THE DIMENSION OF A VECTOR SPACE

KEITH CONRAD

1. Introduction

This handout is a supplementary discussion leading up to the definition of dimension of a vector space and some of its properties. We start by defining the span of a finite set of vectors and linear independence of a finite set of vectors, which are combined to define the all-important concept of a basis.

Definition 1.1. Let V be a vector space over a field F. For any finite subset $\{v_1, \dots, v_n\}$ of V, its *span* is the set of all of its linear combinations:

$$\operatorname{Span}(v_1, \dots, v_n) = \{c_1 v_1 + \cdots + c_n v_n : c_i \in F\}.$$

Example 1.2. In F^3, $\operatorname{Span}((1,0,0),(0,1,0))$ is the xy-plane in F^3.

Example 1.3. If v is a single vector in V then

$$\operatorname{Span}(v) = \{cv : c \in F\} = Fv$$

is the set of scalar multiples of v, which for nonzero v should be thought of geometrically as a line (through the origin, since it includes $0 \cdot v = 0$).

Since sums of linear combinations are linear combinations and the scalar multiple of a linear combination is a linear combination, $\operatorname{Span}(v_1, \dots, v_n)$ is a subspace of V. It may not be all of V, of course.

Definition 1.4. If $\{v_1, \dots, v_n\}$ satisfies $\operatorname{Span}(\{v_1, \dots, v_n\}) = V$, that is, if every vector in V is a linear combination from $\{v_1, \dots, v_n\}$, then we say this set *spans* V or it is a *spanning set* for V.

Example 1.5. In F^2, the set $\{(1,0),(0,1),(1,1)\}$ is a spanning set of F^2. It has some redundancy in it, since removing any one of the vectors leaves behind a spanning set, so remember that spanning sets may be larger than necessary.

Definition 1.6. A finite subset $\{w_1, \dots, w_m\}$ of V is called *linearly independent* when the vanishing of a linear combination only happens in the obvious way:

$$c_1 w_1 + \cdots + c_m w_m = 0 \Longrightarrow \text{ all } c_i = 0.$$

The importance of this concept is that a linear combination of linearly independent vectors has only one possible set of coefficients:

$$c_1 w_1 + \cdots + c_m w_m = c_1' w_1 + \cdots + c_m' w_m \Longrightarrow \text{ all } c_i = c_i'. \tag{1.1}$$

Indeed, subtracting gives $\sum (c_i - c_i') w_i = 0$, so $c_i - c_i' = 0$ for all i by linear independence. Thus $c_i = c_i'$ for all i.

If a subset $\{w_1, \dots, w_m\}$ of V is not linearly independent, it is called *linearly dependent*. What does this condition really mean? Well, to be not linearly independent means there is some set of coefficients $c_1, \dots, c_m$ in F *not all zero* such that

$$c_1 w_1 + \cdots + c_m w_m = 0. \tag{1.2}$$

We don't know which c_i is not zero. If $c_1 \neq 0$ then we can collect all the other terms except $c_1 w_1$ on the other side and multiply by $1/c_1$ to obtain

$$w_1 = -\frac{c_2}{c_1} w_2 - \cdots - \frac{c_m}{c_1} w_m$$

Thus w_1 is a linear combination of $w_2, \dots, w_m$. Conversely, if $w_1 = a_2 w_2 + \cdots + a_m w_m$ is linear combination of $w_2, \dots, w_m$ then $w_1 - a_2 w_2 - \cdots - a_m w_m$ is a linear combination of all the w's that vanishes and the coefficient of w_1 is 1, which is not zero. Similarly, if $c_i \neq 0$ in (1.2) then we can express w_i as a linear combination of the other w_j's, and conversely if w_i is a linear combination of the other w_j's then we obtain a linear combination of all the w's that vanishes and the coefficient of w_i is 1, which is not zero. We have proved the following important result.

Theorem 1.7. *A finite subset $\{w_1, \dots, w_m\}$ in a vector space is linearly dependent precisely when some w_i is a linear combination of the rest.*

Spanning sets for V and linearly independent subsets of V are in some sense opposite concepts:

- Any subset of a linearly independent subset is still linearly independent, but this need not be true of spanning sets.
- Any superset of a spanning set for V is still a spanning set for V, but this need not be true of linearly independent subsets.

Definition 1.8. A finite subset of V that is linearly independent and a spanning set is called a *basis* of V.

Example 1.9. In F^2, the set $\{(1,0),(0,1),(1,1)\}$ is not a basis since it is linearly dependent: the third vector is the sum of the first two. But if we remove any single vector from this set then we get a basis of F^2: $\{(1,0),(0,1)\}$, $\{(1,0),(1,1)\}$, and $\{(0,1),(1,1)\}$ are each a basis of F^2. To show each set is a basis we will show it is a spanning set and it is linearly independent:

(1) For $v = (x, y)$ in V we have $v = x(1,0) + y(0,1)$, so $\{(1,0),(0,1)\}$ spans F^2. To show this set is linearly independent, if $x(1,0) + y(0,1) = (0,0)$ then $(x,y) = (0,0)$, so $x = 0$ and $y = 0$.
(2) For $v = (x, y)$ in V we have $v = (x - y)(1,0) + y(1,1)$, so $\{(1,0),(1,1)\}$ spans F^2. To show this set is linearly independent, if $x(1,0) + y(1,1) = (0,0)$ then $(x + y, y) = (0,0)$, so $x + y = 0$ and $y = 0$. Therefore $x = 0$ and $y = 0$.
(3) For $v = (x, y)$ in V we have $v = (y - x)(0,1) + x(1,1)$, so $\{(0,1),(1,1)\}$ spans F^2. It is left to you to check this set is linearly independent.

Our main goal will be to show if V has a basis then *all bases of V have the same size*, as we saw in Example 1.9 where we found three bases of F^2 all with size 2. That common size of every basis will be called the dimension[1] of V and we will look at how the dimension behaves on subspaces.

[1]Spanning sets, linearly independent sets, and bases can all be extended to infinite sets of vectors, leading to infinite-dimensional vector spaces, but we focus here on the finite-dimensional case.

2. Comparing bases

The following theorem is a first result that links spanning sets in V with linearly independent subsets.

Theorem 2.1. *Suppose $V \neq \{0\}$ and it admits a finite spanning set $\{v_1, \ldots, v_n\}$. Some subset of this spanning set is a linearly independent spanning set.*

The theorem says that once there is a finite spanning set, which could have lots of linear dependence relations, there is a basis for the space. Moreover, the theorem tells us a basis can be found within any spanning set at all.

Proof. While $\{v_1, \ldots, v_n\}$ may not be linearly independent, it contains linearly independent subsets, such as any one single nonzero v_i. Of course, such small linearly independent subsets can hardly be expected to span V. But consider linearly independent subsets of $\{v_1, \ldots, v_n\}$ that are *as large as possible*. Reindexing, without loss of generality, we can write such a subset as $\{v_1, \ldots, v_k\}$.

For $i = k + 1, \ldots, n$, the set $\{v_1, \ldots, v_k, v_i\}$ is not linearly independent (otherwise $\{v_1, \ldots, v_k\}$ is not a maximal linearly independent subset). Thus there is some linear relation

$$c_1v_1 + \cdots + c_kv_k + c_iv_i = 0,$$

where the c's are in F are not all of them are 0. The coefficient c_i cannot be zero, since otherwise we would be left with a linear dependence relation on $v_1, \ldots, v_k$, which does not happen due to their linear independence.

Since $c_i \neq 0$, we see that v_i is in the span of $v_1, \ldots, v_k$. This holds for $i = k+1, \ldots, n$, so any linear combination of $v_1, \ldots, v_n$ is also a linear combination of just $v_1, \ldots, v_k$. As every element of V is a linear combination of $v_1, \ldots, v_n$, we conclude that $v_1, \ldots, v_k$ spans V. By its construction, this is a linearly independent subset of V as well. □

Notice the *non-constructive* character of the proof. If we somehow can check that a (finite) subset of V spans the whole space, Theorem 2.1 says a subset of this is a linearly independent spanning set, but the proof is not constructively telling us which subset of $\{v_1, \ldots, v_n\}$ this might be.

Theorem 2.1 is a "top-down" theorem. It says any (finite) spanning set has a linearly independent spanning set inside of it. It is natural to ask if we can go "bottom-up," and show any linearly independent subset can be enlarged to a linearly independent spanning set. Something along these lines will be proved in Theorem 2.10.

Lemma 2.2. *Suppose $\{v_1, \ldots, v_n\}$ spans V, where $n \geq 2$. Pick any $v \in V$. If some v_i is a linear combination of the other v_j's and v, then V is spanned by the other v_j's and v.*

For example, if V is spanned by v_1, v_2, and v_3, and v_1 is a linear combination of v, v_2, and v_3, where v is another vector in V, then V is spanned by v, v_2, and v_3.

Lemma 2.2 should be geometrically reasonable. See if you can prove it before reading the proof below.

Proof. Reindexing if necessary, we can suppose it is v_1 that is a linear combination of $v, v_2, \ldots, v_n$. We will show every vector in V is a linear combination of $v, v_2, \ldots, v_n$, so these vectors span V.

Pick any $w \in V$. By hypothesis,

$$w = c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

for some $c_i \in F$. Since v_1 is a linear combination of $v, v_2, \ldots, v_n$, we feed this linear combination into the above equation to see w is a linear combination of $v, v_2, \ldots, v_n$. As w was arbitrary in V, we have shown V is spanned by $v, v_2, \ldots, v_m$. ☐

The following central technical result relates spanning sets for V and linearly independent subsets of V. It is called the "exchange theorem" from the process in its proof, which relies on repeated applications of Lemma 2.2.

Theorem 2.3 (Exchange Theorem)**.** *Suppose V is spanned by n vectors, where $n \geq 1$. Every linearly independent subset of V has at most n vectors.*

If you think about linear independence as "degrees of freedom," the exchange theorem makes sense. What makes the theorem somewhat subtle to prove is that the theorem bounds the size of any linearly independent subset once we know the size of one spanning set. Most linearly independent subsets of V are not directly related to the original choice of spanning set, so linking the two sets of vectors is tricky. The proof will show how to link linearly independent sets and spanning sets by an exchange process, one vector at a time.

Proof. First, let's check the result when $n = 1$. In this case, $V = Fv$ for some v (that is, V is spanned by one vector). Two different scalar multiples of v are linearly dependent, so a linearly independent subset of V can have size at most 1.

Now we take $n \geq 2$. We give a proof by *contradiction*. If the theorem is false, then V contains a set of $n + 1$ linearly independent vectors, say $w_1, \ldots, w_{n+1}$.

Step 1: We are told that V can be spanned by n vectors. Let's call such a spanning set $v_1, \ldots, v_n$. We also have the $n + 1$ linearly independent vectors $w_1, \ldots, w_{n+1}$ in V. Write the first vector from our linearly independent set in terms of our spanning set:

$$w_1 = c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

for some $c_i \in F$. Since $w_1 \neq 0$ (a linearly independent set never contains the vector 0), some coefficient c_j is nonzero. Without loss of generality, we can reindex the v's so that c_1 is nonzero. Then the above equation can be solved for v_1 as a linear combination of $w_1, v_2, \ldots, v_n$. By Lemma 2.2,

$$V = \mathrm{Span}(w_1, v_2, \ldots, v_n). \tag{2.1}$$

Notice that we have taken one element from the initial spanning set out and inserted an element from the linearly independent set in its place, retaining the spanning property.

Step 2: Let's repeat the procedure, this time using our new spanning set (2.1). Write w_2 in terms of this new spanning set:

$$w_2 = c_1'w_1 + c_2'v_2 + \cdots + c_n'v_n \tag{2.2}$$

for some c_i' in F. We want to use this equation to show w_2 can be inserted into (2.1) and one of the original vectors can be taken out, without destroying the spanning property. Some care is needed, because we want to keep w_1 in the spanning set rather than accidentally swap it out. (This is an issue that we did not meet in the first step, where no new vectors had yet been placed in the spanning set.)

Certainly one of $c_1', c_2', \ldots, c_n'$ is nonzero, since w_2 is nonzero. But in fact we can say something a bit sharper: regardless of the value of c_1', one of $c_2', \ldots, c_n'$ is nonzero. Indeed, if $c_2', \ldots, c_m'$ are all zero, then $w_2 = c_1'w_1$ is a scalar multiple of w_1, and that violates linear independence (as $\{w_1, \ldots, w_m\}$ is linearly independent, so is the subset $\{w_1, w_2\}$).

Without loss of generality, we can reindex $v_2, \ldots, v_n$ so it is c_2' that is nonzero. Then we can use (2.2) to express v_2 as a linear combination of $w_1, w_2, v_3, \ldots, v_n$. By another application of Lemma 2.2, using our new spanning set in (2.1) and the auxiliary vector w_2, it follows that

$$V = \operatorname{Span}(w_1, w_2, v_3, \ldots, v_n).$$

Step 3: Now that we see how things work, we argue inductively.

Suppose for some k between 1 and $n-1$ that we have shown

$$V = \operatorname{Span}(w_1, \ldots, w_k, v_{k+1}, \ldots, v_n).$$

(This has already been checked for $k = 1$ in Step 1, and $k = 2$ in Step 2, although Step 2 is not logically necessary for what we do; it was just included to see concretely the inductive step we now carry out for any k.)

Using this spanning set for V, write

$$w_{k+1} = a_1 w_1 + \cdots + a_k w_k + a_{k+1} v_{k+1} + \cdots + a_n v_n \tag{2.3}$$

with $a_i \in F$. One of $a_{k+1}, \ldots, a_n$ is nonzero, since otherwise this equation expresses w_{k+1} as a linear combination of $w_1, \ldots, w_k$, and that violates linear independence of the w's.

Reindexing $v_{k+1}, \ldots, v_n$ if necessary, we can suppose it is a_{k+1} that is nonzero. Then (2.3) can be solved for v_{k+1} as a linear combination of $w_1, \ldots, w_k, w_{k+1}, v_{k+2}, \ldots, v_n$. By Lemma 2.2, using the spanning set $\{w_1, \ldots, w_k, v_{k+1}, \ldots, v_n\}$ and the auxiliary vector w_{k+1}, we can swap w_{k+1} into the spanning set in exchange for v_{k+1} without losing the spanning property:

$$V = \operatorname{Span}(w_1, \ldots, w_k, w_{k+1}, v_{k+2}, \ldots, v_n).$$

We have added an new vector to the spanning set and taken one of the original vectors out. Now by induction (or, more loosely, "repeating this step $n-k-1$ more times"), we arrive at the conclusion that

$$V = \operatorname{Span}(w_1, \ldots, w_n). \tag{2.4}$$

However, we were starting with $n+1$ linearly independent vectors $w_1, \ldots, w_{n+1}$, so w_{n+1} is not in the span of $w_1, \ldots, w_n$. That contradicts the meaning of (2.4), which says every vector in V is a linear combination of $w_1, \ldots, w_n$. We have reached a contradiction, so no linearly independent subset of V contains more than n vectors, where n is the size of a spanning set for V. $\square$

Example 2.4. Consider $\mathrm{M}_3(\mathbf{R})$, the 3×3 real matrices. It is a vector space over $\mathbf{R}$ under matrix addition and the usual multiplication of a matrix by a real number. This vector space has a 9-element spanning set, namely the 9 matrices with a 1 in one component and 0 elsewhere. Therefore any linearly independent subset of $\mathrm{M}_3(\mathbf{R})$ has at most 9 elements in it.

Corollary 2.5. *Suppose $V \neq \{0\}$ and V admits a finite basis. Any two bases for V have the same size.*

Proof. Let $\{v_1, \ldots, v_n\}$ and $\{v_1', \ldots, v_m'\}$ be bases for V. Treating the first set as a spanning set for V and the second set as a linearly independent subset of V, the exchange theorem tells us that $m \leq n$. Reversing these roles (which we can do since bases are both linearly independent and span the whole space), we get $n \leq m$. Thus $m = n$. $\square$

Definition 2.6. If V is a vector space over F and V has a finite basis then the (common) size of any basis of V is called the *dimension* of V (over F).

Example 2.7. There are obvious bases of the vector spaces $\mathbf{R}^n$ and $\mathrm{M}_n(\mathbf{R})$: in $\mathbf{R}^n$ one basis is the vectors with 1 in one coordinate and 0 elsewhere, and in $\mathrm{M}_n(\mathbf{R})$ one basis is the matrices with 1 in one entry and 0 elsewhere. Counting the number of terms in a basis, $\mathbf{R}^n$ has dimension n and $\mathrm{M}_n(\mathbf{R})$ has dimension n^2.

Example 2.8. Treating $\mathbf{C}$ as a real vector space, one basis is $\{1, i\}$, so $\mathbf{C}$ has dimension 2 as a vector space over $\mathbf{R}$.

Example 2.9. The vector space $\{0\}$ has no basis, or you might want to say its basis is the empty set. In any event, it is natural to declare the zero vector space to have dimension 0.

Theorem 2.10. *Let V be a vector space with dimension $n \geq 1$. Any spanning set has at least n elements, and contains a basis inside of it. Any linearly independent subset has at most n elements, and can be extended to a basis of V. Finally, an n-element subset of V is a spanning set if and only if it is a linearly independent set.*

Proof. Since V has a basis of n vectors, let's pick such a basis, say $v_1, \ldots, v_n$. We will compare this basis to the spanning sets and the linearly independent sets in V to draw our conclusions, taking advantage of the dual nature of a basis as *both* a linearly independent subset of V and as a spanning set for V.

If $\{u_1, \ldots, u_k\}$ is a spanning set for V, then a comparison with $\{v_1, \ldots, v_n\}$ (interpreted as a linearly independent subset of V) shows $n \leq k$ by the exchange theorem. Equivalently, $k \geq n$. Moreover, Theorem 2.1 says that $\{u_1, \ldots, u_k\}$ contains a basis for V. This settles the first part of the theorem.

For the next part, suppose $\{w_1, \ldots, w_m\}$ is a linearly independent subset of V. A comparison with $\{v_1, \ldots, v_n\}$ (interpreted as a spanning set for V) shows $m \leq n$ by the exchange theorem. To see that the w's can be extended to a basis of V, apply the exchange process from the proof of the exchange theorem, but only m times since we have only m linearly independent w's. We find at the end that

$$V = \mathrm{Span}(w_1, \ldots, w_m, v_{m+1}, \ldots, v_n),$$

which shows the w's can be extended to a spanning set for V. This spanning set contains a basis for V, by Theorem 2.1. Since all bases of V have n elements, this n-element spanning set must be a basis itself.

Taking $m = n$ in the previous paragraph shows any n-element linearly independent subset is a basis (and thus spans V). Conversely, any n-element spanning set is linearly independent, since any linear dependence relation would let us cut down to a spanning set of fewer than n elements, but that violates the first result in this proof: a spanning set for an n-dimensional vector space has at least n elements. □

3. Dimension of subspaces

Theorem 3.1. *If V is an n-dimensional vector space, any subspace is finite-dimensional, with dimension at most n.*

Proof. This theorem is trivial if $V = \{0\}$, so we may assume $V \neq \{0\}$, *i.e.*, $n \geq 1$.

Let W be a subspace of V. Any linearly independent subset of W is also a linearly independent subset of V, and thus has size at most n by Theorem 2.10. Choose a linearly independent subset $\{w_1, \ldots, w_m\}$ of W where m is *maximal*. Then $m \leq n$. We will show $\mathrm{Span}(w_1, \ldots, w_m) = W$.

For any $w \in W$, the set $\{w, w_1, \ldots, w_m\}$ has more than m elements, so it can't be linearly independent. Therefore there is some vanishing linear combination

$$aw + a_1 w_1 + \cdots + a_m w_m = 0$$

where $a, a_1, \ldots, a_m$ are in F and are not all 0. If $a = 0$ then the a_i's all vanish since $w_1, \ldots, w_m$ are linearly independent. Therefore $a \neq 0$, so we can solve for w:

$$w = -\frac{a_1}{a} w_1 - \cdots - \frac{a_m}{a} w_m.$$

Thus w is a linear combination of $w_1, \ldots, w_m$. Since w was arbitrary in W, this shows the w_i's span W. So $\{w_1, \ldots, w_m\}$ is a spanning set for W that is linearly independent by construction. This proves W is finite-dimensional with dimension $m \leq n$. □

Theorem 3.2. *If V has dimension n and W is a subspace with dimension n, then $W = V$.*

Proof. When W has dimension n, any basis for W is a linearly independent subset of V with n elements, so it spans V by Theorem 2.10. The span is also W (by definition of a basis for W), so $W = V$. □

It is important that throughout our calculations (expressing one vector as a linear combination of others when we have a nontrivial linear combination of vectors equal to 0) we can scale a nonzero coefficient of a vector to make the coefficient equal to 1. For example, suppose we tried to do linear algebra over the integers $\mathbf{Z}$ instead of over a field. Then we can't scale a coefficient in $\mathbf{Z}$ to be 1 without possibly needing rational coefficients for other vectors in a linear combination. That suggests results like the ones we have established for vector spaces over fields might not hold for "vector spaces over $\mathbf{Z}$." And it's true: linear algebra over $\mathbf{Z}$ is more subtle than over fields. For example, Theorem 3.2 is *false* if we work with "vector spaces" over $\mathbf{Z}$. Consider the integers $\mathbf{Z}$ and the even integers $2\mathbf{Z}$. By any reasonable definitions, both $\mathbf{Z}$ and $2\mathbf{Z}$ should be considered "one-dimensional" over the integers, where $\mathbf{Z}$ has basis $\{1\}$ and $2\mathbf{Z}$ has basis $\{2\}$ (since every integer is a unique integral multiple of 2). But $2\mathbf{Z} \subset \mathbf{Z}$, so a "one-dimensional vector space over $\mathbf{Z}$" can lie inside another without them being equal. This is a pretty clear failure of Theorem 3.2 when we use scalars from $\mathbf{Z}$ instead of from a field.