner-product space V with inner product $\langle \cdot | \cdot \rangle$. As shown in Figure 3.1, the **projection** of $\underline{v}$ onto $\underline{u}$ is defined to be*

$$\underline{w} = \frac{\langle \underline{v} | \underline{u} \rangle}{\|\underline{u}\|^2} \underline{u}$$

Theorem 3.6.14. *If V is an inner-product space and $\| \cdot \|$ is its associated induced norm, then for any $\underline{v}, \underline{w} \in V$ and any scalar s*

1. $\|s\underline{v}\| = |s| \, \|\underline{v}\|$

2. $\|\underline{v}\| > 0$ *for $\underline{v} \neq \underline{0}$*

3. $|\langle \underline{v} | \underline{w} \rangle| \leq \|\underline{v}\| \, \|\underline{w}\|$ *with equality iff $\underline{v} = \underline{0}$, $\underline{w} = \underline{0}$, or $\underline{v} = s\underline{w}$*

4. $\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\|$ *with equality iff $\underline{v} = \underline{0}$, $\underline{w} = \underline{0}$, or $\underline{v} = s\underline{w}$.*

Proof. The first two items follow immediately from the definitions involved. The third inequality, $|\langle \underline{v} | \underline{w} \rangle| \leq \|\underline{v}\| \, \|\underline{w}\|$, is called the **Cauchy-Schwarz inequality**. When $\underline{v} = \underline{0}$, then clearly $|\langle \underline{v} | \underline{w} \rangle| = \|\underline{v}\| \, \|\underline{w}\| = 0$. Assume $\underline{v} \neq \underline{0}$ and let

$$\underline{u} = \underline{w} - \frac{\langle \underline{w} | \underline{v} \rangle}{\|\underline{v}\|^2} \underline{v}$$

be the residual after projecting $\underline{w}$ onto $\underline{v}$. Then, $\langle \underline{u} | \underline{v} \rangle = 0$ and

$$0 \leq \|\underline{u}\|^2 = \left\langle \underline{u} \ \middle| \ \underline{w} - \frac{\langle \underline{w} | \underline{v} \rangle}{\|\underline{v}\|^2} \underline{v} \right\rangle = \langle \underline{u} \, | \, \underline{w} \rangle = \left\langle \underline{w} - \frac{\langle \underline{w} | \underline{v} \rangle}{\|\underline{v}\|^2} \underline{v} \ \middle| \ \underline{w} \right\rangle$$

$$= \langle \underline{w} | \underline{w} \rangle - \frac{\langle \underline{w} | \underline{v} \rangle \langle \underline{v} | \underline{w} \rangle}{\|\underline{v}\|^2} = \|\underline{w}\|^2 - \frac{|\langle \underline{w} | \underline{v} \rangle|^2}{\|\underline{v}\|^2},$$

where equality holds iff $\underline{u} = \underline{0}$, or equivalently iff $\underline{w} = \underline{0}$ or $\underline{v} = s\underline{w}$. Hence, we find $|\langle \underline{v} | \underline{w} \rangle|^2 = |\langle \underline{w} | \underline{v} \rangle|^2 \leq \|\underline{v}\|^2 \|\underline{w}\|^2$. Note that the $\underline{v} = \underline{0}$ case is checked separately because it is assumed above that $\underline{v} \neq \underline{0}$.

Using this result, we get

$$\|\underline{v} + \underline{w}\|^2 = \|\underline{v}\|^2 + \langle \underline{v} | \underline{w} \rangle + \langle \underline{w} | \underline{v} \rangle + \|\underline{w}\|^2$$

$$= \|\underline{v}\|^2 + 2\mathrm{Re} \, \langle \underline{v} | \underline{w} \rangle + \|\underline{w}\|^2$$

$$\leq \|\underline{v}\|^2 + 2 \, \|\underline{v}\| \, \|\underline{w}\| + \|\underline{w}\|^2 \, ,$$

with equality iff Cauchy-Schwarz holds with equality. Thus, $\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\|$ with equality iff $\underline{v} = \underline{0}$, $\underline{w} = \underline{0}$, or $\underline{v} = s\underline{w}$ (i.e., $\underline{v}$ and $\underline{w}$ are linearly dependent). $\quad\square$

Theorem 3.6.15. *Consider the vector space $\mathbb{R}^n$ with the standard inner product of Example 3.6.3. The function $f \colon V \to F$ defined by $f(\underline{w}) = \langle \underline{w} | \underline{v} \rangle$ is continuous.*

Proof. Let $\underline{w}_1, \underline{w}_2, \ldots$ be a sequence in V converging to $\underline{w}$. Then,

$$|\langle \underline{w}_n | \underline{v} \rangle - \langle \underline{w} | \underline{v} \rangle| = |\langle \underline{w}_n - \underline{w} | \underline{v} \rangle| \leq \|\underline{w}_n - \underline{w}\| \, \|\underline{v}\| \, .$$

Since $\|\underline{w}_n - \underline{w}\| \to 0$, the convergence of $\langle \underline{w}_n, \underline{v} \rangle$ is established. $\quad\square$

3.7 Sets of Orthogonal Vectors

Definition 3.7.1. *Let V be an inner-product space and U, W be subspaces. Then, the subspace U is an **orthogonal** to the subspace W (denoted $U \perp W$) if $\underline{u} \perp \underline{w}$ for all $\underline{u} \in U$ and $\underline{w} \in W$.*

Definition 3.7.2. *A collection W of vectors in V is an **orthogonal set** if all pairs of distinct vectors in W are orthogonal.*

Example 3.7.3. *The standard basis of $\mathbb{R}^n$ is an orthonormal set with respect to the standard inner product.*

Example 3.7.4. *Let V be the vector space (over $\mathbb{C}$) of continuous complex-valued functions on the interval $0 \leq x \leq 1$ with the inner product*

$$\langle f|g \rangle = \int_0^1 f(x)\overline{g(x)}dx.$$

Let $f_n(x) = \sqrt{2}\cos 2\pi nx$ and $g_n(x) = \sqrt{2}\sin 2\pi nx$. Then $\{1, f_1, g_1, f_2, g_2, \ldots\}$ is a countably infinite orthonormal set that is a Schauder basis for this vector space.

Theorem 3.7.5. *An orthogonal set of non-zero vectors is linearly independent.*

Proof. Let W be an orthogonal set of non-zero vectors in a given inner-product space V. Suppose $\underline{w}_1, \ldots, \underline{w}_n$ are distinct vectors in W and consider

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

The inner product $\langle \underline{v}|\underline{w}_i \rangle$ is given by

$$\langle \underline{v}|\underline{w}_i \rangle = \left\langle \sum_j s_j\underline{w}_j | \underline{w}_i \right\rangle = \sum_j s_j \langle \underline{w}_j|\underline{w}_i \rangle = s_i \langle \underline{w}_i|\underline{w}_i \rangle.$$

Since $\langle \underline{w}_i|\underline{w}_i \rangle \neq 0$, it follows that

$$s_i = \frac{\langle \underline{v}|\underline{w}_i \rangle}{\|\underline{w}_i\|^2} \quad 1 \leq i \leq n.$$

In particular, if $\underline{v} = 0$ then $s_j = 0$ for $1 \leq j \leq n$ and the vectors in W are linearly independent. $\square$

Corollary 3.7.6. *If $\underline{v} \in V$ is a linear combination of an orthogonal sequence of distinct, non-zero vectors $\underline{w}_1, \ldots, \underline{w}_n$, then $\underline{v}$ satisfies the identity*

$$\underline{v} = \sum_{i=1}^n \frac{\langle \underline{v}|\underline{w}_i \rangle}{\|\underline{w}_i\|^2}\underline{w}_i,$$

and equals the sum of the projections of $\underline{v}$ onto $\underline{w}_1, \ldots, \underline{w}_n$.

Theorem 3.7.7. *Let V be an inner-product space and assume $\underline{v}_1, \ldots, \underline{v}_n$ are linearly independent vectors in V. Then it is possible to construct an orthogonal sequence of vectors $\underline{w}_1, \ldots, \underline{w}_n \in V$ such that for each $k = 1, \ldots, n$ the set*

$$\{\underline{w}_1, \ldots, \underline{w}_k\}$$

is a basis for the subspace spanned by $\underline{v}_1, \ldots, \underline{v}_k$.

Proof. First, let $\underline{w}_1 = \underline{v}_1$. The remaining vectors are defined inductively as part during the proof. Suppose the vectors

$$\underline{w}_1, \ldots, \underline{w}_m \quad (1 \leq m < n)$$

have been chosen so that for every k

$$\{\underline{w}_1, \ldots, \underline{w}_k\} \quad 1 \leq k \leq m$$

is an orthogonal basis for the subspace spanned by $\underline{v}_1, \ldots, \underline{v}_k$. Let

$$\underline{w}_{m+1} = \underline{v}_{m+1} - \sum_{i=1}^{m} \frac{\langle \underline{v}_{m+1} | \underline{w}_i \rangle}{\|\underline{w}_i\|^2} \underline{w}_i.$$

Then $\underline{w}_{m+1} \neq 0$, for otherwise $\underline{v}_{m+1}$ is a linear combination of $\underline{w}_1, \ldots, \underline{w}_m$ and hence a linear combination of $\underline{v}_1, \ldots, \underline{v}_m$. For $j \in \{1, \ldots, m\}$, we also have

$$\begin{aligned}
\langle \underline{w}_{m+1} | \underline{w}_j \rangle &= \langle \underline{v}_{m+1} | \underline{w}_j \rangle - \sum_{i=1}^{m} \frac{\langle \underline{v}_{m+1} | \underline{w}_i \rangle}{\|\underline{w}_i\|^2} \langle \underline{w}_i | \underline{w}_j \rangle \\
&= \langle \underline{v}_{m+1} | \underline{w}_j \rangle - \frac{\langle \underline{v}_{m+1} | \underline{w}_j \rangle}{\|\underline{w}_j\|^2} \langle \underline{w}_j | \underline{w}_j \rangle \\
&= 0.
\end{aligned}$$

Clearly, $\{\underline{w}_1, \ldots, \underline{w}_{m+1}\}$ is an orthogonal set consisting of $m + 1$ non-zero vectors in the subspace spanned by $\underline{v}_1, \ldots, \underline{v}_{m+1}$. Since the dimension of the latter subspace is $m + 1$, this set is a basis for the subspace. $\square$

The inductive construction of the vectors $\underline{w}_1, \ldots, \underline{w}_n$ is known as the **Gram-Schmidt orthogonalization** process.

Corollary 3.7.8. *Every finite-dimensional inner-product space has a basis of orthonormal vectors.*

Proof. Let V be a finite-dimensional inner-product space. Suppose that $\underline{v}_1, \ldots, \underline{v}_n$ is a basis for V. Apply the Gram-Schmidt process to obtain a basis of orthogonal vectors $\underline{w}_1, \ldots, \underline{w}_n$. Then, a basis of orthonormal vectors is given by

$$\underline{u}_1 = \frac{\underline{w}_1}{\|\underline{w}_1\|}, \ldots, \underline{u}_n = \frac{\underline{w}_n}{\|\underline{w}_n\|}.$$

$\square$

Example 3.7.9. *Consider the vectors*

$$\underline{v}_1 = (2, 2, 1)$$
$$\underline{v}_2 = (3, 6, 0)$$
$$\underline{v}_3 = (6, 3, 9)$$

in $\mathbb{R}^3$ equipped with the standard inner product. Apply the Gram-Schmidt process to $\underline{v}_1, \underline{v}_2, \underline{v}_3$ to obtain an orthogonal basis.

Applying the Gram-Schmidt process to $\underline{v}_1, \underline{v}_2, \underline{v}_3$, we get

$$\underline{w}_1 = (2, 2, 1)$$
$$\underline{w}_2 = (3, 6, 0) - \frac{\langle (3,6,0)|(2,2,1) \rangle}{9}(2, 2, 1)$$
$$= (3, 6, 0) - 2(2, 2, 1) = (-1, 2, -2)$$
$$\underline{w}_3 = (6, 3, 9) - \frac{\langle (6,3,9)|(2,2,1) \rangle}{9}(2, 2, 1) - \frac{\langle (6,3,9)|(-1,2,-2) \rangle}{9}(-1, 2, -2)$$
$$= (6, 3, 9) - 3(2, 2, 1) + 2(-1, 2, -2) = (-2, 1, 2).$$

It is easily verified that $\underline{w}_1, \underline{w}_2, \underline{w}_3$ is an orthogonal set of vectors.

Definition 3.7.10. *Let V be an inner-product space and W be any set of vectors in V. The **orthogonal complement** of W denoted by $W^\perp$ is the set of all vectors in V that are orthogonal to every vector in W or*

$$W^\perp = \left\{ \underline{v} \in V \,\middle|\, \langle \underline{v}|\underline{w} \rangle = 0 \;\forall\; \underline{w} \in W \right\}.$$

Problem 3.7.11. *Let W be any subset of vector space V. Show that $W^\perp$ is a closed subspace of V and that any vector in the subspace spanned by W is orthogonal to any vector in $W^\perp$.*

S 3.7.11. Let $\underline{m}_1, \underline{m}_2 \in W^\perp$ and $s \in F$. For any vector $\underline{w} \in W$, we have

$$\langle \underline{m}_1|\underline{w} \rangle = \langle \underline{m}_2|\underline{w} \rangle = 0.$$

This implies

$$\langle s\underline{m}_1 + \underline{m}_2|\underline{w} \rangle = s\langle \underline{m}_1|\underline{w} \rangle + \langle \underline{m}_2|\underline{w} \rangle = 0.$$

That is, $s\underline{m}_1 + \underline{m}_2 \in W^\perp$. Hence, $W^\perp$ is a subspace of V.

To see that $W^\perp$ is closed, we let $\underline{m}$ be any point in the closure of $W^\perp$ and $\underline{m}_1, \underline{m}_2, \ldots \in W^\perp$ be a sequence that converges to $\underline{m}$. The continuity of the inner product, from Theorem 3.6.15, implies that, for all $\underline{w} \in W$,

$$\langle \underline{m}|\underline{w} \rangle = \left\langle \lim_{n \to \infty} \underline{m}_n | \underline{w} \right\rangle = \lim_{n \to \infty} \langle \underline{m}_n | \underline{w} \rangle = 0.$$

Therefore, $\underline{m} \in W^\perp$ and the orthogonal complement contains all of its limit points.

Notice also that any vector $\underline{w}$ in the subspace spanned by W can be written as $\underline{w} = \sum_i s_i \underline{w}_i$ with $\underline{w}_i \in W$ and $s_i \in F$. Therefore, the inner product of $\underline{w}$ with any $\underline{w}' \in W^\perp$ is given by

$$\langle \underline{w}|\underline{w}' \rangle = \left\langle \sum_i s_i \underline{w}_i \Big| \underline{w}' \right\rangle = \sum_i s_i \langle \underline{w}_i | \underline{w}' \rangle = 0.$$

It follows that the subspace spanned by W is orthogonal to the subspace $W^\perp$.

3.7.1 Hilbert Spaces

Definition 3.7.12. *A complete inner-product space is called a **Hilbert space**.*

Definition 3.7.13. *Recall that a subset $\{\underline{v}_\alpha | \alpha \in A\}$ of a Hilbert space V is said to be orthonormal if $\|\underline{v}_\alpha\| = 1$ for every $\alpha \in A$ and $\langle \underline{v}_\alpha | \underline{v}_\beta \rangle = 0$ for all $\alpha \neq \beta$. If the subspace spanned by the family $\{\underline{v}_\alpha | \alpha \in A\}$ is dense in V, we call this set an **orthonormal basis**.*

Note that, according to this definition, an orthonormal basis for a Hilbert space V is not necessarily a Hamel basis for V. However, it can be shown that any orthogonal basis is a subset of a Hamel basis. In practice it is the orthonormal basis, not the Hamel basis itself, which is of most use. None of these issues arise in finite-dimensional spaces, where an orthogonal basis is always a Hamel basis.

Let $\mathcal{B} = \{\underline{v}_\alpha | \alpha \in A\}$ be an orthonormal basis for Hilbert space V. Each element $\underline{v} \in V$ has the form

$$\underline{v} = \sum_{\alpha \in A} s_\alpha \underline{v}_\alpha,$$

where the sum converges in the (2)-norm. The **Parseval identity**

$$\|\underline{v}\|^2 = \sum_{\alpha \in A} |s_\alpha|^2$$

is obtained by computing $\langle \underline{v}|\underline{v} \rangle$.

Theorem 3.7.14. *Every orthogonal set in a Hilbert space V can be enlarged to an orthonormal basis for V.*

Proof. Let X be the set of orthonormal subsets of V. Furthermore, for $x, y \in X$ consider the strict partial order defined by proper inclusion. If $x = \{\underline{v}_\alpha | \alpha \in A_0\}$ is an element of X, then by the Hausdorff maximal principle there exists a maximal simply ordered subset Z of X containing x. This shows the existence of a maximal orthonormal set $\{\underline{v}_\alpha | \alpha \in A\}$, where $A_0 \subset A$.

Let W be the closed subspace of V generated by $\{\underline{v}_\alpha | \alpha \in A\}$. If $W \neq V$, there is a unit vector $\underline{u} \in W^\perp$, contradicting the maximality of the system $\{\underline{v}_\alpha | \alpha \in A\}$. Thus, $W = V$ and we have an orthonormal basis. $\square$

Theorem 3.7.15. *A Hilbert space V has a countable orthonormal basis if and only if V is separable.*

Sketch of proof. If V is separable, then it contains a countable dense subset. Using the well-ordering theorem, this subset can be ordered into a sequence $\underline{v}_1, \underline{v}_2, \ldots$ such that, for every vector $\underline{v} \in V$ and any $\epsilon > 0$, there exists an n such that $\|\underline{v} - \underline{v}_n\| < \epsilon$. A countable orthonormal basis is generated by applying Gram-Schmidt orthogonalization to this ordered sequence of vectors. Conversely, if V has a countable orthonormal basis, then linear combinations with rational coefficients can be used to construct a countable dense subset. $\square$

3.8   Linear Functionals

Definition 3.8.1. *Let V be a vector space over a field F. A linear transformation f from V into the scalar field F is called a **linear functional** on V.*

That is, f is a functional on V such that

$$f(s\underline{v}_1 + \underline{v}_2) = sf(\underline{v}_1) + f(\underline{v}_2)$$

for all $\underline{v}_1, \underline{v}_2 \in V$ and $s \in F$.

Example 3.8.2. *Let F be a field and let $s_1, \ldots, s_n$ be scalars in F. Then the functional f on F^n defined by*

$$f(v_1, \ldots, v_n) = s_1 v_1 + \cdots + s_n v_n$$

is a linear functional. It is the linear functional which is represented by the matrix

$$\begin{bmatrix} s_1 & s_2 & \cdots & s_n \end{bmatrix}$$

relative to the standard ordered basis for F^n. Every linear functional on F^n is of this form, for some scalars $s_1, \ldots, s_n$.

Definition 3.8.3. *Let n be a positive integer and F a field. If A is an $n \times n$ matrix with entries in F, the **trace** of A is the scalar*

$$\operatorname{tr}(A) = A_{11} + A_{22} + \cdots + A_{nn}.$$

Example 3.8.4. *The trace function is a linear functional on the matrix space $F^{n \times n}$ since*

$$\operatorname{tr}(sA + B) = \sum_{i=1}^{n}(sA_{ii} + B_{ii})$$

$$= s\sum_{i=1}^{n} A_{ii} + \sum_{i=1}^{n} B_{ii}$$

$$= s\operatorname{tr}(A) + \operatorname{tr}(B).$$

Example 3.8.5. *Let $[a, b]$ be a closed interval on the real line and let $C([a, b])$ be the space of continuous real-valued functions on $[a, b]$. Then*

$$L(g) = \int_a^b g(t)dt$$

defines a linear functional L on $C([a, b])$.

Theorem 3.8.6 (Riesz). *Let V be a finite-dimensional Hilbert space and f be a 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. If we choose an orthonormal basis $\mathcal{B} = \underline{v}_1, \ldots, \underline{v}_n$ for V, then the inner product of $\underline{w} = t_1\underline{v}_1 + \cdots + t_n\underline{v}_n$ and $\underline{v} = s_1\underline{v}_1 + \cdots + s_n\underline{v}_n$ will be

$$\langle \underline{w} | \underline{v} \rangle = t_1\bar{s}_1 + \cdots + t_n\bar{s}_n.$$

If f is a linear functional on V, then f has the form

$$f(\underline{w}) = f(t_1\underline{v}_1 + \cdots + t_n\underline{v}_n) = t_1 f(\underline{v}_1) + \cdots + t_n f(\underline{v}_n).$$

Thus, we can choose $\bar{s}_j = f\left(\underline{v}_j\right)$ to get $\langle\underline{w}|\underline{v}\rangle = f\left(\underline{w}\right)$ and this gives

$$\underline{v} = \overline{f\left(\underline{v}_1\right)}\underline{v}_1 + \cdots + \overline{f\left(\underline{v}_n\right)}\underline{v}_n.$$

Let $\underline{v}'$ be any vector that satisfies $f\left(\underline{w}\right) = \langle\underline{w}|\underline{v}'\rangle$ for all $\underline{w} \in V$. Then, we see that $\langle\underline{w}|\underline{v} - \underline{v}'\rangle = 0$ for all $\underline{w} \in V$. This implies that $\underline{v} - \underline{v}' = \underline{0}$.

$\square$