Lecture 7

Observe that we had plenty of freedom to choose the order of equations. This becomes a boon numerical methods. However, it is also necessary to curtail this freedom in order to bring some uniformity, viz., once a set of equations is presented to us in a particular order, we should change the order of the equations only if it is very much necessary. For the same reason, making leading entry equal to 1 is not completely necessary at the beginning. Neither there is any need to sweep the columns above a particular row. With these modification the above method becomes a well defined powerful algorithm thereby providing various advantages. We shall refer to it by Gauss Elimination Method(GEM).

However, we have also observed that the last two sets of operations have their own advantages. Since there is no ambiguity in them, we shall incorporate them in to a second stage (extended) algorithm called Gauss-Jordan Method (GJEM). Below we describe these two methods in algorithmic form. Let us begin with some definitions.

Definition 2.3 *A $m \times n$ matrix $A = ((a_{i,j}))$ is called* echelon matrix *(respectively,* reduced echelon matrix*) if $A = 0$ or if there exists an integer r, $1 \leq r \leq min\{m, n\}$ and integers*

$$1 \leq k(1) < k(2) < \cdots < k(r) \leq m$$

such that
(i) $a_{i,j} = 0$ for all $i > r$ and for all j.
(ii) for each $1 \leq i \leq r$, $a_{i,j} = 0$ for $j < k(i)$.
(iii) for each $1 \leq i \leq r$, $a_{i,k(i)} \neq 0$ (respectively, $= 1$).
(iv) for each $1 \leq i \leq r$, $a_{s,k(i)} = 0$, for all $s > i$ (respectively, for $s \neq i$).

Remark 2.3
(i) Echelon matrices can be defined with respect to columns also. Since by taking transpose, we can easily convert one theory into the other there is no need to study them separately. In the literature there are slightly different conditions in the definition of an echelon matrix and also different names. It would be appropriate to call echelon matrices **Gauss matrices** *and the reduced echelon matrices* **Gauss-Jordan matrices**.
(ii) The integer r occurring in the definition is quite significant. Let us call it number of steps. The columns $C_{k(s)}$ are called the **step columns.** *The $(i, k(i))^{th}$ entries which are non zero in $G(A)$ (and equal to 1 in $J(A)$ are called* **pivots.** *It is convenient for later purpose to introduce a notation for the set of indices j such that the j^{th} column of A is not a step. So, let denote this set by $\rho(A)$ (which is a subset of N).*

Example 2.7 *We have already seen enough examples of echelon matrices. Indeed in the examples of solutions of systems of equations that we discussed, the end product was always an echelon matrix.*

Theorem 2.3 Algorithm For GEM

Let $A = ((a_{ij}))$ be a $m \times n$ matrix. The following algorithm will convert A into an echelon matrix Z.
Step 0 Put $X = A$ and $X = ((x_{i,j}))$.
Step 1. If X is the zero matrix or empty then declare $Z = A$ and stop. Else go to step 2.
Step 2. Look for a non zero entry in the first column of X. If you cannot find one, cut down

the first column of X and call the new matrix as X and go back to step 1. Otherwise find the first non zero entry in the first column say, $=x_{i,1}$.
Step 3. Swap the first row and i^{th} rows of X. (Swap the corresponding rows of A also.)
Step 4. Add $-(x_{j,1}/x_{1,1})R_1$ to R_j, for $j \geq 2$ and perform the corresponding row operations on the matrix A also.
Step 5. Cut down both the first column and the first row of X and call the new matrix as X and go back to step 1.

Theorem 2.4 Algorithm for GJEM *Let A be an echelon matrix. The following row operations on A convert it into a row reduced echelon matrix: For $1 \leq i \leq r$ where r is the number of steps in A, let $a_{i,k(i)}$ be the first non zero entry in the i^{th} row. Divide the i^{th} row of A by $a_{i,k(i)}$. Now for each $1 \leq i \leq r$ add $-a_{j,k(i)}R_i$ to R_j for all $1 \leq j < i$.*

The proof of these two theorems is very straight forward. Indeed we see that the algorithms perform a finite sequence of elementary row operations on the given matrix and then successively on the new resulting matrices until the desired result is obtained. Each such operation can be thought of as multiplication on left by a corresponding elementary matrix, we have:

Theorem 2.5 *Given any matrix A there exist an invertible matrix R such that $RA = G(A)$ (respectively, $= J(A)$)) is a echelon (respectively, reduced echelon) matrix.*

We shall call $G(A)$ the **Gauss form** of A and $J(A)$ the **Gauss-Jordan form** of A. Observe that by the very definition which is algorithmic $G(A)$ and $J(A)$ are well defined.

Finally we sum up the rule for solving a system of linear equations:

HOW TO SOLVE A SYSTEM OF LINEAR EQUATIONS

$$A\mathbf{x} = \mathbf{b}.$$

Step 1 Right down the augmented matrix $[A|\mathbf{b}]$.

Step 2 Apply JGEM and obtain the Gauss-Jordan form of it: $J([A|\mathbf{b}]) = [J(A)|\mathbf{b}']$.
Let r be the number of nonzero rows of $J(A)$ and let
$k(1) < \cdots < k(r)$
be the indices corresponding to the step columns.

Step 3 If $b'_i \neq 0$ for any $i > r$ declare that the system is **inconsistent,** i.e, the system has no solution.
Otherwise proceed.

Step 4 If $r = n$ (the number of columns of A)
declare that the system has a **unique solution**
viz., $(x_1, \ldots, x_n) = (b'_1, \ldots, b'_n)$.
Otherwise proceed.

Step 5 In $[J(A)|\mathbf{b}']$ delete all the zero-rows.

Step 6 Let the indices corresponding to the non-step columns be
$l(1) < l(2) < \cdots, < l(s)$, $(s + r = n)$.
Transfer all the non-step columns to the right side and change their sign to obtain a matrix of the form $[I_r|B']$ where the first column of B' is $\mathbf{b}'$.
The **general solution** is given by putting
$x_{l(i)} = t_i, 1 \leq i \leq s$;
and
$(x_{k(1)}, \ldots, x_{k(r)})^T = B'(1, t_1, \ldots, t_s)^T$.

3 Determinants

We have seen that a rigid motion which fixes the origin is an invertible linear transformation of a euclidean space. It is clear that such a transformation does not change the volume of any solid. Obviously, this cannot be true of all linear transformations. But then how to measure the change in the volume effected by a linear transformation. An answer is in the notion of "determinants."

For example consider the map $\lambda : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by $\mathbf{v} \mapsto \lambda\mathbf{v}$ where $\lambda \in \mathbb{R}$. For $n = 1$ it is clear that a segment of length l in $\mathbb{R}$ (no matter where it is situated in $\mathbb{R}$) is mapped to a segment of length $l\lambda$. Likewise it is easily seen that a cube of volume m in $\mathbb{R}^3$ is mapped to a solid of volume $m\lambda^3$. In $\mathbb{R}^n$, the effect on the volume is to multiply by λ^n. The matrix of λ is the diagonal $\text{diag}(\lambda, \lambda, \ldots, \lambda)$. A little more generally, if we consider the linear transformation which effects a scaling by λ_i in the i^{th} direction, then it is clear that the volume of a n-dimensional cube is changed by a factor of $\lambda_1 \cdots \lambda_n$. These considerations lead us to believe that each linear transformation effects a scaling in the volume. We must hunt for this scaling factor.

So, let us consider a transformation $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ such that $T(1,0) = (a,b) =: \mathbf{v}_1$ and $T(0,1) = (c,d) = \mathbf{v}_2$. This means that T maps the unit square in $\mathbb{R}^2$ to the parallelogram with its vertices $(0,0), (a,b), (c,d), (a+c, b+d)$. The area of this is given by $|ad-bc|$. In terms of vector product it is $|\mathbf{v}_1 \times \mathbf{v}_2|$. Likewise, if $T' : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ is such that $T'(\mathbf{e}_i) = \mathbf{v}_i$ then the volume of the parallelepiped defined by the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is given by the modulus of the scalar triple product $(\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3)$. We can even write down the formula for this in terms of coordinates of $\mathbf{v}_i$, if we wish.

What do we do in the general case? We would like to approach this problem as follows: First we ask what are the basic properties of this 'creature' that we are seeking. After listing sufficiently many of these properties in an economic way, we shall launch a search party to look out for the creature.

Lecture 8

3.1 Axiomatic Approach

In what follows we shall use the symbol $\mathbb{K}$ to denote either the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complx numbers. (Indeed, for many of the concepts and results, we could have taken $\mathbb{K}$ to be just any field.)

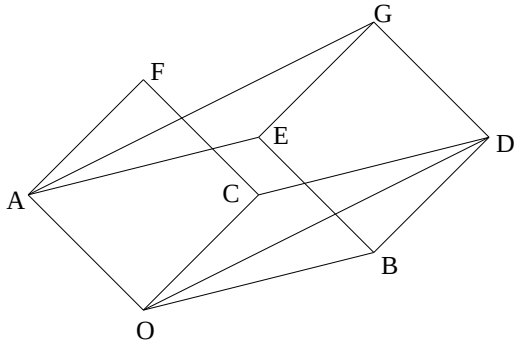
We are looking for a $\mathbb{K}$-valued function defined on the set of $n \times n$ matrices (for each n) say $f : M_n(\mathbb{K}) \longrightarrow \mathbb{K}$ which is likely to represents the volume of the n-dim. parallelepiped defined by the row vectors of a given matrix.

(i) Naturally then if two rows of a matrix A are identical then $f(A) = 0$ since in that case, we do not get a genuine parallelepiped.

(ii) Another important thing we remind ourselves is to look at the sign of the area of the parallelogram. Depending upon the order in which we measure it, the sign of the area could be positive or negative. This is put in a nice way as follows: *If we change any two rows of A then the effect on $f(A)$ must be to multiply by -1.*

(iii) We have already observed that if we scale one of the vectors then the entire volume is also scaled by the same factor.

(iv) The last mentioned property already warms us up for looking for the effect of adding two rows. The following simple diagram illustrates this and affirms that we are on the right track. Namely, we have $(\mathbf{v}_1 + \mathbf{v}_2) \times \mathbf{u} = \mathbf{v}_1 \times \mathbf{u} + \mathbf{v}_2 \times \mathbf{u}$.



OBEA+OCFA=OBEA+BDGE−OBD+AEG=ODGA

`Geometric proof of the additivity of the area`

(Caution: The above figure is not a 3-dimensional one!)

With these things in mind we shall now attack the problem systematically.

Given a matrix $A \in M_n(\mathbb{K})$ say,

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots A_n \end{bmatrix},$$

recall that by arranging the n rows of A one after another we can think of it as an ordered n-tuple of n-vectors: $A = (A_1, \ldots, A_n)$. More generally, this method may be used to identify $M_{n,m}(\mathbb{K})$ with $\mathbb{K}^{mn}$. We are going to exploit it now.

Definition 3.1 *Let $1 \leq i \leq n$. A function $f : M_n(\mathbb{K}) \longrightarrow \mathbb{K}$ is called* linear in the i^{th} row *if,*
$f(A_1, \ldots, A_{i-1}, \alpha B + \beta C, A_{i+1}, \ldots, A_n) =$

$$\alpha f(A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots, A_n) + \beta f(A_1, \ldots, A_{i-1}, C, A_{i+1}, \ldots, A_n) \tag{26}$$

for all rows $A_j, B, C \in \mathbb{K}^n$ and all scalars $\alpha, \beta \in \mathbb{K}$. A function f which is linear in i^{th} row for each $1 \leq i \leq n$ will be called **multi-linear.**

Examples of multi-linear maps that we have are the vector product of two vectors in case $n = 2$ and the scalar triple product in case $n = 3$. We do not know many examples. This should not deter us from our goal. Remember that a linear map was completely determined by its values on the standard vectors $\mathbf{e}_i.'s$ We are going to exploit this now.

Let $N = \{1, 2, \ldots, n\}$ and put $\Lambda = \mathcal{F}(N, N)$ the set of all maps $\sigma : N \longrightarrow N$. (This set can also be thought of as the set of all sequences of length n with values between 1 and n.) Let $E(\sigma)$ denote the matrix with its i^{th} row equal to $\mathbf{e}_{\sigma(i)}$. For example, in this notation if $\sigma = Id$ then $E(\sigma) = I_n$. Write down a few more examples to get familiar with this notation.

Lemma 3.1 Any multi-linear map f is completely determined by its values on $E(\sigma), \sigma \in \Lambda$.

Proof: Let us first illustrate this for $n = 2$. Write

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{27}$$

Then $A = (A_1, A_2)$ where $A_1 = (a, b)$ and $A_2 = (c, d)$. Further $A_1 = a\mathbf{e}_1 + b\mathbf{e}_2$ and $A_2 = c\mathbf{e}_1 + d\mathbf{e}_2$. Therefore

$$\begin{aligned} f(A) &= af(\mathbf{e}_1, A_2) + b(\mathbf{e}_2, A_2) \\ &= a(cf(\mathbf{e}_1, \mathbf{e}_1) + df(\mathbf{e}_1, \mathbf{e}_2)) + b(cf(\mathbf{e}_2, \mathbf{e}_1) + df(\mathbf{e}_2, \mathbf{e}_2)) \\ &= ac\, f(E_{11}) + ad\, f(E_{12}) + bc\, f(E_{21}) + bd\, f(E_{22}). \end{aligned}$$

Thus the pattern in the general case is clear. By applying linearity in each slot at a time we obtain,

$$f(A) = \sum_{\sigma} a_\sigma f(E(\sigma)) \tag{28}$$

Here the scalars a_σ are given by

$$a_\sigma = a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}. \tag{29}$$

The verification is straight forward. ♠

Definition 3.2 *A map $f : M_n(\mathbb{K}) \longrightarrow \mathbb{K}$ is called* alternating *if for every pair $(i,j), i \neq j$,*

$$f(A_1, \ldots, A_i, \ldots, A_j, \ldots, A_n) = -f(A_1, \ldots, A_j, \ldots, A_i, \ldots, A_n). \tag{30}$$

That is, if we interchange the place of any two rows then the sign of the value of the function is changed.

Lemma 3.2 A multi-linear map f is alternating iff it vanishes on all matrices with two equal rows.

Proof: Let f be multi-linear and alternating. If two rows of A are the same then by interchanging them, A does not change. On the other hand the sign of the value of f on A has to change. This means that $f(A) = -f(A)$ and hence $f(A) = 0$. To prove the converse, given a matrix A consider another matrix B whose rows are that of A except in the i^{th} and j^{th} places and $B_i = B_j = A_i + A_j$. Then $f(B) = 0$. On the other hand by multi-linearity, we have

$$\begin{aligned} 0 = f(B) &= f(A_1, \ldots, A_i, \ldots, A_i + A_j, \ldots, A_n) + f(A_1, \ldots, A_j, \ldots, A_i + A_j, \ldots, A_n) \\ &= f(A_1, \ldots, A_i, \ldots, A_j, \ldots, A_n) + f(A_1, \ldots, A_j, \ldots, A_i, \ldots, A_n) \end{aligned}$$

Hence $f(A_1, \ldots, A_i, \ldots, A_j, \ldots, A_n) = -f(A_1, \ldots, A_j, \ldots, A_i, \ldots, A_n)$. ♠

Remark 3.1 *There is more merit in this equivalent formulation of alternating property than meets the eye. It is the right one for certain other types of fields which we have not considered here.*

Lemma 3.3 *Let f be a multi-linear alternating map on $M_n(\mathbb{K})$. Consider the elementary matrix $E = I + \alpha E_{ij}$. Then for all $A \in M_n$,*

$$f(EA) = \begin{cases} f(A) & \text{if } i \neq j \\ (1+\alpha) f(A) & \text{if } i = j \end{cases}$$

Proof: Let $A = (A_1, \ldots, A_n)$. For $i \neq j$, observe that

$$EA = (A_1, \ldots, \ldots, A_j + \alpha A_i, \ldots, A_j, \ldots, A_n).$$

Therefore

$$\begin{aligned} f(EA) &= f(A_1, \ldots, A_i, \ldots, A_j, \ldots, A_n) + \alpha f(A_1, \ldots, A_i, \ldots, A_i, \ldots, A_n) \\ &= f(A). \end{aligned}$$

For $i = j$, we have $EA = (A_1, \ldots, A_{i-1}, (1+\alpha)A_i, A_{i+1}, \ldots, A_n)$. Therefore, $f(EA) = (1 + \alpha) f(A)$, by linearity in the i^{th} slot. This proves the lemma. ♠

Remark 3.2 *In the case of linear maps we were free to choose the value of the function on $\mathbf{e}_i's$. Same is true for a multi-linear map, viz., we are free to choose its value on E_{ij} and then extend it by the formula (28). However, in addition, if f is alternating also, then we are forced to choose its value on $E(\sigma)$ to be zero for all those σ which are not one-to-one.*

Definition 3.3 *Let $\Sigma(n)$ denote those elements $\sigma \in \mathcal{F}(N, N)$ which are one-to-one. These are called* permutations *on n letters. If a permutation changes its value only on two indices then it is called a* transposition.

Given $\sigma \in \Sigma(n)$, let us observe what GEM does on $E(\sigma)$. It is clear that GEM simply consists of a sequence of change of rows and we obtain $G(E(\sigma)) = J(E(\sigma)) = I_n$. Observe that a change of two rows of $E(\sigma)$ is effected by composing σ with a corresponding transposition. Thus GEM yield a sequence of transpositions $\tau_1, \ldots, \tau_k$ such that

$$E(\tau_k \circ \cdots \circ \tau_1 \circ \sigma) = I_n.$$

This means that $\tau_k \circ \cdots \tau_1 \circ \sigma = Id$. Hence $\sigma = \tau_1 \circ \cdots \circ \tau_k$. Thus we have proved:

Lemma 3.4 Every permutation is a (finite) composite of transpositions.

Definition 3.4 *Let $e(\sigma)$, the number of row changes in the GEM applied to $E(\sigma)$ to bring it to $G(E(\sigma)) = I_n$. Let $sgn(\sigma) = (-1)^{e(\sigma)}$.*

We call this the *signature* of σ.

Applying repeatedly the alternating property, it follows that

$$f(E_\sigma) = sgn(\sigma) f(I_n). \tag{31}$$

Combining (28), (29) and (31), we get

Theorem 3.1 *An alternating multi-linear map f is determined by its value on I_n.*

3.2 Existence of Determinant Function

Theorem 3.2 *For every $t \in \mathbb{R}$, there exists a unique $f_t : M_n(\mathbb{R}) \to \mathbb{R}$ which is row-wise multi-linear, alternating and such that $f_t(I_n) = t$. Moreover, for any such f_t we have $f_t(E_\sigma) = sgn(\sigma)t$.*

Note that so far what we have already proved the uniqueness of such a function and the formula $f_t(E_\sigma) = sgn(\sigma)t$.

To show that such a function exists, we just define f_t by the formula

$$f_t(A) = \left(\sum_{\sigma \in \Sigma} sgn(\sigma) a_\sigma \right) t. \tag{32}$$

To verify that f_t is multi-linear is straight forward.

Assuming that formula (32) has alternating property on the set $\{E_\sigma, \sigma \in \Sigma\}$, it follows easily that the same holds for all matrices. Therefore, it is enough to prove:

Lemma 3.5 *Let $\sigma, \tau \in \Sigma(n)$, where τ is a transposition. Then*

$$sgn(\sigma) = -sgn(\sigma \circ \tau). \tag{33}$$

Proof: We prove it by induction on n. If $n = 2$ this is obvious. (For $n = 3$ this can be verified straight forward way, by considering various cases. Try this as an exercise before reading the rest of the proof which will help you to understand the proof better. However, this is not a logical necessity.)

Inductively, let the statement (33) hold for $n-1$. Let $A = E_\sigma,\ B = E_{\sigma\circ\tau}$. We look at the first step of GEM performed on both A and B cut down the first row and first column from both of them compare the results on either side, use induction to obtain the equality (33) for $n-1$ and finally relate it to (33) for n.

Suppose τ interchanges i and j, $i \neq j$. Suppose GEM applied to A interchanges 1^{st} row with k^{th} row. (We do not exclude the case when $k = 1$; it merely means that the first step of GEM does not do anything on A.)

The rest proof is best understood by glancing at the table below which is indeed a schematic representation of the proof.

Case 1: Let $\{1, k\} = \{i, j\}$. Here first step of GEM applied to A is the same as τ applied to A and hence both produce B. Clearly, the first step of GEM applied to B does nothing and all the remaining steps will coincide for both. Hence (33) gets verified here.

Case 2: Suppose the set $\{1, k, i, j\}$ has three elements. Observe that during the first step of GEM on A or B none of the other rows are affected. (Thus by simply ignoring these rows and the corresponding columns, this reduces to the case when $n = 3$ for which you have already verified (33). However, we shall write down a detailed proof.)

We consider first the subcase, where $k = 1$. Then the first step of GEM does not produce any change in A or B and we simply pass to the next stage. The equation (33) for n follows from the same for $n-1$ by induction.

Next, suppose k is equal to one of the indices $\{i, j\}$ say for definiteness, $k = i$. This means that the non zero entry in the first column of A is in i^{th} place and hence the same for B is j^{th} place. Therefore the first step of GEM for A swaps rows $1 \leftrightarrow i$, where as the same for B swaps rows $1 \leftrightarrow j$. Therefore after cutting down the first row and first column, the two matrices are again related by the transposition $\tau' = (i-1, j-1)$. Again the equation (33 for n is obtained by multiplying by -1 the same equation for $n-1$.

Case 3: Suppose $\{1, k, i, j\}$ has four elements. Under this condition, it is clear that the first step of GEM on both A and B is to interchange the first and k^{th} rows. After this the two $(n-1) \times (n-1)$ matrices A', B' are again related by a transposition. Therefore, by induction, (33) is true for A', B' and hence after mutliplying it by -1 the same holds for A, B. ♠

Case	Stage	A	B
2(a)	Initial stage	Y ——— (1) . . . . . . R ——— (i) . . . . . . P ——— (j)	Y ——— (1) . . . . . . P ——— (i) . . . . . . R ——— (j) . . . . . .
2a)	After step 1	Y ——— (1=k) . . . . . . R ——— (i) . . . . . . P ——— (j)	Y ——— (1=k) . . . . . . P ——— (i) . . . . . . R ——— (j) . . . . . .
2(b)	Initial stage	Y ——— (1) . . . . . . R ——— (i=k) . . . . . . P ——— (j)	Y ——— (1) . . . . . . P ——— (i) . . . . . . R ——— (j) . . . . . .
2b)	After Step 1	R ——— (1) . . . . . . Y ——— (i) . . . . . . P ——— (j)	R ——— (1) . . . . . . P ——— (i) . . . . . . Y ——— (j) . . . . . .

Case 3		
Stage	A	B
Initial stage	Y ——— (1) . . . . . . R ——— (i) . . . . . . G ——— (k) . . . . . . P ——— (j)	Y ——— (1) . . . . . . P ——— (i) . . . . . . G ——— (k) . . . . . . R ——— (j) . . . . . .
After Step 1	G ——— (1) . . . . . . R ——— (i) . . . . . . Y ——— (k) . . . . . . P ——— (j)	G ——— (1) . . . . . . P ——— (i) . . . . . . Y ——— (k) . . . . . . R ——— (j) . . . . . .

Definition 3.5 *A multi-linear alternating map* $f : M_n(\mathbb{K}) \longrightarrow \mathbb{K}$ *is called the* **determinant function** *if* $f(I_n) = 1$.

Theorem 3.3 *There exist a determinant function for each* $n \geq 1$.

Proof: Define $f(I_n) = 1$. Use (31) to define f on all permutation matrices $E(\sigma)$. From lemma (3.5), it follows that f is alternating on this set. Extend f to all matrices using (28) and (29). Verifying that f is alternating and multi-linear is straight forward. ♠

Remark 3.3 *The condition in the above definition is called* normalization. *Thus it follows that viz.,A multi-linear, alternating, normalised map is unique. It is denoted by* det . *Often one writes* $\det A$ *or* $|A|$ *to denote the determinant of a square matrix* A. *Clearly it coincides with the concept that we already have in case* $n = 1, 2$. *We shall now establish other properties of* det *which we are familiar with.*

Lecture 9

3.3 GEM Applied to Compute the Determinant

Let us now see how GEM can be applied to compute the determinant of a $n \times n$ matrix.

First consider the case when the row rank of A is equal to n. Then we know that $J(A) = I_n$. Suppose the number of change of rows performed to arrive at $G(A)$ is $e(A)$. By theorem 3.3, it follows that $\det(A) = (-1)^{e(A)} \det(G(A))$. Now look at the diagonal entries of $G(A)$. While arriving at $J(A)$ from $G(A)$, we first divide each row by the corresponding diagonal entry and then sweep each column. Since the sweeping of columns does not change the determinant, it follows that

$$\det(G(A)) = c_1 c_2 \cdots c_n. \tag{34}$$

Combining these two we have

$$\det A = (-1)^{e(A)} c_1 \cdots c_n. \tag{35}$$

where c_i are the diagonal entries of $G(A)$. Now if the row rank of A is $< n$ then clearly some row of $G(A)$ is zero and hence one of the $c_i = 0$. Also, $\det G(A) = 0$ and since $\det A = \pm \det G(A)$, it follows that $\det A = 0$. So, the formula (35) is still valid.

Remark 3.4
(i) In the special case when A is upper triangular, we have $G(A) = A$. Hence $\det A$ is the product of the diagonal entries.
(ii) Consider a $n \times n$ matrix M for which $a_{ij} = 0, k+1 \leq i \leq n, 1 \leq j \leq k$. Thus M can written in the form

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} \tag{36}$$

where A, B, D are respectively of type $k \times k, k \times (n-k), (n-k) \times (n-k)$. Such a matrix is called a block matrix. Let us compute the determinant of such a matrix. Let us apply GEM to M. Observe that unless A has become upper triangular, the process will never involve the rows numbered $> k$. Suppose the number of steps in $G(A)$ is less than k. Then there will be a row of zeros in $G(A)$ and hence $G(M)$ will have a zero on the diagonal. It follows that in this case $\det A = 0 = \det M$. Now suppose the number of steps in $G(A)$ is k. Then GEM applied to M will first make A into the upper triangular matrix and the rest of the operation on M will not involve the first k rows at all. Therefore, these operations can be thought of as two separate sets of row operations on A and D which correspond to operations of GEM on them. Thus we conclude that the diagonal entries of $G(M)$ are those of $G(A)$ and $G(D)$ where as the number of row swaps for M will be the sum of the same for A and D.

Thus we have proved:

Theorem 3.4 *For a block matrix*

$$M = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

we have $\det M = (\det A)(\det B)$.

3.4 Expansion Formula

(**Directly go to the theorem 3.5**.)
Lemma (3.5) has some other interesting applications.

Lemma 3.6 *Let a permutation σ be the composite of s transpositions. Then $sgn(\sigma) = (-1)^s$.*

Proof: Follows by induction on s by successive application of (33). ♠

Lemma 3.7 *For any permutation, $sgn(\sigma) = sgn(\sigma^{-1})$.*

Proof: First observe that for any transposition τ, we have $\tau^{-1} = \tau$. Therefore, if $\sigma = \tau_1 \circ \cdots \circ \tau_k$ then $\sigma^{-1} = \tau_k \circ \cdots \circ \tau_1$. Now apply the previous lemma. ♠

Remark 3.5 *Returning to the formula (28) and (29), observe that all terms corresponding to those σ which are not injective have dropped off, leaving the summation to be taken over $\Sigma(n)$. Using (31) we get:*

Theorem 3.5 Expansion Formula for determinant *For any square matrix A we have*

$$\det A = \sum_{\sigma} sgn(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \tag{37}$$

where the summation is taken over all the permutations σ of the letters $\{1, \ldots, n\}$.

3.5 Determinant of the Transpose

The entire discussion so far was with respect to the rows. We could have carried the whole thing with respect to rows. We then get another normalised alternating multi-linear map with respect to the columns. Let us call it δ. The following theorem tell you that $\delta = \det$.

Theorem 3.6 *For any square matrix A we have* $\det A^T = \det A$.

Proof: Putting $B = A^T$ and observing that $b_{ij} = a_{ji}$ from (37) it follows that

$$\det A^T = \sum_{\sigma} sgn(\sigma) a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n}.$$

Put $\tau = \sigma^{-1}$. Then as σ runs over all possible permutations of $\{1, \ldots, n\}$ so does τ. On the other hand for each i we have $i = \tau(\sigma(i))$ and for each fixed σ as i runs over $\{1, \ldots, n\}$ so does $\sigma(i)$. Therefore

$$a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} = a_{1\tau(1)} \cdots a_{n\tau(n)}.$$

Since we have seen that $sgn(\sigma) = sgn(\sigma^{-1})$, the above summation is the same as the rhs of (37). ♠

Remark 3.6 *This result is extremely useful. For, all the properties of the determinant function that we formulated or verified for rows will now be available for columns also.*

3.6 Laplace Formula for Determinant

(**Skip the proofs**)

Definition 3.6 *Let A_{ij} denote submatrix (minor) of A obtained by cutting down the i^{th} row and j^{th} column. We define the $(i,j)^{th}$ cofactor of A by $cof_{i,j}(A) = (-1)^{i+j} \det A_{ij}$.*

Remark 3.7
(i) Consider a matrix $A = ((a_{ij}))$ in which $a_{1j} = a_{i1} = 0$ for all $i, j > 1$ and $a_{11} = 1$. Then check that $\det A = cof_{11}(A) = \det A_{11}$. *(This is an easy application of GEM).*
(ii) Let now for any square matrix A and (i,j), let A^*_{ij} denote the matrix obtained by replacing the i^{th} row by $\mathbf{e}_j$. In order to compute the determinant of A^*_{ij} we can sweep the j^{th} column and assume that $a_{kj} = 0$ for $k \neq i$. Then by a sequence of interchanging the i^{th} row with

the row just above it we bring this row to the 1^{st} *place. Now the* j^{th} *column is equal to* $(1, 0, \ldots, 0)^T$. *Mean while the determinant has changed its sign to* $(-1)^{i-1}$.

Likewise we now swap this column with the columns left to it and bring it to the first place. The determinant now changed its sign to $(-1)^{i+j-2} = (-1)^{i+j}$. *Thus if we denote this new matrix by* B *then* $\det B = (-1)^{i+j} \det A^*_{ij}$. *Also* $B_{11} = A_{i,j}$ *Hence by (i) above we have* $\det B = \det B_{11} = \det A_{ij}$.

Thus we have proved:

Lemma 3.8 *For any square matrix* A *we have,*

$$det\, A^\star_{ij} = cof_{ij}(A). \tag{38}$$

The following theorem gives an inductive formula to obtain the determinant of a matrix in terms of its entries.

Theorem 3.7 Laplace Expansion Formula For Determinant: *For any* $n \times n$ *matrix* A *we have*

$$\det A = \sum_{j=1}^{n} a_{ij} cof_{ij},\ \forall\ 1 \leq i \leq n. \tag{39}$$

Proof: Using linearity in the i^{th} row we expand the determinant function. Thus

$$\begin{aligned} \det A &= \det(A_1, \ldots, A_i, \ldots A_n) \\ &= \det(A_1, \ldots, \textstyle\sum_j a_{ij}\mathbf{e}_j, \ldots, A_n) \\ &= \textstyle\sum_j a_{ij} \det(A_1, \ldots, \mathbf{e}_j, \ldots, A_n) \\ &= \textstyle\sum_j a_{ij} \det A^*_{ij} \\ &= \textstyle\sum_j a_{ij}\, cof_{ij}(A). \end{aligned}$$

the last equality is taken from (38).

Another Proof of Theorem 3.7: Since $\det A = \det A^T$ we first observe that the row-expansion formula (39) is equivalent to the column expansion:

$$\det A = \sum_{i=1}^{n} a_{ij}\, cof_{ij}(A),\ \forall\ 1 \leq j \leq n. \tag{40}$$

We consider the function f defined by the rhs of (40) and show that it is normalised, alternating multi-linear map. Then by the uniqueness of such a map, the equality (40) follows. It is enough to prove this for $j = 1$ since other equalities can be obtained by using alternating property.
(i) If $A = I_n$ then $A_{i,j}$ has a row of zeros unless $i = j$ and $A_{ii} = I_{n-1}$ it follows that $f(I_n) = 1$.
(ii) Suppose two rows of A are equal, say, $A_r = A_s$. If $r \neq 1, s \neq 1$ then it follows that all A_{i1} have two rows equal. Hence $cof(A_{i1}) = 0$. Therefore $f(A) = 0$.
(iii) In order to prove linearity in the r^{th} row, Let

$$\begin{aligned} X &= (X_1, \ldots, X_{r-1}, aA + bB, X_{r+1}, \ldots, X_n), \\ Y &= (X_1, \ldots, X_{r-1}, A, X_{r+1}, \ldots, X_n), \\ Z &= (X_1, \ldots, X_{r-1}, B, X_{r+1}, \ldots, X_n), \end{aligned}$$

where $A(a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$ are some row vectors. Then for $i \neq r$,

$$cof_{i1}(X) = a\, cof_{i1}(Y) + b\, cof_{i1}(Z).$$

Also $cof_{r1}(X) = cof_{r1}(Y) = cof_{r1}(Z)$. Moreover $x_{r1} = aa_1 + bb_1$. Therefore,

$$\begin{aligned} f(X) &= \textstyle\sum_{i\neq r} x_{i1}\, cof_{i1}(X) + x_{r1} cof(X_{r1}) \\ &= a(\textstyle\sum_{i\neq r} cof_{i1}(Y)) + b(\sum_{i\neq r} cof_{i1}(Z) + (aa_1 + bb_1) cof_{r1}(X) \\ &= a[\textstyle\sum_{i\neq r} cof_{i1}(Y) + a_1\, cof_{r1}(Y)] + b[\sum_{i\neq r} cof_{i1}(Z) + b_1\, cof_{r1}(Z)] \\ &= af(Y) + bf(Z) \end{aligned}$$

This proves the multi-linearity of f. ♠

Remark 3.8
(i) Laplace formula can be used to give an inductive proof of the existence of the determinant. (ii) Given any square matrix A, let M_{ik} denote the matrix obtained by replacing the i^{th} row of A by its k^{th} row. Since two rows of M_{ij} are equal we know that $\det M_{ij} = 0$. *On the other hand (39) applied to M_{ik} gives that*

$$\det M_{ij} = \sum_{j=1}^{n} a_{kj} cof_{ij}(A).$$

Combining it with (39) we obtain

$$\sum_{j=1}^{n} a_k j cof_{ij}(A) = \begin{cases} \det A, & \text{if } k = i; \\ 0, & \text{otherwise.} \end{cases} \tag{41}$$

3.7 Product Formula

(**Skip the proof**)

Theorem 3.8 Product formula for determinant *For any two $n \times n$ matrices A and B, we have* $\det(A) = (\det A)(\det B)$.

Proof: Fix B and consider the map $f(A) = \det(AB)$. Then $f(I_n) = \det B$. If we show that f is multi-linear and alternating then by the uniqueness of such maps the theorem would follow. Observe that the i^{th} row of AB is nothing but A_iB. Hence if two rows of A are equal then so are the corresponding two rows of AB. Hence $f(A) = \det(AB) = 0$. This proves alternating property. Now to check multi-linearity:

$$\begin{aligned} & f(A_1, \ldots, aA_i + a'A_i', \ldots, A_n) \\ = \ & \det(A_1B, \ldots, (aA_i + a'A_i')B, \ldots, A_nB) \\ = \ & a \det(A_1B, \ldots, A_iB, \ldots, A_nB) + a' \det(A_1B, \ldots, A_i'B, \ldots, A_nB) \\ = \ & a\, f(A_1, \ldots, A_i, \ldots, A_n) + a' f(A_1, \ldots, A_i', A_n) \end{aligned}$$

This proves the multi-linearity of f. ♠

3.8 Cramer's Rule

Formula (41) can itself be written as follows. Given a $n \times n$ matrix A let $adj\, A = (m_{ij})$ be the $n \times n$ matrix defined by

$$m_{ij} = \det cof_{ji}(A).$$

Then

Lemma 3.9 $(adj\, A)A = A(adj\, A) = diag(\det A, \ldots, \det A)$

Check that this is just a reformulation of (41).

In particular if $\det A \neq 0$, then it follows that A is invertible with the inverse given by

Lemma 3.10 $A^{-1} = (\det A)^{-1}(adj\, A)$.

As a simple minded application we obtain:

Theorem 3.9 Cramer's Rule: *Given a system of linear equation $A\mathbf{x} = \mathbf{b}$ with an invertible coefficient matrix, let M_j be the matrix obtained from A by replacing the j^{th} column by the column* **b**. *Then*

$$x_i = \frac{\det M_i}{\det A} \tag{42}$$

Proof: We have $\mathbf{x} = A^{-1}\mathbf{b} = (\det A)^{-1}(adj\, A)\mathbf{b}$. Equating the i^{th} entry on either side, we get

$$x_i == (\det A)^{-1}(\sum_{j=1}^{n} m_{ij}b_j) = (\det A)^{-1}\sum_{j=1}^{n} cof_{ji}(A)b_j = (\det A)^{-1}(\det M_i)$$

the last equality is obtained by taking the expansion of $\det M_i$ along the i^{th} column. ♠

Remark 3.9 *Cramer's rule is not good for computational purposes as it involves computation of determinants which required huge number of operations, as compared to GEM and certain modifications of it which we will discuss later.*