Linear Algebra (part 2) : Vector Spaces and Linear Transformations (by Evan Dummit, 2016, v. 2.35)

Contents

3 Vector Spaces and Linear Transformations

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 will study vector spaces from an axiomatic perspective, discuss the notions of span and linear independence, and ultimately explain why every vector space possesses a linearly independent spanning set called a "basis". We will then discuss linear transformations, which are the most natural kind of a map from one vector space to another, and show how they are intimately related with matrices.

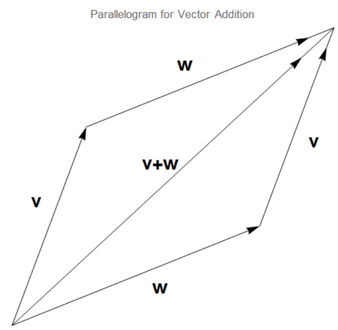
In our discussions we will give concrete examples as often as possible, and use the general properties we have shown about vector spaces to motivate results relevant to solving systems of linear equations and solving differential equations.

3.1 Review of Vectors in $\mathbb{R}^n$

- A vector, as we typically think of it, is a quantity which has both a magnitude and a direction. This is in contrast to a scalar, which carries only a magnitude.
  - Real-valued vectors are extremely useful in just about every aspect of the physical sciences, since just about everything in Newtonian physics is a vector – position, velocity, acceleration, forces, etc. There is also "vector calculus" – namely, calculus in the context of vector fields – which is typically part of a multivariable calculus course; it has many applications to physics as well.
- We often think of vectors geometrically, as a directed line segment (having a starting point and an endpoint).
- Algebraically, we write a vector as an ordered tuple of coordinates: we denote the n-dimensional vector from the origin to the point $(a_1, a_2, \cdots, a_n)$ as $\mathbf{v} = \langle a_1, a_2, \cdots, a_n \rangle$, where the a_i are real-number scalars.

- Some vectors: $\langle 1,2\rangle$, $\langle 3,5,-1\rangle$, $\left\langle -\pi, e^2, 27, 3, \frac{4}{3}, 0, 0, -1\right\rangle$.
  - Notation: We use angle brackets $\langle\cdot\rangle$ rather than parentheses $(\cdot)$ so as to draw a visual distinction between a vector and the coordinates of a point in space. We also draw arrows above vectors (as $\vec{v}$) or typeset them in boldface (as $\mathbf{v}$) in order to set them apart from scalars. Boldface is hard to produce without a computer, so it is highly recommended to use the arrow notation $\vec{v}$ when writing by hand.
  - Note: Vectors are a little bit different from directed line segments, because we don't care where a vector starts: we only care about the difference between the starting and ending positions. Thus: the directed segment whose start is $(0,0)$ and end is $(1,1)$ and the segment starting at $(1,1)$ and ending at $(2,2)$ represent the *same vector* $\langle 1,1\rangle$.

- We can add vectors (provided they are of the same length!) in the obvious way, one component at a time: if $\mathbf{v} = \langle a_1, \cdots, a_n\rangle$ and $\mathbf{w} = \langle b_1, \cdots, b_n\rangle$ then $\mathbf{v}+\mathbf{w} = \langle a_1+b_1, \cdots, a_n+b_n\rangle$.
  - We can justify this using our geometric idea of what a vector does: $\mathbf{v}$ moves us from the origin to the point $(a_1, \cdots, a_n)$. Then $\mathbf{w}$ tells us to add $\langle b_1, \cdots, b_n\rangle$ to the coordinates of our current position, and so $\mathbf{w}$ moves us from $(a_1, \cdots, a_n)$ to $(a_1+b_1, \cdots, a_n+b_n)$. So the net result is that the sum vector $\mathbf{v}+\mathbf{w}$ moves us from the origin to $(a_1+b_1, \cdots, a_n+b_n)$, meaning that it is just the vector $\langle a_1+b_1, \cdots, a_n+b_n\rangle$.
  - Geometrically, we can think of vector addition using a parallelogram whose pairs of parallel sides are $\mathbf{v}$ and $\mathbf{w}$ and whose diagonal is $\mathbf{v}+\mathbf{w}$:



- We can also 'scale' a vector by a scalar, one component at a time: if r is a scalar, then we have $r\,\mathbf{v} = \langle ra_1, \cdots, ra_n\rangle$.
  - Again, we can justify this by our geometric idea of what a vector does: if $\mathbf{v}$ moves us some amount in a direction, then $\frac{1}{2}\mathbf{v}$ should move us half as far in that direction. Analogously, $2\mathbf{v}$ should move us twice as far in that direction, while $-\mathbf{v}$ should move us exactly as far, but in the opposite direction.

- Example: If $\mathbf{v} = \langle -1,2,2\rangle$ and $\mathbf{w} = \langle 3,0,-4\rangle$ then $2\mathbf{w} = \boxed{\langle 6,0,-8\rangle}$, and $\mathbf{v}+\mathbf{w} = \boxed{\langle 2,2,-2\rangle}$. Furthermore, $\mathbf{v}-2\mathbf{w} = \boxed{\langle -7,2,10\rangle}$.

- The arithmetic of vectors in $\mathbb{R}^n$ satisfies several algebraic properties that follow more or less directly from the definition:
  - Addition of vectors is commutative and associative.
  - There is a zero vector (namely, the vector with all entries zero), and every vector has an additive inverse.
  - Scalar multiplication distributes over addition of both vectors and scalars.

3.2 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 $\mathbb{R}^n$. We would like to investigate other collections of things which possess those same properties.

- Definition: A (real) vector space is a collection V of vectors together with two binary operations, addition of vectors ($+$) and scalar multiplication of a vector by a real number ($\cdot$), satisfying the following axioms:
  - Let $\mathbf{v}$, $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ be any vectors in V and α, α_1,α_2 be any (real number) scalars.
  - Note: The statement that $+$ and $\cdot$ are binary operations means that $\mathbf{v}_1 + \mathbf{v}_2$ and $\alpha \cdot \mathbf{v}$ are always defined; that is, they are both vectors in V.
  - [A1] Addition is commutative: $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$.
  - [A2] Addition is associative: $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$.
  - [A3] There exists a zero vector $\mathbf{0}$, with $\mathbf{v} + \mathbf{0} = \mathbf{v}$.
  - [A4] Every vector $\mathbf{v}$ has an additive inverse $-\mathbf{v}$, with $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
  - [M1] Scalar multiplication is consistent with regular multiplication: $\alpha_1 \cdot (\alpha_2 \cdot \mathbf{v}) = (\alpha_1\alpha_2) \cdot \mathbf{v}$.
  - [M2] Addition of scalars distributes: $(\alpha_1 + \alpha_2) \cdot \mathbf{v} = \alpha_1 \cdot \mathbf{v} + \alpha_2 \cdot \mathbf{v}$.
  - [M3] Addition of vectors distributes: $\alpha \cdot (\mathbf{v}_1 + \mathbf{v}_2) = \alpha \cdot \mathbf{v}_1 + \alpha \cdot \mathbf{v}_2$.
  - [M4] The scalar 1 acts like the identity on vectors: $1 \cdot \mathbf{v} = \mathbf{v}$.
- Remark: One may also consider vector spaces where the collection of scalars is something other than the real numbers: for example, there exists an equally important notion of a complex vector space, whose scalars are the complex numbers. (The axioms are the same, except we allow the scalars to be complex numbers.)
  - We will primarily work with real vector spaces, in which the scalars are the real numbers.
  - The most general notion of a vector space involves scalars from a field, which is a collection of numbers which possess addition and multiplication operations which are commutative, associative, and distributive, with an additive identity 0 and multiplicative identity 1, such that every element has an additive inverse and every nonzero element has a multiplicative inverse.
  - Aside from the real and complex numbers, another example of a field is the rational numbers ("fractions").
  - One can formulate an equally interesting theory of vector spaces over any field.
- Here are some examples of vector spaces:
- Example: The vectors in $\mathbb{R}^n$ are a vector space, for any $n > 0$. (This had better be true!)
  - For simplicity we will demonstrate all of the axioms for vectors in $\mathbb{R}^2$; there, the vectors are of the form $\langle x, y\rangle$ and scalar multiplication is defined as $\alpha \cdot \langle x, y\rangle = \langle \alpha x, \alpha y\rangle$.
  - [A1]: We have $\langle x_1, y_1\rangle + \langle x_2, y_2\rangle = \langle x_1 + x_2, y_1 + y_2\rangle = \langle x_2, y_2\rangle + \langle x_1, y_1\rangle$.
  - [A2]: We have $(\langle x_1, y_1\rangle + \langle x_2, y_2\rangle) + \langle x_3, y_3\rangle = \langle x_1 + x_2 + x_3, y_1 + y_2 + y_3\rangle = \langle x_1, y_1\rangle + (\langle x_2, y_2\rangle + \langle x_3, y_3\rangle)$.
  - [A3]: The zero vector is $\langle 0, 0\rangle$, and clearly $\langle x, y\rangle + \langle 0, 0\rangle = \langle x, y\rangle$.
  - [A4]: The additive inverse of $\langle x, y\rangle$ is $\langle -x, -y\rangle$, since $\langle x, y\rangle + \langle -x, -y\rangle = \langle 0, 0\rangle$.
  - [M1]: We have $\alpha_1 \cdot (\alpha_2 \cdot \langle x, y\rangle) = \langle \alpha_1\alpha_2 x, \alpha_1\alpha_2 y\rangle = (\alpha_1\alpha_2) \cdot \langle x, y\rangle$.
  - [M2]: We have $(\alpha_1 + \alpha_2) \cdot \langle x, y\rangle = \langle (\alpha_1 + \alpha_2)x, (\alpha_1 + \alpha_2)y\rangle = \alpha_1 \cdot \langle x, y\rangle + \alpha_2 \cdot \langle x, y\rangle$.
  - [M3]: We have $\alpha \cdot (\langle x_1, y_1\rangle + \langle x_2, y_2\rangle) = \langle \alpha(x_1 + x_2), \alpha(y_1 + y_2)\rangle = \alpha \cdot \langle x_1, y_1\rangle + \alpha \cdot \langle x_2, y_2\rangle$.
  - [M4]: Finally, we have $1 \cdot \langle x, y\rangle = \langle x, y\rangle$.
- Example: The set of $m \times n$ matrices for any m and any n, forms a vector space.
  - 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 (the numbers of the form $a + bi$ for real a and b, and where $i^2 = -1$) are a vector space.
    - The axioms all follow from the standard properties of complex numbers. As might be expected, the "zero vector" is just the complex number $0 = 0 + 0i$.
    - 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: The collection of all real-valued functions on any part of the real line is a vector space, 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 real numbers. 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 $[a, b]$ and any functions f and g, we have
        - [A1]: $(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x)$.
        - [M2]: $\alpha \cdot (f + g)(x) = \alpha\, f(x) + \alpha\, g(x) = (\alpha f)(x) + (\alpha g)(x)$.
        - [M4]: $(1 \cdot f)(x) = f(x)$.
- Example: The "zero space" with a single element $\mathbf{0}$, with $\mathbf{0} + \mathbf{0} = \mathbf{0}$ and $\alpha \cdot \mathbf{0} = \mathbf{0}$ for every α, is a vector space.
    - All of the axioms in this case eventually boil down to $\mathbf{0} = \mathbf{0}$.
    - This space is rather boring: since it only contains one element, there's really not much to say about it.
- There are many simple algebraic properties that can be derived from the axioms, which necessarily hold in every vector space. For example:

    1. Addition has a cancellation law: for any vector $\mathbf{v}$, if $\mathbf{a} + \mathbf{v} = \mathbf{b} + \mathbf{v}$ then $\mathbf{a} = \mathbf{b}$.
        - By [A1]-[A4] we have $(\mathbf{a} + \mathbf{v}) + (-\mathbf{v}) = \mathbf{a} + (\mathbf{v} + (-\mathbf{v})) = \mathbf{a} + \mathbf{0} = \mathbf{a}$.
        - Similarly we also have $(\mathbf{b} + \mathbf{v}) + (-\mathbf{v}) = \mathbf{b} + (\mathbf{v} + (-\mathbf{v})) = \mathbf{b} + \mathbf{0} = \mathbf{b}$.
        - Finally, since $\mathbf{a} + \mathbf{v} = \mathbf{b} + \mathbf{v}$ then $\mathbf{a} = (\mathbf{a} + \mathbf{v}) + (-\mathbf{v}) = (\mathbf{b} + \mathbf{v}) + (-\mathbf{v}) = \mathbf{b}$ so $\mathbf{a} = \mathbf{b}$.
    2. The zero vector is unique: for any vector $\mathbf{v}$, if $\mathbf{a} + \mathbf{v} = \mathbf{v}$, then $\mathbf{a} = \mathbf{0}$.
        - This follows from property (1) applied when $\mathbf{b} = \mathbf{0}$, along with a use of [A3].
    3. The additive inverse is unique: for any vector $\mathbf{v}$, if $\mathbf{a} + \mathbf{v} = \mathbf{0}$ then $\mathbf{a} = -\mathbf{v}$.
        - This follows from property (1) applied when $\mathbf{b} = -\mathbf{v}$, along with a use of [A4].
    4. The scalar 0 times any vector gives the zero vector: $0 \cdot \mathbf{v} = \mathbf{0}$ for any vector $\mathbf{v}$.
        - Expand $\mathbf{v} = (1 + 0) \cdot \mathbf{v} = \mathbf{v} + 0 \cdot \mathbf{v}$ via [M2], [M4] and then apply property (2).
    5. Any scalar times the zero vector is the zero vector: $\alpha \cdot \mathbf{0} = \mathbf{0}$ for any scalar α.
        - Expand $\alpha \cdot \mathbf{0} = \alpha \cdot (\mathbf{0} + \mathbf{0}) = \alpha \cdot \mathbf{0} + \alpha \cdot \mathbf{0}$ via [M1] and then apply property (1).
    6. The scalar -1 times any vector gives the additive inverse: $(-1) \cdot \mathbf{v} = -\mathbf{v}$ for any vector $\mathbf{v}$.
        - Use property (3) and [M2]-[M4] to write $\mathbf{0} = 0 \cdot \mathbf{v} = (1 + (-1)) \cdot \mathbf{v} = \mathbf{v} + (-1) \cdot \mathbf{v}$, and then use property (1) with $\mathbf{a} = -\mathbf{v}$.
    7. The additive inverse of the additive inverse is the original vector: $-(-\mathbf{v}) = \mathbf{v}$ for any vector $\mathbf{v}$.
        - Idea: Use property (5) and [M1], [M4] to write $-(-\mathbf{v}) = (-1)^2 \cdot \mathbf{v} = 1 \cdot \mathbf{v} = \mathbf{v}$.

- At the moment, we cannot say very much about abstract properties of a general vector space.
  - It might seem that the axioms we have imposed do not really impose much structure aside from rather simple properties like the ones listed above: after all, each individual axiom does not say very much on its own.
  - But in fact, we will show that the axioms taken collectively force V to have a very strong and regular structure. In particular, we will be able to describe all of the elements of any vector space in a precise and simple way.

3.3 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.
- Example: Show that the set of diagonal 2×2 matrices is a subspace of the vector space of all 2×2 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: $p \cdot q \cdot \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} = \begin{bmatrix} pqa & 0 \\ 0 & pqb \end{bmatrix} = pq \cdot \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$.
- 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.
  - We will make use of this idea later when we talk about the solutions to a homogeneous linear differential equation: we will show that the solutions form a vector space merely by checking that they are a subspace of the vector space of all real-valued functions, rather than "reinventing the wheel" (so to speak) by going through all of the axioms individually.

3.3.1 The Subspace Criterion

- 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 $\mathbf{w}_1$ and $\mathbf{w}_2$ in W, the vector $\mathbf{w}_1 + \mathbf{w}_2$ is also in W.

- [S3] W is closed under scalar multiplication: For any scalar α and $\mathbf{w}$ in W, the vector $\alpha \cdot \mathbf{w}$ is also in W.

- The reason we don't need to check everything to verify that a collection of vectors forms a subspace is that most of the axioms will automatically be satisfied in W because they're true in V.
  - As long as all of the operations are defined, axioms [A1]-[A2] and [M1]-[M4] will hold in W because they hold in V. But we need to make sure that we can always add and scalar-multiply two elements of W and obtain a result that is in W, which is why we need [S2] and [S3].
  - In order to get axiom [A3] for W, we need to know that the zero vector is in W, which is why we need [S1].
  - In order to get axiom [A4] for W we can use the result that $(-1) \cdot \mathbf{w} = -\mathbf{w}$, to see that the closure under scalar multiplication automatically gives additive inverses.

- 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.

- By using the subspace criterion, it is straightforward to check whether a subset is actually a subspace. 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 $\boxed{\text{is a subspace}}$.

- 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 $2t_1 t_2 = 0$.)
  - 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 $\boxed{\text{not a subspace}}$.
  - Note that all we actually needed to do here was find a single counterexample, of which there are many. Had we noticed earlier that $\langle 1, 1 \rangle$ and $\langle 2, 4 \rangle$ were in the subset but their sum $\langle 3, 5 \rangle$ 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 $\langle s, t, 0 \rangle$ forms a subspace of $\mathbb{R}^3$.
  - We check the parts of the subspace criterion.
  - [S1]: The zero vector is of this form: take $s = t = 0$.
  - [S2]: We have $\langle s_1, t_1, 0 \rangle + \langle s_2, t_2, 0 \rangle = \langle s_1 + s_2, t_1 + t_2, 0 \rangle$, which is again of the same form, if we take $s = s_1 + s_2$ and $t = t_1 + t_2$.
  - [S3]: We have $\alpha \cdot \langle s_1, t_1, 0 \rangle = \langle \alpha s_1, \alpha t_1, 0 \rangle$, which is again of the same form, if we take $s = \alpha s_1$ and $t = \alpha t_1$.
  - All three parts are satisfied, so this subset $\boxed{\text{is a subspace}}$.

- Example: Determine whether the set of vectors of the form $\langle x, y, z \rangle$ where $2x - y + z = 0$ forms a subspace of $\mathbb{R}^3$.
    - [S1]: The zero vector is of this form, since $2(0) - 0 + 0 = 0$.
    - [S2]: If $\langle x_1, y_1, z_1 \rangle$ and $\langle x_2, y_2, z_2 \rangle$ have $2x_1 - y_1 + z_1 = 0$ and $2x_2 - y_2 + z_2 = 0$ then adding the equations shows that the sum $\langle x_1 + x_2, y_1 + y_2, z_1 + z_2 \rangle$ also lies in the space.
    - [S3]: If $\langle x_1, y_1, z_1 \rangle$ has $2x_1 - y_1 + z_1 = 0$ then scaling the equation by α shows that $\langle \alpha x_1, \alpha x_2, \alpha x_3 \rangle$ also lies in the space.
    - All three parts are satisfied, so this subset $\boxed{\text{is a subspace}}$.

- Example: Determine whether the set of vectors of the form $\langle x, y, z \rangle$ where $x, y, z \geq 0$ forms a subspace of $\mathbb{R}^3$.
    - [S1]: The zero vector is of this form: take $t = 0$.
    - [S2]: If $\langle x_1, y_1, z_1 \rangle$ and $\langle x_2, y_2, z_2 \rangle$ have $x_1, y_1, z_1 \geq 0$ and $x_2, y_2, z_2 \geq 0$, then $x_1 + x_2 \geq 0$, $y_1 + y_2 \geq 0$, and $z_1 + z_2 \geq 0$, so $\langle x_1 + x_2, y_1 + y_2, z_1 + z_2 \rangle$ also lies in the space.
    - [S3]: If $\langle x, y, z \rangle$ has $x, y, z \geq 0$, then it is not necessarily true that $\alpha x, \alpha y, \alpha z \geq 0$: specifically, this is not true when $\alpha = -1$.
    - From here we can find a specific counterexample: the vector $\langle 1, 1, 1 \rangle$ is in the subset, but the scalar multiple $-1 \cdot \langle 1, 1, 1 \rangle = \langle -1, -1, -1 \rangle$ is not in the subset. Thus, this subset is $\boxed{\text{not a subspace}}$.

- Example: Determine whether the set of 2×2 matrices of trace zero is a subspace of the space of all 2×2 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 αA.
    - All three parts are satisfied, so this subset $\boxed{\text{is a subspace}}$.

- Example: Determine whether the set of 2×2 matrices of determinant zero is a subspace of the space of all 2×2 matrices.
    - [S1]: The zero matrix has determinant zero.
    - [S3]: Since $\det(\alpha A) = \alpha^2 \det(A)$ when A is a 2×2 matrix, we see that if A has determinant zero then so does αA.
    - [S2]: If A and B have determinant zero, then there does not appear to be a nice way to compute the determinant of $A + B$ in general.
    - We can in fact find a counterexample: if $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ then $\det(A) = \det(B) = 0$, but $\det(A + B) = -1$. Thus, this subset is $\boxed{\text{not a subspace}}$.

3.3.2 Additional Examples of Subspaces

- Here are a few more examples of subspaces of vector spaces which will be of interest to us:
- Example: The collection of solution vectors $\langle x_1, \cdots, x_n \rangle$ to any homogeneous system of linear equations forms a subspace of $\mathbb{R}^n$.
    - It is possible to check this directly by working with equations. But it is much easier to use matrices: write the system in matrix form, as $A\mathbf{x} = \mathbf{0}$, where $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is a solution vector.
    - [S1]: We have $A\mathbf{0} = \mathbf{0}$, by the properties of the zero vector.

- [S2]: If $\mathbf{x}$ and $\mathbf{y}$ are two solutions, the properties of matrix arithmetic imply $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}$ so that $\mathbf{x}+\mathbf{y}$ is also a solution.
- [S3]: If α is a scalar and $\mathbf{x}$ is a solution, then $A(\alpha \cdot \mathbf{x}) = \alpha \cdot (A\mathbf{x}) = \alpha \cdot \mathbf{0} = \mathbf{0}$, so that $\alpha \cdot \mathbf{x}$ is also a solution.

- Example: The collection of continuous functions on $[a, b]$ is a subspace of the space of all functions on $[a, b]$.
  - [S1]: The zero function is continuous.
  - [S2]: The sum of two continuous functions is continuous, from basic calculus.
  - [S3]: The product of continuous functions is continuous, so in particular a constant times a continuous function is continuous.
- Example: The collection of n-times differentiable functions on $[a, b]$ is a subspace of the space of continuous functions on $[a, b]$, for any positive integer n.
  - The zero function is differentiable, as are the sum and product of any two functions which are differentiable n times.
- Example: The collection of all polynomials is a vector space.
  - Observe that polynomials are functions on the entire real line. Therefore, it is sufficient to verify the subspace criteria.
  - The zero function is a polynomial, as is the sum of two polynomials, and any scalar multiple of a polynomial.
- Example: The collection of 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 all 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 α is a scalar and y_1 is a solution, then scaling $y_1'' + 6y_1' + 5y_1 = 0$ by α and using properties of derivatives shows that $(\alpha y_1)'' + 6(\alpha y_1)' + 5(\alpha y_1) = 0$, so αy_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 any constants A and B. (From this description, if we wanted to, we could directly verify that such functions form a vector space.)
- This last example should help explain how the study of vector spaces and linear algebra is useful for the study of differential equations: namely, because the solutions to the given homogeneous linear differential equation form a vector space.
  - It is true more generally that the solutions to an arbitrary homogeneous linear differential equation $y_1^{(n)} + P_n(x) \cdot y_1^{(n-1)} + \cdots + P_1(x) \cdot y_1 = 0$ will form a vector space.
  - Most of the time we cannot explicitly write down the solutions to this differential equation; nevertheless, if we can understand the structure of a general vector space, we can still say something about what the solutions look like.

3.4 Span, Linear Independence, Bases, Dimension

- 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. In order to discuss this idea more precisely, we first need some terminology.

3.4.1 Linear Combinations and Span

- Definition: Given a set $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ of vectors in a vector space V, we say a vector $\mathbf{w}$ in V is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ if there exist scalars $a_1, \cdots, a_n$ such that $\mathbf{w} = a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n$.
  - Example: In $\mathbb{R}^2$, the vector $\langle 1,1\rangle$ is a linear combination of $\langle 1,0\rangle$ and $\langle 0,1\rangle$, because $\langle 1,1\rangle = 1\cdot\langle 1,0\rangle + 1\cdot\langle 0,1\rangle$.
  - Example: In $\mathbb{R}^4$, the vector $\langle 4,0,5,9\rangle$ is a linear combination of $\langle 1,0,0,1\rangle$, $\langle 0,1,0,0\rangle$, and $\langle 1,1,1,2\rangle$, because $\langle 4,0,5,9\rangle = 1\cdot\langle 1,-1,2,3\rangle - 2\cdot\langle 0,1,0,0\rangle + 3\cdot\langle 1,1,1,2\rangle$.
  - Non-Example: In $\mathbb{R}^3$, the vector $\langle 0,0,1\rangle$ is not a linear combination of $\langle 1,1,0\rangle$ and $\langle 0,1,1\rangle$ because there exist no scalars a_1 and a_2 for which $a_1\cdot\langle 1,1,0\rangle + a_2\cdot\langle 0,1,1\rangle = \langle 0,0,1\rangle$: 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 collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ in V, denoted $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$, to be the set W of all vectors which are linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$. Explicitly, the span is the set of vectors of the form $a_1\cdot\mathbf{v}_1 + \cdots + a_n\cdot\mathbf{v}_n$, for some scalars $a_1, \cdots, a_n$.
  - For technical reasons, we define the span of the empty set to be the zero vector.
- Example: The span of the vectors $\langle 1,0,0\rangle$ and $\langle 0,1,0\rangle$ in $\mathbb{R}^3$ is the set of vectors of the form $a\cdot\langle 1,0,0\rangle + b\cdot\langle 0,1,0\rangle = \langle a,b,0\rangle$.
  - Equivalently, the span of these vectors is the set of vectors whose z-coordinate is zero, which (geometrically) forms the plane $z = 0$.
- Example: Determine whether the vectors $\langle 2,3,3\rangle$ and $\langle 4,-1,3\rangle$ are in $\text{span}(\mathbf{v},\mathbf{w})$, where $\mathbf{v} = \langle 1,-1,2\rangle$ and $\mathbf{w} = \langle 2,1,-1\rangle$.
  - For $\langle 2,3,3\rangle$ we must determine whether it is possible to write $\langle 2,3,3\rangle = a\cdot\langle 1,-1,2\rangle + b\cdot\langle 2,1,-1\rangle$ for some a and b.
  - Equivalently, we want to solve the system $2 = a + 2b$, $3 = -a + b$, $3 = 2a - b$.
  - Row-reducing the associated coefficient matrix gives

$$\left[\begin{array}{cc|c} 1 & 2 & 2 \\ -1 & 1 & 3 \\ 2 & -1 & 3 \end{array}\right] \xrightarrow[R_3-2R_1]{R_2+R_1} \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 3 & 5 \\ 0 & -5 & -3 \end{array}\right] \xrightarrow{R_3+\frac{5}{3}R_1} \left[\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 3 & 5 \\ 0 & 0 & 16/3 \end{array}\right]$$

    and we obtain a contradiction. Thus, $\boxed{\langle 2,3,3\rangle \text{ is not in the span}}$.
  - Similarly, for $\langle 4,-1,3\rangle$ we want to solve $\langle 4,-1,3\rangle = c\cdot\langle 1,-1,2\rangle + d\cdot\langle 2,1,-1\rangle$.
  - Row-reducing the associated coefficient matrix gives

$$\left[\begin{array}{cc|c} 1 & 2 & 4 \\ -1 & 1 & -1 \\ 2 & -1 & 3 \end{array}\right] \xrightarrow[R_3-2R_1]{R_2+R_1} \left[\begin{array}{cc|c} 1 & 2 & 4 \\ 0 & 3 & 3 \\ 0 & -5 & -5 \end{array}\right] \xrightarrow{R_3+\frac{5}{3}R_1} \left[\begin{array}{cc|c} 1 & 2 & 4 \\ 0 & 3 & 3 \\ 0 & 0 & 0 \end{array}\right]$$

    from which we can easily obtain the solution $d = 1$, $c = 2$.
  - Since $\langle 4,-1,3\rangle = 2\cdot\langle 1,-1,2\rangle + 1\cdot\langle 2,1,-1\rangle$ we see that $\boxed{\langle 4,-1,3\rangle \text{ is in the span}}$.
- Here are some basic properties of the span:
- Proposition: For any vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ in V, the set $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$ is a subspace of V.
  - Proof: We check the subspace criterion.
  - [S1] The zero vector can be written as $0\cdot\mathbf{v}_1 + \cdots + 0\cdot\mathbf{v}_n$.
  - [S2] The span is closed under addition because $(a_1\cdot\mathbf{v}_1 + \cdots + a_n\cdot\mathbf{v}_n) + (b_1\cdot\mathbf{v}_1 + \cdots + b_n\cdot\mathbf{v}_n) = (a_1+b_1)\cdot\mathbf{v}_1 + \cdots + (a_n+b_n)\cdot\mathbf{v}_n$.

- [S3] The span is closed under scalar multiplication because $\alpha \cdot (a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = (\alpha a_1) \cdot \mathbf{v}_1 + \cdots + (\alpha a_n) \cdot \mathbf{v}_n$.

- Proposition: For any vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in V, if W is any subspace of V that contains $\mathbf{v}_1, \ldots, \mathbf{v}_n$, then W contains $\text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. In other words, the span is the smallest subspace containing the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$.
  - Proof: Consider any element of the span: by definition, it can be written as $w = a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n$ for some scalars $a_1, \cdots, a_n$.
  - Because W is a subspace, it is closed under scalar multiplication, so each of $a_1 \cdot \mathbf{v}_1$, $\cdots$, $a_n \cdot \mathbf{v}_n$ lies in W.
  - Furthermore, also because W is a subspace, it is closed under addition. Thus, the sum $a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n$ lies in W.
  - Thus, every element of the span lies in W, as claimed.

- Definition: Given a vector space V, if the span of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is all of V, we say that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a spanning set (or generating set) for V.
  - There are a number of different phrases we use for this idea: we also say that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span or generate the vector space V.
  - Spanning sets are very useful because they allow us to describe every vector in V in terms of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$.
  - Explicitly: if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ span V, then every vector in V is a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, which is to say, every vector $\mathbf{w}$ in V can be written in the form $\mathbf{w} = a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n$ for some scalars $a_1, \ldots, a_n$.

- Example: Show that the vectors $\langle 1, 0, 0 \rangle$, $\langle 0, 1, 0 \rangle$, and $\langle 0, 0, 1 \rangle$ span $\mathbb{R}^3$.
  - For any vector $\langle a, b, c \rangle$, we can write $\langle a, b, c \rangle = a \cdot \langle 1, 0, 0 \rangle + b \cdot \langle 0, 1, 0 \rangle + c \cdot \langle 0, 0, 1 \rangle$, so it is a linear combination of the three given vectors.

- Example: Show that the matrices $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ span the vector space of 2×2 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 2×2 matrices).
  - A 2×2 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: 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

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & p \\ 2 & 4 & 1 & q \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{ccc|c} 1 & 2 & 3 & p \\ 0 & 0 & -5 & q - 2p \end{array}\right]$$

    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 $\boxed{\text{do span } \mathbb{R}^2}$.

- 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

$$\left[\begin{array}{ccc|c} 1 & 1 & -1 & p \\ -1 & 0 & 2 & q \\ 3 & 1 & -5 & r \end{array}\right] \xrightarrow[R_3-3R_1]{R_2+R_1} \left[\begin{array}{ccc|c} 1 & 1 & -1 & p \\ 0 & 1 & 1 & q+p \\ 0 & -2 & -2 & r-3p \end{array}\right] \xrightarrow{R_3+2R_2} \left[\begin{array}{ccc|c} 1 & 1 & -1 & p \\ 0 & 1 & 1 & q+p \\ 0 & 0 & 0 & r+2q-p \end{array}\right].$$

  - Now, if $r + 2q - p \neq 0$, the final column will have a pivot and the system will be contradictory. 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 $\boxed{\text{do not span } \mathbb{R}^3}$.

- We can generalize the idea in the above examples to give a method for determining whether a collection of vectors in $\mathbb{R}^n$ will span $\mathbb{R}^n$.

- Theorem (Spanning Sets in $\mathbb{R}^n$): A collection of k vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ in $\mathbb{R}^n$ will span $\mathbb{R}^n$ if and only if, for every vector $\mathbf{b}$, there is at least one vector $\mathbf{x}$ satisfying the matrix equation $M\mathbf{x} = \mathbf{b}$, where M is the matrix whose columns are the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. Such a solution $\mathbf{x}$ exists for any $\mathbf{b}$ if and only if M has rank n: that is, when a row-echelon form of M has n pivotal columns.
  - Proof: Write each $\mathbf{v}_i = \begin{bmatrix} m_{1,i} \\ \vdots \\ m_{n,i} \end{bmatrix}$ as a column matrix.
  - Then $a_1 \cdot \mathbf{v}_1 + \cdots + a_k \cdot \mathbf{v}_k = \begin{bmatrix} m_{1,1} \\ \vdots \\ m_{n,1} \end{bmatrix} a_1 + \cdots + \begin{bmatrix} m_{1,k} \\ \vdots \\ m_{n,k} \end{bmatrix} a_k = M \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$, where $M = \begin{bmatrix} m_{1,1} & \cdots & m_{1,k} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,k} \end{bmatrix}$ is the matrix whose columns are the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$.
  - So the statement that, for any $\mathbf{b}$, there exist scalars $a_1, \dots, a_k$ such that $a_1 \cdot \mathbf{v}_1 + \cdots + a_k \cdot \mathbf{v}_k = \mathbf{b}$ is equivalent to the statement that there is a solution $\mathbf{x} = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$ to the matrix equation $M\mathbf{x} = \mathbf{b}$.
  - For the second part of the theorem, consider the matrix equation $M\mathbf{x} = \mathbf{b}$, and perform row operations to put M in row-echelon form.
  - By our theorems on systems of linear equations, this system will have at least one solution precisely when there is no pivot in the augmented column of coefficients.
  - Since $\mathbf{b}$ can be chosen arbitrarily, so can the column of constants in the augmented matrix once we put it in row-echelon form.
  - Since the augmented matrix has at most n pivots (since it has n rows), the only way we can prevent having a pivot in the column of constants is to have all the pivots in the matrix M itself: thus, M must have n pivots. From the definition of rank, this is equivalent to saying M has rank n.

3.4.2 Linear Independence and Linear Dependence

- Definition: We say a finite set of vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ is linearly independent if $a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n = \mathbf{0}$ implies $a_1 = \cdots = a_n = 0$. Otherwise, we say the collection is 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. We will generally focus only on finite sets of vectors in our discussion.

- In other words, $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent precisely when the only way to form the zero vector as a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_n$ is to have all the scalars equal to zero (the "trivial" linear combination). If there is a nontrivial linear combination giving the zero vector, then $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly dependent.
- An equivalent way of thinking of linear (in)dependence is that a set is dependent if one of the vectors is a linear combination of the others (i.e., it "depends" on the others). Explicitly, if $a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n = \mathbf{0}$ and $a_1 \neq 0$, then we can rearrange to see that $\mathbf{v}_1 = -\frac{1}{a_1}(a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n)$.

- Example: Determine whether the vectors $\langle 1, 1, 0 \rangle$, $\langle 0, 2, 1 \rangle$ in $\mathbb{R}^3$ are linearly dependent or linearly independent.
  - Suppose that we had scalars a and b with $a \cdot \langle 1, 1, 0 \rangle + b \cdot \langle 0, 2, 1 \rangle = \langle 0, 0, 0 \rangle$.
  - 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 $\boxed{\text{linearly independent}}$.

- Example: Determine whether the vectors $\langle 1, 1, 0 \rangle$, $\langle 2, 2, 0 \rangle$ in $\mathbb{R}^3$ are linearly dependent or linearly independent.
  - Suppose that we had scalars a and b with $a \cdot \langle 1, 1, 0 \rangle + b \cdot \langle 2, 2, 0 \rangle = \langle 0, 0, 0 \rangle$.
  - 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 $2 \cdot \langle 1, 1, 0 \rangle + (-1) \cdot \langle 2, 2, 0 \rangle = \langle 0, 0, 0 \rangle$, and this is a nontrivial linear combination giving the zero vector meaning that these vectors are $\boxed{\text{linearly dependent}}$.

- There is an easy general way to determine whether two vectors are linearly independent:

- Proposition: In any vector space V, the two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly dependent if one is a scalar multiple of the other, and they are linearly independent otherwise.
  - Proof: If $\mathbf{v}_1 = \alpha \cdot \mathbf{v}_2$ then we can write $1 \cdot \mathbf{v}_1 + (-\alpha) \cdot \mathbf{v}_2 = \mathbf{0}$, and similarly if $\mathbf{v}_2 = \alpha \cdot \mathbf{v}_1$ then we can write $(-\alpha) \cdot \mathbf{v}_1 + 1 \cdot \mathbf{v}_2 = \mathbf{0}$. In either case the vectors are linearly dependent.
  - If the vectors are dependent, then suppose $a \cdot \mathbf{v}_1 + b \cdot \mathbf{v}_2 = \mathbf{0}$ where a, b are not both zero. If $a \neq 0$ then we can write $\mathbf{v}_1 = (-b/a) \cdot \mathbf{v}_2$, and if $b \neq 0$ then we can write $\mathbf{v}_2 = (-a/b) \cdot \mathbf{v}_1$. At least one of these cases must occur, so one of the vectors is a scalar multiple of the other as claimed.

- It is more a delicate problem to determine whether a larger set of vectors is linearly independent.

- Example: Determine whether the vectors $\langle 1, 0, 2, 2 \rangle$, $\langle 2, -2, 3, 0 \rangle$, $\langle 0, 3, 1, 3 \rangle$, and $\langle 0, 4, 1, 2 \rangle$ in $\mathbb{R}^4$ are linearly dependent or linearly independent.
  - 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:

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 0 \\ 0 & -2 & 3 & 4 & 0 \\ 2 & 3 & 1 & 1 & 0 \\ 2 & 0 & 3 & 2 & 0 \end{array}\right] \xrightarrow[R_4 - 2R_1]{R_3 - 2R_1} \left[\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 0 \\ 0 & -2 & 3 & 4 & 0 \\ 0 & -1 & 1 & 1 & 0 \\ 0 & -4 & 3 & 2 & 0 \end{array}\right] \longrightarrow \cdots \longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

    from which we can obtain a nonzero solution $d = 1$, $c = -2$, $b = -1$, $a = 2$.
  - So we see $2 \cdot \langle 1, 0, 2, 2 \rangle + (-1) \cdot \langle 2, -2, 0, 3 \rangle + (-2) \cdot \langle 0, 3, 3, 1 \rangle + 1 \cdot \langle 0, 4, 2, 1 \rangle = \langle 0, 0, 0, 0 \rangle$, and this is a nontrivial linear combination giving the zero vector meaning that these vectors are $\boxed{\text{linearly dependent}}$.

- We can generalize the idea in the above example to give a method for determining whether a collection of vectors in $\mathbb{R}^n$ is linearly independent:

- Theorem (Dependence of Vectors in $\mathbb{R}^n$): A collection of k vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ in $\mathbb{R}^n$ is linearly dependent if and only if there is a nonzero vector $\mathbf{x}$ satisfying the matrix equation $M\mathbf{x} = \mathbf{0}$, where M is the matrix whose columns are the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$.

  - Proof: Write each $\mathbf{v}_i = \begin{bmatrix} m_{1,i} \\ \vdots \\ m_{n,i} \end{bmatrix}$ as a column matrix.
  - Then $a_1 \cdot \mathbf{v}_1 + \cdots + a_k \cdot \mathbf{v}_k = \begin{bmatrix} m_{1,1} \\ \vdots \\ m_{n,1} \end{bmatrix} a_1 + \cdots + \begin{bmatrix} m_{1,k} \\ \vdots \\ m_{n,k} \end{bmatrix} a_k = M \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$, where $M = \begin{bmatrix} m_{1,1} & \cdots & m_{1,k} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,k} \end{bmatrix}$ is the matrix whose columns are the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$.
  - So the linear combination $a_1 \cdot \mathbf{v}_1 + \cdots + a_k \cdot \mathbf{v}_k$ is the zero vector precisely when the matrix product $M\mathbf{x} = \mathbf{0}$, where $\mathbf{x} = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$.
  - By definition, the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ will be linearly dependent when there is a nonzero $\mathbf{x}$ satisfying this matrix equation, and they will be linearly independent when the only solution is $\mathbf{x} = \mathbf{0}$.

- We can also ask about linear independence of functions:

- Example: Determine whether the functions e^x and e^{2x} are linearly independent in the vector space of all real-valued functions.

  - Suppose that we had scalars a and b with $a \cdot e^x + b \cdot e^{2x} = 0$ for all x.
  - Taking the derivative of both sides with respect to x yields $a \cdot e^x + b \cdot 2e^{2x} = 0$.
  - Subtracting the original equation from this one produces $b \cdot e^{2x} = 0$, and since x is a variable we must have $b = 0$.
  - The first equation then gives $a \cdot e^x = 0$ so it must also be true that $a = 0$.
  - Thus, by definition, these functions are $\boxed{\text{linearly independent}}$.

- We can generalize the idea in the above example to give a method for determining whether a collection of functions is linearly independent:

- Definition: For n functions $y_1(x), y_2(x), \cdots, y_n(x)$ which are each differentiable $n-1$ times, their Wronskian is defined to be $W(y_1, y_2, \cdots, y_n) = \det \begin{bmatrix} y_1 & y_2 & \cdots & y_n \\ y_1' & y_2' & \cdots & y_n' \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(n-1)} & y_2^{(n-1)} & \cdots & y_n^{(n-1)} \end{bmatrix}$. (Note that the Wronskian will also be a function of x.)

  - Example: The Wronskian of e^x, e^{2x} is $W(e^x, e^{2x}) = \begin{vmatrix} e^x & e^{2x} \\ e^x & 2e^{2x} \end{vmatrix} = e^{3x}$.

- Theorem (Linear Independence of Functions): Suppose that n functions $y_1, y_2, \dots, y_n$ which are each differentiable $n-1$ times have a Wronskian that is not the zero function. Then the functions are linearly independent in the vector space of real-valued functions.

  - Proof: Suppose that the functions are linearly dependent with $\sum_{j=1}^{n} a_j y_j = 0$, then by differentiating the appropriate number of times we see that $\sum_{j=1}^{n} a_j y_j^{(i)} = 0$ for any $0 \le i \le n$.
  - Hence, in particular, the rows of the Wronskian matrix are linearly dependent (as vectors), and so the determinant of the matrix is zero.

- Therefore, if the Wronskian determinant is *not* zero, the functions cannot be dependent.
- Important Warning: The converse of this theorem is false! As shown below, there exist examples of linearly independent functions $y_1, y_2, \cdots, y_n$ whose Wronskian is the zero function.

- Example: Show that the functions $\sin(x)$ and $\cos(x)$ are linearly independent using the Wronskian.
  - We compute $W(\sin(x), \cos(x)) = \begin{vmatrix} \sin(x) & \cos(x) \\ \cos(x) & -\sin(x) \end{vmatrix} = -1$, which is not the zero function.

- Example: Determine whether the functions $1 + x$, $2 - x$, and $3 + 4x$ are linearly dependent or linearly independent.
  - We compute $W(1, x, 1 + x) = \begin{vmatrix} 1+x & 2-x & 3+4x \\ 1 & -1 & 4 \\ 0 & 0 & 0 \end{vmatrix} = 0$, by expanding along the bottom row.
  - Because these functions are infinitely differentiable and have Wronskian equal to the zero function, they are $\boxed{\text{linearly dependent}}$.
  - A little searching will produce the explicit linear dependence $-11(1 + x) + (2 - x) + 3(3 + 4x) = 0$.

- Example: Compute the Wronskian of the functions x^2 and $x\,|x|$. Are they linearly dependent or linearly independent?
  - We compute $W(x^2, x\,|x|) = \begin{vmatrix} x^2 & x\,|x| \\ 2x & 2\,|x| \end{vmatrix} = 2x^2\,|x| - 2x^2\,|x| = 0$, which is the zero function. (One can verify using the definition of the derivative that $x\,|x|$ is differentiable everywhere and that its derivative is $2\,|x|$.)
  - This suggests these functions are linearly dependent. But in fact, they are linearly independent: if $a \cdot x^2 + b \cdot x\,|x| = 0$, then setting $x = 1$ produces $a + b = 0$ and setting $x = -1$ produces $a - b = 0$, and the only solution is $a = b = 0$.
  - A natural guess is that the issue occurs because $x\,|x|$ is not twice-differentiable. However, there are versions of this counterexample involving functions that are infinitely differentiable.

- 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): If the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly dependent, then for some $1 \le i \le n$ the vector $\mathbf{v}_i$ can be written as a linear combination of the vectors $\mathbf{v}_1, \ldots \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_n$.
  - Proof: By the definition of linear dependence, there is some dependence $a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n = \mathbf{0}$ where not all of the coefficients are zero: say, specifically, that $a_i \neq 0$.
  - Then we can rearrange the statement $a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n = \mathbf{0}$ to read

$$a_i \cdot \mathbf{v}_i = (-a_1) \cdot \mathbf{v}_1 + \cdots + (-a_{i-1}) \cdot \mathbf{v}_{i-1} + (-a_{i+1}) \cdot \mathbf{v}_{i+1} + \cdots + (-a_n) \cdot \mathbf{v}_n$$

    and upon multiplying by $1/a_i$ we see that

$$\mathbf{v}_i = (-\frac{a_1}{a_i}) \cdot \mathbf{v}_1 + \cdots + (-\frac{a_{i-1}}{a_i}) \cdot \mathbf{v}_{i-1} + (-\frac{a_{i+1}}{a_i}) \cdot \mathbf{v}_{i+1} + \cdots + (-\frac{a_n}{a_i}) \cdot \mathbf{v}_n.$$

  - Thus, $\mathbf{v}_i$ can be written as a linear combination of the other vectors, 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 $a \langle 1, -1 \rangle + b \langle 2, 2 \rangle + c \langle 2, 1 \rangle = \langle 0, 0 \rangle$.
  - By row-reducing the appropriate matrix we can find the solution $2 \langle 1, -1 \rangle + 3 \langle 2, 2 \rangle - 4 \langle 2, 1 \rangle = \langle 0, 0 \rangle$.
  - By rearranging we can then write $\boxed{\langle 1, -1 \rangle = -\frac{3}{2} \langle 2, 2 \rangle + 2 \langle 2, 1 \rangle}$.

- Here is a useful result about the span of a linearly independent set of vectors:
- Theorem (Characterization of Linear Independence): The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent if and only if every vector $\mathbf{w}$ in the span of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ may be *uniquely* written as a sum $\mathbf{w} = a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n$: that is, the scalars $a_1, a_2, \ldots, a_n$ are unique.
  - Proof: First suppose the decomposition is always unique: then $a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n = \mathbf{0}$ implies $a_1 = \cdots = a_n = 0$, because $0 \cdot \mathbf{v}_1 + \cdots + 0 \cdot \mathbf{v}_n = \mathbf{0}$ is by assumption the only decomposition of $\mathbf{0}$. So we see that the vectors are linearly independent.
  - Now suppose that we had two ways of decomposing a vector $\mathbf{w}$, say as $\mathbf{w} = a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n$ and as $\mathbf{w} = b_1 \cdot \mathbf{v}_1 + \cdots + b_n \cdot \mathbf{v}_n$.
  - Then subtracting and then rearranging the difference between these two equations yields $\mathbf{w} - \mathbf{w} = (a_1 - b_1) \cdot \mathbf{v}_1 + \cdots + (a_n - b_n) \cdot \mathbf{v}_n$.
  - Now $\mathbf{w} - \mathbf{w}$ is the zero vector, so we have $(a_1 - b_1) \cdot \mathbf{v}_1 + \cdots + (a_n - b_n) \cdot \mathbf{v}_n = \mathbf{0}$.
  - But now because $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 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.

3.4.3 Bases and Dimension

- We will now combine the ideas of a spanning set and a linearly independent set:
- 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.
  - Furthermore, 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 2×3 (real) matrices.
  - A general 2×3 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 $\boxed{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}$ as a basis.

- Indeed, they certainly span the space of all 2×3 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 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ forms a basis of the vector space V if and only if every vector $\mathbf{w}$ in V can be uniquely written in the form $\mathbf{w} = a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n$ for scalars $a_1, a_2, \dots, a_n$.
  - In particular, this proposition says that if we have a basis $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ for V, then we can describe all of the other vectors in V in a particularly simple way (as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$) that is *unique*. A useful way to interpret this idea is to think of the basis vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ as "coordinate directions" and the coefficients $a_1, a_2, \dots, a_n$ as "coordinates".
  - Proof: Suppose $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ is a basis of V. Then by definition, the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ span the vector space V: every vector $\mathbf{w}$ in V can be written in the form $\mathbf{w} = a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n$ for some scalars $a_1, a_2, \dots, a_n$.
  - Furthermore, since the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are linearly independent, by our earlier proposition every vector $\mathbf{w}$ in their span (which is to say, every vector in V) can be uniquely written in the form $\mathbf{w} = a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n$, as claimed.
  - Conversely, suppose every vector $\mathbf{w}$ in V can be uniquely written in the form $\mathbf{w} = a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n$. Then by definition, the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ span V.
  - Furthermore, by our earlier proposition, because every vector in $\mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ can be uniquely written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$, the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 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 polynomials $p(x)$ of degree ≤ 3 such that $p(1) = 0$.
  - We remark that W is a subspace of the vector space V of polynomials, as it satisfies the subspace criterion. (We omit the verification.)
  - A polynomial of degree ≤ 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 $\boxed{\{x^3 - 1,\ x^2 - 1,\ x - 1\}}$ is a basis of W.

- A basis from 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.
  - Proof: We show that we can "reduce" any spanning set by removing elements until the spanning set becomes linearly independent.
  - In the event that the spanning set is infinite, the argument is rather delicate and technical, so we will only treat the case of a finite spanning set consisting of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.
  - Start with an empty collection S of elements. Now, for each $1 \le i \le n$, perform the following procedure:
    * Check whether the vector $\mathbf{v}_i$ is contained in the span of S. (Note that the span of the empty set is the zero vector.)
    * If $\mathbf{v}_i$ is not in the span of S, then add it to S. Otherwise, do nothing.
  - We claim that at the end of this procedure, S 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 see that S is linearly independent, observe that if $\mathbf{v}_i$ is added to the set S, then $\mathbf{v}_i$ is linearly independent from the vectors already in S (as it is not in the span of S). Thus, each vector added to S preserves the linear independence of S, so when the procedure terminates S will be linearly independent.
  - To see that S spans V, the idea is to observe that the span of S is the same as the span of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.
  - Explicitly, consider any vector $\mathbf{v}_i$ that is not in S: it was not added to S during the construction of S, so it must have been contained in the span of the vectors already in S. Therefore, adding $\mathbf{v}_i$ to S will not change the span. Doing this for each vector $\mathbf{v}_i$ not in S will not change the span and will yield the set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, so we conclude $\text{span}(S) = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = V$.
- We can also obtain a basis of a vector space by building up a linearly-independent set of vectors:
- Theorem (Independent Sets and Bases): 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: Let S be a set of linearly independent vectors. (In any vector space, the empty set is always linearly independent.)
    1. If S spans V, then we are done, because then S is a linearly independent generating set; i.e., a basis.
    2. If S does not span V, there is an element $\mathbf{v}$ in V which is not in the span of S. Put $\mathbf{v}$ 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 we will always be able to construct a basis 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.
- By building up from the empty set, we can construct a basis for any vector space. Furthermore, any two bases have the same size:
- Theorem (Bases of Vector Spaces): Every vector space V has a basis, and any two bases of V contain the same number of elements.
  - Remark: That a basis always exists is incredibly useful, and it 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 then we know exactly what the elements of the vector space look like.
  - Proof: To see that every vector space has a basis, let S be any set of linearly independent vectors in V. (One possibility is to take S to be the empty set.) Then since S is linearly independent, there exists a basis of V containing S by our earlier result.

- Another method would be to take S to be any spanning set for V. (One possibility is to take S to be the set of all vectors in V.) Then since S is linearly independent, it contains a basis of V by our earlier result.
- For the second part of the theorem, we will show that if A is a set of vectors with m elements and B is a basis with n elements, with $m > n$, then A is linearly dependent.
- This implies the stated result because A cannot be linearly independent and hence not be a basis if it has more elements than B does. The same argument applies if we interchange the roles of A and B, so A and B must have the same number of elements.
- So suppose B is a basis with n elements and A is a set of m vectors with $m > n$. Then by definition we can write every element $\mathbf{a}_i$ in A as a linear combination of the elements of B, say as $\mathbf{a}_i = \sum_{j=1}^{n} c_{i,j} \cdot \mathbf{b}_j$ for $1 \le i \le m$.
- We would like to see that there is some choice of scalars d_k, not all zero, such that $\sum_{k=1}^{n} d_k \cdot \mathbf{a}_k = \mathbf{0}$: this will show that the vectors $\mathbf{a}_i$ are linearly dependent.
- So consider a linear conbination $\sum_{k=1}^{n} d_k \cdot \mathbf{a}_k = \mathbf{0}$: if we substitute in for the vectors in B, then we obtain a linear combination of the elements of B equalling the zero vector. Since B is a basis, this means each coefficient of $\mathbf{b}_j$ in the resulting expression must be zero.
- If we tabulate the resulting system, we can check that it is equivalent to the matrix equation $C\mathbf{d} = \mathbf{0}$, where C is the $m \times n$ matrix of coefficients with entries $c_{i,j}$, and $\mathbf{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix}$ is the $n \times 1$ matrix with entries the scalars d_k.
- Since C is a matrix which has more rows than columns, by the assumption that $m > n$, we see that the homogeneous system $C\mathbf{d} = \mathbf{0}$ has a nonzero solution for $\mathbf{d}$. (There is at most one pivot per column, and so there must be at least one row that does not have a pivot.)
- But then we have $\sum_{k=1}^{n} d_k \cdot \mathbf{a}_k = \mathbf{0}$ for some scalars d_k not all of which are zero, meaning that the set A is linearly dependent.

- 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 $\dim(V)$.
  - The theorem above assures us that this quantity is always well-defined: every vector space has a basis, and any other basis will have the same number of elements.
- Example: The dimension of $\mathbb{R}^n$ is n, since the n standard unit vectors form a basis.
  - This says 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 ∞, because the (infinite list of) polynomials $1, x, x^2, x^3, \cdots$ are a basis for the space.
- Proposition: If W is a subspace of V, then $\dim(W) \le \dim(V)$.
  - Proof: Consider any basis of W. It is a linearly independent set of vectors in V, so it is contained in some basis of V. The result follows immediately.

3.4.4 Finding Bases for $\mathbb{R}^n$, Row Spaces, Column Spaces, and Nullspaces

- The fact that every vector space has a basis is extremely useful from a theoretical standpoint. We will now discuss some practical methods for finding bases for particular vector spaces that often arise in linear algebra.
  - Our results are very explicit and give two different methods for constructing a basis for a given vector space.
  - One way is to "build" a linearly independent set of vectors into a basis by adding new vectors one at a time (choosing a vector not in the span of the previous vectors) until a basis is obtained.
  - Another way is to "reduce" a spanning set by removing linearly dependent vectors one at a time (finding and removing a vector that is a linear combination of the others) until a basis is obtained.
- Proposition (Bases, Span, Dependence): If V is an n-dimensional vector space, then any set of fewer than n vectors cannot span V, and any set of more than n vectors is linearly dependent.
  - Proof: Suppose first that S is a set of fewer than n vectors in V.
  - Then since S spans span(S) by definition, S contains a basis T for span(S), and T is a linearly independent set of fewer than n vectors in V.
  - Thus, we can extend T to a basis of V, which necessarily contains n elements, strictly more than in T. So there is some vector $\mathbf{v}$ in this extended basis that is not in T: then $\mathbf{v}$ is not in span(S), so S does not span V.
  - Now suppose that S is a set of more than n vectors in V that is linearly independent. We would then be able to extend S to a basis of V, but this is impossible because any basis contains only n elements.
- Example: Determine whether the vectors $\langle 1,2,2,1\rangle$, $\langle 3,-1,2,0\rangle$, $\langle -3,2,1,1\rangle$ span $\mathbb{R}^4$.
  - They $\boxed{\text{do not span}}$: since $\mathbb{R}^4$ is a 4-dimensional space, any spanning set must contain at least 4 vectors.
- Example: Determine whether the vectors $\langle 1,2,1\rangle$, $\langle 1,0,3\rangle$, $\langle -3,2,1\rangle$, $\langle 1,1,4\rangle$ are linearly independent.
  - They $\boxed{\text{are not linearly independent}}$: since $\mathbb{R}^3$ is a 3-dimensional space, any 4 vectors in $\mathbb{R}^3$ are automatically linearly dependent.
- We can also characterize bases of $\mathbb{R}^n$:
- Theorem (Bases of $\mathbb{R}^n$): A collection of k vectors $\mathbf{v}_1,\dots,\mathbf{v}_k$ in $\mathbb{R}^n$ is a basis if and only if $k=n$ and the $n\times n$ matrix B, whose columns are the vectors $\mathbf{v}_1,\dots,\mathbf{v}_n$, is an invertible matrix.
  - Remark: The statement that B is invertible is equivalent to saying that $\det(B)\neq 0$. This gives a rapid computational method for determining whether a given set of vectors forms a basis.
  - Proof: Since $\mathbb{R}^n$ has a basis with n elements, any basis must have n elements by our earlier results, so $k=n$.
  - Now suppose $\mathbf{v}_1,\dots,\mathbf{v}_n$ are vectors in $\mathbb{R}^n$. For any vector $\mathbf{w}$ in $\mathbb{R}^n$, consider the problem of finding scalars $a_1,\cdots,a_n$ such that $a_1\cdot\mathbf{v}_1+\cdots+a_n\cdot\mathbf{v}_n=\mathbf{w}$.
  - This vector equation is the same as the matrix equation $B\mathbf{a}=\mathbf{w}$, where B is the matrix whose columns are the vectors $\mathbf{v}_1,\dots,\mathbf{v}_n$, $\mathbf{a}$ is the column vector whose entries are the scalars $a_1,\dots,a_n$, and $\mathbf{w}$ is thought of as a column vector.
  - By our earlier results, $\mathbf{v}_1,\dots,\mathbf{v}_n$ is a basis of $\mathbb{R}^n$ precisely when the scalars $a_1\dots,a_n$ are unique. In turn this is equivalent to the statement that $B\mathbf{a}=\mathbf{w}$ has a unique solution $\mathbf{a}$ for any $\mathbf{w}$.
  - From our study of matrix equations, this equation has a unique solution precisely when B is an invertible matrix, as claimed.
- Example: Determine whether the vectors $\langle 1,2,1\rangle$, $\langle 2,-1,2\rangle$, $\langle 3,3,1\rangle$ form a basis of $\mathbb{R}^3$.
  - By the theorem, we only need to determine whether the matrix $M=\begin{bmatrix}1&2&3\\2&-1&3\\1&2&1\end{bmatrix}$ is invertible.

- We compute $\det(M) = 1\begin{vmatrix} -1 & 3 \\ 2 & 1 \end{vmatrix} - 2\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} + 3\begin{vmatrix} 2 & -1 \\ 1 & 2 \end{vmatrix} = 10$ which is nonzero.
- Thus, M is invertible, so these vectors $\boxed{\text{do form a basis of } \mathbb{R}^3}$.

- Associated to any matrix M are three spaces that often arise when doing matrix algebra and studying the solutions to systems of linear equations.

- Definition: If M is an $m \times n$ matrix, the row space of M is the subspace of $\mathbb{R}^n$ spanned by the rows of M, the column space of M is the subspace of $\mathbb{R}^m$ spanned by the columns of M, and the nullspace of M is the set of vectors $\mathbf{x}$ in $\mathbb{R}^n$ for which $M\mathbf{x} = \mathbf{0}$.
  - By definition the row space and column spaces are subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, since the span of any set of vectors is a subspace.
  - It is also simple to verify that the nullspace is a subspace of $\mathbb{R}^m$ via the subspace criterion.

- We have already studied in detail the procedure for solving a matrix equation $M\mathbf{x} = \mathbf{0}$, which requires row-reducing the matrix M. It turns out that we can obtain a basis for the row and column spaces from a row-echelon form of M as well:

- Theorem (Bases for Row and Column Spaces): If M is an $m \times n$ matrix, let E be any row-echelon form of M. If k is the number of pivots in E, then the row space and column space are both k-dimensional and the nullspace is $(n-k)$-dimensional. Furthermore, a basis for the row space is given by the nonzero rows of E, while a basis for the column space is given by the columns of M that correspond to the pivotal columns of E.
  - For the column space, we also remark that another option would be to row-reduce the transpose matrix M^T, since the columns of M are the rows of M^T. This in general will produce a basis that is easier to work with, but it is not actually necessary.
  - Proof: First consider the row space, which by definition is spanned by the rows of M.
  - Observe that each elementary row operation does not change the span of the rows of M: for any vectors $\mathbf{v}_i$ and $\mathbf{v}_j$, we have $\text{span}(\mathbf{v}_j, \mathbf{v}_i) = \text{span}(\mathbf{v}_j, \mathbf{v}_i)$, $\text{span}(r\mathbf{v}) = \text{span}(\mathbf{v})$ for any nonzero r, and $\text{span}(\mathbf{v}_i, \mathbf{v}_j) = \text{span}(\mathbf{v}_i + a\mathbf{v}_j, \mathbf{v}_j)$ for any a.
  - So we may put M into a row-echelon form E without altering the span. Now we claim that the nonzero rows $\mathbf{r}_1, \dots, \mathbf{r}_k$ of E are linearly independent. Ultimately, this is because of the presence of the pivot elements: if $a_1 \cdot \mathbf{r}_1 + \cdots + a_k \cdot \mathbf{r}_k = \mathbf{0}$ then each of the vectors $\mathbf{r}_1, \dots, \mathbf{r}_k$ will have a leading coefficient in an entry that is zero in all of the subsequent vectors, so the only solution to the associated system of linear equations is $a_1 = \cdots = a_k = 0$.
  - Now consider the column space. Observe first that the set of solutions $\mathbf{x}$ to the matrix equation $M\mathbf{x} = \mathbf{0}$ is the same as the set of solutions to the equation $E\mathbf{x} = \mathbf{0}$ (by our analysis of row-operations).
  - Now if we write $\mathbf{x} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and expand out each matrix product in terms of the columns of M and E, we will see that $M\mathbf{x} = a_1 \cdot \mathbf{c}_1 + \cdots + a_n \cdot \mathbf{c}_n$ and $E\mathbf{x} = a_1 \cdot \mathbf{e}_1 + \cdots + a_n \cdot \mathbf{e}_n$ where the $\mathbf{c}_i$ are the columns of M and the $\mathbf{e}_i$ are the columns of E.
  - Combining these two observations shows that, for any scalars $a_1, \dots, a_n$, we have $a_1 \cdot \mathbf{c}_1 + \cdots + a_n \cdot \mathbf{c}_n = \mathbf{0}$ if and only if $a_1 \cdot \mathbf{e}_1 + \cdots + a_n \cdot \mathbf{e}_n = \mathbf{0}$.
  - What this means is that any linear dependence between the columns of M gives a linear dependence between the corresponding columns of E (with the same coefficients), and vice versa. So it is enough to determine a basis for the column space of E: then a basis for the column space of M is simply the corresponding columns in M.
  - All that remains is to observe that the set of pivotal columns for E forms a basis for the column space of E: the pivotal columns are linearly independent by the same argument given above for rows, and every other column lies in their span (specifically, any column lies in the span of the pivotal columns to its left, since each row has a pivotal element).

- Finally, the statement about the dimensions of the row and column spaces follows immediately from our descriptions, and the statement about the dimension of the nullspace follows by observing that the matrix equation $M\mathbf{x} = \mathbf{0}$ has k bound variables and $n-k$ free variables.

- Example: Find a basis for the row space, the column space, and the nullspace of the matrix $M = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix}$, as well as the dimension of each space.
  - We begin by row-reducing the matrix M:
  $$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix} \xrightarrow{R_3-R_1} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & 1 & -1 & 2 \end{bmatrix} \xrightarrow{R_3-R_2} \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$
  - The row space is spanned by the two vectors $\boxed{\langle 1,0,2,1\rangle\,,\ \langle 0,1,-1,2\rangle}$.
  - Since columns 1 and 2 have pivots, the first two columns of M give a basis for the column space: $\boxed{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}$.
  - For the nullspace, there are two free variables corresponding to columns 3 and 4. Solving the corresponding system (with variables x_1, x_2, x_3, x_4 and free parameters a, b) yields the solution set $\langle x_1, x_2, x_3, x_4\rangle = \langle -2a-b, a-2b, a, b\rangle = a\langle -2,1,1,0\rangle + b\langle -1,-2,0,1\rangle$.
  - Thus, a basis for the nullspace is given by $\boxed{\langle -2,1,1,0\rangle\,,\ \langle -1,-2,0,1\rangle}$.
  - The row space, column space, and nullspace all have dimension $\boxed{2}$.

- Example: Find a basis for the row space, the column space, and the nullspace of $M = \begin{bmatrix} 1 & -1 & 0 & 2 & 1 \\ -2 & 2 & 0 & -3 & 1 \\ 1 & -1 & 0 & 3 & 8 \end{bmatrix}$ as well as the dimension of each space.
  - We begin by row-reducing the matrix M:
  $$\begin{bmatrix} 1 & -1 & 0 & 2 & 1 \\ -2 & 2 & 0 & -3 & 1 \\ 1 & -1 & 0 & 3 & 8 \end{bmatrix} \underset{R_3-R_1}{\xrightarrow{R_2+2R_1}} \begin{bmatrix} 1 & -1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 2 & 7 \end{bmatrix} \xrightarrow{R_3-2R_2} \begin{bmatrix} 1 & -1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$
  - The row space is spanned by the three vectors $\boxed{\langle 1,-1,0,2,1\rangle\,,\ \langle 0,0,0,1,3\rangle\,,\ \langle 0,0,0,0,1\rangle}$.
  - Since there are pivots in columns 1, 4, and 5, those columns of M give a basis for the column space: $\boxed{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 8 \end{bmatrix}}$.
  - For the nullspace, there are two free variables corresponding to columns 2 and 3. Solving the corresponding system (with variables x_1, x_2, x_3, x_4, x_5 and free parameters a, b) yields the solution set $\langle x_1, x_2, x_3, x_4, x_5\rangle = \langle a, a, b, 0, 0\rangle = a\langle 1,1,0,0,0\rangle + b\langle 0,0,1,0,0\rangle$, so a basis for the nullspace is given by $\boxed{\langle 1,1,0,0,0\rangle\,,\ \langle 0,0,1,0,0\rangle}$.

3.5 Linear Transformations

- Now that we have a reasonably good idea of what the structure of a vector space is, the next natural question is: what do maps from one vector space to another look like?
- It turns out that we don't want to ask about arbitrary functions, but about functions from one vector space to another which preserve the structure (namely, addition and scalar multiplication) of the vector space.

3.5.1 Definition and Examples

- Definition: If V and W are vector spaces, we say a map T from V to W (denoted $T : V \to W$) is a linear transformation if, for any vectors $\mathbf{v}$, $\mathbf{v}_1$, $\mathbf{v}_2$ and any scalar α, the following two properties hold:
  - [T1] The map respects addition of vectors: $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$.
  - [T2] The map respects scalar multiplication: $T(\alpha \cdot \mathbf{v}) = \alpha \cdot \mathbf{v}$.

- Warning: It is important to note that in the statement $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$, the addition on the left-hand side is taking place inside V, whereas the addition on the right-hand side is taking place inside W. Likewise, in the statement $T(\alpha \cdot \mathbf{v}) = \alpha \cdot \mathbf{v}$, the scalar multiplication on the left-hand side is in V while the scalar multiplication on the right-hand side is in W.

- Example: If $V = W = \mathbb{R}^2$, show that the map T which sends $\langle x, y\rangle$ to $\langle x, x + y\rangle$ is a linear transformation from V to W .
  - We simply check the two parts of the definition.
  - Let $\mathbf{v} = \langle x, y\rangle$, $\mathbf{v}_1 = \langle x_1, y_1\rangle$, and $\mathbf{v}_2 = \langle x_2, y_2\rangle$, so that $\mathbf{v}_1 + \mathbf{v}_2 = \langle x_1 + x_2, y_1 + y_2\rangle$.
  - [T1]: We have $T(\mathbf{v}_1+\mathbf{v}_2) = \langle x_1 + x_2, x_1 + x_2 + y_1 + y_2\rangle = \langle x_1, x_1 + y_1\rangle+\langle x_2, x_2 + y_2\rangle = T(\mathbf{v}_1)+T(\mathbf{v}_2)$.
  - [T2]: We have $T(\alpha \cdot \mathbf{v}) = \langle \alpha x, \alpha x + \alpha y\rangle = \alpha \cdot \langle x, x + y\rangle = \alpha \cdot T(\mathbf{v})$.

- Example: If $V = M_{2\times 2}(\mathbb{R})$ and $W = \mathbb{R}$, determine whether the trace map is a linear transformation from V to W .
  - Let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $M_1 = \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}$, $M_2 = \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}$ so $M_1 + M_2 = \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{bmatrix}$.
  - [T1]: We have $\mathrm{tr}(M_1 + M_2) = (a_1 + a_2) + (d_1 + d_2) = (a_1 + d_1) + (a_2 + d_2) = \mathrm{tr}(M_1) + \mathrm{tr}(M_2)$.
  - [T2]: We have $\mathrm{tr}(\alpha \cdot M) = \alpha a + \alpha d = \alpha \cdot (a + d) = \alpha \cdot \mathrm{tr}(M)$.
  - Both parts of the definition are satisfied, so the trace $\boxed{\text{is a linear transformation}}$.

- Example: If $V = M_{2\times 2}(\mathbb{R})$ and $W = \mathbb{R}$, determine whether the determinant map is a linear transformation from V to W.
  - Let $M_1 = \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}$, $M_2 = \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix}$ so $M_1 + M_2 = \begin{bmatrix} a_1 + a_2 & b_1 + b_2 \\ c_1 + c_2 & d_1 + d_2 \end{bmatrix}$.
  - [T1]: We have $\det(M_1 + M_2) = (a_1 + a_2)(d_1 + d_2) - (b_1 + b_2)(c_1 + c_2)$, while $\det(M_1) + \det(M_2) = (a_1d_1 - b_1c_1) + (a_2d_2 - b_2c_2)$.
  - When we expand out the products in $\det(M_1 + M_2)$ we will quickly see that the expression is not the same as $\det(M_1) + \det(M_2)$.
  - An explicit example is $M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $M_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$: $\det(M_1) = \det(M_2) = 0$, while $M_1 + M_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ has determinant 1.
  - The first part of the definition does not hold, so the determinant $\boxed{\text{is not a linear transformation}}$. (In fact, the second part fails as well.)

- Example: If V is the vector space of all differentiable functions and W is the vector space of all functions, determine whether the derivative map D sending a function to its derivative is a linear transformation from V to W.
  - [T1]: We have $D(f_1 + f_2) = (f_1 + f_2)' = f_1' + f_2' = D(f_1) + D(f_2)$.
  - [T2]: Also, $D(\alpha \cdot f) = (\alpha f)' = \alpha \cdot f' = \alpha \cdot D(f)$.
  - Since both parts of the definition are satisfied, the derivative $\boxed{\text{is a linear transformation}}$.

- Remark: More generally, on the vector space of n-times differentiable functions, the map T which sends a function y to the function $y^{(n)} + P_n(x)\, y^{(n-1)} + \cdots + P_2(x)\, y' + P_1(x)\, y$ is a linear transformation, for any functions $P_n(x), \cdots, P_1(x)$.

- Like with the definition of a vector space, we can prove a few simple algebraic properties of linear transformations:

- Proposition: Any linear transformation $T : V \to W$ sends the zero vector of V to the zero vector of W.
  - Proof: Let $\mathbf{v}$ be any vector in V. Since $0 \cdot \mathbf{v} = \mathbf{0}_V$ from basic properties, applying [T2] yields $0 \cdot T(\mathbf{v}) = T(\mathbf{0}_V)$.
  - But $0 \cdot T(\mathbf{v}) = \mathbf{0}_W$ since scaling any vector of W by 0 gives the zero vector of W.
  - Combining these two statements gives $T(\mathbf{0}_V) = 0 \cdot T(\mathbf{v}) = \mathbf{0}_W$, so $T(\mathbf{0}_V) = \mathbf{0}_W$ as claimed.

- Proposition: For any vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and any scalars $a_1, \ldots, a_n$, if T is a linear transformation then $T(a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n) = a_1 \cdot T(\mathbf{v}_1) + \cdots + a_n \cdot T(\mathbf{v}_n)$.
  - This result says that linear transformations can be moved through linear combinations.
  - Proof: By applying [T1] repeatedly, we see that $T(a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n) = T(a_1 \cdot \mathbf{v}_1) + \cdots + T(a_n \cdot \mathbf{v}_n)$.
  - Then by [T2], we have $T(a_i \cdot \mathbf{v}_i) = a_i \cdot T(\mathbf{v}_i)$ for each $1 \le i \le n$.
  - Plugging these relations into the first equation gives $T(a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n) = a_1 \cdot T(\mathbf{v}_1) + \cdots + a_n \cdot T(\mathbf{v}_n)$ as required.

- Proposition: Any linear transformation is uniquely defined by its values on a basis of V.
  - Proof: If V is finite-dimensional, let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be a basis of V. Then any vector $\mathbf{v}$ in V can be written as $\mathbf{v} = a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n$ for some scalars $a_1, \ldots, a_n$.
  - By the previous proposition, we then have $T(\mathbf{v}) = a_1 T(\mathbf{v}_1) + a_2 T(\mathbf{v}_2) + \cdots + a_n T(\mathbf{v}_n)$, so the value of T on any vector $\mathbf{v}$ in V is uniquely determined by the values of $T(\mathbf{v}_1)$, $T(\mathbf{v}_2)$, ... , $T(\mathbf{v}_n)$.
  - Conversely, for any specified values $\mathbf{w}_1, \ldots, \mathbf{w}_n$, it is straightforward to check that the map $T : V \to W$ defined by $T(a_1 \cdot \mathbf{v}_1 + a_2 \cdot \mathbf{v}_2 + \cdots + a_n \cdot \mathbf{v}_n) = a_1 \cdot \mathbf{w}_1 + \cdots + a_n + \mathbf{w}_n$ is a linear transformation.

- Example: If V is the vector space of polynomials of degree ≤ 2 and $T : V \to \mathbb{R}$ is the linear transformation such that $T(1) = 5$, $T(1 + x) = 4$, and $T(2 + x^2) = 3$, find $T(4 + 2x + 2x^2)$.
  - We simply need to express $4 + 2x + 2x^2$ in terms of the basis $\{1, 1 + x, 2 + x^2\}$ for V.
  - A straightforward calculation shows $4 + 2x + 2x^2 = -2(1) + 2(1 + x) + 2(2 + x^2)$.
  - Thus, $T(4 + 2x + 2x^2) = -2T(1) + 2T(1 + x) + 2T(2 + x^2) = -2(5) + 2(4) + 2(3) = \boxed{4}$.

- Here are a few more examples of linear transformations:

- Example: If $V = W = \mathbb{R}^2$, then the map T which sends $\langle x, y\rangle$ to $\langle ax + by, cx + dy\rangle$for any a, b, c, d is a linear transformation.
  - We could simply work out the calculations explicitly. But another way we can think of this map is as a matrix map: T sends the column vector $\begin{bmatrix} x \\ y \end{bmatrix}$ to the column vector $\begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$.
  - So, in fact, this map T is really just left-multiplication by the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. When we think of the map in this way, it is easier to see what is happening:
  - [T1]: We have $T(\mathbf{v}_1 + \mathbf{v}_2) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} (\mathbf{v}_1 + \mathbf{v}_2) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mathbf{v}_1 + \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mathbf{v}_2 = T(\mathbf{v}_1) + T(\mathbf{v}_2)$.
  - [T2]: Also, $T(\alpha \cdot \mathbf{v}) = \begin{bmatrix} a & b \\ c & d \end{bmatrix} (\alpha \mathbf{v}) = \alpha \cdot \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mathbf{v} \right) = \alpha \cdot T(\mathbf{v})$.

- Example: If $V = \mathbb{R}^m$ (thought of as $m \times 1$ matrices) and $W = \mathbb{R}^n$ (thought of as $n \times 1$ matrices) and A is any $n \times m$ matrix, then the map T sending $\mathbf{v}$ to $A\mathbf{v}$ is a linear transformation.
  - The verification is exactly the same as in the previous example.
  - [T1]: We have $T(\mathbf{v}_1 + \mathbf{v}_2) = A(\mathbf{v}_1 + \mathbf{v}_2) = A\mathbf{v}_1 + A\mathbf{v}_2 = T(\mathbf{v}_1) + T(\mathbf{v}_2)$.
  - [T2]: Also, $T(\alpha \cdot \mathbf{v}) = A \cdot \alpha\mathbf{v} = \alpha \cdot (A \cdot \mathbf{v}) = \alpha \cdot T(\mathbf{v})$.
- This last example is very general: in fact, it is so general that *every* linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ is of this form!
- Theorem (Linear Transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$): If T is a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, then there is some $m \times n$ matrix A such that $T(\mathbf{v}) = A\mathbf{v}$ (where we think of $\mathbf{v}$ as a column matrix).
  - This result may seem unexpected at first glance, but it is quite natural: ultimately the idea is that the matrix A is the $m \times n$ matrix whose rows are the vectors $T(e_1)$, $T(e_2)$, $\dots$, $T(e_m)$, where $e_1, \cdots , e_m$ are the standard basis elements of $\mathbb{R}^m$ (e_j is the vector with a 1 in the jth position and 0s elsewhere).
  - Proof: Let the $m \times n$ matrix whose rows are the vectors $T(e_1)$, $T(e_2)$, $\dots$, $T(e_m)$, where $e_1, \cdots , e_m$ are the standard basis elements of $\mathbb{R}^m$ (e_j is the vector with a 1 in the jth position and 0s elsewhere).
  - We claim that for any vector $\mathbf{v}$, $T(\mathbf{v}) = A\mathbf{v}$.
  - To see this, write $\mathbf{v}$ as a linear combination $\mathbf{v} = \displaystyle\sum_{j=1}^{m} a_j \cdot e_j$ of the basis elements.
  - Then since T is a linear transformation, we see that $T(\mathbf{v}) = \displaystyle\sum_{j=1}^{m} a_j \cdot T(e_j)$. If we write down this map one coordinate at a time, we will see that it agrees with the result of computing the matrix product of the matrix A with the "coordinates" of $\mathbf{v}$.
  - Remark: If we write down the map T explicitly, we see that the term in each coordinate in W is a linear function of the coordinates in V: for example, if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ then the linear functions are $ax + by$ and $cx + dy$. This is the reason that "linear transformations" are named so: they are the natural extension of the idea of a linear function.
- This theorem underlines one of the reasons that matrices and vector spaces, which initially seem like they have almost nothing to do with one another, are in fact very closely related: matrices describe the linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. (In fact we will soon generalize this statement even further.)
  - Using this relationship between maps on vector spaces and matrices, it is possible to provide almost trivial proofs of some of the algebraic properties of matrix multiplication which are hard to prove by direct computation.
  - Specifically, the idea is that multiplication of matrices can be viewed as a special case of the composition of functions, so anything that holds for general function composition will hold for matrix multiplication.
  - For example: linear transformations are functions, and function composition is associative. Since multiplication of matrices is a special case of function composition, multiplication of matrices is associative.

3.5.2 Kernel and Image

- We will now study a pair of important subspaces associated to a linear transformation.
- Definition: If $T : V \to W$ is a linear transformation, then the kernel of T, denoted $\ker(T)$, is the set of elements $\mathbf{v}$ in V with $T(\mathbf{v}) = \mathbf{0}$.
  - The kernel is the elements which are sent to zero by T.
  - In the event that $T : \mathbb{R}^n \to \mathbb{R}^m$ is multiplication by a matrix A, then a vector $\mathbf{x}$ is in the kernel precisely when $A\mathbf{x} = \mathbf{0}$: in other words, the kernel of T is the nullspace of the matrix A.

- In general, it is useful to view the kernel as a generalization of the nullspace to arbitrary linear transformations.

- Definition: If $T : V \to W$ is a linear transformation, then image of T (often also called the range of T), denoted $\text{im}(T)$, is the set of elements w in W such that there exists a $\mathbf{v}$ in V with $T(\mathbf{v}) = \mathbf{w}$.
  - The image is the elements in W which can be obtained as output from T.
  - In the event that $T : \mathbb{R}^n \to \mathbb{R}^m$ is multiplication by a matrix A, then a vector $\mathbf{b}$ is in the image precisely when there is a solution $\mathbf{x}$ to the matrix equation $A\mathbf{x} = \mathbf{b}$: in other words, the image of T is the column space of the matrix A.

- The kernel and image are subspaces of the appropriate vector space:

- Proposition: The kernel is a subspace of V.
  - [S1] We have $T(\mathbf{0}) = \mathbf{0}$, by simple properties of linear transformations.
  - [S2] If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in the kernel, then $T(\mathbf{v}_1) = \mathbf{0}$ and $T(\mathbf{v}_2) = \mathbf{0}$. Therefore, $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0}$.
  - [S3] If $\mathbf{v}$ is in the kernel, then $T(\mathbf{v}) = \mathbf{0}$. Hence $T(\alpha \cdot \mathbf{v}) = \alpha \cdot T(\mathbf{v}) = \alpha \cdot \mathbf{0} = \mathbf{0}$.

- Proposition: The image is a subspace of W.
  - [S1] We have $T(\mathbf{0}) = \mathbf{0}$, by simple properties of linear transformations.
  - [S2] If $\mathbf{w}_1$ and $\mathbf{w}_2$ are in the image, then there exist $\mathbf{v}_1$ and $\mathbf{v}_2$ such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Then $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$, so that $\mathbf{w}_1 + \mathbf{w}_2$ is also in the image.
  - [S3] If $\mathbf{w}$ is in the image, then there exists $\mathbf{v}$ with $T(\mathbf{v}) = \mathbf{w}$. Then $T(\alpha \cdot \mathbf{v}) = \alpha \cdot T(\mathbf{v}) = \alpha \cdot \mathbf{w}$, so $\alpha \cdot \mathbf{w}$ is also in the image.

- There is a straightforward way to find a spanning set for the image of a linear transformation:

- Proposition: If $\mathbf{v}_1, \dots, \mathbf{v}_n$ is a basis for V, then $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ spans the image of any linear transformation T.
  - Note that in general the vectors $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ are not necessarily a basis for the image since they need not be linearly independent. (But we have already discussed methods for converting a spanning set into a basis, so it is not hard to find an actual basis for the image.)
  - Proof: Suppose $\mathbf{w}$ is in the image of T. Then by hypothesis, $\mathbf{w} = T(\mathbf{v})$ for some vector $\mathbf{v}$.
  - Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ is a basis for V, there are scalars $a_1, \dots, a_n$ such that $\mathbf{v} = a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n$.
  - Then $\mathbf{w} = T(\mathbf{v}) = a_1 \cdot T(\mathbf{v}_1) + \cdots + a_n \cdot T(\mathbf{v}_n)$ is a linear combination of $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$, so it lies in their span. This is true for any $\mathbf{w}$ in the image of T, so $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ spans the image of T as claimed.
  - Remark: It is natural to wonder whether there is an equally simple way to find a spanning set for the kernel of a linear transformation: unfortunately, there is not. For matrix maps, however, the kernel is the same as the nullspace, so we can compute it using row reductions.

- Example: If $T : \mathbb{R}^2 \to \mathbb{R}^3$ is the linear transformation with $T(x, y) = (x + y, 0, 2x + 2y)$, find a basis for the kernel and for the image of T.
  - For the kernel, we want to find all (x, y, z) such that $T(x, y) = (0, 0, 0)$, so we obtain the three equations $x + y = 0$, $0 = 0$, $2x + 2y = 0$. These equations collectively say $y = -x$, so we see that the kernel is the set of vectors of the form $\langle x, -x \rangle = x \cdot \langle 1, -1 \rangle$, so a basis for the kernel is given by the single vector $\boxed{\langle 1, -1 \rangle}$.
  - For the image, by the proposition above it is enough simply to find the span of $T(\mathbf{v}_1)$, $T(\mathbf{v}_2)$ where $\mathbf{v}_1$ and $\mathbf{v}_2$ are a basis for $\mathbb{R}^2$. Using the standard basis, we compute $T(1, 0) = \langle 1, 0, 2 \rangle$ and $T(0, 1) = \langle 1, 0, 2 \rangle$, so a basis for the image is given by the single vector $\boxed{\langle 1, 0, 2 \rangle}$.

- Proposition (Kernel and One-to-One Maps): For any linear transformation $T : V \to W$, the kernel $\ker(T)$ consists of only the zero vector if and only if the map T is one-to-one: that is, if $T(\mathbf{v}_1) = T(\mathbf{v}_2)$ implies $\mathbf{v}_1 = \mathbf{v}_2$.
  - A one-to-one linear transformation sends different vectors in V to different vectors in W. A one-to-one function of a real variable is one that passes the "vertical line test", and thus has an inverse function f^{-1}.
  - Proof: If T is one-to-one, then (at most) one element of V maps to $\mathbf{0}$. But since the zero vector of V is taken to the zero vector of W, we see that T cannot send anything else to $\mathbf{0}$. Thus $\ker(T) = \{\mathbf{0}\}$.
  - Conversely, if $\ker(T)$ is only the zero vector, then since T is a linear transformation, the statement $T(\mathbf{v}_1) = T(\mathbf{v}_2)$ is equivalent to the statement that $T(\mathbf{v}_1) - T(\mathbf{v}_2) = T(\mathbf{v}_1 - \mathbf{v}_2)$ is the zero vector.
  - But, by the definition of the kernel, $T(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$ precisely when $\mathbf{v}_1 - \mathbf{v}_2$ is in the kernel. However, this means $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$, so $\mathbf{v}_1 = \mathbf{v}_2$. Hence $T(\mathbf{v}_1) = T(\mathbf{v}_2)$ implies $\mathbf{v}_1 = \mathbf{v}_2$, which means T is one-to-one.
- We can give some intuitive explanations for what the kernel and image are measuring.
  - The image of a linear transformation measures how close the map is to giving all of W as output: a linear transformation with a large image hits most of W, while a linear transformation with a small image misses most of W.
  - The kernel of a linear transformation measures how close the map is to being the zero map: a linear transformation with a large kernel sends many vectors to zero, while a linear transformation with a small kernel sends few vectors to zero.
  - We can quantify these notions of "large" and "small" using dimension:
- Definitions: The dimension of $\ker(T)$ is called the nullity of T, and the dimension of $\text{im}(T)$ is called the rank of T.
  - A linear transformation with a large nullity has a large kernel, which means it sends many elements to zero (hence "nullity").
  - There is a very important relationship between the rank and the nullity of a linear transformation:
- Theorem (Nullity-Rank): For any linear transformation $T : V \to W$, $\dim(\ker(T)) + \dim(\text{im}(T)) = \dim(V)$. In words, the nullity plus the rank is equal to the dimension of V.
  - Proof: Let $\mathbf{w}_1, \dots, \mathbf{w}_k$ be a basis for $\text{im}(T)$ in W.
  - Then by the definition of the image, there exist $\mathbf{v}_1, \dots, \mathbf{v}_k$ in V such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for each $1 \le i \le k$.
  - Also let $\mathbf{a}_1, \dots, \mathbf{a}_l$ be a basis for $\ker(T)$. We claim that the set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{a}_1, \dots, \mathbf{a}_l\}$ is a basis for V.
  - To show this, let $\mathbf{v}$ be an element of V. Then since $T(\mathbf{v})$ lies in $\text{im}(T)$, there exist unique scalars $\beta_1, \dots, \beta_k$ such that $T(\mathbf{v}) = \sum_{j=1}^{k} \beta_j \cdot \mathbf{w}_j$.
  - By properties of linear transformations, we then can write

$$T\left(\mathbf{v} - \sum_{j=1}^{k} \beta_j \cdot \mathbf{v}_j\right) = T(\mathbf{v}) - \sum_{j=1}^{k} \beta_j \cdot T(\mathbf{v}_j) = \sum_{j=1}^{k} \beta_j \cdot \mathbf{w}_j - \sum_{j=1}^{k} \beta_j \cdot \mathbf{w}_j = \mathbf{0}.$$

  - Therefore, $\mathbf{v} - \sum_{j=1}^{k} \beta_j \cdot \mathbf{v}_j$ is in $\ker(T)$, so it can be written as a sum $\sum_{i=1}^{l} \gamma_i \cdot \mathbf{a}_i$ for unique scalars γ_i.
  - Putting all this together shows $\mathbf{v} = \sum_{j=1}^{k} \beta_j \cdot \mathbf{v}_j + \sum_{i=1}^{l} \gamma_i \cdot \mathbf{a}_i$ for scalars β_j and γ_i, which says that $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{a}_1, \dots, \mathbf{a}_l\}$ spans V.

- Now suppose we had a dependence $\mathbf{0} = \sum_{j=1}^{k} \beta_j \cdot \mathbf{v}_j + \sum_{i=1}^{l} \gamma_i \cdot \mathbf{a}_i$.
- Applying T to both sides yields $\mathbf{0} = T(\mathbf{0}) = \sum_{j=1}^{k} \beta_j \cdot T(\mathbf{v}_j) + \sum_{i=1}^{l} \gamma_i \cdot T(\mathbf{a}_i) = \sum_{j=1}^{k} \beta_j \cdot \mathbf{w}_j$.
- Since the $\mathbf{w}_j$ are linearly independent we conclude that all the coefficients β_j must be zero.
- We then obtain the relation $\mathbf{0} = \sum_{i=1}^{l} \gamma_i \cdot \mathbf{a}_i$, but now since the $\mathbf{a}_i$ are linearly independent, we conclude that all the coefficients γ_i must also be zero.
- We conclude that $\{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{a}_1, \dots, \mathbf{a}_l\}$ is linearly independent, so since it also spans V, it is a basis for V.

- In the event that the linear transformation is multiplication by a matrix, the nullity-rank theorem reduces to a fact we already knew.
  - Explicitly, if A is an $m \times n$ matrix, the kernel of the multiplication-by-A map is the solution space to the homogeneous system $A\mathbf{x} = \mathbf{0}$, and the image is the set of vectors $\mathbf{c}$ such that there exists a solution to $A\mathbf{x} = \mathbf{c}$.
  - The value of $\dim(\ker(T))$ is the size of a basis for the solutions to the homogeneous equation (i.e., the nullspace), which we know is the number of nonpivotal columns in the reduced row-echelon form of A.
  - The value of $\dim(\operatorname{im}(T))$ is the size of a basis for the collection of row vectors of A, since the row vectors span the image. So the dimension of $\operatorname{im}(T)$ is the number of pivotal columns in the reduced row-echelon form of A.
  - Therefore, the sum of these two numbers is the number of columns of the matrix A, since every column is either pivotal or nonpivotal, which is simply n.
  - Incidentally, we also see that the use of the word "rank" for the the dimension of $\operatorname{im}(T)$ is consistent with our use of the word "rank" to refer to the rank of a matrix (since the rank of a matrix is the same as the number of pivot elements in its row-echelon form).

- Example: If $T : M_{2\times 2}(\mathbb{R}) \to \mathbb{R}$ is the trace map, find the nullity and the rank of T and verify the nullity-rank theorem.
  - We have $T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a + d$.
  - First, we compute the kernel: we see that $T\begin{pmatrix} a & b \\ c & d \end{pmatrix} = 0$ when $d = -a$, so the elements of the kernel have the form $\begin{bmatrix} a & b \\ c & -a \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$.
  - So the kernel has a basis given by the three matrices $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, meaning that the nullity is $\boxed{3}$.
  - For the image, we can clearly obtain any value in $\mathbb{R}$, since $T\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} = a$ for any a. So the image is 1-dimensional, meaning that the rank is $\boxed{1}$.
  - Then the rank plus the nullity is 4, which (per the theorem) is indeed equal to the dimension of the space of 2×2 matrices.

3.5.3 Isomorphisms of Vector Spaces, Matrices Associated to Linear Transformations

- We will now discuss a notion of equivalence of vector spaces.

- Definition: A linear transformation $T : V \to W$ is called an isomorphism if T is one-to-one and onto. Equivalently, T is an isomorphism if $\ker(T) = 0$ and $\mathrm{im}(T) = W$. We say that two vector spaces are isomorphic if there exists an isomorphism between them.
  - Saying that two spaces are isomorphic is a very strong statement: it says that the spaces V and W have exactly the same structure, as vector spaces.
  - Informally, saying that $T : V \to W$ is an isomorphism means that we can use T to relabel the elements of V to have the same names as the elements of W, and that (if we do so) we cannot tell V and W apart at all.
- Example: The space $\mathbb{R}^4$ is isomorphic to the space $M_{2\times 2}$ of 2×2 matrices, with an isomorphism T given by $T(x_1, x_2, x_3, x_4) = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$.
  - This map is a linear transformation; it clearly is additive and respects scalar multiplication.
  - Also, $\ker(T) = 0$ since the only element mapping to the zero matrix is $(0, 0, 0, 0)$. And it is also clear that $\mathrm{im}(T) = M_{2\times 2}$.
  - Thus T is an isomorphism.
- Proposition: If $T : V \to W$ is an isomorphism, the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ in V are linearly independent if and only if $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ are linearly independent in W.
  - Proof: Because T is a linear transformation, we have $a_1 \cdot T(\mathbf{v}_1) + \cdots + a_n \cdot T(\mathbf{v}_n) = T(a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n)$ for any scalars $a_1, \dots, a_n$.
  - To see that $\mathbf{v}_1, \dots, \mathbf{v}_n$ independent implies $T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)$ independent:
    * If $a_1 \cdot T(\mathbf{v}_1) + \cdots + a_n \cdot T(\mathbf{v}_n) = \mathbf{0}$, then by the above we have $T(a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n) = \mathbf{0}$.
    * But now since $\ker(T) = 0$, we get $a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n = \mathbf{0}$, and independence of $\mathbf{v}_1, \dots, \mathbf{v}_n$ then gives $a_1 = \cdots = a_n = 0$.
    * So $T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)$ are linearly independent.
  - To see that $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ independent implies $\mathbf{v}_1, \dots, \mathbf{v}_n$ independent:
    * If $a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n = \mathbf{0}$, then $a_1 \cdot T(\mathbf{v}_1) + \cdots + a_n \cdot T(\mathbf{v}_n) = T(a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n) = T(\mathbf{0}) = \mathbf{0}$.
    * But now the linear independence of $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ gives $a_1 = \cdots = a_n = 0$, so $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent.
- Proposition (Inverse Maps): If T is an isomorphism, then there exists an inverse function $T^{-1} : W \to V$, with $T^{-1}(T(\mathbf{v})) = \mathbf{v}$ and $T(T^{-1}(\mathbf{w})) = \mathbf{w}$ for any $\mathbf{v}$ in V and $\mathbf{w}$ in W. This inverse map $T^{-1}$is also a linear transformation.
  - Proof: The fact that there is an inverse function $T^{-1} : W \to V$ follows immediately because T is one-to-one and onto.
  - Specifically, for any $\mathbf{w}$ in W, by the assumption that T is onto there exists a $\mathbf{v}$ in V with $T(\mathbf{v}) = \mathbf{w}$, and because T is one-to-one, this vector $\mathbf{v}$ is unique. We then define $T^{-1}(\mathbf{w}) = \mathbf{v}$.
  - Now we check the two properties of a linear transformation:
    * [T1] If $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$, then because $T(\mathbf{v}_1 + \mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$, we have $T^{-1}(\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v}_1 + \mathbf{v}_2 = T^{-1}(\mathbf{w}_1) + T^{-1}(\mathbf{w}_2)$.
    * [T2] If $T(\mathbf{v}) = \mathbf{w}$, then because $T(\alpha \cdot \mathbf{v}) = \alpha \cdot \mathbf{w}$, we have $T^{-1}(\alpha \cdot \mathbf{w}) = \alpha \cdot \mathbf{v} = \alpha \cdot T^{-1}(\mathbf{w})$.
- It may seem that isomorphisms are hard to find, but this is not the case.
- Theorem (Isomorphism and Dimension): Two (finite-dimensional) vector spaces V and W are isomorphic if and only if they have the same dimension. In particular, any finite-dimensional vector space is isomorphic to $\mathbb{R}^n$ for some value of n.
  - Remark: This result should be rather unexpected: it certainly doesn't seem obvious, just from the eight axioms of a vector space, that all finite-dimensional vector spaces are essentially "the same" as $\mathbb{R}^n$ for some n. But they are!

- Proof: Isomorphisms preserve linear independence, so two vector spaces can only be isomorphic if they have the same dimension.
- For the other direction, choose a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ for V and a basis $\mathbf{w}_1, \dots, \mathbf{w}_n$ for W. We claim the map T defined by $T(a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n) = a_1 \cdot \mathbf{w}_1 + \cdots + a_1 \cdot \mathbf{w}_n$ is an isomorphism between V and W.
- We need to check five things: that T is unambiguously defined, that T respects addition, that T respects scalar multiplication, that T is one-to-one, and that T is onto.
  - [Well-defined]: We need to make sure that we have not made the definition of T ambiguous: namely, that we have defined T on every element of V, and that we haven't tried to send one element of V to two different elements of W. However, we are safe because $\mathbf{v}_1, \dots, \mathbf{v}_n$ is a basis, which means that for every $\mathbf{v}$ in V, we have a unique way of writing $\mathbf{v}$ as a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_n$.
  - [Addition]: If $\mathbf{v} = a_1 \cdot \mathbf{v}_1 + \cdots + a_1 \cdot \mathbf{v}_n$ and $\tilde{\mathbf{v}} = b_1 \cdot \mathbf{v}_1 + \cdots + b_n \cdot \mathbf{v}_n$, then $T(\mathbf{v} + \tilde{\mathbf{v}}) = (a_1 + b_1) \cdot \mathbf{w}_1 + \cdots + (a_n + b_n) \cdot \mathbf{w}_n = T(\mathbf{v}) + T(\tilde{\mathbf{v}})$ by the distributive law.
  - [Multiplication]: For any scalar β we have $T(\beta \cdot \mathbf{v}) = (\beta a_1) \cdot \mathbf{w}_1 + \cdots + (\beta a_n) \cdot \mathbf{w}_n = \beta \cdot T(\mathbf{v})$ by consistency of multiplication.
  - [One-to-one]: Since $\mathbf{w}_1, \dots, \mathbf{w}_n$ are linearly independent, the only way that $a_1 \cdot \mathbf{w}_1 + \cdots + a_1 \cdot \mathbf{w}_n$ can be the zero vector is if $a_1 = a_2 = \cdots = a_n = 0$, which means $\ker(T) = 0$.
  - [Onto]: Since $\mathbf{w}_1, \dots, \mathbf{w}_n$ span W, every element $\mathbf{w}$ in W can be written as $\mathbf{w} = a_1 \cdot \mathbf{w}_1 + \cdots + a_1 \cdot \mathbf{w}_n$ for some scalars $a_1, \cdots a_n$. Then for $\mathbf{v} = a_1 \cdot \mathbf{v}_1 + \cdots + a_1 \cdot \mathbf{v}_n$, we have $T(\mathbf{v}) = \mathbf{w}$.

- Earlier, we showed that matrices completely describe the linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$. But by the theorem above, every finite-dimensional vector space is isomorphic to $\mathbb{R}^n$ for some value of n: so in fact, matrices will completely describe the linear transformations from any n-dimensional vector space to any m-dimensional vector space.
  - More specifically: let V be any m-dimensional vector space W be any n-dimensional vector space. Choose a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ for V and a basis $\mathbf{w}_1, \dots, \mathbf{w}_m$ for W, and let $\mathbf{v}$ be any element of V.
  - If we write $\mathbf{v} = a_1 \cdot \mathbf{v}_1 + \cdots + a_n \cdot \mathbf{v}_n$ as a linear combination of the basis elements $\mathbf{v}_1, \dots, \mathbf{v}_m$ for V and we write $T(\mathbf{v}) = b_1 \cdot \mathbf{w}_1 + b_2 \cdot \mathbf{w}_2 + \cdots + b_n \cdot \mathbf{w}_m$ as a linear combination of the basis elements $\mathbf{w}_1, \dots, \mathbf{w}_m$ for W then, by combining our theorems, we conclude that there is a matrix C such that $C\mathbf{a} = \mathbf{b}$, where $C = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m,1} & c_{m,1} & \cdots & c_{m,n} \end{bmatrix}$, $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$.
  - What this means is that the coefficients of $\mathbf{v}$ in terms of the basis $\mathbf{v}_1, \dots, \mathbf{v}_m$ and the coefficients of $T(\mathbf{v})$ in terms of the basis $\mathbf{w}_1, \dots, \mathbf{w}_m$ are related by multiplication by the matrix C. In a very concrete sense, the linear transformation behaves in exactly the same way as multiplication by the matrix C.
  - This matrix C is called the matrix associated to the linear transformation T. It implicitly depends on the two bases $\mathbf{v}_1, \dots, \mathbf{v}_n$ and $\mathbf{w}_1, \dots, \mathbf{w}_m$ that we choose for V and W: making a different choice for either basis will yield a different matrix C.

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