Linear Algebra (part 2) : Linear Transformations (by Evan Dummit, 2022, v. 3.00)

Contents

2 Linear Transformations

In this chapter we will study linear transformations, which are structure-preserving maps between vector spaces. Such maps generalize the idea of a linear function and share many properties with linear functions from $\mathbb{R}^m$ to $\mathbb{R}^n$.

We begin by studying linear transformations (in general) and two important subspaces associated to a linear transformation, the kernel and the image. Next, we study the space of linear transformations from one vector space to another, and characterize some algebraic properties of linear transformations. We also analyze invertible linear transformations and isomorphisms and then apply these ideas to establish the rather stunning result that any finite-dimensional F-vector space has structure identical to to the vector space F^n.

We conclude with a lengthy exploration of the various relationships between linear transformations and matrices, and use our understanding of bases to give a concrete "matrix representation" of a linear transformation in the finite-dimensional case. We also analyze the behavior of these matrix representations under change of basis and the correspondence between isomorphisms and invertible matrices.

2.1 Linear Transformations

- Now that we have a reasonably good idea of what a general vector space looks like, the next natural question is: what do maps from one vector space to another look like?
  - Here, 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.

2.1.1 Definition and Basic Examples

- Definition: If V and W are vector spaces having the same scalar field F, we say a function T from V to W (denoted $T : V \to W$) is a linear transformation if the following two properties hold:

  [T1] The map respects addition of vectors: $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$ for any vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V.

  [T2] The map respects scalar multiplication: $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$ for any vector $\mathbf{v}$ in V and any scalar $\alpha \in F$.

  - 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 $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$, the scalar multiplication on the left-hand side is in V while the scalar multiplication on the right-hand side is in W.
  - Remark: We require the vector spaces V and W to have the same field of scalars because condition [T2] would not make sense otherwise.

- Example: If $V = W = \mathbb{R}^2$, show that the map T defined[1] by $T(x, y) = \langle x, x + y\rangle$ 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 $T(\alpha\mathbf{v}) = \langle \alpha x, \alpha x + \alpha y\rangle = \alpha \langle x, x + y\rangle = \alpha T(\mathbf{v})$.

- We can substantially generalize the example above:

- Example: If $V = F^m$ (thought of as $m \times 1$ matrices) and $W = F^n$ (thought of as $n \times 1$ matrices) and A is any $n \times m$ matrix, show that the map T with $T(\mathbf{v}) = A\mathbf{v}$ 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, $T(\alpha\mathbf{v}) = A(\alpha\mathbf{v}) = \alpha(A\mathbf{v}) = \alpha T(\mathbf{v})$.

- Checking a particular map to determine if it is a linear transformation is in general fairly straightforward:

- Example: If $V = M_{2\times 2}(\mathbb{Q})$ and $W = \mathbb{Q}$, 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 $V = M_{2\times 2}(\mathbb{C})$ and $W = \mathbb{C}$, determine whether the determinant map $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$ 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 $\det(M_1) + \det(M_2)$.

[1] In principle here we should actually write $T(\langle x, y\rangle) = \langle x, x + y\rangle$, but this notation looks rather ugly, so we will suppress the vector brackets inside the function notation when writing linear transformations on vectors in F^n.

* 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 this function $\boxed{\text{is not a linear transformation}}$. (In fact, the condition [T2] fails as well.)

- Here are a few additional examples of linear transformations:
  * If V is the vector space of differentiable functions and W is the vector space of real-valued functions, the derivative map D sending a function to its derivative is a linear transformation from V to W.
  * If V is the vector space of all continuous functions on $[a, b]$, then the integral map $I(f) = \int_a^b f(x)\,dx$ is a linear transformation from V to $\mathbb{R}$.
  * The transpose map is a linear transformation from $M_{m\times n}(F)$ to $M_{n\times m}(F)$ for any field F and any positive integers m, n.
  * For any $a \in F$, the evaluation at a map on $F[x]$, defined by $T(p) = p(a)$, is a linear transformation from $F[x]$ to F.
  * If V and W are any vector spaces, the <u>zero map</u> sending all elements of V to the zero vector in W is a linear transformation from V to W.
  * If V is any vector space, the <u>identity map</u> sending all elements of V to themselves is a linear transformation from V to V.
- Here are some basic properties of linear transformations:
- <u>Proposition</u> (Properties of Linear Transformations): Suppose that V and W are F-vector spaces. Then the following hold:
  1. If $T : V \to W$ is linear, then $T(\mathbf{0}_V) = \mathbf{0}_W$.
     * <u>Proof</u>: Let $\mathbf{v}$ be any vector in V. Since $0\mathbf{v} = \mathbf{0}_V$ from basic properties, applying [T2] yields $0T(\mathbf{v}) = T(\mathbf{0}_V)$.
     * But $0T(\mathbf{v}) = \mathbf{0}_W$ since scaling any vector of W by 0 gives the zero vector of W.
     * Combining these two statements gives $T(\mathbf{0}_V) = 0T(\mathbf{v}) = \mathbf{0}_W$, so $T(\mathbf{0}_V) = \mathbf{0}_W$ as claimed.
  2. For any vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$ and any scalars $a_1, \dots, a_n$, if T is linear then $T(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1T(\mathbf{v}_1) + \cdots + a_nT(\mathbf{v}_n)$.
     * This result says that linear transformations can be moved through linear combinations.
     * <u>Proof</u>: By a trivial induction using [T1], we see that $T(a_1\mathbf{v}_1+\cdots+a_n\mathbf{v}_n) = T(a_1\mathbf{v}_1)+\cdots+T(a_n\mathbf{v}_n)$.
     * Then by [T2], we have $T(a_i\mathbf{v}_i) = a_iT(\mathbf{v}_i)$ for each $1 \le i \le n$.
     * Plugging these relations into the first equation gives $T(a_1\mathbf{v}_1+\cdots+a_n\mathbf{v}_n) = a_1T(\mathbf{v}_1)+\cdots+a_nT(\mathbf{v}_n)$ as required.
  3. The map $T : V \to W$ is linear if and only if for any $\mathbf{v}_1$ and $\mathbf{v}_2$ in V and any scalar α, $T(\mathbf{v}_1 + \alpha\mathbf{v}_2) = T(\mathbf{v}_1) + \alpha T(\mathbf{v}_2)$.
     * <u>Proof</u>: If T is linear, then by [T1] and [T2], $T(\mathbf{v}_1 + c\mathbf{v}_2) = T(\mathbf{v}_1) + T(c\mathbf{v}_2) = T(\mathbf{v}_1) + cT(\mathbf{v}_2)$.
     * Conversely, suppose that $T(\mathbf{v}_1 + c\mathbf{v}_2) = T(\mathbf{v}_1) + cT(\mathbf{v}_2)$. Setting $c = 1$ produces $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2)$ so T satisfies [T1].
     * Then taking $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}$ and $c = 1$ yields $T(\mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$, so $T(\mathbf{0}) = \mathbf{0}$.
     * Finally, setting $\mathbf{v}_1 = \mathbf{0}$ yields $T(c\mathbf{v}_2) = T(\mathbf{0}) + cT(\mathbf{v}_2) = cT(\mathbf{v}_2)$ so T satisfies [T2].
- A fundamental result is that a linear transformation is completely determined by its values on a basis:
- <u>Theorem</u> (Linear Transformations and Bases): Any linear transformation $T : V \to W$ is characterized by its values on a basis of V. Conversely, for any basis $B = \{\mathbf{v}_i\}$ of V and any vectors $\{\mathbf{w}_i\}$, there exists a unique linear transformation $T : V \to W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for each i.

- Proof: For the first statement, let $B = \{\mathbf{v}_i\}$ be a basis of V. Then any vector $\mathbf{v}$ in V can be written as $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$ for (unique) vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in B and scalars $a_1, \ldots, a_n$.
- By the previous proposition, $T(\mathbf{v}) = a_1T(\mathbf{v}_1) + a_2T(\mathbf{v}_2) + \cdots + a_nT(\mathbf{v}_n)$, so the value of T is determined by the values $T(\mathbf{v}_1)$, $T(\mathbf{v}_2)$, ... , $T(\mathbf{v}_n)$.
- Conversely, suppose that we are given the values $T(\mathbf{v}_i) = \mathbf{w}_i$ for each $\mathbf{v}_i$ in B. Then the map $T : V \to W$ defined by setting $T(a_1\mathbf{v}_{i_1} + a_2\mathbf{v}_{i_2} + \cdots + a_n\mathbf{v}_{i_n}) = a_1\mathbf{w}_{i_1} + \cdots + a_n\mathbf{w}_{i_n}$ is a well-defined linear transformation from V to W: the map is well-defined because every vector in V can be written in exactly one way as a linear combination of vectors in B, and the linearity properties are both immediate.
- If S were some other linear transformation with $S(\mathbf{v}_i) = \mathbf{w}_i = T(\mathbf{v}_i)$ for each i, then since the $\mathbf{v}_i$ are a basis for V, we immediately see that $S(\mathbf{v}) = T(\mathbf{v})$ for all $\mathbf{v}$ in V, meaning that S and T are the same function.

- The theorem above says that we can reconstruct the entirety of a linear transformation given its values on a basis.

- 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 $T(5 + 2x + 2x^2)$.
  - We simply need to express $5 + 2x + 2x^2$ in terms of the basis $\{1, 1 + x, 2 + x^2\}$ for V.
  - A straightforward calculation shows $5 + 2x + 2x^2 = -1(1) + 2(1 + x) + 2(2 + x^2)$.
  - Thus, $T(5 + 2x + 2x^2) = -T(1) + 2T(1 + x) + 2T(2 + x^2) = -1(5) + 2(4) + 2(3) = \boxed{9}$.

2.1.2 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.

- Definition: If $T : V \to W$ is a linear transformation, then the image of T (often also called the range of T), denoted $\text{im}(T)$, is the set of elements $\mathbf{w}$ 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. If $\text{im}(T) = W$, we say T is onto (or surjective).
  - Even though they mean the same thing, we use the word "image" with linear transformations (rather than "range") to emphasize the additional structure that the image of a linear transformation possesses, relative to the range of a general function.

- Example: If $T : \mathbb{R}^3 \to \mathbb{R}^3$ is the linear transformation with $T(x, y, z) = \langle x + y, z, x + y\rangle$, find the kernel and image of T.
  - For the kernel, we want to find all (x, y, z) such that $T(x, y, z) = \langle 0, 0, 0\rangle$, so we obtain the three equations $x + y = 0$, $z = 0$, $x + y = 0$. These equations collectively say $y = -x$, so we see that the kernel is the set of vectors of the form $\langle x, -x, 0\rangle$.
  - For the image, one possible answer is simply "the set of vectors of the form $\langle x + y, z, x + y\rangle$". A slightly more useful description would be "the vectors of the form $\langle a, b, a\rangle$" since the first and second coordinates can be arbitrary, but the third is always equal to the first.

- The kernel and image are subspaces of V and W respectively:

- Proposition (Kernel and Image are Subspaces): If $T : V \to W$ is linear, then $\ker(T)$ is a subspace of V and $\text{im}(T)$ is a subspace of W.
  - Proof: We simply check the subspace criterion for each.

- For $\ker(T)$, [S1] follows because $T(\mathbf{0}) = \mathbf{0}$ by our properties of linear transformations.
- For [S2], if $\mathbf{v}_1$ and $\mathbf{v}_2$ are in $\ker(T)$, then $T(\mathbf{v}_1) = \mathbf{0}$ and $T(\mathbf{v}_2) = \mathbf{0}$. Therefore, $T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0}$.
- Finally, for [S3], if $\mathbf{v}$ is in $\ker(T)$, then $T(\mathbf{v}) = \mathbf{0}$. Hence $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v}) = \alpha\mathbf{0} = \mathbf{0}$.
- For $\mathrm{im}(T)$, [S1] also follows from $T(\mathbf{0}) = \mathbf{0}$.
- For [S2], if $\mathbf{w}_1$ and $\mathbf{w}_2$ are in $\mathrm{im}(T)$, 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.
- Finally, for [S3], if $\mathbf{w}$ is in $\mathrm{im}(T)$, then there exists $\mathbf{v}$ with $T(\mathbf{v}) = \mathbf{w}$. Then $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v}) = \alpha\mathbf{w}$, so $\alpha\mathbf{w}$ is also in the image.

- There is a straightforward way to find a spanning set for the image of a linear transformation:
- Proposition (Computing Image): If $T : V \to W$ is linear and $\mathbf{v}_1, \dots, \mathbf{v}_n$ is a basis for V, then $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ spans $\mathrm{im}(T)$.
  - 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 (e.g., if T is the zero transformation).
  - 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 $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$.
  - Then $\mathbf{w} = T(\mathbf{v}) = a_1 T(\mathbf{v}_1) + \cdots + a_n T(\mathbf{v}_n)$ 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.
- It is natural to wonder whether there is an equally easy way to find a spanning set for the kernel of a general linear transformation: unfortunately, there is not.
  - In the event that V is finite-dimensional, the general procedure for finding the kernel and image is fairly algorithmic once we choose bases for V and W.
  - Finding the kernel requires solving a system of homogeneous linear equations, which can be done using row-reductions.
  - Finding the image requires reducing the spanning set $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ to a basis. As we have already discussed in our analysis of spanning sets and bases, this reduction can be performed by removing linearly dependent elements until the resulting set is linearly independent. (We will later describe a different procedure using row-reductions.)
  - When the dimensions of V and W are fairly small, these calculations can often be done by inspection, rather than resorting to row-reduction algorithms.
- Example: If $T : \mathbb{R}^2 \to \mathbb{R}^3$ is the linear transformation with $T(x, y) = \langle x + y, 0, 2x + 2y\rangle$, find a basis for $\ker(T)$ and $\mathrm{im}(T)$.
  - For the kernel, we want to find all $\langle x, y\rangle$ such that $T(x, y) = \langle 0, 0, 0\rangle$, which clearly are the 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 $\boxed{\langle 1, 0, 2\rangle}$.
- Example: If $T : P_2(\mathbb{C}) \to \mathbb{C}^2$ is the linear transformation with $T(p) = \langle p(1), p'(1)\rangle$, find a basis for $\ker(T)$ and $\mathrm{im}(T)$.
  - Notice that $T(a + bx + cx^2) = \langle a + b + c, b + 2c\rangle$.
  - For the kernel, we want to find all p such that $T(p) = \langle 0, 0\rangle$, which is equivalent to requiring $a+b+c = 0$ and $b + 2c = 0$, so that $b = -2c$ and $a = c$. Thus, the kernel is the set of polynomials of the form $p(x) = c - 2cx + cx^2$, which is spanned by the polynomial $\boxed{1 - 2x + x^2}$.

- The image is spanned by $T(1) = \langle 1,0\rangle$, $T(x) = \langle 1,1\rangle$, $T(x^2) = \langle 1,2\rangle$. Since these vectors clearly span $\mathbb{C}^2$, we may take the first two vectors $\boxed{\langle 1,0\rangle, \langle 1,1\rangle}$ as a basis.

- 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 $\mathrm{im}(T)$ 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 $T : V \to W$, $\dim(\ker(T)) + \dim(\mathrm{im}(T)) = \dim(V)$. In words, the nullity plus the rank is equal to the dimension of V.
  - Proof: Let $\{\mathbf{x}_i\}_{i\in I}$ be a basis for $\ker(T)$ and let $\{\mathbf{w}_j\}_{j\in J}$ be a basis for $\mathrm{im}(T)$ in W, where I and J are indexing sets.
  - Then by the definition of the image, there exist $\{\mathbf{v}_j\}_{j\in J}$ in V such that $T(\mathbf{v}_j) = \mathbf{w}_j$ for each $j \in J$.
  - We claim that the (multi)set of vectors $S = \{\mathbf{x}_i\}_{i\in I} \cup \{\mathbf{v}_j\}_{j\in J}$ is a basis for V; since the cardinality of S is clearly $\dim(\ker(T)) + \dim(\mathrm{im}(T))$, this will establish the theorem.
  - To see that S spans V, let $\mathbf{v}$ be an element of V. Since $T(\mathbf{v})$ lies in $\mathrm{im}(T)$, there exist scalars $b_1, \dots, b_k$ and vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ such that $T(\mathbf{v}) = \sum_{j=1}^k b_j \mathbf{w}_j$.
  - Then $T\left[\mathbf{v} - \sum_{j=1}^k b_j\mathbf{v}_j\right] = T(\mathbf{v}) - \sum_{j=1}^k b_j T(\mathbf{v}_j) = \sum_{j=1}^k b_j\mathbf{w}_j - \sum_{j=1}^k b_j\mathbf{w}_j = \mathbf{0}$, meaning that $\mathbf{v} - \sum_{j=1}^k b_j\mathbf{v}_j$ is in $\ker(T)$ hence can be written as a sum $\sum_{i=1}^l c_i\mathbf{x}_i$ for unique scalars c_i and some vectors $\mathbf{x}_1, \dots \mathbf{x}_l$.
  - But this means $\mathbf{v} = \sum_{j=1}^k b_j\mathbf{v}_j + \sum_{i=1}^l c_i\mathbf{x}_i$ for some scalars b_j and c_i, and so S spans V.
  - To see that S is linearly independent, suppose we had a dependence $\mathbf{0} = \sum_{j=1}^k b_j\mathbf{v}_j + \sum_{i=1}^l c_i\mathbf{x}_i$.
  - Applying T to both sides yields $\mathbf{0} = T(\mathbf{0}) = \sum_{j=1}^k b_j T(\mathbf{v}_j) + \sum_{i=1}^l c_i T(\mathbf{x}_i) = \sum_{j=1}^k b_j\mathbf{w}_j$.
  - Since the $\mathbf{w}_j$ are linearly independent, we conclude that all the coefficients b_j must be zero.
  - We then obtain the relation $\mathbf{0} = \sum_{i=1}^l c_i\mathbf{x}_i$, but now since the $\mathbf{x}_i$ are linearly independent, we conclude that all the coefficients c_i must also be zero. Hence S is linearly independent, as claimed.

- 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 conclusion of 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.

2.1.3 Algebraic Operations on Linear Transformations

- Now that we have established some basic properties of individual linear transformations, we will study the ways that linear transformations can interact with one another.
- Functions sharing the same domain and range have a vector space structure, as we have already discussed.
  - Recall that if $f : S \to T$ and $g : S \to T$ are functions, then we have defined the sum $f + g$, which is also a function from S to T, by setting $(f + g)(x) = f(x) + g(x)$ for each x in S.
  - Likewise, we have defined the scalar multiple cf as the function with $(cf)(x) = cf(x)$, for each x in S.
  - With these operations of (function) addition and scalar multiplication, the set of functions from S to T forms a vector space.
- We would now like to analyze this structure when the collection of functions is the set of linear transformations from one vector space to another.
  - The key observation is that the sum of two linear transformations is also a linear transformation, as is any scalar multiple of a linear transformation. Explicitly:
- Theorem (Space of Linear Transformations): Let V and W be vector spaces with the same field of scalars. Then the set $\mathcal{L}(V, W)$ of all linear transformations from V to W is a subspace of the space of functions from V to W.
  - Proof: We verify the subspace criterion.
  - [S1]: The zero map is a linear transformation.
  - [S2]: Suppose that T_1 and T_2 be linear transformations: we must show that $T_1 + T_2$ is also a linear transformation. This follows from the observations that

$$\begin{aligned}(T_1 + T_2)(\mathbf{v}_1 + \mathbf{v}_2) &= T_1(\mathbf{v}_1 + \mathbf{v}_2) + T_2(\mathbf{v}_1 + \mathbf{v}_2) = [T_1(\mathbf{v}_1) + T_1(\mathbf{v}_2)] + [T_2(\mathbf{v}_1) + T_2(\mathbf{v}_2)] \\ &= [T_1(\mathbf{v}_1) + T_2(\mathbf{v}_1)] + [T_1(\mathbf{v}_2) + T_2(\mathbf{v}_2)] = (T_1 + T_2)(\mathbf{v}_1) + (T_1 + T_2)(\mathbf{v}_2)\end{aligned}$$

    and that

$$(T_1 + T_2)(\alpha\mathbf{v}) = T_1(\alpha\mathbf{v}) + T_2(\alpha\mathbf{v}) = \alpha T_1(\mathbf{v}) + \alpha T_2(\mathbf{v}) = \alpha[(T_1 + T_2)(\mathbf{v})].$$

  - [S3]: Suppose that T is a linear transformation: we must show that cT is also a linear transformation. This follows from the observations that

$$(cT)(\mathbf{v}_1 + \mathbf{v}_2) = cT(\mathbf{v}_1 + \mathbf{v}_2) = c\,[T(\mathbf{v}_1) + T(\mathbf{v}_2)] = cT(\mathbf{v}_1) + cT(\mathbf{v}_2)$$

    and

$$(cT)(\alpha\mathbf{v}) = cT(\alpha\mathbf{v}) = c[\alpha T(\mathbf{v})] = \alpha[cT(\mathbf{v})].$$

- We can go further if we also allow ourselves to compose linear transformations. As with sums and scalar multiple of linear transformations, the composition of two linear transformations is also a linear transformation. Explicitly:
- Proposition (Composition of Linear Transformations): If $T_2 : U \to V$ and $T_1 : V \to W$ are two linear transformations, then the composite function $T_1 \circ T_2 : U \to W$, written T_1T_2, is also a linear transformation.

- Proof: Observe that $T_1T_2(\mathbf{v}_1+\mathbf{v}_2) = T_1[T_2(\mathbf{v}_1+\mathbf{v}_2)] = T_1[T_2(\mathbf{v}_1)+T_2(\mathbf{v}_2)] = T_1T_2(\mathbf{v}_1)+T_1T_2(\mathbf{v}_2)$ and that $T_1T_2(\alpha\mathbf{v}) = T_1[T_2(\alpha\mathbf{v})] = T_1[\alpha T_2(\mathbf{v})] = \alpha T_1T_2(\mathbf{v})$.

- Example: For $T_1(x,y) = \langle x, x-y, y\rangle$ and $T_2(x,y,z) = \langle x-z, y-z\rangle$, find T_1T_2 and T_2T_1.
  - We simply compute $T_1(T_2(x,y,z)) = T_1(x-z, y-z) = \langle x-z, x-y, y-z\rangle$ and $T_2(T_1(x,y)) = T_2(x, x-y, y) = \langle x-y, x-2y\rangle$.
  - It is quite easy to see from this explicit description that both compositions are linear.

- Example: For $T_1(x,y) = \langle 3x, x-y\rangle$ and $T_2(x,y) = \langle x+y, 2x\rangle$, find T_1T_2 and T_2T_1.
  - We compute $T_1(T_2(x,y)) = T_1(x+y, 2x) = \langle 3x+3y, -x+y\rangle$ and $T_2(T_1(x,y)) = T_2(3x, x-y) = \langle 4x-y, 6x\rangle$.

- Composition of linear transformations is not commutative (in much the same way that composition of general functions is not commutative): if T_1T_2 is defined, it need not even be the case that T_2T_1 is defined.
  - In the event that both compositions T_1T_2 and T_2T_1 are defined (which requires $T_1 : V \to W$ and $T_2 : W \to V$), the composite function $T_1T_2 : W \to W$ and the composite function $T_2T_1 : V \to V$ are maps on different vector spaces.
  - Finally, even in the very special case where T_1 and T_2 are both linear transformations from a vector space V to itself, T_1T_2 is not generally equal to T_2T_1, as we saw in the second example above.

- On the other hand, except for commutativity (which fails), the operations of composition, addition, and scalar multiplication of linear transformations possess most of the other algebraic properties reminiscent of those of a field or vector space.

- Proposition (Identity for Linear Transformations): Suppose V, W are vector spaces with the same field of scalars, and $T : V \to W$ is a linear transformation. If I_V is the identity map on V and I_W is the identity map on W (each mapping every vector to itself), then $TI_V = T$ and $I_WT = T$.
  - Proof: Trivial, since $TI_V(\mathbf{v}) = T(\mathbf{v}) = I_WT(\mathbf{v})$ for every $\mathbf{v}$ in V, by definition of I_V and I_W.

- Proposition (Associativity of Linear Transformations): Suppose U, V, W, X are vector spaces with the same field of scalars, and $T_3 : U \to V$, $T_2 : V \to W$, and $T_1 : W \to X$ are linear transformations. Then $T_1(T_2T_3) = (T_1T_2)T_3$.
  - Proof: This follows from the fact that function composition is always associative.
  - Explicitly, for any $\mathbf{v}$ in U, we have $[T_1(T_2T_3)](\mathbf{v}) = T_1[(T_2T_3)(\mathbf{v})] = T_1[T_2(T_3(\mathbf{v}))]$, while $[(T_1T_2)T_3](\mathbf{v}) = [T_1T_2](T_3(\mathbf{v})) = T_1(T_2(T_3(\mathbf{v})))$.
  - Since these expressions are both equal to $T_1(T_2(T_3(\mathbf{v}))$ for any $\mathbf{v}$ in U, these functions are the same.

- Since composition of linear transformations is associative, for a linear transformation $T : V \to V$ we can define powers of T:

- Definition: For a linear transformation $T : V \to V$, we define $T^0 = I_V$, the identity transformation, and for $n \geq 1$ we set $T^n = T^{n-1}T$.
  - By an easy induction using the associative law, one can see that $T^{m+n} = T^mT^n$ for all $m, n \geq 0$.

- We also have versions of the "distributive law" for linear transformations:

- Proposition (Distributivity of Linear Transformations): Suppose U, V, W are vector spaces with the same field of scalars, $T_1 : U \to V$ and $T_2 : U \to V$ are linear transformations, and c is any scalar.
  1. If $S : V \to W$ is any linear transformation, then $S(T_1+T_2) = ST_1 + ST_2$ and $(cS)T_1 = c(ST_1)$.
  2. If $S : W \to U$ is any linear transformation, then $(T_1+T_2)S = T_1S + T_2S$ and $(cT_1)S = c(T_1S)$.

- Proof: For the first statement of (1), for any vector $\mathbf{v}$ in U we have

$$[S(T_1+T_2)](\mathbf{v}) = S[(T_1+T_2)(\mathbf{v})] = S[T_1(\mathbf{v})+T_2(\mathbf{v})] = ST_1(\mathbf{v})+ST_2(\mathbf{v}) = [ST_1+ST_2](\mathbf{v})$$

  so $S(T_1+T_2)$ and ST_1+ST_2 behave identically on every vector $\mathbf{v}$.
- Similarly, for the second statement of (1), we have

$$[(cS)T_1](\mathbf{v}) = (cS)[T_1(\mathbf{v})] = cS(T_1(\mathbf{v})) = c(ST_1)(\mathbf{v})$$

  and so $(cS)T_1$ and $c(ST_1)$ behave identically on every vector $\mathbf{v}$.
- The statements in (2) follow in exactly the same way.

2.1.4 One-to-One Linear Transformations

- Now that we have characterized some of the algebraic properties of linear transformations, we ask the next natural question: when does a linear transformation $T : V \to W$ have an inverse?
  - To be precise, we are seeking a two-sided inverse for T: a function $T^{-1} : \text{im}(T) \to V$ such that TT^{-1} is the identity on $\text{im}(T)$ (meaning that $TT^{-1}(\mathbf{w}) = \mathbf{w}$ for every $\mathbf{w}$ in $\text{im}(T)$) and that $T^{-1}T$ is the identity on V (meaning that $T^{-1}T(\mathbf{v}) = \mathbf{v}$ for every $\mathbf{v}$ in V).
  - For general functions $f : S \to T$, f will possess a two-sided inverse precisely when f is one-to-one, meaning that $f(x_1) = f(x_2)$ implies $x_1 = x_2$. Equivalently, a one-to-one function is one that sends distinct elements in its domain to distinct elements in its range.
  - The construction is straightforward (intuitively, we simply "reverse" the action of f).
  - Explicitly, suppose $f : S \to T$ is one-to-one. For each $y \in \text{range}(f)$, there exists an element $x \in S$ such that $f(x) = y$, and by the assumption that f is one-to-one, there is exactly one such x. Now define the function $f^{-1} : \text{range}(f) \to S$ as follows: for each $y \in \text{range}(f)$ with $f(x) = y$, set $f^{-1}(y) = x$.
  - This definition is well-posed because x is unique, and it is easy to verify that $f^{-1}(f(x)) = x$ for each x in S and that $f(f^{-1}(y)) = y$ for each y in $\text{range}(f)$.
- We can give a simple characterization of whether a particular linear transformation possesses an inverse by looking at its kernel:
- Proposition (Kernel and One-to-One Maps): If $T : V \to W$ is a linear transformation, the following are equivalent:
  1. T is a one-to-one function on V.
  2. T possesses a (two-sided) inverse function $T^{-1} : \text{im}(T) \to V$ that is a linear transformation.
  3. The kernel $\ker(T)$ consists of only the zero vector.

  - Proof: We show that (1) implies (2), that (2) implies (3), and that (3) implies (1).
  - $1 \implies 2$: If T is a one-to-one function then by the discussion above, we know that T has a two-sided inverse function T^{-1}, so we just need to show that T^{-1} is a linear transformation.
    * [T1]: If $\mathbf{w}_1$ and $\mathbf{w}_2$ are two elements in $\text{im}(T)$, then by hypothesis there exist vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Then $T(\mathbf{v}_1+\mathbf{v}_2) = \mathbf{w}_1+\mathbf{w}_2$ so applying T^{-1} yields $T^{-1}(\mathbf{w}_1+\mathbf{w}_2) = T^{-1}T(\mathbf{v}_1+\mathbf{v}_2) = \mathbf{v}_1+\mathbf{v}_2 = T^{-1}(\mathbf{w}_1)+T^{-1}(\mathbf{w}_2)$.
    * [T2]: If $\mathbf{w}$ is any element in $\text{im}(T)$, then by hypothesis there exists a vector $\mathbf{v}$ in V such that $T(\mathbf{v}) = \mathbf{w}$. Then $T(c\mathbf{v}) = c\mathbf{w}$, so applying T^{-1} yields $T^{-1}(c\mathbf{w}) = T^{-1}T(c\mathbf{v}) = c\mathbf{v} = cT^{-1}(\mathbf{w})$.
  - $2 \implies 3$: Suppose T has an inverse transformation T^{-1} and that $T(\mathbf{v}) = \mathbf{0}$. Applying T^{-1} to both sides yields $\mathbf{v} = T^{-1}(T(\mathbf{v})) = T^{-1}(\mathbf{0}) = \mathbf{0}$. Thus, the only element in $\ker(T)$ is the zero vector.
  - $3 \implies 1$: Suppose $\ker(T) = \{\mathbf{0}\}$ and that $T(\mathbf{v}_1) = T(\mathbf{v}_2)$. Then since T is a linear transformation, we can write $\mathbf{0} = T(\mathbf{v}_1) - T(\mathbf{v}_2) = T(\mathbf{v}_1 - \mathbf{v}_2)$, hence $\mathbf{v}_1 - \mathbf{v}_2$ is in $\ker(T)$. But since $\ker(T) = \{\mathbf{0}\}$, we conclude that $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}$, so that $\mathbf{v}_1 = \mathbf{v}_2$. Hence $T(\mathbf{v}_1) = T(\mathbf{v}_2)$ implies $\mathbf{v}_1 = \mathbf{v}_2$, which means T is one-to-one.

- Per the result above, we can easily determine whether a given linear transformation is one-to-one by computing its kernel.
- Example: Determine whether $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(x, y, z) = \langle x - y, z, 2x - 2y \rangle$ is one-to-one.
  - We simply find $\ker(T)$: if $T(x, y, z) = \langle 0, 0, 0 \rangle$, then we must have $x - y = 0$, $z = 0$, and $2x - 2y = 0$, so we easily see that $\ker(T)$ is the set of vectors $\langle x, y, z \rangle$ with $x = y$.
  - Since the nonzero vector $\langle 1, 1, 0 \rangle$ is in $\ker(T)$, for instance, we see that T is not one-to-one and hence $\boxed{\text{not one-to-one}}$.
- Example: Determine whether the linear transformation T on the space of polynomials with real coefficients defined by $T(p) = \int_0^x p(t)\,dt$ is one-to-one.
  - We simply find $\ker(T)$: if $p(x) = a_0 + a_1 x + \cdots + a_n x^n$, then $T(p) = \int_0^x (a_0 + a_1 x + \cdots + a_n x^n)\,dx = a_0 x + \dfrac{a_1}{2} x^2 + \cdots + \dfrac{a_n}{n+1} x^{n+1}$.
  - Since $T(p) = 0$ only when $a_0 = a_1 = \cdots = a_n = 0$, we conclude that $\ker(T) = 0$ and hence that T is $\boxed{\text{one-to-one}}$.
  - In this case, we can even write down the inverse of T: it is the derivative map $D(p) = p'(x)$.
- One-to-one maps have a number of useful properties. Here are a few properties that follow essentially from the definition:
  - If $T : V \to W$ is one-to-one, then its inverse $T^{-1} : \mathrm{im}(T) \to V$ is unique.
  - If $T : V \to W$ is one-to-one, then $T^{-1} : \mathrm{im}(T) \to V$ is also one-to-one, and $(T^{-1})^{-1} = T$.
  - If $T : U \to V$ and $S : V \to W$ are one-to-one, then so is ST, and $(ST)^{-1} = T^{-1} S^{-1}$.
- Another important property of one-to-one maps is that they preserve linear independence:
- Proposition (One-to-One Maps Preserve Independence): If $T : V \to W$ is a one-to-one linear transformation, 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 $a_1 T(\mathbf{v}_1) + \cdots + a_n T(\mathbf{v}_n) = T(a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n)$ for any scalars $a_1, \dots, a_n$.
  - First suppose that $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent, and consider a dependence $a_1 T(\mathbf{v}_1) + \cdots + a_n T(\mathbf{v}_n) = \mathbf{0}$.
  - By the above, we see that $T(a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n) = \mathbf{0}$, so since $\ker(T) = \{\mathbf{0}\}$, this implies $a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n = \mathbf{0}$. But since $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent, we must have $a_1 = \cdots = a_n = 0$.
  - For the other direction, suppose that $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ are linearly independent, and that $b_1 \mathbf{v}_1 + \cdots + b_n \mathbf{v}_n = \mathbf{0}$. Then $b_1 T(\mathbf{v}_1) + \cdots + b_n T(\mathbf{v}_n) = T(b_1 \mathbf{v}_1 + \cdots + b_n \mathbf{v}_n) = T(\mathbf{0}) = \mathbf{0}$.
  - But since $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ are linearly independent, we must have $b_1 = \cdots = b_n = 0$.
- Corollary: If $T : V \to W$ is a one-to-one linear transformation and B is a basis for V, then the vectors $T(\mathbf{v}_i)$ for $\mathbf{v}_i$ in B form a basis for $\mathrm{im}(T)$.
  - Proof: We showed earlier that the vectors $T(\mathbf{v}_i)$ span $\mathrm{im}(T)$, and the previous proposition shows that the $T(\mathbf{v}_i)$ are also linearly independent, so they are a basis.

2.1.5 Isomorphisms of Vector Spaces

- We will now discuss a notion of equivalence of vector spaces, building on our results about one-to-one maps.
  - Recall that by an earlier proposition, $T : V \to W$ is one-to-one (or injective) if $\ker(T) = \{\mathbf{0}\}$.
  - Also recall that we defined earlier that $T : V \to W$ is onto (or surjective) if $\operatorname{im}(T) = W$.
  - A function $f : S \to T$ which is both injective and surjective is called a bijection.
- We showed already that a one-to-one linear transformation $T : V \to W$ possesses a (unique) inverse map $T^{-1} : \operatorname{im}(T) \to V$.
  - However, what we would really like is for this inverse map to be a function on all of W, not just on the image of T.
  - In order to ensure this, we simply need to require that $\operatorname{im}(T) = W$, which is the same as saying that T is onto.
- 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 $\ker(T) = \{\mathbf{0}\}$ and $\operatorname{im}(T) = W$. 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, as we will see: it says that the vector spaces V and W have exactly the same structure.
  - More specifically, 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: Show that the map $T : \mathbb{R}^4 \to M_{2\times 2}(\mathbb{R})$ given by $T(x_1, x_2, x_3, x_4) = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}$ is an isomorphism.
  - 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 $\operatorname{im}(T) = M_{2\times 2}$.
  - Thus T is an isomorphism.
- Example: Show that the map $T : \mathbb{C}^3 \to P_2(\mathbb{C})$ given by $T(a, b, c) = (a + b) + (a + c)x + (b + c)x^2$ is an isomorphism.
  - This map is a linear transformation; it clearly is additive and respects scalar multiplication.
  - Also, $\ker(T) = \{\mathbf{0}\}$ since $T(a, b, c) = \mathbf{0}$ requires $a + b = a + c = b + c = 0$, and the only solution to this system is $a = b = c = 0$.
  - Finally, a brief calculation will show that $T\left(\frac{a_0 + a_1 - a_2}{2}, \frac{a_0 + a_2 - a_1}{2}, \frac{a_1 + a_2 - a_0}{2}\right) = a_0 + a_1 x + a_2 x^2$, so $\operatorname{im}(T) = P_2(\mathbb{C})$.
  - Thus T is an isomorphism.
  - Remark: Alternatively, after computing $\ker(T) = \{\mathbf{0}\}$, we could have used the nullity-rank theorem to conclude that the dimension of $\operatorname{im}(T)$ was $3 - 0 = 3$, hence necessarily all of $P_2(\mathbb{C})$.
- Proposition (Isomorphisms and Inverses): If $T : V \to W$ is a linear transformation, the following are equivalent:
  1. T is one-to-one and onto.
  2. T possesses an inverse function $T^{-1} : W \to V$ that is a linear transformation.
  3. T possesses an inverse function $T^{-1} : W \to V$.
  - Notice that the condition (3) does not include the requirement that T^{-1} be a linear transformation: the point is that it is sufficient for T to have an inverse *function* only.

- Proof: We show that (1) implies (2) and that (3) implies (1), since (2) clearly implies (3).
- $1 \implies 2$: If T is one-to-one, then by our earlier results on invertible transformations, we know that T has a two-sided inverse function $T^{-1} : \mathrm{im}(T) \to V$ that is a linear transformation. If T is also onto, then $\mathrm{im}(T) = W$, so the inverse function T^{-1} is a map from W to V.
- $3 \implies 1$: Suppose T possesses an inverse function T^{-1}. Then by hypothesis, $T(T^{-1}(\mathbf{w})) = \mathbf{w}$ for every $\mathbf{w}$ in W, so T is onto. Furthermore, since $T(\mathbf{0}) = \mathbf{0}$ and T^{-1} is well-defined, we must have $T^{-1}(\mathbf{0}) = \mathbf{0}$. Then if $T(\mathbf{v}) = \mathbf{0}$, applying T^{-1} to both sides yields $\mathbf{v} = T^{-1}T(\mathbf{v})) = T^{-1}(\mathbf{0}) = \mathbf{0}$, so $\ker(T) = \{\mathbf{0}\}$ meaning that T is one-to-one.

• Here are a few other properties of isomorphisms that follow essentially immediately from the definition and our earlier results about one-to-one maps:

- The identity map $I_V : V \to V$ is an isomorphism.
- If $T : V \to W$ is an isomorphism, then $T^{-1} : W \to V$ is also an isomorphism.
- If $T : U \to V$ and $S : V \to W$ are isomorphisms, then $ST : U \to W$ is an isomorphism.
- Remark: These three properties above collectively show that "being isomorphic" is an equivalence relation[2] on vector spaces. (For example, if U is isomorphic to V and V is isomorphic to W, then U is isomorphic to W.)
- Isomorphisms preserve linear independence: 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.
- Isomorphisms preserve span: if $T : V \to W$ is an isomorphism and $S = \{\mathbf{v}_i\}_{i \in I}$ is a subset of V with $T(S) = \{T(\mathbf{v}_i)\}_{i \in I}$, then $T(\operatorname{span} S) = \operatorname{span} T(S)$.

• 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 V with scalar field F is isomorphic to F^n, where $n = \dim_F V$.

- This result should be rather unexpected: it certainly doesn't seem obvious, just from the eight axioms of a vector space, that any finite-dimensional vector space is essentially "the same" as the vector space F^n for some n. But they are!
- Proof: Isomorphisms preserve linear independence (since they and their inverses are one-to-one), 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 $T(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1\mathbf{w}_1 + \cdots + a_1\mathbf{w}_n$ is an isomorphism between V and W.
- We need to check five things: that T is well-defined, that T respects addition, that T respects scalar multiplication, that T is one-to-one, and that T is onto.
- T is well-defined: The description above defines T on every element $\mathbf{v}$ of V because $\mathbf{v}_1, \dots, \mathbf{v}_n$ spans V, and the definition is unique because there is only way of writing $\mathbf{v}$ as a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_n$ (because $\mathbf{v}_1, \dots, \mathbf{v}_n$ is linearly independent).
- T respects addition: If $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_1\mathbf{v}_n$ and $\tilde{\mathbf{v}} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$, then $T(\mathbf{v} + \tilde{\mathbf{v}}) = (a_1 + b_1)\mathbf{w}_1 + \cdots + (a_n + b_n)\mathbf{w}_n = T(\mathbf{v}) + T(\tilde{\mathbf{v}})$ by the distributive law.
- T respects scalar multiplication: For any scalar β we have $T(\beta\mathbf{v}) = (\beta a_1)\mathbf{w}_1 + \cdots + (\beta a_n)\mathbf{w}_n = \beta T(\mathbf{v})$.
- T is one-to-one: Since $\mathbf{w}_1, \dots, \mathbf{w}_n$ are linearly independent, the only way that $a_1\mathbf{w}_1 + \cdots + a_1\mathbf{w}_n$ can be the zero vector is if $a_1 = a_2 = \cdots = a_n = 0$, which means $\ker(T) = 0$.
- T is onto: Since $\mathbf{w}_1, \dots, \mathbf{w}_n$ span W, every element $\mathbf{w}$ in W can be written as $\mathbf{w} = a_1\mathbf{w}_1 + \cdots + a_1\mathbf{w}_n$ for some scalars $a_1, \cdots a_n$. Then for $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_1\mathbf{v}_n$, we have $T(\mathbf{v}) = \mathbf{w}$.

[2] Recall that a relation $\sim$ on a set S is an equivalence relation if $x \sim x$ for every x, if $x \sim y$ implies $y \sim x$, and if $x \sim y$ and $y \sim z$ together imply $x \sim z$. Equivalence relations behave similarly to equalities, and (indeed) equality of two numbers is the prototypical example of an equivalence relation.

- In much the same way that being linearly independent, being a spanning set, and being a basis are equivalent in a finite-dimensional vector space, we have a similar relationship between being one-to-one, being onto, and being an isomorphism for a map between two finite-dimensional vector spaces of the same dimension:

- <u>Proposition</u> (One-to-One, Onto, Isomorphism Equivalences): Suppose $T : V \to W$ is a linear transformation where V and W are finite-dimensional vector spaces of the same dimension. Then T is one-to-one if and only if T is onto, if and only if T is an isomorphism.

  - <u>Proof</u>: Under the given hypotheses, T is one-to-one $\iff$ $\ker(T) = \{\mathbf{0}\}$ $\iff$ $\dim(\ker(T)) = 0 \overset{\text{nullity-rank}}{\iff}$ $\dim(\text{im}(T)) = \dim(V) \overset{\dim(V)=\dim(W)}{\iff} \dim(\text{im}(T)) = \dim(W) \iff \text{im}(T) = W \iff T$ is onto.
  - Thus, T is one-to-one if and only if T is onto. These two statements together are equivalent to T being an isomorphism, so all three statements are equivalent to one another.

- It is important to note that the result above is not true for infinite-dimensional vector spaces:

- <u>Example</u>: Let $V = \mathbb{R}[x]$. Show that the linear transformation $I : V \to V$ defined by $I[f(x)] = \int_0^x f(t)\,dt$ is one-to-one but not onto, while the linear transformation $D : V \to V$ defined by $D[f(x)] = f'(x)$ is onto but not one-to-one.

  - Notice that $I[a_0 + a_1x + \cdots + a_nx^n] = a_0x + \dfrac{a_1}{2}x^2 + \cdots + \dfrac{a_n}{n+1}x^{n+1}$, so the only polynomial that I maps to zero is the zero polynomial. However, I is not onto, since any polynomial in the image of I is divisible by x. (In fact, the image of I is exactly the set of polynomials divisible by x.)
  - In a similar way, $D[a_0 + a_1x + \cdots + a_nx^n] = a_1 + 2a_2x + \cdots + na_nx^{n-1}$: then $D[1] = 0 = D[0]$ so D is not one-to-one. However, D is onto, since as is easy to check directly (or as follows from the fundamental theorem of calculus), $D(I(f)) = f$.
  - <u>Remark</u>: Note here that $D(I(f)) = f$ for every polynomial f, but $I(D(f)) = f - a_0$. The linear transformations I and D cancel one another when composed in the order $D \circ I$, but do not cancel when composed in the order $I \circ D$.

2.2 Matrices Associated to Linear Transformations

- So far, we have studied linear transformations $T : V \to W$ in a fairly generic way, without much reference to the structure of V or W.

  - If we choose a basis for V and a basis for W, however, we can describe the behavior of T with respect to this basis, and it turns out that T behaves exactly like multiplication by a matrix[3], at least when V and W are finite-dimensional.

- To illustrate the idea, consider the map from $T : \mathbb{R}^3 \to \mathbb{R}^2$ with $T(x, y, z) = \langle 2x - y + z, 3x + 4y - 5z\rangle$.

  - Let us choose the standard basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \{\langle 1, 0, 0\rangle, \langle 0, 1, 0\rangle, \langle 0, 0, 1\rangle\}$ for V and the standard basis $\{\mathbf{w}_1, \mathbf{w}_2\} = \{\langle 1, 0\rangle, \langle 0, 1\rangle\}$ for W.
  - Then $T(\mathbf{v}_1) = 2\mathbf{w}_1 + 3\mathbf{w}_2$, $T(\mathbf{v}_2) = -\mathbf{w}_1 + 4\mathbf{w}_2$, and $T(\mathbf{v}_3) = \mathbf{w}_1 - 5\mathbf{w}_2$.
  - We can summarize this by saying that $T(a\mathbf{v}_1 + b\mathbf{v}_2 + c\mathbf{v}_3) = (2a - b + c)\mathbf{w}_1 + (3a + 4b - 5c)\mathbf{w}_2$.
  - Notice that the coefficients of $\mathbf{w}_1$ and $\mathbf{w}_2$ are given by the entries in the matrix product $\begin{bmatrix} 2 & -1 & 1 \\ 3 & 4 & -5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.
  - Furthermore, as we proved earlier, the behavior of T on V is completely characterized by its behavior on a basis of V, and by the definition of a basis, any vector in W is completely characterized by the coefficients when it is written as a linear combination of the basis elements of W.
  - In other words, the entries in the matrix $\begin{bmatrix} 2 & -1 & 1 \\ 3 & 4 & -5 \end{bmatrix}$ completely characterize the behavior of the linear transformation T, once we have chosen the bases $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for V and $\{\mathbf{w}_1, \mathbf{w}_2\}$ for W.

[3] In fact, the correspondence between linear transformations and matrix multiplication is the reason that matrix multiplication is defined the way it is.

- Observe that the columns of this matrix are simply the coefficient vectors for the basis elements of V in terms of the basis elements of W.

- By choosing particular bases for V and for W, we obtain a correspondence between linear transformations from V to W and matrices: this will allow us to analyze both types of objects together, and to study each one using our understanding of the other.
  - For example, by using properties of linear transformations, 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.
  - Conversely, we will be able to prove a number of things about linear transformations by using properties of matrix arithmetic.

2.2.1 The Matrix Associated to a Linear Transformation

- To define matrices associated to linear transformations, we first need to define the objects we will use for the construction:

- Definition: If V is a finite-dimensional vector space, an ordered basis for V is a basis of V equipped with a particular ordering.
  - We will write an ordered basis in the same way as we write a generic set, and it is to be taken for granted the fact that when we write an expression like $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, we intend β to be an ordered basis unless specifically stated otherwise.
  - Example: The pairs $\beta_1 = \{\langle 1, 0\rangle, \langle 0, 1\rangle\}$, $\beta_2 = \{\langle 1, 1\rangle, \langle 0, 2\rangle\}$, and $\beta_3 = \{\langle 0, 2\rangle, \langle 1, 1\rangle\}$ are three different ordered bases of $\mathbb{R}^2$. (Note that $\beta_2 \neq \beta_3$ because the ordering is different.)

- Definition: Let V be a finite-dimensional vector space with scalar field F, and let $\beta = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ be an ordered basis for V. For a vector $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$, we define the coordinate vector of $\mathbf{v}$ relative to β to be the vector $[\mathbf{v}]_\beta = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ in F^n.
  - Note that because β is a basis of V, the coefficients $a_1, a_2, \dots, a_n$ exist and are unique.
  - Example: If V is the space of polynomials of degree ≤ 3 with ordered basis $\beta = \{1, x, x^2, x^3\}$, then the coordinate vectors of $3 - 4x + x^3$ and $-x$ are $\begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}$ respectively.
  - Example: If $V = \mathbb{R}^2$ with ordered basis $\beta = \{\langle 1, 1\rangle, \langle 0, 2\rangle\}$, then the coordinate vectors of $\langle 1, 1\rangle$, $\langle 1, 5\rangle$, and $\langle 4, -2\rangle$ relative to β are $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, and $\begin{bmatrix} 4 \\ -3 \end{bmatrix}$ respectively.

- By working with coordinate vectors, we can essentially transport our discussion from the general vector space V into the more concrete setting of F^n. Explicitly:

- Proposition: Let V be a finite-dimensional vector space with scalar field F, and let β be an ordered basis of V. Then the map $\varphi : V \to F^n$ defined by $\varphi(\mathbf{v}) = [\mathbf{v}]_\beta$ is an isomorphism.
  - Proof: It is easy to see that φ is linear, since $[\mathbf{v} + \mathbf{w}]_\beta = [\mathbf{v}]_\beta + [\mathbf{w}]_\beta$ and $[c\mathbf{v}]_\beta = c[\mathbf{v}]_\beta$.
  - Furthermore, since β is linearly independent, the only vector $\mathbf{v}$ whose coordinate vector is the zero vector is $\mathbf{v} = \mathbf{0}$, so φ is one-to-one. Finally, since β spans V, the map φ is onto.

- Given a linear transformation $T : V \to W$, if we choose ordered bases $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for V and $\{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ for W, we can represent the behavior of T by writing down the coordinate vectors for the elements $T(\mathbf{v}_j)$ with respect to the vectors $\mathbf{w}_i$.

- Definition: Let V and W be finite-dimensional vector spaces with ordered bases $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ respectively. If $T : V \to W$ is a linear transformation, for each $1 \le j \le n$ and $1 \le i \le m$ there exist unique scalars $a_{i,j}$ such that $T(\mathbf{v}_j) = \sum_{i=1}^{m} a_{i,j}\mathbf{w}_i$ for each $1 \le j \le n$. The $m \times n$ matrix $[T]_\beta^\gamma$ whose (i,j)-entry is $a_{i,j}$ is called the matrix representation of T with respect to the ordered bases β and γ.
    - The definition is rather lengthy, but the basic idea is the same as the one we described above: the jth column of the matrix $[T]_\beta^\gamma$ is $[T(\mathbf{v}_j)]_\gamma$, the coordinate vector of $T(\mathbf{v}_j)$ with respect to the basis γ (of W).

- Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ with $T(x,y,z) = \langle 2x - y + z, 3x + 4y - 5z\rangle$. Find the matrix associated to T with respect to the standard bases $\beta = \{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 0,0,1\rangle\}$ and $\gamma = \{\langle 1,0\rangle, \langle 0,1\rangle\}$ of $\mathbb{R}^3$ and $\mathbb{R}^2$ respectively.
    - We have $T(1,0,0) = 2\langle 1,0\rangle + 3\langle 0,1\rangle$, $T(0,1,0) = -1\langle 1,0\rangle + 4\langle 0,1\rangle$, and $T(0,0,1) = 1\langle 1,0\rangle - 5\langle 0,1\rangle$.
    - Therefore, the matrix associated to T is $[T]_\beta^\gamma = \boxed{\begin{bmatrix} 2 & -1 & 1 \\ 3 & 4 & -5 \end{bmatrix}}$.

- Example: Let $T : \mathbb{C}^3 \to P_2(\mathbb{C})$ be defined by $T(a,b,c) = (a+b) + (a-2c)x + (a+b+c)x^2$. Find the matrix associated to T with respect to the standard bases $\beta = \{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 0,0,1\rangle\}$ and $\gamma = \{1, x, x^2\}$ of $\mathbb{C}^3$ and $P_2(\mathbb{C})$ respectively.
    - We have $T(1,0,0) = 1 + x + x^2$, $T(0,1,0) = 1 + x^2$, and $T(0,0,1) = -2x + x^2$.
    - Therefore, the matrix associated to T is $[T]_\beta^\gamma = \boxed{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix}}$.

- Example: Let $T : P_2(\mathbb{Q}) \to M_{2\times 2}(\mathbb{Q})$ be defined by $T(p) = \begin{bmatrix} p(0) & p(1) \\ p'(0) & p'(1) \end{bmatrix}$. Find the matrix associated to T with respect to the standard bases $\beta = \{1, x, x^2\}$ and $\gamma = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ of $P_2(\mathbb{Q})$ and $M_{2\times 2}(\mathbb{Q})$ respectively.
    - We have $T(1) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = 1e_{1,1} + 1e_{1,2} + 0e_{2,1} + 0e_{2,2}$, $T(x) = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = 0e_{1,1} + 1e_{1,2} + 1e_{2,1} + 1e_{2,2}$, and $T(x^2) = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} = 0e_{1,1} + 1e_{1,2} + 0e_{2,1} + 2e_{2,2}$.
    - Therefore, the matrix associated to T is $[T]_\beta^\gamma = \boxed{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}}$.

- Example: Let $T : P_3(\mathbb{R}) \to P_3(\mathbb{R})$ be defined by $T(p) = \dfrac{12}{x-1}\int_1^x p(t)\,dt$. Find the matrix associated to T with respect to the standard basis $\beta = \gamma = \{1, x, x^2, x^3\}$.
    - We have $T(1) = 12$, $T(x) = 6 + 6x$, $T(x^2) = 4 + 4x + 4x^2$, and $T(x^3) = 3 + 3x + 3x^2 + 3x^3$.
    - Therefore, the matrix associated to T is $[T]_\beta^\gamma = \boxed{\begin{bmatrix} 12 & 6 & 4 & 3 \\ 0 & 6 & 4 & 3 \\ 0 & 0 & 4 & 3 \\ 0 & 0 & 0 & 3 \end{bmatrix}}$.

- We note in particular that if we use different bases, the same linear transformation will generally have different associated matrices:

- Example: Let $I : \mathbb{R}^2 \to \mathbb{R}^2$ be the identity transformation $I(a, b) = \langle a, b\rangle$. Find the matrix associated to I with respect to the standard basis $\beta_1 = \gamma_1 = \{\langle 1, 0\rangle, \langle 0, 1\rangle\}$ of $\mathbb{R}^2$.
  - We have $I(1, 0) = 1\langle 1, 0\rangle + 0\langle 0, 1\rangle$ and $I(0, 1) = 1\langle 1, 0\rangle + 0\langle 0, 1\rangle$.
  - Therefore, the matrix associated to I is $[I]_{\beta_1}^{\gamma_1} = \boxed{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}$, which we recognize as the 2×2 identity matrix.
- Example: Let $I : \mathbb{R}^2 \to \mathbb{R}^2$ be the identity transformation $I(a, b) = \langle a, b\rangle$. Find the matrix associated to I with respect to the bases $\beta_2 = \{\langle 2, -2\rangle, \langle 3, 1\rangle\}$ and $\gamma_2 = \{\langle 1, -1\rangle, \langle 1, 1\rangle\}$ of $\mathbb{R}^2$.
  - We have $I(2, -2) = 2\langle 1, -1\rangle + 0\langle 1, 1\rangle$ and $I(3, 1) = 1\langle 1, -1\rangle + 2\langle 1, 1\rangle$.
  - Therefore, the matrix associated to I is $[I]_{\beta_2}^{\gamma_2} = \boxed{\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}}$.
  - Note that the matrix for this linear transformation is different from the one given above: this should not be surprising, since we are using different bases.

2.2.2 Algebraic Properties of Matrices Associated to Linear Transformations

- We can use the matrix associated to a linear transformation to evaluate the linear transformation on arbitrary vectors, using matrix multiplication.
  - Recall that if A is an $m \times n$ matrix and B is an $n \times q$ matrix, then the matrix product AB is the $m \times q$ matrix whose (i, j)-entry is the sum $(AB)_{i,j} = \sum_{k=1}^{n} a_{i,k}b_{k,j}$.
- Proposition (Associated Matrix Action): Suppose that $\dim(V) = n$ with an ordered basis $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, that $\dim(W) = m$ with an ordered basis $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$, and that $T : V \to W$ is linear. If $M = [T]_\beta^\gamma$ and $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ is a vector in V, then $T(\mathbf{v}) = y_1\mathbf{w}_1 + \cdots + y_m\mathbf{w}_m$, where $y_i = \sum_{k=1}^{n} m_{i,k}x_k$. Equivalently, the coordinate vector $[T(\mathbf{v})]_\gamma$ is given by the matrix product $M\,[\mathbf{v}]_\beta$.
  - Roughly speaking, this proposition says that the linear transformation T acts as left-multiplication by its associated matrix $[T]_\beta^\gamma$, when considered on the level of coordinate vectors.
  - Proof: By properties of linear transformations and the fact that $T(\mathbf{v}_i) = \sum_{j=1}^{n} m_{i,k}\mathbf{w}_i$, we can write

$$T(\mathbf{v}) = T(\sum_{k=1}^{n} x_i\mathbf{v}_i) = \sum_{k=1}^{n} x_i T(\mathbf{v}_i) = \sum_{k=1}^{n} x_i \left[\sum_{i=1}^{n} m_{i,k}\mathbf{w}_i\right] = \sum_{i=1}^{n} \left[\sum_{k=1}^{n} m_{i,k}x_i\right] \mathbf{w}_i$$

    from which we see that $y_i = \sum_{k=1}^{n} m_{i,k}x_k$ as claimed.
- Example: For $T : \mathbb{R}^3 \to \mathbb{R}^2$ with $T(x, y, z) = \langle 2x - y + z, 3x + 4y - 5z\rangle$, and with the standard bases β and γ of $\mathbb{R}^3$ and $\mathbb{R}^2$ respectively, verify that $[T(\mathbf{v})]_\gamma = [T]_\beta^\gamma\,[\mathbf{v}]_\beta$ for $\mathbf{v} = \langle 2, 3, 5\rangle$.
  - We computed earlier that $[T]_\beta^\gamma = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 4 & -5 \end{bmatrix}$ for this transformation.
  - For $\mathbf{v} = \langle 2, 3, 5\rangle$, we have $[\mathbf{v}]_\beta = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$, so $[T]_\beta^\gamma\,[\mathbf{v}]_\beta = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 4 & -5 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 6 \\ -7 \end{bmatrix}$.

- Since $T(\mathbf{v}) = T(2,3,5) = \langle 6, -7 \rangle$, we indeed see that $[T(\mathbf{v})]_\gamma = \begin{bmatrix} 6 \\ -7 \end{bmatrix} = [T]_\beta^\gamma [\mathbf{v}]_\beta$.

- By applying this result to a composition of linear transformations, we can deduce that the matrix associated to a composition of linear transformations is the matrix product of the associated matrices. Indeed, as we mentioned previously the fact that matrix multiplication models the composition of linear transformations is the reason that matrix multiplication is defined the way it is.

- Corollary (Linear Transformations and Matrix Multiplication): Suppose that U, V, and W are finite-dimensional and have ordered bases α, β, and γ respectively, and that $T : U \to V$ and $S : V \to W$ are linear transformations. Then $[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta$, or, in words, the matrix associated to ST is the product of the matrix associated to S with the matrix associated to T.
  - Proof: Let $\mathbf{v}$ be any vector in U. Then by the previous proposition, $[ST]_\alpha^\gamma [\mathbf{v}]_\alpha = [ST(\mathbf{v})]_\gamma$, while $[S]_\beta^\gamma [T]_\alpha^\beta [\mathbf{v}]_\alpha = [S]_\beta^\gamma [T(\mathbf{v})]_\beta = [ST(\mathbf{v})]_\gamma$.
  - Since these two expressions are equal for every vector $\mathbf{v}$ in U, the matrices $[ST]_\alpha^\gamma$ and $[S]_\beta^\gamma [T]_\alpha^\beta$ are equal.

- Example: Let $T : \mathbb{R}^3 \to P_2(\mathbb{R})$ be defined by $T(a,b,c) = (a+b) + (a-2c)x + (a+b+c)x^2$ and $S : P_2(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ be defined by $S(p) = \begin{bmatrix} p(0) & p(1) \\ p'(0) & p'(1) \end{bmatrix}$. For the standard bases $\alpha = \{\langle 1,0,0 \rangle, \langle 0,1,0 \rangle, \langle 0,0,1 \rangle\}$ of $\mathbb{R}^3$, $\beta = \{1, x, x^2\}$ of $P_2(\mathbb{R})$, and $\gamma = \{e_{1,1}, e_{1,2}, e_{2,1}, e_{2,2}\}$ of $M_{2\times 2}(\mathbb{R})$, verify that $[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta$.
  - We computed earlier that $[T]_\alpha^\beta = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix}$ and that $[S]_\beta^\gamma = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$.
  - Thus, $[S]_\beta^\gamma [T]_\alpha^\beta = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 2 & -1 \\ 1 & 0 & -2 \\ 3 & 2 & 0 \end{bmatrix}$.
  - We can also see that $ST(a,b,c) = \begin{bmatrix} a+b & 3a+2b-c \\ a-2c & 3a+2b \end{bmatrix}$ from a direct calculation, so $[ST]_\alpha^\gamma = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 2 & -1 \\ 1 & 0 & -2 \\ 3 & 2 & 0 \end{bmatrix}$. This is indeed equal to $[S]_\beta^\gamma [T]_\alpha^\beta$.

- Once we choose ordered bases β and γ for V and W, we can in fact view linear transformations $T : V \to W$ completely interchangeably with their associated matrices. More explicitly:

- Theorem (Matrices and Linear Spaces): Suppose that $\dim(V) = n$ with V having an ordered basis $\beta = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and that $\dim(W) = m$ with W having an ordered basis $\gamma = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$, where V and W have the same scalar field F. Then the map $\Phi : \mathcal{L}(V,W) \to M_{m\times n}(F)$ defined by $\Phi(T) = [T]_\beta^\gamma$ is an isomorphism.
  - This theorem says that the space $\mathcal{L}(V,W)$ of linear transformations from V to W is isomorphic to the space of $m \times n$ matrices, where the correspondence is given by writing down the associated matrix with respect to the fixed ordered bases β and γ.
  - Proof: By our characterization of isomorphisms, it is enough to show that Φ is linear, one-to-one, and onto.
  - For [T1]: Suppose S and T are elements of $\mathcal{L}(V,W)$. Then by definition, $(S+T)(\mathbf{v}_j) = S(\mathbf{v}_j) + T(\mathbf{v}_j) = \sum_{i=1}^m s_{i,j}\mathbf{w}_i + \sum_{i=1}^m t_{i,j}\mathbf{w}_i = \sum_{i=1}^m (s_{i,j} + t_{i,j})\mathbf{w}_i$, so the (i,j)-entry of $[S+T]_\beta^\gamma$ is the (i,j)-entry of $[S]_\beta^\gamma$ plus the (i,j)-entry of $[T]_\beta^\gamma$. Thus, $[S+T]_\beta^\gamma = [S]_\beta^\gamma + [T]_\beta^\gamma$.
  - For [T2]: Suppose T is an element of $\mathcal{L}(V,W)$ and c is a scalar. Then by definition, $(cT)(\mathbf{v}_j) = cT(\mathbf{v}_j) = c\sum_{i=1}^m t_{i,j}\mathbf{w}_i = \sum_{i=1}^m (ct_{i,j})\mathbf{w}_i$, so the (i,j)-entry of $[cT]_\beta^\gamma$ is c times the (i,j)-entry of $[T]_\beta^\gamma$. Thus, $[cT]_\beta^\gamma = c[T]_\beta^\gamma$.

- The fact that Φ is one-to-one is immediate: if $[T]_\beta^\gamma$ is the zero matrix, then for any $\mathbf{v}$ in V, the coordinate vector $[T(\mathbf{v})]_\gamma$ is the zero vector. Thus, $T(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v}$ in V, so T is the zero transformation.
- Finally, Φ is onto, because a linear transformation is characterized by its values on a basis (and these values can be arbitrary). Explicitly, for any matrix M in $M_{m\times n}(F)$, the linear transformation T specified by choosing $T(\mathbf{v}_j) = \sum_{i=1}^{m} m_{i,j}\mathbf{w}_i$ for each $1 \le j \le n$ has $[T]_\beta^\gamma = M$.

- Corollary: If $\dim(V) = n$ and $\dim(W) = m$, then the dimension of $\mathcal{L}(V, W)$ is mn.
  - Proof: Isomorphisms preserve dimension; the theorem above says that $\mathcal{L}(V, W)$ is isomorphic to $M_{m\times n}(\mathbb{F})$, and the latter has dimension mn.

- We will remark at this juncture that, although we have used matrix multiplication as an ingredient in the proofs above, we did not actually invoke any of the algebraic properties of matrix multiplication (e.g., associativity or distributivity).
  - In fact, we can use the results above to *prove* that matrix multiplication is associative and distributive, by invoking the corresponding facts about linear transformations (which we have already established).

- Theorem (Algebra of Matrix Operations): For any matrices A, B, and C such that the appropriate products are defined, we have $A(BC) = (AB)C$, $(A + B)C = AC + BC$, and $A(B + C) = AB + AC$.
  - Proof: To show $A(BC) = (AB)C$, let U, V, W, X be vector spaces such that $T_3 : U \to V$ has associated matrix A, $T_2 : V \to W$ has associated matrix B, and $T_1 : W \to X$ has associated matrix C, for fixed bases of each space. (The theorems above guarantee that we can make such choices.)
  - Then, since composition of linear transformations is associative, we see that $T_1(T_2T_3) = (T_1T_2)T_3$.
  - Writing down the associated matrix to each transformation, using the theorems above, then immediately yields $A(BC) = (AB)C$.
  - The other properties follow in a similar way.
  - Remark: This method provides a much cleaner proof of these algebraic properties of matrix multiplication than the standard method of multiplying everything out from the definition.

2.2.3 The Rank of a Linear Transformation

- Definition: If A is an $m \times n$ matrix, the rank of A is defined to be the rank of the linear transformation $T : F^n \to F^m$ of left-multiplication by A (namely, $T(\mathbf{v}) = A\mathbf{v}$). Equivalently, the rank is the dimension of $\mathrm{im}(T)$.
  - The most important property of rank is that it is unchanged upon multiplication by an invertible matrix.
  - We will prove this fact for linear transformations in general.

- Proposition (Rank and Isomorphisms): If $T : V \to W$, $L : W \to W$, and $R : V \to V$ are linear transformations with L and R isomorphisms, then $\mathrm{rank}(LT) = \mathrm{rank}(T) = \mathrm{rank}(TR)$.
  - Proof: First consider $LT : V \to W$. Since L is an isomorphism (and therefore one-to-one), when we restrict L to $\mathrm{im}(T)$, it remains one-to-one. Therefore, this restricted map $L_{\mathrm{im}(T)}$ is an isomorphism of $\mathrm{im}(T)$ with $\mathrm{im}(LT)$. Hence $\dim(\mathrm{im}(LT)) = \dim(\mathrm{im}(T))$, and so $\mathrm{rank}(LT) = \mathrm{rank}(T)$.
  - Now consider $TR : V \to W$. Since R is an isomorphism (and therefore onto), $\mathrm{im}(TR) = \mathrm{im}(T)$. Thus, $\mathrm{rank}(AQ) = \dim(\mathrm{im}(TR)) = \dim(\mathrm{im}(T)) = \mathrm{rank}(A)$.

- Corollary (Rank and Invertible Matrices): If A is an $m \times n$ matrix, P is an invertible $m \times m$ matrix, and Q is an invertible $n \times n$ matrix, then $\mathrm{rank}(PA) = \mathrm{rank}(A) = \mathrm{rank}(AQ)$.
  - Proof: Let $V = F^n$ and $W = F^m$, and take $T : V \to W$ to have associated matrix A, $L : W \to W$ to have associated matrix P, and $R : V \to V$ to have associated matrix Q.

- Since P and Q are invertible matrices, L and R are isomorphisms. The previous proposition then yields the results immediately.

- We also remark that the rank of a matrix can be computed from any linear transformation having that associated matrix:

- Corollary: If $T : V \to W$ is a linear transformation and β and γ are ordered bases for V and W respectively, then $\text{rank}(T) = \text{rank}([T]_\beta^\gamma)$.
  - Proof: Let $\dim(V) = n$ and $\dim(W) = m$. The coordinate vector maps associated to β and γ yield isomorphisms $L : \mathbb{F}^n \to W$ and $R : V \to \mathbb{F}^m$.
  - Then $[T]_\beta^\gamma = LTR$ by definition, so $\text{rank}([T]_\beta^\gamma) = \text{rank}(LTR) = \text{rank}(LT) = \text{rank}(T)$ by the proposition above.

- We are often interested in the rank when solving systems of linear equations. The key fact is that row operations do not alter the rank:

- Corollary: Applying elementary row operations to a matrix does not alter its rank.
  - Proof: Each elementary matrix is invertible, and a product of invertible matrices is also invertible.

- The rank of a matrix is easy to compute from its row-echelon form:

- Proposition (Rank and Row-Echelon Form): The rank of any matrix is equal to the number of pivots in (any) row-echelon form.
  - Proof: As noted above, applying row operations to a matrix A does not alter its rank, so the rank of A is equal to the rank of (any) row-echelon form.
  - Thus we may assume B is in row-echelon form. By definition, the rank of B is $\dim(\text{im}(T))$ where $T : F^n \to F^m$ is the linear transformation $T(\mathbf{v}) = B\mathbf{v}$.
  - Furthermore, $\text{im}(T)$ is spanned by $\{T(e_1), T(e_2), \dots, T(e_n)\}$, where $\{e_1, \dots, e_n\}$ is the standard basis of F^n. However, $T(e_j)$ is simply the jth column of B.
  - Now we simply observe that the pivotal columns in a row-echelon matrix are linearly independent (each pivotal column has one more nonzero entry than the previous one) and the nonpivotal columns are spanned by the pivotal columns (since all of their nonzero entries lie above the pivot entries).
  - Thus, the rank of A, which is the same as the dimension of $\text{im}(T)$, is the number of pivots in the row-echelon form of A.

2.2.4 Inverse Transformations and Inverse Matrices

- With minimal additional effort, we can also study the matrices associated to a linear transformation $T : V \to W$ having an inverse $T^{-1} : W \to V$.
  - Recall that if A is an $n \times n$ matrix, we say that A is invertible if there exists an $n \times n$ matrix B such that $AB = I_n = BA$, and we call $B = A^{-1}$ the inverse of A.
  - As we have already mentioned, determining whether a particular matrix is invertible can be achieved by evaluating its determinant, and the inverse matrix itself can be computed using row-reduction.
  - Our focus at the moment is primarily on the theoretical properties of invertible matrices.

- Theorem (Invertible Transformations): Suppose V and W are finite-dimensional with ordered bases β and γ, and that $T : V \to W$ is linear. Then T has an inverse transformation $T^{-1} : W \to V$ if and only if $[T]_\beta^\gamma$ is an invertible matrix, and in such a case, $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.
  - Proof: Let $\dim(V) = n$. First, if T has an inverse T^{-1}, then T is an isomorphism, so $\dim(W) = \dim(V)$.
  - Next, by definition $T^{-1}T(\mathbf{v}) = \mathbf{v}$ for every $\mathbf{v}$ in V, and $TT^{-1}(\mathbf{w}) = \mathbf{w}$ for every $\mathbf{w}$ in W, so $[T^{-1}T]_\beta^\beta$ and $[TT^{-1}]_\gamma^\gamma$ are both equal to the $n \times n$ identity matrix.

- But since $[T^{-1}T]_\beta^\beta = [T^{-1}]_\gamma^\beta [T]_\beta^\gamma$ and $[TT^{-1}]_\gamma^\gamma = [T]_\beta^\gamma [T^{-1}]_\gamma^\beta$, we immediately conclude that $[T]_\beta^\gamma$ is invertible and has inverse $[T^{-1}]_\gamma^\beta$.
- Conversely, suppose that $[T]_\beta^\gamma$ is an invertible matrix, and let S be the linear transformation with associated matrix $([T]_\beta^\gamma)^{-1}$.
- Then $[ST]_\beta^\beta = [S]_\gamma^\beta [T]_\beta^\gamma = I_n$ and similarly $[TS]_\gamma^\gamma = [T]_\beta^\gamma [S]_\gamma^\beta = I_n$. Thus, both ST and TS have associated matrix equal to the identity matrix, so they are both the identity transformation. Hence, T has a two-sided inverse function.

- It turns out that we can weaken the definition of "invertible matrix" slightly without any loss:
- Proposition (One-Sided Inverses): If A and B are $n \times n$ matrices such that $AB = I_n$, then A and B are both invertible and their inverses are each other.
  - What this result says is that a matrix possessing either a left inverse or a right inverse automatically has a two-sided inverse.
  - Proof: Choose an n-dimensional vector space V with a basis β, and let $T : V \to V$ and $S : V \to V$ have associated matrices A and B.
  - Then the associated matrix for ST is the identity matrix, so $ST = I$. But this implies that T is one-to-one, hence (since V is finite-dimensional) an isomorphism. Then $S = (ST)T^{-1} = T^{-1}$ is also an isomorphism, so by the above results, the associated matrices A and B are both invertible.
  - Finally, using associativity, we have $A^{-1} = A^{-1}AB = B$ and similarly $B^{-1} = ABB^{-1} = A$.
- There are many equivalent criteria for when a matrix has an inverse. Here are some of them:
- Proposition (Invertible Matrices): If V is an n-dimensional vector space with ordered basis β, and $T : V \to V$ has associated matrix $A = [T]_\beta^\beta$, the following are equivalent:

1. The matrix A is invertible: there exists an $n \times n$ matrix B with $AB = I_n = BA$.
2. The matrix A has a right inverse: there exists an $n \times n$ matrix B with $AB = I_n$.
3. The matrix A has a left inverse: there exists an $n \times n$ matrix B with $CA = I_n$.
4. The linear transformation T is an isomorphism.
5. The kernel of T consists only of the zero vector.
6. The rank of T (equivalently, the rank of A) is equal to n.
7. The linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.
8. The linear system $A\mathbf{x} = \mathbf{c}$ has exactly one solution $\mathbf{x}$, for any $\mathbf{c}$.
9. The matrix A is row-equivalent to the identity matrix.
10. The determinant of A is nonzero.

  - Proof: The proposition above shows that (1), (2), and (3) are equivalent. Propositions from earlier show (4), (5), and (6) are equivalent, and that (1) and (4) are equivalent, so (1)-(6) are all equivalent.
  - For (7), note that $A\mathbf{x} = \mathbf{0}$ is equivalent to saying that $T(\mathbf{x}) = \mathbf{0}$, so the system $A\mathbf{x} = \mathbf{0}$ has only the solution $\mathbf{x} = \mathbf{0}$ precisely when the kernel of T consists only of the zero vector. Thus, (5) and (7) are equivalent.
  - For (8), note that $A\mathbf{x} = \mathbf{c}$ is equivalent to saying that $T(\mathbf{x}) = \mathbf{c}$, so the system $A\mathbf{x} = \mathbf{c}$ has a unique solution precisely when $T(\mathbf{x}) = \mathbf{c}$ has exactly one solution for $\mathbf{x}$: but this is equivalent to saying that T has an inverse function, which is equivalent to (4).
  - For (9), since row operations do not change the solutions to a system, saying that A is row-equivalent to the identity matrix is equivalent to saying that the system $A\mathbf{x} = \mathbf{c}$ has a unique solution, which is (8).
  - Finally, for (10), we showed previously that $\det(A) \neq 0$ is equivalent to saying that A is row-equivalent to the identity matrix, which is (9).

2.2.5 Change of Basis, Similarity

- As an application of inverse matrices, we can discuss change of basis.
  - As motivation, consider the graph of the equation $6x^2 + 4xy + 9y^2 = 1$ in the plane. Without modifying the equation, it is difficult to determine the shape of the graph of this curve.
  - If, however, we define new variables $s = \frac{1}{\sqrt{5}}x + \frac{2}{\sqrt{5}}y$ and $t = \frac{2}{\sqrt{5}}x - \frac{1}{\sqrt{5}}y$, a short computation will show that the equation $6x^2 + 4xy + 9y^2 = 1$ is equivalent to $2s^2 + t^2 = 5$.
  - Since the vectors $\mathbf{s} = \frac{1}{\sqrt{5}}\langle 1, 2\rangle$ and $\mathbf{t} = \frac{1}{\sqrt{5}}\langle 2, -1\rangle$ both have length 1 and are orthogonal (as their dot product is zero), we can see that the equation $2s^2 + t^2 = 5$ therefore represents an ellipse whose two axes have lengths $\sqrt{5}$ (in the $\mathbf{t}$-direction) and $\sqrt{5/2}$ (in the $\mathbf{s}$-direction).
  - By using the basis $\{\mathbf{s}, \mathbf{t}\}$ for $\mathbb{R}^2$ rather than the standard basis $\{\langle 1, 0\rangle, \langle 0, 1\rangle\}$, we obtain a more useful description of the curve $6x^2 + 4xy + 9y^2 = 1$.
- We now describe, in general, how coordinates of vectors change when we write them in terms of a new basis.
- <u>Definition</u>: Suppose β and γ are two ordered bases of the finite-dimensional vector space V. The <u>change-of-basis matrix</u> is defined to be $Q = [I]_\beta^\gamma$, where I is the identity transformation on V.
  - Note $[I]_\beta^\gamma$ is the matrix whose columns represent vectors in β as linear combinations of vectors in γ.
- <u>Proposition</u> (Change of Basis): Suppose β and γ are two ordered bases of the finite-dimensional vector space V. Then the change-of-basis matrix $[I]_\beta^\gamma$ is invertible, and for any vector $\mathbf{v}$ in V, we have $[\mathbf{v}]_\gamma = [I]_\beta^\gamma [\mathbf{v}]_\beta$.
  - <u>Proof</u>: Since I is invertible, by our earlier results we immediately see that the change-of-basis matrix is invertible and its inverse is $[I]_\gamma^\beta$.
  - Furthermore, by our proposition about the associated matrix action, we have $[I]_\beta^\gamma [\mathbf{v}]_\beta = [I\mathbf{v}]_\gamma = [\mathbf{v}]_\gamma$.
- <u>Example</u>: In $\mathbb{R}^3$, let $\beta = \{\langle 2, 1, 2\rangle, \langle -1, 1, 0\rangle, \langle 3, 1, 3\rangle\}$ and $\gamma = \{\langle 1, 0, 0\rangle, \langle 1, 1, 0\rangle, \langle 1, 1, 1\rangle\}$. Find the change-of-basis matrix $[I]_\beta^\gamma$ and verify that $[\mathbf{v}]_\gamma = [I]_\beta^\gamma [\mathbf{v}]_\beta$ for $\mathbf{v} = \langle 13, 9, 16\rangle$.
  - We compute $\langle 2, 1, 2\rangle = 1\langle 1, 0, 0\rangle - 1\langle 1, 1, 0\rangle + 2\langle 1, 1, 1\rangle$, $\langle -1, 1, 0\rangle = -2\langle 1, 0, 0\rangle + 1\langle 1, 1, 0\rangle + 0\langle 1, 1, 1\rangle$, and $\langle 3, 1, 3\rangle = 2\langle 1, 0, 0\rangle - 2\langle 1, 1, 0\rangle + 3\langle 1, 1, 1\rangle$.
  - Thus, the change-of-basis matrix is $[I]_\beta^\gamma = \boxed{\begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & -2 \\ 2 & 0 & 3 \end{bmatrix}}$.
  - We also calculate $\langle 13, 9, 16\rangle = 2\langle 2, 1, 2\rangle + 3\langle -1, 1, 0\rangle + 4\langle 3, 1, 3\rangle = 4\langle 1, 0, 0\rangle - 7\langle 1, 1, 0\rangle + 16\langle 1, 1, 1\rangle$.
  - Then $[I]_\beta^\gamma [\mathbf{v}]_\beta = \begin{bmatrix} 1 & -2 & 2 \\ -1 & 1 & -2 \\ 2 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ -7 \\ 16 \end{bmatrix} = [\mathbf{v}]_\gamma$, as required.
- If we have a linear transformation $T : V \to W$, we can change basis in both V and W to obtain a new matrix associated to T. This matrix is a product of the original matrix with the appropriate change-of-basis matrices in a natural way:
- <u>Proposition</u> (Change of Basis): Suppose α and β are ordered bases of the finite-dimensional vector space V, that γ and δ are ordered bases of the finite-dimensional vector space W, and that $T : V \to W$ is linear. Then $[T]_\beta^\delta = P[T]_\alpha^\gamma Q^{-1}$, where $P = [I]_\gamma^\delta$ is the change of basis matrix from γ to δ (i.e., from old to new in W) and and $Q = [I]_\alpha^\beta$ is the change of basis matrix from α to β (i.e., from old to new in V).
  - <u>Proof</u>: By the previous proposition on the change of basis matrix, $Q^{-1} = [I]_\beta^\alpha$. Then $P[T]_\alpha^\gamma Q^{-1} = [I]_\gamma^\delta [T]_\alpha^\gamma [I]_\beta^\alpha = [ITI]_\beta^\delta = [T]_\beta^\delta$, as claimed.
- <u>Corollary</u>: Suppose α and β are ordered bases of the finite-dimensional vector space V and $T : V \to V$ is linear. Then $[T]_\beta^\beta = Q[T]_\alpha^\alpha Q^{-1}$ where $Q = [I]_\alpha^\beta$ is the change of basis matrix from α to β.

- Proof: Apply the previous result when $\gamma = \alpha$ and $\delta = \beta$.

- Example: In $P_1(\mathbb{R})$, let $\beta = \{1, 1-x\}$ and $\gamma = \{1+x, x\}$. For $T(p) = p(1) + xp'(x)$, find $[T]_\beta^\beta$ and $[T]_\gamma^\gamma$ and verify that $[T]_\gamma^\gamma = Q[T]_\beta^\beta Q^{-1}$ where $Q = [I]_\beta^\gamma$.
  - We have $T(1) = 1$ and $T(1-x) = -1 + (1-x)$, so $[T]_\beta^\beta = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$.
  - Also, $T(1+x) = 2(1+x) - x$ and $T(x) = 1 + x$, so $[T]_\gamma^\gamma = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$.
  - Since $1 = (1+x) - x$ and $1 - x = (1+x) - 2x$, $Q = [I]_\beta^\gamma = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$.
  - Inversely, since $1 + x = 2 - (1-x)$ and $x = 1 - (1-x)$, we have $Q^{-1} = [I]_\gamma^\beta = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$.
  - Then $Q[T]_\beta^\beta Q^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} = [T]_\gamma^\gamma$ as claimed.

- For a variety of reasons, we will be interested in studying classes of matrices which represent the same linear transformation in different bases. Such matrices have a particular name:

- Definition: We say two $n \times n$ matrices A and B are similar (or conjugate) if there exists an invertible $n \times n$ matrix Q such that $B = QAQ^{-1}$. (We refer to QAQ^{-1} as the conjugate of A by Q.)
  - Notice that if $B = QAQ^{-1}$ then $A = Q^{-1}BQ$, so similarity is a symmetric relation.
  - Example: The matrices $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -1 \\ 2 & -1 \end{bmatrix}$ are similar with $Q = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}$. Explicitly, we can compute that $Q^{-1} = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$, and then see $\begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$, so that $QAQ^{-1} = B$.
  - Remark: The matrix $Q = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$ also has $B = QAQ^{-1}$. In general, if two matrices A and B are similar, then there can be many different matrices Q with $B = QAQ^{-1}$.

- Proposition (Similar Matrices): If A and B are similar $n \times n$ matrices, then there exists a linear transformation $T : V \to V$ on an n-dimensional vector space and ordered bases α and β of V such that $A = [T]_\alpha^\alpha$ and $B = [T]_\beta^\beta$.
  - More simply: If A and B are similar $n \times n$ matrices, then A and B are the associated matrices to some shared linear transformation.
  - Proof: Suppose $B = QAQ^{-1}$ for some Q. Choose any n-dimensional vector space V with ordered basis α, and let $T : V \to V$ be the linear transformation with $A = [T]_\alpha^\alpha$.
  - Take β to be the ordered basis such that $Q = [I]_\alpha^\beta$: in other words, with $\alpha_j = \sum_{i=1}^n Q_{i,j}\beta_i$. (Since Q is invertible, these β_j are a basis for V.) Then $B = QAQ^{-1} = Q[T]_\alpha^\alpha Q^{-1} = [T]_\beta^\beta$ by our results above.

- Similar matrices, owing to the fact that they represent the same linear transformation in different bases, share many algebraic properties.
  - Explicitly, if $B = QAQ^{-1}$ and $D = QCQ^{-1}$, then $B + D = Q(A+C)Q^{-1}$, $BD = Q(AC)Q^{-1}$, and $B^{-1} = Q(A^{-1})Q^{-1}$. This tells us that the sum, product, or inverse of conjugates is the conjugate of the corresponding sum, product, or inverse.

- We later study the following question: given a matrix A, what is the simplest matrix B that A is similar to?

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