1 Vector Spaces

In this chapter we will introduce the notion of an abstract vector space, which is, ultimately, a generalization of the ideas inherent in studying vectors in 2- or 3-dimensional space. We introduce vector spaces from an axiomatic perspective, deriving a number of basic properties using only the axioms. We then develop the general theory of vector spaces: we discuss subspaces, span, linear dependence and independence, and then prove that every vector space possesses a linearly independent spanning set called a “basis”. We close with a discussion of vector spaces having an additional kind of structure called an “inner product”, which generalizes the idea of the dot product of vectors in \( \mathbb{R}^n \).

1.1 The Formal Definition of a Vector Space

- The two operations of addition and scalar multiplication (and the various algebraic properties they satisfy) are the key properties of vectors in \( \mathbb{R}^n \) and of matrices. We would like to investigate other collections of things which possess those same properties.

- **Definition:** Let \( F \) be a field, and refer to the elements of \( F \) as scalars. A **vector space over** \( F \) is a triple \( (V, +, \cdot) \) of a collection \( V \) of elements called “vectors”, together with two binary operations\(^1\), addition of vectors \( (+) \) and scalar multiplication of a vector by a scalar \( (\cdot) \), satisfying the following axioms:

  - **[A1]** Addition is commutative: \( v + w = w + v \) for any vectors \( v \) and \( w \).
  - **[A2]** Addition is associative: \( (u + v) + w = u + (v + w) \) for any vectors \( u \), \( v \), and \( w \).
  - **[A3]** There exists a zero vector \( 0 \), with \( v + 0 = v = 0 + v \) for any vector \( v \).
  - **[A4]** Every vector \( v \) has an additive inverse \(-v\), with \( v + (-v) = 0 = (-v) + v \).

\(^1\)The result of adding vectors \( v \) and \( w \) is denoted as \( v + w \), and the result of scalar-multiplying \( v \) by \( \alpha \) is denoted as \( \alpha \cdot v \) (or often simply \( \alpha v \)). The definition of “binary operation” means that \( v + w \) and \( \alpha \cdot v \) are also vectors in \( V \).
Scalar multiplication is consistent with regular multiplication: \( \alpha \cdot (\beta \cdot \mathbf{v}) = (\alpha \beta) \cdot \mathbf{v} \) for any scalars \( \alpha, \beta \) and vector \( \mathbf{v} \).

Addition of scalars distributes: \( (\alpha + \beta) \cdot \mathbf{v} = \alpha \cdot \mathbf{v} + \beta \cdot \mathbf{v} \) for any scalars \( \alpha, \beta \) and vector \( \mathbf{v} \).

Addition of vectors distributes: \( \alpha \cdot (\mathbf{v} + \mathbf{w}) = \alpha \cdot \mathbf{v} + \alpha \cdot \mathbf{w} \) for any scalar \( \alpha \) and vectors \( \mathbf{v} \) and \( \mathbf{w} \).

The scalar 1 acts like the identity on vectors: \( 1 \cdot \mathbf{v} = \mathbf{v} \) for any vector \( \mathbf{v} \).

We will primarily consider vector spaces where the collection of scalars (namely, the field \( F \)) is either the set of real numbers or the set of complex numbers: we refer to such vector spaces as real vector spaces or complex vector spaces, respectively.

However, all of the general theory of vector spaces will hold over any field. Some complications can arise in certain kinds of fields (such as the two-element field \( F_2 = \{0, 1\} \) where \( 1 + 1 = 0 \)) where adding 1 a finite number of times to itself yields 0; we will generally seek to gloss over such complications.

Here are some examples of vector spaces:

**Example:** The vectors in \( \mathbb{R}^n \) are a real vector space, for any \( n > 0 \).

- For simplicity we will demonstrate all of the axioms for vectors in \( \mathbb{R}^2 \); there, the vectors are of the form \( (x, y) \) and scalar multiplication is defined as \( \alpha \cdot (x, y) = (\alpha x, \alpha y) \). (Note that the dot here is not the dot product.)
- [A1]: We have \( (x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) = (x_2, y_2) + (x_1, y_1) \).
- [A2]: We have \( ((x_1, y_1) + (x_2, y_2)) + (x_3, y_3) = (x_1 + x_2 + x_3, y_1 + y_2 + y_3) = (x_1, y_1) + (x_2, y_2) + (x_3, y_3) \).
- [A3]: The zero vector is \((0, 0)\), and clearly \( (x, y) + (0, 0) = (x, y) \).
- [A4]: The additive inverse of \((x, y)\) is \((-x, -y)\), since \((x, y) + (-x, -y) = (0, 0)\).
- [M1]: We have \( \alpha_1 \cdot (\alpha_2 \cdot (x, y)) = (\alpha_1 \alpha_2 x, \alpha_1 \alpha_2 y) = (\alpha_1 \alpha_2) \cdot (x, y) \).
- [M2]: We have \( (\alpha_1 + \alpha_2) \cdot (x, y) = ((\alpha_1 + \alpha_2) x, (\alpha_1 + \alpha_2) y) = \alpha_1 \cdot (x, y) + \alpha_2 \cdot (x, y) \).
- [M3]: We have \( \alpha \cdot ((x_1, y_1) + (x_2, y_2)) = \alpha (x_1 + x_2), \alpha (y_1 + y_2)) = \alpha \cdot (x_1, y_1) + \alpha \cdot (x_2, y_2) \).
- [M4]: Finally, we have \( 1 \cdot (x, y) = (x, y) \).

**Example:** The set \( M_{m\times n}(F) \) of \( m \times n \) matrices, for any fixed \( m \) and \( n \), forms a vector space over \( F \).

- The various algebraic properties we know about matrix addition give [A1] and [A2] along with [M1], [M2], [M3], and [M4].
- The “zero vector” in this vector space is the zero matrix (with all entries zero), and [A3] and [A4] follow easily.
- Note of course that in some cases we can also multiply matrices by other matrices. However, the requirements for being a vector space don’t care that we can multiply matrices by other matrices! (All we need to be able to do is add them and multiply them by scalars.)

**Example:** The complex numbers are a real vector space under normal addition and multiplication.

- The axioms all follow from the standard properties of complex numbers: the “zero vector” is \( 0 = 0 + 0i \), and the additive inverse of \( a + bi \) is \( -a - bi \).
- Again, note that the complex numbers have “more structure” to them, because we can also multiply two complex numbers, and the multiplication is also commutative, associative, and distributive over addition. However, the requirements for being a vector space don’t care that the complex numbers have these additional properties.

**Example:** If \( F \) is any field and \( S \) is any set, the collection of all functions from \( S \) to \( F \) is a vector space over \( F \), where we define the “sum” of two functions as \( (f + g)(x) = f(x) + g(x) \) for every \( x \), and “scalar multiplication” as \( (\alpha \cdot f)(x) = \alpha f(x) \).

- To illustrate: if \( f(x) = x \) and \( g(x) = x^2 \), then \( f + g \) is the function with \( (f + g)(x) = x + x^2 \), and \( 2f \) is the function with \( (2f)(x) = 2x \).
○ The axioms follow from the properties of functions and the properties of the field \( F \): we simply verify that each axiom holds at every value \( x \) in \( S \). The “zero vector” in this space is the zero function; namely, the function \( Z \) which has \( Z(x) = 0 \) for every \( x \).

○ For example (just to demonstrate a few of the axioms), for any value \( x \) in \( S \) and any functions \( f \) and \( g \), we have

\[
\begin{align*}
\text{[A1]: } (f + g)(x) &= f(x) + g(x) = g(x) + f(x) = (g + f)(x). \\
\text{[M2]: } \alpha \cdot (f + g)(x) &= \alpha f(x) + \alpha g(x) = (\alpha f)(x) + (\alpha g)(x). \\
\text{[M4]: } (1 \cdot f)(x) &= f(x).
\end{align*}
\]

- **Example:** If \( F \) is any field, the set of polynomials \( P(F) \) with coefficients in \( F \) is a vector space over \( F \).

○ This follows in the same way as the verification for general functions.

- **Example:** The zero space with a single element \( 0 \), with \( 0 + 0 = 0 \) and \( \alpha \cdot 0 = 0 \) for every \( \alpha \), is a vector space.

○ All of the axioms in this case eventually boil down to \( 0 = 0 \).

○ This space is rather boring: since it only contains one element, there’s really not much to say about it.

- Purely for ease of notation, it will be useful to define subtraction:

- **Definition:** The difference of two vectors \( v, w \) in a vector space \( V \) is defined to be \( v - w = v + (-w) \).

○ The difference has the fundamental property we would expect: by axioms [A2] and [A3], we can write

\[
(v - w) + w = (v + (-w)) + w = v + ((-w) + w) = v + 0 = v.
\]

○ There are many simple algebraic properties that can be derived from the axioms which (therefore) hold in every vector space.

- **Theorem** (Basic Properties of Vector Spaces): In any vector space \( V \), the following are true:

  1. Addition has a cancellation law: for any vector \( v \), if \( a + v = b + v \) then \( a = b \).

     ○ **Proof:** By [A1]-[A4] we have \( (a + v) + (-v) = a + (v + (-v)) = a + 0 = a \).

     ○ Similarly we also have \( (b + v) + (-v) = b + (v + (-v)) = b + 0 = b \).

     ○ Finally, since \( a + v = b + v \) then \( a = (a + v) + (-v) = (b + v) + (-v) = b \) so \( a = b \).

  2. The zero vector is unique: if \( a + v = v \) for some vector \( v \), then \( a = 0 \).

     ○ **Proof:** By [A3], \( v = 0 + v \), so we have \( a + v = 0 + v \). Then by property (1) we conclude \( a = 0 \).

  3. The additive inverse is unique: for any vector \( v \), if \( a + v = 0 \) then \( a = -v \).

     ○ **Proof:** By [A4], \( 0 = (-v) + v \), so \( a + v = (-v) + v \). Then by property (1) we conclude \( a = -v \).

  4. The scalar 0 times any vector gives the zero vector: \( 0 \cdot v = 0 \) for any vector \( v \).

     ○ **Proof:** By [M2] and [M4] we have \( v = 1 \cdot v = (0 + 1) \cdot v = 0 \cdot v + 1 \cdot v = 0 \cdot v + v \).

     ○ Thus, by [A3], we have \( 0 + v = 0 \cdot v + v \) so by property (1) we conclude \( 0 = 0 \cdot v \).

  5. Any scalar times the zero vector is the zero vector: \( \alpha \cdot 0 = 0 \) for any scalar \( \alpha \).

     ○ **Proof:** By [M1] and [M4] we have \( \alpha \cdot 0 = \alpha \cdot (0 + 0) = \alpha \cdot 0 + \alpha \cdot 0 \).

     ○ Thus, by [A3], we have \( 0 + \alpha \cdot 0 = \alpha \cdot (0 + 0) = \alpha \cdot 0 + \alpha \cdot 0 \), so by property (1) we conclude \( 0 = \alpha \cdot 0 \).

  6. The scalar \( -1 \) times any vector gives the additive inverse: \( (-1) \cdot v = -v \) for any vector \( v \).

     ○ **Proof:** By property (4) and [M2]-[M4] we have \( v + (-1) \cdot v = 1 \cdot v + (-1) \cdot v = (1 + (-1)) \cdot v = 0 \cdot v = 0 \).

     ○ But now by property (3), since \( v + (-1) \cdot v = 0 \), we see that \( (-1) \cdot v = -v \).

  7. The additive inverse of the additive inverse is the original vector: \( -(v) = v \) for any vector \( v \).

     ○ **Proof:** By property (5) twice and [M3]-[M4], \( -(v) = (-1) \cdot (v) = (-1) \cdot ( -1) \cdot v = (-1)^2 \cdot v = 1 \cdot v = v \).

  8. The only scalar multiples equal to the zero vector are the trivial ones: if \( \alpha \cdot v = 0 \), then either \( \alpha = 0 \) or \( v = 0 \).
1.2 Subspaces

- **Definition:** A **subspace** \( W \) of a vector space \( V \) is a subset of the vector space \( V \) which, under the same addition and scalar multiplication operations as \( V \), is itself a vector space.

- Any vector space automatically has two subspaces: the entire space \( V \), and the “trivial” subspace consisting only of the zero vector.

- These examples are rather uninteresting, since we already know \( V \) is a vector space, and the subspace consisting only of the zero vector has very little structure.

- **Example:** Show that the set of diagonal \( 2 \times 2 \) matrices is a subspace of the vector space of all \( 2 \times 2 \) matrices.

  - To do this directly from the definition, we need to verify that all of the vector space axioms hold for the matrices of the form \[
  \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix}
\]
  for some \( a, b \).

  - First we need to check that the addition operation and scalar multiplication operations actually make sense; we see that \[
  \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix} + \begin{bmatrix}
  c & 0 \\
  0 & d
  \end{bmatrix} = \begin{bmatrix}
  a+c & 0 \\
  0 & b+d
  \end{bmatrix}
\] is also a diagonal matrix, and \[
  p \cdot \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix} = \begin{bmatrix}
  pa & 0 \\
  0 & pb
  \end{bmatrix}
\] is a diagonal matrix too, so the sum and scalar multiplication operations are well-defined.

  - Then we have to check the axioms, which is rather tedious. Here are some of the verifications:

    - **[A1]** Addition is commutative: \[
  \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix} + \begin{bmatrix}
  c & 0 \\
  0 & d
  \end{bmatrix} = \begin{bmatrix}
  c & 0 \\
  0 & d
  \end{bmatrix} + \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix}.
\]

    - **[A3]** The zero element is the zero matrix, since \[
  \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix} + \begin{bmatrix}
  0 & 0 \\
  0 & 0
  \end{bmatrix} = \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix}.
\]

    - **[A4]** The additive inverse of \[
  \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix}
\] is \[
  \begin{bmatrix}
  -a & 0 \\
  0 & -b
  \end{bmatrix}
\] since \[
  \begin{bmatrix}
  a & 0 \\
  0 & b
  \end{bmatrix} + \begin{bmatrix}
  -a & 0 \\
  0 & -b
  \end{bmatrix} = \begin{bmatrix}
  0 & 0 \\
  0 & 0
  \end{bmatrix}.
\]


- [M1] Scalar multiplication is consistent with regular multiplication: \( p \cdot q \cdot \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} = \begin{bmatrix} pqa & 0 \\ 0 & pqb \end{bmatrix} \).

- It is very time-consuming to verify all of the axioms for a subspace, and much of the work seems to be redundant. It would be convenient if we could clean up the repetitive nature of the verifications.

- **Theorem (Subspace Criterion):** A subset \( W \) of a vector space \( V \) is a subspace of \( V \) if and only if \( W \) has the following three properties:

  - [S1] \( W \) contains the zero vector of \( V \).
  - [S2] \( W \) is closed under addition: for any \( w_1 \) and \( w_2 \) in \( W \), the vector \( w_1 + w_2 \) is also in \( W \).
  - [S3] \( W \) is closed under scalar multiplication: for any scalar \( \alpha \) and \( w \) in \( W \), the vector \( \alpha \cdot w \) is also in \( W \).

  - **Proof:** Each of these conditions is necessary for \( W \) to be a subspace: the definition of binary operation requires [S2] and [S3] to hold, because when we add or scalar-multiply elements of \( W \), we must obtain a result that is in \( W \). For [S1], \( W \) must contain a zero vector \( 0_W \), and then we can write \( 0_V = 0_V + 0_W = 0_W \), so \( W \) contains the zero vector of \( V \).

  - Now suppose each of [S1]-[S3] holds for \( W \). Since all of the operations are therefore defined, axioms [A1]-[A2] and [M1]-[M4] will hold in \( W \) because they hold in \( V \). Axiom [A3] for \( W \) follows from [S1] since \( 0_W = 0_V \). Finally, for [A4], for any vector \( w \) in \( W \), by our basic properties we know that \( (-1) \cdot w = -w \), so since \( (-1) \cdot w \) is in \( W \) by [S3], we see that \(-w\) is in \( W \).

- Very often, if we want to check that something is a vector space, it is often much easier to verify that it is a subspace of something else we already know is a vector space, which is easily done using the subspace criterion. In order to show that a subset is *not* a subspace, it is sufficient to find a single example in which any one of the three criteria fails.

- **Example:** Determine whether the set of vectors of the form \( \langle t, t, t \rangle \) forms a subspace of \( \mathbb{R}^3 \).

  - We check the parts of the subspace criterion.
    - [S1]: The zero vector is of this form: take \( t = 0 \).
    - [S2]: We have \( \langle t_1, t_1, t_1 \rangle + \langle t_2, t_2, t_2 \rangle = \langle t_1 + t_2, t_1 + t_2, t_1 + t_2 \rangle \), which is again of the same form if we take \( t = t_1 + t_2 \).
    - [S3]: We have \( \alpha \cdot \langle t_1, t_1, t_1 \rangle = \langle \alpha t_1, \alpha t_1, \alpha t_1 \rangle \), which is again of the same form if we take \( t = \alpha t_1 \).
    - All three parts are satisfied, so this subset is a subspace.

- **Example:** Determine whether the set of \( n \times n \) matrices of trace zero is a subspace of the space of all \( n \times n \) matrices.

  - [S1]: The zero matrix has trace zero.
  - [S2]: Since \( \text{tr}(A + B) = \text{tr}(A) + \text{tr}(B) \), we see that if \( A \) and \( B \) have trace zero then so does \( A + B \).
  - [S3]: Since \( \text{tr}(\alpha A) = \alpha \text{tr}(A) \), we see that if \( A \) has trace zero then so does \( \alpha A \).
  - All three parts are satisfied, so this subset is a subspace.

- **Example:** Determine whether the set of vectors of the form \( \langle t, t^2 \rangle \) forms a subspace of \( \mathbb{R}^2 \).

  - We try checking the parts of the subspace criterion.
    - [S1]: The zero vector is of this form: take \( t = 0 \).
    - [S2]: For this criterion we try to write \( \langle t_1, t_1^2 \rangle + \langle t_2, t_2^2 \rangle = \langle t_1 + t_2, t_1^2 + t_2^2 \rangle \), but this does not have the correct form, because in general \( t_1^2 + t_2^2 \neq (t_1 + t_2)^2 \). (These quantities are only equal if \( 2t_1t_2 = 0 \).)
    - From here we can find a specific counterexample: the vectors \( \langle 1, 1 \rangle \) and \( \langle 2, 4 \rangle \) are in the subset, but their sum \( \langle 3, 5 \rangle \) is not. Thus, this subset is not a subspace.

5
○ Note that all we actually needed to do here was find a single counterexample, of which there are many. Had we noticed earlier that \((1, 1)\) and \((2, 4)\) were in the subset but their sum \((3, 5)\) was not, that would have been sufficient to conclude that the given set was not a subspace.

• **Example:** Determine whether the set of vectors of the form \((x, y, z)\) where \(x, y, z \geq 0\) forms a subspace of \(\mathbb{R}^3\).
  
  ○ It is **not a subspace** the vector \((1, 1, 1)\) is in the subset, but the scalar multiple \(-1 \cdot (1, 1, 1) = (-1, -1, -1)\) is not.

• There are a few more general subspaces that serve as important examples.

• **Proposition:** For any interval \([a, b]\), the collection of continuous functions on \([a, b]\) is a subspace of the space of all functions on \([a, b]\), as is the set of \(n\)-times differentiable functions on \([a, b]\).

  ○ **Proof:** We show each of these sets is a subspace of the collection of all (real-valued) functions on the interval \([a, b]\), which we already know is a vector space.

  ○ For the first statement, observe that the zero function is continuous, that the sum of two continuous functions is continuous, and that any scalar multiple of a continuous function is continuous.

  ○ The second statement follows in the same way; the zero function is also \(n\)-times differentiable, as is the sum of two \(n\)-times differentiable functions and any scalar multiple of an \(n\)-times differentiable function.

• **Example:** Show that the solutions to the (homogeneous, linear) differential equation \(y'' + 6y' + 5y = 0\) form a vector space.

  ○ We show this by verifying that the solutions form a subspace of the space of real-valued functions.

  ○ [S1]: The zero function is a solution.

  ○ [S2]: If \(y_1\) and \(y_2\) are solutions, then \(y_1'' + 6y_1' + 5y_1 = 0\) and \(y_2'' + 6y_2' + 5y_2 = 0\), so adding and using properties of derivatives shows that \((y_1 + y_2)'' + 6(y_1 + y_2)' + 5(y_1 + y_2) = 0\), so \(y_1 + y_2\) is also a solution.

  ○ [S3]: If \(\alpha\) is a scalar and \(y_1\) is a solution, then scaling \(y_1'' + 6y_1' + 5y_1 = 0\) by \(\alpha\) and using properties of derivatives shows that \((\alpha y_1)'' + 6(\alpha y_1)' + 5(\alpha y_1) = 0\), so \(\alpha y_1\) is also a solution.

  ○ Note that we did not need to know how to solve the differential equation to answer the question. (For completeness, the solutions are \(y = Ae^{-x} + Be^{-5x}\) for arbitrary constants \(A\) and \(B\).)

• We can use the subspace criterion to give easier proofs of a number of results about subspaces, such as the following.

• **Proposition (Intersection of Subspaces):** If \(V\) is a vector space, the intersection of any collection of subspaces of \(V\) is also a subspace of \(V\).

  ○ **Remark:** Unlike the intersection of subspaces, the union of two subspaces will not generally be a subspace.

  ○ **Proof:** Let \(S\) be a collection of subspaces of \(V\) and take \(I = \bigcap_{W \in S} W\) to be the intersection of the subspaces in \(S\). By the subspace criterion, the zero vector of \(V\) is in each subspace in \(S\), so it also is contained in \(I\).

  ○ Now let \(w_1\) and \(w_2\) be any vectors in \(I\), and \(\alpha\) be any scalar. By the definition of \(I\), the vectors \(w_1\) and \(w_2\) are in each subspace \(W\) in \(S\).

  ○ So by the subspace criterion, \(w_1 + w_2\) and \(\alpha \cdot w_1\) are also in each subspace \(W\) in \(S\); but this means both \(w_1 + w_2\) and \(\alpha \cdot w_1\) are in \(I\).

  ○ Thus, \(I\) satisfies each component of the subspace criterion, so it is a subspace of \(V\).

• One thing we would like to know, now that we have the definition of a vector space and a subspace, is what else we can say about elements of a vector space: that is, we would like to know what kind of structure the elements of a vector space have.

  ○ In some of the earlier examples we saw that, in \(\mathbb{R}^n\) and a few other vector spaces, subspaces could all be written down in terms of one or more parameters. We will develop this idea in the next few sections.
### 1.3 Linear Combinations and Span

- **Definition:** Given a set \( v_1, v_2, \ldots, v_n \) of vectors in a vector space \( V \), we say a vector \( w \) in \( V \) is a **linear combination** of \( v_1, v_2, \ldots, v_n \) if there exist scalars \( a_1, \ldots, a_n \) such that \( w = a_1 \cdot v_1 + a_2 \cdot v_2 + \cdots + a_n \cdot v_n \).

- **Example:** In \( \mathbb{R}^2 \), the vector \( (1, 1) \) is a linear combination of \( (1, 0) \) and \( (0, 1) \), because \( (1, 1) = 1 \cdot (1, 0) + 1 \cdot (0, 1) \).

- **Example:** In \( \mathbb{R}^4 \), the vector \( (4, 0, 5, 9) \) is a linear combination of \( (1, 0, 0, 1), (0, 1, 0, 0), \) and \( (1, 1, 1, 2) \), because \( (4, 0, 5, 9) = 1 \cdot (1, -1, 2, 3) - 2 \cdot (0, 1, 0, 0) + 3 \cdot (1, 1, 1, 2) \).

- **Non-Example:** In \( \mathbb{R}^3 \), the vector \( (0, 0, 1) \) is not a linear combination of \( (1, 1, 0) \) and \( (0, 1, 1) \) because there exist no scalars \( a_1 \) and \( a_2 \) for which \( a_1 \cdot (1, 1, 0) + a_2 \cdot (0, 1, 1) = (0, 0, 1) \): this would require a common solution to the three equations \( a_1 = 0, a_1 + a_2 = 0, \) and \( a_2 = 1, \) and this system has no solution.

- **Definition:** We define the **span** of a finite set of vectors \( \{v_1, v_2, \ldots, v_n\} \) in \( V \), denoted \( \text{span}(v_1, v_2, \ldots, v_n) \), to be the set of all vectors which are linear combinations of \( v_1, v_2, \ldots, v_n \). In other words, the span is the set of vectors of the form \( a_1 \cdot v_1 + \cdots + a_n \cdot v_n \), for scalars \( a_1, \ldots, a_n \). For an infinite collection of vectors, we define the span to be the set of all linear combinations of finitely many of the vectors.

- **Note:** For technical reasons, we define the span of the empty set to be the zero vector.

- **Example:** The span of the vectors \( (1, 0, 0) \) and \( (0, 1, 0) \) in \( \mathbb{R}^3 \) is the set of vectors of the form \( a \cdot (1, 0, 0) + b \cdot (0, 1, 0) = (a, b, 0) \). Equivalently, the span of these vectors is the set of vectors whose \( z \)-coordinate is zero, which (geometrically) forms the plane \( z = 0 \).

- **Example:** The span of the polynomials \( \{1, x, x^2, x^3\} \) is the set of polynomials of degree at most 3.

- **Example:** Determine whether the vectors \( (2, 3, 3) \) and \( (4, -1, 3) \) are in \( \text{span}(v, w) \), where \( v = (1, -1, 2) \) and \( w = (2, 1, -1) \).

- For \( (2, 3, 3) \) we must determine whether it is possible to write \( (2, 3, 3) = a \cdot (1, -1, 2) + b \cdot (2, 1, -1) \) for some \( a \) and \( b \).

- Equivalently, we want to solve the system \( 2 = a + 2b, 3 = -a + b, 3 = 2a - b \).

- Adding the first two equations yields \( 5 = 3b \) so that \( b = 5/3 \). The second equation then yields \( a = -4/3 \). However, this does not satisfy the third equation. So there are no such \( a \) and \( b \), meaning that \( (2, 3, 3) \) is not in the span.

- Similarly, for \( (4, -1, 3) \) we want to solve \( (4, -1, 3) = c \cdot (1, -1, 2) + d \cdot (2, 1, -1) \), or \( 4 = c + 2d, -1 = -c + d, 3 = 2c - d \).

- Using a similar procedure as above shows that \( d = 1, c = 2 \) is a solution: thus, we have \( (4, -1, 3) = 2 \cdot (1, -1, 2) + 1 \cdot (2, 1, -1) \), meaning that \( (4, -1, 3) \) is in the span.

- **Proposition** (Span is a Subspace): For any set \( S \) of vectors in \( V \), the set \( \text{span}(S) \) is a subspace of \( V \).

- **Proof:** We check the subspace criterion. If \( S \) is empty, then by definition \( \text{span}(S) = \{0\} \) and \( \{0\} \) is a subspace of \( V \).

- Now assume \( S \) is not empty. Let \( v \) be any vector in \( S \): then \( 0 \cdot v = 0 \) is in \( \text{span}(S) \).

- The span is closed under addition because we can write the sum of any two linear combinations as another linear combination: \( (a_1 \cdot v_1 + \cdots + a_n \cdot v_n) + (b_1 \cdot v_1 + \cdots + b_n \cdot v_n) = (a_1 + b_1) \cdot v_1 + \cdots + (a_n + b_n) \cdot v_n \).

- Finally, we can write any scalar multiple of a linear combination as a linear combination: \( \alpha \cdot (a_1 v_1 + \cdots + a_n v_n) = (\alpha a_1) \cdot v_1 + \cdots + (\alpha a_n) \cdot v_n \).

- **Proposition:** For any vectors \( v_1, \ldots, v_n \) in \( V \), if \( W \) is any subspace of \( V \) that contains \( v_1, \ldots, v_n \), then \( W \) contains \( \text{span}(v_1, \ldots, v_n) \). In other words, the span is the smallest subspace containing the vectors \( v_1, \ldots, v_n \).
\( \text{Proof:} \) Consider any element \( w \) in \( \text{span}(v_1, v_2, \ldots, v_n) \): by definition, we can write \( w = a_1 v_1 + \cdots + a_n v_n \) for some scalars \( a_1, \ldots, a_n \).

- Because \( W \) is a subspace, it is closed under scalar multiplication, so each of \( a_1 v_1, \ldots, a_n v_n \) lies in \( W \).
- Furthermore, also because \( W \) is a subspace, it is closed under addition. Thus, the sum \( a_1 v_1 + \cdots + a_n v_n \) lies in \( W \).
- Thus, every element of the span lies in \( W \), as claimed.

- **Corollary:** If \( S \) and \( T \) are two sets of vectors in \( V \) with \( S \subseteq T \), then \( \text{span}(S) \) is a subspace of \( \text{span}(T) \).

  - **Proof:** Since the span is always a subspace, we know that \( \text{span}(T) \) is a subspace of \( V \) containing all the vectors in \( S \). By the previous proposition, \( \text{span}(T) \) therefore contains every linear combination of vectors from \( S \), which is to say, \( \text{span}(T) \) contains \( \text{span}(S) \).

- Here are a pair of results involving how span interacts with adjoining a new vector to a set:

  - **Proposition:** If \( S \) is any set of vectors in \( V \) and \( T = S \cup \{ w \} \) for some vector \( w \) in \( V \), then \( \text{span}(T) = \text{span}(S) \) if and only if \( w \) is in \( \text{span}(S) \).

    - **Proof:** By the previous corollary, since \( T \) contains \( S \), \( \text{span}(T) \) contains \( \text{span}(S) \).
    - If \( \text{span}(T) = \text{span}(S) \), then since \( w \) is in \( T \) (hence in \( \text{span}(T) \)) we conclude \( w \) is in \( \text{span}(S) \).
    - Conversely, if \( w \) is in \( \text{span}(S) \), then we can eliminate \( w \) from any linear combination of vectors in \( T \) to obtain a linear combination of vectors only in \( S \).
    - Explicitly: suppose \( w = a_1 v_1 + \cdots + a_n v_n \) where \( v_1, \ldots, v_n \) are in \( S \). Then any \( x \) in \( \text{span}(T) \) is some linear combination \( x = c_1 w + b_1 v_1 + \cdots + b_n v_n + b_{n+1} v_{n+1} + \cdots + b_m v_m \) for some \( v_1, \ldots, v_m \) in \( S \).
    - But then \( x = (b_1 + c a_1) v_1 + \cdots + (b_n + c a_n) v_n + b_{n+1} v_{n+1} + \cdots + b_m v_m \) can be written as a linear combination only involving vectors in \( S \), so \( x \) is in \( \text{span}(S) \). Thus, \( \text{span}(S) = \text{span}(T) \).

  - **Proposition:** If \( S \) and \( T \) are any subsets of \( V \) with \( \text{span}(S) = \text{span}(T) \) and \( w \) is any vector in \( V \), then \( \text{span}(S \cup \{ w \}) = \text{span}(T \cup \{ w \}) \).

    - **Proof:** By hypothesis, every vector in \( S \) lies in \( \text{span}(S) = \text{span}(T) \), and since \( \text{span}(T) \) is contained in \( \text{span}(T \cup \{ w \}) \), every vector in \( S \) is contained in \( \text{span}(T \cup \{ w \}) \).
    - Since \( w \) is also in \( \text{span}(T \cup \{ w \}) \), \( S \cup \{ w \} \) is in \( \text{span}(T \cup \{ w \}) \). But now since \( \text{span}(T \cup \{ w \}) \) is a subspace of \( V \), it contains \( \text{span}(S \cup \{ w \}) \).
    - Therefore, \( \text{span}(S \cup \{ w \}) \subseteq \text{span}(T \cup \{ w \}) \). By the same argument with \( S \) and \( T \) interchanged, \( \text{span}(T \cup \{ w \}) \subseteq \text{span}(S \cup \{ w \}) \). Therefore, we must have equality: \( \text{span}(S \cup \{ w \}) = \text{span}(T \cup \{ w \}) \).

- Sets whose span is the entire space have a special name:

  - **Definition:** Given a set \( S \) of vectors in a vector space \( V \), if \( \text{span}(S) = V \) then we say that \( S \) is a **spanning set** (or generating set) for \( V \).

    - Spanning sets are very useful because they allow us to describe every vector in \( V \) in terms of the vectors in \( S \).
    - Explicitly, every vector in \( V \) is a linear combination of the vectors in \( S \), which is to say, every vector \( w \) in \( V \) can be written in the form \( w = a_1 v_1 + \cdots + a_n v_n \) for some scalars \( a_1, \ldots, a_n \) and some vectors \( v_1, v_2, \ldots, v_n \) in \( S \).

  - **Example:** Show that the matrices \[
  \begin{bmatrix}
  1 & 0 \\
  0 & -1
  \end{bmatrix}, \quad
  \begin{bmatrix}
  0 & 1 \\
  0 & 0
  \end{bmatrix}, \quad
  \begin{bmatrix}
  0 & 0 \\
  1 & 0
  \end{bmatrix}
  \] span the vector space of \( 2 \times 2 \) matrices of trace zero.

    - Recall that we showed earlier that the space of matrices of trace zero is a vector space (since it is a subspace of the vector space of all \( 2 \times 2 \) matrices).
    - A \( 2 \times 2 \) matrix \[
  \begin{bmatrix}
  a & b \\
  c & d
  \end{bmatrix}
  \] has trace zero when \( a + d = 0 \), or equivalently when \( d = -a \).
• So any matrix of trace zero has the form 
\[
\begin{bmatrix}
a & b \\
c & -a
\end{bmatrix} = a \begin{bmatrix}1 & 0 \\
0 & -1\end{bmatrix} + b \begin{bmatrix}0 & 1 \\
0 & 0\end{bmatrix} + c \begin{bmatrix}0 & 0 \\
1 & 0\end{bmatrix}.
\]

• Since any matrix of trace zero is therefore a linear combination of the matrices
\[
\begin{bmatrix}
1 & 0 \\
0 & -1
\end{bmatrix}, \begin{bmatrix}0 & 1 \\
0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0 \\
1 & 0\end{bmatrix},
\]
we conclude that they are a spanning set.

• Example: Show that the matrices
\[
\begin{bmatrix}1 & 0 \\
1 & -1
\end{bmatrix}, \begin{bmatrix}0 & 1 \\
0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0 \\
1 & 0\end{bmatrix}
\]
also span the vector space of 2 × 2 matrices of trace zero.

• We can write
\[
\begin{bmatrix}a & b \\
c & -a
\end{bmatrix} = a \begin{bmatrix}1 & 0 \\
0 & -1\end{bmatrix} + b \begin{bmatrix}0 & 1 \\
0 & 0\end{bmatrix} + (c - a) \begin{bmatrix}0 & 0 \\
1 & 0\end{bmatrix}.
\]

• This set of matrices is different from the spanning set in the previous example, which underlines an important point: any given vector space may have many different spanning sets.

• Example: Determine whether the polynomials 1, 1 + x^2, x^4, 1 + x^2 + x^4 span the space \(W\) of polynomials with complex coefficients having degree at most 4 and satisfying \(p(x) = p(-x)\).

• It is straightforward to verify that this set of polynomials is a subspace of the polynomials with complex coefficients.

• A polynomial of degree at most 4 has the form \(p(x) = a + bx + cx^2 + dx^3 + ex^4\), and having \(p(x) = p(-x)\) requires \(a - bx + cx^2 - dx^3 + ex^4 = a + bx + cx^2 + dx^3 + ex^4\), or equivalently \(b = d = 0\).

• Thus, the desired polynomials are those of the form \(p(x) = a + cx^2 + ex^4\) for arbitrary complex numbers \(a, c,\) and \(e\).

• Since we can write \(a + cx^2 + ex^4 = (a - c) \cdot 1 + c \cdot (1 + x^2) + e \cdot x^4 + 0 \cdot (1 + x^2 + x^4)\), the given polynomials do span \(W\).

• Note that we could also have written \(a + cx^2 + ex^4 = (a - c) \cdot 1 + (c - e) \cdot (1 + x^2) + 0 \cdot x^4 + e \cdot (1 + x^2 + x^4)\), so the polynomials in \(W\) can be written as a linear combination of the vectors in the spanning set in more than one way. (In fact, they can be written as a linear combination in infinitely many ways.)

• This example underlines another important point: if \(\text{span}(S) = V\), it is possible that any given vector in \(V\) can be written as a linear combination of vectors in \(S\) in many different ways.

• Example: Determine whether the vectors \(\langle 1, 2 \rangle, \langle 2, 4 \rangle, \langle 3, 1 \rangle\) span \(\mathbb{R}^2\).

• For any vector \(\langle p, q \rangle\), we want to determine whether there exist some scalars \(a, b, c\) such that \(\langle p, q \rangle = a \cdot \langle 1, 2 \rangle + b \cdot \langle 2, 4 \rangle + c \cdot \langle 3, 1 \rangle\).

• Equivalently, we want to check whether the system \(p = a + 2b + 3c, q = 2a + 4b + c\) has solutions for any \(p, q\).

• Row-reducing the associated coefficient matrix gives
\[
\begin{bmatrix}
1 & 2 & 3 \\
2 & 4 & 1
\end{bmatrix} \begin{bmatrix}
p \\
q
\end{bmatrix} \rightarrow \begin{bmatrix}
1 & 2 & 3 \\
0 & 0 & -5
\end{bmatrix} \begin{bmatrix}
p \\
q - 2p
\end{bmatrix}
\]
and since this system is non-contradictory, there is always a solution: indeed, there are infinitely many. (One solution is \(c = \frac{2}{5} p - \frac{1}{5} q, b = 0, a = -\frac{1}{5} p + \frac{3}{5} q\).)

• Since there is always a solution for any \(p, q\), we conclude that these vectors do span \(\mathbb{R}^2\).

• Example: Determine whether the vectors \(\langle 1, -1, 3 \rangle, \langle 2, 2, -1 \rangle, \langle 3, 4, 7 \rangle\) span \(\mathbb{R}^3\).

• For any vector \(\langle p, q, r \rangle\), we want to determine whether there exist some scalars \(a, b, c\) such that \(\langle p, q, r \rangle = a \cdot \langle 1, -1, 3 \rangle + b \cdot \langle 2, 2, -1 \rangle + c \cdot \langle 3, 1, 2 \rangle\).
Row-reducing the associated coefficient matrix gives

\[
\begin{bmatrix}
1 & 1 & -1 & p \\
-1 & 0 & 2 & q \\
3 & 1 & -5 & r
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 1 & -1 & p \\
0 & 1 & 1 & q+p \\
0 & -2 & -2 & r - 3p
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 1 & -1 & p \\
0 & 1 & 1 & q+p \\
0 & 0 & 1 & r + 2q - p
\end{bmatrix}.
\]

Now, if \( r + 2q - p \neq 0 \), the last row will be a contradictory equation. This can certainly occur: for example, we could take \( r = 1 \) and \( p = q = 0 \).

Since there is no way to write an arbitrary vector in \( \mathbb{R}^3 \) as a linear combination of the given vectors, we conclude that these vectors do not span \( \mathbb{R}^3 \).

- Ultimately, the example above requires knowing when a particular set of linear equations has a solution.
- We will postpone further discussion of this question about finding spanning sets for \( \mathbb{R}^n \) (and other vector spaces) until we have the necessary background to solve general systems of linear equations.

### 1.4 Linear Independence and Linear Dependence

- **Definition:** We say a finite set of vectors \( v_1, \ldots, v_n \) is linearly independent if \( a_1 \cdot v_1 + \cdots + a_n \cdot v_n = 0 \) implies \( a_1 = \cdots = a_n = 0 \). Otherwise, we say the collection is linearly dependent. (The empty set of vectors is by definition linearly independent.)

  - In other words, \( v_1, \ldots, v_n \) are linearly independent precisely when the only way to form the zero vector as a linear combination of \( v_1, \ldots, v_n \) is to have all the scalars equal to zero (the “trivial” linear combination).

  - If there is a non trivial linear combination giving the zero vector, then \( v_1, \ldots, v_n \) are linearly dependent.

- **Note:** For an infinite set of vectors, we say it is linearly independent if every finite subset is linearly independent, per the definition above. Otherwise, if some finite subset displays a dependence, we say the infinite set is dependent.

- **Example:** The matrices
  \[
  \begin{bmatrix}
  2 & 3 \\
  -1 & -1 \\
  \end{bmatrix}, \quad \begin{bmatrix}
  0 & 3 \\
  -1 & 2 \\
  \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix}
  0 & 0 \\
  0 & 0 \\
  \end{bmatrix}
  \]
  are linearly dependent, because \( 3 \cdot \begin{bmatrix} 2 & 3 \\ 2 & -4 \end{bmatrix} = 6 \cdot \begin{bmatrix} -1 & -1 \\ -1 & 2 \end{bmatrix} + (-1) \cdot \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \).

- **Example:** Determine whether the vectors \( (1, 1, 0), (0, 2, 1) \) in \( \mathbb{R}^3 \) are linearly dependent or linearly independent.
  - Suppose that we had scalars \( a \) and \( b \) with \( a \cdot (1, 1, 0) + b \cdot (0, 2, 1) = (0, 0, 0) \).
  - Comparing the two sides requires \( a = 0, a + 2b = 0, b = 0 \), which has only the solution \( a = b = 0 \).
  - Thus, by definition, these vectors are linearly independent.

- **Example:** Determine whether the vectors \( (1, 1, 0), (2, 2, 0) \) in \( \mathbb{R}^3 \) are linearly dependent or linearly independent.
  - Suppose that we had scalars \( a \) and \( b \) with \( a \cdot (1, 1, 0) + b \cdot (2, 2, 0) = (0, 0, 0) \).
  - Comparing the two sides requires \( a + 2b = 0, a + 2b = 0, 0 = 0 \), which has (for example) the nontrivial solution \( a = 1, b = -2 \).
  - Thus, we see that we can write \( 2 \cdot (1, 1, 0) + (-1) \cdot (2, 2, 0) = (0, 0, 0) \), and this is a nontrivial linear combination giving the zero vector meaning that these vectors are linearly dependent.

- Here are a few basic properties of linear dependence and independence that follow from the definition or the proposition above:
  - Any set containing the zero vector is linearly dependent. (The zero vector is always a linear combination of the other vectors in the set.)
  - Any set containing a linearly dependent set is linearly dependent. (Any dependence in the smaller set gives a dependence in the larger set.)
In a vector space, the terminology of linear dependence arises from the fact that if a set of vectors is linearly dependent, one of the vectors is necessarily a linear combination of the others (i.e., it depends on the others): answering this question will reduce to determining whether a set of linear equations has a solution.

Suppose that we had scalars $a, b, c, d$ with $a \cdot \langle 1, 0, 2, 2 \rangle + b \cdot \langle 2, -2, 3, 0 \rangle + c \cdot \langle 0, 3, 1, 3 \rangle + d \cdot \langle 0, 4, 1, 2 \rangle = \langle 0, 0, 0, 0 \rangle$.

This is equivalent to saying $a + 2b = 0$, $-2b + 3c + 4d = 0$, $2a + 3b + c + d = 0$, and $2a + 3c + 2d = 0$.

To search for solutions we can convert this system into matrix form and then row-reduce it:

$$
\begin{bmatrix}
1 & 2 & 0 & 0 & 0 \\
0 & -2 & 3 & 4 & 0 \\
2 & 3 & 1 & 1 & 0 \\
2 & 0 & 3 & 2 & 0
\end{bmatrix}
\rightarrow
\begin{bmatrix}
r_{3-2r_1} \\
r_{4-2r_2}
\end{bmatrix}
\rightarrow
\cdots
\rightarrow
\begin{bmatrix}
1 & 0 & 0 & -2 & 0 \\
0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 2 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

from which we can obtain a nonzero solution $d = 1$, $c = -2$, $b = -1$, $a = 2$.

So we see $2 \cdot \langle 1, 0, 2, 2 \rangle + (-1) \cdot \langle 2, -2, 0, 3 \rangle + (-2) \cdot \langle 0, 3, 1, 3 \rangle + 1 \cdot \langle 0, 4, 2, 1 \rangle = \langle 0, 0, 0, 0 \rangle$. This is a nontrivial linear combination giving the zero vector, so these vectors are linearly dependent.

The terminology of “linear dependence” arises from the fact that if a set of vectors is linearly dependent, one of the vectors is necessarily a linear combination of the others (i.e., it “depends” on the others):

Proposition (Dependence and Linear Combinations): A set $S$ of vectors is linearly dependent if and only if one of the vectors is a linear combination of (some of) the others.

To avoid trivialities, we remark here that if $S = \{0\}$ then the result is still correct, since the set of linear combinations (i.e., the span) of the empty set is the zero vector.

Proof: If $v$ is a linear combination of other vectors in $S$, say $v = a_1 \cdot v_1 + a_2 \cdot v_2 + \cdots + a_n \cdot v_n$, then we have a nontrivial linear combination yielding the zero vector, namely $(-1) \cdot v + a_1 \cdot v_1 + \cdots + a_n \cdot v_n = 0$.

Conversely, suppose there is a nontrivial linear combination of vectors in $S$ giving the zero vector, say, $b_1 \cdot v_1 + b_2 \cdot v_2 + \cdots + b_n \cdot v_n = 0$. Since the linear combination is nontrivial, at least one of the coefficients is nonzero, say, $b_i$.

Then $b_i \cdot v_i = (-b_1) \cdot v_1 + \cdots + (-b_{i-1}) \cdot v_{i-1} + (-b_{i+1}) \cdot v_{i+1} + \cdots + (-b_n) \cdot v_n$, and by scalar-multiplying both sides by $\frac{1}{b_i}$ (which exists because $b_i \neq 0$ by assumption) we see $v_i = \frac{-b_1}{b_i} \cdot v_1 + \cdots + \frac{-b_{i-1}}{b_i} \cdot v_{i-1} + \frac{-b_{i+1}}{b_i} \cdot v_{i+1} + \cdots + \frac{-b_n}{b_i} \cdot v_n$.

Thus, one of the vectors is a linear combination of the others, as claimed.
Example: Write one of the linearly dependent vectors \( \langle 1, -1 \rangle, \langle 2, 2 \rangle, \langle 2, 1 \rangle \) as a linear combination of the others.

- If we search for a linear dependence, we require \( a \cdot \langle 1, -1 \rangle + b \cdot \langle 2, 2 \rangle + c \cdot \langle 2, 1 \rangle = \langle 0, 0 \rangle \).
- By row-reducing the appropriate matrix we can find the solution \( 2 \cdot \langle 1, -1 \rangle + 3 \cdot \langle 2, 2 \rangle - 4 \cdot \langle 2, 1 \rangle = \langle 0, 0 \rangle \).
- By rearranging we can then write \( \langle 1, -1 \rangle = -\frac{3}{2} \langle 2, 2 \rangle + 2 \cdot \langle 2, 1 \rangle \) (Of course, this is not the only possible answer: any of the vectors can be written as a linear combination of the other two, since all of the coefficients in the linear dependence are nonzero.)

Linear independence and span are related in a number of ways. Here are a few:

Theorem (Independence and Span): Let \( S \) be a linearly independent subset of the vector space \( V \), and \( v \) be any vector of \( V \) not in \( S \). Then the set \( S \cup \{ v \} \) is linearly dependent if and only if \( v \) is in \( \text{span}(S) \).

- Proof: If \( v \) is in \( \text{span}(S) \), then one vector (namely \( v \)) in \( S \cup \{ v \} \) can be written as a linear combination of the other vectors (namely, the vectors in \( S \)). So by our earlier proposition, \( S \cup \{ v \} \) is linearly dependent.
- Conversely, suppose that \( S \cup \{ v \} \) is linearly dependent, and consider a nontrivial dependence. If the coefficient of \( v \) is zero, then we would obtain a nontrivial dependence among the vectors in \( S \) (impossible, since \( S \) is linearly independent), so the coefficient of \( v \) is not zero: say, \( a \cdot v + b_1 \cdot v_1 + \cdots + b_n \cdot v_n = 0 \) with \( a \neq 0 \) and for some \( v_1, v_2, \ldots, v_n \) in \( S \).
  - Then \( v = \left( -\frac{b_1}{a} \right) \cdot v_1 + \cdots + \left( -\frac{b_n}{a} \right) \cdot v_n \) is a linear combination of the vectors in \( S \), so \( v \) is in \( \text{span}(S) \).

We can also characterize linear independence using the span:

Theorem (Characterization of Linear Independence): A set \( S \) of vectors is linearly independent if and only if every vector \( w \) in \( \text{span}(S) \) may be uniquely written as a sum \( w = a_1 \cdot v_1 + \cdots + a_n \cdot v_n \) for unique scalars \( a_1, a_2, \ldots, a_n \) and unique vectors \( v_1, v_2, \ldots, v_n \) in \( S \).

- Proof: First suppose the decomposition is always unique: then for any \( v_1, v_2, \ldots, v_n \) in \( S \), \( a_1 \cdot v_1 + \cdots + a_n \cdot v_n = 0 \) implies \( a_1 = \cdots = a_n = 0 \), because \( 0 \cdot v_1 + \cdots + 0 \cdot v_n = 0 \) is by assumption the only decomposition of \( 0 \). So we see that the vectors are linearly independent.
- Now suppose that we had two ways of decomposing a vector \( w \), say as \( w = a_1 \cdot v_1 + \cdots + a_n \cdot v_n \) and as \( w = b_1 \cdot v_1 + \cdots + b_n \cdot v_n \).
  - By subtracting, we obtain \( (a_1 - b_1) \cdot v_1 + \cdots + (a_n - b_n) \cdot v_n = w - w = 0 \).
  - But now because \( v_1, \ldots, v_n \) are linearly independent, we see that all of the scalar coefficients \( a_1 - b_1, \cdots, a_n - b_n \) are zero. But this says \( a_1 = b_1, a_2 = b_2, \ldots, a_n = b_n \), which is to say that the two decompositions are actually the same.

1.5 Bases and Dimension

- We will now combine the ideas of a spanning set and a linearly independent set, and use the resulting objects to study the structure of vector spaces.

1.5.1 Definition and Basic Properties of Bases

- Definition: A linearly independent set of vectors which spans \( V \) is called a basis for \( V \).
  - Terminology Note: The plural form of the (singular) word “basis” is “bases”.
- Example: Show that the vectors \( \langle 1, 0, 0 \rangle, \langle 0, 1, 0 \rangle, \langle 0, 0, 1 \rangle \) form a basis for \( \mathbb{R}^3 \).
  - The vectors certainly span \( \mathbb{R}^3 \), since we can write any vector \( \langle a, b, c \rangle = a \cdot \langle 1, 0, 0 \rangle + b \cdot \langle 0, 1, 0 \rangle + c \cdot \langle 0, 0, 1 \rangle \) as a linear combination of these vectors.
Further, the vectors are linearly independent, because $a \cdot \langle 1,0,0 \rangle + b \cdot \langle 0,1,0 \rangle + c \cdot \langle 0,0,1 \rangle = \langle a, b, c \rangle$
is the zero vector only when $a = b = c = 0$.

Thus, these three vectors are a linearly independent spanning set for $\mathbb{R}^3$, so they form a basis.

- A particular vector space can have several different bases:

- **Example:** Show that the vectors $\langle 1,1,1 \rangle$, $\langle 2,-1,1 \rangle$, $\langle 1,2,1 \rangle$ also form a basis for $\mathbb{R}^3$.

  - Solving the system of linear equations determined by $x \cdot \langle 1,1,1 \rangle + y \cdot \langle 2,-1,1 \rangle + z \cdot \langle 1,2,1 \rangle = \langle a, b, c \rangle$ for $x, y, z$ will yield the solution $x = -3a - b + 5c$, $y = a - c$, $z = 2a + b - 3c$.
  
  - Therefore, $\langle a, b, c \rangle = (-3a - b + 5c) \cdot \langle 1,1,1 \rangle + (a - c) \cdot \langle 2,-1,1 \rangle + (2a + b - 3c) \cdot \langle 1,2,1 \rangle$, so these three vectors span $\mathbb{R}^3$.
  
  - Furthermore, solving the system $x \cdot \langle 1,1,1 \rangle + y \cdot \langle 2,-1,1 \rangle + z \cdot \langle 1,2,1 \rangle = \langle 0,0,0 \rangle$ yields only the solution $x = y = z = 0$, so these three vectors are also linearly independent.
  
  - So $\langle 1,1,1 \rangle$, $\langle 2,-1,1 \rangle$, $\langle 1,2,1 \rangle$ are a linearly independent spanning set for $\mathbb{R}^3$, meaning that they form a basis.

- **Example:** Find a basis for the vector space of $2 \times 3$ (real) matrices.

  - A general $2 \times 3$ matrix has the form $\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} = a \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} + e \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + f \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

  - This decomposition suggests that we can take the set of six matrices $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ as a basis.

  - Indeed, they certainly span the space of all $2 \times 3$ matrices, and they are also linearly independent, because the only linear combination giving the zero matrix is the one with $a = b = c = d = e = f = 0$.

- **Non-Example:** Show that the vectors $\langle 1,1,0 \rangle$ and $\langle 1,1,1 \rangle$ are not a basis for $\mathbb{R}^3$.

  - These vectors are linearly independent, since neither is a scalar multiple of the other.

  - However, they do not span $\mathbb{R}^3$ since, for example, it is not possible to obtain the vector $\langle 1,0,0 \rangle$ as a linear combination of $\langle 1,1,0 \rangle$ and $\langle 1,1,1 \rangle$.

  - Explicitly, since $a \cdot \langle 1,1,0 \rangle + b \cdot \langle 1,1,1 \rangle = \langle a + b, a + b, b \rangle$, there are no possible $a, b$ for which this vector can equal $\langle 1,0,0 \rangle$, since this would require $a + b = 1$ and $a + b = 0$ simultaneously.

- **Non-Example:** Show that the vectors $\langle 1,0,0 \rangle$, $\langle 0,1,0 \rangle$, $\langle 0,0,1 \rangle$, $\langle 1,1,1 \rangle$ are not a basis for $\mathbb{R}^3$.

  - These vectors do span $V$, since we can write any vector $\langle a, b, c \rangle = a \cdot \langle 1,0,0 \rangle + b \cdot \langle 0,1,0 \rangle + c \cdot \langle 0,0,1 \rangle + 0 \cdot \langle 1,1,1 \rangle$.

  - However, these vectors are not linearly independent, since we have the explicit linear dependence $1 \cdot \langle 1,0,0 \rangle + 1 \cdot \langle 0,1,0 \rangle + 1 \cdot \langle 0,0,1 \rangle + (-1) \cdot \langle 1,1,1 \rangle = \langle 0,0,0 \rangle$.

- Having a basis allows us to describe all the elements of a vector space in a particularly convenient way:

- **Proposition (Characterization of Bases):** The set of vectors $v_1, v_2, \ldots, v_n$ forms a basis of the vector space $V$ if and only if every vector $w$ in $V$ can be written in the form $w = a_1 \cdot v_1 + a_2 \cdot v_2 + \cdots + a_n \cdot v_n$ for *unique* scalars $a_1, a_2, \ldots, a_n$.

  - In particular, this proposition says that if we have a basis $v_1, v_2, \ldots, v_n$ for $V$, then we can describe all of the other vectors in $V$ in a particularly simple way (as a linear combination of $v_1, v_2, \ldots, v_n$) that is *unique*. A useful way to interpret this idea is to think of the basis vectors $v_1, v_2, \ldots, v_n$ as “coordinate directions” and the coefficients $a_1, a_2, \ldots, a_n$ as “coordinates”.

13
○ **Proof:** Suppose $v_1, v_2, \ldots, v_n$ is a basis of $V$. Then by definition, the vectors $v_1, v_2, \ldots, v_n$ span the vector space $V$: every vector $w$ in $V$ can be written in the form $w = a_1 \cdot v_1 + a_2 \cdot v_2 + \cdots + a_n \cdot v_n$ for some scalars $a_1, a_2, \ldots, a_n$.

○ Furthermore, since the vectors $v_1, v_2, \ldots, v_n$ are linearly independent, by our earlier proposition every vector $w$ in their span (which is to say, every vector in $V$) can be uniquely written in the form $w = a_1 \cdot v_1 + a_2 \cdot v_2 + \cdots + a_n \cdot v_n$, as claimed.

○ Conversely, suppose every vector $w$ in $V$ can be uniquely written in the form $w = a_1 \cdot v_1 + a_2 \cdot v_2 + \cdots + a_n \cdot v_n$. Then by definition, the vectors $v_1, v_2, \ldots, v_n$ span $V$.

○ Furthermore, by our earlier proposition, because every vector in span$(v_1, v_2, \ldots, v_n)$ can be uniquely written as a linear combination of $v_1, v_2, \ldots, v_n$, the vectors $v_1, v_2, \ldots, v_n$ are linearly independent: thus, they are a linearly independent spanning set for $V$, so they form a basis.

- If we have a general description of the elements of a vector space, we can often extract a basis by direct analysis.

- **Example:** Find a basis for the space $W$ of real polynomials $p(x)$ of degree $\leq 3$ such that $p(1) = 0$.

  ○ Notice that $W$ is a subspace of the vector space $V$ of all polynomials with real coefficients, as it satisfies the subspace criterion. (We omit the verification.)

  ○ A polynomial of degree $\leq 3$ has the form $p(x) = ax^3 + bx^2 + cx + d$ for constants $a, b, c, d$.

  ○ Since $p(1) = a + b + c + d$, the condition $p(1) = 0$ gives $a + b + c + d = 0$, so $d = -a - b - c$.

  ○ Thus, we can write $p(x) = ax^3 + bx^2 + cx + (-a - b - c) = a(x^3 - 1) + b(x^2 - 1) + c(x - 1)$, and conversely, any such polynomial has $p(1) = 0$.

  ○ Since every polynomial in $W$ can be uniquely written as $a(x^3 - 1) + b(x^2 - 1) + c(x - 1)$, we conclude that the set $\{x^3 - 1, x^2 - 1, x - 1\}$ is a basis of $W$.

### 1.5.2 Existence and Construction of Bases

- **A basis for a vector space can be obtained from a spanning set:**

  - **Theorem** (Spanning Sets and Bases): If $V$ is a vector space, then any spanning set for $V$ contains a basis of $V$.

    ○ In the event that the spanning set is infinite, the argument is rather delicate and technical (and requires an ingredient known as the axiom of choice), so we will only treat the case of a finite spanning set consisting of the vectors $v_1, v_2, \ldots, v_n$.

    ○ **Proof** (finite spanning set case): Suppose $\{v_1, \ldots, v_n\}$ spans $V$. We construct an explicit subset that is a basis for $V$.

    ○ Start with an empty collection $S_0$ of elements. Now, for each $1 \leq k \leq n$, perform the following procedure:

      * Check whether the vector $v_k$ is contained in the span of $S$. (Note that the span of the empty set is the zero vector.)
      * If $v_k$ is not in the span of $S_{k-1}$, let $S_k = S_{k-1} \cup \{v_k\}$. Otherwise, let $S_k = S_{k-1}$.

    ○ We claim that the set $S_n$ is a basis for $V$. Roughly speaking, the idea is that the collection of elements which we have not thrown away will still be a generating set (since removing a dependent element will not change the span), but the collection will also now be linearly independent (since we threw away elements which were dependent).

    ○ To show that $S_n$ is linearly independent, we use induction on $k$ to show that $S_k$ is linearly independent for each $1 \leq k \leq n$.

      * For the base case we take $k = 0$: clearly, $S_0$ (the empty set) is linearly independent.
      * For the inductive step, suppose $k \geq 1$ and that $S_{k-1}$ is linearly independent.
      * If $v_k$ is in the span of $S_{k-1}$, then $S_k = S_{k-1}$ is linearly independent.
* If \(v_k\) is not in the span of \(S_{k-1}\), then \(S_k = S_{k-1} \cup \{v_k\}\) is linearly independent by our proposition about span and linear independence.
* In both cases, \(S_k\) is linearly independent, so by induction, \(S_n\) is linearly independent.

○ To show that \(S_n\) spans \(V\), let \(T_k = \{v_1, \ldots, v_k\}\). We use induction on \(k\) to show that \(\text{span}(S_k) = \text{span}(T_k)\) for each \(0 \leq k \leq n\).
  * For the base case we take \(k = 0\): clearly, \(\text{span}(S_0) = \text{span}(T_0)\), since both \(S_0\) and \(T_0\) are empty.
  * For the inductive step, suppose \(k \geq 1\) and that \(\text{span}(S_{k-1}) = \text{span}(T_{k-1})\).

○ Proof: Let \(S\) be any spanning set for \(V\). (For example, we could take \(S\) to be the set of all vectors in \(V\).) Then since \(S\) spans \(V\), it contains a basis for \(V\).

○ Remark: That a basis always exists is incredibly helpful, and is without a doubt the most useful fact about vector spaces. Vector spaces in the abstract are very hard to think about, but a vector space with a basis is something very concrete, since the existence of a basis allows us to describe all the vectors in a precise and regular form.

• The above procedure allows us to construct a basis for a vector space by “dropping down” by removing linearly dependent vectors from a spanning set. We can also construct bases for vector spaces by “building up” from a linearly independent set.

• Theorem (Replacement Theorem): Suppose that \(S = \{v_1, v_2, \ldots, v_n\}\) is a basis for \(V\) and \(\{w_1, w_2, \ldots, w_m\}\) is a linearly independent subset of \(V\). Then there is a reordering of the basis \(S\), say \(\{a_1, a_2, \ldots, a_n\}\) such that for each \(1 \leq k \leq m\), the set \(\{w_1, w_2, \ldots, w_k, a_{k+1}, a_{k+2}, \ldots, a_n\}\) is a basis for \(V\). Equivalently, the elements \(\{w_1, w_2, \ldots, w_m\}\) can be used to successively replace the elements of the basis, with each replacement remaining a basis of \(V\).

○ Proof: We prove the result by induction on \(k\). For the base case, we take \(k = 0\), and there is nothing to prove.

○ For the inductive step, suppose that \(B_k = \{w_1, w_2, \ldots, w_k, a_{k+1}, a_{k+2}, \ldots, a_n\}\) is a basis for \(V\): we must show that we can remove one of the vectors \(a_i\) and reorder the others to produce a basis \(B_{k+1} = \{w_1, w_2, \ldots, w_k, w_{k+1}, a'_{k+2}, \ldots, a'_n\}\) for \(V\).

○ By hypothesis, since \(B_k\) spans \(V\), we can write \(w_{k+1} = c_1 \cdot w_1 + \cdots + c_k \cdot w_k + d_{k+1} \cdot a_{k+1} + \cdots + d_n \cdot a_n\) for some scalars \(c_i\) and \(d_i\).

○ If all of the \(d_i\) were zero, then \(w_{k+1}\) would be a linear combination of \(w_1, \ldots, w_k\), contradicting the assumption that \(\{w_1, w_2, \ldots, w_m\}\) is a linearly independent set of vectors.

○ Thus, at least one \(d_i\) is not zero. Rearrange the vectors \(a_i\) so that \(d_{k+1} \neq 0\); then \(w_{k+1} = c_1 \cdot w_1 + \cdots + c_k \cdot w_k + d_{k+1} \cdot a'_{k+1} + \cdots + d_n \cdot a'_n\).

○ We claim now that \(B_{k+1} = \{w_1, w_2, \ldots, w_k, w_{k+1}, a'_{k+2}, \ldots, a'_n\}\) is a basis for \(V\).

○ To see that \(B_{k+1}\) spans \(V\), since \(d_{k+1} \neq 0\), we can solve for \(a'_{k+1}\) as a linear combination of the vectors \(w_1, \ldots, w_{k+1}, a'_{k+2}, \ldots, a'_n\). (The exact expression is cumbersome, and the only fact we require is to note that the coefficient of \(w_{k+1}\) is not zero.)

○ If \(x\) is any vector in \(V\), since \(B_k\) spans \(V\) we can write \(x = e_1 \cdot w_1 + \cdots + e_k \cdot w_k + e_{k+1} \cdot a'_{k+1} + \cdots + e_n \cdot a'_n\).

○ Plugging in the expression for \(a'_{k+1}\) in terms of \(w_1, \ldots, w_{k+1}, a'_{k+2}, \ldots, a'_n\) then shows that \(x\) is a linear combination of \(w_1, \ldots, w_{k+1}, a'_{k+2}, \ldots, a'_n\).
○ To see that \( B_{k+1} \) is linearly independent, suppose we had a dependence \( 0 = f_1 \cdot w_1 + \cdots + f_k \cdot w_k + f_{k+1} \cdot a_{k+1}' + \cdots + f_n \cdot a_n' \).

○ Now plug in the expression for \( a_{k+1}' \) in terms of \( w_1, \ldots, w_{k+1}, a_{k+2}', \ldots, a_n' \): all of the coefficients must be zero because \( B_k \) is linearly independent. But the coefficient of \( w_{k+1} \) is \( f_{k+1} \) times a nonzero scalar, so \( f_{k+1} = 0 \).

○ But this implies \( 0 = f_1 \cdot w_1 + \cdots + f_k \cdot w_k + f_{k+2} \cdot a_{k+2}' + \cdots + f_n \cdot a_n' \), and this is a dependence involving the vectors in \( B_k \). Since \( B_k \) is (again) linearly independent, all of the coefficients are zero. Thus \( f_1 = f_2 = \cdots = f_n = 0 \), and so \( B_{k+1} \) is linearly independent.

○ Finally, since we have shown \( B_{k+1} \) is linearly independent and spans \( V \), it is a basis for \( V \). By induction, we have the desired result for all \( 1 \leq k \leq m \).

• Although the proof of the Replacement Theorem is rather cumbersome, we obtain a number of useful corollaries.

• **Corollary**: Suppose \( V \) has a basis with \( n \) elements. If \( m > n \), then any set of \( m \) vectors of \( V \) is linearly dependent.

  ○ **Proof**: Suppose otherwise, so that \( \{w_1, w_2, \ldots, w_m\} \) is a linearly independent subset of \( V \).

  ○ Apply the Replacement Theorem with the given basis of \( V \): at the \( n \)th step we have replaced all the elements of the original basis with those in our new set, so by the conclusion of the theorem we see that \( \{w_1, \ldots, w_n\} \) is a basis for \( V \).

  ○ Then \( w_{n+1} \) is necessarily a linear combination of \( \{w_1, \ldots, w_n\} \), meaning that \( \{w_1, \ldots, w_n, w_{n+1}\} \) is linearly dependent. Thus \( \{w_1, w_2, \ldots, w_m\} \) is linearly dependent.

• **Corollary**: Any two bases of a vector space have the same number of elements.

  ○ **Proof**: If every basis is infinite, we are already done, so now suppose that \( V \) has some finite basis, and choose \( B \) to be a basis of minimal size.\(^2\)

  ○ Suppose \( B \) has \( n \) elements, and consider any other basis \( B' \) of \( V \). By the previous corollary, if \( B' \) contains more than \( n \) vectors, it would be linearly dependent (impossible). Thus, \( B' \) also has \( n \) elements, so every basis of \( V \) has \( n \) elements.

• **Theorem** (Building-Up Theorem): Given any linearly independent set of vectors in \( V \), there exists a basis of \( V \) containing those vectors. In short, any linearly independent set of vectors can be extended to a basis.

  ○ **Proof** (finite basis case): Let \( S \) be a set of linearly independent vectors and let \( B \) be any basis of \( V \) (we have already shown that \( V \) has a basis). Apply the Replacement Theorem to \( B \) and \( S \): this produces a new basis of \( V \) containing \( S \).

  ○ **Remark**: Although we appealed to the Replacement Theorem here, we can also give a slightly different, more constructive argument like the one we gave for obtaining a basis from a spanning set.

    1. Start with a linearly independent set \( S \) of vectors in \( V \). If \( S \) spans \( V \), then we are done.
    2. If \( S \) does not span \( V \), there is an element \( v \) in \( V \) which is not in the span of \( S \). Put \( v \) in \( S \); then by hypothesis, the new \( S \) will still be linearly independent.
    3. Repeat the above two steps until \( S \) spans \( V \).

  ○ If \( V \) is “finite-dimensional” (see below), then this procedure will always terminate in a finite number of steps. In the case where \( V \) is “infinite-dimensional”, matters are trickier, and we will omit the very delicate technical details required to deal with this case.

---

\(^2\)The size of a basis is either a nonnegative integer or \( \infty \). The fact that a basis of smallest size must exist follows from an axiom (the well-ordering principle) that any nonempty set of nonnegative integers has a smallest element.
1.5.3 Dimension

- **Definition:** If $V$ is a vector space, the number of elements in any basis of $V$ is called the dimension of $V$ and is denoted $\dim(V)$. If the dimension of $V$ is a finite number, we say that $V$ is **finite-dimensional**; otherwise, we say $V$ is **infinite-dimensional**.

  - Our results above assure us that the dimension of a vector space is always well-defined: every vector space has a basis, and any other basis will have the same number of elements.

- **Here are a few examples:**
  - **Example:** The dimension of $\mathbb{R}^n$ is $n$, since the $n$ standard unit vectors form a basis. (This at least suggests that the term “dimension” is reasonable, since it is the same as our usual notion of dimension.)
  - **Example:** The dimension of the vector space of $m \times n$ matrices is $mn$, because there is a basis consisting of the $mn$ matrices $E_{i,j}$, where $E_{i,j}$ is the matrix with a 1 in the $(i,j)$-entry and 0s elsewhere.
  - **Example:** The dimension of the vector space of all polynomials is $\infty$, because the (infinite list of) polynomials $1, x, x^2, x^3, \ldots$ are a basis for the space.
  - **Example:** Over the field of real numbers, the vector space of complex numbers has dimension 2, since the set $\{1, i\}$ forms a basis.
  - **Example:** Over the field of complex numbers, the vector space of complex numbers has dimension 1, since the set $\{1\}$ forms a basis.

- As the last two examples indicate, the dimension of a vector space depends on the field we are using.
  - To avoid ambiguities, it is best to indicate which field we are using when we discuss dimensions: we often do this by writing a subscript to indicate the field, so that $\dim_F V$ denotes the dimension of $V$ as a vector space over the field $F$.
  - Thus, $\dim_{\mathbb{R}}(\mathbb{C}) = 2$ while $\dim_{\mathbb{C}}(\mathbb{C}) = 1$.
  - When the field is implied by context (or not relevant to the discussion) we will omit it.

- **Here are a few basic properties of dimension:**
  - If $W$ is a subspace of $V$, then $\dim(W) \leq \dim(V)$. (Choose any basis of $W$: it is a linearly independent set of vectors in $V$, so it is contained in some basis of $V$ by the Building-Up Theorem.)
  - If $\dim(V) = n$, then any linearly independent set of vectors has at most $n$ elements. (This was a corollary to the Replacement Theorem.)
  - If $\dim(V) = n$, then any linearly independent set of $n$ vectors is a basis for $V$. (Apply the Replacement Theorem.)
  - If $\dim(V) = n$, then any spanning set of $V$ has at least $n$ elements. (Any spanning set contains a basis.)
  - If $\dim(V) = n$, then any spanning set of $V$ having exactly $n$ elements is a basis for $V$. (The spanning set contains a basis, but since the basis must have $n$ elements, the basis is the entire spanning set.)

- **Example:** Find the dimension of the complex vector space of $3 \times 3$ matrices $A$ satisfying $A^T = -A$.
  - If $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ is such a matrix, then $A^T = -A$ requires
  
  \[
  \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} = - \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},
  \]
  so that
  
  \[
  a = e = i = 0, b = d, c = g, \text{ and } h = f.
  \]
  - Thus, $A = \begin{bmatrix} 0 & b & c \\ -b & 0 & f \\ -c & -f & 0 \end{bmatrix} = b \cdot \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + c \cdot \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + f \cdot \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.
  
  - Thus, the three matrices
  \[
  \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},
  \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},
  \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}
  \]
  form a basis for the space, so the dimension is 3.
• In general, finite-dimensional vector spaces are much better-behaved than infinite-dimensional vector spaces. We will therefore usually focus our attention on finite-dimensional spaces, since infinite-dimensional spaces can have occasional counterintuitive properties. For example:

• Example: The dimension of the vector space of all real-valued functions on the interval $[0, 1]$ is $\infty$, because it contains the infinite-dimensional vector space of polynomials.

  ◦ We have not actually written down a basis for the vector space of all real-valued functions on the interval $[0, 1]$, although (per our earlier results) this vector space does have a basis.
  ◦ There is a good reason for this: it is not possible to give a simple description of such a basis.
  ◦ The set of functions $f_a(x) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{if } x \neq a \end{cases}$ for real numbers $a$, does not form a basis for the space of real-valued functions: although this infinite set of vectors is linearly independent, it does not span the space, since (for example) the constant function $f(x) = 1$ cannot be written as a finite linear combination of these functions.

1.6 Inner Product Spaces

• Now that we have described some of the fundamental properties shared by all vector spaces, we will turn our attention to vector spaces possessing an additional kind of structure that resembles the dot product in $\mathbb{R}^n$, and which will allow us to formulate notions of “length” and “angle” in more general vector spaces.

  ◦ As motivation, recall, for instance, that if $v \in \mathbb{R}^n$, then the dot product $v \cdot v = ||v||^2$ is the square of the length of $v$.
  ◦ However, if we try to use the same definition of dot product to find the “length” of a vector of complex numbers, we will obtain less sensible results. For example, the dot product of the vector $^3 (1, 0, i)$ with itself is 0, but this vector certainly doesn’t have “length zero”!
  ◦ The most natural choice of “length” for the vector $(1, 0, i)$ is $\sqrt{2}$: the first component has absolute value 1, the second has absolute value 0, and the last has absolute value 1, for an overall “length” of $\sqrt{1^2 + 0^2 + 1^2}$, in analogy with the real vector $(1, 0, 1)$ which also has length $\sqrt{2}$.
  ◦ One way to obtain a nonnegative function that seems to capture this idea of “length” for a complex vector is to include a conjugation: notice that for any complex vector $v$, the dot product $v \cdot \overline{v}$ will always be a nonnegative real number.
  ◦ Using the example above, we can compute that $(1, 0, i) \cdot (1, 0, i) = 1^2 + 0^2 + 1^2 = 2$, so this “modified dot product” seems to give the square of the length of a complex vector, at least in this one case.
  ◦ All of this suggests that the right analogy of the “$\mathbb{R}^n$ dot product” for a pair of complex vectors $v, w$ might be $v \cdot \overline{w}$, not $v \cdot w$.

• With this in mind, let us now discuss how to generalize the dot product to more general vector spaces.

1.6.1 Inner Products

• Definition: If $V$ is a real vector space, an inner product on $V$ is a pairing that assigns a scalar in $F$ to each ordered pair $(v, w)$ of vectors in $V$. This pairing is denoted $\langle v, w \rangle$ and must satisfy the following properties:

  [I1] Linearity in the first argument: $\langle v_1 + cv_2, w \rangle = \langle v_1, w \rangle + c\langle v_2, w \rangle$.
  [I2] Symmetry: $\langle v, w \rangle = \langle w, v \rangle$.
  [I3] Positive-definiteness: $\langle v, v \rangle \geq 0$ for all $v$, and $\langle v, v \rangle = 0$ only when $v = 0$.

---

3In this section, we will use parentheses around vectors rather than angle brackets, since we will shortly be using angle brackets to denote an inner product. We will also avoid using dots when discussing scalar multiplication, and reserve the dot notation for the dot product of two vectors.
○ The linearity and symmetry properties are fairly clear: if we fix the second component, the inner product behaves like a linear function in the first component, and we want both components to behave in the same way.
○ The positive-definiteness property is intended to capture an idea about “length”: namely, the length of a vector \( \mathbf{v} \) should be the inner product of \( \mathbf{v} \) with itself, and lengths are supposed to be nonnegative. Furthermore, the only vector of length zero should be the zero vector.

- **Definition:** A vector space \( V \) together with an inner product \( \langle \cdot , \cdot \rangle \) on \( V \) is called an inner product space.
  - Any given vector space may have many different inner products.
  - When we say “suppose \( V \) is an inner product space”, we intend this to mean that \( V \) is equipped with a particular (fixed) inner product.

- The entire purpose of defining an inner product is to generalize the notion of the dot product to more general vector spaces, so we should check that the dot product on \( \mathbb{R}^n \) is actually an inner product:
- **Example:** Show that the standard dot product on \( \mathbb{R}^n \), defined as \( \langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = x_1 y_1 + \cdots + x_n y_n \) is an inner product.
  - [I1]-[I2]: It is an easy algebraic computation to verify the linearity and symmetry properties.
  - [I3]: If \( \mathbf{v} = (x_1, \ldots, x_n) \) then \( \mathbf{v} \cdot \mathbf{v} = x_1^2 + x_2^2 + \cdots + x_n^2 \). Since each square is nonnegative, \( \mathbf{v} \cdot \mathbf{v} \geq 0 \), and \( \mathbf{v} \cdot \mathbf{v} = 0 \) only when all of the components of \( \mathbf{v} \) are zero.

- There are other examples of inner products on \( \mathbb{R}^n \) beyond the standard dot product.
- **Example:** Show that the pairing \( \langle (x_1, y_1), (x_2, y_2) \rangle = 3x_1 x_2 + 2x_1 y_2 + 2x_2 y_1 + 4y_1 y_2 \) on \( \mathbb{R}^2 \) is an inner product.
  - [I1]-[I2]: It is an easy algebraic computation to verify the linearity and symmetry properties.
  - [I3]: We have \( \langle (x, y), (x, y) \rangle = 3x^2 + 4xy + 4y^2 = 2x^2 + (x + 2y)^2 \), and since each square is nonnegative, the inner product is always nonnegative. Furthermore, it equals zero only when both squares are zero, and this clearly only occurs for \( x = y = 0 \).

- Another class of very important examples of an inner product are defined on function spaces:
- **Example:** Let \( V \) be the vector space of continuous (real-valued) functions on the interval \([a, b]\), and \( w(x) \) be any continuous function. Show that \( \langle f, g \rangle = \int_a^b f(x) g(x) \, dx \) is an inner product on \( V \).
  - [I1]: We have \( \langle f_1 + cf_2, g \rangle = \int_a^b \left[ f_1(x) + cf_2(x) \right] g(x) \, dx = \int_a^b f_1(x) g(x) \, dx + c \int_a^b f_2(x) g(x) \, dx = \langle f_1, g \rangle + c \langle f_2, g \rangle \).
  - [I2]: Observe that \( \langle g, f \rangle = \int_a^b g(x) f(x) \, dx = \int_a^b f(x) g(x) \, dx = \langle f, g \rangle \).
  - [I3]: Notice that \( \langle f, f \rangle = \int_a^b f(x)^2 \, dx \) is the integral of a nonnegative function, so it is always nonnegative. Furthermore (since \( f \) is assumed to be continuous) the integral of \( f^2 \) cannot be zero unless \( f \) is identically zero.
  - **Remark:** More generally, if \( w(x) \) is any fixed positive (“weight”) function that is continuous on \([a, b]\), \( \langle f, g \rangle = \int_a^b f(x) g(x) \cdot w(x) \, dx \) is an inner product on \( V \).

- With a minor modification, we can extend the idea of an inner product to a complex vector space:
- **Definition:** If \( V \) is a (real or) complex vector space, an inner product on \( V \) is a pairing that assigns a scalar in \( F \) to each ordered pair \( (\mathbf{v}, \mathbf{w}) \) of vectors in \( V \). This pairing is denoted \( \langle \mathbf{v}, \mathbf{w} \rangle \) and must satisfy the following properties:
  - [I1] Linearity in the first argument: \( \langle v_1 + cv_2, w \rangle = \langle v_1, w \rangle + c \langle v_2, w \rangle \).
  - [I2] Conjugate-symmetry: \( \langle v, w \rangle = \overline{\langle w, v \rangle} \) (where the bar denotes the complex conjugate).
  - [I3] Positive-definiteness: \( \langle v, v \rangle \geq 0 \) for all \( v \), and \( \langle v, v \rangle = 0 \) only when \( v = 0 \).
This definition generalizes the one we gave for a real vector space earlier: if \( V \) is a real vector space then \( \langle w, v \rangle = \langle w, v \rangle \) since \( \langle w, v \rangle \) is always a real number.

The only difference between this definition and the one we gave for real vector spaces is that the symmetry property has been replaced with conjugate-symmetry. Ultimately, the reason for this choice is our desire to extend the idea of the dot product to vectors of complex numbers.

Another (perhaps more compelling) reason is that the only "symmetric complex inner product", by which we mean a nonzero complex vector space \( V \) with an inner product \( (\cdot, \cdot) \) satisfying the three properties given, but with \([I2]\) replaced with the symmetric version \( (v, w) = \langle w, v \rangle \), is the trivial inner product on the zero space. Therefore, because we want to develop an interesting theory, we require conjugate-symmetry.

Important Warning: In other disciplines (particularly physics), inner products are often defined to be linear in the second argument, rather than the first. With this convention, the roles of the first and second component will be reversed (relative to our definition). This does not make any difference in the general theory, but can be extremely confusing since the definitions in mathematics and physics are otherwise identical, and the properties will have very similar statements.

- The chief reason for using this definition is that our modified dot product for complex vectors is an inner product:

- **Example:** For complex numbers \( v = (x_1, \ldots, x_n) \) and \( w = (y_1, \ldots, y_n) \), show that the map \( \langle v, w \rangle = x_1 \overline{y_1} + \cdots + x_n \overline{y_n} \) is an inner product on \( \mathbb{C}^n \).

  - **[I1]:** If \( v_1 = (x_1, \ldots, x_n) \), \( v_2 = (z_1, \ldots, z_n) \), and \( w = (y_1, \ldots, y_n) \), then
    \[
    \langle v_1 + c v_2, w \rangle = (x_1 + c z_1) \overline{y_1} + \cdots + (x_n + c z_n) \overline{y_n} \]
    \[
    = (x_1 \overline{y_1} + \cdots + x_n \overline{y_n}) + c (z_1 \overline{y_1} + \cdots + z_n \overline{y_n})
    \]
    \[
    = \langle v_1, w \rangle + c \langle v_2, w \rangle.
    \]

  - **[I2]:** Observe that \( \langle w, v \rangle = \overline{y_1 x_1} + \cdots + \overline{y_n x_n} = \overline{y_1} x_1 + \cdots + \overline{y_n} x_n = \langle v, w \rangle \).

  - **[I3]:** If \( v = (x_1, \ldots, x_n) \) then \( \langle v, v \rangle = x_1 \overline{x_1} + x_2 \overline{x_2} + \cdots + x_n \overline{x_n} = |x_1|^2 + \cdots + |x_n|^2 \). Each term is nonnegative, so \( v \cdot v \geq 0 \), and clearly \( v \cdot v = 0 \) only when all of the components of \( v \) are zero.

  - This map is often called the "standard inner product" on \( \mathbb{C}^n \), since it is fairly natural.

- Here is another, more complicated example:

- **Example:** Let \( V = M_{n \times n}(\mathbb{C}) \) be the vector space of complex \( n \times n \) matrices. Show that \( \langle A, B \rangle = \text{tr}(AB^*) \) is an inner product on \( V \), where \( M^* = \overline{M}^T \) is the complex conjugate of the transpose of \( M \) (often called the conjugate transpose or the adjoint of \( M \)).

  - **[I1]:** We have \( \langle A + c C, B \rangle = \text{tr}[(A + c C)B^*] = \text{tr}[AB^* + c CB^*] = \text{tr}(AB^*) + c \text{tr}(CB^*) = \langle A, B \rangle + c \langle C, B \rangle \), where we used the facts that \( \text{tr}(M + N) = \text{tr}(M) + \text{tr}(N) \) and \( \text{tr}(c M) = c \text{tr}(M) \).

  - **[I2]:** Observe that \( \langle B, A \rangle = \overline{\text{tr}(BA^*)} = \text{tr}(B^* A) = \text{tr}(AB^*) = \langle A, B \rangle \), where we used the facts that \( \text{tr}(MN) = \text{tr}(M^* N^*) \) and that \( \text{tr}(MN) = \text{tr}(NM) \), both of which are easy algebraic calculations.

  - **[I3]:** We have \( \langle A, A \rangle = \sum_{j=1}^{n} \sum_{k=1}^{n} A_{j,k}^* A_{k,j} = \sum_{j=1}^{n} \sum_{k=1}^{n} A_{j,k} A_{j,k}^* = \sum_{j=1}^{n} |A_{j,k}|^2 \geq 0 \), and equality can only occur when each element of \( A \) has absolute value zero (i.e., is zero).

- **Remark:** This inner product is often called the Frobenius inner product.

### 1.6.2 Properties of Inner Products, Norms

- Our fundamental goal in studying inner products is to extend the notion of length in \( \mathbb{R}^n \) to a more general setting. Using the positive-definiteness property, we can define a notion of length in an inner product space.

- **Definition:** If \( V \) is an inner product space, we define the norm (or length) of a vector \( v \) to be \( ||v|| = \sqrt{\langle v, v \rangle} \).
• When $V = \mathbb{R}^n$ with the standard dot product, the norm on $V$ reduces to the standard notion of “length” of a vector in $\mathbb{R}^n$.

• Here are a few basic properties of inner products and norms:

• Proposition: If $V$ is a (real or complex) inner product space with inner product $(\cdot, \cdot)$, then the following are true:

1. For any vectors $v, w_1,$ and $w_2$, $(v, w_1 + w_2) = (v, w_1) + (v, w_2)$.
   - Proof: We have $(v, w_1) = (v, w_1 + w_2) - (v, w_2)$ by $[I1]$ and $[I2]$.

2. For any vectors $v$ and $w$ and scalar $c$, $(v, cw) = c(v, w)$.
   - Proof: We have $(v, cw) = c(v, w)$ by $[I1]$ and $[I2]$.

3. For any vector $v$, $(v, 0) = 0$.
   - Proof: Apply property (3) and $[I2]$ with $c = 0$, using the fact that $0w = 0$ for any $w$.

4. For any vector $v$, $||v||$ is a nonnegative real number, and $||v|| = 0$ if and only if $v = 0$.
   - Proof: Immediate from $[I3]$.

5. For any vector $v$ and scalar $c$, $||cv|| = |c| ||v||$.
   - Proof: We have $||c \cdot v|| = \sqrt{(c \cdot v, c \cdot v)} = \sqrt{c^2(v, v)} = |c| ||v||$, using $[I2]$ and property (2).

• In $\mathbb{R}^n$, there are a number of fundamental inequalities about lengths, which generalize quite pleasantly to general inner product spaces.

• The following result, in particular, is one of the most fundamental inequalities in all of mathematics:

• Theorem (Cauchy-Schwarz Inequality): For any $v$ and $w$ in an inner product space $V$, we have $|\langle v, w \rangle| \leq ||v|| \cdot ||w||$, with equality if and only if the set $\{v, w\}$ is linearly dependent.
   - Proof: If $w = 0$ then the result is trivial (since both sides are zero, and $\{v, 0\}$ is always dependent), so now assume $w \neq 0$.
   - Let $t = \frac{\langle v, w \rangle}{\langle w, w \rangle}$. By properties of inner products and norms, we can write
     
     $||v - tw||^2 = \langle v - tw, v - tw \rangle = \langle v, v \rangle - t \langle w, v \rangle - t \langle v, w \rangle + t^2 \langle w, w \rangle$
     
     $= \langle v, v \rangle - \frac{\langle v, w \rangle \langle v, v \rangle}{\langle w, w \rangle} - \frac{\langle v, w \rangle^2}{\langle w, w \rangle} + \frac{\langle v, w \rangle^2}{\langle w, w \rangle}^2$ = $\langle v, v \rangle - \frac{\langle v, w \rangle^2}{\langle w, w \rangle}$.

   - Therefore, since $||v - tw||^2 \geq 0$ and $\langle v, w \rangle \geq 0$, clearing denominators and rearranging yields $|\langle v, w \rangle|^2 \leq \langle v, v \rangle \cdot \langle w, w \rangle$. Taking the square root yields the stated result.

   - Furthermore, we will have equality if and only if $||v - tw||^2 = 0$, which is in turn equivalent to $v - tw = 0$; namely, when $v$ is a multiple of $w$. Since we also get equality if $w = 0$, equality occurs precisely when the set $\{v, w\}$ is linearly dependent.

   - Remark: As written, this proof is completely mysterious: why does making that particular choice for $t$ work? Here is some motivation: in the special case where $V$ is a real vector space, we can write $||v - tw||^2 = \langle v, v \rangle - 2t \langle v, w \rangle + t^2 \langle w, w \rangle$, which is a quadratic function of $t$ that is always nonnegative.

   - To decide whether a quadratic function is always nonnegative, we complete the square to see that
     
     $\langle v, v \rangle - 2t \langle v, w \rangle + t^2 \langle w, w \rangle = \langle w, w \rangle \left(t - \frac{\langle v, w \rangle}{\langle w, w \rangle}\right)^2 + \left(\langle v, v \rangle - \frac{\langle v, w \rangle^2}{\langle w, w \rangle}\right)$.

   - Thus, the minimum value of the quadratic function is $\langle v, v \rangle - \frac{\langle v, w \rangle^2}{\langle w, w \rangle}$, and it occurs when $t = \frac{\langle v, w \rangle}{\langle w, w \rangle}$.
• The Cauchy-Schwarz inequality has many applications (most of which are, naturally, proving other inequalities). Here are a few such applications:

• **Theorem (Triangle Inequality):** For any vectors \( \mathbf{v} \) and \( \mathbf{w} \) in an inner product space \( V \), we have \( ||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}|| \), with equality if and only if one vector is a positive-real scalar multiple of the other.

  ○ **Proof:** Using the Cauchy-Schwarz inequality and the fact that for any complex number \( z \), \( \text{Re}(z) \leq |z| \), we have

  \[ ||\mathbf{v} + \mathbf{w}||^2 = \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle + 2\text{Re}(\langle \mathbf{v}, \mathbf{w} \rangle) + \langle \mathbf{w}, \mathbf{w} \rangle \leq (\mathbf{v}, \mathbf{v}) + 2||\mathbf{v}|| ||\mathbf{w}|| + (\mathbf{w}, \mathbf{w}) \leq ||\mathbf{v}||^2 + 2||\mathbf{v}|| ||\mathbf{w}|| + ||\mathbf{w}||^2. \]

  Taking the square root of both sides yields the desired result.

  ○ Equality will hold if and only if \( \{\mathbf{v}, \mathbf{w}\} \) is linearly dependent (for equality in the Cauchy-Schwarz inequality) and \( \langle \mathbf{v}, \mathbf{w} \rangle \) is a nonnegative real number. If either vector is zero, equality always holds. Otherwise, we must have \( \mathbf{v} = c \cdot \mathbf{w} \) for some nonzero constant \( c \); then \( \langle \mathbf{v}, \mathbf{w} \rangle = c \langle \mathbf{w}, \mathbf{w} \rangle \) will be a nonnegative real number if and only if \( c \) is a nonnegative real number.

• **Example:** Show that for any continuous function \( f \) on \([0, 3]\), it is true that \( \int_0^3 x f(x) \, dx \leq 3\sqrt{\int_0^3 f(x)^2 \, dx} \).

  ○ Simply apply the Cauchy-Schwarz inequality to \( f \) and \( g(x) = x \) in the inner product space of continuous functions on \([0, 3]\) with inner product \( \langle f, g \rangle = \int_0^3 f(x)g(x) \, dx \).

  ○ We obtain \( ||f|| \leq ||f|| ||g|| \), or, explicitly, \( ||\int_0^3 x f(x) \, dx|| \leq \sqrt{\int_0^3 f(x)^2 \, dx} \cdot \sqrt{\int_0^3 x^2 \, dx} = 3\sqrt{\int_0^3 f(x)^2 \, dx} \).

  ○ Since any real number is less than or equal to its absolute value, we immediately obtain the required inequality \( \int_0^3 x f(x) \, dx \leq 3\sqrt{\int_0^3 f(x)^2 \, dx} \).

• **Example:** Show that for any positive reals \( a, b, c \), it is true that \( \sqrt{\frac{a+2b}{a+b+c}} + \sqrt{\frac{b+2c}{a+b+c}} + \sqrt{\frac{c+2a}{a+b+c}} \leq 3 \).

  ○ Let \( \mathbf{v} = (\sqrt{a+2b}, \sqrt{b+2c}, \sqrt{c+2a}) \) and \( \mathbf{w} = (1, 1, 1) \) in \( \mathbb{R}^3 \). By the Cauchy-Schwarz inequality, \( \mathbf{v} \cdot \mathbf{w} \leq ||\mathbf{v}|| ||\mathbf{w}|| \).

  ○ We compute \( \mathbf{v} \cdot \mathbf{w} = \sqrt{a+2b} + \sqrt{b+2c} + \sqrt{c+2a} \), along with \( ||\mathbf{v}||^2 = (a+2b) + (b+2c) + (c+2a) = 3(a+b+c) \) and \( ||\mathbf{w}||^2 = 3 \).

  ○ Thus, we see \( \sqrt{a+2b} + \sqrt{b+2c} + \sqrt{c+2a} \leq 3\sqrt{3(a+b+c)} \cdot \sqrt{3} \), and upon dividing through by \( \sqrt{a+b+c} \) we obtain the required inequality.

• **Example** (for those who like quantum mechanics): Prove the momentum-position formulation of Heisenberg’s uncertainty principle: \( \sigma_x \sigma_p \geq \hbar/2 \). (In words: the product of uncertainties of position and momentum is greater than or equal to half of the reduced Planck constant.)

  ○ It is a straightforward computation that, for two (complex-valued) observables \( X \) and \( Y \), the pairing \( \langle X, Y \rangle = E[X \overline{Y}] \), the expected value of \( X \overline{Y} \), is an inner product on the space of observables.

  ○ Assume (for simplicity) that \( x \) and \( p \) both have expected value 0.

  ○ We assume as given the commutation relation \( xp - px = i\hbar \).

  ○ By definition, \( \sigma_x^2 = E[x^2] = \langle x, x \rangle \) and \( \sigma_p^2 = E[p^2] = E[\overline{p}p] = \langle \overline{p}, p \rangle \) are the variances of \( x \) and \( p \) respectively (in the statistical sense).

  ○ By the Cauchy-Schwarz inequality, we can therefore write \( \sigma_x^2 \sigma_p^2 = \langle x, x \rangle \langle \overline{p}, p \rangle \geq |\langle x, \overline{p} \rangle|^2 = |E[xp]|^2 \).

  ○ We can write \( xp = \frac{1}{2}(xp + px) + \frac{1}{2}(xp - px) \), where the first component is real and the second is imaginary, so taking expectations yields \( E[xp] = \frac{1}{2}E[xp + px] + \frac{1}{2}E[xp - px] \), and therefore, \( |E[xp]| \geq \frac{1}{2} |E[xp - px]| = \frac{1}{2} |i\hbar| = \frac{\hbar}{2} \).

  ○ Combining with the inequality above yields \( \sigma_x^2 \sigma_p^2 \geq \hbar^2/4 \), and taking square roots yields \( \sigma_x \sigma_p \geq \hbar/2 \).
1.6.3 Orthogonality, Orthonormal Bases and the Gram-Schmidt Procedure

- Motivated by the Cauchy-Schwarz inequality, we can define a notion of angle between two nonzero vectors in a real inner product space:

- **Definition:** If $V$ is a real inner product space, we define the **angle** between two nonzero vectors $v$ and $w$ to be the real number $\theta$ in $[0, \pi]$ satisfying $\cos \theta = \frac{\langle v, w \rangle}{||v|| \cdot ||w||}$.
  
  - By the Cauchy-Schwarz inequality, the quotient on the right is a real number lying in $[-1, 1]$, so there is exactly one such angle $\theta$.

- A particular case of interest is when two vectors have inner product 0.

- **Definition:** We say two vectors in an inner product space are **orthogonal** if their inner product is zero. We say a set $S$ of vectors is an **orthogonal set** if every pair of vectors in $S$ is orthogonal.
  
  - By our basic properties, the zero vector is orthogonal to every vector. Two nonzero vectors will be orthogonal if and only if the angle between them is $\pi/2$. (This generalizes the idea of two vectors being “perpendicular”.)

- **Example:** In $\mathbb{R}^3$ with the standard dot product, the vectors $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ are orthogonal.

- **Example:** In $\mathbb{R}^3$ with the standard dot product, the three vectors $(-1, 1, 2)$, $(2, 0, 1)$, and $(1, 5, -2)$ form an orthogonal set, since the dot product of each pair is zero.

- The first orthogonal set above seems more natural than the second.
  
  - One reason for this is that the vectors in the first set each have length 1, while the vectors in the second set have various different lengths ($\sqrt{5}$, $\sqrt{3}$, and $\sqrt{30}$ respectively).

- **Definition:** We say a set $S$ of vectors is an **orthonormal set** if every pair of vectors in $S$ is orthogonal, and every vector in $S$ has norm 1.

- **Example:** In $\mathbb{R}^3$, $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is an orthonormal set, but $\{(-1, 1, 2), (2, 0, 1), (1, 5, -2)\}$ is not.

- In both examples above, notice that the given orthogonal sets are also linearly independent. This feature is not an accident:

- **Proposition** (Orthogonality and Independence): In any inner product space, every orthogonal set of nonzero vectors is linearly independent.

  - **Proof:** Suppose we had a linear dependence $a_1v_1 + \cdots + a_nv_n = 0$ for an orthogonal set of nonzero vectors $\{v_1, \ldots, v_n\}$.
    
    - Then, for any $j$, $0 = \langle 0, v_j \rangle = \langle a_1v_1 + \cdots + a_nv_n, v_j \rangle = a_1\langle v_1, v_j \rangle + \cdots + a_n\langle v_n, v_j \rangle = a_j\langle v_j, v_j \rangle$, since each of the inner products $\langle v_i, v_j \rangle$ for $i \neq j$ is equal to zero.
    
    - But now, since $v_j$ is not the zero vector, $\langle v_j, v_j \rangle$ is positive, so it must be the case that $a_j = 0$. This holds for every $j$, so all the coefficients of the linear dependence are zero. Hence there can be no nontrivial linear dependence, so any orthogonal set is linearly independent.

- **Corollary:** If $V$ is an $n$-dimensional vector space and $S$ is an orthogonal set of $n$ nonzero vectors, then $S$ is a basis for $V$. (We refer to such a basis as an **orthogonal basis**.)

  - **Proof:** By the proposition above, $S$ is linearly independent, and by our earlier results, a linearly-independent set of $n$ vectors in an $n$-dimensional vector space is necessarily a basis.

- If we have a basis of $V$, then (essentially by definition) every vector in $V$ can be written as a unique linear combination of the basis vectors.

  - However, as we have seen, computing the coefficients of the linear combination can be quite cumbersome.
• **Theorem (Orthogonal Decomposition):** If \( V \) is an \( n \)-dimensional vector space and \( S = \{e_1, \ldots, e_n\} \) is an orthogonal basis, then for any \( v \in S \), we can write \( v = c_1 e_1 + \cdots + c_n e_n \), where \( c_k = \langle v, e_k \rangle / \langle e_k, e_k \rangle \) for each \( 1 \leq k \leq n \). In particular, if \( S \) is an orthonormal basis, then each \( c_k = \langle v, e_k \rangle \).

  • **Proof:** Since \( S \) is a basis, there do exist such coefficients \( c_i \) and they are unique.
  
  • We then compute \( \langle v, e_k \rangle = \langle e_1, e_k \rangle + \cdots + c_n e_n, e_k \rangle = c_1 \langle e_1, e_k \rangle + \cdots + c_n \langle e_n, e_k \rangle = c_k \langle e_k, e_k \rangle \) since each of the inner products \( \langle e_j, e_k \rangle \) for \( j \neq k \) is equal to zero.
  
  • Therefore, we must have \( c_k = \langle v, e_k \rangle / \langle e_k, e_k \rangle \) for each \( 1 \leq k \leq n \).
  
  • If \( S \) is an orthonormal basis, then \( \langle e_k, e_k \rangle = 1 \) for each \( k \), so we get the simpler expression \( c_k = \langle v, e_k \rangle \).

• **Example:** Write \( v = (7, 3, -4) \) as a linear combination of the basis vectors \( \{(−1, 1, 2), (2, 0, 1), (1, 5, −2)\} \) of \( \mathbb{R}^3 \).

  • We saw above that this set is an orthogonal basis, so let \( e_1 = (−1, 1, 2), e_2 = (2, 0, 1), \) and \( e_3 = (1, 5, −2) \).
  
  • We compute \( v \cdot e_1 = −12, v \cdot e_2 = 10, v \cdot e_3 = 30, e_1 \cdot e_1 = 6, e_2 \cdot e_2 = 5, \) and \( e_3 \cdot e_3 = 30 \).
  
  • Thus, per the theorem, \( v = c_1 e_1 + c_2 e_2 + c_3 e_3 \) where \( c_1 = −12/6 = −2, c_2 = 10/5 = 2, \) and \( c_3 = 30/30 = 1 \).
  
  • Indeed, we can verify that \( (7, 3, −4) = −2(−1, 1, 2) + 2(2, 0, 1) + 1(1, 5, −2) \).

• **Theorem (Gram-Schmidt Procedure):** Let \( S = \{v_1, v_2, \ldots\} \) be a basis of the inner product space \( V \), and set \( V_k = \text{span}(v_1, \ldots, v_k) \). Then there exists an orthogonal set of vectors \( \{w_1, w_2, \ldots\} \) such that, for each \( k \geq 1 \), \( \text{span}(w_1, \ldots, w_k) = \text{span}(V_k) \) and \( w_k \) is orthogonal to every vector in \( V_{k−1} \). Furthermore, this sequence is unique up to multiplying the elements by nonzero scalars.

  • **Proof:** We construct the sequence \( \{w_1, w_2, \ldots\} \) by induction. For the base case, we set \( w_1 = v_1 \); clearly this choice has both of the required properties (since \( V_0 \) is the zero subspace).
  
  • For the inductive step, suppose we have constructed \( \{w_1, w_2, \ldots, w_{k−1}\} \), where \( \text{span}(w_1, \ldots, w_{k−1}) = \text{span}(V_{k−1}) \).
  
  • Define \( w_k = v_k − a_1 w_1 − a_2 w_2 − \cdots − a_{k−1} w_{k−1} \), where \( a_j = \langle v_k, w_j \rangle / \langle w_j, w_j \rangle \).
  
  • By hypothesis, each of \( w_1, \ldots, w_{k−1} \) is a linear combination of \( v_1, \ldots, v_{k−1}, v_k \), as is \( w_k \) (upon plugging in these expressions to the definition of \( w_k \)). Thus, by properties of the span, \( \text{span}(w_1, \ldots, w_k) \subseteq V_k \).
  
  • Inversely, since each of \( v_1, \ldots, v_{k−1} \) is a linear combination of \( w_1, \ldots, w_{k−1}, w_k \), as is \( v_k \) (upon solving for \( v_k \) in the definition of \( w_k \) above). Thus, by properties of the span, \( V_k \subseteq \text{span}(w_1, \ldots, w_k) \). Hence, \( \text{span}(w_1, \ldots, w_k) = V_k \).
  
  • Furthermore, we have

\[
\langle w_k, w_j \rangle = \langle v_k − a_1 w_1 − a_2 w_2 − \cdots − a_{k−1} w_{k−1}, w_j \rangle
\]

\[
= \langle v_k, w_j \rangle − a_1 \langle w_1, w_j \rangle − a_2 \langle w_2, w_j \rangle − \cdots − a_{k−1} \langle w_{k−1}, w_j \rangle
\]

\[
= \langle v_k, w_j \rangle − a_j \langle w_j, w_j \rangle = 0
\]

because all of the inner products \( \langle w_i, w_j \rangle \) are zero except for \( \langle w_j, w_j \rangle \).

  • Therefore, \( w_k \) is orthogonal to each of \( w_1, \ldots, w_{k−1} \), and is therefore orthogonal to all linear combinations of these vectors.

  • For the uniqueness, we also use induction. The base case \( k = 1 \) is trivial, since \( \text{span}(w_1) \) will equal \( \text{span}(v_1) \) if and only if \( w_1 \) is a nonzero scalar multiple of \( v_1 \).
○ For the inductive step, suppose that we have some other sequence of elements \( \{y_1, y_2, \ldots, y_k\} \) with \( \text{span}(y_1, \ldots, y_k) = \text{span}(V_k) = \text{span}(w_1, \ldots, w_k) \) and with \( y_k \) orthogonal to every vector in \( V_{k-1} = \text{span}(w_1, \ldots, w_{k-1}) \).

○ Then \( y_k = c_1 w_1 + \cdots + c_k w_k \). By hypothesis, for any \( 1 \leq j \leq k-1 \), the vector \( y_k \) is orthogonal to \( w_j \), so we can write \( 0 = \langle y_k, w_j \rangle = \langle c_1 w_1 + \cdots + c_k w_k, w_j \rangle = c_j \langle w_j, w_j \rangle \).

○ But since \( \langle w_j, w_j \rangle \neq 0 \), we conclude \( c_j = 0 \) for all \( 1 \leq j \leq k-1 \). Thus, \( y_k = c_k w_k \) for some constant \( c_k \) (which clearly cannot be zero, since \( y_k \) is not zero), as required.

• **Corollary:** Every finite-dimensional inner product space has an orthonormal basis.

○ **Proof:** Choose any basis \( \{v_1, \ldots, v_n\} \) for \( V \) and apply the Gram-Schmidt procedure: this yields an orthogonal basis \( \{w_1, \ldots, w_n\} \) for \( V \).

○ Now simply normalize each vector in \( \{w_1, \ldots, w_n\} \) by dividing by its norm: this preserves orthogonality, but rescales each vector to have norm 1, thus yielding an orthonormal basis for \( V \).

• The proof of the Gram-Schmidt procedure may seem involved, but applying it in practice is fairly straightforward.

○ We remark here that, although our algorithm above gives an orthogonal basis, it is also possible to perform the normalization at each step during the procedure, to construct an orthonormal basis one vector at a time.

○ When performing computations by hand, it is generally disadvantageous to normalize at each step, because the norm of a vector will often involve square roots (which will then be carried into subsequent steps of the computation).

○ When using a computer (with approximate arithmetic), however, normalizing at each step can avoid certain numerical instability issues. The particular description of the algorithm we have discussed turns out not to be especially numerically stable, but it is possible to modify the algorithm to avoid magnifying the error as substantially when iterating the procedure.

• **Example:** Apply the Gram-Schmidt procedure to the vectors \( v_1 = (2, 1, 2), v_2 = (5, 4, 2), v_3 = (-1, 2, 1) \). Then use the result to find an orthonormal basis for \( \mathbb{R}^3 \).

○ We start with \( w_1 = v_1 = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} \).

○ Next, \( w_2 = v_2 - a_1 w_1 \), where \( a_1 = \frac{v_2 \cdot w_1}{w_1 \cdot w_1} = \frac{(5, 4, 2) \cdot (2, 1, 2)}{(2, 1, 2) \cdot (2, 1, 2)} = \frac{18}{9} = 2 \). Thus, \( w_2 = \begin{pmatrix} 1 \\ 2 \\ -2 \end{pmatrix} \).

○ Finally, \( w_3 = v_3 - b_1 w_1 - b_2 w_2 \) where \( b_1 = \frac{v_3 \cdot w_1}{w_1 \cdot w_1} = \frac{(1, 2, 1) \cdot (2, 1, 2)}{(2, 1, 2) \cdot (2, 1, 2)} = \frac{2}{9} \), and \( b_2 = \frac{v_3 \cdot w_2}{w_2 \cdot w_2} = \frac{(1, 2, 1) \cdot (1, 2, -2)}{(1, 2, -2) \cdot (1, 2, -2)} = \frac{1}{9} \). Thus, \( w_3 = \begin{pmatrix} -14/9 \\ 14/9 \\ -7/9 \end{pmatrix} \).

○ For the orthonormal basis, we simply divide each vector by its length.

○ We get \( \frac{w_1}{||w_1||} = \frac{1}{3} \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \frac{w_2}{||w_2||} = \frac{1}{3} \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix}, \) and \( \frac{w_3}{||w_3||} = \frac{1}{3} \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \).

### 1.6.4 Some Remarks about Fourier Series

• An extremely useful application of the general theory of orthonormal bases is that of Fourier series.

○ Fourier analysis, broadly speaking, studies the problem of approximating a function on an interval by trigonometric functions. This problem is very similar to the question, studied in calculus, of approximating a function by a polynomial in \( x \) (the typical method is to use Taylor polynomials).

○ Fourier series have a tremendously wide variety of applications, ranging from to solving partial differential equations (in particular, the famous wave equation and heat equation), studying acoustics and optics (decomposing an acoustic or optical waveform into simpler waves of particular frequencies), electrical engineering, and quantum mechanics.
Although a full discussion of Fourier series belongs more properly to analysis, we can give some of the basic ideas.

- A typical scenario in Fourier analysis is to approximate a continuous function on $[0, 2\pi]$ using a trigonometric polynomial: a function that is a polynomial in $\sin(x)$ and $\cos(x)$.
- Using trigonometric identities, this question is equivalent to approximating a function $f(x)$ by a (finite) Fourier series of the form $s(x) = a_0 + b_1 \cos(x) + b_2 \cos(2x) + \cdots + b_k \cos(kx) + c_1 \sin(x) + c_2 \sin(2x) + \cdots + c_k \sin(kx)$.
- Notice that, in the expression above, $s(0) = s(2\pi)$ since each function in the sum has period $2\pi$. Thus, we can only realistically hope to get close approximations to functions satisfying $f(0) = f(2\pi)$.
- Let $V$ be the vector space of continuous, real-valued functions on the interval $[0, 2\pi]$ having equal values at 0 and $2\pi$, and define an inner product on $V$ via $\langle f, g \rangle = \int_0^{2\pi} f(x)g(x) \, dx$.

**Proposition:** The functions $\{\varphi_0, \varphi_1, \varphi_2, \ldots\}$ are an orthonormal set on $V$, where $\varphi_0(x) = \frac{1}{\sqrt{2\pi}}$ and $\varphi_{2k-1}(x) = \frac{1}{\sqrt{\pi}} \cos(kx)$ and $\varphi_{2k}(x) = \frac{1}{\sqrt{\pi}} \sin(kx)$ for each $k \geq 1$.

- **Proof:** Using the product-to-sum identities, such as $\sin(ax) \sin(bx) = \frac{1}{2} [\cos(a-b)x - \cos(a+b)x]$, it is a straightforward exercise in integration to verify that $\langle \varphi_i, \varphi_j \rangle = 0$ for each $i \neq j$.
- Furthermore, we have $\langle \varphi_0, \varphi_0 \rangle = \frac{1}{2\pi} \int_0^{2\pi} 1 \, dx = 1$, $\langle \varphi_{2k-1}, \varphi_{2k-1} \rangle = \frac{1}{\pi} \int_0^{2\pi} \cos^2(kx) \, dx = 1$, and $\langle \varphi_{2k}, \varphi_{2k} \rangle = \frac{1}{\pi} \int_0^{2\pi} \sin^2(kx) \, dx = 1$. Thus, the set is orthonormal.

- If it were the case that $S = \{\varphi_0, \varphi_1, \varphi_2, \ldots\}$ were an orthonormal basis for $V$, then, given any other function $f(x)$ in $V$, we could write $f$ as a linear combination of functions in $\{\varphi_0, \varphi_1, \varphi_2, \ldots\}$, where we can compute the appropriate coefficients using the inner product on $V$.

- Unfortunately, $S$ does not span $V$: we cannot, for example, write the function $g(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} \sin(nx)$ as a finite linear combination of $\{\varphi_0, \varphi_1, \varphi_2, \ldots\}$, since doing so would require each of the infinitely many terms in the sum.
- Ultimately, the problem, as exemplified by the function $g(x)$ above, is that the definition of “basis” only allows us to write down finite linear combinations.
- If, however, we relax our approach and allow “infinite sums” of our orthonormal basis elements, then (assuming some results from analysis demonstrating the convergence), our formulas for the coefficients do give a formula for $f(x)$ as an infinite sum.

**Theorem (Fourier Series):** Let $f(x)$ be a twice-differentiable function on $[0, 2\pi]$ satisfying $f(0) = f(2\pi)$, and define the Fourier coefficients of $f$ as $a_j = \langle f, \varphi_j \rangle = \int_0^{2\pi} f(x)\varphi_j(x) \, dx$, for the trigonometric functions $\varphi_j(x)$ defined above. Then $f(x) = \sum_{j=0}^{\infty} a_j \varphi_j(x)$ for every $x$ in $[0, 2\pi]$.

- We will not prove this theorem. However, it can be interpreted as a “limiting version” of the theorem we stated earlier giving the coefficients for the linear combination of a vector in terms of an orthonormal basis.
- The theorem gives us an explicit way to write the function $f(x)$ as an “infinite linear combination” of the orthonormal basis elements $\{\varphi_0, \varphi_1, \varphi_2, \ldots\}$.

Well, you’re at the end of my handout. Hope it was helpful.

Copyright notice: This material is copyright Evan Dummit, 2012-2017. You may not reproduce or distribute this material without my express permission.

26