Notes on Determinants

The symmetric group $S_n$. It is not necessary to know group theory for this discussion but it might help.

**Definition.** The symmetric group $S_n$ is the set of all 1-1 maps of $\{1,2,\ldots,n\}$ into itself. Elements of $S_n$ are called permutations. Any rearrangement $\{i_1, i_2, \ldots, i_n\}$ of $\{1,2,\ldots,n\}$ is regarded as the element of $S_n$ which maps $k$ into $i_k$.

If $\sigma \in S_n$ then $\sigma$ is onto as well as 1-1, since $\{1,2,\ldots,n\}$ is finite. Multiplication of group elements in $S_n$ is simply composition. Thus $\rho = \sigma \tau$ is defined by the condition $\rho(i) = (\sigma \tau)(i) = (\sigma(\tau(i))$ for all $i \in \{1,2,\ldots,n\}$. The identity map is called $I$, and as usual, we have $\sigma \sigma^{-1} = \sigma^{-1} \sigma = I$.

It is convenient to regard $S_n \subseteq S_{n+1}$. If $\sigma \in S_n$, simply define $\sigma(n+1) = n+1$, and the newly define $\sigma$ is in $S_{n+1}$. More accurately, we have defined a 1-1 map from $S_n$ into $S_{n+1}$ which clearly preserves multiplication.

**Definition.** (Cycle notation) Let $i_i, i_2, \ldots, i_k$ be distinct elements of $\{1,2,\ldots,n\}$. Then the cycle $(i_1 \ i_2 \ \cdots \ i_k)$ is the map $\sigma$ defined by the conditions

1. $\sigma(i_\nu) = i_{\nu+1}$ for $\nu = 1, \ldots, k - 1$.
2. $\sigma(i_k) = i_1$.
3. $\sigma(x) = x$ if $x \not\in \{i_i, i_2, \ldots, i_k\}$.

The numbers $i_i, i_2, \ldots, i_k$ are called elements of the cycle. The cycle $\sigma$ is said to have length (or order) $k$. It is an easy matter to see that $\sigma^k = I$. In fact, $\sigma(i_1) = i_2, \sigma^2(i_1) = i_3, \ldots, \sigma^{k-1}(i_1) = i_k$.

For example, if $n = 5$, the cycle $\sigma = (2 \ 5 \ 4)$ of length 3 satisfies

$$\sigma(1) = 1, \ \sigma(2) = 5, \ \sigma(3) = 3, \ \sigma(4) = 2, \ \sigma(5) = 4$$

In short, $2 \mapsto 5 \mapsto 4 \mapsto 2$ (and $1 \mapsto 1, \ 3 \mapsto 3$.) The elements of $\sigma$ are 2, 4, and 5.

**Definition.** A transposition (or interchange) is a cycle $(i \ j)$ of length 2. Note that for transpositions $\tau$, we have $\tau^2 = I$.

**Theorem.** Any permutation in $S_n$ is a product of transpositions.

We prove this by induction on $n$. It is clearly true for $n = 2$. Assume that it is true for $n$ and let $\sigma \in S_{n+1}$. If $\sigma(n+1) = n+1$ we have already noted that we may regard $\sigma \in S_n$ and the result is true. Otherwise, $\sigma(n+1) = i$ where $1 \leq i \leq n$. Then $\tau = (n+1 \ i)\sigma$ maps $n+1$
into \( n + 1 \), and so we may regard \((n + 1 \ i)\sigma\) as an element of \( S_n \). By induction \((n + 1 \ i)\sigma = \rho\) is a product of transpositions. Multiplying by \((n + 1 \ i)\) and using \((n + 1 \ i)^2 = I\), we find that \(\sigma = (n + 1 \ i)\rho\), a product of transpositions. Note that the count of transpositions goes up by 1 in the inductive step, so we can state that every permutation in \( S_n \) is a product of at most \( n - 1 \) transpositions. Note also that we don’t really need \( S_n \subseteq S_{n+1}\). The induction could have been made on the largest element \( x \) such that \(\sigma(x) \neq x\), assuming \(\sigma \neq I\).

We’ll need the following lemma in what follows (Transpositions almost commute with permutations):

**Lemma.** Let \(\sigma \in S_n\). Then if \(\tau\) is a transposition in \( S_n\), we have \(\sigma\tau = \tau'\sigma\) for some transposition in \( S_n\).

Proof of Lemma. Suppose \(\tau = (i \ j)\). Let \(\sigma(i) = i'\) and \(\sigma(j) = j'\), and consider the transformation \(\rho = \sigma\tau\sigma^{-1}\). A simple check shows that

\[
\rho(i') = j'; \quad \rho(j') = j'; \quad \rho(k) = k \text{ if } k \neq i', j'
\]

Thus \(\rho\) is a transposition and since \(\rho = \sigma\tau\sigma^{-1}\), we have \(\rho\sigma = \sigma\tau\).

We now prove an important and non-trivial result:

**Theorem:** There is a transformation \(\text{sign}: S_n \to \{-1, 1\}\) with the following properties:

1) If \(\tau\) is a transposition then \(\text{sign}(\tau) = -1\).
2) For any \(\rho, \sigma \in S_n\), \(\text{sign}(\rho\sigma) = \text{sign}(\rho)\text{sign}(\sigma)\).

**Remark.** Since the sign of a transposition is \(-1\), and every permutation is a product of transpositions, it follows that the sign assignment is unique. The idea of this result is that if a permutation is written as a product of \( k \) transpositions, then the parity of \( k \) is unique.

We prove this by induction on \( n\), using the natural embedding of \( S_n \) in \( S_{n+1}\). It is clearly true for \( n = 1\), where \(\text{sign}(I) = 1\), and for \( n = 2\), where \(\text{sign}(I) = 1\) and \(\text{sign}(1\ 2) = -1\). Now assume this result true for \( n\), and we want to show how to extend the definition of sign to \( S_{n+1}\). If \(\sigma \in S_{n+1}\), we have \(\sigma(n + 1) = k\) for some \( k \) with \(1 \leq k \leq n + 1\). Thus if \( k = n + 1\), \(\sigma \in S_n\) and otherwise \((k + n + 1)\sigma \in S_n\). In this case \(\sigma \in (k + n + 1)S_n\). Define \(\tau_k = (k + n + 1)\) for \(1 \leq k \leq n\). For convenience, also define \(\tau_{n+1} = I\). Thus \( S_{n+1}\) is partitioned as

\[
S_{n+1} = \tau_1S_n \cup \tau_2S_n \cup \ldots \cup \tau_nS_n \cup \tau_{n+1}S_n
\]

We can now define \(\text{sign}\) on \( S_{n+1}\). If \(\rho \in S_n\), \(\text{sign}(\rho)\) is already defined. Otherwise, \(\rho = \tau_k\sigma\), where \(\sigma \in S_n\). In this case we naturally define \(\text{sign}(\rho) = -\text{sign}(\sigma)\).

Under this definition, it is clear that the sign of any transposition is \(-1\). We must prove that the sign of a product is the product of the signs. Let \(\rho_1 = \tau_r\sigma_1\) and let \(\rho_2 = \tau_s\sigma_s\), where \(1 \leq r, s \leq n + 1\) and each \(\sigma_i\) is in \( S_n\). Then

\[
\rho_1\rho_2 = \tau_r\sigma_1\tau_s\sigma_s = \tau_r\tau\sigma_1\sigma_s
\]
where, using the lemma, \( \tau \) is a transposition.

These notes present a somewhat more abstract version than the text. As usual, \( V \) is a \( n \)-dimensional vector space over a field \( F \).

**Definition.** Let \( f \) be an \( F \) valued function of \( k \) vector variables in \( V \). We write \( f : V^k \to F \). 
\( f \) is said to be \( k \)-linear if it is linear in each of its variables. Thus, for each \( i \) (\( 1 \leq i \leq k \)), and fixed vectors \( v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k \), and arbitrary vectors \( v \) and \( w \), and arbitrary constants \( a \) and \( b \), we have

\[
f(v_1, \ldots, v_{i-1}, av + bw, v_{i+1}, \ldots, v_k) = af(\ldots, v, \ldots) + bf(\ldots, w, \ldots)
\]

where the dots represent the fixed vectors \( v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k \).

The following is a generalization of the corresponding result when \( k = 1 \).

**Theorem.** Let \( e_1, \ldots, e_n \) be a basis for \( V \), and let \( k \geq 1 \) be given. Let \( a_{i_1i_2\ldots i_k} \in F \) for each \( k \)-tuple \( (i_1, \ldots, i_k) \) with \( 1 \leq i_1, \ldots, i_k \leq n \). Then there is one and only one \( k \)-linear functional \( f : V^k \to F \) such that

\[
f(e_{i_1}, \ldots, e_{i_k}) = a_{i_1i_2\ldots i_k}
\]

The proof is straightforward. Let’s first illustrate it for \( k = 2 \) and arbitrary \( n \). In this case, we have \( a_{ij} \in F \) for \( 1 \leq i, j \leq n \). Now if \( v, w \in V \), we have, uniquely,

\[
v = \sum_{i=1}^{n} x_i e_i, \quad w = \sum_{j=1}^{n} y_j e_j
\]

If \( f \) were bilinear in \( v \) and \( w \) with \( f(e_i, e_j) = a_{ij} \), we would have

\[
f(v, w) = f(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j) = \sum_{i,j=1}^{n} x_i y_j f(e_i, e_j) = \sum_{i,j=1}^{n} a_{ij} x_i y_j
\]

Conversely, if for given \( a_{ij} \), this formula is taken as the definition of \( f(v, w) \), it is easy to see it is bilinear (since it is linear in each of its variables) and that \( f(e_i, e_j) = a_{ij} \).

The general case is similar. We are burdened because we can’t use dummy variables \( i, j \), so we have to resort to the indices \( i_1, i_2, \ldots, i_k \).

Note the basis form of \( f(v_1, \ldots, v_k) \). If the vector \( v_k \) has coordinates \( (x_{k_1}, x_{k_2}, \ldots, x_{kn}) \) then we have

\[
f(v_1, \ldots, v_k) = \sum_{i_1, \ldots, i_k} a_{i_1 \ldots i_k} x_{i_1} \ldots x_{ki_k}
\]

In addition to \( k \)-linearity, it is necessary to consider skew symmetry.
**Definition.** Let $f$ be $k$-linear with $f : V^k \rightarrow F$. $f$ is said to be skew symmetric if $f = 0$ whenever two of the variables are equal.

In this case we also have, when $i < j$,

$$ f(\ldots, v_i, \ldots, v_j, \ldots) = -f(\ldots, v_j, \ldots, v_i, \ldots) $$

This follows from $k$-linearity as follows. (We take two variables for simplicity.) Since $f(v + w, v + w) = 0$, we have by $k$-linearity

$$ f(v, v) + f(v, w) + f(w, v) + f(w, w) = 0 $$

which implies $f(v, w) = -f(w, v)$, since $f(v, v) = f(w, w) = 0$ by skew symmetry. The condition $f(v, w) = -f(w, v)$ is often taken as the definition of skew symmetry.

We have seen that a $k$-linear function $f$ is given by a $k$-dimensional array $a_{i_1i_2\ldots i_k}$. Since this can be interpreted as the value of $f(e_{i_1}, \ldots, e_{i_k})$, where $e_1, \ldots, e_n$ are a basis, we can easily see what happens if $f$ is skew symmetric. We would have to have

$$ a_{i_1i_2\ldots i_k} = 0 \text{ if the } a_i \text{ are not distinct} $$

and

$$ a_{j_1j_2\ldots j_k} \text{ changes sign if two of indices are interchanged.} $$

It is easy to show that these conditions are sufficient for the corresponding $k$-linear function to be skew symmetric.

Here a little algebra is very helpful. If $k$ numbers $i_1, \ldots, i_k$ are rearranged to $j_1, \ldots, j_k$, the rearrangement is said to be an even permutation if the rearrangement can be achieved using an even number of interchanges. It is odd if can be achieved in an odd number of interchanges. An important result in algebra is that you can’t have a permutation which is even and odd. The sign of a permutation is said to be 1 if it is even, and $-1$ if odd. Thus the skew symmetry condition on $a_{i_1i_2\ldots i_k}$ is that

$$ a_{i_1i_2\ldots i_k} = \epsilon a_{j_1j_2\ldots j_k} $$

where $\epsilon$ is the sign of the permutation from the $i$’s to the $j$’s.

**Definition.** A volume element $K$ on an $n$ dimensional vector space $V$ is a skew symmetric, $n$-linear function of $n$ variables.

Note that this is the first time we set the number $k$ of vector variables equal to the dimension of the vector space. We may think of $K(v_1, \ldots, v_n)$ as the oriented volume of the $n$-dimensional parallelepiped formed by the $v_1, \ldots, v_n$. In the plane, we have an idea that the direction from $v_1$ to $v_2$ is opposite from the direction from $v_2$ to $v_1$. (Think counterclockwise vs clockwise.) This is determined by the sign of $K(v_1, v_2)$. Similarly in 3-space, think
right handed vs left handed system. And even on the lowly line, think left or right. Thus, 
n-linearity and skew symmetry have geometric meaning and they are partially discussed in 
the text.

**Theorem.** The volume elements on $V$ form a one-dimensional space.

**Proof.** Take a basis $e_1, \ldots, e_n$. Define the volume element $K_0$ by the condition $K_0(e_1, \ldots, e_n) = 1$. This uniquely defines $K_0$ since it implies

$$a_{1\ldots n} = 1, \quad a_{i_1 \ldots i_n} = \text{sign}(i_1 \ldots i_n) \text{ and } a_{i_1 \ldots i_n} = 0 \text{ if the } i_k \text{'s are not distinct.}$$

It is now an easy matter to verify that if $K$ is any volume element on $V$, and $k = K(e_1, \ldots, e_n)$, then $K = kK_0$.

We can now define the determinant of an $n \times n$ matrix $A$.

**Definition.** Let $\{e_1, \ldots, e_n\}$ be the standard basis in $F^n$, and let $K_0$ be the unique volume element of $F^n$ with $K_0(e_1, \ldots, e_n) = 1$. Then if $A$ is any $n \times n$ matrix with columns $A_1, \ldots, A_n$, we define

$$\det(A) = K_0(A_1, \ldots, A_n) = |A_1 \ldots A_n|$$

Here, the absolute value notation is used as a shorthand for the determinant. We often write $\det(A) = |A|$.

It follows directly from the definition that $\det I = 1$.

For example let’s compute

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$ 

Using the bracket notation, the two columns, and bilinearity and skew symmetry, we have

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = |ae_1 + ce_2, be_1 + de_2| = ab|e_1, e_1| + ad|e_1, e_2| + bc|e_2, e_1| + cd|e_2, e_2|$$

$$= ad|e_1, e_2| - bc|e_1, e_2| = ad - bc$$

We now derive an array of theorems. Initially, they concern the columns of a matrix, but as we shall soon see, $\det(A) = \det(A^t)$, so they will apply as well to the rows. Note the parallel with the elementary column operations.

**The Column Operations.**

(1) If two columns of a matrix are interchanged, the determinant changes its sign.

For a proof, note that this is precisely the skew symmetry condition of a volume element.

$$\ldots A_i \ldots A_j \ldots = -|\ldots A_j \ldots A_i \ldots|.$$ 

(2) If a column of a matrix is multiplied by a constant $c$, the determinant is multiplied by that constant. This follows from $n$-linearity: $|\ldots cA_i \ldots| = c|\ldots A_i \ldots|$.
(3) If a multiple of a column of $A$ is added to another column, the determinant is unchanged. Thus, $|\ldots A_i \ldots A_j \ldots| = |\ldots A_i \ldots (cA_i + A_j) \ldots|$. This follows from $n$-linearity and skew symmetry.

By definition, we have $\det(A) = |Ae_1, Ae_2, \ldots, Ae_n|$, where the $e_i$ are the standard basis vectors in $F^n$. We now go one step further with the following important theorem.

**Theorem.** If $A$ is an $n \times n$ matrix, and $v_1, \ldots, v_n$ are $n$ vectors, then

$$|Av_1, Av_2, \ldots, Av_n| = \det(A)|v_1, v_2, \ldots, v_n|$$

Note the geometric content here. $A$ maps a parallelepiped into another parallelepiped. Its volume gets magnified by $\det(A)$. So $\det(A)$, originally conceived as the volume of the image of the “unit cube” (defined to have volume 1), is now seen to be a magnifier of the volume of any parallelepiped.

**Proof.** Define the volume element $K(v_1, \ldots, v_n) = |Av_1, Av_2, \ldots, Av_n|$. (It is clearly $n$-linear and skew symmetric.) Since the volumes form a one dimensional space, there is a constant $c$ satisfying $K(v_1, \ldots, V_n) = c|v_1, v_2, \ldots, v_n|$. Thus,

$$|Av_1, Av_2, \ldots, Av_n| = c|v_1, v_2, \ldots, v_n|$$

To find $c$, substitute $v_i = e_i$ to obtain

$$|Ae_1, Ae_2, \ldots, Ae_n| = c|e_1, e_2, \ldots, e_n|$$

or

$$\det(A) = c$$

Thus,

$$|Av_1, Av_2, \ldots, Av_n| = \det(A)|v_1, v_2, \ldots, v_n|$$

which is the result.

We have an important corollary.

**Corollary.** If $A$ and $B$ are $n \times n$ matrices,

$$\det(AB) = \det(A) \det(B)$$

**Proof.**

$$\det(AB) = |ABe_1, \ldots, AB e_n| = \det(A)|Be_1, \ldots, Be_n| = \det(A) \det(B)$$

**Corollary.** If $A$ has an inverse then $\det(A) \neq 0$ and $\det(A^{-1}) = 1/\det(A)$. 


To see this, we have $AA^{-1} = I$. Taking determinants, we have $\det(A) \det(A^{-1}) = 1$, which is the result.

We now wish to show that $\det(A) = \det(A^t)$. This will allow us to use row operations as well as column operations as on page 5, since the transpose converts columns into rows, and vice versa. We prove this using elementary row and column matrices. The three types are as follows:

1. Row and column interchanges (illustrated for the interchange of row 1 and 2). The corresponding elementary matrix is

\[
E_r = E_c = \begin{bmatrix}
0 & 1 & 0 & \ldots \\
1 & 0 & 0 & \ldots \\
0 & 0 & 1 & \ldots \\
\vdots & \vdots & \vdots & \ddots
\end{bmatrix}
\]

In this case $E = E_r = E_c$ is symmetric, and we have $\det(E) = \det(E^t) = -1$, by skew symmetry.

2. Multiplying a row or column by $c$ (illustrated for row 1). The corresponding elementary matrix for row and column are the same:

\[
E_r = E_c = \begin{bmatrix}
c & O \\
O & I
\end{bmatrix}
\]

In this case $E = E_r = E_c$ is symmetric, and we have $\det(E) = \det(E^t) = c$ by $n$-linearity.

3. Adding a constant times a row (here 1) to another row (here 2), with the similar operation for columns. We have

\[
E_r = \begin{bmatrix}
1 & 0 & O \\
1 & 1 & O \\
O & O & I
\end{bmatrix}, \text{ and } E_c = \begin{bmatrix}
1 & c & O \\
0 & 1 & O \\
O & O & I
\end{bmatrix}
\]

Using simple column operations, we easily see that $\det(E_r) = \det(E_c) = 1$. Further, $E_r^t = E_c$. So in this case, we have $\det(E) = \det(E^t)$.

Thus we have the theorem:

**Theorem.** If $P$ is a product of elementary row or column matrices, then $\det(P) = \det(P^t)$.

To see this, note that $P = E_1E_2\ldots E_k$, so $P^t = E_2^tE_3^t\ldots E_k^tE_1^t$. The result follows by taking determinants and using the result for elementary matrices.

We can now prove the general theorem.
Theorem. If $A$ is any $n \times n$ matrix, then $\det(A) = \det(A^t)$.

Proof. By the section on row and column transformations, we know that by consecutive row and column transformations, it is possible to transform $A$ into a matrix $A'$ whose entries are 0 off the main diagonal. So $A'$ is symmetric: $(A')^t = A$. Thus, in matrix form,

$$A' = P A Q$$

where $P$ and $Q$ are respectively products of elementary row and column matrices. Taking determinants, we get

$$\det(A') = \det(P) \det(A) \det(Q)$$

Now take transposes and determinants:

$$A' = (A')^t = Q^t A^t P^t$$

so

$$\det(A') = \det(Q^t) \det(A^t) \det(P^t) = \det(Q) \det(A^t) \det(P)$$

Thus, $\det(P) \det(A) \det(Q) = \det(Q) \det(A^t) \det(P)$. Since $P$ and $Q$ are invertible, both $\det(P)$ and $\det(Q)$ are not zeros, and we can cancel the factor $\det(P) \det(Q)$ from both sides to get the result.

Evaluation of determinants. In principle, the definition itself will let us evaluate a determinant. However, for large matrices, the number of computations is prohibitive. The expansion of an $n \times n$ determinant has $n!$ terms. Thus, even a 5 by 5 has over a hundred terms, each term being a factor of five entries. The use of column and row transformations make life easier. For the record, we give the definition, using the definition. If $A$ is $n \times n$ whose entry in the $i$-th row and $j$-th column is $a_{ij}$, we have

$$\det(A) = \sum_{j_1 j_2 \ldots j_n} \text{sign}(j_1, j_2, \ldots, j_n) a_{1j_1} a_{2j_2} \ldots a_{nj_n}$$

The sum is taken over all permutations of $1, 2, \ldots, n$. This is simplified if many of the terms are 0, and for this we use row and column operations. We first take a useful step:

Theorem. If a matrix $A$ is an upper diagonal matrix, then $\det(A)$ is the product of the elements along the diagonal.

To see this, we have using the definition,

$$\det(A) = |a_{11} e_1, a_{12} e_2 + a_{22} e_2, a_{13} e_1 + a_{23} e_2 + a_{33} e_3, \ldots|$$

In this expansion, the second term’s $a_{12} e_1$ will contribute 0, because of the first term, and so only the $e_2$ term gives any contribution. Similarly, only the $e_3$ term in the third entry gives a contribution, etc. The required expansion is thus simply

$$\det(A) = |a_{11} e_1, a_{22} e_2, a_{33} e_3, \ldots| = a_{11} a_{22} a_{33} \ldots |e_1, e_2, e_3, \ldots| = a_{11} a_{22} \ldots a_{nn}$$

\(^1\)That is, all entries below the main diagonal are 0.
By transposing, we see that the same result is true for lower diagonal matrices. We can now augment the previous result that invertible matrices have non-zero determinants, by proving the converse.

**Theorem.** If det(A) ≠ 0, then A is invertible. For the proof, we know that by row and column operations, we can transform A into a diagonal matrix A′ with one’s and zero’s down the main diagonal. We thus have A′ = PAQ. Since P and Q are invertible, their determinants are not 0, and so det(A′) = det(P) det(A) det(Q) ≠ 0. This means that A′ has no zeros on the main diagonal, and so A′ = I. Thus I = PAQ, and A = P⁻¹Q⁻¹ is invertible. Summarizing, A is invertible if and only if its determinant is not 0.

The following result reduces the determination of an n × n determinant to a determinant of lower order.

**Theorem.** Suppose an n × n matrix A is of the form

\[
A = \begin{bmatrix}
a & x \\
0 & B
\end{bmatrix}
\]

where 0 is the (n − 1) × 1 zero vector and x is a 1 × (n − 1) row vector. Then det(A) = a det B.

**Proof.** Apply row and column transformations to A, subject to the conditions that they are used only on the last n − 1 rows or columns, and they do not use any of type (2) (multiplying a column or row by c ≠ 0.) If there are s interchanges of rows or columns, the resulting matrix A′ will have determinant (−1)^s det(A). In this way, we can transform A to a diagonal matrix A′:

\[
A' = \begin{bmatrix}
a & 0 & 0 & \ldots & 0 \\
0 & b_1 & 0 & \ldots & 0 \\
0 & 0 & b_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & b_{n-1}
\end{bmatrix} = \begin{bmatrix}
a & 0 \\
0 & B'
\end{bmatrix}
\]

Then (−1)^s det(A) = det(A′) = ab_1b_2...b_{n-1}. Since the row and column operations operated on the matrix B, we also have det(B′) = (−1)^s det(B) = b_1b_2...b_{n-1}. Thus, (−1)^s det(A) = a(−1)^s det(B). Canceling (−1)^s, we have the result.

If a matrix has only one non-zero element anywhere in the first column, this result can be used to calculate the determinant. A row transformation puts it into the first row, and the above theorem can be applied. The result is as follows:

**Theorem:** Suppose the first column of an n × n matrix is ae_i, and the matrix obtained by eliminating the first column and the i-th row is the (n − 1) × (n − 1) matrix B. Then det(A) = (−1)^{1+i} det(B).
For the proof, suppose the matrix is

$$A = \begin{bmatrix} 0 & r_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ a & r_i \\ \cdot & \cdot \\ 0 & r_n \end{bmatrix}$$

where $r_i$ is the $i$-th row to the right of the 0 or the $a$. Then, interchanging rows, we find

$$\det(A) = \begin{vmatrix} a & r_i \\ 0 & r_2 \\ \cdot & \cdot \\ 0 & r_{i-1} \\ a & r_i \\ \cdot & \cdot \\ 0 & r_{i+1} \\ \cdot & \cdot \\ 0 & r_n \end{vmatrix} = -a = -(-1)^{i-2}a = (-1)^{i+1}a$$

The extra factor $(-1)^{i-2}$ is needed, because it takes $i - 2$ steps to pull $r_1$ back to its rightful place before $r_2$, one step at a time.

A similar argument proves the following generalization:

If the $j$-th column of an $n \times n$ matrix $A$ is 0, except possibly for the $i$-th row whose entry is $a_{ij}$, and if $A_{ij}$ is the determinant of the $(n-1) \times (n-1)$ matrix obtained by eliminating the $i$-th row and $j$-th column of $A$, then $|A| = (-1)^{i+j}a_{ij}A_{ij}$.

**Definition.** Let $A$ be an $n \times n$ matrix, and $A_{ij}$ the determinant of the $(n-1) \times (n-1)$ matrix obtained by eliminating the $i$-th row and $j$-th column of $A$. The cofactor $C_{ij}$ is defined as $(-1)^{i+j}A_{ij}$. The sign introduced as a factor of the determinant $A_{ij}$ is governed by the familiar checkerboard pattern:

$$\begin{bmatrix} + & - & + & \ldots & \ldots \\ - & + & - & \ldots & \ldots \\ + & - & + & \ldots & \ldots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ldots & \ldots & \ldots & \ldots & + \end{bmatrix}$$

Thus the previous result can be stated: If all the entries of a matrix in a column (or row) are 0, with the exception of $a_{ij}$, then the determinant of the matrix is $a_{ij}C_{ij}$, where $C_{ij}$ is the cofactor of $a_{ij}$.  

10
By linearity, since the $j$-th row is $\sum_j a_{ij}e_i$, we have the following

**Expansion by Rows or Columns.** If $A$ is an $n \times n$ matrix with entries $a_{ij}$ and corresponding cofactors $C_{ij}$, then (expanding along a column $j$)

$$\det(A) = \sum_i a_{ij}C_{ij}$$

Similarly (expanding along a row $i$)

$$\det(A) = \sum_j a_{ij}C_{ij}$$

This has an interesting matrix version. Define the *adjoint matrix* $B = \text{adj}(A)$ by

$$b_{ij} = C_{ji}$$

where the $C$'s are the cofactors of the matrix $A$. (Note the interchange of row and column here.) Then the following is true:

$$A(\text{adj}(A)) = (\text{adj}(A))A = (\det A)I$$

(1)

and as a consequence, if $\det(A) \neq 0$, we have

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

To prove equation (1), set $B = \text{adj}(A)$, and $D=AB$. Then by definition,

$$d_{ij} = \sum_k a_{ik}b_{kj} = \sum_k a_{ik}C_{jk}$$

When $i = j$, this latter sum is simply the expansion by row $i$ of the determinant of $A$. Thus, the diagonal elements of $D$ are each $\det(A)$. When $i \neq j$, the sum is 0. For in this case, replace the $j$-th row of $A$ by its $i$-th row. In this case the determinant of the resulting matrix is 0, since its $i$-th and $j$-th row are the same. Thus, using the expansion along the $j$-th row we have $0 = \sum_k a_{jk}C_{jk} = \sum_k a_{ik}C_{jk}$, since the cofactor of the changed matrix is the same along row $j$. A similar argument applies to the product $BA$, proving the result. In principle this gives a formula for the determinant:

**Theorem.** If $A$ is invertible, then $A^{-1} = \frac{1}{\det(A)}\text{adj}(A)$.

For example, a familiar and useful formula is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

\(^2\text{The text calls it the classical adjoint.}\)
We finally note the well-known *Cramer’s rule* for solving linear equations.

**Cramer’s Rule.** Let $A$ be an $n \times n$ invertible matrix, and $b$ an $n \times 1$ column vector. Then the equation $Ax = b$ has the unique solution $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, with

$$x_i = \frac{\det(A_i)}{\det(A)}$$

where $A_i$ is the matrix $A$ in which the $i$-th column is replaced by the $B$, the column of constants, and as usual $x_i$ is the $i$-th component of $x$.

**Proof.** We know that the equation has a unique solution $x$. In vector form, the equation is

$$x_1c_1 + x_2c_2 + \cdots + x_nc_n = b$$

where $c_i$ is the $i$-th column of $A$. Thus,

$$|c_1, \ldots, c_{i-1}, b, c_{i+1}, \ldots, c_n| = |c_1, \ldots, c_{i-1}, x_1c_1 + x_2c_2 + \cdots + x_nc_n, c_{i+1}, \ldots, c_n|$$

$$= x_i |c_1, c_2, \ldots, c_n|$$

This gives the result.