Linear Spaces

We have seen (12.1-12.3 of Apostol) that n-tuple space V_n has the following properties:

Addition:

1. (Commutativity) $A + B = B + A$.
2. (Associativity) $A + (B+C) = (A+B) + C$.
3. (Existence of zero) There is an element $\underline{0}$ such that $A + \underline{0} = A$ for all A.
4. (Existence of negatives) Given A, there is a B such that $A + B = \underline{0}$.

Scalar multiplication:

5. (Associativity) $c(dA) = (cd)A$.
6. (Distributivity) $(c+d)A = cA + dA$, $c(A+B) = cA + cB$.
7. (Multiplication by unity) $1A = A$.

Definition. More generally, let V be *any* set of objects (which we call vectors). And suppose there are two operations on V, as follows: The first is an operation (denoted $+$) that assigns to each pair A, B of vectors, a vector denoted $A + B$. The second is an operation that assigns to each real number c and each vector A, a vector denoted cA. Suppose also that the seven preceding properties hold. Then V, with these two operations, is called a linear space (or a *vector space*). The seven properties are called the *axioms for a linear space*.

There are many examples of linear spaces besides n-tuple space V_n. The study of linear spaces and their properties is dealt with in a subject called Linear Algebra. We shall treat only those aspects of linear algebra needed for calculus. Therefore we will be concerned only with n-tuple space V_n and with certain of its subsets called "linear subspaces" :

Definition. Let W be a non-empty subset of V_n ; suppose W is closed under vector addition and scalar multiplication. Then W is called a linear subspace of V_n (or sometimes simply a subspace of V_n.)

To say W is closed under vector addition and scalar multiplication means that for every pair A, B of vectors of W, and every scalar c, the vectors $A + B$ and cA belong to W. Note that it is automatic that the zero vector $\underline{0}$ belongs to W, since for any A in W, we have $\underline{0} = 0A$. Furthermore, for each A in W, the vector $-A$ is also in W. This means (as you can readily check) that W is a linear space in its own right (i.e., it satisfies all the axioms for a linear space).

Subspaces of V_n may be specified in many different ways, as we shall see.

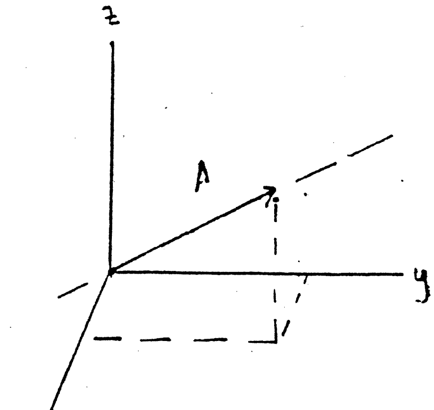
Example 1. The subset of V_n consisting of the 0-tuple alone is a subspace of V_n; it is the "smallest possible" subspace. And of course V_n is by definition a subspace of V_n; it is the "largest possible" subspace.

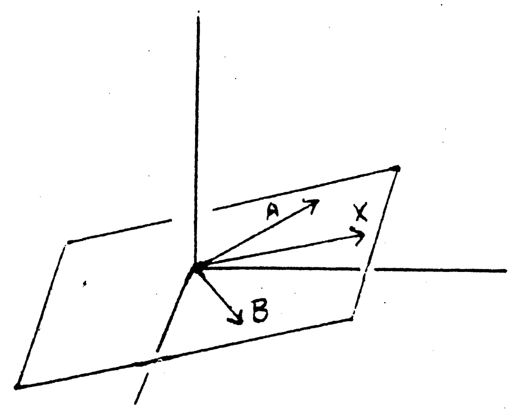
Example 2. Let A be a fixed non-zero vector. The subset of V_n consisting of all vectors X of the form $X = cA$ is a subspace of V_n. It is called the subspace spanned by A. In the case $n = 2$ or 3, it can be pictured as consisting of all vectors lying on a line through the origin.

Example 3. Let A and B be given non-zero vectors that are not parallel. The subset of V_n consisting of all vectors of the form

$$X = cA + dB$$

is a subspace of V_n. It is called the subspace spanned by A and B. In the case $n = 3$, it can be pictured as consisting of all vectors lying in the plane through the origin that contains A and B.





We generalize the construction given in the preceding examples as follows:

Definition. Let $S = \{A_1, \ldots, A_k\}$ be a set of vectors in V_n. A vector X of V_n of the form

$$X = c_1A_1 + \ldots + c_kA_k$$

is called a linear combination of the vectors $A_1,\ldots,A_k$. The set W of all such vectors X is a subspace of V_n, as we will see; it is said to be the subspace spanned by the vectors $A_1,\ldots,A_k$. It is also called the linear span of $A_1,\ldots,A_k$ and denoted by $L(S)$.

Let us show that W is a subspace of V_n. If X and Y belong to W, then

$$X = c_1A_1 + \cdots + c_kA_k \quad \text{and} \quad Y = d_1A_1 + \cdots + d_kA_k,$$

for some scalars c_i and d_i. We compute

$$X + Y = (c_1+d_1)A_1 + \cdots + (c_k+d_k)A_k,$$
$$aX = (ac_1)A_1 + \cdots + (ac_k)A_k,$$

so both $X + Y$ and aX belong to W by definition. Thus W is a subspace of V_n.

Giving a spanning set for W is one standard way of specifying W. Different spanning sets can of course give the same subspace. For example, it is intuitively clear that, for the plane through the origin in Example 3, <u>any</u> two non-zero vectors C and D that are not parallel and lie in this plane will span it. We shall give a proof of this fact shortly.

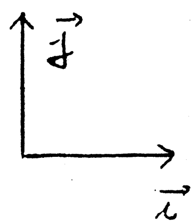
<u>Example 4.</u> The n-tuple space V_n has a natural spanning set, namely the vectors

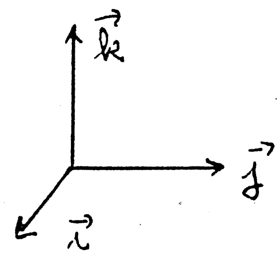
$$E_1 = (1,0,0,\ldots,0),$$
$$E_2 = (0,1,0,\ldots,0),$$
$$\cdots$$
$$E_n = (0,0,0,\ldots,1).$$

These are often called the <u>unit coordinate vectors</u> in V_n. It is easy to see that they span V_n, for if $X = (x_1,\ldots,x_n)$ is an element of V_n, then

$$X = x_1E_1 + \cdots + x_nE_n.$$

In the case where $n = 2$, we often denote the unit coordinate vectors E_1 and E_2 in V_2 by $\vec{i}$ and $\vec{j}$, respectively. In the case where $n = 3$, we often denote E_1, E_2, and E_3 by $\vec{i}$, $\vec{j}$, and $\vec{k}$ respectively. They are pictured as in the accompanying figure.





<u>Example 5</u>. The subset W of V_3 consisting of all vectors of the form $(a,b,0)$ is a subspace of V_3. For if X and y are 3-tuples whose third component is 0, so are $X + Y$ and cX. It is easy to see that W is the linear span of $(1,0,0)$ and $(0,1,0)$.

<u>Example 6</u>. The subset of V_3 consisting of all vectors of the form $X = (3a+2b, a-b, a+7b)$ is a subspace of V_3. It consists of all vectors of the form

$$X = a(3,1,1) + b(2,-1,7),$$

so it is the linear span of $(3,1,1)$ and $(2,-1,7)$.

<u>Example 7.</u> The set W of all tuples (x_1, x_2, x_3, x_4) such that

$$3x_1 - x_2 + 5x_3 + x_4 = 0$$

is a subspace of V_4, as you can check. Solving this equation for x_4, we see that a 4-tuple belongs to W if and only if it has the form

$$X = (x_1, x_2, x_3, -3x_1 + x_2 - 5x_3),$$

where x_1 and x_2 and x_3 are arbitrary. This element can be written in the form

$$X = x_1(1,0,0,-3) + x_2(0,1,0,1) + x_3(0,0,1,-5).$$

It follows that $(1,0,0,-3)$ and $(0,1,0,1)$ and $(0,0,1,-5)$ span W.

Exercises

1. Show that the subset of V_3 specified in Example 5 is a subspace of V_3. Do the same for the subset of V_4 specified in Example 7. What can you say about the set of all $(x_1,\ldots,x_n)$ such that $a_1x_1 + \ldots + a_nx_n = 0$ in general? (Here we assume $A = (a_1,\ldots,a_n)$ is not the zero vector.) Can you give a geometric interpretation?

2. In each of the following, let W denote the set of all vectors (x,y,z) in V_3 satisfying the condition given. (Here we use (x,y,z) instead of (x_1,x_2,x_3) for the general element of V_3.) Determine whether W is a subspace of V_3. If it is, draw a picture of it or describe it geometrically, and find a spanning set for W.

(a) $x = 0$.

(b) $x + y = 0$.

(c) $x + y = 1$.

(d) $x = y$ and $2x = z$.

(e) $x = y$ or $2x = z$.

(f) $x^2 - y^2 = 0$.

(g) $x^2 + y^2 = 0$.

3. Consider the set F of all real-valued functions defined on the interval $[a,b]$.

(a) Show that F is a linear space if $f + g$ denotes the usual sum of functions and cf denotes the usual product of a function by a real number. What is the zero vector?

(b) Which of the following are subspaces of F?

(i) All continuous functions.

(ii) All integrable functions.

(iii) All piecewise-monotonic functions.

(iv) All differentiable functions.

(v) All functions f such that $f(a) = 0$.

(vi) All polynomial functions.

Linear independence

Definition. We say that the set $S = \{A_1,\ldots,A_k\}$ of vectors of V_n spans the vector X if X belongs to $L(S)$, that is, if

$$X = c_1A_1 + \ldots + c_kA_k$$

for some scalars c_i. If S spans the vector X, we say that S spans X uniquely if the equations

$$X = \sum_{i=1}^{k} c_iA_i \qquad \text{and} \qquad X = \sum_{i=1}^{k} d_iA_i$$

imply that $c_i = d_i$ for all i.

It is easy to check the following:

Theorem 1. Let $S = \{A_1,\ldots,A_k\}$ be a set of vectors of V_n; let X be a vector in $L(S)$. Then S spans X uniquely if and only if S spans the zero vector $\underline{0}$ uniquely.

Proof. Note that $\underline{0} = \sum 0A_i$. This means that S spans the zero vector uniquely if and only if the equation

$$\underline{0} = \sum_{i=1}^{k} c_i A_i$$

implies that $c_i = 0$ for all i.

Suppose S spans $\underline{0}$ uniquely. To show S spans X uniquely, suppose

$$X = \sum_{i=1}^{k} c_i A_i \quad \text{and} \quad X = \sum_{i=1}^{k} d_i A_i \ .$$

Subtracting, we see that

$$\underline{0} = \sum_{i=1}^{k} (c_i - d_i) A_i \ ,$$

whence $c_i - d_i = 0$, or $c_i = d_i$, for all i.

Conversely, suppose S spans X uniquely. Then

$$X = \sum_{i=1}^{k} x_i A_i$$

for some (unique) scalars x_i. Now if

$$\underline{0} = \sum_{i=1}^{k} c_i A_i \ ,$$

it follows that

$$X = X + \underline{0} = \sum_{i=1}^{k} (x_i + c_i) A_i \ .$$

Since S spans X uniquely, we must have $x_i = x_i + c_i$, or $c_i = 0$, for all i. □

This theorem implies that if S spans one vector of $L(S)$ uniquely, then it spans the zero vector uniquely, whence it spans every vector of $L(S)$ uniquely. This condition is important enough to be given a special name:

Definition. The set $S = \{A_1, \ldots, A_k\}$ of vectors of V_n is said to be linearly independent (or simply, independent) if it spans the zero vector uniquely. The vectors themselves are also said to be independent in this situation.

If a set is not independent, it is said to be dependent.

Example 8. If a subset T of a set S is dependent, then S itself is dependent. For if T spans $\underline{0}$ non-trivially, so does S. (Just add on the additional vectors with zero coefficients.)

This statement is equivalent to the statement that if S is independent, then so is any subset of S.

Example 9. Any set containing the zero vector $\underline{0}$ is dependent. For example, if $S = \{A_1,\ldots,A_k\}$ and $A_1 = \underline{0}$, then

$$\underline{0} = 1A_1 + 0A_2 + \ldots + 0A_k .$$

Example 10. The unit coordinate vectors $E_1,\ldots,E_n$ in V_n span $\underline{0}$ uniquely, so they are independent.

Example 11. Let $S = \{A_1,\ldots,A_k\}$. If the vectors A_i are non-zero and mutually orthogonal, then S is independent. For suppose

$$\underline{0} = \sum_{i=1}^{k} c_i A_i .$$

Taking the dot product of both sides of this equation with A_1 gives the equation

$$0 = c_1\, A_1 \cdot A_1$$

(since $A_i \cdot A_1 = 0$ for $i \neq 1$). Now $A_1 \neq \underline{0}$ by hypothesis, whence $A_1 \cdot A_1 \neq 0$, whence $c_1 = 0$. Similarly, taking the dot product with A_j for the fixed index j shows that $c_j = 0$.

Sometimes it is convenient to replace the vectors A_i by the vectors $B_i = A_i/\|A_i\|$. Then the vectors $B_1,\ldots,B_k$ are of unit length and are mutually orthogonal. Such a set of vectors is called an orthonormal set. The coordinate vectors $E_1,\ldots,E_n$ form such a set.

Example 12. A set consisting of a single vector A is independent if $A \neq \underline{0}$. A set consisting of two non-zero vectors A, B is independent if and only if the vectors are not parallel. More generally, one has the following result:

Theorem 2. The set $S = \{A_1, \ldots, A_k\}$ is independent if and only if none of the vectors A_j can be written as a linear combination of the others.

Proof. Suppose first that one of the vectors equals a linear combination of the others. For instance, suppose that

$$A_1 = c_2 A_2 + \cdots + c_k A_k;$$

then the following non-trivial linear combination equals zero:

$$A_1 - c_2 A_2 - \cdots - c_k A_k = \underline{0}.$$

Conversely, if

$$c_1 A_1 + c_2 A_2 + \cdots + c_k A_k = \underline{0},$$

where not all the c_i are equal to zero, we can choose m so that $c_m \neq 0$, and obtain the equation

$$A_m = -(c_1/c_m) A_1 - \cdots - (c_k/c_m) A_k,$$

where the sum on the right extends over all indices different from m. □

Given a subspace W of V_n, there is a very important relation that holds between spanning sets for W and independent sets in W:

Theorem 3. Let W be a subspace of V_n that is spanned by the k vectors $A_1, \ldots, A_k$. Then any independent set of vectors in W contains at most k vectors.

Proof. Let $\{B_1,\ldots,B_m\}$ be a set of vectors of W; let $m > k$. We wish to show that these vectors are dependent. That is, we wish to find scalars $x_1,\ldots,x_m$, not all zero, such that

$$\sum_{j=1}^{m} x_j B_j = \underline{0} .$$

Since each vector B_j belongs to W, we can write it as a linear combination of the vectors A_i. We do so, using a "double-indexing" notation for the coefficents, as follows:

$$B_j = a_{1j} A_1 + a_{2j} A_2 + \cdots + a_{kj} A_k .$$

Multiplying the equation by x_j and summing over j, and collecting terms, we have the equation

$$\sum_{j=1}^{m} x_j B_j = \left(\sum_{j=1}^{m} x_j a_{1j}\right) A_1 + \left(\sum_{j=1}^{m} x_j a_{2j}\right) A_2 + \cdots + \left(\sum_{j=1}^{m} x_j a_{kj}\right) A_k .$$

In order for $\sum x_j B_j$ to equal $\underline{0}$, it will suffice if we can choose the x_j so that coefficient of each vector A_i in this equation equals 0. Now the numbers a_{ij} are given, so that finding the x_j is just a matter of solving a (homogeneous) system consisting of k equations in m unknowns. Since $m > k$, there are more unknowns than equations. In this case the system always has a non-trivial solution X (i.e., one different from the zero vector). This is a standard fact about linear equations, which we now prove. □

First, we need a definition.

Definition. Given a homogeneous system of linear equations, as in (*) following, a solution of the system is a vector $(x_1,\ldots,x_n)$ that satisfies each equation of the system. The set of all solutions is a linear subspace of V_n (as you can check). It is called the solution space of the system.

It is easy to see that the solution set is a subspace. If we let

$$A_j = (a_{j_1}, a_{j_2}, \ldots, a_{j_n})$$

be the n-tuple whose components are the coefficents appearing in the j^{th} equation of the system, then the solution set consists of those X such that $A_j \cdot X = 0$ for all j. If X and Y are two solutions, then

$$A_j \cdot (X + Y) = A_j \cdot X + A_j \cdot Y = \underline{0}$$

and

$$A_j \cdot (cX) = c(A_j \cdot X) = 0$$

Thus X + Y and cX are also solutions, as claimed.

Theorem 4. Given a homogeneous system of k linear equations in n unknowns. If k is less than n, then the solution space contains some vector other than **0**.

Proof. We are concerned here only with proving the *existence* of some solution other than **0**, not with actually finding such a solution in practice, nor with finding all possible solutions. (We will study the practical problem in much greater detail in a later section.)

We start with a system of k equations in n unknowns:

$$(*) \qquad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0, \\ &\vdots \\ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n &= 0. \end{aligned}$$

Our procedure will be to reduce the size of this system step-by-step by eliminating first x_1, then x_2, and so on. After $k - 1$ steps, we will be reduced to solving just one equation and this will be easy. But a certain amount of care is needed in the description—for instance, if $a_{11} = \cdots = a_{k1} = 0$, it is nonsense to speak of "eliminating" x_1, since all its coefficients are zero. We have to allow for this possibility.

To begin then, if all the coefficients of x_1 are zero, you may verify that the vector $(1, 0, \ldots, 0)$ is a solution of the system which is different from **0**, and you are done. Otherwise, at least one of the coefficients of x_1 is nonzero, and we may suppose for convenience that the equations have been arranged so that this happens in the first equation, with the result that $a_{11} \neq 0$. We multiply the first equation by the scalar a_{21}/a_{11} and then subtract it from the second, eliminating the x_1-term from the second equation. Similarly, we eliminate the x_1-term in each of the remaining equations. The result is a *new* system of linear equations of the form

$$(**)\qquad a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0,$$

$$\boxed{\begin{aligned} b_{22}x_2 + \cdots + b_{2n}x_n &= 0,\\ &\vdots\\ b_{k2}x_2 + \cdots + b_{kn}x_n &= 0. \end{aligned}}$$

Now any solution of this new system of equations is also a solution of the old system (*), because we can recover the old system from the new one: we merely multiply the first equation of the system (**) by the same scalars we used before, and then we *add* it to the corresponding later equations of this system.

The crucial thing about what we have done is contained in the following statement: If the smaller system enclosed in the box above has a solution other than the zero vector, then the larger system (**) also has a solution other than the zero vector [so that the original system (*) we started with has a solution other than the zero vector]. We prove this as follows: Suppose $(d_2, \ldots, d_n)$ is a solution of the smaller system, different from $(0, \ldots, 0)$. We substitute into the first equation and solve for x_1, thereby obtaining the following vector,

$$\big((-1/a_{11})(a_{12}d_2 + \cdots + a_{1n}d_n), d_2, \ldots, d_n\big),$$

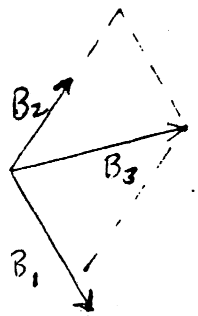
which you may verify is a solution of the larger system (**).

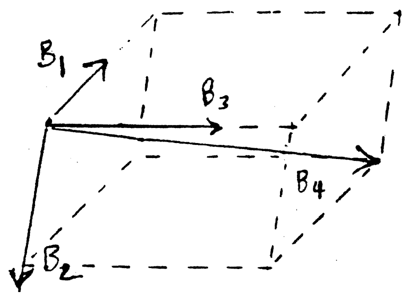
In this way we have reduced the size of our problem; we now need only to prove our theorem for a system of $k - 1$ equations in $n - 1$ unknowns. If we apply this reduction a second time, we reduce the problem to proving the theorem for a system of $k - 2$ equations in $n - 2$ unknowns. Continuing in this way, after $k - 1$ elimination steps in all, we will be down to a system consisting of only one equation, in $n - k + 1$ unknowns. Now $n - k + 1 \geq 2$, because we assumed as our hypothesis that $n > k$; thus our problem reduces to proving the following statement: a "system" consisting of *one* linear homogeneous equation in *two or more* unknowns always has a solution other than **0**.

We leave it to you to show that this statement holds.(Be sure you consider the case where one or more or all of the coefficents are zero.) □

Example 13. We have already noted that the vectors $E_1, \ldots, E_n$ span all of V_n. It follows, for example, that any three vectors in V_2 are dependent, that is, one of them equals a linear combination of the others. The same holds for any four vectors in V_3. The accompanying picture makes these facts plausible.

Similarly, since the vectors $E_1, \ldots, E_n$ are independent, any spanning set of V_n must contain at least n vectors. Thus no two vectors can span V_3, and no set of three vectors can span V_4.





<u>Theorem 5.</u> Let W be a subspace of V_n that does not consist of $\underline{0}$ alone. Then:

(a) The space W has a linearly independent spanning set.

(b) Any two linearly independent spanning sets for W have the same number k of elements; furthermore, $k < n$ unless W is all of V_n.

<u>Proof.</u> (a) Choose $A_1 \neq \underline{0}$ in W. Then the set $\{A_1\}$ is independent. In general, suppose $\{A_1, \ldots, A_i\}$ is an independent set of vectors of W. If this set spans W, we are finished. Otherwise, we can choose a vector A_{i+1} of W that is not in $L(A_1, \ldots, A_i)$. Then the set $\{A_1, \ldots, A_i, A_{i+1}\}$ is independent: For suppose that

$$c_1 A_1 + \ldots + c_i A_i + c_{i+1} A_{i+1} = \underline{0}$$

for some scalars c_i not all zero. If $c_{i+1} = 0$, this equation contradicts independence of $\{A_1, \ldots, A_i\}$, while if $c_{i+1} \neq 0$, we can solve this equation for A_{i+1}, contradicting the fact that A_{i+1} does not belong to $L(A_1, \ldots, A_i)$.

Continuing the process just described, we can find larger and larger independent sets of vectors in W. The process stops only when the set we obtain spans W. Does it ever stop? Yes, for W is contained in V_n, and V_n contains no more than n independent vectors. So the process <u>cannot</u> be repeated indefinitely!

(b) Suppose $S = \{A_1,\ldots,A_k\}$ and $T = \{B_1,\ldots,B_j\}$ are two linearly independent spanning sets for W. Because S is independent and T spans W, we must have $k \leq j$, by the preceding theorem. Because S spans W and T is independent, we must have $k \geq j$. Thus $k = j$.

Now V_n contains no more than n independent vectors; therefore we must have $k \leq n$. Suppose that W is not all of V_n. Then we can choose a vector A_{k+1} of V_n that is not in W. By the argument just given, the set $\{A_1,\ldots,A_k,A_{k+1}\}$ is independent. It follows that $k+1 \leq n$, so that $k < n$. □

<u>Definition.</u> Given a subspace W of V_n that does not consist of $\underline{0}$ alone, it has a linearly independent spanning set. Any such set is called a <u>basis</u> for W, and the number of elements in this set is called the <u>dimension</u> of W. We make the convention that if W consists of $\underline{0}$ alone, then the dimension of W is zero.

<u>Example 14.</u> The space V_n has a "natural" basis consisting of the vectors $E_1,\ldots,E_n$. It follows that V_n has dimension n. (Surprise!) There are many other bases for V_n. For instance, the vectors

$$A_1 = (1,0,0,\ldots,0)$$
$$A_2 = (1,1,0,\ldots,0)$$
$$A_3 = (1,1,1,\ldots,0)$$
$$\ldots$$
$$A_n = (1,1,1,\ldots,1)$$

form a basis for V_n, as you can check.

Exercises

1. Consider the subspaces of V_3 listed in Exercise 2, p. A6. Find bases for each of these subspaces, and find spanning sets for them that are not bases.

2. Check the details of Example 14.

3. Suppose W has dimension k. (a) Show that any independent set in W consisting of k vectors spans W. (b) Show that any spanning set for W consisting of k vectors is independent.

4. Let $S = \{A_1,\ldots,A_m\}$ be a spanning set for W. Show that S contains a basis for W. [Hint: Use the argument of Theorem 5.]

5. Let $\{A_1,\ldots,A_k\}$ be an independent set in V_n. Show that this set can be extended to a basis for V_n. [Hint: Use the argument of Theorem 5.]

6. If V and W are subspaces of V_n and V_k, respectively, a function $T : V \to W$ is called a linear transformation if it satisfes the usual linearity properties:

$$T(X + Y) = T(X) + T(Y),$$

$$T(cX) = cT(X).$$

If T is one-to-one and carries V onto W, it is called a linear isomorphism of vector spaces.

Suppose $A_1,\ldots,A_k$ is a basis for V; let $B_1,\ldots,B_k$ be arbitrary vectors of W. (a) Show there exists a linear transformation $T : V \to W$ such that $T(A_i) = B_i$ for all i. (b) Show this linear transformation is unique.

7. Let W be a subspace of V_n; let $A_1,\ldots,A_k$ be a basis for W. Let X, Y be vectors of W. Then $X = \sum x_i A_i$ and $Y = \sum y_i A_i$ for unique scalars x_i and y_i. These scalars are called the components of X and Y, respectively, relative to the basis $A_1,\ldots,A_k$.

(a) Note that $X + Y = \sum (x_i + y_i)A_i$ and $cX = \sum (cx_i)A_i$. Conclude that the function $T : V_k \to W$ defined by $T(x_1,\ldots,x_k) = \sum x_i A_i$ is a linear isomorphism.

(b) Suppose that the basis $A_1,\ldots,A_k$ is an orthonormal basis. Show that $X\cdot Y = \sum x_i y_i$. Conclude that the isomorphism T of (a) preserves the dot product, that is, $T(X)\cdot T(Y) = X\cdot Y$.

8. Prove the following:

Theorem. If W is a subspace of positive dimension of V_n, then W has an orthonormal basis.

Proof. Step 1. Let $B_1,\ldots,B_m$ be mutually orthogonal non-zero vectors in V_n; let A_{m+1} be a vector not in $L(B_1,\ldots,B_m)$. Given scalars $c_1,\ldots,c_m$, let

$$B_{m+1} = A_{m+1} + c_1B_1 + \ldots + c_mB_m .$$

Show that B_{m+1} is different from $\underline{0}$ and that $L(B_1,\ldots,B_m,B_{m+1}) = L(B_1,\ldots,B_m,A_{m+1})$. Then show that the c_i may be so chosen that B_{m+1} is orthogonal to each of $B_1,\ldots,B_m$.

Step 2. Show that if W is a subspace of V_n of positive dimension, then W has a basis consisting of vectors that are mutually orthogonal. [Hint: Proceed by induction on the dimension of W.]

Step 3. Prove the theorem.

Gauss-Jordan elimination

If W is a subspace of V_n, specified by giving a spanning set for W, we have at present no constructive process for determining the dimension of W nor of finding a basis for W, although we know these exist. There is a simple procedure for carrying out this process; we describe it now.

Definition. The rectangular array of numbers

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \cdots & \\ a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix}$$

is called a <u>matrix</u> of size k by n. The number a_{ij} is called the <u>entry</u> of A in the $i^{\underline{th}}$ row and $j^{\underline{th}}$ column. Suppose we let A_i be the vector

$$A_i = (a_{i1}, a_{i2}, \ldots, a_{in})$$

for $i = 1,\ldots k$. Then A_i is just the i^{th} row of the matrix A. The subspace of V_n spanned by the vectors $A_1,\ldots,A_k$ is called the <u>row space</u> of the matrix A.

We now describe a procedure for determining the dimension of this space. It involves applying operations to the matrix A, of the following types:

(1) Interchange two rows of A.

(2) Replace row i of A by itself plus a scalar multiple of another row, say row m.

(3) Multiply row i of A by a non-zero scalar.

These operations are called the <u>elementary row operations.</u> Their usefulness comes from the following fact:

<u>Theorem 6.</u> Suppose B is the matrix obtained by applying a sequence of elementary row operations to A, successively. Then the row spaces of A and B are the same.

<u>Proof.</u> It suffices to consider the case where B is obtained by applying a single row operation to A. Let $A_1,\ldots,A_k$ be the rows of A, and let $B_1,\ldots,B_k$ be the rows of B.

If the operation is of type (1), these two sets of vectors are the same (only their order is changed), so the spaces they span are the same. If the operation is of type (3), then

$$B_i = cA_i \quad \text{and} \quad B_j = A_j \quad \text{for } j \neq i.$$

Clearly, any linear combination of $B_1,\ldots,B_k$ can be written as a linear combination of $A_1,\ldots,A_k$. Because $c \neq 0$, the converse is also true. Finally, suppose the operation is of type (2). Then

$$B_i = A_i + dA_m \quad \text{and} \quad B_j = A_j \quad \text{for } j \neq i.$$

Again, any linear combination of $B_1,\ldots,B_k$ can be written as a linear combination of $A_1,\ldots,A_k$. Because

$$A_i = B_i - dA_m = B_i - dB_m ,$$

and

$$A_j = B_j \quad \text{for } j \neq i ,$$

the converse is also true. □

The Gauss-Jordan procedure consists of applying elementary row operations to the matrix A until it is brought into a form where the dimension of its row space is obvious. It is the following:

<u>Gauss-Jordan</u> <u>elimination.</u> Examine the first column of your matrix.

(I) If this column consists entirely of zeros, nothing needs to be done. Restrict your attention now to the matrix obtained by deleting the first column, and begin again.

(II) If this column has a non-zero entry, exchange rows if necessary to bring it to the top row. Then add multiplesof the top row to the lower rows so as to make all remaining entries in the first column into zeros. Restrict your attention now to the matrix obtained by deleting the first column and first row, and begin again.

The procedure stops when the matrix remaining has only one row.

Let us illustrate the procedure with an example.

Problem. Find the dimension of the row space of the matrix

$$A = \begin{bmatrix} 0 & 1 & 4 & 1 & 2 \\ -1 & -2 & 0 & 9 & -1 \\ 1 & 2 & 0 & -6 & 1 \\ 2 & 5 & 4 & -10 & 4 \end{bmatrix}.$$

Solution. First step. Alternative (II) applies. Exchange rows (1) and (2), obtaining

$$\begin{bmatrix} -1 & -2 & 0 & 9 & -1 \\ 0 & 1 & 4 & 1 & 2 \\ 1 & 2 & 0 & -6 & 1 \\ 2 & 5 & 4 & -10 & 4 \end{bmatrix}$$

Replace row (3) by row (3) + row (1); then replace (4) by (4) + 2 times (1).

$$\begin{bmatrix} -1 & -2 & 0 & 9 & -1 \\ 0 & 1 & 4 & 1 & 2 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 1 & 4 & 8 & 2 \end{bmatrix}$$

Second step. Restrict attention to the matrix in the box. (II) applies. Replace row (4) by row (4) - row (2), obtaining

$$\begin{bmatrix} -1 & -2 & 0 & 9 & -1 \\ 0 & 1 & 4 & 1 & 2 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 7 & 0 \end{bmatrix}$$

Third step. Restrict attention to the matrix in the box. (I) applies, so nothing needs be done. One obtains the matrix

$$\begin{bmatrix} -1 & -2 & 0 & 9 & -1 \\ 0 & 1 & 4 & 1 & 2 \\ 0 & 0 & 0 & \boxed{3} & 0 \\ 0 & 0 & 0 & 7 & 0 \end{bmatrix}$$

Fourth step. Restrict attention to the matrix in the box. (II) applies. Replace row (4) by row (4) $- \frac{7}{3}$ row (3), obtaining

$$B = \begin{bmatrix} \textcircled{-1} & -2 & 0 & 9 & -1 \\ 0 & \textcircled{1} & 4 & 1 & 2 \\ 0 & 0 & 0 & \textcircled{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The procedure is now finished. The matrix B we end up with is in what is called echelon or "stair-step"form. The entries beneath the steps are zero. And the entries -1, 1, and 3 that appear at the "inside corners" of the stairsteps are non-zero. These entries that appear at the "inside corners" of the stairsteps are often called the pivots in the echelon form.

You can check readily that the non-zero rows of the matrix B are independent. (We shall prove this fact later.) It follows that the non-zero rows of the matrix B form a basis for the row space of B, and hence a basis for the row space of the original matrix A. Thus this row space has dimension 3.

The same result holds in general. If by elementary operations you reduce the matrix A to the echelon form B, then the non-zero rows are B are independent, so they form a basis for the row space of B, and hence a basis for the row space of A.

Now we discuss how one can continue to apply elementary operations to reduce the matrix B to an even nicer form. The procedure is this:

Begin by considering the last non-zero row. By adding multiples of this row to each row above it, one can bring the matrix to the form where each entry lying above the pivot in this row is zero. Then continue the process, working now with the next-to-last non-zero row. Because all the entries above the last pivot are already zero, they remain zero as you add multiples of the next-to-last non-zero row to the rows above it. Similarly one continues. Eventually the matrix reaches the form where all the entries that are directly above the pivots are zero. (Note that the stairsteps do not change during this process, nor do the pivots themselves.)

Applying this procedure in the example considered earlier, one brings the matrix B into the form

$$C = \begin{bmatrix} -1 & 0 & 8 & 0 & 3 \\ 0 & 1 & 4 & 0 & 2 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note that up to this point in the reduction process , we have used only elementary row operations of types (1) and (2). It has not been necessary to multiply a row by a non-zero scalar. This fact will be important later on.

We are not yet finished. The final step is to multiply each non-zero row by an appropriate non-zero scalar, chosen so as to make the pivot entry into 1. This we can do, because the pivots are non-zero. At the end of this process, the matrix is in what is called <u>reduced echelon form.</u>

The reduced echelon form of the matrix C above is the matrix

$$D = \begin{bmatrix} 1 & 0 & -8 & 0 & -3 \\ 0 & 1 & 4 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

As we have indicated, the importance of this process comes from the following theorem:

Theorem 7. Let A be a matrix; let W be its row space. Suppose we transform A by elementary row operations into the echelon matrix B, or into the reduced echelon matrix D. Then the non-zero rows of B are a basis for W, and so are the non-zero rows of D.

Proof. The rows of B span W, as we noted before; and so do the rows of D. It is easy to see that no non-trivial linear combination of the non-zero rows of D equals the zero vector, because each of these rows has an entry of 1 in a position where the others all have entries of 0. Thus the dimension of W equals the number r of non-zero rows of D. This is the same as the number of non-zero rows of B. If the rows of B were not independent, then one would equal a linear combination of the others. This would imply that the row space of B could be spanned by fewer than r rows, which would imply that its dimension is less than r.

Exercises

1. Find bases for the row spaces of the following matrices:

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 2 & -1 & 4 \\ 0 & -1 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 3 & 2 & 1 \\ 5 & 3 & 3 \\ 7 & 4 & 5 \\ 1 & 1 & -1 \end{bmatrix}$$

$$B = \begin{bmatrix} 3 & 2 & 1 \\ 5 & 3 & 3 \\ 1 & 1 & -1 \end{bmatrix} \qquad E = \begin{bmatrix} 1 & -2 & 1 & 2 \\ 2 & 3 & -1 & -5 \\ 4 & -1 & 1 & -1 \\ 5 & -3 & 2 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 3 & 2 & 1 \\ 5 & 3 & 3 \\ 7 & 4 & 5 \end{bmatrix}$$

2. Reduce the matrices in Exercise 1 to reduced echelon form.
Save your answers for later use!

*3. Prove the following:

Theorem. The reduced echelon form of a matrix is unique.

Proof. Let D and D' be two reduced echelon matrices, whose rows span the same subspace W of V_n. We show that $D = D'$.

Let $R_1, \ldots, R_k$ be the non-zero rows of D ; and suppose that the pivots (first non-zero entries) in these rows occur in columns $j_1, \ldots, j_k$, respectively.

(a) Show that the pivots of D' occur in the columns $j_1, \ldots, j_k$. [Hint: Let R be a row of D'; suppose its pivot occurs in column p. We have $R = c_1R_1 + \cdots + c_kR_k$ for some scalars c_i . (Why?) Show that $c_i = 0$ if $j_i < p$. Derive a contradiction if p is not equal to any of $j_1, \ldots, j_k$.]

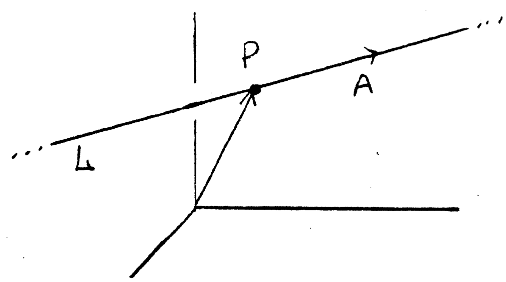
(b) If R is a row of D' whose pivot occurs in column j_m , show that $R = R_m$. [Hint: We have $R = c_1R_1 + \cdots + c_kR_k$ for some scalars c_i . Show that $c_i = 0$ for $i \neq m$, and $c_m = 1$.]

Parametric equations of lines and planes in V_n

Given n-tuples P and A, with $A \neq \underline{0}$, the line through P determined by A is defined to be the set of all points X such that

$$(*) \qquad X = P + tA$$

for some scalar t. It is denoted by L(P;A). The vector A is called a direction vector for the line. Note that if $P = \underline{0}$, then L is simply the 1-dimensional subspace of V_n spanned by A.



The equation (*) is often called a parametric equation for the line, and t is called the parameter in this equation. As t ranges over all real numbers, the corresponding point X ranges over all points of the line L. When t = 0, then X = P; when t = 1, then X = P + A; when $t = \frac{1}{2}$, then $X = P + \frac{1}{2}A$; and so on. All these are points of L.

Occasionally, one writes the vector equation out in scalar form as follows:

$$x_1 = p_1 + ta_1$$
$$x_2 = p_2 + ta_2$$
$$\cdots$$
$$x_n = p_n + ta_n$$

where $P = (p_1,\ldots,p_n)$ and $A = (a_1,\ldots,a_n)$. These are called the <u>scalar parametric equations</u> for the line.

Of course, there is no uniqueness here; a given line can be represented by many different parametric equations. The following theorem makes this result precise:

<u>Theorem 8</u>. <u>The lines $L(P;A)$ and $L(Q;B)$ are equal if and only if they have a point in common and A is parallel to B.</u>

<u>Proof.</u> If $L(P;A) = L(Q;B)$, then the lines obviously have a point in common. Since P and $P + A$ lie on the first line they also lie on the second line, so that

$$P = Q + t_1B \quad \text{and} \quad P + A = Q + t_2B$$

for distinct scalars t_1 and t_2. Subtracting, we have $A = (t_2-t_1)B$, so A is parallel to B.

Conversely, suppose the lines intersect in a point R, and suppose A and B are parallel. We are given that

$$P + t_1A = R = Q + t_2B$$

for some scalars t_1 and t_2, and that $A = cB$ for some $c \neq 0$. We can solve these equations for P in terms of Q and B:

$$P = Q + t_2B - t_1A = Q + (t_2-t_1c)B.$$

Now, given any point $X = P + tA$ of the line $L(P;A)$, we can write

$$X = P + tA = Q + (t_2-t_1c)B + tcB.$$

Thus X belongs to the line $L(Q;B)$.

Thus every point of $L(P;A)$ belongs to $L(Q;B)$. The symmetry of the argument shows that the reverse holds as well. □

<u>Definition</u>. It follows from the preceding theorem that given a line, its direction vector is uniquely determined up to a non-zero scalar multiple. We define two lines to be <u>parallel</u>

if their direction vectors are parallel.

Corollary 9. Distinct parallel lines cannot intersect.

Corollary 10. *Given a line* L *and a point* Q, *there is exactly one line containing* Q *that is parallel to* L.

Proof. Suppose L is the line L(P;A). Then the line L(Q;A) contains Q and is parallel to L. By Theorem 8, any other line containing Q and parallel to L is equal to this one. □

Theorem 11. *Given two distinct points* P *and* Q, *there is exactly one line containing them.*

Proof. Let $A = Q - P$; then $A \neq \underline{0}$. The line L(P;A) contains both P (since $P = P + 0A$) and Q (since $Q = P + 1A$).

Now suppose L(R;B) is some other line containing P and Q. Then

$$P = R + t_1 B,$$
$$Q = R + t_2 B,$$

for distinct scalars t_1 and t_2. It follows that

$$Q - P = (t_2 - t_1)B,$$

so that the vector $A = Q - P$ is parallel to B. It follows from Theorem 8 that

$$L(R;B) = L(P;A). \quad \square$$

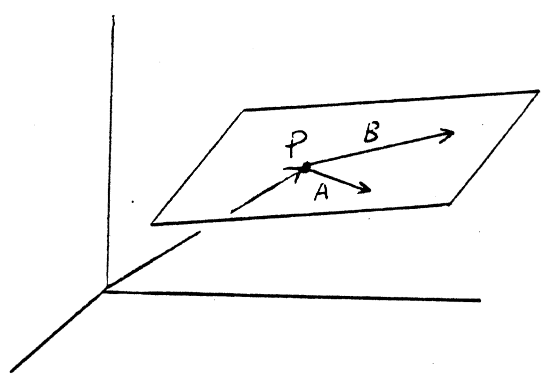
Now we study planes in V_n.

<u>Definition</u>. If P is a point of V_n and if A and B are independent vectors of V_n, we define the <u>plane through P determined by A and B</u> to be the set of all points X of the form

$$(*) \qquad X = P + sA + tB,$$

where s and t run through all real numbers. We denote this plane by M(P;A,B).

The equation (*) is called a <u>parametric equation</u> for the plane, and s and t are called the <u>parameters</u> in this equation. It may be written out as n scalar equations, if desired. When $s = t = 0$, then $X = P$; when $s = 1$ and $t = 0$, then $X = P + A$; when $s = 0$ and $t = 1$, then $X = P + B$; and so on.



Note that if $P = \underline{0}$, then this plane is just the 2-dimensional subspace of V_n spanned by A and B.

Just as for lines, a plane has many different parametric representations. More precisely, one has the following theorem:

<u>Theorem 12</u>. <u>The planes</u> M(P;A,B) <u>and</u> M(Q;C,D) <u>are equal if and only if they have a point in common and the linear span of</u> A <u>and</u> B <u>equals the linear span of</u> C <u>and</u> D.

<u>Proof</u>. If the planes are equal, they obviously have a

point in common. Furthermore, since P and $P + A$ and $P + B$ all lie on the first plane, they lie on the second plane as well. Then

$$P = Q + s_1C + t_1D,$$

$$P + A = Q + s_2C + t_2D,$$

$$P + B = Q + s_3C + t_3D,$$

are some scalars s_i and t_i. Subtracting, we see that

$$A = (s_2-s_1)C + (t_2-t_1)D,$$

$$B = (s_3-s_1)C + (t_3-t_1)D.$$

Thus A and B lie in the linear span of C and D. Symmetry shows that C and D lie in the linear span of A and B as well. Thus these linear spans are the same.

Conversely, suppose that the planes intersect in a point R and that $L(A,B) = L(C,D)$. Then

$$P + s_1A + t_1B = R = Q + s_2C + t_2D$$

for some scalars s_i and t_i. We can solve this equation for P as follows:

$$P = Q + (\text{linear combination of } A,B,C,D).$$

Then if X is any point of the first plane $M(P;A,B)$, we have

$$\begin{aligned} X &= P + sA + tB \qquad \text{for some scalars } s \text{ and } t, \\ &= Q + (\text{linear combination of } A,B,C,D) + sA + tB \\ &= Q + (\text{linear combination of } C,D), \end{aligned}$$

since A and B belong to $L(C,D)$.

Thus X belongs to $M(Q;C,D)$.

Symmetry of the argument shows that every point of $M(Q;C,D)$ belongs to $M(P;A,B)$ as well. □

Definition. Given a plane $M = M(P;A,B)$, the vectors A and B are not uniquely determined by M, but their linear span is. We say the planes $M(P;A,B)$ and $M(Q;C,D)$ are parallel if $L(A,B) = L(C,D)$.

Corollary 13. Two distinct parallel planes cannot intersect.

Corollary 14. Given a plane M and a point Q, there is exactly one plane containing Q that is parallel to M.

Proof. Suppose $M = M(P;A,B)$. Then $M(Q;A,B)$ is a plane that contains Q and is parallel to M. By Theorem 12 any other plane containing Q parallel to M is equal to this one. □

Definition. We say three points P,Q,R are collinear if they lie on a line.

Lemma 15. The points P,Q,R are collinear if and only if the vectors Q-P and R-P are dependent (i.e., parallel).

Proof. The line $L(P; Q-P)$ is the one containing P and Q, and the line $L(P;R-P)$ is the one containing P and R. If Q-P and R-P are parallel, these lines are the same, by Theorem 8, so P, Q, and R are collinear. Conversely, if P, Q, and R are collinear, these lines must be the same, so that Q-P and R-P must be parallel. □

Theorem 16. Given three non-collinear points P, Q, R, there is exactly one plane containing them.

Proof. Let $A = Q - P$ and $B = R - P$; then A and B are independent. The plane $M(P; A,B)$ contains P and $P + A = Q$ and $P + B = R$.

Now suppose $M(S;C,D)$ is another plane containing P, Q, and R. Then

$$P = S + s_1C + t_1D$$
$$Q = S + s_2C + t_2D$$
$$R = S + s_3C + t_3D$$

for some scalars s_i and t_i. Subtracting, we see that the vectors $Q - P = A$ and $R - P = B$ belong to the linear span of C and D. By symmetry, C and D belong to the linear span of A and B. Then Theorem 12 implies that these two planes are equal.

Exercises

1. We say the line L is parallel to the plane $M = M(P;A,B)$ if the direction vector of L belongs to $L(A,B)$. Show that if L is parallel to M and intersects M, then L is contained in M.

2. Show that two vectors A_1 and A_2 in V_n are linearly dependent if and only if they lie on a line through the origin.

3. Show that three vectors A_1, A_2, A_3 in V_n are linearly dependent if and only if they lie on some plane through the origin.

4. Let $P = (1,0,-1)$, $Q = (0,0,0)$, $R = (-2,5,0)$. Let $A = (1,-1,0)$, $B = (2,0,1)$.

(a) Find parametric equations for the line through P and Q, and for the line through R with direction vector A. Do these lines intersect?

(b) Find parametric equations for the plane through P, Q, and R, and for the plane through P determined by A and B.

5. Let L be the line in V_3 through the points $P = (1,0,2)$ and $Q = (-1,1,3)$. Let L' be the line through $\underline{0}$ parallel to the vector $A = (3,1,-1)$. Find parametric equations for the line that intersects both L and L' and is orthogonal to both of them.

Parametric equations for k-planes in V_n.

Following the pattern for lines and planes, one can define, more generally, a k-plane in V_n as follows:

Definition. Given a point P of V_n and a set $A_1,\ldots,A_k$ of k independent vectors in V_n, we define the k-plane through P determined by $A_1,\ldots,A_k$ to be the set of all vectors X of the form

$$X = P + t_1A_1 + \cdots + t_kA_k,$$

for some scalars t_i. We denote this set of points by $M(P;A_1,\ldots,A_k)$.

Said differently, X is in the k-plane $M(P;A_1,\ldots,A_k)$ if and only if $X - P$ is in the linear span of $A_1,\ldots,A_k$.

Note that if $P = \underline{0}$, then this k-plane is just the k- dimensional linear subspace of V_n spanned by $A_1,\ldots,A_k$.

Just as with the case of lines (1-planes) and planes (2-planes), one has the following results:

Theorem 17. Let $M_1 = M(P;A_1,\ldots,A_k)$ and $M_2 = M(Q;B_1,\ldots,B_k)$ be two k-planes in V_n. Then $M_1 = M_2$ if and only if they have a point in common and the linear span of $A_1,\ldots,A_k$ equals the linear span of $B_1,\ldots,B_k$.

Definition. We say that the k-planes M_1 and M_2 of this theorem are parallel if the linear span of $A_1,\ldots,A_k$ equals the linear span of $B_1,\ldots,B_k$.

Theorem 18. Given a k-plane M in V_n and a point Q, there is exactly one k-plane in V_n containing Q and parallel to M.

Lemma 19. Given points $P_0,\ldots,P_k$ in V_n, they are contained in a plane of dimension less than k if and only if the vectors

$P_1 - P_0, \ldots, P_k - P_0$ are dependent.

Theorem 20. Given $k+1$ distinct points $P_0, \ldots, P_k$ in V_n. If these points do not lie in any plane of dimension less than k, then there is exactly one k-plane containing them; it is the k-plane

$$M(P_0;\ P_1-P_0, \ldots, P_k-P_0).$$

More generally, we make the following definition:

Definition. If $M_1 = M(P;A_1,\ldots,A_k)$ is a k-plane, and $M_2 = M(Q;B_1,\ldots,B_m)$ is an m-plane, in V_n, and if $k \leq m$, we say M_1 is parallel to M_2 if the linear span of $A_1,\ldots,A_k$ is contained in the linear span of $B_1,\ldots,B_m$.

Exercises

1. Prove Theorems 17 and 18.

2. Prove Theorems 19 and 20.

3. Given the line $L = L(\underline{0};A)$ in V_3, where $A = (1,-1,2)$. Find parametric equations for a 2-plane containing the point $P = (1,1,1)$ that is parallel to L. Is it unique? Can you find such a plane containing both the point P and the point $Q = (-1,0,2)$?

4. Given the 2-plane M_1 in V_4 containing the points $P = (1,-1,2,-1)$ and $Q = (0,1,1,0)$ and $R = (1,1,0,3)$. Find parametric equations for a 3-plane in V_4 that contains the point $S = (1,1,1,1)$ and is parallel to M_1. Is it unique? Can you find such a 3-plane that contains both S and the point $T = (0,1,0,2)$?