

In this section we begin the material from Chapters 3 and 4 of **Artin**.

We have defined the notion of a binary operation on a set X ; this is a function from $X \times X$ to X . There is another kind of operation that occurs frequently, where we combine elements from sets X and Y and obtain an element of Y . The classic example, which is the model for our definition below, is the multiplication of a vector by a scalar. In this case, we take a number $r \in \mathbb{R}$ and a vector $v \in \mathbb{R}^n$, and we obtain a new vector $rv = r \cdot v \in \mathbb{R}^n$. An operation of this sort is a function $X \times Y \rightarrow Y$. **Artin** calls this an *external law of composition*. We will call it an *operation of X on Y* or an *action of X on Y* , and we will adopt the convention of writing the result of applying the function to (x, y) as either $x.y$ or simply xy .

Let R be a ring and let $(M, +)$ be an *Abelian group*. We say M is a (left) *module* over R (or an *R -module*) if there is an operation of R on M satisfying the following conditions for all $a, b \in R$, $m, n \in M$. The first condition is an associative law, while the final two conditions are distributive laws.

1. $a.(b.m) = (ab).m$;
2. $1_R.m = m$;
3. $(a + b).m = a.m + b.m$;
4. $a.(m + n) = a.m + a.n$.

When R is a field, we use the name *vector space* instead of module. We refer to the operation of R on M as *scalar multiplication*.

We will give some general examples of modules for rings which are not necessarily fields, but after that we will only consider vector spaces. When R is not a field, the theory of modules is quite a bit more complicated.

- Example 7.1.**
1. Let $R = \mathbb{Z}$ and let M be an Abelian group. Then there is a unique way to make M into a \mathbb{Z} -module. To see this, note that the condition that $1m = m$ for all $m \in M$ together with distributivity implies nm equals the sum of m with itself n times for any positive integer n . We will shortly see that $(-r)m = -(rm)$ always holds, and so it follows that there is only one way to define nm that might make M into a \mathbb{Z} -module. This definition does make M into a \mathbb{Z} -module: in fact, the axioms (1)–(4) above are just the laws of exponents for M written in additive form.
 2. Let R be any ring. Then R^n , whether regarded as the set of $n \times 1$ column vectors, the set of $1 \times n$ row vectors, or the set of n -tuples from R , is an R -module with componentwise operations. Thus $r\mathbf{x}$ is just the product of the scalar r and the matrix \mathbf{x} that we used in Section 1 on matrices. The properties required for M to be an R -module all follow from the corresponding properties in R .
 3. The preceding example is a special case of the following. Given any positive integers m, n , the set $M_{m \times n}(R)$ of all $m \times n$ matrices over R is an R -module with the usual definition of scalar multiplication: $r.A = rA$.
 4. Let $R \subseteq S$ be rings of numbers. Then S is an R -module when we use the multiplication in S , that is, $r.s = rs$.

Let's record some expected, and easily proven, properties of modules.

Lemma 7.2. *Let M be an R -module. Then for all $r \in R, m \in M$, we have the following.*

1. $0_R m = 0_M$.
2. $r 0_M = 0_M$.
3. $(-1)m = -m$.

Proof. Exercise. *Hints:* to show (1), write $0m = (0 + 0)m = 0m + 0m$ and conclude $0m = 0$. The proof of (2) is similar. Use (1) to prove (3). ■

Because M is an Abelian group, we know cancellation holds in M , with respect to $+$. Cancellation with respect to \cdot doesn't always hold, but we do have the following.

Lemma 7.3. *Let F be a field and let V be a vector space over F .*

1. *If $a \in F, v \in V$ and $av = 0$, then $a = 0$ or $v = 0$.*
2. *If $a, b \in F, v \in V$ and $v \neq 0$, then $av = bv$ implies $a = b$.*
3. *If $a \in F, v, w \in V$ and $a \neq 0$, then $av = aw$ implies $v = w$.*

Proof. (1) If $a \neq 0$ and $av = 0$, then $v = 1v = a^{-1}av = a^{-1}0 = 0$.

(2) and (3) are corollaries of (1) — Exercise. ■

We will be interested in a much more general kind of cancellation. We say $v_1, \dots, v_n \in V$ are *linearly independent* if whenever $\sum_{i=1}^n a_i v_i = 0$ for $a_1, \dots, a_n \in F$, we have $a_i = 0$ for each $i = 1, \dots, n$. We say $v_1, \dots, v_n \in V$ are *linearly dependent* if they are not linearly independent, that is, if there exist $a_1, \dots, a_n \in F$, not all of which are 0, such that $a_1 v_1 + \dots + a_n v_n = 0$.

Lemma 7.3 says that any single nonzero $v \in V$ is linearly independent. If we consider the ordinary plane \mathbb{R}^2 , we see that two vectors v, w are linearly independent iff they do not lie on the same line through $\mathbf{0} = (0, 0)$. However, any three vectors in \mathbb{R}^2 are linearly dependent. The reason for this last claim is that if two vectors v, w in \mathbb{R}^2 are independent, then *any* vector $u \in \mathbb{R}^2$ can be written as $u = av + bw$ for some $a, b \in \mathbb{R}$, whence $av + bw + (-1)u = 0$. Thus the condition of linear independence is related to another condition, that of expressing other vectors as a linear combination of the given vectors. Our goal in this section is to explore this link. This will lead us to the notion of a basis and the dimension of a vector space.

Before we give the necessary formal definitions, let us consider the ideas discussed in the last paragraph. When we deal with ordinary n -space \mathbb{R}^n , it is usually crucial for us to know we have a set of co-ordinate axes. For example, in three dimensions, we express points, functions, and so on, in terms of the co-ordinates (x, y, z) , which in turn are defined in terms of the x, y, z -axes. Every point has a unique set of co-ordinates relative to these axes. If $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are the normal unit vectors, then the point (x, y, z) corresponds to the vector $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$.

A basis of a vector space V can be thought of as the same thing — a set of vectors that define co-ordinate axes, such that every vector can be written as a unique combination of the basis vectors.

If v_1, \dots, v_n are elements of the vector space V over the field F , a *linear combination* of v_1, \dots, v_n is a sum of the form $\sum_{i=1}^n a_i v_i$ for some $a_1, \dots, a_n \in F$. We will be ambiguous here: the term linear combination can refer either to the sum or to the vector that is the result of that sum. Thus we will refer to a_1, \dots, a_n as the *coefficients* in the linear combination, even though different coefficients could yield the same vector. We leave it to the reader to make our ambiguity clear in any given situation.

A *trivial linear combination* is one in which every coefficient is 0. We say v_1, \dots, v_n are *linearly independent* if the only linear combination yielding 0 is the trivial one. (This is the same as the definition given earlier in this section.) We say v_1, \dots, v_n *span* V if every element of V can be written as a linear combination of v_1, \dots, v_n . We say v_1, \dots, v_n form a *basis* of V if every element of V can be written *uniquely* as a linear combination of v_1, \dots, v_n . (This means every $v = \sum_{i=1}^n a_i v_i$ for some $a_1, \dots, a_n \in F$, and the n -tuple (a_1, \dots, a_n) is unique.)

Thus a basis is a set of elements of V that can serve as a set of co-ordinate axes.

- Example 7.4.** 1. The quintessential example of a basis is the set of vectors $\mathbf{e}_1 = (1, 0, \dots, 0), \dots, \mathbf{e}_n = (0, 0, \dots, 1)$ in F^n .
2. More generally, the matrix units E_{ij} , $1 \leq i \leq m$, $1 \leq j \leq n$ form a basis for the vector space $M_{m \times n}(F)$ of $m \times n$ matrices over F .
3. Let $F = \mathbb{R}$ and $V = \mathbb{C}$. Then V is a vector space over F , and every element of V can be written uniquely in the form $a + bi = a1 + bi$ for some $a, b \in F$. This says precisely that $1, i$ is a basis for V over F . It is not the only basis: there are uncountably many bases. Another example of a basis is $2 + 3i, 5 - 7i$.

Lemma 7.5. *Let $v_1, \dots, v_n \in V$. Then v_1, \dots, v_n form a basis for V if and only if they are linearly independent and span V .*

Proof. Suppose v_1, \dots, v_n form a basis. Clearly they span V . We always have $0 = \sum_{i=1}^n 0v_i$, so if $\sum_{i=1}^n a_i v_i = 0$, then $a_i = 0$ for all i by uniqueness. Thus v_1, \dots, v_n are linearly independent.

Conversely, if v_1, \dots, v_n span V , any $v \in V$ can be written $v = \sum_{i=1}^n a_i v_i$ for some scalars $a_i \in F$. If also $v = \sum_{i=1}^n b_i v_i$, then $0 = \sum_{i=1}^n (a_i - b_i) v_i$. Hence if v_1, \dots, v_n are linearly independent, we conclude that $a_i = b_i$ for each i . This proves v_1, \dots, v_n form a basis. ■

Remark 7.6. Above we have applied the terms *linearly independent*, *span*, *basis* to a group of objects v_1, \dots, v_n and so we have used plural language (“are”, “span”, “form”). We frequently think of v_1, \dots, v_n as the set $X = \{v_1, \dots, v_n\}$, in which case we use the singular: X is linearly independent, X spans V , X is a basis. In what follows we will mix these modes of usage.

There is still another way in which we regard a basis. If our goal is to put a co-ordinate system on V , then we will presumably associate the n -tuple (a_1, \dots, a_n) to the vector $\sum_{i=1}^n a_i v_i$. This implies an ordering. Thus when we wish to use explicit co-ordinates, we need to speak of an *ordered basis*, which is an n -tuple (v_1, \dots, v_n) . As usual, we generally leave it to the reader to figure out what we are talking about at any given moment!

Once we decide to apply the terms “linearly independent”, “span”, and “basis” to sets, it becomes natural to allow infinite sets, and hence to modify the definitions slightly. Thus if X is a set, a *linear combination* of elements of X is a sum (or its result) $\sum_{i=1}^n a_i v_i$ for some finite collection v_1, \dots, v_n of distinct elements of X and some $a_1, \dots, a_n \in F$. Linear independence, spanning, and basis are defined solely in terms of such finite linear combinations.

Our goal for the rest of this section is straightforward. We wish to show that every vector space has a basis, and that any two bases have the same number of elements. This common number will be called the *dimension* of the vector space.

In pursuit of this goal, it is convenient to introduce other notions, which fortuitously are natural and useful in their own right.

If $W \subseteq V$, we say W is a *subspace* of V if it is a subgroup under $+$ (i.e., is closed under $+$ and $-$ and contains 0) and it is closed under scalar multiplication, i.e., $a \in F, w \in W$ implies $aw \in W$.

Lemma 7.7. *Let $W \subseteq V$. Then W is a subspace of V if and only if (a) $W \neq \emptyset$; (b) If $w, w' \in W$, then $w + w' \in W$; and (c) If $a \in F, w \in W$, then $aw \in W$.*

Proof. What must be shown is that if $w \in W$, then $-w \in W$. We leave this as an exercise. ■

If V is the plane \mathbb{R}^2 , then the subspaces of V are precisely $\{0\}$, V , and all of the lines through the origin. It is obvious that these are subspaces. It is also obvious that if a subspace contains a vector v , then it contains the line through v and the origin. Thus all that remains is to convince yourself that if a subspace contains two vectors that do not line on the same line through the origin, it contains the whole plane.

In general, a subspace of the n -dimensional vector space \mathbb{R}^n is a flat or linear space, containing the origin, of smaller dimension. In fact, let us define $X \subseteq \mathbb{R}^n$ to be a *linear subset* if X contains the entire line through any two points of X . The subspaces of \mathbb{R}^n are precisely the linear subsets containing the origin. We will say more about this later.

Let $X \subseteq V$. The *span of X* is defined to be the set of all linear combinations of finitely many elements from X . Thus

$$\text{span } X = \left\{ \sum_{i=1}^n a_i v_i \mid a_1, \dots, a_n \in F, v_1, \dots, v_n \in X, n \geq 0 \right\}.$$

In this last equation, we allowed the possibility that $n = 0$. We will make the convention that a sum of 0 terms is 0_V . This convention is only necessary to deal with the case $X = \emptyset$, and so our convention amounts to defining $\text{span } \emptyset = \{0\}$.

Lemma 7.8. *Let V be a vector space over F and let $X \subseteq V$. Then $\text{span } X$ is the smallest subspace of V containing X .*

Proof. We need to show two things. First, that $\text{span } X$ is a subspace of V (plainly $X \subseteq \text{span } X$) and second, that if $W \supseteq X$ is a subspace of V , then $\text{span } X \subseteq W$.

We leave both as exercises. ■

If W is a subspace of V and $X \subseteq V$ and $\text{span } X = W$, then we say X *spans* W .

In terms of our vector analogy, the subspace spanned by a set of vectors is the smallest linear space through the origin containing all the vectors.

The following lemma links linear independence and spanning.

Lemma 7.9. *Let V be a vector space over F , let $X \subseteq V$ be linearly independent, and let $y \in V$. Then the following statements are true.*

1. $X \cup \{y\}$ is linearly independent iff $y \notin \text{span } X$.
2. If $y \notin X$, then $X \cup \{y\}$ is linearly dependent iff $y \in \text{span } X$.

Proof. (1) and (2) are plainly equivalent; we will prove (2).

First, suppose $y \in \text{span } X$. Then we have $y = \sum_{i=1}^n a_i x_i$ for some $a_1, \dots, a_n \in F$, $x_1, \dots, x_n \in X$. Thus $a_1 x_1 + \dots + a_n x_n + (-1)y = 0$ is a nontrivial linear combination from $X \cup \{y\}$ that yields 0, so $X \cup \{y\}$ is linearly dependent.

Next suppose $X \cup \{y\}$ is linearly dependent, i.e., suppose that there is a non-trivial linear combination of elements from $X \cup \{y\}$ that equals 0. Since X is linearly independent, this combination must involve y in a non-trivial way. That is, there must exist $a_1, \dots, a_n, a \in F$ with $a \neq 0$ such that $a_1 x_1 + \dots + a_n x_n + ay = 0$. We can solve this equation for y and we find $y = \sum_{i=1}^n -(a_i/a)x_i \in \text{span } X$. ■

We are now ready to pursue our goal of showing bases exist and the cardinality of a basis of V is uniquely determined by V . We begin with one of the key results.

Theorem 7.10. *Let V be a vector space over a field F and let $X, Y \subseteq V$. If X is linearly independent and Y spans V , then there is a subset Y' of Y such that $X \cup Y'$ is a basis of X .*

Proof. Assume Y is finite: we will discuss the case where Y is infinite in an appendix to this section.

Choose a subset Y' of Y with the following two properties: (1) $X \cup Y'$ is linearly independent, and (2) Y' is the largest subset of Y satisfying condition (1). Note that \emptyset satisfies condition (1), so there are subsets of Y satisfying (1). Since Y is finite, there is a largest such subset. (It is here where we have to use more advanced techniques if Y is infinite.)

We claim $X' = X \cup Y'$ is a basis of V . It is linearly independent by definition, so we must show X' spans V . We will first show that $Y \subseteq \text{span } X'$.

Let $y \in Y$ and suppose $y \notin \text{span } X'$. Then $y \notin X'$ and by Lemma 7.9, $X' \cup \{y\}$ is linearly independent. But if we set $Y'' = Y' \cup \{y\}$, we have $Y' \subsetneq Y''$ and $X \cup Y''$ linearly independent. This contradicts our choice of Y' . It follows that $Y \subseteq \text{span } X'$.

Now $\text{span } X'$ is a subspace containing Y , so by Lemma 7.8, $\text{span } X'$ contains the subspace $\text{span } Y = V$. Thus X' spans V , and we have proven that X' is a basis of V . ■

Corollary 7.11. *Let V be a vector space. Then any linearly independent subset of V can be expanded to a basis, and any subset that spans V can be contracted to a basis.*

Proof. If X is linearly independent, we can take $Y = V$ and apply Theorem 7.10.

If Y spans V , we can take $X = \emptyset$ and apply Theorem 7.10. ■

Remark 7.12. There is one problem with the proof of the preceding corollary, and hence with the proof of the next corollary. In the proof, we used the set V as a spanning set, but V is likely to be infinite. Thus we are forced to confront the “infinite case” we tried to avoid. One way to avoid this problem is to prove all results only for *finitely spanned* vector spaces, that is, vector spaces which have a finite spanning set. (Such vector spaces are precisely the *finite dimensional vector spaces*.) The reader may either make this restriction or read the appendix to this section, where the “infinite” problem is discussed.

Corollary 7.13. *Every vector space has a basis.*

Proof. Apply Corollary 7.11 either to the linearly independent set \emptyset or the spanning set V . ■

Problem 7.A. Let X be a subset of a vector space V . Show that the following statements are equivalent.

1. X is a basis of V .
2. X is a maximal linearly independent subset of V . (That is, X is linearly independent and if $X \subsetneq Y \subseteq V$, then Y is linearly dependent.)
3. X is a minimal spanning set in V . (That is, X spans V and if $Y \subsetneq X$, then Y does not span V .)

Our next goal is to compare the sizes of bases. This requires a lemma, which is related to Theorem 7.10, but is not quite the same.

Lemma 7.14 (Exchange Lemma). *Let V be a vector space, let $X \subseteq V$ be a linearly independent set, and let $Y \subseteq V$ span V .*

1. *Either $X \subseteq Y$ or there are $x \in X \setminus Y$, $y \in Y \setminus X$ such that if we create sets X' and Y' by exchanging x and y , that is $X' = (X \setminus \{x\}) \cup \{y\}$ and $Y' = (Y \setminus \{y\}) \cup \{x\}$, then X' is linearly independent and Y' spans V .*
2. *If X, Y are bases, then either $X = Y$ or we can choose x, y as in (1) such that X', Y' are bases.*

Proof. (1) Suppose $X \not\subseteq Y$: then there exists an $x \in X \setminus Y$. Since X is linearly independent, so is $X_1 = X \setminus \{x\}$, and moreover, $x \notin \text{span } X_1$ by Lemma 7.9. As in the proof of Theorem 7.10, this implies there exists a $y \in Y$ such that $y \notin \text{span } X_1$. Again by Lemma 7.9, we see that $X' = X_1 \cup \{y\}$ is linearly independent.

The linear independence of X' is all that we will actually need below. However, we claimed we could choose y so that Y' spans V ; to do this, we must be a little more careful.

We can still take any $x \in X \setminus Y$. Since Y spans V , we can write $x = \sum_{i=1}^n a_i y_i$ for some $y_1, \dots, y_n \in Y$ and some nonzero $a_1, \dots, a_n \in F$. Since X is linearly independent, there must be at least one y_j such that $y_j \notin \text{span } X_1$. Set $y = y_j$ for this j , and $Y' = (Y \setminus \{y_j\}) \cup \{x\}$

Then X' is linearly independent as above. We can write $y = y_j = (1/a_j)x + \sum_{i \neq j} -(a_i/a_j)y_i$, so $y \in \text{span } Y'$. Obviously for any other $z \in Y$, we have $z \in \text{span } Y'$. It follows (as in the proof of Theorem 7.10) that Y' spans V .

(2) Exercise. ■

Theorem 7.15. *Let X, Y be subsets of a vector space V and suppose that X is linearly independent and Y spans V . Then $|X| \leq |Y|$.*

Proof. Again we will assume Y is finite, say $|Y| = n$; we will discuss the infinite case in the appendix to this section.

Suppose the theorem is false and that V, X, Y give us a counterexample. Keeping Y, V fixed and *changing X if necessary*, we can assume that $|X \cap Y|$ is as large as possible, that is, if $X' \subseteq Y$ is such that V, X', Y give us a counterexample, then $|X' \cap Y| \leq |X \cap Y|$. (This is possible because all the numbers involved are no greater than $n = |Y|$.)

If $X \not\subseteq Y$, then by the Exchange Lemma, there are $x \in X \setminus Y$, $y \in Y \setminus X$ such that $X' = (X \setminus \{x\}) \cup \{y\}$ is linearly independent. Moreover, $X' \cap Y = (X \cap Y) \cup \{y\}$ is strictly larger than $X \cap Y$. This contradicts our choice of X .

Thus we must have $X \subseteq Y$ and so we can conclude that $|X| \leq |Y|$. ■

Corollary 7.16. *Any two bases of a vector space have the same number of elements. That is, if V is a vector space over a field F and X, Y are bases of V , then $|X| = |Y|$.*

Proof. By Theorem 7.15, we have $|X| \leq |Y|$ and $|Y| \leq |X|$. Thus $|X| = |Y|$.

This result can also be proved directly using part (2) of the Exchange Lemma. ■

We define the *dimension* of a vector space V to be the size of any (and hence every) basis of V , and we denote it $\dim V$.

Thus for example, $\dim F^n = n$ and $\dim M_{m \times n}(F) = mn$. We have $\dim_{\mathbb{R}} \mathbb{C} = 2$. (If there are different fields in use, we sometimes write $\dim_F V$ to make clear that V is a vector space over F .)

Corollary 7.17. *Let V be a vector space over a field F and let $n = \dim V$ be finite.*

1. *Any linearly independent subset of V with n elements is a basis.*
2. *Any subset of n elements that spans V is a basis.*

Proof. (1) Let X be a linearly independent set of n elements. By Corollary 7.11, there is a basis $X' \supseteq X$. But $|X'| = \dim V = n = |X|$, so we must have $X = X'$.

(2) This proof is similar to the proof of (1).

Note that this proof would fail if $\dim V$ were infinite, and in that case, the corollary is not true. ■

This last result is very useful in deciding whether a given set is a basis. For example, we know F^2 has dimension 2, since it has the standard basis $\mathbf{e}_1, \mathbf{e}_2$. Thus by Corollary 7.17,

two vectors $v, w \in F^2$ form a basis iff they are linearly independent. It is easy to tell when two vectors are linearly independent. We conclude that $v, w \in F^2$ form a basis iff neither v nor w is a multiple of the other.

The following result is another very useful application.

Corollary 7.18. *Let F be a field and let $A \in M_n(F)$. Then the following conditions are equivalent.*

1. A is invertible.
2. The columns of A form a basis for F^n .
3. The columns of A are linearly independent.
4. The columns of A span F^n .
5. The rows of A form a basis for F^n .
6. The rows of A are linearly independent.
7. The rows of A span F^n .

Proof. Since F^n is a vector space of dimension n — we know the standard basis has n elements — the equivalence of (2),(3),(4) and the equivalence of (5),(6),(7) follow immediately from Corollary 7.17.

Recall from Section 1 that (1) holds if and only if the equation $A\mathbf{x} = \mathbf{b}$ can be solved

for any \mathbf{b} . If $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and if A_i is column i of A , then $A\mathbf{x} = \sum_{i=1}^n x_i A_i$. Thus $A\mathbf{x}$ is

a linear combination of the columns of A , and so the statement that $A\mathbf{x} = \mathbf{b}$ can always be solved is equivalent to the statement that the columns of A span F^n . This proves (1) is equivalent to (4).

A similar proof shows (1) is equivalent to (7), and this completes the proof. ■

Here is another nice application of bases. In Artin, this result is used instead of the Exchange Lemma to prove Theorem 7.15. We get it as a corollary.

Corollary 7.19. *A homogeneous system of m linear equations in n unknowns always has a nonzero solution if $n > m$.*

Put in matrix terms, if A is an $m \times n$ matrix with $n > m$, then there is a $\mathbf{x} \in F^n$ with $\mathbf{x} \neq \mathbf{0}$ but $A\mathbf{x} = \mathbf{0}$.

Proof. As in the proof of Corollary 7.18, the product $A\mathbf{x}$ is a linear combination of the columns of A . There are n of these columns, and they are elements of F^m , a vector space of dimension $m < n$. Thus the set of columns must be linearly dependent, that is, some non-trivial linear combination of them must be $\mathbf{0}$. This says $A\mathbf{x} = \mathbf{0}$ for some nonzero \mathbf{x} . ■

APPENDIX: Infinite-dimensional Vector Spaces

At two points in this section we made the assumption that spanning sets or bases were finite. In this appendix we will briefly discuss the general case.

The first place where the finiteness assumption was used was in the proof of Theorem 7.10. We had a linearly independent set $X \subseteq V$ and a spanning set $Y \subseteq V$, and we needed the existence of a largest subset Y' of Y such that $X \cup Y'$ remained linearly independent. In the proof what we needed for “largest” was that if $Y' \subsetneq Y''$, then $X \cup Y''$ is linearly dependent. We usually express this by saying that $Y' \subseteq Y$ is *maximal with respect to the property* that $X \cup Y'$ is linearly dependent. When Y is finite, we know such maximal sets exist because we can take a subset Y' satisfying this property that has as

many elements as possible. When Y is infinite, however, there will be larger and larger subsets in a never-ending chain.

Instead, we have to appeal to a fundamental principle of “infinite” mathematics, Zorn’s Lemma. This lemma asserts the existence of objects without giving any means of constructing them, and so it is viewed with disfavor by some. If one is willing to use it, however, it is extremely powerful. (Indeed, many results cannot be proven without Zorn’s Lemma.) We will state it below but not prove it. The proof involves the Axiom of Choice and some form of transfinite induction — in fact, Zorn’s Lemma is equivalent to the Axiom of Choice.

We first state a special version for sets, and then discuss the more general form.

Lemma (Special Zorn’s Lemma for Sets). *Let \mathcal{S} be a non-empty collection of sets. A non-empty subset \mathcal{C} of \mathcal{S} is said to be a chain if for any $A, B \in \mathcal{C}$, either $A \subseteq B$ or $B \subseteq A$. Suppose that whenever \mathcal{C} is a chain in \mathcal{S} , the union $\cup_{C \in \mathcal{C}} C$ is an element of \mathcal{S} . Then \mathcal{S} contains a maximal element.*

Proof. Sorry, you’ll have to look this one up. ■

Let us show that this applies to our situation. We are given $X, Y \subseteq V$ where X is linearly independent and Y spans V . We let \mathcal{S} be the collection of all subsets Y' of V such that $X \cup Y'$ is linearly independent. We need to show the hypothesis of Zorn’s Lemma applies, and then we will be able to conclude that \mathcal{S} contains a maximal element Y' , which is exactly what we want.

Let \mathcal{C} be a chain in \mathcal{S} , and put $Z = \cup_{C \in \mathcal{C}} C$. Clearly $Z \subseteq Y$; we must show $X \cup Z$ is linearly independent. Suppose not: then there are $x_1, \dots, x_n \in X$, $y_1, \dots, y_m \in Z$ such that some non-trivial linear combination of all these elements is 0. Each $y_i \in C_i$ for some $C_i \in \mathcal{C} \subseteq \mathcal{S}$. Since \mathcal{C} is a chain, there is an index j , $1 \leq j \leq m$, such that $C_i \subseteq C_j$ for all $i = 1, \dots, m$. The elements $x_1, \dots, x_n, y_1, \dots, y_m$ all lie in $X \cup C_j$, and since $C_j \in \mathcal{S}$, these elements must be linearly independent. This contradicts the choice of these elements and shows that $X \cup Z$ must be linearly independent after all. Thus $Z \in \mathcal{S}$, as required.

The only properties of sets that occur in Zorn’s Lemma above are their properties relative to the partial order \subseteq . Thus it is reasonable to try to formulate Zorn’s Lemma in a more general setting, and it turns out that it is both true and useful.

Let \leq be a partial order on a set S . We say $C \subseteq S$ is a *chain* if C is linearly ordered under \leq , that is, if for any $c, d \in C$, either $c \leq d$ or $d \leq c$. We say $x \in S$ is an *upper bound* for $C \subseteq S$ if $c \leq x$ for every $c \in C$. Finally, we say $x \in S$ is a *maximal element* (or simply *maximal*) if $x \leq s$ for $s \in S$ implies $s = x$. (Note that we do not require $s \leq x$ for all $s \in S$. We are not assuming S is totally ordered. In particular, S may have many maximal elements — or it may have none at all.)

If we take for S our collection of sets \mathcal{S} and we take for \leq the inclusion relation \subseteq , then a chain in \mathcal{S} relative to \subseteq is precisely a chain in \mathcal{S} in the sense we defined above. Moreover, $\cup_{C \in \mathcal{C}} C$ is an upper bound for any $\mathcal{C} \subseteq \mathcal{S}$. Our hypothesis in the special Zorn’s Lemma for sets was that this union is in \mathcal{S} for any chain \mathcal{C} , and hence every chain in \mathcal{S} has an upper bound in \mathcal{S} . This is a special case of the general form of Zorn’s Lemma.

Lemma (General Zorn’s Lemma). *Let \leq be a partial order on a non-empty set S and suppose that every chain in S has an upper bound. Then S has a maximal element.*

Proof. This version is potentially far more powerful than the version we gave for sets, but we can prove the general version using the special one. Let \mathcal{S} be the set of all chains in S . If \mathcal{C} is a chain in \mathcal{S} , then \mathcal{C} is a collection of chains in S that are linearly ordered under \subseteq . It follows that $\cup_{C \in \mathcal{C}} C$ is a chain in S . (*Good exercise.*) Thus $\cup_{C \in \mathcal{C}} C \in \mathcal{S}$.

By the special version of Zorn's Lemma for sets, it follows that \mathcal{S} contains a maximal element C , that is, a chain C that cannot be added onto. By our hypothesis for the general Zorn's Lemma, this chain C has an upper bound x . We claim x is a maximal element in S .

If this is false, there is an element $y \in S$ with $x < y$. But then $s < y$ for every $s \in C$, so $C \cup \{y\}$ is a chain that properly contains C . This is impossible, and so x must be maximal. ■

The other place we appealed to finiteness was in our proof of Theorem 7.15. We were given $X, Y \subseteq V$ where X is linearly independent and Y spans V and we wished to show $|X| \leq |Y|$. We showed this under the hypothesis that Y is finite, using the Exchange Lemma, by fixing V, Y and taking a counterexample V, X, Y with $X \cap Y$ maximal.

By Zorn's Lemma (verify!) we can find a maximal X such that V, X, Y is a counterexample, but can we find one with $X \cap Y$ maximal? It is not clear how to order our counterexamples X so that $X \cap Y$ is maximized. For example, it seems quite possible to have counterexamples X, X' with $X \subsetneq X'$ but $X \cap Y = X' \cap Y$ (if there are any counterexamples at all!).

We can instead solve this particular problem by a counting argument. We need two facts about infinite sets. If S is a set, let $\mathcal{F}(S)$ denote the set of all *finite* subsets of S . The first fact we need is that if S is infinite, then $|S| = |\mathcal{F}(S)|$.

Suppose $f : S \rightarrow T$. The second fact we need is that if S is infinite and $|S| > |T|$, then there exists a $t \in T$ such that $\overleftarrow{f}(t) = \{s \in S \mid f(s) = t\}$ is infinite.

Both of these facts are consequences of the fact that for non-empty sets A, B , if at least one of them is infinite, then $|A \times B| = \max(|A|, |B|)$. (If \aleph, \beth are infinite cardinals, then $\aleph + \beth = \aleph \cdot \beth = \max(\aleph, \beth)$.) This implies, for example, that if \aleph is an infinite cardinal, a countable union of sets of cardinality \aleph has cardinality \aleph .

Assume these two facts and assume Y is infinite. Since Y spans V , for every $x \in X$, there is a finite subset $Z \subseteq Y$ such that $x = \sum_{z \in Z} a_z z$ for some scalars $a_z \in F$. For each x , pick a particular set Z — or shrink Y to a basis in which case Z is unique if we assume each $a_z \neq 0$ — and define a function $f : X \rightarrow \mathcal{F}(Y)$ by letting $f(x)$ be the chosen set Z .

Since Y is infinite, our first fact above tell us $|\mathcal{F}(Y)| = |Y|$. If $|X| > |Y|$, our second fact tells us there is a finite $Z \subseteq Y$ and an infinite $X' \subseteq X$ with $f(x) = Z$ for all $x \in X'$. Thus each element of X' lies in the finite dimensional subspace of V spanned by Z . Since $X' \subseteq X$, we know X' is linearly independent. Thus we have an infinite linearly independent set contained in a vector space of finite dimension. We proved this is impossible in the finite case of Theorem 7.15.