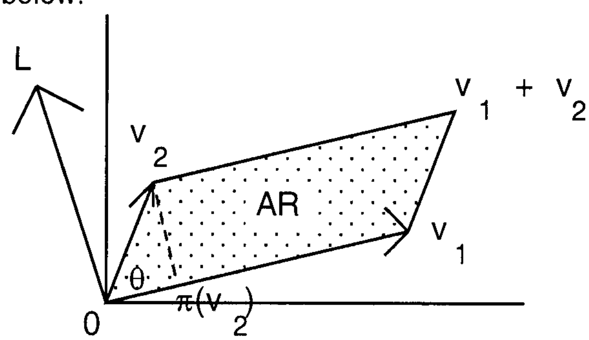
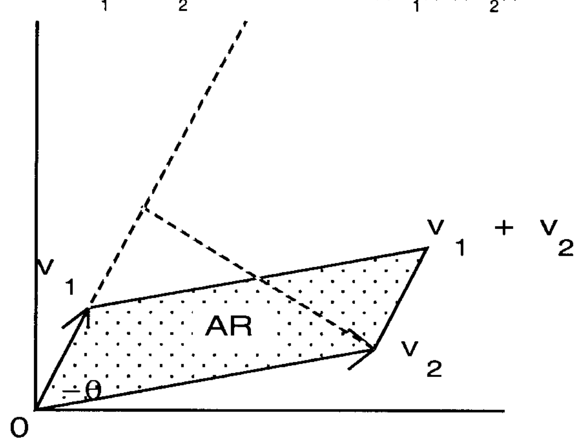
Lecture #3

Determinants

Consider the rows of a 2x2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ as vectors in R^2, say v_1 and v_2. The matrix A is non-singular or invertible if and only if these vectors are linearly independent, which means that they do not lie on the same line. They do not lie on the same line if and only if the parallelogram they generate has positive area, so that this area may be interpreted as a measure of the invertibility of A. The parallelogram v_1 and v_2 generate and its area, AR, are pictured in the diagram below.



This area, AR, equals the length of v_1 time the distance from v_2 to the line through v_1. That distance is the length of the dotted line in the diagram. The length of the dotted line is $||v_2||\sin(\theta)$, where $||v_2||$ is the length of v_2 and θ is the angle from v_1 to v_2 going in the counterclockwise direction. Therefore the area is $AR = ||v_1||\ ||v_2|| \sin\theta$. Notice that this definition of the area can be negative. For instance, if we reverse the labelling of v_1 and v_2, as in the next diagram, then the angle from v_1 to v_2 is $-\theta$, so that the area of the parallelogram generated by v_1 to v_2 is $AR = ||v_1||\ ||v_2|| \sin(-\theta) = -||v_1||\ ||v_2|| \sin\theta < 0$. We might therefore



be tempted to define the area to be the absolute value of $||v_1|| \, ||v_2|| \sin\theta$, namely $\left| ||v_1|| \, ||v_2|| \sin\theta \right|$. If we use this definition, however, we lose an advantage of $||v_1|| \, ||v_2|| \sin\theta$, which is that it respects the algebraic structure of R^2. Since the matrix A is used to do algebra, it makes intuitive sense to have any measure of its degree of invertibility respect its algebraic structure.

I show in what sense the formula $||v_1|| \, ||v_2|| \sin\theta$ respects the algebraic structure of R^2, by finding another formula for the signed area of the parallelogram determined by v_1 and v_2. Let π be the orthogonal projection of R^2 onto the one dimensional subspace that is the line through 0 and v_1 and let L be the vector of unit length that makes an angle of $90°$ with v_1 when the angle is measured going from v_1 to L in the counterclockwise direction. The vector $v_2 - \pi(v_2)$ is parallel to L and either points in the same direction as L or the opposite direction. Therefore $v_2 - \pi(v_2) = tL$, for some number t. Since L has unit length, $|t|$ equals the length of $v_2 - \pi(v_2)$. If $\sin\theta > 0$, then L and $v_2 - \pi(v_2)$ point in the same direction and $t > 0$. If $\sin\theta < 0$, then L and $v_2 - \pi(v_2)$ point in opposite directions and $t < 0$. If $\sin\theta = 0$, then $v_2 - \pi(v_2) = 0$ and $t = 0$. Since $||v_2|| \sin\theta = t$ and $L.\pi(v_2) = 0$, we have that

$$||v_1|| \, ||v_2|| \sin\theta = ||v_1||t = ||v_1||t(L.L) = ||v_1||[L.(v_2 - \pi(v_2))] = ||v_1||(L.v_2).$$

Therefore our algebraic expression for the signed area of the parallelogram determined by v_1 and v_2 is $||v_1||(L.v_2)$, which we see is linear with respect to v_2 when v_1 is held fixed.

We could just as well have held v_2 fixed and let L be a vector of unit length perpendicular to v_2, so that $||v_1|| \, ||v_2|| \sin\theta = ||v_2||(L.v_1)$. We then see that f is linear with respect to v_1 when v_2 is held fixed.

Let $f(v_1, v_2) = ||v_1|| \, ||v_2|| \sin\theta$. We have seen that f is linear with respect to each of v_1 and v_2 separately. Such a function is said to be <u>bilinear</u>. Since f is bilinear,

$$f(0, v_2) = f(-0, v_2) = -f(0, v_2),$$

so that $f(0, v_2) = 0$, for any v_2. Similarly $f(v_1, 0) = 0$, for any v_1. It should be clear that $f(v_1, v_2) = -f(v_2, v_1)$, for all 2-vectors v_1 and v_2, since the sin of the angle going from v_1 to v_2 is minus the sin of the angle going from v_2 to v_1. If $f(v_1, v_2) = -f(v_2, v_1)$, for all v_1 and v_2, f is said to be <u>alternating</u>. Notice that if f is alternating, then $f(v, v) = 0$, for any 2-vector v, which certainly must be true of the area of the parallelogram generated by v and v. Notice also that since f is bilinear and alternating, then for any number b and any 2-vectors v_1 and v_2,

$$f(v_1 + bv_2, v_2) = f(v_1, v_2) + bf(v_2, v_2) = f(v_1, v_2) + b0 = f(v_1, v_2) .$$

I now use the facts that f is alternating and bilinear to derive a formula for $f(v_1, v_2)$. Recall that v_1 and v_2 are the first and second rows, respectively, of the matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Assume for the moment that $a_{11} \neq 0$. Then by subtracting $a_{11}^{-1}a_{21}$ Times the first row from the second, we can change A to $\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} - a_{11}^{-1}a_{21}a_{12} \end{pmatrix}$. Suppose that $a_{22} - a_{11}^{-1}a_{21}a_{12} \neq 0$. Then by subtracting $(a_{22} - a_{11}^{-1}a_{21}a_{12})^{-1}a_{12}$ times the second row of this matrix from the first, we obtain the matrix $\begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} - a_{11}^{-1}a_{21}a_{12} \end{pmatrix}$. Since f is bilinear and alternating, we conclude that

$$f((a_{11}, a_{12}), (a_{21}, a_{22})) = f((a_{11}, a_{12}), (0, a_{22} - a_{11}^{-1}a_{21}a_{12}))$$

$$= f((a_{11}, 0), (0, a_{22} - a_{11}^{-1}a_{21}a_{12}))$$

$$= a_{11}f((1, 0), (0, a_{22} - a_{11}^{-1}a_{21}a_{12}))$$

$$= a_{11}(a_{22} - a_{11}^{-1}a_{21}a_{12})f((1, 0), (0, 1))$$

$$= (a_{11}a_{22} - a_{21}a_{12})f((1, 0), (0, 1)) .$$

If $a_{22} - a_{11}^{-1}a_{21}a_{12} = 0$, so that $a_{11}a_{22} - a_{21}a_{12} = 0$, then

$$f((a_{11}, a_{12}), (a_{21}, a_{22})) = f((a_{11}, a_{12}), (0, 0)) = 0$$

$$= (a_{11}a_{22} - a_{21}a_{12})f((1, 0), (0, 1)) .$$

If $a_{11} = 0$ and $a_{12} \neq 0$, then by subtracting $a_{12}^{-1}a_{22}$ times the first row of A from the second, we see that

$$f((a_{11}, a_{12}), (a_{21}, a_{22})) = f((0, a_{12}), (a_{21}, a_{22}))$$
$$= f((0, a_{12}), (a_{21}, 0)) = a_{12}a_{21} f((0, 1), (1, 0))$$
$$= -a_{21}a_{12} f((1, 0), (0, 1)) = (a_{11}a_{22} - a_{21}a_{12}) f((1, 0), (0, 1)) .$$

If the first row of A is zero, then

$$f((a_{11}, a_{12}), (a_{21}, a_{22})) = f((0, 0), (a_{21}, a_{22})) = 0$$
$$= (a_{11}a_{22} - a_{21}a_{12}) f((1, 0), (0, 1)) .$$

So in every case, the equation

$$f((a_{11}, a_{12}), (a_{21}, a_{22})) = (a_{11}a_{22} - a_{21}a_{12}) f((1, 0), (0, 1))$$

holds. Since f((1, 0), (0, 1)) is the area defined by the natural basis vectors of R^2 with the first vector coming first, we know that it is plus the area of a unit square and hence equals +1. Therefore $f((a_{11}, a_{12}), (a_{21}, a_{22})) = a_{11}a_{22} - a_{21}a_{12}$. The expression $a_{11}a_{22} - a_{21}a_{12}$ is called the determinant of the 2x2 matrix A, and is the signed area of the parallelogram generated by the rows of A. The sign is positive if one goes counterclockwise from the first row to the second to reach the second via the smallest angle between these two vectors. If one must go clockwise to go through the smallest angle, the area is negative. If there is no smallest angle, because the rows point in the same or opposite directions, then the area is 0.

I turn from the intuition behind the definition of the determinant to its formal definition. Let A be an NxN matrix and consider the rows of A to be N-vectors, that is, as belonging to R^N. The determinant is an alternating multilinear form from R^N to R, where these terms are defined as follows.

Definition: If S is a set and K is a positive integer, let

$$S^K = \underbrace{S \times S \times .... \times S}_{\leftarrow \text{ K times } \rightarrow},$$

where "$\times$" is the symbol for the Cartesian product defined in lecture 2.

The determinant is a function from $(R^N)^N$ to R, that is, a real-valued function defined on the Cartesian product of R^N with itself N times.

Definition: A multilinear form on a vector space V is a function $f: V^K \rightarrow R$ such that, for $k = 1, ... , K$, $f(v_1, ... , v_{k-1}, v_k, v_{k+1}, ... , v_K)$ is a linear function of v_k when v_n is held fixed for $n \neq k$.

Definition: A multilinear form $f: V^K \to R$ is alternating if

$$f(v_1, \ldots, v_{k-1}, v_k, v_{k+1}, \ldots, v_{n-1}, v_n, v_{n+1}, \ldots, v_K)$$

$$= -f(v_1, \ldots, v_{k-1}, v_n, v_{k+1}, \ldots, v_{n-1}, v_k, v_{n+1}, \ldots, v_K),$$

for any $k < n$. That is, interchanging two variables of f changes its sign.

Lemma 3.1: If $f: V^K \to R$ is an alternating, multilinear form, then

1) $f(v_1, \ldots, v_N) = 0$, if $v_k = v_n$, where $k \neq n$, and

2) $f(v_1, \ldots v_{n-1}, cv_k + v_n, v_{n+1}, \ldots, v_K) = f(v_1, \ldots, v_K)$, if $k \neq n$ and c is a number.

Proof: The proof was given earlier when describing the determinant as an alternating bilinear form on R^2. ■

Consider the N rows of an NxN matrix A to be N-successive vectors in R^N, so that the set of all NxN matrices may be viewed as the space $(R^N)^N$. A member of $(R^N)^N$ will be written as A, where A is an NxN matrix.

Theorem 3.2: If $f: (R^N)^N \to R$ is an alternating multilinear form, then f(A) = (det A)f(I), where I is the NxN identity matrix, and

$$\det A = \sum_{\substack{\sigma \text{ is a permutation} \\ \text{of } \{1, \ldots, N\}}} (\operatorname{sgn} \sigma)\, a_{1,\sigma(1)}\, a_{2,\sigma(2)} \ldots a_{N,\sigma(N)}.$$

Proof: No formal will be given of this theorem. It is proved by using elementary row operations to reduce A to the identity matrix and keeping track of what happens to the value of f at each step. An argument of this kind was made above in discussing the definition of the determinant for 2x2 matrices. ■

The function det A is called the determinant if A. This theorem asserts that the determinant of A is the unique alternating multilinear form on $(R^N)^N$ that assigns the value 1 to the identity matrix I.

In case the reader does not know the definitions of a permutation or its sign, I explain these.

Definition: A permutation of {1, ... , N} is a one to one and onto function from {1, ... , N} to {1, ... , N}.

Every permutation can be expressed as a succession of interchanges of neighboring pairs. For instance, consider the permutation that takes (1, 2, 3, 4) to (4, 1, 2, 3). This may be achieved as the succession of pairwise switches (1, 2, 3, 4) to (1, 2, 4, 3) to (1, 4, 2, 3) to (4, 1, 2, 3). There are many ways to express one permutation as a succession of pairwise interchanges, but it can be proved that for a given permutation the number of pairwise interchanges needed is either always odd or always even. (No proof is given here.) A permutation is said to be odd if the number of interchanges is odd. Otherwise the permutation is said to be even. The sign of a permutation σ is defined to be one if σ is even and -1 if σ is odd. The sign of σ is denoted by sgn σ.

The determinant of an NxN matrix A may be described verbally as follows. Pick one entry from each row, every time from a different column and form the product of these N numbers. The choice of columns defines a permutation of $\{1, \ldots, N\}$. Multiply the product by the sign of this permutation. The determinant of A is the sum of these signed products.

If we return to theorem 3.2, we see that the determinant is the only alternating multilinear form on $(R^N)^N$ that assigns the value 1 to the NxN identity matrix I. This makes sense geometrically, because 1 is the volume of the N-dimensional cube with edges all of length 1, and the parallelepiped determined by the N natural basis vectors of R^N is such a cube. The determinant of an NxN matrix A is the signed volume of the N-dimensional parallelepiped determined by the N rows of A. Reflections like those made above in the two dimensional case show that this volume should be an alternating multilinear form on the rows of A. The determinant is the volume with a positive or negative sign, and it is nearly impossible to visualize this sign in dimensions higher than 3. Perhaps the best way to think about it is to assume that the volume determined by the standard basis vectors is positive if these are taken in their natural order. The sign of all other volumes is then determined by the alternating nature of the multilinear form that is the determinant.

The following proposition is a consequence of the fact that the determinant is an alternating multilinear form.

Proposition 3.3: Let A be an NxN matrix.

1) If B is obtained from A by multiplying one row of A by a number c, then det B = c(det A).

2) If B is obtained from A by adding a multiple of one row to another, then det B = det A.

3) If B is obtained from A by interchanging two rows of A, then det B = –det A.

4) If two rows of A are equal, then det A = 0.

Theorem 3.4: if A and B are NxN matrices, then det AB = (det A)(det B).

Proof: This theorem is proved by verifying that the function f(A) = det AB is an alternating multilinear form on $(R^N)^N$. Theorem 3.2 then implies that f(A) = (det A)f(I), where I is the NxN identity matrix. But f(I) = det IB = det B, so that det AB = (det A)f(I) = (det A)(det B). ∎

Corollary 3.5: If A is an invertible matrix, then $\det A \neq 0$ and $\det(A^{-1}) = (\det A)^{-1}$.

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

Definition: If A is an MxN matrix, the transpose of A, denoted A^T, is the NxM matrix with $(m, n)^{th}$ entry equal to the $(n, m)^{th}$ entry of A.

Example: $\begin{pmatrix} 6 & 1 \\ 4 & 2 \\ 2 & 5 \end{pmatrix}^T = \begin{pmatrix} 6 & 4 & 2 \\ 1 & 2 & 5 \end{pmatrix}$.

Theorem 3.6: If A is an NxN matrix, then $\det A^T = \det A$.

Proof: If σ is a permutation of $\{1, \ldots, N\}$, then $\operatorname{sgn} \sigma = \operatorname{sgn}(\sigma^{-1})$.

$$\det A = \sum_{\substack{\sigma \text{ is a permutation} \\ \text{of } \{1, \ldots, N\}}} (\operatorname{sgn} \sigma)\, a_{1,\sigma(1)}\, a_{2,\sigma(2)} \cdots a_{N,\sigma(N)}$$

$$= \sum_{\substack{\sigma \text{ is a permutation} \\ \text{of } \{1, \ldots, N\}}} (\operatorname{sgn} \sigma^{-1})\, a_{\sigma^{-1}(1),1}\, a_{\sigma^{-1}(2),2} \cdots a_{\sigma^{-1}(N),N}$$

$$= \sum_{\substack{\sigma^{-1} \text{ is a permutation} \\ \text{of } \{1, \ldots, N\}}} (\operatorname{sgn} \sigma^{-1})\, a_{\sigma^{-1}(1),1}\, a_{\sigma^{-1}(2),2} \cdots a_{\sigma^{-1}(N),N}$$

$$= \sum_{\substack{\sigma \text{ is a permutation} \\ \text{of } \{1, \ldots, N\}}} (\operatorname{sgn} \sigma)\, a_{\sigma(1),1}\, a_{\sigma(2),2} \cdots a_{\sigma(N),N} = \det A^T.$$ ■

This theorem implies that the determinant could be defined as a multilinear function of the columns of a matrix.

I next introduce the concept of cofactors of an NxN matrix, which are useful in computing its determinant and inverse. If A is an NxN matrix and m and n are such that $1 \leq m \leq N$ and $1 \leq n \leq N$, let $A(m \mid n)$ be the $(N-1)\mathrm{x}(N-1)$ matrix obtained from A by eliminating its m^{th} row and n^{th} column.

Definition: The $(m, n)^{th}$ cofactor of A is $C_{mn} = (-1)^{m+n} \det A(m \mid n)$.

Theorem 3.7: If A is an NxN matrix and $1 \leq n \leq N$, then $\det A = \sum_{m=1}^{N} a_{mn} C_{mn} = \sum_{m=1}^{N} a_{nm} C_{nm}$.

Proof: This theorem follows directly from the equation

$$\det A = \sum_{\substack{\sigma \text{ is a permutation} \\ \text{of } \{1, \ldots, N\}}} (\operatorname{sgn} \sigma)\, a_{1,\sigma(1)}\, a_{2,\sigma(2)} \cdots a_{N,\sigma(N)},$$

though it requires some work to see that this is so. ■

The equation $\det A = \sum_{m=1}^{N} a_{mn} C_{mn}$ is the expansion in cofactors of the determinant of A along the n^{th} column of A. The equation $\det A = \sum_{m=1}^{N} a_{nm} C_{nm}$ is the expansion in cofactors of the determinant of A along the n^{th} row of A. These expansions may be used to calculate the determinant by induction on the size of a matrix. For instance, expansion along the first column of a 3x3 matrix yields the equation

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

Theorem 3.8: If A is an NxN matrix, then $\sum_{m=1}^{N} a_{mn} C_{mk} = 0$, if $k \neq n$.

Proof: Replace the k^{th} column of A by its n^{th} column, obtaining the matrix B. Then $B(m \mid k) = A(m \mid k)$, for all m. Since B has two equal columns, $\det B = 0$ and so

$$0 = \det B = \sum_{m=1}^{N} b_{mk} (-1)^{m+k} \det B(m \mid k) = \sum_{m=1}^{N} a_{mn} (-1)^{m+k} \det A(m \mid k) = \sum_{m=1}^{N} a_{mn} C_{mk},$$

where b_{mk} is the $(m, k)^{th}$ entry of the matrix B. ■

To summarize, $\sum_{m=1}^{N} a_{mn} C_{mk} = \delta_{nk} \det A$,, where δ_{nk} is the Kronecker delta function defined by the equation

$$\delta_{nk} = \begin{cases} 1, & \text{if } n = k \\ 0, & \text{otherwise.} \end{cases}$$

Because $\det A = \det A^T$,

$$\sum_{n=1}^{N} a_{mn} C_{kn} = \delta_{mk} \det A^T = \delta_{mk} \det A,$$

for all m and k.

These equations make it possible to use cofactors to calculate the inverse of a matrix. In order to describe how to do so, I introduce the concept of the adjoint of a matrix.

Definition: The adjoint matrix of an NxN matrix A is the matrix adj A, the (m, n)th entry of which is

$$(\operatorname{adj} A)_{mn} = C_{nm} = (-1)^{n+m} \det A(n \mid m),$$

for m and n such that $1 \le m \le N$ and $1 \le n \le N$.

The adjoint of A is the transpose of the matrix of cofactors of A, where the (m, n)th entry of the matrix of cofactors is C_{mn}. Notice that by theorems 3.7 and 3.8, the (m, n)th entry of the product (adj A)A is

$$\sum_{j=1}^{N} (\operatorname{adj} A)_{mj} a_{jn} = \sum_{j=1}^{N} C_{jm} a_{jn} = \sum_{j=1}^{N} a_{jn} C_{jm} = \delta_{nm} \det A.$$

Therefore (adj A)A = (det A)$\mathbf{I}$, where $\mathbf{I}$ is the NxN identity matrix. Hence

$$\frac{1}{\det A} (\operatorname{adj} A) A = \mathbf{I},$$

if det A $\ne$ 0. That is, if det A $\ne$ 0, then A has a left inverse and so is invertible and

$$A^{-1} = \frac{1}{\det A} (\operatorname{adj} A).$$

By corollary 3.5 if A is invertible, then det A $\ne$ 0 and

$$\det A^{-1} = \frac{1}{\det A}.$$

This proves the following theorem.

Theorem 3.9: An NxN matrix A is invertible if and only if det A $\ne$ 0, and in this case

$$A^{-1} = \frac{1}{\det A} (\operatorname{adj} A).$$

Singular Matrices and Linear Transformations

We have already seen in theorem 2.12 that a linear transformation is invertible if and only if it is non-singular.

Definition: Let V be a finite dimensional vector space. The linear transformation $T:V \to V$ is singular if $T(v) = 0$, for some $v \neq 0$.

That is, T is singular if the dimension of its nullity exceeds 0, that is, if it is not non-singular. So theorem 2.12 implies that T is singular if and only if it is not invertible.

Definition: An NxN matrix A is singular if $Ax = 0$, for some non-zero N-vector x.

Proposition 3.10: Let A be an NxN matrix and let $T: V \to V$ be a linear transformation from the N-dimensional vector space V to itself. If A is the matrix representation of T with respect to some basis $v_1, \ldots, v_N$ of V, T is singular if and only if A is singular.

Proof: The proof of this proposition should be obvious. ■

Proposition 3.11: An NxN matrix A is singular if and only if det A = 0.

Proof: By theorem 3.9, det A = 0 if and only if A is not invertible. By theorem 2.6, A is not invertible if and only if A is singular. ■

Complex Numbers

The next topic to be covered in linear algebra is characteristic values, and for this I need the concept of a complex number. Mathematicians introduced complex numbers in order to have a realm in which all polynomial equations have solutions. Not all polynomial equations have solutions in the real numbers. For instance, there is no real number satisfying the equation $x^2 + 1 = 0$. To meet this problem, we can make up an "imaginary" number, i, that solves this equation. That is $i = \sqrt{-1}$. Almost magically, it turns out that if this number is adjoined to the real numbers, then the resulting set, called the complex numbers, contains the solution of any polynomial equation.

A complex number is of the form a + bi, where a and b are real numbers and $i = \sqrt{-1}$. Complex numbers may be added, subtracted, multiplied, and divided. Thus

$(a + bi) + (c + di) = (a + c) + (b + d)i,$
$(a + bi) - (c + di) = (a - c) + (b - d)i,$
$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$

The complex number $(a + bi)/(c + di)$ is found by solving the equation

$$a + bi = (c + di)(x + yi)$$

for x and y. This equation becomes

$$a + bi = (cx - dy) + (cy + dx)i,$$

which implies the two simultaneous linear equations

$$cx - dy = a$$
$$dx + cy = b.$$

The solution of these equations is

$$x = \frac{ac + bd}{c^2 + d^2} \text{ and } y = \frac{bc - ad}{c^2 + d^2}.$$

Notice that any real number r is also complex in that it may be written as $r + 0i$, so that when we speak of a number as complex, we include the possibility that it be real.

The key property of complex numbers is that any polynomial equation

$$0 = p(x) = a_N x^N + a_{N-1} x^{N-1} + \ldots. + a_1 x + a_0$$

has a complex solution if the coefficients $a_N, \ldots.., a_0,$ are real or complex. Furthermore if $a_N \neq 0$, then there are N such solutions, $b_1, \ldots., b_N,$ and

$$p(x) = a_N(x - b_1)(x - b_2)\ldots.(x - b_N).$$

The solutions $b_1, \ldots., b_N$ are called the <u>roots</u> of p. Some of the roots $b_1, \ldots., b_N$ may be equal. A complex number b is said to be a multiple root of p, if $b = b_n$, for more than one value of n. If $a_N \neq 0$, the positive integer N is said to be the <u>degree</u> of p.

<u>Characteristic Values and Vectors</u>

A deep understanding of linear transformations and NxN matrices is based on use of characteristic values and vectors, though I will not have time to say much about them.

<u>Definition</u>: If A is an NxN matrix, a complex number λ is a <u>characteristic value</u> of A, if the matrix $A - \lambda I$ is singular, where I is the NxN identity matrix. The non-zero N-vector x is a <u>characteristic vector</u> of A if $(A - \lambda I)x = 0$. If λ is not real, the components of x may be complex numbers, even if the entries of A are real.

There are other terms for characteristic values and vectors. For instance, they are termed eigenvalues and eigenvectors.

Remark: The same terminology applies to linear transformations $T: V \to V$, where V is a finite dimensional vector space. Thus λ is a characteristic value of T if $T - \lambda id_V$ is singular, where $id_V : V \to V$ is the identity function defined by $id_V(v) = v$, for all $v \in V$.

By proposition 3.11, λ is a characteristic value of an NxN matrix A if and only if $\det(A - \lambda I) = 0$. Hence λ is a characteristic value of A, if and only if λ is a root of the polynomial equation $p(x) = \det(xI - A)$. The polynomial $p(x) = \det(xI - A)$ is called the characteristic polynomial of A. This polynomial has degree N, since it is of the form $x^N + a_{N-1}x^{N-1} + ... + a_1 x + a_0$, for some numbers $a_{N-1}, a_{N-2}, ... , a_1$. Its N roots, counting multiplicities, are the characteristic values of A.

Example: Let $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, where a and b are real numbers and $b \neq 0$. The characteristic equation of A is

$$0 = \det(xI - A) = \det\begin{pmatrix} x-a & -b \\ b & x-a \end{pmatrix} = (x-a)^2 + b^2 = x^2 - 2ax + a^2 + b^2.$$

Using the quadratic formula, we find that

$$x = \frac{2a \pm \sqrt{4a^2 - 4a^2 - 4b^2}}{2} = \frac{2a \pm 2bi}{2} = a \pm bi.$$

So $a + bi$ and $a - bi$ are the characteristic values of A. The characteristic vectors corresponding to the characteristic value $a + bi$ are the vectors $x = t(-i, 1)$, for any non-zero complex number t. The characteristic vectors corresponding to the characteristic value $a - bi$ are the vectors $x = t(i, i)$, for any non-zero complex number t.

Theorem 3.12: If A is an NxN matrix, then $\det(A) = \lambda_1\lambda_2....\lambda_N$, where $\lambda_1, ... , \lambda_N$ are the characteristic values of A.

Proof: We know that $\det(xI - A) = (x - \lambda_1)(x - \lambda_2)....(x - \lambda_N)$. Let $x = 0$ in this equation. Then $(-1)^N\det(A) = \det(-A) = (-1)^N\lambda_1\lambda_2....\lambda_N$, so that $\det(A) = \lambda_1\lambda_2....\lambda_N$. ■

The next theorem gives an idea of the use that can be made of characteristic values. Its statement requires the notion of the absolute value of a complex number.

Definition: If $\lambda = a + bi$ is a complex number, where a and b are real numbers, then the absolute value of λ is $|\lambda| = \sqrt{a^2 + b^2}$.

Theorem 3.13: Let A be an NxN matrix with characteristic values $\lambda_1, \ldots, \lambda_N$. If $|\lambda_n| < 1$, for all n, then $\lim_{k \to \infty} A^k = 0$, where $A^k = \underbrace{AA.....A}_{k \text{ times}}$ is the product of A with itself k times.

I provide no proof of this theorem.

Bilinear Forms

We will use bilinear forms in defining sufficient conditions for maxima and minima in terms of the second derivative.

Definition: If V is a vector space, $f: V \times V \to R$ is a bilinear form on V if for each $\underline{v} \in V$, $f(v, \underline{v})$ and $f(\underline{v}, v)$ are linear functions of v.

Bilinear forms on a finite dimensional vector space may be represented by a square matrix. Let $v_1, \ldots, v_N$ be a basis for V and let $f: V \times V \to R$ be a bilinear form. For integers m and n from 1 to N, let $a_{mn} = f(v_m, v_n)$. Let A be the NxN matrix with (m, n)th entry a_{mn}. If $v = \sum_{m=1}^{N} b_m v_m$ and $w = \sum_{n=1}^{N} c_n v_n$ belong to V, then

$$f(v, w) = f\left(\sum_{m=1}^{N} b_m v_m, \sum_{n=1}^{N} c_n v_n\right) = \sum_{m=1}^{N}\sum_{n=1}^{N} b_m c_n f(v_m, v_n) = \sum_{m=1}^{N}\sum_{n=1}^{N} b_m c_n a_{mn} = (b_1, \ldots, b_N) A \begin{pmatrix} c_1 \\ \cdot \\ \cdot \\ \cdot \\ c_N \end{pmatrix}.$$

The matrix A represents f in the sense that given the basis $v_1, \ldots, v_N$ there exists one and only matrix A such that $f(v, w) = \sum_{m=1}^{N}\sum_{n=1}^{N} b_m c_n a_{mn}$, when $v = \sum_{m=1}^{N} b_m v_m$ and $w = \sum_{n=1}^{N} c_n v_n$.

Definition: An NxN matrix A is symmetric if $A = A^T$.

Definition: A bilinear form f on V is symmetric if f(v, w) = f(w, v), for all v and w in V.

Remark: The bilinear form f on V is symmetric if and only if the matrix A representing it with respect to some basis is symmetric.

Definition: The quadratic form associated with a symmetric bilinear form f is the function q defined by the equation $q(v) = f(v, v)$.

Definition: A symmetric bilinear form f on V or its associated quadratic form is

1) positive definite if $f(v, v) > 0$, for all $v \neq 0$,

2) positive semi-definite if $f(v, v) \geq 0$, for all v,

3) negative definite if $f(v, v) < 0$, for all $v \neq 0$, and

4) negative semi-definite if $f(v, v) \leq 0$, for all v.

The same definitions apply to symmetric NxN matrices, for each of these represents a symmetric bilinear form. For instance, the symmetric NxN matrix A is positive definite if $v^TAv > 0$, for any non-zero N-vector.

Example: The inner or dot product is a positive definite symmetric bilinear form on R^N. Its matrix representation with respect to the standard basis is the NxN identity matrix I.

If A is an NxN matrix and $1 \leq k \leq N$, let A_k be the kxk submatrix obtained by eliminating the last N – k rows and columns of A. That is,

$$A_k = \begin{pmatrix} a_{11} & . & . & a_{1k} \\ . & & & . \\ . & & & . \\ a_{k1} & . & . & a_{kk} \end{pmatrix}.$$

Theorem 3.14: The NXN symmetric matrix A and any bilinear form it represents is

1) negative definite, if and only if $(-1)^k \det A_k > 0$, for all k = 1, ... , N, and

2) positive definite, if and only if $\det A_k > 0$, for all k = 1, ... , N.

I provide no proof of this theorem.

Theorem 3.15: Let A be a symmetric NxN matrix. A is positive definite if and only if all of its characteristic values are positive. A is positive semi-definite if and only if all of its characteristic values are non-negative. A is negative semi-definite if and only if all of its characteristic values are non-positive. A is negative definite if and only if all of its characteristic values are negative.