4. Linear Subspaces

There are many subsets of $\mathbf{R}^n$ which mimic $\mathbf{R}^n$. For example, a plane L passing through the origin in $\mathbf{R}^3$ actually mimics $\mathbf{R}^2$ in many ways. First, L contains zero vector O as $\mathbf{R}^2$ does. Second, the sum of any two vectors in the plane L remains in the plane. Third, any scalar multiple of a vector in L remains in L. The plane L is an example of a linear subspace of $\mathbf{R}^3$.

4.1. Addition and scaling

Definition 4.1. *A subset V of $\mathbf{R}^n$ is called a linear subspace of $\mathbf{R}^n$ if V contains the zero vector O, and is closed under vector addition and scaling. That is, for $X, Y \in V$ and $c \in \mathbf{R}$, we have $X + Y \in V$ and $cX \in V$.*

What would be the smallest possible linear subspace V of $\mathbf{R}^n$? The singleton $V = \{O\}$ has all three properties that are required of a linear subspace. Thus it is a (and the smallest possible) linear subspace which we call *the zero subspace.* As a subspace, it shall be denoted as (O).

What would be a linear subspace V of "one size" up? There must be at least one additional (nonzero) vector A in V besides O. All its scalar multiples cA, $c \in \mathbf{R}$, must also be members of V. So, V must contain the line $\{cA|c \in \mathbf{R}\}$. Now note that this line does possess all three properties required of a linear subspace, hence this line is a linear subspace of $\mathbf{R}^n$. Thus we have shown that for any given *nonzero* vector $A \in \mathbf{R}^n$, the line $\{cA|c \in \mathbf{R}\}$ is a linear subspace of $\mathbf{R}^n$.

Exercise. Show that there is no subspace in between (O) and the line $V^1 := \{cA|c \in \mathbf{R}\}$. In other words, if V is a subspace such that $(O) \subset V \subset V^1$, then either $V = (O)$ or $V = V^1$.

How about subspaces of other larger "sizes" besides (0) and lines?

Example. *The span.* Let $S = \{A_1, .., A_k\}$ be a set of vectors in $\mathbf{R}^n$. Recall that a linear combination of S is a vector of the form

$$\sum_{i=1}^{k} x_i A_i = x_1 A_1 + \cdots + x_k A_k$$

where the x_i are numbers. The zero vector O is always a linear combination:

$$O = \sum_{i=1}^{k} 0 A_k.$$

Adding two linear combinations $\sum_i x_i A_i$ and $\sum y_i A_i$, we get

$$\sum x_i A_i + \sum y_i A_i = \sum (x_i + y_i) A_i,$$

which is also a linear combination. Here we have used the properties V1 and V3 for vectors in $\mathbf{R}^n$. Scaling the linear combination $\sum x_i A_i$ by a number c, we get

$$c \sum x_i A_i = \sum c x_i A_i,$$

which is also a linear combination. Here we have used properties V2 and V4 Thus we have shown that *the set of linear combinations* contains O, and is closed under vector addition and scaling, hence this set is a subspace of $\mathbf{R}^n$. This subspace is called *the span of* S, and is denoted by $Span(S)$. In this case, we say that V *is spanned by* S, or that S *spans* V.

Exercise. Is $(1, 0, 0)$ in the span of $\{(1, 1, 1), (1, 1, -1)\}$? How about $(1, 1, 0)$?

Exercise. Let $A_1, .., A_k \in \mathbf{R}^n$ be k column vectors. Show that the following are equivalent:

i. $V = Span\{A_1, ..., A_k\}$.

ii. $A_i \in V$ for all i and every vector B in V is a linear combination of the A_i.

iii. V is the image of the linear map $\mathbf{R}^k \to \mathbf{R}^n$, $X \mapsto AX$, where $A = [A_1, .., A_k]$.

Question. Let V be a linear subspace of $\mathbf{R}^n$. Is V spanned by some $A_1, ..., A_k \in V$? If so, what is the minimum k possible?

Next, let try to find linear subspaces of $\mathbf{R}^n$ from the opposite extreme: what is the largest possible subspace of $\mathbf{R}^n$? The set $\mathbf{R}^n$ is itself clearly the largest possible subset of $\mathbf{R}^n$ and it possesses all three required properties of a subspace. So, $V = \mathbf{R}^n$ is the largest possible subspace of $\mathbf{R}^n$. What would be a subspace "one size" down?

Let A be a nonzero vector in $\mathbf{R}^n$. Let $A^\perp$ denote the set of vectors X orthogonal to A, ie.

$$A^\perp = \{X \in \mathbf{R}^n | A \cdot X = 0\}.$$

This is called the *hyperplane* orthogonal to A. Since $A \cdot O = 0$, O is orthogonal to A. If X, Y are orthogonal to A, then

$$A \cdot (X + Y) = A \cdot X + A \cdot Y = O + O = O.$$

Hence $X + Y$ is orthogonal to A. Also if c is a number, then

$$A \cdot (cX) = c(A \cdot X) = 0.$$

Hence cX is also orthogonal to A. Thus $A^\perp$ contains the zero vector O, and is closed under vector addition and scaling. So $A^\perp$ is a linear subspace of $\mathbf{R}^n$.

Exercise. Let $S = \{A_1, .., A_m\}$ be vectors in $\mathbf{R}^n$. Let $S^\perp$ be the set of vectors X orthogonal to all $A_1, .., A_m$. The set $S^\perp$ is called the *orthogonal complement* of S. Verify that $S^\perp$ is a linear subspace of $\mathbf{R}^n$. Show that if $m < n$ then $S^\perp$ contains a nonzero vector. (Hint: Theorem 1.11.)

Exercise. Is $(1, 0, 0)$ in the orthogonal complement of $\{(0, 1, 1),\ (1, 0, 1)\}$? How about $(1, 1, -1)$? How about $(t, t, -t)$ for any scalar t? Are there any others?

Question. Let V be a linear subspace of $\mathbf{R}^n$. Are there vectors $S = \{A_1, .., A_k\}$ such that $V = S^\perp$? If so, what is the minimum k possible? Note that this question amounts to finding a linear system with the prescribed solution set V.

The two questions we posed above will be answered later in this chapter.

4.2. Matrices and linear subspaces

Recall that a homogeneous linear system of m equations in n variables can be written in the form (chapter 3):

$$AX = O$$

where $A = (a_{ij})$ is a given $m \times n$ matrix, and X is the column vector with the variable entries $x_1, ..., x_n$.

Definition 4.2. *We denote by $Null(A)$ (the null space of A) the set of solutions to the homogeneous linear system $AX = O$. We denote by $Row(A)$ (the row space of A) the set of linear combinations of the rows of A. We denote by $Col(A)$ (the column space of A) the set of linear combinations of the columns of A.*

Theorem 4.3. *Let A be an $m \times n$ matrix. Then both $Null(A)$, $Row(A)$ are linear subspaces of $\mathbf{R}^n$, and $Col(A)$ is a linear subspace of $\mathbf{R}^m$.*

Proof: Obviously $AO = O$. Thus $Null(A)$ contains the zero vector O. Let X, Y be elements of $Null(A)$, ie.

$$AX = O, \quad AY = O.$$

Then by Theorem 3.1

$$A(X + Y) = AX + AY = O + O = O.$$

Thus $Null(A)$ is closed under vector addition. Similarly, if c is a scalar then

$$A(cX) = c(AX) = cO = O.$$

Thus $Null(A)$ is closed under scaling.

Note that $Row(A) = Span(S)$ where S is the set of row vectors of A. We saw earlier that the span of any set of vectors in $\mathbf{R}^n$ is a linear subspace of $\mathbf{R}^n$.

Finally, observe that $Col(A) = Row(A^t)$, which is a linear subspace of $\mathbf{R}^m$. $\square$

Exercise. Let $L : \mathbf{R}^n \to \mathbf{R}^m$ be a linear map, represented by the matrix A. Show that the image of L is $Col(A)$. Show that L is one-to-one iff $Null(A) = \{O\}$.

4.3. Linear independence

Definition 4.4. *Let $\{A_1, .., A_k\}$ be a set of vectors in $\mathbf{R}^n$. A list of numbers $\{x_1, .., x_k\}$ is called a linear relation of $\{A_1, .., A_k\}$ if*

$$(*) \qquad x_1A_1 + \cdots + x_kA_k = O$$

holds. Abusing terminology, we often call $(*)$ *a linear relation.*

Example. Given any set of vectors $\{A_1, .., A_k\}$, there is always a linear relation $\{0, .., 0\}$, since

$$0A_1 + \cdots + 0A_k = O.$$

This is called the *trivial relation.*

Example. The set $\{(1,1), (1,-1), (1,0)\}$ has a nontrivial linear relation $\{1, 1, -2\}$:

$$(1,1) + (1,-1) - 2(1,0) = O.$$

Exercise. Find a nontrivial linear relation of the set $\{(1,1,-2), (1,-2,1), (-2,1,1)\}$.

Exercise. Find all the linear relations of $\{(1,1), (1,-1)\}$.

Definition 4.5. *A set* $\{A_1, .., A_k\}$ *of vectors in* $\mathbf{R}^n$ *is said to be linearly dependent if it has a nontrivial linear relation. The set is said to be linearly independent if it has no nontrivial linear relation.*

Exercise. Is $\{(1,1),\ (1,-1)\}$ linearly independent?

Exercise. Is $\{(1,-1),\ (\pi,-\pi)\}$ linearly independent?

Exercise. Is $\{(1,1),\ (1,-1),\ (1,2)\}$ linearly independent?

Example. Consider the set $\{E_1, .., E_n\}$ of standard unit vectors in $\mathbf{R}^n$. What are the linear relations for this set? Let

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

The vector on the left hand side has entries $(x_1, .., x_n)$. So this equation says that $x_1 = \cdots = x_n = 0$. Thus the set $\{E_1, .., E_n\}$ has only one linear relation – the trivial relation. So this set is linearly independent.

Exercise. Write a linear relation of $\{(1,1),\ (1,-1),\ (1,2)\}$ as a system of 2 equations.

Exercise. Let $\{A_1, A_2, .., A_k\}$ be a set of vectors in $\mathbf{R}^2$. Write a linear relation

$$x_1A_1 + \cdots + x_kA_k = O$$

as a system of 2 equations in k variables. More generally, let $\{A_1, A_2, .., A_k\}$ be a set of vectors in $\mathbf{R}^n$. Then a linear relation can be written as

$$AX = O$$

where A is the $n \times k$ matrix with columns $A_1, ..., A_k$, and X is the column vector in $\mathbf{R}^k$ with entries $x_1, .., x_k$. Thus *a linear relation can be thought of as a solution to a linear system.*

Theorem 4.6. *A set $\{A_1, .., A_k\}$ of more than n vectors in $\mathbf{R}^n$ is linearly dependent.*

Proof: As just mentioned, finding a linear relation for $\{A_1, .., A_k\}$ means solving the linear system

$$AX = O$$

of n equations in k variables. Thus if $k > n$, then there is a nontrivial solution to the linear system by Theorem 1.11. $\square$

4.4. Bases and dimension

Definition 4.7. *Let V be a linear subspace of $\mathbf{R}^n$. A set $\{A_1, .., A_k\}$ of vectors in V is called a basis of V if the set is linearly independent and it spans V. In this case, we say that V is k-dimensional. By definition, if $V = \{O\}$ then the empty set is the basis of V.*

Example. Every vector X in $\mathbf{R}^n$ is a linear combination of the set $\{E_1, .., E_n\}$. Thus this set spans $\mathbf{R}^n$. We have also seen that this set is linearly independent. Thus this set is a basis of $\mathbf{R}^n$. It is called the standard basis of $\mathbf{R}^n$. Thus $\mathbf{R}^n$ is n-dimensional, by definition.

Exercise. Let $\{A_1, .., A_k\}$ be a linearly independent set of vectors in $\mathbf{R}^n$. Prove that if B is not a linear combination of $\{A_1, .., A_k\}$, then $\{A_1, .., A_k, B\}$ is linearly independent.

Theorem 4.8. *Every linear subspace V of $\mathbf{R}^n$ has a basis.*

Proof: If $V = \{O\}$, then there is nothing to prove. So let's begin with a nonzero vector A_1 in V. The set $\{A_1\}$ is linearly independent. If this set spans V, then it is a basis of V.

If it doesn't span V, then there is some vector A_2 in V but not in $Span\{A_1\}$, so that the set $\{A_1, A_2\}$ is linearly independent (preceding exercise). Continue this way by adjoining more vectors $A_3, .., A_k$ in V if possible, while maintaining that $\{A_1, A_2, .., A_k\}$ is linearly independent. This process terminates when $\{A_1, A_2, .., A_k\}$ spans V. If this process were to continue indefinitely, then we would be able to find a linearly independent set with more than n vectors in $\mathbf{R}^n$, contradicting Theorem 4.6. So this process must terminate. □

The argument actually proves much more than the theorem asserts. It shows that given an independent set J of vectors in V, we can always grow J into a basis – by appending one vector at a time from V. More generally, it shows that if S is a set spanning V and if $J \subset S$ is an independent subset, then we can also grow J into a basis by appending vectors from S. To summarize, we state

Theorem 4.9. *(Basis Theorem) Let V be a linear subspace of $\mathbf{R}^n$ spanned by a set S, and $J \subset S$ is an independent subset. Then there is a basis of V that contains J.*

Example. The proof above actually tells us a way to find a basis of V, namely, by stepwise enlarging a linearly independent set using vectors in V, until the set is big enough to span V. Let's find a basis of $\mathbf{R}^2$ by starting from $(1, 1)$. Now take, say $(1, 0)$, which is not in $Span\{(1, 1)\}$. So we get $\{(1, 1), (1, 0)\}$. By Theorem 4.6, we need no more vectors. So $\{(1, 1), (1, 0)\}$ is a basis of $\mathbf{R}^2$.

Exercise. Find a basis of $\mathbf{R}^3$ by starting from $(1, 1, 1)$.

Exercise. Express $(1, 0, 1)$ as a linear combination of the basis you have just found. How many ways can you do it?

Exercise. Find a basis of solutions to the linear system:

$$\begin{array}{ccccccccccc} x_1 & + & & - & x_3 & + & x_4 & + & x_5 & = & 0 \\ & & x_2 & + & x_3 & - & x_4 & + & x_5 & = & 0. \end{array}$$

What is the dimension of your space of solutions? Express the solution $(-2, 0, 0, 1, 1)$ as a linear combination of your basis.

For the rest of this section, V will be a linear subspace of $\mathbf{R}^n$.

Theorem 4.10. *(Uniqueness of Coefficients) Let $\{A_1, .., A_k\}$ be basis of V. Then every vector in V can be expressed as a linear combination of the basis in just one way.*

Proof: Let B be a vector in V, and $\sum_i x_i A_i = B = \sum_i y_i A_i$ be two ways to express B as linear combination of the basis. Then we have

$$O = \sum_i x_i A_i - \sum_i y_i A_i = \sum_i (x_i - y_i) A_i.$$

Since the A's are linearly independent, we have $x_i - y_i = 0$ for all i. Thus $x_i = y_i$ for all i. $\Box$

Theorem 4.11. *(Dimension Theorem) Suppose V is a linear subspace of $\mathbf{R}^n$ that has a basis of k vectors. Then the following holds:*

(a) Any set of more than k vectors in V is linearly dependent.

(b) Any set of k linearly independent vectors in V is a basis of V.

(c) Any set of less than k vectors in V does not span V.

(d) Any set of k vectors which spans V is a basis of V.

Therefore, any two bases of V have the same number of vectors.

Proof: By assumption, V has a basis $A_1, .., A_k$, which we regard as column vectors in $\mathbf{R}^n$. Let $B_1, .., B_m$ be a given list of m vectors in V.

Part (a). Put $A = [A_1, .., A_k]$. Since $A_1, ..., A_k$ span V, each B_i is a linear combination of $A_1, ..., A_k$, which means that

$$B_1 = AC_1, ..., B_m = AC_m$$

where $C_1, .., C_m$ are m column vectors in $\mathbf{R}^k$. We can write

$$B = AC$$

where $B = [B_1, .., B_m]$ and $C = [C_1, .., C_m]$. If $m > k$, then the system $CX = O$ has more variables then equations, hence it has a nontrivial solution. In this case, $BX = ACX = O$ has a nontrivial solution, implying that $B_1, .., B_m$ are dependent.

Part (b). Suppose $m = k$ and $B_1, .., B_k$ are independent. We want to show that it spans V. If $B_1, .., B_k$ do not span V, then there would be a vector B in V, which is not a linear combination of $B_1, .., B_k$, and hence $B_1, .., B_k, B$ would be $k+1$ independent vectors in V. This would contradict (a).

Part (c). Suppose $m < k$. If $\{B_1, .., B_m\}$ spans V, then it would have a subset which is a basis of V by the Basis Theorem. This basis would have some $p < k$ elements. By (a), the set $\{A_1, .., A_k\}$ would be dependent because $k > p$. This is a contradiction. Thus the set $\{B_1, ..., B_m\}$ cannot span V.

Part (d). Let $\{B_1, .., B_k\}$ be a set which spans V. We want to show that it is linearly independent. By the Basis Theorem, it has a subset S which is a basis of V. Since S spans V, it cannot have less than k elements, by (c). Thus S is all of $\{B_1, .., B_k\}$. $\square$

Definition 4.12. *(Dimension) If V has a basis of k vectors, we say that V is k-dimensional, and we write $dim(V) = k$.*

Corollary 4.13. *Let W, V be linear subspaces of $\mathbf{R}^n$ such that $W \subset V$. Then $dim(W) \leq dim(V)$. If we also have $dim(W) = dim(V)$, then $W = V$.*

Proof: Put $k = dim(V)$, $l = dim(W)$, and let $B_1, .., B_l$ form a basis of W. Since the B's are also independent vectors in V, $l \leq k$ by the Dimension Theorem (a).

Next, suppose $W \subsetneq V$. We will show that $l < k$, proving our second assertion. Since $Span\{B_1, .., B_l\} = W \subsetneq V$, we can find $B \in V$ such that $B_1, .., B_l, B$ are independent. It follows that $l + 1 \leq k$, by the Dimension Theorem (a) again. So, $l < k$. $\square$

Exercise. Prove that if a subspace V of $\mathbf{R}^n$ has $dim(V) = n$, then $V = \mathbf{R}^n$.

Exercise. Let $L : \mathbf{R}^n \to \mathbf{R}^m$ be a linear map, represented by the matrix A. Show that L is onto iff $dim\ Col(A) = m$. (Hint: L is onto iff $Col(A) = \mathbf{R}^m$.)

Exercise. Now answer a question we posed earlier. Let V be a linear subspace of $\mathbf{R}^n$. Is V spanned by some $A_1, .., A_k \in V$? If so, what is the minimum k possible?

Exercise. Let $A_1, .., A_k$ be a basis of V. Let A be the matrix with columns $A_1, .., A_k$. Show that $Null(A) = \{O\}$, hence conclude that the linear map $L : \mathbf{R}^k \to \mathbf{R}^n$, $X \mapsto AX$, is one-to-one. Show that the image of the map is V. *The map L is called a parametric description of V.*

4.5. Matrices and bases

Theorem 4.14. *A square matrix is invertible iff its columns are independent.*

Proof: By Theorem 3.6, a square matrix A is invertible iff the linear system $AX = O$ has the unique solution $X = O$. Write $A = [A_1, .., A_n]$, $X = (x_1, .., x_n)^t$. Then $AX = O$ reads $x_1 A_1 + \cdots + x_n A_n = O$, a linear relation of the columns of A. So, A is invertible iff the columns of A has no nontrivial linear relation. □

Thus we have shown that the following are equivalent, for a square matrix A:

- A is invertible.
- the reduced row echelon of A is I.
- $AB = I$ for some matrix B.
- the columns of A are linearly independent.
- $Null(A)$ is the zero subspace.
- A^t is invertible.
- the rows of A are linearly independent.

Exercise. Is the set of row vectors in the matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

linearly independent? What is the dimension of $Row(A)$?

Theorem 4.15. *Let A be a row echelon. Then the set of nonzero row vectors in A is linearly independent.*

Proof: Let $A_1, .., A_k$ be the nonzero rows of A. Since A is a row echelon, the addresses of these rows are strictly increasing:

$$p_1 < p_2 < \cdots < p_k.$$

Consider a linear relation

$$(*) \qquad x_1 A_1 + \cdots + x_k A_k = O.$$

Call the left hand side Y. We will show that the x's are all zero.

Observe that the p_1th entry of Y is x_1 times the pivot in ${}_1A$. Thus (*) implies that $x_1 = 0$. Thus (*) becomes

$$x_2 A_2 + \cdots + x_k A_k = O.$$

Now repeat the same argument as before, we see that $x_2 = 0$. Continuing this way, we conclude that $x_1 = x_2 = \cdots = x_k = 0$. $\Box$

Corollary 4.16. *If A is a row echelon, then the set of nonzero row vectors in A form a basis of $Row(A)$.*

Proof: The subspace $Row(A)$ is spanned by the nonzero rows of A. By the preceding theorem, the nonzero rows form a basis of $Row(A)$. $\Box$

Theorem 4.17. *If A, B are row equivalent matrices, then*

$$Row(A) = Row(B)$$

ie. row operations do not change row space.

Proof: It suffices to show that if A transforms to B under a single row operation, then they have the same row space. Suppose A transforms to B under R1, R2, or R3. *We'll show that each row of B is in $Row(A)$.* For then it follows that $Row(B) \subset Row(A)$ (why?). The reverse inclusion is similar.

Since B is obtained from A under R1-R3, each row of B is one of the following:

(a) a row of A;

(b) a scalar multiple of a row of A;

(c) one row of A plus a scalar multiple of a another row of A.

Each of of these is a vector in $Row(A)$. Therefore each row of B is in $Row(A)$. $\square$

The converse is also true: if A and B have the same row space, then A, B are row equivalent (see Homework).

Exercise. Write down your favorite 3×4 matrix A and find its reduced row echelon B. Verify that the nonzero row vectors in B form a linearly independent set. Express every row vector in A as a linear combination of this set.

Corollary 4.18. *Suppose A, B are row equivalent matrices. Then the row vectors in A are linearly independent iff the row vectors in B are linearly independent.*

Proof: Let A, B be $m \times n$. Suppose the rows of A are linearly independent. Then they form a basis of $Row(A)$, so that $Row(A)$ is m-dimensional. By the preceding theorem $Row(B)$ is m-dimensional. Since $Row(B)$ is spanned by its m rows, these rows form a basis of $Row(B)$ by the Dimension Theorem (d). Hence the rows of B are linearly independent. The converse is similar. $\square$

Corollary 4.19. *Suppose A, B are row equivalent matrices, and that B is a row echelon. Then the rows of A are linearly dependent iff B has a zero row.*

Proof: If B has a zero row, then the rows of B are linearly dependent (why?). By the preceding corollary, it follows that the row vectors in A are also linearly dependent. Conversely, suppose B has no zero row. Then the row vectors of B are linearly independent by Theorem 4.15. It follows that the row vectors in A are also linearly independent by the preceding corollary. $\square$

Exercise. For the matrix A you wrote down in the previous exercise, decide if the row vectors are linearly independent.

Corollary 4.20. *Suppose A, B are row equivalent matrices, and that B is a row echelon. Then the nonzero row vectors of B form a basis of $Row(A)$.*

Proof: Since A, B are row equivalent,

$$Row(A) = Row(B)$$

by the preceding theorem. Since B is a row echelon, $Row(B)$ is spanned by the set of nonzero row vectors in B. It follows that $Row(A)$ is spanned by the same set. $\square$

Exercise. Suppose A is a 6×5 matrix and B is a row echelon of A with 5 nonzero rows. What is the dimension of $Row(A)$?

Given a set $S = \{A_1, .., A_k\}$ of row vectors in $\mathbf{R}^n$, the preceding corollary gives a procedure for determining whether S is linearly independent, and it finds a basis of the linear subspace $Span(S)$. We call this procedure the *basis test* in $\mathbf{R}^n$.

L1. Form a $k \times n$ matrix A with the vectors in S.

L2. Find a row echelon B of A.

L3. S is linearly independent iff B has no zero rows.

L4. The nonzero rows in B form a basis of $Span(S) = Row(A)$.

Exercise. Use the preceding corollary to find a basis of the subspace spanned by $(1, -1, 0, 0)$, $(1, 0, -1, 0)$, $(1, 0, 0, -1)$, $(0, 1, -1, 0)$, $(0, 1, 0, -1)$ in $\mathbf{R}^4$.

4.6. The rank of a matrix

Definition 4.21. *To every matrix A, we assign a number $rank(A)$, called the rank of A, defined as the dimension of $Col(A)$.*

Exercise. What is the rank of

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}?$$

Theorem 4.22. *If A is a reduced row echelon, then*

$$rank(A) = dim\ Row(A) = \#pivots.$$

Proof: If the entries of A are all zero, then there is nothing to prove. Suppose that A is $m \times n$, and that there are k nonzero row vectors in A, so that $\#pivots = k$. By Theorem 4.15, $dim\ Row(A) = k$. It remains to show that $rank(A) = k$.

Let $p_1, .., p_k$ be the addresses of the first k rows. Since A is reduced, the columns containing the pivots are the standard vectors $E_1, .., E_k$ in $\mathbf{R}^m$. Because rows $(k+1)$ to m are all zeros, entries $(k+1)$ to m of each column are all zeros. This means that each column of A is a linear combination of the set $S = \{E_1, .., E_k\}$. Hence S is a basis of $Col(A)$, and so $rank(A) = k$. $\square$

We have seen that $Row(A)$ remains unchanged by row operations on A. However, this is not so for $Col(A)$. For example,

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

are row equivalent. But $Col(A)$ is the line spanned by $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, while $Col(B)$ is the line spanned by $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Nevertheless, we have

Theorem 4.23. *If A, B are row equivalent matrices, then*

$$rank(A) = rank(B),$$

ie. row operations do not change the rank.

Proof: Suppose $A \sim B$. Then the two linear systems

$$AX = O, \quad BX = O$$

have the same solutions, by Theorem 1.10. In terms of the column vectors in A, B, the two systems read

$$x_1 A_1 + \cdots + x_n A_n = O, \quad x_1 B_1 + \cdots + x_n B_n = O.$$

This means that the column vectors in A and those in B have the same linear relations.

Let $r = rank(A)$. Since $Col(A)$ has dimension r and is spanned by the columns $A_1, .., A_n$ of A, by the Basis Theorem we can find a basis $\{A_{i_1}, .., A_{i_r}\}$ of $Col(A)$. Since the basis is independent, the only linear relation of the form

$$x_{i_1} A_{i_1} + \cdots + x_{i_r} A_{i_r} = O$$

is the trivial relation with $x_{i_1} = \cdots = x_{i_r} = 0$. This shows that the only linear relation

$$x_{i_1} B_{i_1} + \cdots + x_{i_r} B_{i_r} = O$$

is also the trivial relation, implying that the set of column vectors $\{B_{i_1}, .., B_{i_r}\}$ in B is linearly independent. We can grow it to a basis of $Col(B)$, by the Basis Theorem. This shows that $dim(Col(B)) = rank(B) \geq r$, i.e.

$$rank(B) \geq rank(A).$$

Now interchange the roles of A and B, we see that $rank(A) \geq rank(B)$. So we conclude that $rank(A) = rank(B)$. $\square$

Corollary 4.24. *For any matrix A, $rank(A^t) = rank(A)$.*

Proof: By the preceding theorem $rank(A)$ is unchanged under row operations. Since $Row(A)$ is unchanged under row operations, so is $rank(A^t) = dim\ Row(A)$. By Theorem 4.22, the asserted equality $rank(A^t) = rank(A)$ holds if A is a reduced row echelon. It follows that the same equality holds for an arbitrary A. $\square$

The preceding theorem also gives us a way to find the rank of a given matrix A. Namely, *find the reduced row echelon of A, and then read off the number of pivots.* (Compare this with the basis test.)

Exercise. Find the rank of

$$\begin{bmatrix} 1 & 1 & -1 & -1 & 0 \\ -1 & 1 & -1 & 0 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ 1 & 0 & -1 & -1 & 1 \end{bmatrix}$$

by first finding its reduced row echelon. Also use the reduced row echelon to find $dim\ Null(A)$.

Exercise. Pick your favorite 4×5 matrix A. Find $rank(A)$ and $dim\ Null(A)$.

Exercise. What is $rank(A) + dim\ Null(A)$ in each of the two exercises above?

Given an $m \times n$ matrix A, we know that row operations do not change the subspace $Null(A)$. In particular the number $dim\ Null(A)$ is unchanged. The preceding theorem says that the number $rank(A)$ is unchanged either. In particular, the sum $rank(A) + dim\ Null(A)$ remains unchanged. Each summand depends, of course on the matrix A. But remarkably, the sum depends only on *the size* of A!

Theorem 4.25. *(Rank-nullity Relation) For any $m \times n$ matrix A,*

$$rank(A) + dim\ Null(A) = n.$$

Proof: Put $r = rank(A)$. Then $Col(A)$ is a subspace of $\mathbf{R}^m$ of dimension r. By the Basis Theorem, there are r columns of A that form a basis of $Col(A)$. Let $i_1 < \cdots < i_r$ be the position of those columns. Let $B_1, .., B_k$ be a basis of $Null(A)$. We will prove that the standard unit vectors $E_{i_1}, .., E_{i_r}$ in $\mathbf{R}^n$ together with $B_1, .., B_k$ form a basis of $\mathbf{R}^n$, so that $r + k = n$, as desired.

First, we check that the vectors $E_{i_1}, .., E_{i_r}, B_1, .., B_k$ are independent. Consider a linear relation

$$x_1 E_{i_1} + \cdots + x_r E_{i_r} + y_1 B_1 + \cdots + y_k B_k = O.$$

Regard both sides as column vectors and multiply each with A. Since $AE_j = A_j$, the jth column of A, and since $AB_1 = \cdots = AB_k = O$, the result is $x_1 A_{i_1} + \cdots + x_r A_{i_r} = O$. Since $A_{i_1}, .., A_{i_r}$ are independent, $x_1 = \cdots = x_r = 0$. This leaves $y_1 B_1 + \cdots + y_k B_k = 0$. Since $B_1, .., B_k$ are independent, $y_1 = \cdots = y_k = 0$, too. So, the vectors $E_{i_1}, .., E_{i_r}, B_1, .., B_k$ are independent.

Next, we check that these vectors span $\mathbf{R}^n$. Given $X \in \mathbf{R}^n$, $AX \in Col(A)$. Since $A_{i_1}, .., A_{i_r}$ span $Col(A)$, it follows that

$$AX = z_1 A_{i_1} + \cdots + z_r A_{i_r}$$

for some $z_1, .., z_r \in \mathbf{R}$. The right side is equal to $A(z_1 E_{i_1} + \cdots + z_r E_{i_r})$. It follows that $X - (z_1 E_{i_1} + \cdots + z_r E_{i_r}) \in Null(A)$, hence this vector is a linear combination of $B_1, .., B_k$, i.e.

$$X - (z_1 E_{i_1} + \cdots + z_r E_{i_r}) = y_1 B_1 + \cdots + y_k B_k$$

implying that X is a linear combination of $E_{i_1}, .., E_{i_r}, B_1, .., B_k$. $\square$

Exercise. By inspection, find the dimension of the solution space to

$$x + 3y - z + w = 0$$

$$x - y + z - w = 0.$$

Exercise. What is the rank of an invertible $n \times n$ matrix?

Exercise. Suppose that A is an $m \times n$ reduced row echelon with k pivots. What is the dimension of the space of solutions to the linear system

$$AX = O?$$

How many free parameters are there in the general solutions?

Corollary 4.26. *Let A be $m \times n$ reduced row echelon of rank r. Then the $n - r$ basic solutions to $AX = O$ of Chapter 1 form a basis of $Null(A)$.*

Proof: Recall that r coincides the the number of pivots of A, whose addresses we denote by $p_1 < \cdots < p_r$. Solving $AX = O$, we find that the r "pivot" variables $x_{p_1}, .., x_{p_r}$ can be expressed uniquely in terms of the "nonpivot" variables x_i, $i \notin \{p_1, .., p_r\}$, and the general solution is a linear combination of $n - r$ basic solutions. In particular, these basic solutions span $Null(A)$. By the preceding theorem, $dim\ Null(A) = n - r$. By the Dimension Theorem, the basic solutions form a basis of $Null(A)$. $\square$

4.7. Orthogonal complement

In one of the previous exercises, we introduced the notion of the orthogonal complement $S^\perp$ of a finite set $S = \{A_1, .., A_k\}$ of vectors in $\mathbf{R}^n$. Namely, $S^\perp$ consists of vectors X which are orthogonal to all $A_1, .., A_k$. We have also posed the question: *Is every linear subspace of $\mathbf{R}^n$ the orthogonal complement of a finite set?* We now answer this question.

Throughout this section, let V be a linear subspace of $\mathbf{R}^n$. The *orthogonal complement* of V is defined to be

$$V^\perp = \{X \in \mathbf{R}^n | Y \cdot X = 0, \quad for\ all\ Y \in V\}.$$

Exercise. Verify that $V^\perp$ is a linear subspace of $\mathbf{R}^n$.

Theorem 4.27. *Let $B_1, ..., B_l$ be a basis of V, and let B be the $l \times n$ matrix whose rows are $B_1, .., B_l$. Then $V^\perp = Null(B)$.*

Proof: If $X \in V^\perp$, then $Y \cdot X = 0$ for all $Y \in V$. In particular, $B_i \cdot X = 0$ for $i = 1, .., l$, hence $BX = O$ and $X \in Null(B)$. Conversely, let $X \in Null(B)$. Then $B_i \cdot X = 0$ for $i = 1, .., l$. Given $Y \in V$, we can express it as a linear combination, say $Y = y_1 B_1 + \cdots + y_l B_l$. So,

$$Y \cdot X = (y_1 B_1 + \cdots + y_l B_l) \cdot X = 0.$$

This shows that $X \in V^\perp$. $\square$

Corollary 4.28. *$dim\ V + dim\ V^\perp = n$.*

Proof: Let $B_1, .., B_l$ be a basis of V, and$B$ be the $l \times n$ matrix whose *rows* are $B_1, .., B_l$. By the preceding theorem, $Null(B) = V^\perp$. But $rank(B) = l = dim\ V$. Now our assertion follows from the Rank-nullity Relation, applied to B. $\square$

Corollary 4.29. *If $V^\perp = \{O\}$, then $V = \mathbf{R}^n$.*

Proof: By the preceding corollary, $dim\ V = n = dim\ \mathbf{R}^n$. It follows that $V = \mathbf{R}^n$, by a corollary to the Dimension Theorem. □

Exercise. Show that for any linear subspace V of $\mathbf{R}^n$, $V \subset (V^\perp)^\perp$ and $V \cap V^\perp = \{O\}$.

Corollary 4.30. $(V^\perp)^\perp = V$.

Proof: Since the relation $dim\ V + dim\ V^\perp = n$ holds for any subspace V of $\mathbf{R}^n$, we can apply it to $V^\perp$ as well, i.e. $dim\ V^\perp + dim\ (V^\perp)^\perp = n$. This implies that

$$dim\ (V^\perp)^\perp = dim\ V.$$

Since $V \subset (V^\perp)^\perp$, it follows that $V = (V^\perp)^\perp$, by a corollary to the Dimension Theorem. □

Corollary 4.31. *Every vector $X \in \mathbf{R}^n$ can be uniquely expressed as $X = A + B$, where $A \in V$ and $B \in V^\perp$.*

Proof: We prove uniqueness first. If $A, A' \in V$ and $B, B' \in V^\perp$ and $X = A + B = A' + B'$, then

$$A - A' = B' - B.$$

The left side is in V and the right side is in $V^\perp$. Since $V \cap V^\perp\{O\}$, both sides are zero, hence $A = A'$ and $B = B'$.

Let $V + V^\perp$ be the set consisting of all vectors $A + B$ with $A \in V$, $B \in V^\perp$ (cf. section 4.9). This is a linear subspace of $\mathbf{R}^n$ (exercise). To complete the proof, we will show that $V + V^\perp = \mathbf{R}^n$. By a corollary to the Dimension Theorem, it suffices to prove that $\dim(V + V^\perp) = n$. Let $A_1, .., A_k$ form a basis of V, and $B_1, .., B_l$ form a basis of $V^\perp$. By corollary above, $l + k = n$. Note that $A_1, .., A_k, B_1, .., B_l$ span $V + V^\perp$. It remains to show that they are independent. Let $A = x_1A_1 + \cdots + x_kA_k$ and $B = y_1B_1 + \cdots + y_lB_l$ with $x_j, y_j \in \mathbf{R}$, and assume that $A + B = O$. By the uniqueness argument, $A = B = O$. Since the A_j are independent, this implies that $x_j = 0$ for all j. Likewise $y_j = 0$ for all j. This completes the proof. □

Corollary 4.32. *For any given linear subspace V of $\mathbf{R}^n$, there is a $k \times n$ matrix A such that $V = Null(A)$. Moreover, the smallest possible value of k is $n - dim\ V$.*

Proof: Let $A_1, .., A_k$ be a basis of $V^\perp$. Note that $k = dim\ V^\perp = n - dim\ V$. We have

$$V = (V^\perp)^\perp = Null(A)$$

by Theorem 4.27 (applied to the subspace $V^\perp$.)

Let A' be an $l \times n$ matrix such that $V = Null(A')$. We will show that $l \geq k$. Applying the Rank-nullity Relation to A', we see that $rank(A') = n - dim\ V = k$. Since $rank(A') = dim\ Row(A')$, this implies that A' must have at least k rows, by the Dimension Theorem. So, $l \geq k$. $\square$

Exercise. Let V be the line spanned by $(1, 1, 1)$ in $\mathbf{R}^3$. Find a smallest matrix A such that $V = Null(A)$. Repeat this for $V = Span\{(1, 1, -1, -1), (1, -1, 1 - 1)\}$.

Corollary 4.33. *(Best Approximation) Let V be a linear subspace of $\mathbf{R}^n$. For any given $B \in \mathbf{R}^n$, there is a unique point $C \in V$ such that*

$$\|B - C\| < \|B - D\|$$

for each $D \in V$ not equal to C. Moreover $B - C \in V^\perp$. The point C is called the projection of B along V.

Proof: By a corollary above, B can be uniquely expressed as

$$B = C + C'$$

where $C \in V$ and $C' \in V^\perp$. Let $D \in V$. Since $C - D \in V$ and $C' = B - C \in V^\perp$, it follows by Pythagoras that,

$$\|B - D\|^2 = \|(B - C) + (C - D)\|^2 = \|B - C\|^2 + \|C - D\|^2.$$

For $D \neq C$, it follows that

$$\|B - D\|^2 > \|B - C\|^2.$$

Taking square root yields our asserted inequality.

Next, we show uniqueness: there is no more than one point $C \in V$ with the minimizing property that

$$\|B - C\| < \|B - D\|$$

for each $D \in V$ not equal to C. Suppose C_1, C_2 are two such points. If they are not equal, then their minimizing property implies that

$$\|B - C_1\| < \|B - C_2\| \quad \& \quad \|B - C_2\| < \|B - C_1\|$$

which is absurd. $\square$

4.8. Coordinates and change of basis

Throughout this section, V will be a k dimensional linear subspace of $\mathbf{R}^n$.

Let $\{A_1, .., A_k\}$ be a basis of V. Then any given vector X in V can be expressed as a linear combination of this basis in just one way, by Theorem 4.10:

$$X = y_1 A_1 + \cdots + y_k A_k.$$

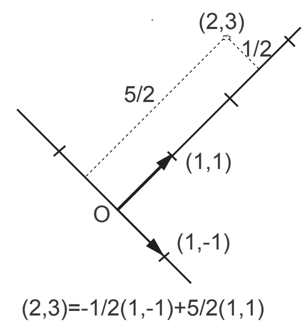
Definition 4.34. *The scalar coefficients $(y_1, .., y_k)$ above are called the coordinates of X relative to the basis $\{A_1, .., A_k\}$.*

Example. The coordinates of a vector $X = (x_1, .., x_n)$ relative to the standard basis $\{E_1, .., E_n\}$ of $\mathbf{R}^n$ are $(x_1, .., x_n)$ since

$$X = x_1 E_1 + \cdots + x_n E_n.$$

These coordinates are called the *Cartesian coordinates.*

Example. The following picture depicts the coordinates of $(2, 3)$ relative to the basis $\{(1, -1),\ (1, 1)\}$ of $V = \mathbf{R}^2$.



Exercise. Find the coordinates of $(2,3)$ relative to the basis $\{(1,2),\ (2,1)\}$ of $V = \mathbf{R}^2$.

Exercise. *Coordinates depend on the order of the basis vectors.* Find the coordinates of $(2,3)$ relative to the basis $\{(2,1),\ (1,2)\}$ of $V = \mathbf{R}^2$.

Exercise. Verify that $X = (1,1,-2)$ lies in $V = Span(S)$ where $S = \{(1,0,-1),\ (1,-1,0)\}$. What are the coordinates of X relative to the basis S?

Let $P = \{A_1, .., A_k\}$ and $Q = \{B_1, .., B_k\}$ be two bases of V. We shall regard vectors as *column* vectors, and put

$$A = [A_1, .., A_k], \quad B = [B_1, .., B_k]$$

which are $n \times k$ matrices. Each B_i is a unique linear combination of the first basis. So, there is a unique $k \times k$ matrix T such that

$$B = AT.$$

T is called the *transition matrix* from P to Q. Similarly, each A_i is a unique linear combination of the second basis Q, and so we have a transition matrix T' from Q to P:

$$A = BT'.$$

Theorem 4.35. *$TT' = I$.*

Proof: We have

$$A = BT' = ATT'.$$

So $A(TT' - I)$ is the zero matrix, where I is the $k \times k$ identity matrix. This means that each column of the matrix $TT' - I$ is a solution to the linear system $AX = O$. But the columns of A are independent. So, $AX = O$ has no nontrivial solution, implying that each column of $TT' - I$ is zero. Thus, $TT' = I$. $\square$

Exercise. Find the transition matrix from the standard basis $P = \{E_1, E_2, E_3\}$ to the basis $Q = \{(1, 1, 1),\ (1, -1, 0),\ (1, 0, -1)\}$. of $\mathbf{R}^3$ (regarded as column vectors.) Find the transition matrix from Q to P.

Theorem 4.36. *Let $\{A_1, ..., A_k\}$ be a linearly independent set in $\mathbf{R}^n$, and U be an $n \times n$ invertible matrix. Then $\{UA_1, ..., UA_k\}$ is linearly independent.*

Proof: Put $A = [A_1, .., A_k]$ and consider the linear relation

$$UAX = O$$

of the vectors $UA_1, .., UA_k$. Since U is invertible, the linear relation becomes $AX = O$. Since the columns of A are independent, $X = O$. It follows that $UA_1, .., UA_k$ have no nontrivial linear relation. $\square$

Corollary 4.37. *If $\{A_1, ..., A_n\}$ is a basis of $\mathbf{R}^n$, and U an $n \times n$ invertible matrix, then $\{UA_1, ..., UA_n\}$ is also a basis of $\mathbf{R}^n$.*

Proof: By the preceding theorem, the set $\{UA_1, ..., UA_n\}$ is linearly independent. By (corollary to) the Dimension theorem, this is a basis of $\mathbf{R}^n$. $\square$

Theorem 4.38. *The transition matrix from one basis $\{A_1, ..., A_n\}$ of $\mathbf{R}^n$ to another basis $\{B_1, ..., B_n\}$ is $A^{-1}B$ where $A = [A_1, ..., A_n]$ and $B = [B_1, ..., B_n]$.*

Proof: If T is the transition matrix, then $AT = B$. Since the columns of A are independent, A is invertible, by Theorem 4.14. It follows that $T = A^{-1}B$. $\square$

4.9. Sums and direct sums

In this section, we generalize results of an earlier section involving the pair of subspaces $V, V^{\perp}$ of $\mathbf{R}^n$.

Definition 4.39. *Let U, V be linear subspaces of $\mathbf{R}^n$. The sum of U and V is defined to be the set*

$$U + V = \{A + B | A \in U,\ \ B \in V\}.$$

Exercise. Verify that $U + V$ is a linear subspace of $\mathbf{R}^n$.

Theorem 4.40. *Let U, V be subspaces of $\mathbf{R}^n$. Suppose that U is spanned by $A_1, .., A_k$ and that V is spanned by $B_1, .., B_l$. Then $U + V$ is spanned by $A_1, .., A_k, B_1, .., B_l$.*

Proof: A vector in $U + V$ has the form $A + B$ with $A \in U$ and $B \in V$. By supposition, A, B have the forms

$$A = \sum_i x_i A_i, \qquad B = \sum_j y_j B_j.$$

So

$$A + B = \sum_i x_i A_i + \sum_j y_j B_j$$

which is a linear combination of $\{A_1, .., A_k, B_1, .., B_l\}$. Thus we have shown that every vector in $U + V$ is a linear combination of the set $\{A_1, .., A_k, B_1, .., B_l\}$. $\square$

Theorem 4.41. *Let U, V be subspaces of $\mathbf{R}^n$. Then*

$$dim(U + V) = dim(U) + dim(V) - dim(U \cap V).$$

Proof: Let $A_1, .., A_k$ form a basis of U, and $B_1, .., B_l$ a basis of V, and consider the $n \times k, n \times l$ matrices:

$$A = [A_1, .., A_k], \quad B = [B_1, .., B_l].$$

A vector in $\mathbf{R}^{k+l}$ can be written in the form of a column $\begin{bmatrix} X \\ Y \end{bmatrix}$ where $X \in \mathbf{R}^k$, $Y \in \mathbf{R}^l$. Define the map

$$L : \mathbf{R}^{k+l} \to \mathbf{R}^n, \qquad \begin{bmatrix} X \\ Y \end{bmatrix} \mapsto AX + BY$$

which is clearly linear (and is represented by the matrix $[A|B]$). By the preceding theorem $Col([A|B]) = U + V$. By the Rank-nullity relation,

$$dim(U + V) = rank(C) = k + l - dim(Null(C)).$$

Thus it suffices to show that $m := dim(Null(C)) = dim(U \cap V)$. Let $P_1, ..., P_m$ be a basis of $Null(C)$. To complete the proof, we will construct a basis of $U \cap V$ with m elements. Each P_i is a column vector of the form

$$P_i = \begin{bmatrix} X_i \\ Y_i \end{bmatrix}$$

where $X_i \in \mathbf{R}^k$ and $Y_i \in \mathbf{R}^l$. Since $O = CP_i = AX_i + BY_i$, we have $AX_i = -BY_i = U \cap V$ for all i. The AX_i are independent. For if $O = \sum_i z_i AX_i = A \sum_i z_i X_i$ then $\sum_i z_i X_i = 0$, since the columns of A are independent. Likewise $\sum_i z_i Y_i = 0$, hence $\sum_i z_i P_i = O$ implying that $z_i = 0$ for all i since the P_i are independent. Thus $AX_1, .., AX_m$ are independent. The AX_i also span $U \cap V$. For if $Z \in U \cap V$, then there exist (unique) $X \in \mathbf{R}^k$ and $Y \in \mathbf{R}^l$ such that $Z = AX = -BY$ since the columns of A span U (likewise for V), implying that $O = CP = AX + BY$ where $P = \begin{bmatrix} X \\ Y \end{bmatrix}$. Thus $P \in Null(C)$, so that $P = \sum_i z_i P_i$ for some $z_i \in \mathbf{R}$, implying that $X = \sum_i z_i X_i$ and $Y = \sum_i z_i Y_i$. It follows that

$$Z = AX = \sum_i z_i AX_i.$$

Thus we have shown that the $AX_1, .., AX_m$ form a basis of $U \cap V$. □

Corollary 4.42. *Let $C = [A|B]$ and let $P_i = \begin{bmatrix} X_i \\ Y_i \end{bmatrix}$ $(1 \le i \le m)$ form a basis of $Null(C)$ as in the preceding proof. Then $AX_1, .., AX_m$ form a basis of $U \cap V$.*

Corollary 4.43. *Let U, V be subspaces of $\mathbf{R}^n$. Suppose that U has basis $A_1, .., A_k$ and that V has basis $B_1, .., B_l$. If $U \cap V = \{O\}$, then $U + V$ has basis $A_1, .., A_k, B_1, .., B_l$. The converse is also true.*

Proof: Suppose $U \cap V = \{O\}$. By the preceding two theorems, $U + V$ is spanned by $A_1, .., A_k, B_1, .., B_l$, and has dimension $k + l$. By the Dimension Theorem, those vectors form a basis of $U + V$. Conversely suppose those vectors form a basis of $U + V$. Then the preceding theorem implies that $dim(U \cap V) = 0$, i.e. $U \cap V = \{O\}$. □

Exercise. Let U be the span of $\{(1,1,-1,-1),\ (1,0,-1,0)\ ,(-1,1,1,-1)\}$, and V be the span of $\{(0,1,0,-1),\ (1,-1,1,-1)\}$. Find a basis of $U+V$. Do the same for $U \cap V$.

Exercise. Let U be the span of $\{(1,1,-1,-1),\ (1,0,-1,0)\}$, and V be the span of $\{(1,-1,1,-1)\}$. What is $U \cap V$? What is $dim(U+V)$?

Definition 4.44. *Let U, V be subspaces of $\mathbf{R}^n$. If $U \cap V = \{O\}$, we call $U+V$ the direct sum of U and V.*

Theorem 4.45. *Let U, V be subspaces of $\mathbf{R}^n$. Then the following are equivalent:*

(a) (Zero overlap) $U \cap V = \{O\}$.

(b) (Independence) If $A \in U$, $B \in V$ and $A+B=O$, then $A=B=O$.

(c) (Unique decomposition) Every vector $C \in U+V$ can be written uniquely as $A+B$ with $A \in U$, $B \in V$.

(d) (Dimension additivity) $dim(U+V) = dim(U) + dim(V)$

Proof: Assume (a), and let $A \in U$, $B \in V$ and $A+B=O$. Then $A=-B \in U \cap V$, hence $A=-B=O$ by (a), proving (b). Thus (a) implies (b).

Assume (b), and let $C \in U+V$. Then $C=A+B$ for some $A \in U$ and $B \in V$, by definition. To show (c), we must show that A, B are uniquely determined by C. Thus let $C=A'+B'$ where $A' \in U$ and $B' \in V$. Then $A+B=C=A'+B'$, implying that $A-A'=B'-B \in U \cap V$, hence $A-A'=B'-B=O$ by (a), proving that (c) holds. Thus (b) implies (c).

Assume (c), and let $C \in U \cap V$. Then $C=2C-C=3C-2C$ with $2C, 3C \in U$ and $-C, -2C \in V$. It follows that $2C=3C$ (and $-C=-2C$) by (c), implying that $C=O$, proving that (a) holds. Thus (c) implies (a).

Finally, the preceding theorem implies that (a) and (d) are also equivalent. $\square$

Definition 4.46. *Let $V_1, .., V_k$ be subspaces of $\mathbf{R}^n$. Define their sum to be the subspace (verify it is indeed a subspace!)*

$$\sum_{i=1}^{k} V_i = V_1 + \cdots + V_k = \{A_1 + \cdots + A_k | A_i \in V_i,\ i = 1, .., k\}.$$

We say that this sum is a direct sum if $A_i \in V_i$ for $i = 1, .., k$ and $A_1 + \cdots + A_k = O$ implies that $A_i = O$ for all i.

Note that if $k \geq 2$, then

$$V_1 + \cdots + V_k = (V_1 + \cdots + V_{k-1}) + V_k.$$

Exercise. Show that if $dim(\sum_{i=1}^k V_i) \leq \sum_{i=1}^k dim(V_i)$. Moreover equality holds iff $\sum_{i=1}^k V_i$ is a direct sum.

4.10. Orthonormal bases

In chapter 2, we saw that when a set $\{A_1, .., A_k\}$ is orthonormal, then a vector B which is a linear combination of this set has a nice universal expression

$$B = \sum (B \cdot A_i) A_i.$$

So if $\{A_1, .., A_n\}$ is orthonormal and a basis of $\mathbf{R}^n$, then *every* vector in $\mathbf{R}^n$ has a similar expression in terms of that basis. In this section, we will develop an algorithm to find an orthogonal basis, starting from a given basis. This algorithm is known as the *Gram-Schmidt orthogonalization process*. Note that to get an orthonormal basis from an orthogonal basis, it is enough to normalize each of the basis vectors to length one.

Let $\{A_1, .., A_n\}$ be a given basis of $\mathbf{R}^n$. Put

$$A_1' = A_1.$$

It is nonzero, so that the set $\{A_1'\}$ is linearly independent.

We adjust A_2 so that we get a new vector A_2' which is nonzero and orthogonal to A_1'. More precisely, let $A_2' = A_2 - cA_1'$ and demand that $A_2' \cdot A_1' = 0$. This gives $c = \frac{A_2 \cdot A_1'}{A_1' \cdot A_1'}$. Thus we put

$$A_2' = A_2 - \frac{A_2 \cdot A_1'}{A_1' \cdot A_1'} A_1'.$$

Note that A_2' is nonzero, for otherwise A_2 would be a multiple of $A_1' = A_1$. So, we get an orthogonal set $\{A_1', A_2'\}$ of nonzero vectors.

We adjust A_3 so that we get a new vector A_3' which is nonzero and orthogonal to A_1', A_2'. More precisely, let $A_3' = A_3 - c_2 A_2' - c_1 A_1'$ and demand that $A_3' \cdot A_1' = A_3' \cdot A_2' = 0$. This gives $c_1 = \frac{A_3 \cdot A_1'}{A_1' \cdot A_1'}$ and $c_2 = \frac{A_3 \cdot A_2'}{A_2' \cdot A_2'}$. Thus we put

$$A_3' = A_3 - \frac{A_3 \cdot A_2'}{A_2' \cdot A_2'} A_2' - \frac{A_3 \cdot A_1'}{A_1' \cdot A_1'} A_1'.$$

Note that A_3' is also nonzero, for otherwise A_3 would be a linear combination of A_1', A_2'. This would mean that A_3 is a linear combination of A_1, A_2, contradicting linear independence of $\{A_1, A_2, A_3\}$.

More generally, we put

$$A_k' = A_k - \sum_{i=1}^{k-1} \frac{A_k \cdot A_i'}{A_i' \cdot A_i'} A_i'$$

for $k = 1, 2, .., n$. Then A_k' is nonzero and is orthogonal to $A_1', .., A_{k-1}'$, for each k. Thus the end result of Gram-Schmidt is an orthogonal set $\{A_1', .., A_n'\}$ of nonzero vectors in $\mathbf{R}^n$.

Note that A_k' is a linear combination of $\{A_1, .., A_k\}$. Thus $\{A_1', .., A_k'\}$ is a linearly independent set in $Span\{A_1, .., A_k\}$, which has dimension k. It follows that $\{A_1', .., A_k'\}$ is also a basis of $Span\{A_1, .., A_k\}$. Let V be a linear subspace of $\mathbf{R}^n$ and $\{A_1, .., A_k\}$ be a basis of V. Then the Gram-Schmidt process gives us an orthogonal basis $\{A_1', .., A_k'\}$ of V.

Theorem 4.47. *Every subspace V of $\mathbf{R}^n$ has an orthonormal basis.*

Exercise. Apply Gram-Schmidt to $\{(1,1),(1,0)\}$.

Exercise. Apply Gram-Schmidt to $\{(1,1,1),(1,1,0),(1,0,0)\}$.

Exercise. Let $P = \{A_1, .., A_n\}$ and $Q = \{B_1, .., B_n\}$ be two orthonormal bases of $\mathbf{R}^n$. Recall that the transition matrix from P to Q is $T = B^t(A^{-1})t$ where

$$A = [A_1, .., A_n], \qquad B = [B_1, .., B_n].$$

Explain why T is an orthogonal matrix.

4.11. Least Square Problems

Where the problems come from. In science, we often try to find a theory to fit or to explain a set of experimental data. For example in physics, we might be given a spring and asked to find a relationship between the displacement x of the spring and the force y exerted on it by pulling its ends. Thus we might hang one end of the spring to the ceiling, and then attach various weights to the other end, and then record how much the spring stretches for each test weight. Thus we have a series of given weights representing forces $y_1, .., y_m$ exerted on the spring. The corresponding displacements $x_1, .., x_m$, are then recorded. We can plot the data points $(x_1, y_1), ..., (x_m, y_m)$ on a graph paper. If the spring is reasonably elastic, and the stretches made are not too large, then one discovers that those data points lie almost on a straight line. One might then conjecture the following functional relation between y and x:

$$y = kx + c.$$

This is called Hooke's law. *What are the best values of the constants k, c?* If all the data points (x_i, y_i) were to lie exactly on a single line (they never do in practice), $y = kx + c$, then we would have the equations

$$\begin{gathered} c + kx_1 = y_1 \\ \vdots \\ c + kx_m = y_m, \end{gathered}$$

or in matrix form,

$$AC = Y, \qquad where \;\; A = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}, \quad C = \begin{bmatrix} c \\ k \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

In reality, of course, given the experimental data A, Y, one will not find a vector $C \in \mathbf{R}^2$ such that $AC = Y$ *exactly.* Instead, the next best thing to find is a vector C such that the "error" $\|AC - Y\|^2$ is as small as possible. Finding such an error-minimizing vector C is called a least square problem. Obviously we need more than one data point, ie. $m > 1$, to make a convincing experiment. To avoid redundancy, we may as well also assume that the x_i are all different. Under these assumptions, the rank of the $m \times 2$ matrix A is 2. (Why?)

More generally, a theory might call for fitting a collection of data points $(x_1, y_1), ..., (x_m, y_m)$, using a polynomial functional relation

$$y = c_0 + c_1 x + \cdots + c_n x^n,$$

instead of a linear one. As before, an exact fit would have resulted in the equations

$$\begin{aligned} c_0 + c_1 x_1 + \cdots + c_n x_1^n &= y_1 \\ &\vdots \\ c_0 + c_1 x_m + \cdots + c_n x_m^n &= y_m, \end{aligned}$$

or in matrix form

$$AC = Y, \qquad A = \begin{bmatrix} 1 & x_1 & \cdots & x_1^n \\ & \vdots & & \\ 1 & x_m & \cdots & x_m^n \end{bmatrix}, \quad C = \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}.$$

Thus given the data A, Y, now the least square problem is to find a vector $C = (c_0, .., c_n) \in \mathbf{R}^{n+1}$ such that the error $\|AC - Y\|^2$ is minimum. Again, to make the problem interesting, we assume that $m \geq n+1$ and that the x_i are all distinct. In Chapter 5 when we study the Vandemonde determinants, we will see that under these assumptions the first $n+1$ rows of A above are linearly independent. It follows that $rank(A^t) = rank(A) = n+1$. Let's abstract this problem one step further.

Least Square Problem. Given any $m \times n$ matrix A of rank n with $m \geq n$, and any vector $Y \in \mathbf{R}^m$, find a vector $C = (c_1, .., c_n) \in \mathbf{R}^n$ which minimizes the value $\|AC - Y\|^2$.

4.12. Solutions

Theorem 4.48. *The Least Square Problem has a unique solution.*

Proof: Write A in terms of its columns $A = [A_1, .., A_n]$, and let $V = Col(A) \subset \mathbf{R}^m$ be the subspace spanned by the A_i. By the Best Approximation Theorem above, there is a unique point X in V closest to Y, ie.

$$(*) \qquad \|X - Y\| < \|Z - Y\|$$

for each $Z \in V$ not equal to X. Since the columns A_i span V, each vector in V has the form AE for some $E \in \mathbf{R}^n$. So, we can write $X = AC$ for some $C \in \mathbf{R}^n$, and (*) becomes

$$\|AC - Y\| < \|AD - Y\|$$

for all $D \in \mathbf{R}^n$ such that $AC \neq AD$. Since $rank(A) = n$, $Null(A) = \{O\}$ by the Rank-nullity Theorem. It follows that $AC \neq AD$ (which is equivalent to $A(C - D) \neq O$) is equivalent to $C \neq D$. This shows that the vector C is the solution to the Least Square Problem. □

Theorem 4.49. *(a) The solution to the Least Square Problem is given by*

$$C = (A^t A)^{-1} A^t Y.$$

(b) The projection of $Y \in \mathbf{R}^m$ along the subspace $V = Row(A^t) \subset \mathbf{R}^m$ is given by

$$AC = A\ (A^t A)^{-1} A^t Y.$$

(c) The map $L : \mathbf{R}^m \to \mathbf{R}^m$, $L(Y) = AC$, is a linear map represented by the matrix $A(A^t A)^{-1} A^t$.

Proof: (a) In the preceding proof, we found that there is a unique $C \in \mathbf{R}^n$ such that $X = AC$ is the point in V closest to Y. Recall that in our proof of the Best Approximation Theorem, as a corollary to the Rank-Nullity relation, we found that $X - Y \in V^{\perp}$. It follows that

$$O = A^t(X - Y) = A^t AC - A^t Y.$$

It suffices to show that $A^t A$ is invertible. For then it follows that $C - (A^t A)^{-1}\ A^t Y = O$, which is the assertion (a). We will show that $Null(A^t A) = \{O\}$. Let $Z = (z_1, ..., z_n) \in Null(A^t A)$, ie. $A^t AZ = O$. Dot this with Z, we get

$$0 = Z \cdot A^t AZ = Z^t A^t AZ = (AZ) \cdot (AZ).$$

It follows that $AZ = O$, i.e. $Z \in Null(A) = \{O\}$, so $Z = O$.

Part (b) follows immediately from (a), and part (c) follows immediately from (b).

□

Exercise. You are given the data points $(1,1),(2,2),(3,4),(5,4)$. Find the best line in $\mathbf{R}^2$ that fits these data.

4.13. Homework

1. Let V be a linear subspace of $\mathbf{R}^n$. Decide whether each of the following is TRUE of FALSE. Justify your answer.

   (a) If $dim\ V = 3$, then any list of 4 vectors in V is linearly dependent.

   (b) If $dim\ V = 3$, then any list of 2 vectors in V is linearly independent.

   (c) If $dim\ V = 3$, then any list of 3 vectors in V is a basis.

   (d) If $dim\ V = 3$, then some list of 3 vectors in V is a basis.

   (e) If $dim\ V = 3$, then V contains a linear subspace W with $dim\ W = 2$.

   (f) $(1,\pi)$, $(\pi,1)$ form a basis of $\mathbf{R}^2$.

   (g) $(1,0,0)$, $(0,1,0)$ do not form a basis of the plane $x-y-z=0$.

   (h) $(1,1,0)$, $(1,0,1)$ form a basis of the plane $x-y-z=0$.

   (i) If A is a 3×4 matrix, then the row space $Row(A)$ is at most 3 dimensional.

   (j) If A is a 4×3 matrix, then the row space $Row(A)$ is at most 3 dimensional.

2. Find a basis for the hyperplane in $\mathbf{R}^4$

   $$x-y+2z+t=0.$$

   Find an orthonormal basis for the same hyperplane.

3. Find a basis for the linear subspace of $\mathbf{R}^4$ consisting of all vectors of the form

   $$(a+b,a,c,b+c).$$

Find an orthonormal basis for the same subspace.

4. Find a basis for each of the subspaces $Null(A), Row(A), Row(A^t)$ of $\mathbf{R}^4$, where A is the matrix

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

5. Let A be a 4×6 matrix. What is the maximum possible rank of A? Show that the columns of A are linearly dependent.

6. Prove that if $X \in \mathbf{R}^n$ is a nonzero column vector, then the $n \times n$ matrix XX^t has rank 1.

7. Let $X \in \mathbf{R}^n$. Prove that there exists an orthogonal matrix A such that $AX = \|X\|E_1$. Conclude that for any unit vectors $X, Y \in \mathbf{R}^n$, there exists an orthogonal matrix A such that $AX = Y$. Thus we say that orthogonal matrices act *transitively* on the unit sphere $\|X\| = 1$.

8. Let V be the subspace of $\mathbf{R}^4$ consisting of vectors X orthogonal to $v = (1, -1, 1, -1)$.

   (a) Find a basis of V.

   (b) What is $dim(V)$?

   (c) Give a basis of $\mathbf{R}^4$ which contains a basis of V.

   (d) Find an orthonormal basis of V, and an orthonormal basis of $\mathbf{R}^4$ which contains a basis of V.

9. Let

$$P_1 = (1, -1, 2, -2)$$
$$P_2 = (1, 1, 2, 2)$$
$$P_3 = (1, 0, 2, 0)$$
$$P_4 = (0, 1, 0, 2).$$

Let V be the subspace spanned by the P_i. Find a basis for the orthogonal complement $V^{\perp}$.

10. Find a 3×5 matrix A so that the solution set to $AX = O$ is spanned by

$$P_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

11. Let

$$P_1 = (1, 2, 2, 4)$$
$$P_2 = (2, -1, 4, -2)$$
$$P_3 = (4, -2, -2, 1)$$
$$P_4 = (2, 4, -1, -2)$$
$$Q = (1, 1, 1, 1).$$

(a) Verify that $\{P_1, P_2, P_3, P_4\}$ is an orthogonal set.

(b) Write Q as a linear combination of the P_i.

(c) Let V be the subspace spanned by P_1, P_2. Find the point of V closest to Q.

(d) Find the shortest distance between V and Q.

12. (a) Let U be the span of the row vectors $(1, 1, 1, 1, 1)$ and $(3, 2, -1, -4, 3)$. Let V be the span of $(5, 4, 1, -2, 5)$ and $(2, -1, -9, 1, 9)$. Find a basis of the linear subspace $U + V$ in $\mathbf{R}^5$.

(b) Let U be the span of $A_1, .., A_k \in \mathbf{R}^n$. Let V be the span of $B_1, .., B_l \in \mathbf{R}^n$. Design an algorithm *using row operations* for finding a basis of the linear subspace $U + V$ in $\mathbf{R}^n$.

13. What is the rank of an $n \times n$ upper triangular matrix where the diagonal entries are all nonzero? Explain.

14. Let U, V be subspaces of $\mathbf{R}^n$ such that U contains V. Prove that $V^\perp$ contains $U^\perp$.

15. * Let A be an $m \times n$ matrix, and let B be an $n \times r$ matrix.

(a) Prove that the columns of AB are linear combinations of the columns of A. Thus prove that

$$rank(AB) \leq rank(A).$$

(b) Prove that

$$rank(AB) \leq rank(B).$$

(Hint: $rank(AB) = rank\ (AB)^t$ and $rank(B) = rank(B^t)$.)

16. * Suppose that A, B are reduced row echelons of the same size and that $Row(A) = Row(B)$.

(a) Show that A, B have the same number of nonzero rows.

(b) Denote the addresses of A, B by $p_1 < \cdots < p_k$, $q_1 < \cdots < q_k$, respectively. Denote the nonzero rows of A, B by $A_1, .., A_k$, $B_1, .., B_k$, respectively. Show that $p_1 \leq q_1$. Conclude that $p_1 = q_1$. (Hint: Write $A_1 = x_1 B_1 + \cdots + x_k B_k$ and dot both sides with E_{p_1}. Show that $A = B$.)

(c) Show that $A_1 = B_1$. By induction, show that $A_i = B_i$ for all i.

(d) Show that if A, B are reduced row echelons of a matrix C, then $A = B$. This shows that a reduced row echelon of any given matrix is unique.

(e) Suppose that A, B are matrices of the same size and that $Row(A) = Row(B)$. Show that A, B are row equivalent. (Hint: Reduce to the case when A, B are reduced row echelons.)

17. Let U and V be linear subspaces of $\mathbf{R}^n$. Prove that

$$U \cap V = (U^\perp + V^\perp)^\perp.$$

18. * Suppose you are given two subspaces U, V of $\mathbf{R}^n$ and given their respective bases $\{A_1, .., A_k\}$ and $\{B_1, .., B_l\}$. Design an algorithm to find a basis of $U \cap V$.

19. * Let U, V be subspaces of $\mathbf{R}^n$ such that $U \subset V$. Prove that $(V \cap U^\perp) + U$ is a direct sum and it is equal to V. Hence conclude that

$$\dim(V \cap U^\perp) = \dim V - \dim U.$$

(Hint: To show $V \subset (V \cap U^\perp) + U$, for $A \in V$, show that $A - B \in V \cap U^\perp$ where B is the best approximation of A in U, hence $A = (A - B) + B$.)

20. Let P, Q, R be three bases of a linear subspace V in $\mathbf{R}^n$. Let T, T', T'' be the respective transition matrices from P to Q, from Q to R, and P to R. Prove that

$$T'' = T'T.$$

21. Let $V = \mathbf{R}^n$, $P = \{A_1, .., A_n\}$ be a basis, and Q the standard basis. Regard the A_i as column vectors and let $A = [A_1, .., A_n]$. Show directly (with Theorem 4.38) that the transition matrix from Q to P is A^t. Conclude that the transition matrix from P to Q is $(A^{-1})^t$.

22. Prove that if A is an $m \times n$ matrix of rank 1, then $A = BC^t$ for some column vectors $B \in \mathbf{R}^m$, $C \in \mathbf{R}^n$. (Hint: Any two rows of A are linearly dependent.)

23. * Let A, B be $n \times n$ matrix such that $AB = O$. Prove that $rank(A) + rank(B) \leq n$.

24. Prove that in $\mathbf{R}^n$, any n points are coplanar. In other words, there is a hyperplane which contains all n points. If the n points are linearly independent, prove that the hyperplane containing them is unique.

25. Let V be a given linear subspace of $\mathbf{R}^n$. Define the map

$$L : \mathbf{R}^n \to \mathbf{R}^n, \quad X \mapsto X'$$

where X' is the closest point in V to X. Show that L is linear. Fix an orthonormal basis $B_1, .., B_k$ of V. Find the matrix of L in terms of these column vectors. (Hint: The best approximation theorem.)

26. In a physics experiment, you measure the time t it takes for a water-filled balloon to hit the ground when you release it from various height h in a tall building. Suppose that the measurements you've made for (h, t) give the following data points: $(1, 5), (2, 6), (3, 8), (4, 9)$. It seems that h and t bear a quadratic relation $h = at^2 + bt + c$. Use the Least Square method to find the best values of a, b, c that fit your data.

5. Determinants

Recall that a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if $ad - bc \neq 0$. Conversely if A is invertible, then $ad - bc \neq 0$. This is a single numerical criterion for something that starts out with four numbers a, b, c, d. The number $ad - bc$ is called the determinant of A.

Question. Is there an analogous numerical criterion for $n \times n$ matrices?

• When $ad - bc \neq 0$, we know that $A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. This is a formula for A^{-1} expressed explicitly in terms of the entries of A.

Question. When A is invertible, is there an explicit formula for A^{-1} in terms of the entries of A?

We will study the determinant of $n \times n$ matrices, and answer these questions in the affirmative. We will also use it to study volume preserving linear transformations.

5.1. Permutations

Definition 5.1. *A permutation of n letters is a rearrangement of the first n positive integers. We denote a permutation by a list $K = (k_1, .., k_n)$ where $k_1, .., k_n$ are distinct positive integers between 1 and n. The special permutation $(1, 2, .., n)$ is called the identity permutation and is denoted by Id.*

Example. There is just one permutation of 1 letter: (1). There are two permutations of 2 letters: $(1, 2)$ and $(2, 1)$. There are 6 permutations of 3 letters: $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$, $(3, 2, 1)$.

Exercise. List all permutations of 4 letters.

Theorem 5.2. *There are exactly $n!$ permutations of n letters.*

Proof: We want to find the total number T_n of ways to fill n slots

$$(-,-,...,-)$$

with the n distinct letters $1,2,..,n$. There are clearly n different choices to fill the first slot. After the first slot is filled, there are $n-1$ slots left to be filled with the remaining $n-1$ distinct letters. So

$$T_n = n \cdot T_{n-1}.$$

Since $T_1 = 1$, we have $T_2 = 2\cdot 1 = 2$, $T_3 = 3\cdot 2\cdot 1 = 3!$, and so on. This shows that $T_n = n!$.
$\square$

5.2. The sign function

Define the *Vandemonde function* of n variables:

$$\begin{aligned} V(x_1,.,,x_n) &= (x_2-x_1)\times(x_3-x_1)(x_3-x_2)\times\cdots\times(x_n-x_1)(x_n-x_2)\cdots(x_n-x_{n-1}) \\ &= \prod_{1\le i<j\le n}(x_j-x_i). \end{aligned}$$

It follows immediately from this definition that:

• In this product, there is one factor $(x_j - x_i)$ for each pair i,j with $i<j$, and so there are exactly $\frac{1}{2}n(n-1)$ such factors.

• No two factors are equal, even up to sign.

Let $K = (k_1,..,k_n)$ be a permutation, and consider

$$V(x_{k_1},x_{k_2},..,x_{k_n}) = \prod_{1\le i<j\le n}(x_{k_j}-x_{k_i}).$$

Thus, this is also a product of $\frac{1}{2}n(n-1)$ factors of the form $(x_b - x_a)$ with $a\neq b$. Again, no two factors are equal, even up to sign. Thus the factors occuring here must be those occuring in $V(x_1,..,x_n)$ above, up to signs. In other words, $(x_b - x_a)$ occurs in $V(x_1,..,x_n)$ iff $(x_a - x_b)$ or $(x_b - x_a)$ (but not both) occurs in $V(x_{k_1}, x_{k_2},..,x_{k_n})$. It follows that

$$V(x_{k_1}, x_{k_2},..,x_{k_n}) = \pm V(x_1,..,x_n)$$

where the sign $\pm$ depends only on K.

Definition 5.3. *For each permutation of n letters $K = (k_1,..,k_n)$, we define $sign(K)$ to be the number ± 1 such that*

$$V(x_{k_1}, x_{k_2},..,x_{k_n}) = sign(K)V(x_1,..,x_n).$$

Example. Obviously $sign(Id) = +1$.

Example. For two letters, $V(x_1, x_2) = x_2 - x_1$. So

$$V(x_2, x_1) = (x_1 - x_2) = -(x_2 - x_1) = -V(x_1, x_2),$$

and $sign(2,1) = -1$. For three letters, $V(x_1, x_2, x_3) = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$. So

$$V(x_3, x_1, x_2) = (x_1 - x_3)(x_2 - x_3)(x_2 - x_1) = V(x_1, x_2, x_3),$$

and $sign(3,1,2) = +1$.

Exercise. Find all $sign(K)$ for three letters.

Given a permutation $K = (k_1,..,k_n)$, we can *swap two neighboring* entries of K and get a new permutation L. Clearly we can apply a series of such swaps to transform $K = (k_1,..,k_n)$ to $Id = (1,2,..,n)$. By reversing the swaps, we transform Id to K. The following theorem is proved in the Appendix of this chapter.

Theorem 5.4. *(Sign Theorem) If K transforms to L by a swap, then*

$$sign(L) = -sign(K).$$

In particular if K transforms to Id by a series of m swaps, then

$$sign(K) = (-1)^m.$$

Example. The series of swaps

$$(3,1,2) \to (1,3,2) \to (1,2,3)$$

has length 2. So $sign(3,1,2) = (-1)^2 = 1$ as before.

Example. The series of swaps

$$(2,4,1,3) \to (2,1,4,3) \to (1,2,4,3) \to (1,2,3,4)$$

has length 3. So $sign(2,4,1,3) = (-1)^3 = -1$.

Exercise. Find $sign(5,1,4,2,3)$.

Corollary 5.5. *If $J = (j_1, .., j_n)$ and $K = (k_1, .., k_n)$ differ exactly by two entries, then they have different sign, ie. $sign(K) = -sign(J)$.*

Proof: Suppose $b < c$, and $k_a = j_a$ for all $a \neq b, c$, and $k_b = j_c,\ k_c = j_b$, ie.

$$(k_1, .., k_b, .., k_c, .., k_n) = (j_1, .., j_c, .., j_b, .., j_n).$$

Then we can transform J to K by first moving j_b to the j_c slot via $b - c$ swaps, and then followed by moving j_c back to the j_b slot via $b - c - 1$ swaps. So in this manner, it takes $2(b - c) - 1$ swaps to transform J to K. By the Sign Theorem,

$$sign(K) = (-1)^{2(b-c)-1} sign(J) = -sign(J). \quad \square$$

Exercise. Find $sign(50, 2, 3, ..., 49, 1)$.

Exercise. Find $sign(50, 49, ..., 2, 1)$.

Theorem 5.6. *(Even Permutations) There are exactly $n!/2$ permutations of n letters K with $sign(K) = +1$.*

Proof: Let S_n denotes the set of all permutations of n letters. Define a map

$$\sigma : S_n \to S_n, \quad \sigma(k_1, k_2, k_3, ..., k_n) = (k_2, k_1, k_3, ..., k_n).$$

This map is invertible with itself being the inverse. By the sign theorem, $sign(\sigma(K)) = -sign(K)$. Now $S_n = A_n \cup B_n$ where A_n is the set of permutations with plus signs, and B_n those with minus signs. Therefore $\sigma(A_n) \subset B_n$ and $\sigma(B_n) \subset A_n$. Since σ is one-to-one, it follows that A_n and B_n have the same number of elements. Since S_n is the disjoint union of A_n and B_n, each must have $n!/2$ elements. $\square$

5.3. Sum over permutations

Definition 5.7. *Let $A = (a_{ij})$ be an $n \times n$ matrix. We define the determinant $det(A)$ of A to be the sum*

$$\sum_K sign(K) a_{1k_1} a_{2k_2} \cdots a_{nk_n},$$

where K ranges over all permutations of n letters.

So $det(A)$ is a sum of $n!$ terms – one for each permutation K. Each term is a product which comes from selecting (according to K) one entry from each row, ie. the product

$$sign(K) a_{1k_1} a_{2k_2} \cdots a_{nk_n}$$

comes from selecting the k_1th entry from row 1, k_2th entry from row 2, and so on. Note that the k's are all *different* integers between 1 and n. Note that the coefficient of this product of the a's is always $sign(K) = \pm 1$.

Example. If A is 2×2, then

$$det(A) = sign(12) a_{11} a_{22} + sign(21) a_{12} a_{21} = +a_{11} a_{22} - a_{12} a_{21}.$$

Example. If A is 3×3, then

$$\begin{aligned} det(A) = &+ a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} \\ &- a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

Exercise. Find $det(I)$ for the 3×3 identity matrix I.

Exercise. What is $det(I)$ for the $n \times n$ identity matrix I? Write $I = (\delta_{ij})$. Then $det(I)$ is the sum of $n!$ products of the form

$$sign(K) \delta_{1k_1} \delta_{2k_2} \cdots \delta_{nk_n}.$$

What is this product when $K \neq Id$? $K = Id$? Conclude that

$$det(I) = 1.$$

Exercise. Find $det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}$.

Exercise. A square matrix A is said to be upper triangular if the entries below the diagonal are all zero. Prove that if $A = (a_{ij})$ is an $n \times n$ upper triangular matrix, then

$$det(A) = a_{11}a_{22} \cdots a_{nn}.$$

Exercise. Find $det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$.

Exercise. Prove that if $A = (a_{ij})$ is an $n \times n$ matrix having a zero row, then

$$det(A) = 0.$$

Theorem 5.8. $det(A^t) = det(A)$.

The proof is given in the Appendix of this chapter.

Exercise. Verify this theorem for 3×3 matrices using the formula for $det(A)$ above.

5.4. Determinant as a multilinear alternating function

We now regard an $n \times n$ matrix A as a list of n row vectors in $\mathbf{R}^n$. Then the determinant is now a function which takes n row vectors $A_1, .., A_n$ as the input and yield the number $det(A)$ as the output. Symbolically, we write

$$det(A) = det \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix}.$$

We now discuss a few important properties of det:

(i) *Alternating property.* Suppose $A_j = A_i$ for some $j > i$. Then

$$det(A) = 0.$$

Proof: Fixed i, j with $j > i$. Let S_n be the set of permutations of n letters, A_n be the subset with plus signs, and B_n those with minus signs. Define a map (cf. proof of Even Permutations Theorem)

$$\sigma : S_n \to S_n, \quad \sigma(k_1, .., k_i, .., k_j, .., k_n) = (k_1, .., k_j, .., k_i, .., k_n)$$

ie. $\sigma(K)$ is obtained from K by interchanging the i, j entries. The map σ is one-to-one, and $sign(\sigma(K)) = -sign(K)$ by the Sign Theorem. Thus $\sigma(A_n) = B_n$.

Put

$$f(K) := sign(K) a_{1k_1} \cdots a_{i,k_i} \cdots a_{j,k_j} \cdots a_{nk_n}.$$

Then

$$det(A) = \sum_{K \in S_n} f(K) = \sum_{K \in A_n} f(K) + \sum_{K \in B_n} f(K).$$

The last sum is equal to $\sum_{K \in \sigma(A_n)} f(K) = \sum_{K \in A_n} f(\sigma(K))$, so that

$$det(A) = \sum_{K \in A_n} (f(K) + f(\sigma(K))).$$

Now

$$\begin{aligned} f(\sigma(K)) &= sign(\sigma(K)) a_{1k_1} \cdots a_{i,k_j} \cdots a_{j,k_i} \cdots a_{nk_n} \\ &= -sign(K) a_{1k_1} \cdots a_{i,k_j} \cdots a_{j,k_i} \cdots a_{nk_n}. \end{aligned}$$

Since row i and j of A are the same, we have $a_{i,k_j} = a_{j,k_j}$ and $a_{j,k_i} = a_{i,k_i}$. So

$$\begin{aligned} f(\sigma(K)) &= -sign(K) a_{1k_1} \cdots a_{j,k_j} \cdots a_{i,k_i} \cdots a_{nk_n} \\ &= -sign(K) a_{1k_1} \cdots a_{i,k_i} \cdots a_{i,k_j} \cdots a_{nk_n} \\ &= -f(K) \end{aligned}$$

Thus $det(A) = 0$. $\square$

(ii) *Scaling property.*

$$det \begin{pmatrix} A_1 \\ \vdots \\ cA_i \\ \vdots \\ A_n \end{pmatrix} = c \; det \begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix}.$$

Proof: In the definition of $det(A)$, each term

$$sign(K)a_{1k_1}a_{2k_2}\cdots a_{nk_n} \qquad (*)$$

has a factor a_{ik_i} which is an entry in row i. Thus if row i is scaled by a number c, each term (*) is scaled by the same number. Thus $det(A)$ is scaled by an overall factor c. $\square$

(iii) *Additive property.*

$$det\begin{pmatrix} A_1 \\ \vdots \\ A_i + V \\ \vdots \\ A_n \end{pmatrix} = det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix} + det\begin{pmatrix} A_1 \\ \vdots \\ V \\ \vdots \\ A_n \end{pmatrix}.$$

Proof: The argument is similar to proving the scaling property. $\square$

(iv) *Unity property.*

$$det(I) = 1.$$

(v) Since $det(A^t) = det(A)$, each of the properties (i)-(iii) can be stated in terms of columns, by regarding $det(A)$ as a function of n column vectors.

5.5. Computational consequences

We now study how determinant changes under a row operation. Let A be an $n \times n$ matrix. Consider the matrix obtained from A by replacing the two rows A_i, A_j with $j > i$, by $A_i + A_j,\ A_i + A_j$. Then applying properties (i) and (iii), we have

$$0 = det\begin{pmatrix} A_1 \\ \vdots \\ A_i + A_j \\ \vdots \\ A_i + A_j \\ \vdots \\ A_n \end{pmatrix} = det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_j \\ \vdots \\ A_n \end{pmatrix} + det\begin{pmatrix} A_1 \\ \vdots \\ A_j \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix}.$$

This shows that if A' is obtained from A by interchanging two rows, then $det(A') = -det(A)$.

Now suppose A' is obtained from A by adding c times row i to another row $j > i$. Then by properties (ii)-(iii), we have

$$det(A') = det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_j + cA_i \\ \vdots \\ A_n \end{pmatrix} = det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_j \\ \vdots \\ A_n \end{pmatrix} + c\ det\begin{pmatrix} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_i \\ \vdots \\ A_n \end{pmatrix}.$$

The first term on the right hand side is just $det(A)$, and the second term is zero by the alternating property. Thus $det(A') = det(A)$.

Theorem 5.9. *Let A, A' be $n \times n$ matrices.*

(i) If A transforms to A' by row operation R1, then

$$det(A') = -det(A).$$

(ii) If A transforms to A' by row operation R2 with scalar c, then

$$det(A') = c\ det(A).$$

(iii) If A transforms to A' by row operation R3, then

$$det(A') = det(A).$$

Thus given an square matrix A, each row operation affects $det(A)$ only by a *nonzero* scaling factor. By row reduction, we have a sequence of some m row operations transforming A to a reduced row echelon B. The change in $det(A)$ after each row operation is a nonzero scaling factor c_i. At the end of the m row operations, we get

$$det(B) = c_1 c_2 \cdots c_m det(A).$$

Note that the numbers $c_1, .., c_m$ are determined using properties (i)-(iii) *alone.*

Consider the reduced row echelon B. If B has a zero row, then B is unchanged if we scale that zero row by -1. By property (ii), we get $det(B) = -det(B)$. Hence $det(B) = 0$

in this case. If B has no zero row then, by Theorem 3.3, $B = I$. Hence $det(B) = 1$ by (iv). This shows that $det(A)$ can be computed by by performing row reduction on A and determining the numbers $c_1, .., c_m$.

Exercise. Write down your favorite 3×3 matrix A and find $det(A)$ by row reduction. Verify that your answer agree with the formula for $det(A)$ above.

Exercise. Find $det \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}$.

5.6. Theoretical consequences

Theorem 5.10. *If $K = (k_1, ..., k_n)$ is a permutation and $A_1, .., A_n \in \mathbf{R}^n$ are column vectors, then*

$$det[A_{k_1}, .., A_{k_n}] = sign(K) det[A_1, .., A_n].$$

And there is a similar statement for rows.

Proof: We can perform a sequence of swaps on the columns $[A_{k_1}, .., A_{k_n}]$ to transform it to $[A_1, .., A_n]$. The same sequence of swaps transforms the permutation $(k_1, k_2, .., k_n)$ to $(1, 2, .., n)$. Each swap on the matrix and on the permutation result in a sign change. By the Sign Theorem, the net sign change is $sign(K)$. $\square$

Theorem 5.11. *(Determinant Recognition) det is the only function on square matrices which has properties (i)-(iv).*

Proof: Let A be a square matrix. By using properties (i)-(iii), we have shown that

$$det(B) = c_1 c_2 \cdots c_m det(A)$$

where $c_1, .., c_m$ are nonzero numbers *determined using (i)-(iii) alone*, while a sequence of m row operations transform A to a reduced row echelon B. Therefore if F is a function on square matrices which has properties (i)-(iv), then

$$F(B) = c_1 c_2 \cdots c_m F(A).$$

If B has a zero row, by using property (ii) alone we have seen that $det(B) = 0$. Thus similarly $F(B) = 0 = det(B)$. If B has no zero row then, by Theorem 3.3, $B = I$. So $F(B) = 1 = det(B)$ by property (iv). Thus in general, we have

$$F(A) = \frac{1}{c_1 \cdots c_m} F(B) = \frac{1}{c_1 \cdots c_m} det(B) = det(A). \qquad \square$$

Theorem 5.12. *If $det(A) \neq 0$, then A is invertible.*

Proof: Let B be a reduced row echelon of A. Since $det(B)$ is $det(A)$ times a nonzero number, it follows that $det(B) \neq 0$. This shows that B has no zero rows. By Theorem 3.3, it follows that $B = I$. Thus A is row equivalent to I, hence invertible, by Theorem 3.6. $\square$

Theorem 5.13. *(Multiplicative Property) If A, B are $n \times n$ matrices, then $det(AB) = det(A)det(B)$.*

Proof: (Sketch) We will sketch two proofs, the first being more computational, and the second conceptual.

First proof. We will use the column versions of properties (i)-(iii). For simplicity, Let's consider the case of $n = 3$. Put $B = [B_1, B_2, B_3] = (b_{ij})$. Then

$$\begin{aligned} det(AB) &= det[AB_1, AB_2, AB_3] \\ &= det[b_{11}A_1 + b_{21}A_2 + b_{31}A_3, b_{12}A_1 + b_{22}A_2 + b_{32}A_3, b_{13}A_1 + b_{23}A_2 + b_{33}A_3] \end{aligned}$$

Expanded using properties (ii)-(iii), the right hand side is the sum

$$\sum_i \sum_j \sum_k b_{i1} b_{j2} b_{k3} det[A_i, A_j, A_k]$$

where i, j, k takes values in $\{1, 2, 3\}$. By the alternating property (i), $det[A_i, A_j, A_k]$ is zero unless i, j, k are distinct. So

$$det(AB) = \sum_{(i,j,k)} b_{i1} b_{j2} b_{k3} det[A_i, A_j, A_k]$$

where we sum over all distinct i, j, k. But a triple (i, j, k) with distinct entries is precisely a permutation of 3 letters. Given such a triple, we also have $det[A_i, A_j, A_k] = sign(i, j, k) det[A_1, A_2, A_3]$. So we get

$$det(AB) = \sum_{(i,j,k)} sign(i, j, k) b_{i1} b_{j2} b_{k3}\ det[A_1, A_2, A_3] = det(B)det(A). \qquad \square$$

Second proof. Fix B and define the function

$$F_B(A) := det(AB)$$

on n row vectors $A = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}$, $A_i \in \mathbf{R}^n$. Using the fact that $AB = \begin{bmatrix} A_1 B \\ \vdots \\ A_n B \end{bmatrix}$ and properties i.–iv. of the determinant in Theorem 5.11, it is easy to check that the function $F_B(A)$ too has the following properties:

i. F_B has the alternating property, i.e. if A contains two equal rows then $F_B(A) = 0$;

ii. F_B has the additivity property in each row;

iii. F_B has the scaling property in each row;

iv. $F_B(I) = det(B)$.

We now argue that F_B is the only function with properties i.–iv. Let $A = P_0 \to P_1 \to P_2 \to \cdots \to P_m = A'$ be a sequence of row operations transforming A to its reduced row echelon A'. Then either $A' = I$ (if A is invertible) or A' has a zero row (if A is not invertible). By properties i.–iii., we find $F_B(A) = c_1 F_B(P_1)$ where $c_1 = -1$ if $A \to P_1$ by row operation R1, $c_1 = 1$ if $A \to P_1$ is row operation R2, or $c_1 = c$ if $A \to P_1$ is row operation R3, i.e. scaling a row of A by a constant $c \neq 0$. Likewise, we have

$$F_B(P_{i-1}) = c_i F_B(P_i), \quad i = 1, .., m$$

where c_i is the constant determined by properties i.–iii. and the choice of row operation $P_{i-1} \to P_i$. Thus we have

$$F_B(A) = c_1 \cdots c_m F_B(A').$$

Finally, $F_B(A') = 1$ if $A' = I$ by iv., or $F_B(A') = 0$ if A' has a zero row by iii. This determines $F_B(A)$. It follows that if F is any other function with properties i.–iv., then we must also have

$$F(A) = c_1 \cdots c_m F(A')$$

with $F(A') = F_B(A')$. This shows that $F_B(A) = F(A)$. In other words, F_B is the only function with properties i.–iv.

It is easy to verify that the function

$$A \mapsto det(A)det(B)$$

also has properties i.–iv. It follows that $det(AB) = F_B(A) = det(A)det(B)$ by uniqueness proven above. $\square$

Corollary 5.14. *If A is invertible, then $det(A) \neq 0$. In this case $det(A^{-1}) = 1/det(A)$.*

Proof: $det(AA^{-1}) = det(A) \cdot det(A^{-1}) = 1$. $\square$

5.7. Minors

We now discuss a recursive approach to determinants. Throughout this section, we assume that $n \geq 2$.

Definition 5.15. *Let A be an $n \times n$ matrix. The (ij) minor of A is the number $|M_{ij}| = det(M_{ij})$ where M_{ij} is the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j.*

Theorem 5.16. *(Expansion along row i)*

$$det(A) = (-1)^{i+1} a_{i1}|M_{i1}| + (-1)^{i+2} a_{i2}|M_{i2}| + \cdots + (-1)^{i+n} a_{in}|M_{in}|.$$

This can be proved as follows. Regard each $|M_{ij}|$ as a function on $n \times n$ matrices. Show that the sum on the right hand side satisfies all four properties (i)-(iv). Now use the fact that det is the only such function. We omit the details.

Example. Let $A = (a_{ij})$ be 3×3. By expanding along row 1, we get

$$det(A) = +a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Exercise. There is a similar theorem for expansion along a column. Formulate this theorem. (cf. $det(A^t) = det(A)$.)

Exercise. Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Find $det(A)$ by expansion along a row. Which row is the easiest?

Definition 5.17. *Let A be an $n \times n$ matrix, and let $|M_{ij}|$ be its (ij) minor. The adjoint matrix of A is the $n \times n$ matrix A^* whose (ij) entry is $(-1)^{j+i}|M_{ji}|$.*

Example. If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then

$$A^* = \begin{bmatrix} +a_{22} & -a_{12} \\ -a_{21} & +a_{11} \end{bmatrix}.$$

Example. If $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, then

$$A^* = \begin{bmatrix} + \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & - \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & + \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ - \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & + \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & - \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} \\ + \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} & - \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} & + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix}.$$

Exercise. Find AA^* and A^*A.

Theorem 5.18. *(Cramer's rule) $AA^* = det(A)I$.*

Proof: The formula for expansion along row i above can be written in terms of dot product as:

$$det(A) = (a_{i1}, .., a_{in}) \cdot \left((-1)^{i+1}|M_{i1}|, ..., (-1)^{i+n}|M_{in}|\right) = {}_iA \cdot A_i^*$$

where ${}_iA$ is the ith row of A, and A_i^* is ith column of A^*. This shows that all the diagonal entries of AA^* are equal to $det(A)$.

We now show that all the (ij) entries, with $i \neq j$, of AA^* are zero. Let C be the matrix obtained from A by replacing its row i by its row j, so that rows i, j of C are identical. Thus $det(C) = 0$. Expanding along row i, we get

$$0 = det(C) = (-1)^{i+1}c_{i1}|M_{i1}| + (-1)^{i+2}c_{i2}|M_{i2}| + \cdots + (-1)^{i+n}c_{in}|M_{in}|$$
$$= (a_{j1}, .., a_{jn}) \cdot \left((-1)^{i+1}|M_{i1}|, ..., (-1)^{i+n}|M_{in}|\right) \qquad (*).$$

Here $|M_{ij}|$ denotes the minors of C. Since A differs from C by just row i, A and C have the same minors $|M_{i1}|, .., |M_{in}|$. So (*) becomes

$$0 = {}_jA \cdot A_i^*. \qquad \square$$

Exercise. Use Cramer's rule to find the inverse of

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Find A^{-1} by row reduction. Which method do you find more efficient?

Exercise. Use Cramer's rule to give another proof of Theorem 5.12.

5.8. Geometry of determinants

Definition 5.19. *Let u, v be two vectors in $\mathbf{R}^n$. The set*

$$\{t_1u + t_2v | 0 \le t_i \le 1\}$$

is called the parallellogram generated by u, v.

Exercise. Draw the parallelogram generated by $(1, 0)$ and $(0, 1)$.

Exercise. Draw the parallelogram generated by $(1, -1)$ and $(2, 1)$.

Definition 5.20. *In $\mathbf{R}^2$ the signed area of a parallellogram generated by u, v is defined to be $det(u, v)$. The absolute value is called the area.*

Recall that the determinant function on two vectors in $\mathbf{R}^2$ has the following algebraic properties:

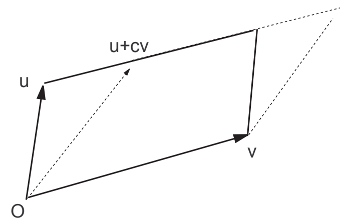
(i) $det(u, v) = -det(v, u)$.

(ii) (scaling) $det(cu, v) = cdet(u, v)$.

(iii) (shearing) $det(u + cv, v) = det(u, v)$.

(iv) (unit square) $det(E_1, E_2) = 1$.

Each of them corresponds to a simple geometric properties of area. For example, (iv) corresponds to the fact that the standard square in $\mathbf{R}^2$ has area 1. Property (iii) says that the area of a parallellogram doesn't change under "shearing" along an edge, as depicted here.



Example. Let's find the area of the triangle with vertices $(0,0),\ (1,1),\ (1,2)$.

Consider the parallelogram generated by $(1,1),\ (1,2)$. Its area is twice the area we want. Thus the answer is $\frac{1}{2}|det\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}| = \frac{1}{2}$.

Definition 5.21. *Let $u_1,..,u_n$ be vectors in $\mathbf{R}^n$. Then set*

$$\{t_1u_1 + \cdots + t_nu_n | 0 \le t_i \le 1\}$$

is called the parallelopiped generated by $u_1,..,u_n$.

Definition 5.22. *In $\mathbf{R}^n$ the signed volume of a parallelopiped generated by $u_1,..,u_n$ is defined to be $det(u_1,..,u_n)$. The absolute value is called the volume.*

Exercise. Find the volume of the parallelopiped generated by $(1,1,1),\ (1,-1,0),\ (0,0,1)$.

We now return to a question we posed in chapter 3: *when does an $n \times n$ matrix preserve volume?*

Theorem 5.23. *A square matrix A preserves volume iff $|det(A)| = 1$.*

Proof: Again, for clarity, let's just consider the case of 2×2.

Let B_1, B_2 be arbitrary vectors in $\mathbf{R}^2$. Recall that the area (volume) of the parallelogram generated by them is $|det(B)|$. Under the transformation A, the new parallelogram is generated by the transform vectors AB_1, AB_2. So the new volume is

$$(*) \quad |det[AB_1, AB_2]| = |det(AB)| = |det(A)||det(B)|.$$

Thus, if A preserves volume, then the right hand side is equal to $|det(B)|$, implying that $|det(A)| = 1$. Conversely, if $|det(A)| = 1$, then the $(*)$ says that A preserves volume. $\square$

Example. Consider $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Note that $\frac{1}{\sqrt{2}}A$ is an orthogonal matrix. So, A preserves angle (chapter 3). But $det(A) = -2$, and so A does not preserve volume.

Cross product. This is a *bilinear* operation on $\mathbf{R}^3$. It takes two vectors A, B as the input and yields one vector $A \times B$ as the output. It is defined by the formula

$$A \times B = det \begin{bmatrix} E_1 & E_2 & E_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = (a_2b_3 - a_3b_2)E_1 - (a_1b_3 - a_3b_1)E_2 + (a_1b_2 - a_2b_1)E_3$$

for $A = (a_1, a_2, a_3)$ and $B = (b_1, b_2, b_3)$. The vector $A \times B$ is called the cross product of A and B. We will study a few properties of this operation. Let A, B, C be row vectors in $\mathbf{R}^3$.

Claim: $(A \times B) \cdot C = det \begin{bmatrix} A \\ B \\ C \end{bmatrix}$.

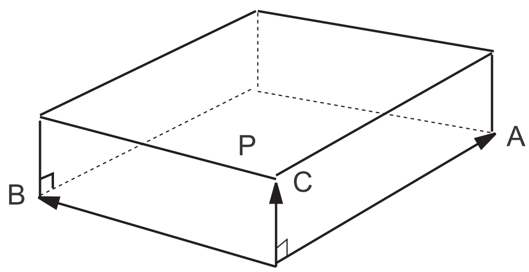
Proof: By the formula for $A \times B$ above, we have

$$\begin{aligned}(A \times B) \cdot C &= ((a_2b_3 - a_3b_2)E_1 - (a_1b_3 - a_3b_1)E_2 + (a_1b_2 - a_2b_1)E_3) \cdot (c_1E_1 + c_2E_2 + c_3E_3) \\ &= (a_2b_3 - a_3b_2)c_1 - (a_1b_3 - a_3b_1)c_2 + (a_1b_2 - a_2b_1)c_3 \\ &= det \begin{bmatrix} A \\ B \\ C \end{bmatrix}.\end{aligned}$$

$\square$

If $C = A$ or $C = B$, then the alternating property of determinant implies that $(A \times B) \cdot C = 0$. Thus $A \times B$ is orthogonal to both A and B. It follows that $A \times B$ is orthogonal to any linear combination $aA + bB$. Now suppose A, B are linearly independent, so that they span a plane in $\mathbf{R}^3$. Then $A \times B$ is orthogonal to any vector in that plane. In other words, *$A \times B$ is a vector perpendicular to the plane spanned by A and B.*

What does the number $\|A \times B\|$ represents geometrically? The vectors A, B generate a parallelogram P in the plane they span. We will see that *$\|A \times B\|$ is the area of P* (according to the standard meter stick in $\mathbf{R}^3$). Let C be any vector perpendicular to plane spanned by A, B. Consider the volume of the parallelopiped generated by A, B, C, which we can find in two ways. First this volume is $Area(P)\|C\|$, as the picture here shows.



Vol=Area(P).||C||

On the other hand, the volume is also $|det \begin{bmatrix} A \\ B \\ C \end{bmatrix}|$. In particular, if $C = \frac{A \times B}{\|A \times B\|}$ we have $\|C\| = 1$, and hence

$$\begin{aligned} Area(P) &= Area(P)\|C\| \\ &= |det \begin{bmatrix} A \\ B \\ C \end{bmatrix}| \\ &= |(A \times B) \cdot C| \\ &= \frac{(A \times B) \cdot (A \times B)}{\|A \times B\|} \\ &= \|A \times B\|. \end{aligned}$$

Exercise. Find the area of the parallelogram generated by $A = (1, 1, 0)$ and $B = (1, 0, -1)$ in $\mathbf{R}^3$.

5.9. Appendix

Theorem 5.24. *(Sign Theorem) If K transforms to L by a swap, then*

$$sign(L) = -sign(K).$$

In particular if K transforms to Id by a series of m swaps, then

$$sign(K) = (-1)^m.$$

Proof: Suppose

$$l_i = \begin{cases} k_i & if \quad i \neq a, a+1 \\ k_{a+1} & if \quad i = a \\ k_a & if \quad i = a+1 \end{cases}$$

ie. $(l_1, .., l_a, l_{a+1}, .., l_n) = (k_1, .., k_{a+1}, k_a, .., k_n)$. We will show that

$$V(x_{l_1}, .., x_{l_n}) = -V(x_{k_1}, .., x_{k_n}).$$

By comparing the factors of $V(x_{l_1}, .., x_{l_n})$ with those of $V(x_{k_1}, .., x_{k_n})$, we will show that they differ by exactly one factor, and that factor differs by a sign. Now, the factors of $V(x_{l_1}, .., x_{l_n})$ are

$$x_{l_j} - x_{l_i}, \quad j > i.$$

Case 1. $j \neq a+1$, $i \neq a$. Then

$$x_{l_j} - x_{l_i} = x_{k_j} - x_{k_i}.$$

Since $j > i$, the right hand side is a factor of $V(x_{k_1}, .., x_{k_n})$.

Case 2. $j = a+1$, $i < a$. Then

$$x_{l_j} - x_{l_i} = x_{k_a} - x_{k_i}.$$

Since $a > i$, the right hand side is a factor of $V(x_{k_1}, .., x_{k_n})$.

Case 3. $j > a+1$, $i = a$. Then

$$x_{l_j} - x_{l_i} = x_{k_j} - x_{k_{a+1}}.$$

Since $j > a+1$, the right hand side is a factor of $V(x_{k_1}, .., x_{k_n})$.

Case 4. $j = a+1$, $i = a$. Then

$$x_{l_j} - x_{l_i} = x_{k_a} - x_{k_{a+1}}.$$

Since $j > a+1$, this differs from the factor $x_{k_{a+1}} - x_{k_a}$ of $V(x_{k_1}, .., x_{k_n})$ by a sign.

Thus we conclude that $V(x_{l_1}, .., x_{l_n})$ and $V(x_{k_1}, .., x_{k_n})$ differ by a sign. This proves our first assertion that $sign(L) = -sign(K)$. If

$$K \to K_1 \to \cdots \to K_m = Id$$

is a series of swaps transforming K to Id, then by the first assertion,

$$sign(K) = (-1)sign(K_1) = (-1)(-1)sign(K_2) = \cdots = (-1)^m sign(Id) = (-1)^m. \qquad \square$$

To prove the next theorem, we begin with some rudiments of functions on a finite set. Let S be a finite set, and $f : S \to \mathbf{R}$ be a function. Consider the sum

$$\sum_{x\in S} f(x)$$

which means the sum over all values $f(x)$ as x ranges over S. Let $h : S \to S$ be a one-to-one map. Then

$$\sum_{x\in S} f(h(x)) = \sum_{x\in S} f(x).$$

In the discussion below, S will be the set of permutations of n letters. The map h will be constructed below.

Theorem 5.25. *$det(A^t) = det(A)$.*

Proof: Let A be an $n \times n$ matrix. For any permutation K, we define

$$f(K) := sign(K)a_{k_1 1} \cdots a_{k_n n}$$
$$g(K) := sign(K)a_{1k_1} \cdots a_{nk_n}.$$

By definition

$$det(A) = \sum_K f(K)$$
$$det(A^t) = \sum_K g(K).$$

By applying a series of some m swaps, we can transform K to Id. For each such swap, we interchange the two corresponding neighboring a's in our product $f(K)$. After that series of m interchanges, the product $f(K)$ becomes

$$f(K) = sign(K)a_{1l_1} \cdots a_{nl_n}$$

where $L = (l_1, .., l_n)$ is the permutation we get from Id via that same series of swaps. By the Sign Theorem, Thus $sign(K) = (-1)^m = sign(L)$. So the product becomes

$$f(K) = sign(L)a_{1l_1} \cdots a_{nl_n} = g(L).$$

Now the correspondence $K \mapsto L$ is one-to-one (because K can be recovered from L by reversing the swaps). This correspondence defines a one-to-one map $h : K \mapsto L$ on the set of permutations.

By definition $g(h(K)) = f(K)$. Thus we get

$$det(A) = \sum_K f(K) = \sum_K g(h(K)) = \sum_K g(K) = det(A^t). \qquad \square$$

5.10. Homework

1. Find the determinants of

$$(a) \begin{bmatrix} -1 & 5 & 3 \\ 4 & 0 & 0 \\ 2 & 7 & 8 \end{bmatrix} \qquad (b) \begin{bmatrix} 4 & -9 & 2 \\ 4 & -9 & 2 \\ 3 & 1 & 0 \end{bmatrix}$$

$$(c) \begin{bmatrix} -1 & 2 & 3 \\ 0 & 4 & 1 \\ 0 & 0 & -2 \end{bmatrix} \qquad (d) \begin{bmatrix} 4 & 0 & 0 \\ 1 & -9 & 0 \\ 37 & 22 & 1 \end{bmatrix}$$

$$(e) \begin{bmatrix} 2 & -1 & 4 \\ 3 & 1 & 5 \\ 1 & 2 & 3 \end{bmatrix} \qquad (f) \begin{bmatrix} 1 & 1 & -2 & 4 \\ 0 & 1 & 1 & 3 \\ 2 & -1 & 1 & 0 \\ 3 & 1 & 2 & 5 \end{bmatrix}$$

2. Find the areas of the parallellograms generated by the following vectors in $\mathbf{R}^2$.

   (a) $(1, 1)$, $(3, 4)$.

   (b) $(-1, 1)$, $(0, 5)$.

3. Find the area of the parallellogram with vertices $(2, 3)$, $(5, 3)$, $(4, 5)$, $(7, 5)$.

4. Find the areas of the triangles with the following vertices in $\mathbf{R}^2$.

   (a) $(0, 0)$, $(3, 2)$, $(2, 3)$.

   (b) $(-1, 2)$, $(3, 4)$, $(-2, 1)$.

5. Find the volumes of the parallellopipeds generated by the following vectors in $\mathbf{R}^3$.

   (a) $(1,1,3)$, $(1,0,-1)$, $(1,-1,0)$.

   (b) $(-1,1,2)$, $(1,0,0)$, $(2,0,-1)$.

6. Find the areas of the parallellograms generated by the following vectors in $\mathbf{R}^3$.

   (a) $(1,2,3)$, $(3,2,1)$.

   (b) $(1,0,\pi)$, $(-1,2,0)$.

7. Find $sign(50,1,2,..,49)$, $sign(47,48,49,50,1,2,..,46)$.

8. Find $sign(k,k+1,..,n,1,2,..,k-1)$ for $1 \le k \le n$.

9. Let A be the matrix

$$\begin{bmatrix} x & y & z & t \\ y & z & t & x \\ z & t & x & y \\ t & x & y & z \end{bmatrix}.$$

   The coefficient of x^4 in $det(A)$ is _____.

   The coefficient of y^4 in $det(A)$ is _____.

   The coefficient of z^4 in $det(A)$ is _____.

   The coefficient of t^4 in $det(A)$ is _____.

   The coefficient of x^2yt in $det(A)$ is _____.

10. What is the determinant of the matrix

$$\begin{bmatrix} 1 & a & a^2 & a^3 \\ 1 & b & b^2 & b^3 \\ 1 & c & c^2 & c^3 \\ 1 & d & d^2 & d^3 \end{bmatrix}.$$

When is this matrix singular (ie. not invertible)?

11. Is there a real matrix A with

$$A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}?$$

Explain.

12. (a) Find the adjoint matrix of

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

(b) Find A^{-1} using the adjoint matrix of A.

13. Draw a picture, and use determinant to find the area of the polygon with vertices $A = (0,0)$, $B = (2,4)$, $C = (6,6)$, $D = (4,10)$. The four edges of the polygons are the line segments AB, BC, CD, and DA.

14. * Let

$$V_n = det \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ 1 & x_2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix}.$$

(a) Prove the following recursion formula:

$$V_n = (x_n - x_1)(x_n - x_2) \cdots (x_n - x_{n-1}) V_{n-1}.$$

(Hint: Subtract $x_n \times$column $(n-1)$ from column n; Subtract $x_n \times$column $(n-2)$ from column $(n-1)$, and so on.)

(b) Use (a) to show that V_n is equal to the Vandermonde function $V(x_1, .., x_n)$.

15. * Let $(x_1, y_1), .., (x_n, y_n) \in \mathbf{R}^2$ be points such that the x_i are pairwise distinct.

(a) Show that there is one and only one polynomial function f of degree at most $n-1$ whose graph passes through those given n points. In other words, there is a unique polynomial function of the form

$$f(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}$$

such that $f(x_i) = y_i$ for all i.

(b) Show that if f is a polynomial function with $f(x) = 0$ for all x, then f is the zero polynomial, i.e. the defining coefficients a_i are all zero.

(c) Find the quadratic polynomial function f whose graph passes through $(0,0), (1,2), (2,1)$.

16. * Let $(x_1, y_1), .., (x_n, y_n) \in \mathbf{R}^2$ be points such that the x_i are pairwise distinct.

(a) For each i, consider the product $f_i(x) := \prod_{1 \leq j \leq n, j \neq i} \frac{x - x_j}{x_i - x_j}$. Argue that each f_i is a polynomial function of degree at most $n-1$, and satisfies

$$f_i(x_j) = \delta_{ij}, \quad 1 \leq i, j \leq n.$$

(b) Conclude that the polynomial function given by

$$f(x) = f_1(x) y_1 + \cdots + f_n(x) y_n$$

satisfies $f(x_i) = y_i$ for all i. The formula for $f(x)$ above is called the Lagrange interpolation formula.

(c) Apply this formula to (b) in the preceding problem.

17. (With calculus) Let $f(t), g(t)$ be two functions having derivatives of all orders. Let

$$W(t) = det \begin{bmatrix} f(t) & g(t) \\ f'(t) & g'(t) \end{bmatrix}.$$

Prove that

$$W'(t) = det \begin{bmatrix} f(t) & g(t) \\ f''(t) & g''(t) \end{bmatrix}.$$

18. Show that for any scalar c, and any 3×3 matrix A,

$$det(cA) = c^3 \; det(A).$$

19. Show that for any scalar c, and any $n \times n$ matrix A,

$$det(cA) = c^n \; det(A).$$

20. Let $A, B \in \mathbf{R}^3$.

(a) If A, B are orthogonal, argue geometrically that

$$A \times B = \|A\| \; \|B\| C$$

where C is a unit vector orthogonal to both A, B.

(b) For arbitrary A, B,

$$\|A \times B\| = \sqrt{\|A\|^2 \|B\|^2 - (A \cdot B)^2}.$$

(Hint: If cA is the projection of B along A, then A and $B - cA$ are orthogonal.)

Note that the right hand side of the formula in (b) makes sense for A, B in $\mathbf{R}^n$. This expression can therefore be used to define the area of a parallellogram in $\mathbf{R}^n$.

21. Let A be a $(m+n) \times (m+n)$ matrix of the block form

$$A = \begin{bmatrix} A_1 & O_1 \\ O_2 & A_2 \end{bmatrix}$$

where A_1, A_2 are $m \times m$ and $n \times n$ matrices respectively, O_1, O_2 are the $m \times n$ and $n \times m$ zero matrix respectively. Show that

$$A = \begin{bmatrix} A_1 & O_1 \\ O_2 & I_2 \end{bmatrix} \begin{bmatrix} I_1 & O_1 \\ O_2 & A_2 \end{bmatrix}$$

where I_1, I_2 are the $m \times m$ and $n \times n$ identity matrices respectively. Use this to show that

$$det(A) = det(A_1) det(A_2).$$

Give a second proof by induction on m.

22. * Let A be an $n \times n$ matrix with $det(A) = 0$, and let A^* be its adjoint matrix. Show that

$$rank(A^*) + rank(A) \leq n.$$

Give an example to show that the two sides need not be equal. (Hint: Cramer's rule.)

23. If A is an invertible matrix, prove that $A^* = det(A)A^{-1}$.

24. Prove that for any $n \times n$ matrix A, we have $det(A^*) = det(A)^{n-1}$.

25. * Prove that for any $n \times n$ square matrices A, B, we have $(AB)^* = B^*A^*$. (Hint: Consider each entry of $(AB)^* - B^*A^*$ as a polynomial function f of the entries of A and B. Show that this function vanishes whenever $g = det(AB)$ is nonzero. Then consider the polynomial function fg. Now use the following fact from Algebra: if f, g are polynomial functions such that fg is identically zero, then either f or g is identically zero.)

26. * If A is an $n \times n$ matrix with $det(A) \neq 0$, prove that $A^{**} = det(A)^{n-2}A$. Now prove that the same holds even when $det(A) = 0$. (Hint: Use the same idea as in the preceding problem.)

27. * Let $X = (x_1, .., x_n)$, viewed as a column vector. Put $p(X) = det(I - 2XX^t)$. Prove that $p(X) = 1 - 2\|X\|^2$. (Hint: What is $p(AX)$ if A is an orthogonal matrix, and what is $E_1E_1^t$?)

28. * Let J be the $n \times n$ matrix whose (ij) entry is $\frac{1}{\|X\|^2}(\delta_{ij} - 2\frac{x_ix_j}{\|X\|^2})$ where $X = (x_1, .., x_n)$ is a nonzero vector. Prove that

$$det\ J = -\frac{1}{\|X\|^{2n}}.$$

29. * (A college math contest question) Let $a_1, .., a_n, b_1, .., b_n \in \mathbf{R}$ such that $a_i + b_j \neq 0$. Put $c_{ij} := \frac{1}{a_i + b_i}$ and let C be the $n \times n$ matrix (c_{ij}). Prove that

$$det(C) = \frac{\prod_{1 \leq i < j \leq n}(a_i - a_j)(b_i - b_j)}{\prod_{1 \leq i,j \leq n}(a_i + b_j)}.$$

6. Eigenvalue Problems

Let A be an $n \times n$ matrix. We say that A is *diagonal* if the entries are all zero except along the diagonal, ie. it has the form

$$A = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \cdots & & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix}.$$

In this case, we have

$$AE_i = \lambda_i E_i, \quad i = 1, .., n$$

ie. A transforms each of the standard vectors E_i of $\mathbf{R}^n$ by a scaling factor λ_i.

In general if A is not diagonal, but if we could find a basis $\{B_1, .., B_n\}$ of $\mathbf{R}^n$ so that

$$AB_i = \lambda_i B_i, \quad i = 1, .., n$$

then A would seem as if it were diagonal. The problem of finding such a basis and the scaling factors λ_i is called an *eigenvalue problem.*

6.1. Characteristic polynomial

Definition 6.1. *A number λ is called an eigenvalue of A if the matrix $A - \lambda I$ is singular, ie. not invertible.*

The statement that λ is an eigenvalue of A can be restated in any one of the following equivalent ways:

(a) The matrix $A - \lambda I$ is singular.

(b) The function $det(A - xI)$ vanishes at $x = \lambda$.

(c) $Null(A - \lambda I)$ is not the zero space. This subspace is called the *eigenspace* for λ.

(d) There is a nontrivial solution X to the linear system $AX = \lambda X$. Such a *nonzero* vector X is called an *eigenvector* for λ.

Definition 6.2. *The function $det(A - xI)$ is called the characteristic polynomial of A.*

What is the general form of the function $det(A-xI)$? Put $A = (a_{ij})$ and $I = (\delta_{ij})$, so that $A - xI = (a_{ij} - x\delta_{ij})$. Here

$$\delta_{ij} = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j. \end{cases}$$

The determinant $det(A - xI)$ is the sum of $n!$ terms of the form

$$sign(K)(a_{1k_1} - x\delta_{1k_1}) \cdots (a_{nk_n} - x\delta_{nk_n}) \qquad (*).$$

This term, when expanded out, is clearly a linear combination of the functions $1, x, .., x^n$. So $det(A - xI)$ is also a linear combination of those functions, ie. it is of the form

$$det(A - xI) = c_0 + c_1 x + \cdots + c_n x^n$$

where $c_0, .., c_n$ are numbers. Observe that when $K = (k_1, ..., k_n) \neq (1, 2, ..., n)$, then some of the entries $\delta_{1k_1}, .., \delta_{nk_n}$ are zero. In this case, the highest power in x occuring in the term $(*)$ is *less* than n. The only term $(*)$ where the power x^n occurs corresponds to $K = (1, 2, ..., n)$. In this case, $(*)$ is

$$(a_{11} - x) \cdots (a_{nn} - x)$$

and x^n occurs with coefficient $(-1)^n$. This shows that

$$c_n = (-1)^n.$$

On the other hand, when $x = 0$ we have $det(A - xI) = det(A)$.

Theorem 6.3. *If A is an $n \times n$ matrix, then its characteristic polynomial is of the shape*

$$det(A - xI) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1} + c_n x^n$$

where $c_0 = det(A)$ and $c_n = (-1)^n$.

Corollary 6.4. *If n is odd, then A has at least one eigenvalue.*

Proof: By the preceding theorem, $det(A-xI)$ is dominated by the leading term $(-1)^n x^n = -x^n$ as $|x| \to \infty$. In particular $det(A - xI) < 0$ for large enough x, and $det(A - xI) > 0$ for large enough $-x$. Polynomial functions are continuous functions. By the intermediate value theorem of calculus, the function $det(A - xI)$ vanishes at some x. $\square$

An eigenvalue problem is a *nonlinear* problem because finding eigenvalues involves solving the nonlinear equation

$$c_0 + c_1 x + \cdots + c_n x^n = 0.$$

However, once an eigenvalue λ is found, finding the corresponding eigenspace means solving the homogeneous *linear system*

$$(A - \lambda I)X = O.$$

Example. Let $A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$. Its characteristic polynomial is

$$det \begin{bmatrix} 1-x & 3 \\ 0 & 2-x \end{bmatrix} = (1-x)(2-x).$$

This function vanishes exactly at $x = 1, 2$. Thus the eigenvalues of A are $1, 2$.

Example. *Not every matrix has a real eigenvalue.* Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Its characteristic polynomial $x^2 + 1$. This function does not vanish for any real x. Thus A has no real-valued eigenvalue.

Exercise. Give a 3×3 matrix with characteristic polynomial $-(x^2 + 1)(x - 1)$.

Exercise. Give a 3×3 matrix with characteristic polynomial $-(x^2 + 1)x$.

Exercise. *Warning.* Row equivalent matrices don't have to have the same characteristic polynomials. Give a 2×2 example to illustrate this.

Exercise. Show that if $\lambda = 0$ is an eigenvalue of A then A is singular. Conversely, if A is singular then $\lambda = 0$ is an eigenvalue of A.

Exercise. Show that if $A = (a_{ij})$ is an upper triangular matrix, then its characteristic polynomial is

$$(a_{11} - x) \cdots (a_{nn} - x).$$

What are the eigenvalues of A?

Theorem 6.5. *Let $u_1, .., u_k$ be eigenvectors of A. If their corresponding eigenvalues are distinct, then the eigenvectors are linearly independent.*

Proof: We will do induction. Since u_1 is nonzero, the set $\{u_1\}$ is linearly independent.

Inductive hypothesis: $\{u_1, .., u_{k-1}\}$ is linearly independent.

Let $\lambda_1, .., \lambda_k$ be the distinct eigenvalues corresponding to the eigenvectors $u_1, .., u_k$. We want to show that they are linearly independent. Consider a linear relation

$$x_1 u_1 + \cdots + x_k u_k = O \qquad (*).$$

Applying A to this, we get

$$x_1 \lambda_1 u_1 + \cdots + x_k \lambda_k u_k = O.$$

Substracting from the left hand side $\lambda_k(x_1 u_1 + \cdots + x_k u_k) = O$, we get

$$x_1(\lambda_1 - \lambda_k) u_1 + \cdots + x_{k-1}(\lambda_{k-1} - \lambda_k) u_{k-1} = O.$$

By the inductive hypothesis, the coefficients of $u_1, .., u_{k-1}$ are zero. Since the λ are distinct, it follows that $x_1 = \cdots = x_{k-1} = 0$. Substitute this back into $(*)$, we get $x_k u_k = 0$. Hence $x_k = 0$. Thus $\{u_1, ..., u_k\}$ has no nontrivial linear relation, and is therefore linearly independent. $\square$

Corollary 6.6. *Let A be an $n \times n$ matrix. There are no more than n distinct eigenvalues.*

Proof: If there were $n+1$ distinct eigenvalues, then there would be $n+1$ linearly independent eigenvectors by the preceding theorem. But having $n+1$ linearly independent vectors in $\mathbf{R}^n$ contradicts the Dimension Theorem. $\square$

Corollary 6.7. *Let A be an $n \times n$ matrix having n distinct eigenvalues. Then there is a basis of $\mathbf{R}^n$ consisting of eigenvectors of A.*

Proof: Let $u_1, .., u_n$ be eigenvectors corresponding to the n distinct eigenvalues of A. By the preceding theorem these eigenvectors are linearly independent. Thus they form a basis of $\mathbf{R}^n$ by a corollary to the Dimension Theorem. $\square$

Exercise. Find the eigenvalues and bases of eigenspaces for

$$A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}.$$

6.2. Diagonalizable matrices

Definition 6.8. *An $n \times n$ matrix A is said to be diagonalizable if $\mathbf{R}^n$ has a basis consisting of eigenvectors of A. Such a basis is called an eigenbasis of A.*

Recall that if A is a diagonal matrix with diagonal entries $\lambda_1, .., \lambda_n$, then

$$AE_i = \lambda_i E_i.$$

Thus $\{E_1, .., E_n\}$ is an eigenbasis of any $n \times n$ diagonal matrix.

If A has an eigenbasis $\{B_1, .., B_n\}$ with corresponding eigenvalues $\lambda_1, .., \lambda_n$, then

$$AB_i = \lambda_i B_i.$$

Thus A behaves as if it were a diagonal matrix, albeit relative to a non-standard basis $\{B_1, .., B_n\}$. Given a matrix A, the process of finding an eigenbasis and the corresponding eigenvalues is called *diagonalization*. This process may or may not end in success depending on A.

Theorem 6.9. *Suppose A is diagonalizable with eigenbasis $\{B_1, .., B_n\}$ and corresponding eigenvalues $\lambda_1, .., \lambda_n$. Let D be the diagonal matrix with diagonal entries $\lambda_1, .., \lambda_n$, and put $B = [B_1, .., B_n]$. Then*

$$D = B^{-1}AB.$$

This is called a diagonal form of A.

Proof: Since the column vectors $B_1, .., B_n$ form a basis of $\mathbf{R}^n$, B is invertible. We want to show that $BD = AB$. Now, the ith column of BD is $B_i \lambda_i$, while the ith column of AB is AB_i. So column by column BD and AB agree. $\square$

The steps in the proof above can be reverse to prove the converse:

Theorem 6.10. *If there is an invertible matrix B such that $D = B^{-1}AB$ is diagonal, then A is diagonalizable with eigenvalues given by the diagonal entries of D and eigenvectors given by the columns of B.*

Corollary 6.11. *A is diagonalizable iff A^t is diagonalizable.*

Proof: By the first of the two theorems, if A is diagonalizable, we have

$$D = B^{-1}AB$$

where D and B are as defined above. Taking transpose on both sides, we get

$$D^t = B^t A^t (B^t)^{-1}.$$

Since D^t is also diagonal, it follows that A^t is diagonalizable by the second of the two theorems. The converse is similar. $\square$

Exercise. Find the diagonal form of the matrix A in the preceding two exercises if possible.

Exercise. Suppose A is a diagonalizable matrix with a diagonal form

$$D = B^{-1}AB.$$

Is A^2 diagonalizable? Is A^k diagonalizable? If so, what are their diagonal forms? What is $det(A)$? If A is invertible, is A^{-1} diagonalizable? If so, what is the diagonal form of A^{-1}?

It is often useful to shift our focus from eigenvectors and eigenvalues to eigenspaces. Let A be an $n \times n$ matrix. We say that a linear subspace V of $\mathbf{R}^n$ is *stabilized* by A if $AX \in V$ for all $X \in V$. Let λ be an eigenvalue of A. We write

$$V_\lambda = Null(A - \lambda I).$$

Thus V_λ is the eigenspace of A corresponding to λ. Clearly, V_λ is stabilized by A. Furthermore, as an immediate consequence of Theorem 6.5, we have

Theorem 6.12. *Let $\lambda_1, .., \lambda_k$ be pairwise distinct eigenvalues of A. Then*

$$V_{\lambda_1} + \cdots + V_{\lambda_k}$$

is a direct sum. In particular, $\dim V_{\lambda_1} + \cdots + \dim V_{\lambda_k} \leq n$.

We can now restate diagonalizability as follows.

Theorem 6.13. *A is diagonalizable iff $V_{\lambda_1} + \cdots + V_{\lambda_k} = \mathbf{R}^n$, where $\lambda_1, .., \lambda_k$ are the pairwise distinct eigenvalues of A.*

Proof: Consider the "if" part first and assume that $\mathbf{R}^n$ is the direct sum of the eigenspaces V_{λ_i} of A as stated. Let $d_i = \dim V_{\lambda_i}$. Then $d_1 + \cdots + d_k = n$. Let $u_1, .., u_{d_1}$ form a basis of V_{λ_1} and $u_{d_1+\cdots+d_{i-1}+1}, .., u_{d_1+\cdots+d_i}$ form a basis of V_{λ_i}, $i = 2, .., k$. Then it is straightforward to check that

$$u_1, .., u_n$$

form an eigenbasis of A, hence A is diagonalizable.

Conversely, suppose A is diagonalizable with eigenbasis $u_1, .., u_n$. We will show that

$$V_{\lambda_1} + \cdots + V_{\lambda_k} = \mathbf{R}^n.$$

Let $v \in \mathbf{R}^n$. It suffices to show that $v = v_1 + \cdots + v_k$ for some $v_i \in V_{\lambda_i}$, $i = 1, .., k$. Since the u_j form a basis of $\mathbf{R}^n$, we can write $v = x_1 u_1 + \cdots + x_n u_n$ for some $x_j \in \mathbf{R}$, $j = 1, .., n$. For each j, the eigenvalue corresponding to the eigenvector u_j must be in the list $\lambda_1, .., \lambda_k$, say

$$Au_j = \lambda_{j'} u_j$$

i.e. $u_j \in V_{\lambda_{j'}}$ for some $j' \in \{1, .., k\}$. This shows that each term $x_j u_j$ lies in exactly one $V_{\lambda_{j'}}$. We can group all such terms which lie in V_{λ_1} and name it v_1, i.e.

$$v_1 = \sum_{u_j \in V_{\lambda_1}} x_j u_j \in V_{\lambda_1}.$$

Likewise

$$v_i = \sum_{u_j \in V_{\lambda_i}} x_j u_j \in V_{\lambda_i}.$$

for $i = 1, .., k$. This yields

$$v = v_1 + \cdots + v_k$$

as desired. □

Exercise. Fill in the details for the "if" part in the preceding proof.

Corollary 6.14. *A is diagonalizable iff* $\dim V_{\lambda_1} + \cdots + \dim V_{\lambda_k} = n$.

Proof: This follows from the theorem and an exercise in section 4.9. □

Exercise. Find the eigenvalues and bases of eigenspaces for the following matrix, and decide if it is diagonalizable:

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

6.3. Symmetric matrices

Question. When is a matrix A diagonalizable?

We discuss the following partial answer here. Recall that a matrix A is symmetric if $A^t = A$. We will prove that if A is symmetric, then A is diagonalizable.

Exercise. *Warning.* A diagonalizable matrix does not have to be symmetric. Give an example to illustrate this.

Exercise. Prove that if A is symmetric, then

$$X \cdot AY = Y \cdot AX$$

for all vectors X, Y.

Theorem 6.15. *Let A be a symmetric matrix. If v_1, v_2 are eigenvectors of A corresponding to distinct eigenvalues λ_1, λ_2, then $v_1 \cdot v_2 = 0$.*

Proof: Since A is symmetric, we have

$$v_1 \cdot Av_2 = v_2 \cdot Av_1.$$

By assumption,

$$Av_1 = \lambda_1 v_1, \quad Av_2 = \lambda_2 v_2.$$

So we get

$$\lambda_2 v_1 \cdot v_2 = \lambda_1 v_2 \cdot v_1.$$

Since $\lambda_1 \neq \lambda_2$, we conclude that $v_1 \cdot v_2 = 0$. $\square$

Definition 6.16. *Given a symmetric matrix A, we define a function of n variables $X = (x_1, .., x_n)$ by*

$$f(X) = X \cdot AX.$$

This is called the quadratic form associated with A.

Maximum, Minimum. Let V be a linear subspace of $\mathbf{R}^n$. The quadratic form f can also be thought of as a function defined on the unit sphere S in V, i.e. the set of unit vectors in V. A quadratic form is an example of a continuous function on S. A theorem in multivariable calculus says that any continuous function on a sphere has a *maximum.* In other words, there is a unit vector P in V such that

$$f(P) \geq f(X)$$

for all unit vectors X in V. Likewise for *minimum*: there is a unit vector Q in V such that

$$f(Q) \leq f(X)$$

for all unit vectors X in V. We shall assume these facts without proof here.

Note that if λ is an eigenvalue of A with eigenvector P, we can normalize P so that it has unit length. Then

$$f(P) = P \cdot AP = \lambda P \cdot P = \lambda.$$

This shows that every eigenvalue of A is a value of f on the unit sphere in $\mathbf{R}^n$. We will see that the maximum and minimum values of f are both eigenvalues of A.

6.4. Diagonalizability of symmetric matrices

Let $A = (a_{ij})$ be an $n \times n$ symmetric matrix. We will prove that A is diagonalizable. Let V be a linear subspace of $\mathbf{R}^n$. We shall say that A *stabilizes* V if $AX \in V$ for any $X \in V$. In this case, we can think of the quadratic form f of A as a (continuous) function defined on the unit sphere $S \cap V$ in V. We denote this function by f_V.

Theorem 6.17. *(Min-Max Theorem) If A stabilizes the subspace V, and if $P \in S \cap V$ is a maximum point of the function f_V, then P is an eigenvector of A whose corresponding eigenvalue is $f(P)$. Likewise for minimum.*

Proof: We will show that

$$(*) \quad X \cdot (AP - f(P)P) = 0$$

for any $X \in \mathbf{R}^n$, but we will first reduce this to the case $X \in S \cap V \cap P^{\perp}$.

By a corollary to the Rank-nullity theorem, any vector in $\mathbf{R}^n$ can be uniquely written as $Y + Z$, with $Y \in V$ and $Z \in V^{\perp}$. Since V is stabilized by A, it follows that $AP - f(P)P \in V$, and hence (*) holds automatically for all $X \in V^{\perp}$. So, it is enough to prove (*) for all $X \in V$. Note that (*) holds automatically if X is any multiple of P (check this!) But given $X \in V$, we can further decompose it as

$$[X - (X \cdot P)P] + (X \cdot P)P.$$

Note that the first term lie in $V \cap P^{\perp}$, and the second term is a multiple of P. So, to prove (*) for all $X \in V$, it is enough to prove (*) for all $X \in V \cap P^{\perp}$. For this, we may as well assume that $X = Q$ is a unit vector. Thus, we will prove that for $Q \in S \cap V \cap P^{\perp}$, $Q \cdot AP = 0$.

Consider the parameterized curve

$$C(t) = (cos\ t)P + (sin\ t)Q.$$

Since $P, Q \in V$, this curve lies in V. Since $C(t) \cdot C(t) = 1$, we have $C(t) \in S \cap V$, $\forall t$. Let's evaluate f at $C(t)$:

$$f(C(t)) = (cos^2\ t)P \cdot AP + 2(cos\ t)(sin\ t)Q \cdot AP + (sin^2\ t)Q \cdot AQ.$$

Here we have use the fact that $Q \cdot AP = P \cdot AQ$ (why?) Since P is a maximum point of $f(X)$ on $S \cap V$, and since $C(t) \in S \cap V$, $\forall t$, we have $f(C(0)) = f(P) \geq f(C(t))$ for all t, i.e. $t = 0$ is a maximum point of the function $f(C(t))$. It follows that

$$\frac{d}{dt} f(C(t))|_{t=0} = 0.$$

Computing this derivative, we get the condition

$$Q \cdot AP = 0.$$

This completes the proof. $\Box$

Corollary 6.18. *The maximum value of the quadratic form $f(X) = X \cdot AX$ on the standard unit sphere in $\mathbf{R}^n$ is the largest eigenvalue of A. Likewise for the minimum value of f and smallest eigenvalue of A.*

Proof: By the preceding theorem, applied to the case $V = \mathbf{R}^n$, the maximum value $f(P)$ of $f(X)$ on the unit sphere is an eigenvalue of A. We know that A has at most n eigenvalues, and we saw that each one of them is of the form $f(Q)$ for some unit vector Q. But $f(P) \geq f(Q)$ by maximality of $f(P)$. So, $f(P)$ is the largest eigenvalue of A. Likewise in the minimum case. $\Box$

Exercise. Let V be a linear subspace of $\mathbf{R}^n$ and $P \in V$. Show that $V + P^{\perp} = \mathbf{R}^n$. (Hint: Show that $V^{\perp} \subset P^{\perp}$, hence $\mathbf{R}^n = V + V^{\perp} \subset V + P^{\perp}$.)

Theorem 6.19. *(Spectral Theorem) If A is symmetric, then A is diagonalizable. In fact there is an orthonormal basis of $\mathbf{R}^n$ consisting of eigenvectors of A.*

Proof: We will prove the following more general

Claim: If A stabilizes V then V has an orthonormal basis consisting of eigenvectors of A.

To prove the claim, we shall do induction on $dim\ V$. When $dim\ V = 1$, then a unit vector in V is an eigenvector of V (why?), and our claim follows in this case. Suppose the claim holds for $dim\ V = k$. Let V be a $k+1$ dimensional subspace stabilized by A.

Let P be a maximum point of the quadratic form f_V. By the Min-Max theorem, P is a (unit) eigenvector of A. By the preceding exercise, we have $\dim(V \cap P^\perp) = \dim V + \dim P^\perp - \dim(V + P^\perp) = k + 1 + n - 1 + n = k$. Observe that $V \cap P^\perp$ is stabilized by A. For if $X \in V$ and $P \cdot X = 0$, then $P \cdot AX = AP \cdot X = f(P)P \cdot X = 0$, implying that $AX \in P^\perp$. Since V is stabilized by A, we have $AX \in V$ also. Now applying our inductive hypothesis to the k dimensional subspace $V \cap P^\perp$ stabilized by A, it follows that $V \cap P^\perp$ has an orthonormal basis $P_1, .., P_k$ consisting of eigenvectors of A. This implies that $P, P_1, .., P_k$ are eigenvectors of A which form an orthonormal basis of V. $\square$

Corollary 6.20. *If A is symmetric, then A has the shape $A = BDB^t$ where B is an orthogonal matrix and D is diagonal.*

Exercise. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Find the eigenvalues of A. What are the maximum and the minimum values of the quadratic form $f(X) = X \cdot AX$ on the unit circle $\|X\| = 1$?

6.5. Homework

1. Find the characteristic polynomial, eigenvalues, and bases for each eigenspace:

(a) $\begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 2 \\ 1 & 2 & -1 \\ -1 & 1 & 4 \end{bmatrix}$.

(c) $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$.

2. Diagonalize the matrices

(a) $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$.

3. Find a 3×3 matrix A with eigenvalues $1, -1, 0$ corresponding to the respective eigenvectors $(1, 1, -1)$, $(1, 0, 1)$, $(2, 1, 1)$. How many such matrices are there?

4. Explain why you can't have a 3×3 matrix with eigenvalues $1, -1, 0$ corresponding to the respective eigenvectors $(1, 1, -1)$, $(1, 0, 1)$, $(2, 1, 0)$.

5. Show that the function
$$f(x, y) = 3x^2 + 5xy - 4y^2$$
can be written as
$$f(X) = X \cdot AX$$
where A is some symmetric matrix. Find the maximum and minimum of f on the unit circle.

6. Consider the function
$$f(x, y, z) = x^2 + z^2 + 2xy + 2yz$$
on the unit sphere in $\mathbf{R}^3$.

   (a) Find the maximum and the minimum values of f.

   (b) Find the maximum and the minimum points.

7. Let $A = \begin{bmatrix} 2 & 1 \\ -3 & 6 \end{bmatrix}$.

   (a) Find eigenvalues and corresponding eigenvectors for A.

   (b) Do the same for A^2.

   (c) Find A^{10}.

8. Let $A = \begin{bmatrix} 46 & 38 \\ -19 & -11 \end{bmatrix}$. Diagonalize A. Find a matrix B such that $B^3 = A$.

9. Suppose A is an $n \times n$ matrix such that $A^2 - A - I = O$. Find all possible eigenvalues of A.

10. (a) Find the eigenvalues of the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}?$$

(b) Find an orthogonal basis v_1, v_2, v_3 of $\mathbf{R}^3$ consisting of eigenvectors of A.

(c) What are the maximum and minimum values of the quadratic form $f(X) = X \cdot AX$ on the unit sphere $X \cdot X = 1$?

11. Let $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -2 \\ 0 & -2 & 1 \end{bmatrix}$. Find the eigenvalues and corresponding eigenvectors of A.

12. Which ones of the following matrices are diagonalizable? Explain.

$$(a) \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 5 & 6 & 7 \\ 3 & 6 & 8 & 9 \\ 4 & 7 & 9 & 0 \end{bmatrix} \qquad (b) \quad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

(c) $A + A^t$ where A is an $n \times n$ matrix.

13. Let A be square matrix.

(a) Prove that A and A^t have the same characteristic polynomial, ie.

$$det(A - xI) = det(A^t - xI).$$

(b) Conclude that A and A^t have the same eigenvalues.

14. Recall that we call two square matrices A, B similar if $B = CAC^{-1}$ for some invertible matrix C.

(a) Prove that similar matrices have the same characteristic polynomial.

(b) Conclude that similar matrices have the same eigenvalues.

15. For any real number θ, diagonalize the matrix

$$A = \begin{bmatrix} cos\ \theta & sin\ \theta \\ sin\ \theta & -cos\ \theta \end{bmatrix}.$$

16. For what value of θ, is the matrix

$$A = \begin{bmatrix} cos\ \theta & -sin\ \theta \\ sin\ \theta & cos\ \theta \end{bmatrix}$$

diagonalizable?

17. Let A and B be two square matrices of the same size. We say that A, B *commute* if $AB = BA$. Show that if A, B commute, and if X is an eigenvector of A with eigenvalue λ, then BX is also an eigenvector of A with the same eigenvalue.

18. * Let A and B be two square matrices of the same size. Show that the eigenvalues of AB are the same as the eigenvalues of BA. (Hint: Multiply $ABX = \lambda X$ by B.)

19. Let A be an invertible matrix. Show that if λ is an eigenvalue of A, then $\lambda \neq 0$, and that λ^{-1} is an eigenvalue of A^{-1}.

20. Fix a nonzero column vector $B \in \mathbf{R}^n$.

(a) Show that the $n \times n$ matrix $A = BB^t$ is symmetric.

(b) Show that B is an eigenvector of A with eigenvalue $\|B\|^2$.

(c) Show that any nonzero vector X orthogonal to B is an eigenvector of A with eigenvalues 0.

(d) Find a way to construct an eigenbasis of A.

21. * Prove that a square matrix A of rank 1 is diagonalizable iff $A^2 \neq O$. (Hint: $A = BC^t$ for some $B, C \in \mathbf{R}^n$.)

22. * Prove that if A is a square matrix of rank 1, then its characteristic polynomial is of the form $(-x)^{n-1}(-x + \lambda)$ for some number λ. (Hint: Use the preceding problem: $\lambda = C \cdot B$.)

23. * Suppose A is an $n \times n$ matrix with characteristic polynomial

$$(-1)^n (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

where $\lambda_1, .., \lambda_k$ are pairwise distinct eigenvalues of A. Prove that A is diagonalizable iff $\dim V_{\lambda_i} = m_i$ for all $1 \leq i \leq k$. (Hint: If A is diagonalizable, can you find a relationship between the eigenspaces of A and those of the diagonal form D of A?)

7. Abstract Vector Spaces

There are many objects in mathematics which behave just like $\mathbf{R}^n$: these objects are equipped with operations like vector addition and scaling as in $\mathbf{R}^n$. Linear subspaces of $\mathbf{R}^n$ are such examples, as we have seen in chapter 4. It is, therefore, worthwhile to develop an abstract approach which is applicable to the variety of cases all at once.

We have the operation of dot product for $\mathbf{R}^n$. We will also study the abstract version of this.

7.1. Basic definition

Definition 7.1. *A vector space V is a set containing a distinguished element O, called the zero element, and equipped with three operations, called vector addition, scaling, and negation. Addition takes two elements u, v of V as the input, and yields an element $u + v$ as the output. Vector scaling takes one element u of V and one number c as the input, and yields an element cu as the output. Negation takes one element u of V and yields an element $-u$ as output. The zero vector, addition, scaling, and negation are required to satisfy the following properties: for any $u, v, w \in V$ and $a, b \in \mathbf{R}$:*

V1. $(u + v) + w = u + (v + w)$

V2. $u + v = v + u$

V3. $u + O = u$

V4. $u + (-u) = O$.

V5. $a(u + v) = au + av$

V6. $(a + b)u = au + bu$

V7. $(ab)u = a(bu)$

V8. $1u = u$.

Our intuition suggests that in an abstract vector space, we should have $O = 0u$. Likewise we expect that $-u = (-1)u$, and $cO = O$ for any number c. We will prove the first equality, and leave the other two as exercises.

Proof: By V6, we have

$$0u = 0u + 0u.$$

Adding $-(0u)$ to both sides and using V4, we get

$$O = (0u + 0u) + (-(0u)).$$

Using V1, we get

$$O = 0u + (0u + (-(0u))).$$

Using V4 and V3, we get

$$O = 0u + O = 0u.$$

This completes the proof. $\square$

This shows that O can always be obtained as a special scalar multiple of a vector. Thus to specify a vector space, one need not specify the zero vector once scaling is specified.

Exercise. Prove that in a vector space V, $-u = (-1)u$ for any $u \in V$.

Exercise. Prove that in a vector space V, $cO = O$ for any scalar c.

The identity $-u = (-1)u$ shows that negation operation is actually a special case of the scaling operation. Thus to specify a vector space, one need not specify the negation operation once scaling is specified.

Example. $\mathbf{R}^n$ is a set containing the element $O = (0, .., 0)$ and equipped with the operations of addition and scaling, as defined in chapter 2. The element O and the two operations satisfy the eight properties V1-V8, as we have seen in chapter 2.

Example. Let $M(m, n)$ be the set of matrices of a given size $m \times n$. There is a zero matrix $\mathbf{O}$ in $M(m, n)$. We equip $M(m, n)$ with the usual entrywise addition and scaling. These are the operations on matrices we introduce in chapter 3. Their formal properties (see chapter 3) include V1-V8. Thus $M(m, n)$ is a vector space.

Example. Let $\mathbf{R}^{\mathbf{R}}$ be the set of functions $f : \mathbf{R} \to \mathbf{R}$, $x \mapsto f(x)$. We want to make $\mathbf{R}^{\mathbf{R}}$ a vector space. We declare the zero function $O : \mathbf{R} \to \mathbf{R}$, $x \mapsto 0$, to be our zero vector. For $f, g \in \mathbf{R}^{\mathbf{R}}$, $c \in \mathbf{R}$, we declare that $f + g$ is the function $x \mapsto f(x) + g(x)$, and cf is the function $x \mapsto cf(x)$. It remains to show that these three declared ingredients satisfy properties V1.-V8. This is a straightforward exercise to be left to the reader.

Exercise. Let U, V be vector spaces. We define $U \oplus V$ to be the set consisting of all pairs (u, v) of elements $u \in U$ and $v \in V$. We define addition and scaling on $U \oplus V$ as follows:

$$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2)$$
$$c(u, v) = (cu, cv).$$

Verify the properties V1-V8. The vector space $U \oplus V$ is called the *direct product* of U and V. Likewise, if $V_1, .., V_r$ are vector spaces, we can define their direct product $V_1 \oplus \cdots \oplus V_r$.

Definition 7.2. *Let V be a vector space. A subset W of V is called a linear subspace of V if W contains the zero vector O, and is closed under vector addition and scaling.*

Example. *A linear subspace W of a vector space V is a vector space.*

Proof: By definition, W contains a zero vector, and W being a subset of V, inherits the two vector operations from V. The properties V1-V8 hold regardless of W, because each of the properties is an equation involving the very same operations which are defined on V. Thus W is a vector space. $\square$

Example. Let $S(n)$ be the set of symmetric $n \times n$ matrices. Thus $S(n)$ is a subset of $M(n, n)$. The zero matrix $\mathbf{O}$ is obviously symmetric. The sum of $A + B$ of two symmetric matrices A, B is symmetric because

$$(A + B)^t = A^t + B^t$$

(chapter 3). The multiple cA of a symmetric matrix by a scalar c is symmetric because

$$(cA)^t = c\ A^t = cA.$$

Thus the subset $S(n)$ of $M(n,n)$ contains the zero element $\mathbf{O}$ and is closed under vector addition and scaling. Hence $S(n)$ is a linear subspace of $M(n,n)$.

Example. (With calculus) Let $C^0 \subset \mathbf{R}^{\mathbf{R}}$ be the set of continuous functions. We have the zero function O, which is continuous. In calculus, we learn that if f, g are two continuous functions, then their sum $f+g$ is continuous; and if c is a number, then the scalar multiple cf is also continuous. Thus C^0 is a vector subspace of $\mathbf{R}^{\mathbf{R}}$.

Exercise. Let P be a given $n \times n$ matrix. Let K be the set of $n \times n$ matrices A such that

$$PA = \mathbf{O}.$$

Show that K is a linear subspace of $M(n,n)$.

Exercise. (With calculus) A polynomial function f is a function of the form

$$f(t) = a_0 + a_1 t + \cdots + a_n t^n$$

where the a's are given real numbers. We denote by $\mathcal{P}$ the set of polynomial functions. Verify that $\mathcal{P}$ is a linear subspace of C^0.

Example. (With calculus) A smooth function is a function having derivatives of all order. We denote by C^∞ the set of smooth functions. Verify that C^∞ is a linear subspace of C^0.

Definition 7.3. *Let V be a vector space, $\{u_1, .., u_k\}$ be a set of elements in V, and $x_1, .., x_k$ be numbers. We call*

$$x_1 u_1 + \cdots + x_k u_k$$

a linear combination of $\{u_1, .., u_k\}$. More generally, let S be an arbitrary subset (possibly infinite) of V. If $u_1, .., u_k$ are elements of S, and $x_1, .., x_k$ are numbers, we call

$$x_1 u_1 + \cdots + x_k u_k$$

a linear combination of S.

Exercise. *Span.* (cf. chapter 4) Let V be vector space and S a subset of V. Let $Span(S)$ be the set of all linear combinations of S. Verify that $Span(S)$ is a linear subspace of V. It is called the span of S.

Exercise. Write $\begin{bmatrix} 1 & -2 \\ 1 & 0 \end{bmatrix}$ as a linear combination of $S = \{\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}\}$. Is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ a linear combination of S?

Definition 7.4. *Let $\{u_1, .., u_k\}$ be elements of a vector space V. A list of numbers $x_1, .., x_k$ is called a linear relation of $\{u_1, .., u_k\}$ if*

$$(*) \qquad x_1 u_1 + \cdots + x_k u_k = O.$$

The linear relation $0, .., 0$ (k zeros) is called the trivial relation. Abusing terminology, we often call the equation (*) a linear relation of $\{u_1, .., u_k\}$.

Example. Consider the elements $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}$ of $M(2,2)$. They have a nontrivial linear relation $1, -1, -1, -1$:

$$1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + (-1) \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + (-1) \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} + (-1) \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} = \mathbf{O}.$$

Definition 7.5. *Let S be a set of elements in a vector space V. We say that S is linearly dependent if there is a finite subset $\{u_1, .., u_k\} \subset S$ having a nontrivial linear relation. We say that S is linearly independent if S is not linearly dependent. By convention, the empty set is linearly independent.*

In the definition above, we do allow S to be an infinite set, as this may occur in some examples below. If S is finite, then it is linearly dependent iff it has a nontrivial linear relation.

Example. For each pair of integers (i,j) with $1 \le i \le m$, $1 \le j \le n$, let $E(i,j)$ be the $m \times n$ matrix whose (ij) entry is 1 and all other entries are zero. We claim that the set of all $E(i,j)$ is linearly independent. A linear relation is of the form

$$(*) \qquad \sum x_{ij} E(i,j) = \mathbf{O}$$

where we sum over all pairs of integers (i,j) with $1 \le i \le m$, $1 \le j \le n$. Note that $\sum x_{ij} E(i,j)$ is the $m \times n$ matrix whose (ij) entry is the number x_{ij}. Thus $(*)$ says that $x_{ij} = 0$ for all i, j. This shows that the set of all $E(i,j)$ is linearly independent.

Example. (With calculus) Consider the set $\{1, t, t^2\}$ of polynomial functions. What are the linear relations for this set? Let

$$x_0 \cdot 1 + x_1 \cdot t + x_2 \cdot t^2 = O.$$

Here O is the zero function. Differentiating this with respect to t twice, we get

$$2x_2 \cdot 1 = O.$$

This implies that $x_2 = 0$. So we get

$$x_0 \cdot 1 + x_1 \cdot t = O.$$

Differentiating this with respect to t once, we get

$$x_1 \cdot 1 = O,$$

which implies that $x_1 = 0$. So we get

$$x_0 \cdot 1 = O,$$

which implies that $x_0 = 0$. Thus the set $\{1, t, t^2\}$ has no nontrivial linear relation, hence it is linearly independent.

Exercise. (With calculus) Show that for any $n \geq 0$, $\{1, t, .., t^n\}$ is linearly independent. Conclude that the set of all monomial functions $\{1, t, t^2, ...\}$ is linearly independent.

Example. (With calculus) Consider the set $\{e^t,\ e^{2t}\}$ of exponential functions. What are the linear relations for this set? Let

$$(*) \quad x_1 e^t + x_2 e^{2t} = O.$$

Differentiating this once, we get

$$x_1 e^t + 2x_2 e^{2t} = O.$$

Subtracting 2 times (*) from this, we get

$$-x_1 e^t = O.$$

Since e^t is never zero, it follows that $x_1 = 0$. So (*) becomes

$$x_2 e^{2t} = O,$$

which implies that $x_2 = 0$. Thus the set $\{e^t,\ e^{2t}\}$ is linear independent.

Exercise. (With calculus) Let $a_1, .., a_n$ be distinct numbers. Show that the set $\{e^{a_1 t}, .., e^{a_n t}\}$ of exponential functions is linearly independent.

7.2. Bases and dimension

Throughout this section, V will be a vector space.

Definition 7.6. *A subset S of V is called a basis of V if it is linearly independent and it spans V. By convention, the empty set is the basis of the zero space $\{O\}$.*

Example. We have seen that the matrices $E(i,j)$ in $M(m,n)$ form a linearly independent set. This set also spans $M(m,n)$ because every $m \times n$ matrix $A = (a_{ij})$ is a linear combination of the E's:

$$A = \sum a_{ij} E(i,j).$$

Example. In an exercise above, we have seen that the set $S = \{1, t, t^2, ...\}$ of monomial functions is linearly independent. By definition, it spans the space $\mathcal{P}$ of polynomial functions. Thus S is a basis of $\mathcal{P}$.

Exercise. Let U, V be vector spaces, and let $\{u_1, .., u_r\}$, $\{v_1, .., v_s\}$ be bases of U, V respectively. Show that the set $\{(u_1, O), .., (u_r, O), (O, v_1), .., (O, v_s)\}$ is a basis of the direct product $U \oplus V$.

Lemma 7.7. *Let S be a linearly dependent set in V. Then there is a proper subset $S' \subset S$ such that $Span(S') = Span(S)$. In other words, we can remove some elements from S and still get the same span.*

Proof: Suppose $\{u_1, .., u_k\} \subset S$ has a nontrivial relation

$$x_1 u_1 + \cdots + x_k u_k = O,$$

say with $x_1 \neq 0$. Let S' be the set S with u_1 removed. Since $S' \subset S$, it follows that $Span(S') \subset Span(S)$. We will show the reverse inclusion $Span(S') \supset Span(S)$.

Let $v_1, .., v_l$ be elements in S, and $c_1, .., c_l$ be numbers. We will show that the linear combination $c_1 v_1 + \cdots + c_l v_l$ is in $Span(S')$. We can assume that the v's are all distinct. If u_1 is not one of the v's, then $v_1, .., v_l$ are all in S'. So $c_1 v_1 + \cdots + c_l v_l$ is in $Span(S')$. If u_1 is one of the v's, say $u_1 = v_1$, then

$$c_1 v_1 + \cdots + c_l v_l = c_1(-\frac{x_2}{x_1}u_2 - \cdots - \frac{x_k}{x_1}u_k) + c_2 v_2 + \cdots + c_l v_l.$$

This is a linear combination of $u_2, .., u_k, v_2, .., v_l$, which are all in S'. Thus it is in $Span(S')$. $\Box$.

Exercise. Verify that the set $\{(1,1,-1,-1),\ (1,-1,1,-1),\ (1,-1,-1,1),\ (1,0,0,-1)\}$ is linearly dependent in $\mathbf{R}^4$. Which vector can you remove and still get the same span?

Theorem 7.8. *(Finite Basis Theorem) Let S be a finite set that spans V. Then there is a subset $R \subset S$ which is a basis of V.*

Proof: If S is linearly independent, then $R = S$ is a basis of V. If S is linearly dependent, then, by the preceding lemma, we can remove some elements from S and the span of the remaining set S' is still V. We can continue to remove elements from S', while maintaining the span. Because S is finite, we will eventually reach a linearly independent set R. $\Box$

Warning. The argument above will not work if S happens to be infinite. Proving the statement of the theorem but with S being an infinite set requires foundation of set theory, which is beyond the scope of this book. What is needed is something called the *Axiom of Choice.*

Let S be a basis of V. Since V is spanned by S, every element v in V is of the form

$$(*) \quad v = a_1 u_1 + \cdots + a_k u_k$$

where the u's are distinct elements of S and the a's are numbers. Suppose that

$$v = b_1 u_1 + \cdots + b_k u_k.$$

Then we have

$$(a_1 - b_1)u_1 + \cdots + (a_k - b_k)u_k = O.$$

Since $u_1, .., u_k$ form a linearly independent set, it follows that $a_i - b_i = 0$, ie. $a_i = b_i$, for all i. This shows that, for a given v, the coefficient a_i of each element $u_i \in S$ appearing in the expression $(*)$ is unique. We call the number a_i the *coordinate* of v along u_i. (Note that when an element $u \in S$ does not occur in the expression (*), then the coordinate of v along u is 0 by definition.) We denote the coordinates of v relative to the basis S by (a_S).

Example. Every vector X in $\mathbf{R}^n$ can be written as a linear combination of the standard basis $\{E_1, ..., E_n\}$:

$$X = \sum_i x_i E_i.$$

The numbers x_i are called the standard coordinates of X.

Example. Consider the vector space $\mathcal{P}$ of all polynomial functions, and the basis $\{1, t, t^2, ...\}$. If $f(t) = a_0 + a_1 t + \cdots + a_n t^n$, then the coordinate of f along t^i is a_i for $i = 0, 1, .., n$, and is 0 for $i > n$.

Definition 7.9. *Let $k \geq 0$ be an integer. We say that V is k-dimensional if V has a basis with k vectors. If V has no finite basis, we say that V is infinite dimensional.*

Theorem 7.10. *(Uniqueness of Coefficients) Let $\{v_1, .., v_k\}$ be basis of V. Then every vector in V can be expressed as a linear combination of the basis in just one way.*

Theorem 7.11. *(Dimension Theorem) Let $k \geq 0$ be an integer. If V is k-dimensional, then the following holds:*

(a) Any set of more than k vectors in V is linearly dependent.

(b) Any set of k linearly independent vectors in V is a basis of V.

(c) Any set of less than k vectors in V does not span V.

(d) Any set of k vectors which spans V is a basis of V.

The proofs are word for word the same as in the case of a linear subspaces of $\mathbf{R}^n$ in Chapter 4.

Example. We have seen that the space $M(m,n)$ of $m \times n$ matrices has a basis consisting of the matrices $E(i,j)$. So

$$dim\ M(m,n) = mn.$$

Example. In an exercise above, we have seen that if U, V are finite dimensional vector spaces, then

$$dim(U \oplus V) = dim\ U + dim\ V.$$

Exercise. Show that the space of $n \times n$ symmetric matrices has dimension $\frac{n(n+1)}{2}$.

Exercise. (With calculus) Show that the space C^0 of continuous functions is infinite dimensional.

Exercise. Suppose that $dim\ V = k$ is finite and that W is a linear subspace of V. Show that $dim\ W \leq dim\ V$. Show that if $dim\ W = dim\ V$ then $W = V$. (Hint: The same exercise has been given in Chapter 4.)

Exercise. Suppose S, T are linear subspaces of a vector space V. Verify that $S \cap T$ is also a linear subspace of V.

7.3. Inner Products

We now discuss the abstraction of the dot product in $\mathbf{R}^n$. Recall that this is an operation which assigns a number to a pair of vectors in $\mathbf{R}^n$.

Definition 7.12. *Let V be a vector space. A function $\langle , \rangle : V \times V \to \mathbf{R}$ is called an inner product if for $v_1, v_2, v \in V$, $c \in \mathbf{R}$:*

- *(Symmetric)* $\langle v_2, v_1 \rangle = \langle v_1, v_2 \rangle$.

- *(Additive)* $\langle v, v_1 + v_2 \rangle = \langle v, v_1 \rangle + \langle v, v_2 \rangle$.

- *(Scaling)* $\langle v_1, cv_2 \rangle = c\langle v_1, v_2 \rangle$.

- *(Positive)* $\langle v, v\rangle > 0$ *if* $v \neq O$.

The additive and scaling property together say that an inner product is linear in the second slot. By the symmetric property, an inner product is also linear in the second slot. Thus one often says that an inner product is a symmetric *bilinear* form which is positive definite. In some books, the notion of a symmetric bilinear form is discussed without imposing the positivity assumption.

Example. We have seen that the dot product $\langle X, Y\rangle = X \cdot Y$ on $\mathbf{R}^n$ is an operation with all those four properties above (D1-D4 in chapter 2). Thus the dot product on $\mathbf{R}^n$ is an example of an inner product. A vector space with an inner product is an abstraction of $\mathbf{R}^n$ equipped with the dot product. Much of what we learn in this case will carry over to the general case. A vector space that comes equipped with an inner product is called an *inner product space.*

Example. (With calculus) Let V be the set of continuous functions on a fixed interval $[a, b]$. Define

$$\langle f, g\rangle = \int_a^b f(t)g(t)dt.$$

In calculus, we learn that integration has all those four properties which make $\langle,\rangle$ an inner product on the vector space V.

Example. Let V be a vector space with an inner product $\langle,\rangle$, and let W be a linear subspace. Then we can still assign the number $\langle u, v\rangle$ to every pair u, v of elements in W. This defines an inner product on W. We call this the restriction of $\langle,\rangle$ to W.

Exercise. Let $A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$. Define a new operation on $\mathbf{R}^2$: for $X, Y \in \mathbf{R}^2$,

$$\langle X, Y\rangle = X \cdot AY.$$

Verify that this is an inner product on the vector space $\mathbf{R}^2$.

Exercise. Explain why the following operation $*$ defined below on $\mathbf{R}^2$ fails to be an inner product:

$$(x_1, x_2) * (y_1, y_2) = |x_1 y_1| + |x_2 y_2|.$$

Exercise. Explain why the following operation $*$ defined below on $\mathbf{R}^2$ fails to be an inner product:

$$(x_1, x_2) * (y_1, y_2) = x_1 y_1 - x_2 y_2.$$

7.4. Lengths, angles and basic inequalities

Definition 7.13. *Let V be a vector space with inner product $\langle , \rangle$. We say that $v, w \in V$ are orthogonal if $\langle v, w \rangle = 0$. We define the length $\|v\|$ of a vector v to be the number*

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

We call v a unit element if $\|v\| = 1$. We define the distance between v, w to be $\|w - v\|$.

Throughout the following, unless stated otherwise, V will be a vector space with inner product $\langle , \rangle$.

Example. (With calculus) Let V be the space of continuous functions on the interval $[-1, 1]$, equipped with the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(t) g(t) dt.$$

Let's find the length of the constant function 1.

$$\|1\|^2 = \langle 1, 1 \rangle = \int_{-1}^{1} dt = 2.$$

Thus $\|1\| = \sqrt{2}$. The unit element in the direction of 1 is therefore the constant function $1/\sqrt{2}$.

Exercise. (With calculus) Let V be as in the preceding example. Find the length of the function t. Find the unit element in the direction of f.

Exercise. Use the symmetric and additive properties to derive the identity: for any $v, w \in V$,

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2 + 2\langle v, w \rangle.$$

Exercise. Show that if c is a number, then $\|cv\| = |c|\|v\|$. In particular $\|v - w\| = \|w - v\|$.

Exercise. Let $v, w \in V$ be any elements with $w \neq O$. Prove that $v - cw$ is orthogonal to w iff $c = \frac{\langle v,w\rangle}{\langle w,w\rangle}$. This number $c = \frac{\langle v,w\rangle}{\langle w,w\rangle}$ is called the *component* of v along w, and element cw is called the *projection* of v along w.

Exercise. (With calculus) Let V be the space of continuous functions on the interval $[-1, 1]$, equipped with the inner product as in an example above. Find the component of t along 1.

Theorem 7.14. *(Pythagoras theorem) If v, w are orthogonal elements in V, then*

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2.$$

Theorem 7.15. *(Schwarz' inequality) For $v, w \in V$,*

$$|\langle v, w\rangle| \leq \|v\| \; \|w\|.$$

Theorem 7.16. *(Triangle inequality) For $v, w \in V$,*

$$\|v + w\| \leq \|v\| + \|w\|.$$

Exercise. Prove the last three theorems by imitating the proofs in the case of $\mathbf{R}^n$.

By Schwarz inequality, we have

$$-\|v\| \; \|w\| \leq \langle v, w\rangle \leq \|v\| \; \|w\|.$$

Now for $v \neq O$, we have $\langle v, v\rangle > 0$ and so $\|v\| = \sqrt{\langle v, v\rangle} > 0$. Thus if v, w are nonzero, we can divide all three terms of the inequality above by $\|v\| \; \|w\| > 0$ and get

$$-1 \leq \frac{\langle v, w\rangle}{\|v\| \; \|w\|} \leq 1.$$

Now the function *cos* on the interval $[0, \pi]$ is a one-to-one and onto correspondence between $[0, \pi]$ and $[-1, 1]$. Thus given a value $\frac{\langle v,w\rangle}{\|v\| \ \|w\|}$ in $[-1, 1]$ there is a unique number θ in $[0, \pi]$ such that

$$cos\ \theta = \frac{\langle v, w\rangle}{\|v\| \ \|w\|}.$$

Definition 7.17. *If v, w be nonzero elements in V, we define their angle to be the number θ between 0 and π such that*

$$cos\ \theta = \frac{\langle v, w\rangle}{\|v\| \ \|w\|}.$$

Exercise. What is the angle of between v, w in V if $\langle v, w\rangle = 0$? if $v = cw$ for some number $c > 0$? if $v = cw$ for some number $c < 0$?

Exercise. What is the cosine of the angle between v, w in V if $\|v\| = \|w\| = 1$, and $\|v + w\| = \frac{3}{2}$?

7.5. Orthogonal sets

Throughout the section, V will be a vector space with inner product $\langle , \rangle$.

Let S be a set of elements in V. We say that S is *orthogonal* if $\langle v, w\rangle = 0$ for any distinct $v, w \in S$. We say that S is *orthonormal* if it is orthogonal and every element of S has length 1.

Theorem 7.18. *(Orthogonal sum) Let $\{u_1, .., u_k\}$ be an orthonormal set of elements in V, and v be a linear combination of this set. Then*

$$v = \sum_{i=1}^{k} \langle v, u_i\rangle u_i$$
$$\|v\|^2 = \sum_{i=1}^{k} \langle v, u_i\rangle^2.$$

Theorem 7.19. *(Best approximation I) best approximation Let $\{u_1, .., u_k\}$ be an orthonormal set, and $v \in V$. Then*

$$\|v - \sum_{i=1}^{k} \langle v, u_i \rangle u_i\| < \|v - \sum_{i=1}^{k} x_i u_i\|$$

for any $(x_1, .., x_k) \neq (\langle v, u_1 \rangle, .., \langle v, u_k \rangle)$.

Theorem 7.20. *(Best approximation II) Let U be a linear subspace of V, and $v \in V$. Then there is a unique vector $u \in U$ such that*

$$\|v - u\| < \|v - w\|$$

for all $w \in U$ not equal to u. The point u is called the projection of v along U.

Theorem 7.21. *(Bessel's inequality) Let $\{u_1, .., u_k\}$ be an orthonormal set, and v be any vector in V. Then*

$$\sum_{i=1}^{k} \langle v, u_i \rangle^2 \leq \|v\|^2.$$

Exercise. Prove the last four theorems by imitating the case of $\mathbf{R}^n$ in Chapters 2 and 4.

Exercise. Let $\{u_1, .., u_k\}$ be an orthonormal set. Prove that

$$\langle \sum_{i=1}^{k} x_i u_i, \sum_{i=1}^{k} y_i u_i \rangle = \sum_{i=1}^{k} x_i y_i.$$

Exercise. Let V be the space of continuous functions on $[0, 1]$. Verify that that functions 1 and $1 - 2t$ are orthogonal with respect to the inner product $\int_0^1 fg$. Write $1 + t$ as a linear combination of $\{1, 1 - 2t\}$.

Exercise. (With calculus) Let V be the space of continuous functions on $[-\pi, \pi]$ with inner product given in an exercise above. Verify that *cos* t and *sin* t are orthogonal with respect to this inner product. Find the linear combination of *cos* t, *sin* t which best approximate t.

Exercise. (With calculus) Let f be a continuous function on the interval $[-1,1]$. It is said to be odd if $f(-t) = -f(t)$ for all t. It is said to be even if $f(-t) = f(t)$ for all t. Show that every even function is orthogonal to every odd function, with respect to the inner product $\int_{-1}^{1} fg$.

7.6. Orthonormal bases

Throughout this section, V will be a finite dimensional vector space with an inner product $\langle , \rangle$. We will see that the *Gram-Schmidt orthogonalization process* for $\mathbf{R}^n$ carries over to a general inner product space quite easily. All we have to do is to replace the dot product by an abstract inner product $\langle , \rangle$.

Lemma 7.22. *If $\{u_1, .., u_k\}$ is a set of nonzero vectors in V which is orthogonal, then V is linearly independent.*

Proof: Consider a linear relation

$$\sum_{i=1}^{k} x_i u_i = O.$$

Take the inner product of both sides with u_j, we get

$$x_j \langle u_j, u_j \rangle = 0.$$

Since $u_j \neq O$, it follows that $x_j = 0$. This holds for any j. $\square$

Let $\{v_1, .., v_n\}$ be a basis of V. Thus *dim* $V = n$. From this, we will construct an orthogonal basis $\{v_1', .., v_n'\}$. Note that to get an orthonormal basis from this, it is enough to normalize each element to length one.

Put

$$v_1' = v_1.$$

It is nonzero, so that the set $\{v_1'\}$ is linearly independent. We adjust v_2 so that we get a new element v_2' which is nonzero and orthogonal to v_1'. More precisely, let $v_2' = v_2 - cv_1'$ and demand that $\langle v_2', A_1'\rangle = 0$. This gives $c = \frac{\langle v_2, v_1'\rangle}{\langle v_1', v_1'\rangle}$. Thus we put

$$v_2' = v_2 - \frac{\langle v_2, v_1'\rangle}{\langle v_1', v_1'\rangle} v_1'.$$

Note that v_2' is nonzero, for otherwise v_2 would be a multiple of $v_1' = v_1$. So, we get an orthogonal set $\{v_1', v_2'\}$ of nonzero elements in V.

We adjust v_3 so that we get a new vector v_3' which is nonzero and orthogonal to v_1', v_2'. More precisely, let $v_2' = v_3 - c_2 v_2' - c_1 v_1'$ and demand that $\langle v_3', v_1'\rangle = \langle v_3', v_2'\rangle = 0$. This gives $c_1 = \frac{\langle v_3, v_1'\rangle}{\langle v_1', v_1'\rangle}$ and $c_2 = \frac{\langle v_3, v_2'\rangle}{\langle v_2', v_2'\rangle}$. Thus we put

$$v_3' = v_3 - \frac{\langle v_3, v_2'\rangle}{\langle v_2', v_2'\rangle} v_2' - \frac{\langle v_3, v_1'\rangle}{\langle v_1', v_1'\rangle} v_1'.$$

Note that v_3' is also nonzero, for otherwise v_3 would be a linear combination of v_1', v_2'. This would mean that v_3 is a linear combination of v_1, v_2, contradicting linear independence of $\{v_1, v_2, v_3\}$.

More generally, we put

$$v_k' = v_k - \sum_{i=1}^{k-1} \frac{\langle v_k, v_i'\rangle}{\langle v_i', v_i'\rangle} v_i'$$

for $k = 1, 2, .., n$. Then v_k' is nonzero and is orthogonal to $v_1', .., v_{k-1}'$, for each k. Thus the end result of Gram-Schmidt is an orthogonal set $\{v_1', .., v_n'\}$ of nonzero vectors in V. By the lemma above, this set is linearly independent. Since $\mathit{dim}\ V = n$, this set is a basis of V. We have therefore proven

Theorem 7.23. *Every finite dimensional inner product space has an orthonormal basis.*

Exercise. (With calculus) Let V be the space of functions spanned by $\{1, t, t^2, t^3\}$ defined on the interval $[-1, 1]$. We give this space our usual inner product:

$$\langle f, g\rangle = \int_{-1}^{1} fg.$$

Apply Gram-Schmidt to the basis $\{1, t, t^2, t^3\}$ of V.

7.7. Orthogonal complement

In this section, V will continue to be a finite dimensional vector space with a given inner product $\langle\ ,\ \rangle$.

Definition 7.24. *Let W be a linear subspace of V. The orthogonal complement of W is the set*

$$W^{\perp} = \{v \in V | \langle v, w\rangle = 0,\ \forall w \in W\}.$$

Theorem 7.25. *$W^{\perp}$ is a linear subspace of V. Moreover $W \cap W^{\perp} = \{O\}$.*

Theorem 7.26. *If $W^{\perp} = \{O\}$, then $W = V$.*

Theorem 7.27. *(a) Every vector $v \in V$ can be written uniquely as $v = w + x$ where $w \in W,\ x \in W^{\perp}$.*

(b) $dim(V) = dim(W) + dim(W^{\perp})$.

(c) $(W^{\perp})^{\perp} = W$.

Exercise. Prove the last three theorems by imitating the case of $\mathbf{R}^n$.

• **Warning.** The second and third theorems above do not hold for infinite dimensional inner product spaces in general. However, the first theorem does hold in general.

7.8. Homework

1. Let S, T be two linear subspaces of a vector space V. Define the set $S + T$ by

$$S + T = \{s + t | s \in S,\quad t \in T\}.$$

(a) Show that $S+T$ is a linear subspace of V. It is called the *sum* of S,T.

(b) We say that $S+T$ a *direct sum* if

$$s \in S,\ t \in T,\ s+t=O \implies s=t=O.$$

Show that $S+T$ is a direct sum iff

$$S \cap T = \{O\}.$$

If S,T are finite dimentional, show that $S+T$ is a direct sum iff

$$dim(S+T) = dim(S) + dim(T).$$

(c) Likewise, if $V_1,..,V_r$ are linear subspaces of V, then we can define their sum $V_1+\cdots+V_r$. Write down the definition.

(d) Guess the right notion of the direct sum in this case.

2. Let $S \subset V$ be a linear subspace of a finite dimensional vector space V. Show that there is a subspace $T \subset V$ such that $S \cap T = \{O\}$ and $S+T=V$. The space T is called a *complementary* subspace of S in V.

3. Formulate and prove the abstract versions of Theorems 4.31 and 4.32.

4. Let V be the space of $n \times n$ matrices. Define the trace function $tr : V \to \mathbf{R}$,

$$tr(A) = a_{11} + \cdots + a_{nn}$$

where $A = (a_{ij})$.

(a) Define $\langle\ ,\ \rangle : V \times V \to \mathbf{R}$ by $\langle A,B\rangle = tr(A^tB)$. Show that this defines an inner product on V.

(b) Carry out Gram-Schmidt for the space of 2×2 matrices starting from the basis consisting of

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},\ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},\ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix},\ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

(c) Give an example to show that $tr(AB)$ does not define an inner product on V.

5. Let $X, Y \in \mathbf{R}^n$ be column vectors, and A be any $n \times n$ matrix.

(a) Show that

$$tr(XY^t) = X \cdot Y.$$

(b) Show that if $K = XX^t$, then

$$tr(KAKA) = (tr(KA))^2.$$

In the following exercises, V will be a finite dimensional vector space with a given inner product $\langle\ ,\ \rangle$.

6. (With calculus) Let V be the space of continuous functions on the interval $[-1, 1]$ with the usual inner product. Find the best approximation of the function e^t as a linear combination of $1, t, t^2, t^3$.

7. Consider an inhomogeneous linear system

$$(*) \qquad AX = B$$

in n variables $X = (x_1, .., x_n)$. Let X_0 be a given solution.

(a) Show that if $AY = O$, then $X = X_0 + Y$ is a solution to $(*)$.

(b) Show that conversely, every solution X to $(*)$ is of the form $X_0 + Y$ where $AY = O$.

(c) Let S be the solution set to $(*)$. Give an example to show that S is *not* a linear subspace of $\mathbf{R}^n$ unless $B = O$.

(d) Define new addition $\oplus$ and new scaling $\otimes$ on S by

$$(X_0 + Y_1) \oplus (X_0 + Y_2) = X_0 + Y_1 + Y_2$$
$$c \otimes (X_0 + Y) = X_0 + cY.$$

Verify that S is a vector space under these two operations.

(e) Show that if $\{Y_1, ..., Y_k\}$ is a basis of $Null(A)$, then $\{X_0 + Y_1, ..., X_0 + Y_k\}$ is a basis of S above.

8. Let X_0 be a given point in $\mathbf{R}^n$. Define a new addition $\oplus$ and scaling $\otimes$ on the set $\mathbf{R}^n$ so that
$$X_0 \oplus X_0 = X_0$$
$$c \otimes X_0 = X_0$$
for all scalar c. Thus X_0 is the "origin" of $\mathbf{R}^n$ relative to these new operations.

9. Let S be any finite set. Let V be the set of real-valued functions on S. Thus an element f of V is a rule which assigns a real number $f(s)$ to every element $s \in S$. Prove the following.

   (a) V has the structure of a vector space.

   (b) Now suppose that $S = \{1, 2, .., n\}$ is the list of the first n integers. Show that there is a one-to-one correspondence between V and $\mathbf{R}^n$, namely a function f corresponds to the vector $(f(1), .., f(n))$ in $\mathbf{R}^n$. You will show that this allows you to *identify* V with $\mathbf{R}^n$ as a vector space.

   (c) The zero function corresponds to $(0, ..., 0)$.

   (d) If the functions f, g correspond respectively to the vectors X, Y, then the function $f + g$ corresponds to $X + Y$.

   (e) If the functions f correspond to the vectors X, then the function cf corresponds to cX for any scalar c.

   (f) Show that, in general, $dim(V)$ is the number of elements in S.

10. Continue with the preceding exercise. Suppose that S is the set of integer pairs (i, j) with $i = 1, .., m$, $j = 1, .., n$, and V is the set of real-valued functions on S. Can you identify V to a vector space you have studied?

In the following, unless stated otherwise, V is a vector space with inner product $\langle , \rangle$ and length function $\|A\| = \sqrt{\langle A, A\rangle}$.

11. Let A, B, C, D be vectors in V.

(a) If $\|C\|^2 = \|D\|^2 = 1$, and $\|C+D\|^2 = 3/2$, find the cosine of the angle between C, D.

(b) Show that if A, B are orthogonal, then $\|A - B\| = \|A + B\|$.

(c) Show that if $\|A - B\| = \|A + B\|$, then A, B are orthogonal.

12. Prove that if A is orthogonal to every vector in V, then $A = O$.

13. Suppose A, B are nonzero elements in V. Prove that $A = cB$ for some number c iff $|\langle A, B\rangle| = \|A\|\|B\|$.

14. Let A, B be any vectors in V. Prove that

(a) $\|A + B\|^2 + \|A - B\|^2 = 2\|A\|^2 + 2\|B\|^2$.

(b) $\|A - B\|^2 = \|A\|^2 + \|B\|^2 - 2\|A\|\ \|B\|\ cos\ \theta$ where θ is the angle between A and B.

15. * Suppose A, B are nonzero vectors in V. Prove that $A = cB$ for some number $c > 0$ iff $\|A + B\| = \|A\| + \|B\|$.

16. Let c be the component of A along B in V. Prove that

$$\|A - cB\| \leq \|A - xB\|$$

for any number x. That is, c is the number that minimizes $\|A - cB\|$.

17. * Let A be a symmetric $n \times n$ matrix. Define a new operation on $\mathbf{R}^n$ by

$$\langle X, Y\rangle = X \cdot AY.$$

(a) Show that $\langle , \rangle$ is symmetric and bilinear.

(b) A is said to be *positive definite* if $X \cdot AX > 0$ for any nonzero vector X in $\mathbf{R}^n$. Show that for any $n \times n$ matrix B, the matrix $B^t B$ is positive definite iff B is invertible.

(c) Prove that a symmetric A is positive definite iff all its eigenvalues are positive.

(d) Show that $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$ is positive definite iff $a + d > 0$ and $ad - b^2 > 0$.

18. * (You should do problem 3 first before this one.) Let M be a given symmetric positive definite $n \times n$ matrix. Let V be the space of $n \times n$ matrices. Define $\langle \;,\; \rangle : V \times V \to \mathbf{R}$ by $\langle A, B \rangle = tr(A^t M B)$. Show that this defines an inner product on V.

19. Let V be a finite dimensional vector space with an inner product $\langle , \rangle$, and V_1, V_2 be linear subspaces which are orthogonal, ie. for any $v \in V_1$, $u \in V_2$, we have $\langle v, u \rangle = 0$. Show that $V_1 \cap V_2 = \{O\}$.

20. Let V be a finite dimensional vector space with an inner product $\langle , \rangle$, and $V_1, .., V_k$ be linear subspaces which are pairwise orthogonal, ie. if $i \neq j$, then for any $v \in V_i$, $u \in V_j$, we have $\langle v, u \rangle = 0$. Show that if $V = V_1 + \cdots + V_k$, then
$$dim(V) = dim(V_1) + \cdots + dim(V_k).$$

21. * Prove that any $n \times n$ matrix A satisfies a polynomial relation
$$a_0 I + a_1 A + \cdots + a_N A^N = O$$
for some numbers $a_0, .., a_N$, not all zero. (Hint: What is the dimension of $M(n, n)$?)

22. A subset C of a vector space V is said to be convex if for any vectors $u, v \in C$, the line segment connecting u, v lies in C. That is,
$$tu + (1 - t)v \in C$$

for $0 \leq t \leq 1$. Show that in $V = M(n,n)$, the set of all positive definite matrices is convex.

23. * Let U, V be linear subspaces of a finite dimensional inner product space W with $dim\ U = dim\ V$. Thus we have the direct sums $U + U^\perp = V + V^\perp = W$. Let $\pi_U, \pi_{U^\perp}$ be the orthogonal projection map from W onto $U, U^\perp$ respectively. Show that $\pi_U V = U$ iff $\pi_{U^\perp} V^\perp = U^\perp$. (Hint: Note that $Ker\ \pi_U = U^\perp$, $Ker\ \pi_{U^\perp} = U$, and that $\pi_U V = U \Leftrightarrow Ker\ \pi_U \cap V = 0$. Use $(A+B)^\perp = A^\perp \cap B^\perp$.)

24. Let U, V be linear subspaces of a finite dimensional vector space W with $dim\ U = dim\ V$. Let U', V' be complementary subspaces of U, V respectively, in W, so that we have the direct sums $U + U' = V + V' = W$. Let $\pi_U, \pi_{U'}$ be the projection map from W onto U, U' respectively with respect to the direct sum $U + U'$. Give a counterexample to show that $\pi_U V = U$ does not always imply $\pi_{U'} V' = U'$. (Hint: Consider 4 lines U, U', V, V' in $\mathbf{R}^2$ with $V' = U$.)