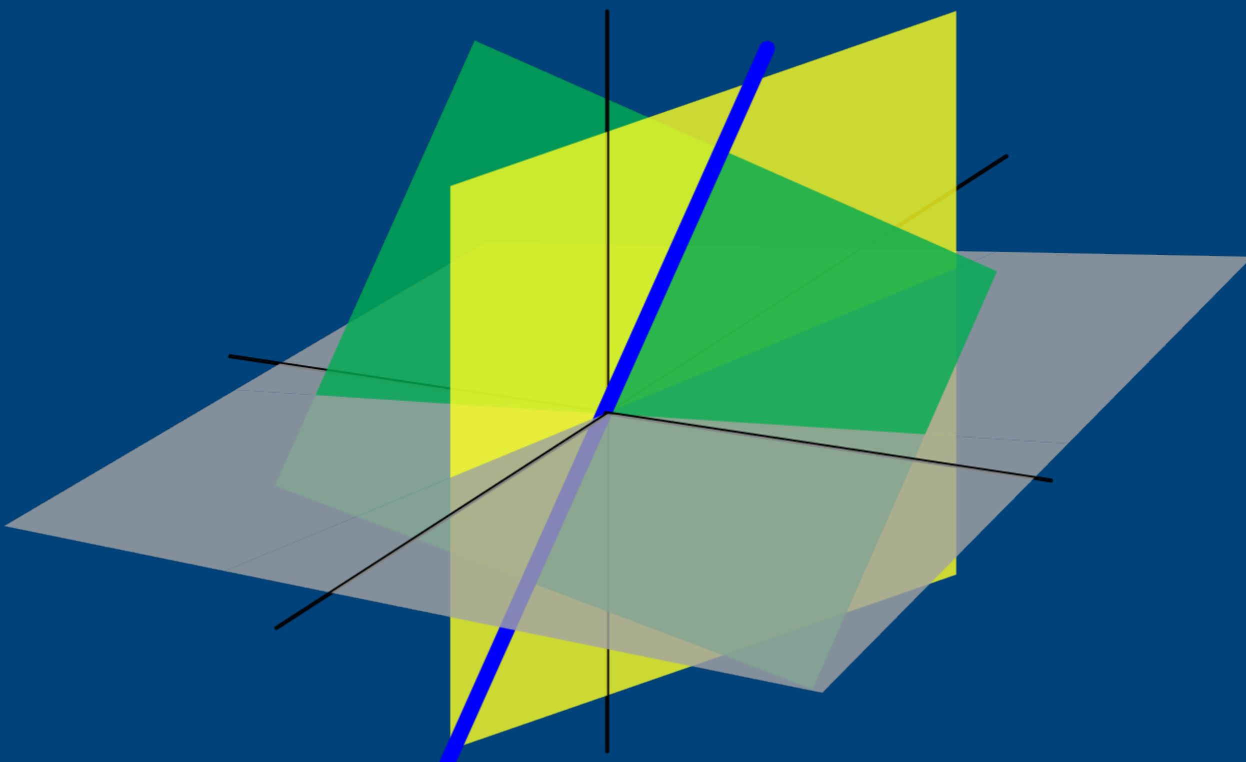
Linear Algebra Done Right

Lecture Notes

"It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment."

CARL FRIEDRICH GAUSS



Lectures by:
Sheldon Axler
Notes by:
Rohit Agarwal
rohaga@berkeley.edu

Contents

1 Vector Spaces

1.1 $\mathbb{R}^n$ and $\mathbb{C}^n$

Definition 1.1 ($\mathbb{R}$) $\mathbb{R}$ denotes the field of real numbers.

Some nonconstant polynomials with real coefficients have no real zeroes. Example: the equation:

$$x^2 + 1 = 0$$

has no real solutions. Thus, we invent a solution called i, such that $i^2 = -1$.

Definition 1.2 (Complex Numbers)

- A complex number is an ordered pair (a, b), where $a, b \in \mathbb{R}$, but we will write this as $a + bi$.
- The set of all complex numbers is denoted by $\mathbb{C}$:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$$

- Addition and multiplication on $\mathbb{C}$ are defined as follows

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

Note 1.1 If $a \in \mathbb{R}$, we identify $a + 0i$ with the real number a. Thus we think of $\mathbb{R}$ as a subset of $\mathbb{C}$. We also usually write $0 + bi$ as just bi, and we usually write $0 + 1i$ as just i. From the definition of multiplication above, we have that $i^2 = -1$.

Note 1.2 (Properties of Complex Arithmetic) $\forall \alpha, \beta \in \mathbb{C}$

- Commutativity

$$\alpha + \beta = \beta + \alpha \text{ and } \alpha\beta = \beta\alpha$$

- Associativity

$$(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda) \text{ and } (\alpha\beta)\lambda = \alpha(\beta\lambda)$$

- Identities

$$\lambda + 0 = \lambda \text{ and } \lambda 1 = \lambda$$

- Additive Inverse

$$\text{For every } \alpha \in \mathbb{C} \text{ there exists a unique } \beta \in \mathbb{C} \text{ such that } \alpha + \beta = 0$$

- Multiplicative Inverse

$$\text{For every } \alpha \in \mathbb{C} \setminus \{0\}, \text{ there exists a unique } \beta \in \mathbb{C} \text{ such that } \alpha\beta = 1$$

- Distributivity

$$\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$$

Definition 1.3 ($\mathbb{F}$) $\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$

Elements of $\mathbb{F}$ are sometimes called scalars. We call it $\mathbb{F}$ because those are both fields.

Now we discuss the idea of a "list." To understand the idea, here some examples of simple sets we have already seen in other mathematics:

- The set $\mathbb{R}^2$, which you can think of as a plane, is the set of all ordered pairs of real numbers:

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$

- The set $\mathbb{R}^3$, which you can think of as ordinary space, is the set of all ordered triples of real numbers:

$$\mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}$$

Definition 1.4 (List) A *list* of *length* n is an ordered collection of n numbers separated by commas and surrounded by parenthesis.

$$\text{i.e. } (x_1, \ldots, x_n)$$

Two lists are equal if an only if they have the same length and the same elements.

Here are some examples of lists from sets we are familiar with:

1. $(7, 3)$ is a list of length 2. Thus $(7, 3) \in \mathbb{R}^2$.
2. $(5, 9, -2)$ is a list of length 3. Thus $(5, 9, -2) \in \mathbb{R}^3$

Definition 1.5 ($\mathbb{F}^n$) $\mathbb{F}^n$ is the set of all lists of length n of elements of $\mathbb{F}$:

$$\mathbb{F}^n = \{(x_1, \ldots, x_n) : x_j \in \mathbb{F} \text{ for } j = 1, \ldots, n\}$$

Elements of $\mathbb{F}^n$ are often called *points* or *vectors*.

It does not matter if these sets have geometric sense. We can manipulate them algebraically. This is where the name linear algebra comes from.

Definition 1.6 (Addition in $\mathbb{F}^n$) Addition in $\mathbb{F}^n$ is defined by adding the corresponding coordinates:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

Definition 1.7 (Scalar Multiplication in $\mathbb{F}^n$) The product of a number $\lambda \in \mathbb{F}$ and a vector $\mathbb{F}^n$ is defined by multiplying each coordinate of the vector by λ:

$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n)$$

Single letters can denote elements of $\mathbb{F}^n$ efficiently. You can say $x + y = z$ instead of saying e.g.

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (z_1, \ldots, z_n)$$

Definition 1.8 (0 list) Let 0 denote the list of length n whose coordinates are all 0:

$$0 = (0, \ldots, 0)$$

It should always be clear from context which 0 you're talking about. For example: we have the following:

Theorem If $x \in \mathbb{F}^n$, then $0x = 0$.

The 0 on the LHS is a scalar in $\mathbb{F}$. The 0 on the RHS is a vector in $\mathbb{F}^n$.

1.2 Definition of a Vector Space

The motivation for the definition of a vectors space comes from the properties of addition and scalar multiplication in $\mathbb{F}^n$:

- Addition is commutative, associative, and has an identity.
- Every element has an additive inverse.
- Scalar multiplication is associative.
- Scalar multiplication by 1 acts as expected.
- Addition and scalar multiplication are connected by distributive properties.

First, let us define what addition/scalar multiplication is.

Definition 1.9 (Addition, Scalar Multiplication)

- An *addition* on a set V is a function that assigns an element $u + w \in V$ to each pair of elements $u, w \in V$
- A *scalar multiplication* on a set V is a function that assigns an element $\lambda u \in V$ to each $\lambda \in \mathbb{F}$ and each $u \in V$

Example 1.1 Suppose V is the set of real valued functions on the interval $[0, 1]$.
For $f, g \in V$ and $\lambda \in \mathbb{R}$, define $f + g$ and λf by:

$$(f + g)(x) = f(x) + g(x)$$

and

$$(\lambda f)(x) = \lambda f(x)$$

Thus $f + g \in V$ and $\lambda f \in V$.

Now, we can define a vector space V. These are based off the properties of $\mathbb{F}^n$:

Definition 1.10 (Vector Space) A *vector space* is a set V along with an addition on V and a scalar multiplication on V such that the following properties hold:

- $u + w = w + u$ for all $u, w \in V$
- $(u + v) + w = u + (v + w)$ and $(ab)u = a(bu)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{F}$
- There exists $0 \in V$ such that $u + 0 = u$ for all $u \in V$
- For every $u \in V$, there exists $w \in V$ such that $u + w = 0$
- $1u = u$ for all $u \in V$
- $a(u + w) = au + aw$ and $(a + b)u = au + bu$ for all $a, b \in \mathbb{F}$ and all $u, w \in V$

Example 1.2 Vector Spaces:

- $\mathbb{F}^n$ with the usual operations of addition and scalar multiplications is a vector space.
- $\mathbb{F}^\infty$ is defined to be the set of all sequences of elements of $\mathbb{F}$:

  $$\{(x_1, x_2, \ldots) : x_j \in \mathbb{F} \text{ for } j = 1, 2, \ldots\}$$

  Addition and scalar multiplication are also defined coordinate-wise. This is also a vector space.
- More generally, if S is a set, let $\mathbb{F}^S$ denote the set of functions from S to $\mathbb{F}$. For $f, g \in \mathbb{F}^S$, the sum $f + g \in \mathbb{F}^S$ is the function defined by:

  $$(f + g)(x) = f(x) + g(x)$$

  for all $x \in S$. For $\lambda \in \mathbb{F}$ and $f \in \mathbb{F}^S$, the product $\lambda f \in \mathbb{F}^S$ is the function defined by

  $$(\lambda f)(x) = \lambda f(x)$$

  for all $x \in S$. With these definitions, $\mathbb{F}^S$ becomes a vector space.

Our first theorem then follows:

Theorem 1.1 (A Number 0 Times a Vector) If V is a vector space, $\forall u \in V$, $0u = 0$.

Proof For arbitrary $u \in V$, we have:

$$\begin{aligned} 0u &= (0+0)u \\ &= 0u + 0u \end{aligned}$$

Adding the additive inverse of $0u$, denoted $-0u$, to both sides of the equation above gives:

$$\begin{aligned} 0u + (-0u) &= 0u + 0u + (-0u) \\ 0 &= 0u \end{aligned}$$

as desired. ∎

Advantages of the abstract approach to vector spaces:

- Can apply what was done in multiple new situations.
- Stripping away inessential properties leads to greater understanding.

If V is a vector space, it would be incorrect to prove that $0u = 0$ for $u \in V$ by writing: Let $u = (x_1, \ldots, x_n)$, thus...

Note 1.3 An element of V is not necessarily of the form $(x_1, \ldots, x_n)$.

1.3 Subspaces

Let's add a new convention. From now on, V denotes a vector space over $\mathbb{F}$ for brevity.

Definition 1.11 (Subspace) A subset U of V is called a subspace of V if U is also a vector space (using the same addition and scalar multiplication as on V).

Example 1.3 $\{(x_1, x_2, 0) : x_1, x_2, \in \mathbb{F}\}$ is a subspace of $\mathbb{F}^3$

Definition 1.12 (Conditions for a Subspace) A subset U of V is a subspace of V if and only if U satisfies the following three conditions:

- $0 \in U$
- $u, w \in U \implies u + w \in U$
- $\lambda \in \mathbb{F}, u \in U \implies \lambda u \in U$

Note that we do not need to check any of the other properties of a vector space because we know that they will hold. Most of these properties are related to the addition and multiplication properties, which we know hold since we're using the same ones.

Example 1.4 Examples of subspaces:

- If $b \in \mathbb{F}$, then
$$\{(x_1, x_2, x_3, x_4) \in \mathbb{F}^4 : x_3 = 5x_4 + b\}$$
is a subspace of $\mathbb{F}^4$ if and only if $b = 0$, in order to have the additive identity in the set.
- The set of continuous real-valued functions on the interval [0,1] is a subspace of $\mathbb{R}^{[0,1]}$. (The zero function is the identity in this case)

- The set of differentiable real-valued functions on $\mathbb{R}$ is a subspace of $\mathbb{R}^{\mathbb{R}}$ (a sum of two differentiable functions is differentiable)
- The set of all sequences of complex numbers with limit 0 is a subspace of $\mathbb{C}^{\infty}$
- The subspaces of $\mathbb{R}^2$ are precisely $\{0\}, \mathbb{R}^2$ and all lines in $\mathbb{R}^2$ through the origin.

Definition 1.13 (Sum of Subsets) Suppose $U_1, \ldots, U_m$ are subsets of V. The sum of $U_1, \ldots, U_m$, denoted $U_1 + \ldots + U_m$, is the set of all possible sums of elements of $U_1, \ldots, U_m$.

$$U_1 + \ldots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \ldots, u_m \in U_m\}$$

Theorem 1.2 (Sum of Subspaces is the Smallest Containing Subspace) Suppose $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \ldots U_m$ is the smallest subspace of V containing $U_1, \ldots, U_m$.

Definition 1.14 (Direct Sum) Suppose $U_1, \ldots, U_m$ are subspaces of V.
The sum $U_1 + \ldots + U_m$ is called a direct sum if each element of $U_1 + \ldots + U_m$ can be written in only one way as a sum $u_1 + \ldots + u_m$ where each u_j is in U_j
If the sum is indeed a direct sum, we use $\oplus$ between the symbols to denote that it is a direct sum.

Example 1.5 Suppose

$$U = \{(x, y, 0) \in \mathbb{F}^3 : x, y \in \mathbb{F}\}, W = \{(0, 0, z) \in \mathbb{F}^3 : z \in \mathbb{F}\}$$

Thus, $\mathbb{F}^3 = U \oplus W$.

These two theorems make it easy to see whether something is a direct sum.

Theorem 1.3 (Condition for a Direct Sum) Suppose $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \ldots + U_m$ is a direct sum if and only if the only way to write 0 as a sum $u_1 + \ldots + u_m$, where each u_j is in U_j, is by taking each u_j equal to 0.

Theorem 1.4 (Direct Sum of Two Subspaces) Suppose U and W are subspaces of V. Then $U + W$ is a direct sum if and only if $U \cap W = \{0\}$.

2 Finite-Dimensional Vector Spaces

2.1 Span and Linear Independence

Definition 2.1 (List) A list of length n is an ordered collection of n elements (which might be numbers, other lists, or more abstract entities) separated by commas (and perhaps surrounded by parentheses).

Definition 2.2 (Linear Combination) A linear combination of a list $v_1, \ldots, v_m$ of vectors in V is a vector of the form

$$a_1 v_1 + \ldots + a_m v_m$$

where $a_1, \ldots, a_m \in \mathbb{F}$.

- $(13, -1, 7)$ is a linear combination of $(2, 1, -1), (1, -2, 4)$ because:

$$5(2, 1, -1) + 3(1, -2, 4) = (13, -1, 7)$$

- $(13, -1, 6)$ is a linear combination of $(2, 1, -1), (1, -2, 4)$ because: there do not exist numbers $a_1, a_2 \in \mathbb{F}$ such that:

$$a_1(2, 1, -1) + a_2(1, -2, 4) = (13, -1, 7)$$

Definition 2.3 (Span) The set of all linear combinations of a list of vectors $v_1, \ldots, v_m$ in V is called the span of $v_1, \ldots, v_m$, denoted span$(v_1, \ldots, v_m)$. In other words

$$\text{span}(v_1, \ldots, v_m) = \{a_1 v_1, + \cdots + a_m v_m : a_1, \ldots, a_m \in \mathbb{F}\}$$

The previous example shows that in $\mathbb{F}^3$:

- $(13, -1, 7) \in \text{span}((2, 1, -1), (1, -2, 4))$
- $(13, -1, 6) \notin \text{span}((2, 1, -1), (1, -2, 4))$

Definition 2.4 The span of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.

Definition 2.5 (Finite-Dimensional Vector Space) A vector space is called **finite-dimensional** if the span of some list of vectors in it is the entire vector space.

Note that we have defined all lists to be finite (have a length that is a natural n). This means that we do not have specify this.

Example 2.1 $\mathbb{F}^3$ is finite-dimensional because:

$$\mathbb{F}^3 = \text{span}((1, 0, 0), (0, 1, 0), (0, 0, 1))$$

Definition 2.6 (Infinite-Dimensional Vector Space) A vector space is called **infinite-dimensional** if it is not finite-dimensional.

Example 2.2 $\mathbb{F}^\infty$ is infinite-dimensional.

Linear algebra is the study of linear maps on finite-dimensional vector spaces.

Definition 2.7 (Linear Independence) A list $v_1, \ldots, v_m$ of vectors in V is called **linearly independent** if the only choice of $a_1, \ldots, a_m \in \mathbb{F}$ that makes $a_1 v_1 + \cdots + a_m v_m$ equal 0 is $a_1 = \cdots = a_m = 0$.

Example 2.3 Examples of linearly independent lists:

- A list v of one vector $v \in V$ is linearly independent iff $v \neq 0$
- A list of two vectors in V is linearly independent iff neither vector is scalar multiple of the other.
- The list $1, x, \ldots, x^m$ is linearly independent in $\mathbb{R}^{\mathbb{R}}$ for each nonnegative integer m. The reason is that the subspace spanned by these vectors represent all the polynomials of degree up to m. The only way that a polynomial can be identically 0 (the identity) is if all the coefficients are 0.

Definition 2.8 (Linear Dependence) A list of vectors in V is called **linearly dependent** if it is not linearly independent. Alternatively, a list $v_1, v_2, \ldots, v_n$ of vectors in V is linearly dependent if there exist $a_1, a_2, \ldots, a_n \in \mathbb{F}$, not all 0, such that $a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = 0$.

Example 2.4 Examples of linearly dependent lists:

- $(2, 3, 1), (1, -1, 2), (7, 3, 8)$ is linearly dependent in $\mathbb{F}^3$ because
$$2(2, 3, 1) + 3(1, -1, 2) + (-1)(7, 3, 8) = (0, 0, 0)$$
- Every list of vectors in V containing the 0 vector is linearly dependent.
- If some vector in a list of vectors in V is a linear combination of the other vectors, then the list in linearly dependent.

Theorem 2.1 (Linear Dependence Lemma) Suppose $v_1, v_2, \ldots, v_n$ is a linearly dependent list in V. Then there exists $j \in \{1, 2, \ldots, m\}$ such that the following hold:

- $v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$
- If the jth term is removed from $v_1, v_2, \ldots, v_n$, the span of the remaining list equals $\operatorname{span}(v_1, v_2, \ldots, v_n)$.

This captures the idea of redundancy. Let's look at an example:

$(2, 3, 1), (1, -1, 2), (7, 3, 8)$ is linearly dependent in $\mathbb{F}^3$.

So, $(7, 3, 8) \in \operatorname{span}(2, 3, 1), (1, -1, 2)$ and we also see that without it, the span remains the same.

Theorem 2.2 (Length of linearly independent list ≤ length of spanning list) In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

Proof Suppose $u_1, u_2, \ldots, u_m$ is linearly independent in V. Suppose also that $w_1, w_2, \ldots, w_n$ spans V. We need to prove that $m \leq n$. We do so through the multi-step process described below.

Step 1
Let B be the list $w_1, w_2, \ldots, w_n$, which spans V. Thus adjoining any vector in V to this list produces a linearly dependent list (because the newly adjoined vector can be written as a linear combination of the other vectors). In particular

$$u_1, w_1, w_2, \ldots, w_n$$

is linearly dependent. Thus by the Linear Dependence Lemma, we can remove one of the w's so that the new list B (of length n) consisting of u_1 and the reamining w's spans V.

Step j
The list B (of length n) from step $j - 1$ spans V. Thus adjoining any vector to this list produces a linearly dependent list. In particular, the list of length $n + 1$ is obtained by adjoining u_j to B, placing it just after $u_1, \ldots, u_{j-1}$, is linearly

dependent. By the Linear Dependence Lemma, one of the vectors in this list is in the span of the previous ones, and because $u_1, u_2, \ldots, u_j$ is linearly independent, this vector is one of the w's, note one of the u's. We can remove that w from B so that the new list B (of length n) consisting of $u_1, \ldots, u_j$ and the remaining w's spans V.

After step m, we have added all the u's and the process stops. At each step as we add a u to B, the Linear Dependence Lemma implies that there is some w to remove. Thus there are at least as many w's as u's. ■

Let's apply the theorem to claim the following:

Example 2.5 Applications of Theorem 2.2

- The list $(1,2,3)$, $(4,5,8)$, $(9,6,7)$, $(-3,2,8)$ is not linearly independent because the list $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ spans $\mathbb{R}^3$. The theorem is applied here because something bigger than a spanning list cannot possibly be a linearly independent list.
- The list $(1,2,3,-5)$, $(4,5,8,3)$, $(9,6,7,-1)$ does not span $\mathbb{R}^4$ because the list $(1,0,0,0)$, $(0,1,0,0)$, $(0,0,1,0)$, $(0,0,0,1)$ is linearly independent in $\mathbb{R}^4$. The theorem applied here because the contrapositive is that no spanning list can be smaller than a linearly independent one, which shows that the smaller list spanning is impossible.

2.2 Bases

Definition 2.9 ($\mathcal{P}_m(\mathbb{F})$) For m a nonnegative integer, $\mathcal{P}_m(\mathbb{F})$ denotes the set of polynomials with coefficients $\mathbb{F}$ and degree at most m.

Example: $(3+2i)z^2 + 4iz + 9 \in \mathcal{P}_{20}(\mathbb{C})$

Definition 2.10 (basis) A **basis** of V is a list of vectors in V that is linearly independent and spans V.

Example 2.6 Examples of bases:

- The list $(1,0,\ldots,0)$, $(0,1,0,\ldots,0)$, $\ldots$, $(0,\ldots,0,1)$ is a basis of $\mathbb{F}^n$, called the **standard basis** of $\mathbb{F}^n$.
- The list $(1,1,0)$,$(0,0,1)$ is a basis of $\{(x,x,y) \in \mathbb{F}^3 : x, y \in \mathbb{F}\}$
- The list $(1,-1,0)$, $(1,0,-1)$ is a basis of

$$\{(x,y,z) \in \mathbb{F}^3 : x+y+z=0\}.$$

- The list $1, z, \ldots, z^m$ is a basis of $\mathcal{P}_m(\mathbb{F})$.

Sometimes it's useful to see some non-examples of bases.

Example 2.7 Non-example of Basis:

- The list $(1,2,-4)$, $(7,-5,6)$ is linearly independent in $\mathbb{F}^3$ but is not a basis of $\mathbb{F}^3$ because it does not span $\mathbb{F}^3$

Theorem 2.3 (Basis Gives Unique Representation as Linear Combination) A list $v_1, v_2, \ldots, v_n$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form:

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$$

where $a_1, \ldots, a_n \in \mathbb{F}$.

Example 2.8 The list $(1,-1,0),(1,0,-1)$ is a basis of

$$\{(x,y,z)\in\mathbb{F}^3 : x+y+z=0\}.$$

If $(x,y,z)\in V$, then

$$(x,y,z) = -y(1,-1,0)+(-z)(1,0,-1)$$

Theorem 2.4 (Every Spanning List Contains a Basis) Every spanning list in a vector can be reduced to a basis of the vector space.

Theorem 2.5 (Basis of Finite-Dimensional Vector Space) Every finite-dimensional vector space has a basis.

Theorem 2.6 (Every Linearly Independent List Can be Extended to a Basis) Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.

2.3 Dimension

Intuitively, we want the dimension of $\mathbb{F}^n$ to be n. Perhaps we should define it by the size of the basis of a vector space. However, we need to make sure there aren't multiple bases with the same length.

Theorem 2.7 (Basis Length does not depend on Basis) Any two bases of a vector space have the same length.

Proof Suppose B_1 and B_2 are two bases of V.
Since these are bases, B_1 is linearly independent in V and B_2 spans V, so $\text{length}(B_1)\le\text{length}(B_2)$.
Swapping the roles, we also know $\text{length}(B_1)\ge\text{length}(B_2)$.
Thus, $\text{length}(B_1)=\text{length}(B_2)$.

Definition 2.11 (Dimension)

- The **dimension** of a finite-dimensional vector space is the length of any basis of the vector space.
- The dimension of V (if V is finite-dimensional) is denoted by $\dim V$.

Example 2.9 Examples of dimension:

- $\dim\mathbb{F}^n = n$ because the standard basis of $\mathbb{F}^n$ has length n.
- $\dim\mathcal{P}_m(\mathbb{F}) = m+1$ beacause the basis $1,z,\ldots,z^m$ has $m+1$ vectors.

Theorem 2.8 (Linearly Independent List of the Right Length is a Basis) Suppose V is finite-dimensional. Then every linearly independent list of vectors in V with length $\dim V$ is a basis of V.

Proof Suppose $\dim V = n$ and $v_1,v_2,\ldots,v_n$ is linearly independent in V.
We know $v_1,v_2,\ldots,v_n$ can be extended to a basis of V. However, every basis of V has length n, so in this case the extension is the trivial one, meaning that no elements are adjoined to $v_1,v_2,\ldots,v_n$.
In other words, $v_1,v_2,\ldots,v_n$ is a basis of V.

Theorem 2.9 (Spanning List of the Right Length is a Basis) Suppose V is finite-dimensional. Then every spanning list of vectors in V with length $\dim V$ is a basis of V.

Proof Suppose $\dim V = n$ and $v_1,v_2,\ldots,v_n$ spans V.
The list $v_1,v_2,\ldots,v_n$ can be reduced to a basis of V. However, every basis of V has length n, so in this case the reduction

is the trivial one, meaning that no elements are removed from $v_1, v_2, \ldots, v_n$
In other words, $v_1, v_2, \ldots, v_n$ is a basis of V.

Theorem 2.10 (Dimension of a Sum) If U_1 and U_2 are subspaces of a finite-dimensional vector space, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

3 Linear Maps

3.1 Vector Space of Linear Maps

Now, we may need more vector spaces, so let V AND W denoting vector spaces over $\mathbb{F}$.

Definition 3.1 ($\mathcal{P}(\mathbb{F})$) $\mathcal{P}(\mathbb{F})$ is the vector space of all polynomials with coefficients in $\mathbb{F}$.

Definition 3.2 (Linear Map) A **linear map** from V to W is a function $T : V \to W$ with the following properties:

- additivity: $T(u_1 + u_2) = Tu_1 + Tu_2$ for all $u_1, u_2 \in V$
- homogeneity: $T(\lambda u) = \lambda(Tu)$ for all $\lambda \in \mathbb{F}$ and all $u \in V$

For linear maps, we often use the notation Tu as well as the more standard functional notation $T(u)$.

Definition 3.3 ($\mathcal{L}(V, W)$) The set of all linear maps from V to W is denoted $\mathcal{L}(V, W)$.

Example 3.1 Examples of Linear Maps:

- Zero: Define $0 \in \mathcal{L}(V,W)$ by $0u = 0$ for all $u \in V$.
- Identity map: Define $I \in \mathcal{L}(V,V)$ by $Iu = u$ for all $u \in V$.
- Differentiation: Define $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ by $Dp = p'$.
- Integration: Define $T \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathbb{R})$ by $Tp = \int_0^1 p(x)\, \mathrm{d}x$.
- Multiplication by x^2: Define $T \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ by
$$(Tp)(x) = x^2 p(x)$$
for $x \in \mathbb{R}$.
- Backward shift: Define $T \in \mathcal{L}(\mathbb{F}^\infty, \mathbb{F}^\infty)$ by
$$T(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots).$$
- From $\mathbb{R}^3$ to $\mathbb{R}^2$: Define $T \in \mathcal{L}\left(\mathbb{R}^3, \mathbb{R}^2\right)$ by
$$T(x, y, z) = (2x - y + 3z, 7x + 5y - 6z).$$

Theorem 3.1 Suppose $v_1, v_2, \dots, v_n$ is a basis of V and $w_1, w_2, \dots, w_n \in W$. Then there exists a unique linear map $T : V \to W$ such that
$$Tv_j = w_j$$
for each $j = 1, \dots, n$.

Proof Define $T : V \to W$ by
$$T(a_1 v_1 + a_2 v_2 + \cdots + a_n v_n) = a_1 w_1 + \cdots + a_n w_n,$$
where $a_1, a_2, \dots, a_n$ are arbitrary elements of $\mathbb{F}$.
It is straightforward to check the above map is additive, just take all the coefficients except a_i to be 0. The distributive property handles homogeneity.
There cannot be another such map because if you add all the constraints together, you get precisely this relation. ■

Definition 3.4 (Addition and Scalar Multiplication on $\mathcal{L}(V, W)$) Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$. The sum $S + T$ is defined as:

$$(S + T)(u) = Su + Tu$$

and the product λT is defined as:

$$(\lambda T)(u) = \lambda(Tu)$$

for all $u \in V$.
Clearly, these maps are also linear maps, thus stay in the set.

Note 3.1 ($\mathcal{L}(V, W)$ is a Vector Space) With the operations of addition and scalar multiplication as defined above, $\mathcal{L}(V, W)$ is a vector space.

Definition 3.5 (Product of Linear Maps) If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then the product $ST \in \mathcal{L}(U, W)$ is defined by

$$(ST)(u) = S(Tu)$$

for $u \in U$.

Note 3.2 (Algebraic Properties of Products of Linear Maps)

- Associativity: $(T_1T_2)T_3 = T_1(T_2T_3)$
- Identity: $TI = IT = T$ (note this may be two different I's)
- Distributive Properties: $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$

Theorem 3.2 (Linear Maps take 0 to 0) Suppose T is a linear map from V to W. Then $T(0) = 0$.

There's a tricky bit about the word "linear". In calculus, we say any $f(x) = mx + b$, this is termed linear. However, in the sense of vector spaces, this function is only linear if and only if $b = 0$.

3.2 Null Spaces and Ranges

Definition 3.6 (Null Space) For $T \in \mathcal{L}(V, W)$, the **null space** of T, denoted T, is the subset of V containing those vectors that T maps to 0:

$$\operatorname{null} T = \{u \in V : Tu = 0\}$$

Example 3.2 Examples of Null Spaces:

- Suppose T is the zero map form V to W; in other words, $Tu = 0$ for every $u \in V$. Then $\operatorname{null} T = V$.
- Suppose $\phi \in \mathcal{L}\left(\mathbb{C}^3, \mathbb{C}\right)$ is defined by $\phi(z_1, z_2, z_3) = z_1 + 2z_2 + 3z_3$. Then $\operatorname{null} \phi = \{(z_1, z_2, z_3) \in \mathbb{C}^3 : z_1 + 2z_2 + 3z_3 = 0\}$.
- Consider D, the differentiation map. The only functions whose derivative equals zero is the constant functions. Thus, the null space is the set of all constant functions.

Theorem 3.3 (The Null Space is a Subspace) Suppose $T \in \mathcal{L}(V, W)$. Then $\operatorname{null} T$ is a subspace of V.

Definition 3.7 (Injective) A function $T : V \to W$ is called **injective** or **one-to-one** if $Tu = Tv$ implies $u = v$.

Theorem 3.4 Let $T \in \mathcal{L}(V, W)$. Then T is injective if and only if $\operatorname{null} T = 0$.

Definition 3.8 (Range) For $T \in \mathcal{L}(V, W)$, the **range** of T is the subset of W consisting of those vectors that are of the form Tu for some $u \in V$:

$$\operatorname{range} T = \{Tu : u \in V\}$$

Example 3.3 Ranges:

- Suppose T is the zero map from V to W; in other words, $Tu = 0$ for every $u \in V$. Then $\operatorname{range} T = \{0\}$.
- Suppose $T \in \mathcal{L}\left(\mathbb{R}^2, \mathbb{R}^3\right)$ is defined by $T(x,y) = (2x, 5y, x + y)$, then $\operatorname{range} T = \{(2x, 5y, x + y) : x, y \in \mathbb{R}\}$. A basis of $\operatorname{range} T$ is $(2, 0, 1)$, $(0, 5, 1)$.
- Consider the differentiation map $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}),\mathcal{P}(\mathbb{R}))$. Since every polynomial $q \in \mathcal{P}(\mathbb{R})$ has a polynomial $p \in \mathcal{P}(\mathbb{R})$ such that $p' = q$, the range of D is $\mathcal{P}(\mathbb{R})$.

Theorem 3.5 (The Range is a Subspace) If $T \in \mathcal{L}(V, W)$, then $\operatorname{range} T$ is a subspace of W.

Definition 3.9 (Surjective) A function $T : V \to W$ is called **surjective** or **onto** if its range equals W.

Theorem 3.6 (Fundamental Theorem of Linear Maps) Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Proof (Proof Sketch)
Let $u_1, \ldots, u_m$ be a basis of $\operatorname{null} T$; thus $\dim \operatorname{null} T = m$.
The linear independent list $u_1, \ldots u_m$ can be extended to a basis

$$u_1, u_2, \ldots, u_m, v_1, v_2, \ldots, v_n$$

Thus $\dim V = m + n$. To complete the proof, we need to show that $\dim \operatorname{range} T = n$. We do this by proving that $Tv_1, \ldots, TV_n$ is a basis of $\operatorname{range} T$. ■

Theorem 3.7 (A Map to a Smaller Dimensional Space is Not Injective) Suppose V and W are finite-dimensional vector spaces such that $\dim V > \dim W$. Then no linear map from V to W is injective.

Proof Suppose $T \in \mathcal{L}(V, W)$. Because

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$$

and

$$\dim V > \dim W \geq \dim \operatorname{range} T$$

we have $\dim \operatorname{null} T > 0$. Thus T is not injective. ■

Theorem 3.8 (A Map to a Larger Dimensional Space is Not Surjective) Suppose V and W are finite-dimensional vector spaces such that $\dim V < \dim W$. Then no linear map from V to W is surjective.

Proof Suppose $T \in \mathcal{L}(V, W)$. Because

$$\dim V = \dim T + \dim \operatorname{range} T$$

we have

$$\dim \operatorname{range} T \leq \dim V < \dim W$$

Thus, T is not surjective. ■

Now we can use these results to prove some facts about a related subject, the theory of systems of linear equations.

Definition 3.10 (Homogenous Linear Equations) Fix positive integers m and n and let $A_{j,k} \in \mathbb{F}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$. Consider the homogeneous system of linear equations

$$\sum_{k=1}^{n} A_{1,k} x_k = 0$$
$$\vdots \sum_{k=1}^{n} A_{m,k} x_k \qquad = 0$$

These are called homogenous because the constant terms are all 0.

We wish to ask the following: do there exist solutions other than the trivial solution, i.e. $x_1 = \cdots = x_n = 0$?

Define $T : \mathbb{F}^n \to \mathbb{F}^m$ by

$$T(x_1, \ldots, x_n) = \left(\sum_{k=1}^{n} A_{1,k} x_k, \ldots, \sum_{k=1}^{n} A_{m,k} x_k \right)$$

The equation $T(x_1, \ldots, x_n) = 0$ is the same as the homogeneous system of linear equations above. This is asking if $T = 0$, which is the same asking: is T injecive?

Well, we know T is not injective if $\dim \mathbb{F}^n > \dim \mathbb{F}^m$, in other words if $n > m$, so we have the following result:

Theorem 3.9 (Homogenous System of Linear Equations) A homogeneous system of linear equations with more variables than equations has nonzero solutions.

Now, let us talk about other types of systems of linear equations.

Definition 3.11 (Inhomogenous Linear Equations) Fix positive integers m and n and let $A_{j,k} \in \mathbb{F}$ for $j = 1, \ldots, m$ and $k = 1, \ldots, n$. Consider the inhomogeneous system of linear equations

$$\sum_{k=1}^{n} A_{1,k} x_k = c_1$$
$$\vdots \sum_{k=1}^{n} A_{m,k} x_k \qquad = c_m$$

These are called inhomogenous because the constant terms are not all 0.

Now we wonder the following: is there some choice of $c_1, c_2, \ldots, c_m \in \mathbb{F}$ such that no solution exists?

Define $T : \mathbb{F}^n \to \mathbb{F}^m$ by

$$T(x_1, \ldots, x_n) = \left(\sum k = 1^n A_{1,k} x_k, \ldots, \sum k = 1^n A_{m,k} x_k \right)$$

The equation $T(x_1, \ldots, x_n) = (c_1, c_2, \ldots, c_m)$ is the same as the inhomogeneous system of linear equations above. This is the same as asking: is T surjective?

We know T is not surjective if $m > n$ (similar to previous logic), so we have the following result:

Theorem 3.10 (Inhomogenous System of Linear Equations) An inhomogeneous system of linear equations with more equations than variables has no solution for some choice of constant terms.

3.3 Matrices

Definition 3.12 (Matrix) Let m and n denote positive integers. An m-by-n **matrix** A is a rectangular array of elements of $\mathbb{F}$ with m rows and n columns:

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix}$$

The notation $A_{j,k}$ denotes the entry in row j, column k of A.

The first index refers to the row numbers and the second index refers to column numbers.

Thus $A_{2,3}$ refers to the entry in the second row, third column of A.

Definition 3.13 (Matrix of a Linear Map, $\mathcal{M}(T)$) Suppose $T \in \mathbb{L}[V,W]$ and $v_1, v_2, \dots, v_n$ is a basis of V and $w_1, w_2, \dots, w_m$ is a basis of W. The **matrix** of T with respect to these bases is the m-by-n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \dots + A_{m,k}w_m$$

If the bases are not clear from the context, then the notation

$$\mathcal{M}(T, (v_1, v_2, \dots, v_n), (w_1, w_2, \dots, w_m))$$

is used.

$$\mathcal{M}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & \dots & v_k & \dots & v_n \end{array} \\ \left(\begin{array}{ccccc} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{array} \right) \end{array}.$$

To understand what the matrix really means, fix a column k the kth column of $\mathcal{M}(T)$ consists of the scalars needed to write

$$Tv_k = \sum_{j=1}^{m} A_{j,k}w_j.$$

The picture above should remind you that Tv_k can be computed from $\mathcal{M}(T)$ by multiplying each entry in the kth column by the corresponding w_j from the left column, and then adding up the resulting vectors.

Example 3.4 (Matrices) Suppose $T \in \mathcal{L}\left(\mathbb{F}^2, \mathbb{F}^3\right)$ is defined by:

$$T(x,y) = (x + 3y, 2x + 5y, 7x + 9y)$$

Because $T(1,0) = (1,2,7)$ and $T(0,1) = (3,5,9)$, the matrix of T with respect to the standard bases is the 3-by-2 matrix

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{pmatrix}$$

Suppose $D \in \mathcal{L}(\mathcal{P}_3(\mathbb{R}), \mathcal{P}_2(\mathbb{R}))$ is the differentiation map. Beacuse $(x^n)' = nx^{n-1}$, the matrix of D with respect to the standard bases of $\mathcal{P}_3(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$ is the 3-by-4 matrix

$$\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

Note that $\mathcal{M}(T)$ contains all the information about T. The matrix contains all the information about how the basis vectors transform under T, however, under linearity, we can write every vector in V as a linear combination of the basis and since T is linear we can see how any vector transforms.

Definition 3.14 (Matrix Addition) The sum of two matrices of the same size is the matrix obtained by adding corresponding entries in the matrices:

$$\begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix} + \begin{pmatrix} C_{1,1} & \dots & C_{1,n} \\ \vdots & & \vdots \\ C_{m,1} & \dots & C_{m,n} \end{pmatrix}$$

$$= \begin{pmatrix} A_{1,1} + C_{1,1} & \dots & A_{1,n} + C_{1,n} \\ \vdots & & \vdots \\ A_{m,1} + C_{m,1} & \dots & A_{m,n} + C_{m,n} \end{pmatrix}$$

In other words, $(A + C)_{j,k} = A_{j,k} + C_{j,k}$.

In the following result, we assume that the same bases are used for $\mathcal{M}(S + T)$, $\mathcal{M}(S)$, and $\mathcal{M}(T)$.

Theorem 3.11 (Addition of Matrices) Suppose $S, T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$.

Definition 3.15 (Scalar Multiplication of a Matrix) The produce of a scalar and a matrix is the matrix obtained by multiplying each entry in the matrix by the scalar:

$$\lambda \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1} & \dots & \lambda A_{1,n} \\ \vdots & & \vdots \\ \lambda A_{m,1} & \dots & \lambda A_{m,n} \end{pmatrix}$$

In other words, $(\lambda A)_{j,k} = \lambda A_{j,k}$.

In the following result, we assume that the same bases are used for $\mathcal{M}(\lambda T)$ and $\mathcal{M}(T)$.

Theorem 3.12 (The Matrix of a Scalar Times a Linear Map) Suppose $\lambda \in \mathbb{F}$ and $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.

Note 3.3 ($\mathbb{F}^{m,n}$) For m and n positive integers, the set of all m-by-n matrices with entries in $\mathbb{F}$ is denoted by $\mathbb{F}^{m,n}$.

Theorem 3.13 ($\dim \mathbb{F}^{m,n} = mn$) Suppose m and n are positive integers. With addition and scalar multiplication defined as above, $\mathbb{F}^{m,n}$ is a vector space with dimension mn.

Now we want to define matrix multiplication.

Consider the following vector spaces:

- V with basis $v_1, v_2, \ldots, v_n$.
- W with basis $w_1, w_2, \ldots, w_m$.
- U with basis $u_1, u_2, \ldots, u_p$.

Consider linear maps $T : U \to V$ and $S : V \to W$. We want to define matrix multiplication such that

$$\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$$

We can try $(AB)_{j,k} = A_{jk} \times B_{j,k}$, however this will not line up with our definition for linear maps.

Instead consider the following, suppose $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = C$.

For $1 \leq k \leq p$, we have

$$\begin{aligned}(ST)u_k &= S\Big(\sum r = 1^n C_{r,k} v_r\Big)\\ &= \sum r = 1^n C_{r,k} S v_r\\ &= \sum r = 1^n C_{r,k} \sum j = 1^m A_{j,r} w_j\\ &= \sum r = 1^n \Big(\sum j = 1^m A_{j,r} C_{r,k}\Big) w_j.\end{aligned}$$

Thus, $\mathcal{M}(ST)$ is the m-by-p matrix whose entry in row j, column k, is

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}$$

We can collect this in the following result:

Definition 3.16 Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then, AC is the m-by-p matrix whose entry in row j, column k, is

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}$$

The entry in row j, column k, of AC is computed by taking row j of A and column k of C, multiplying together corresponding entries, and then summing.

In addition, matrix multiplication is only defined if the amount of columns of A is the same as the amount of columns of C.

In the following result, the same bases are used in considering linear maps with shared vector spaces.

Definition 3.17 Suppose A is an m-by-n matrix.

- If $1 \leq j \leq m$, then $A_{j,\cdot}$ denotes the 1-by-n matrix consisting of row j of A.
- If $1 \leq k \leq n$, then $A_{\cdot,k}$ denotes the m-by-1 matrix consisting of column k of A.

Here are some alternate ways to think about matrix multiplication.

Theorem 3.14 Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then

$$(AC)_{j,k} = A_{j,\cdot} C_{\cdot,k}$$

$$(AC)_{\cdot,k} = AC_{\cdot,k}$$

$$(AC)_{j,\cdot} = A_{j,\cdot}C$$

Theorem 3.15 (Linear Combination of Columns) Suppose A is an m-by-n matrix and $c = \begin{pmatrix} c_1 \\ \dots \\ c_n \end{pmatrix}$ is an n-by-1 matrix. Then $Ac = c_1 A_{\cdot,1} + \cdots + c_n A_{\cdot,n}$.

In other words, Ac is a linear combination of the columns of A, with the scalars that multiply the columns coming from c.

Theorem 3.16 Suppose $a = \begin{pmatrix} a_1 & \dots & a_n \end{pmatrix}$ is a 1-by-n matrix and C is an n-by-p matrix. Then

$$aC = a_1 C_{1,\cdot} + \cdots + a_n C_{n,\cdot}$$

In other words, aC is a linear combination of the rows of C, with the scalars that multiply the rows coming from a.

3.4 Invertibility and Isomorphic Vector Spaces