Matrix - Determinant

From Leibniz formula to Laplace expansion

Determinant is defined **(only for square matrix)** by Leibniz formula, from which Laplace expansion can be derived. Leibniz formula together with Laplace expansion are the two most common representations of determinant. We will firstly present Leibniz formula, and then derive Laplace expansion using permutation (group theory). Leibniz formula states the determinant of a N×N square matrix A :

$$
\begin{aligned}
A &= (a_{i,j})_{i,j\in[1,N]} && \text{where } a_{ij} \text{ is the element in the i}^{\text{th}} \text{ row and the j}^{\text{th}} \text{ column} \\
\det(A) &= \sum_{\sigma\in S_N} \operatorname{sgn}(\sigma)\prod_{n=1}^{N} a_{\sigma(n),n} && \text{where } S_N = Perm(\{n, n\in[1,N]\}) \text{, all permutations of an integer set} \\
&= \sum_{\pi\in S_N} \operatorname{sgn}(\pi^{-1})\prod_{\pi(m)=1}^{N} a_{m,\pi(m)} && \text{where } m=\sigma(n) \text{ and } n=\pi(m) \text{, thus } \pi=\sigma^{-1} \text{, i.e. } \sigma,\pi\in S_N \\
&= \sum_{\pi\in S_N} \operatorname{sgn}(\pi^{-1})\prod_{m=1}^{N} a_{m,\pi(m)} && \text{since } \pi\in S_N \text{ is bijective function} \\
&= \sum_{\pi\in S_N} \operatorname{sgn}(\pi)\prod_{m=1}^{N} a_{m,\pi(m)} && \text{since } \operatorname{sgn}(\pi)=\operatorname{sgn}(\pi^{-1}) \\
&= \det(A^T)
\end{aligned}
$$

Therefore, determinant of a matrix equals to determinant of its inverse. Now, lets look at the determinant when N = 3. There are 3! = 6 different permutations.

$$
\begin{aligned}
\det(A) &= \sum_{\sigma\in S_3} \operatorname{sgn}(\sigma)\prod_{n=1}^{3} a_{\sigma(n),n} \\
&= \begin{array}{l}
\operatorname{sgn}(\sigma_A)\times a_{\sigma_A(1),1}\times a_{\sigma_A(2),2}\times a_{\sigma_A(3),3}+\operatorname{sgn}(\sigma_B)\times a_{\sigma_B(1),1}\times a_{\sigma_B(2),2}\times a_{\sigma_B(3),3}+ \\
\operatorname{sgn}(\sigma_C)\times a_{\sigma_C(1),1}\times a_{\sigma_C(2),2}\times a_{\sigma_C(3),3}+\operatorname{sgn}(\sigma_D)\times a_{\sigma_D(1),1}\times a_{\sigma_D(2),2}\times a_{\sigma_D(3),3}+ \\
\operatorname{sgn}(\sigma_E)\times a_{\sigma_E(1),1}\times a_{\sigma_E(2),2}\times a_{\sigma_E(3),3}+\operatorname{sgn}(\sigma_F)\times a_{\sigma_F(1),1}\times a_{\sigma_F(2),2}\times a_{\sigma_F(3),3}
\end{array}
\end{aligned}
$$

where :

$$
\begin{aligned}
\sigma_A &= \begin{pmatrix}1&2&3\\1&2&3\end{pmatrix} & \sigma_B &= \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix} \\
\sigma_C &= \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix} & \sigma_D &= \begin{pmatrix}1&2&3\\2&1&3\end{pmatrix} \\
\sigma_E &= \begin{pmatrix}1&2&3\\3&1&2\end{pmatrix} & \sigma_F &= \begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}
\end{aligned}
$$

Now, lets define submatrix, minor and cofactor. Given of a matrix A, submatrix $A_{p,q}$ is defined as :

$$
\begin{aligned}
A &= (a_{i,j})_{i,j\in[1,N]} \\
A_{p,q} &= (a_{i,j})_{i,j\in[1,N],i\neq p,j\neq q} \\
&= (b_{i,j})_{i,j\in[1,N-1]}
\end{aligned}
$$

$$
\text{where} \quad b_{i,j} = \begin{cases}
a_{i,j} & 1\le i<p \quad and \quad 1\le j<q \\
a_{i,j+1} & 1\le i<p \quad and \quad q\le j\le N-1 \\
a_{i+1,j} & p\le i\le N-1 \quad and \quad 1\le j<q \\
a_{i+1,j+1} & p\le i\le N-1 \quad and \quad q\le j\le N-1
\end{cases} \qquad \text{(equation 1)}
$$

Minor $M_{p,q}$ and cofactor $C_{p,q}$ are defined as determinant of submatrix and signed determinant of submatrix :

$$
\begin{aligned}
M_{p,q} &= \det(A_{p,q}) \\
C_{p,q} &= (-1)^{p+q} M_{p,q} \\
&= (-1)^{p+q}\det(A_{p,q})
\end{aligned}
$$

Lets we derive the Laplace expansion. According to Leibniz formula, determinant is sum over all permutations. Now we randomly pick one number p from 1 to N, i.e. $p\in[1,N]$, then we have :

$$\begin{aligned}
\det(A) &= \sum_{\sigma\in S_N}\operatorname{sgn}(\sigma)\prod_{n=1}^{N} a_{n,\sigma(n)} \\
&= \sum_{q=1}^{N}\left(\sum_{\sigma\in S_N, s.t.\sigma(p)=q}\operatorname{sgn}(\sigma)\prod_{n=1}^{N} a_{n,\sigma(n)}\right) \\
&= \sum_{q=1}^{N}\left(\sum_{\sigma\in S_N, s.t.\sigma(p)=q}\operatorname{sgn}(\sigma)a_{p,\sigma(p)}\prod_{n=1,n\neq p}^{N} a_{n,\sigma(n)}\right) \\
&= \sum_{q=1}^{N}\left(\sum_{\sigma\in S_N, s.t.\sigma(p)=q}\operatorname{sgn}(\sigma)a_{p,q}\prod_{n=1}^{N-1} b_{n,\pi(n)}\right) && \text{(see remark 1)} \\
&= \sum_{q=1}^{N}\left(a_{p,q}\sum_{\pi\in S_{N-1}}\operatorname{sgn}(\sigma)\prod_{n=1}^{N-1} b_{n,\pi(n)}\right) && \text{(see remark 2)} \\
&= \sum_{q=1}^{N}\left(a_{p,q}(-1)^{p+q}\sum_{\pi\in S_{N-1}}\operatorname{sgn}(\pi)\prod_{n=1}^{N-1} b_{n,\pi(n)}\right) && \text{(see remark 3)} \\
&= \sum_{q=1}^{N}a_{p,q}(-1)^{p+q}\det(A_{p,q}) && \text{(since } \det(A_{p,q})=\sum_{\pi\in S_{N-1}}\operatorname{sgn}(\pi)\prod_{n=1}^{N-1} b_{n,\pi(n)}\text{)} \\
&= \sum_{q=1}^{N}a_{p,q}C_{p,q} && \text{(where } C_{p,q} \text{ is cofactor)}
\end{aligned}$$

<u>Remark 1</u>

For every permutation $\sigma\in S_N$ s.t. $\sigma(p)=q$, we can always find permutation $\pi\in S_{N-1}$, so that element $(n, \sigma(n))$ in matrix A and element $(n, \pi(n))$ in submatrix $A_{p,q}$ form one to one correspondence.

$$\text{i.e.}\quad \pi(n) = \begin{cases}
\sigma(n) & 1\le n<p \quad and \quad 1\le\sigma(n)<q \quad or \quad 1\le\pi(n)<q \\
\sigma(n)-1 & 1\le n<p \quad and \quad q<\sigma(n)\le N \quad or \quad q\le\pi(n)\le N-1 \\
\sigma(n+1) & p\le n\le N-1 \quad and \quad 1\le\sigma(n+1)<q \quad or \quad 1\le\pi(n)<q \\
\sigma(n+1)-1 & p\le n\le N-1 \quad and \quad q<\sigma(n+1)\le N \quad or \quad q\le\pi(n)\le N-1
\end{cases} \qquad \text{(equation 2)}$$

$$\text{or}\quad \sigma(n) = \begin{cases}
\pi(n) & 1\le n<p \quad and \quad 1\le\sigma(n)<q \quad or \quad 1\le\pi(n)<q \\
\pi(n)+1 & 1\le n<p \quad and \quad q<\sigma(n)\le N \quad or \quad q\le\pi(n)\le N-1 \\
\pi(n-1) & p<n\le N \quad and \quad 1\le\sigma(n)<q \quad or \quad 1\le\pi(n-1)<q \\
\pi(n-1)+1 & p<n\le N \quad and \quad q<\sigma(n)\le N \quad or \quad q\le\pi(n-1)\le N-1
\end{cases} \qquad \text{(equation 3)}$$

According to definition of submatrix in equation 1, by putting i = n and j = π(n), we have :

$$b_{i,j} = \begin{cases}
a_{i,j} & 1\le i<p \quad and \quad 1\le j<q \\
a_{i,j+1} & 1\le i<p \quad and \quad q\le j\le N-1 \\
a_{i+1,j} & p\le i\le N-1 \quad and \quad 1\le j<q \\
a_{i+1,j+1} & p\le i\le N-1 \quad and \quad q\le j\le N-1
\end{cases} \qquad \text{(by equation 1)}$$

$$\begin{aligned}
\Rightarrow\quad b_{n,\pi(n)} &= \begin{cases}
a_{n,\pi(n)} & 1\le n<p \quad and \quad 1\le\pi(n)<q \\
a_{n,\pi(n)+1} & 1\le n<p \quad and \quad q\le\pi(n)\le N-1 \\
a_{n+1,\pi(n)} & p\le n\le N-1 \quad and \quad 1\le\pi(n)<q \\
a_{n+1,\pi(n)+1} & p\le n\le N-1 \quad and \quad q\le\pi(n)\le N-1
\end{cases} \\
&= \begin{cases}
a_{n,\sigma(n)} & 1\le n<p \quad and \quad 1\le\sigma(n)<q \\
a_{n,\sigma(n)} & 1\le n<p \quad and \quad q<\sigma(n)\le N \\
a_{n+1,\sigma(n+1)} & p\le n\le N-1 \quad and \quad 1\le\sigma(n+1)<q \\
a_{n+1,\sigma(n+1)} & p\le n\le N-1 \quad and \quad q<\sigma(n+1)\le N
\end{cases} && \text{(by equation 2)} \\
&= \begin{cases}
a_{n,\sigma(n)} & 1\le n<p \quad and \quad \sigma(n)\neq q \\
a_{n+1,\sigma(n+1)} & p\le n\le N-1 \quad and \quad \sigma(n+1)\neq q
\end{cases} && \text{(since } \sigma(n)\neq q \text{ except } \sigma(p)=q\text{)}
\end{aligned}$$

<u>Remark 2</u>

Firstly element $a_{p,q}$ is independent of permutation σ, hence it can be taken out of the summation over σ. Secondly, there is one to one correspondence between $\sigma\in S_N$ s.t. $\sigma(p)=q$ and $\pi\in S_{N-1}$, thus the summation over all $\sigma\in S_N$ s.t. $\sigma(p)=q$ is the same as summation over all $\pi\in S_{N-1}$.

Remark 3

For every $\sigma \in S_N$ s.t. $\sigma(p)=q$, we can always express it in terms of $\pi \in S_{N-1}$ as :

σ	=	$\mu \circ \pi \circ \eta$	where μ, η are cyclic permutations
η	=	$(N, N-1, N-2, ..., p+1, p)$	cycle notation, apply this mapping first
μ	=	$(q, q+1, q+2, ..., N-1, N)$	cycle notation, apply this mapping last

Lets verify it by expanding the permutation composition (start with η, followed with π, and finally with μ) using the Cauchy's two line notation.

$$\mu \circ \pi \circ \eta = \begin{pmatrix} 1 & \to & 1 & \to & \pi(1) & \to & \mu(\pi(1)) \\ 2 & \to & 2 & \to & \pi(2) & \to & \mu(\pi(2)) \\ 3 & \to & 3 & \to & \pi(3) & \to & \mu(\pi(3)) \\ ... & & & & & & \\ p-1 & \to & p-1 & \to & \pi(p-1) & \to & \mu(\pi(p-1)) \\ p & \to & N & \to & N & \to & \mu(N)=q \\ p+1 & \to & p & \to & \pi(p) & \to & \mu(\pi(p)) \\ ... & & & & & & \\ q-1 & \to & q-2 & \to & \pi(q-2) & \to & \mu(\pi(q-2)) \\ q & \to & q-1 & \to & \pi(q-1) & \to & \mu(\pi(q-1)) \\ q+1 & \to & q & \to & \pi(q) & \to & \mu(\pi(q)) \\ ... & & & & & & \\ N-1 & \to & N-2 & \to & \pi(N-2) & \to & \mu(\pi(N-2)) \\ N & \to & N-1 & \to & \pi(N-1) & \to & \mu(\pi(N-1)) \end{pmatrix} \qquad \text{(recall } \pi(N) \text{ is undefined as } \pi \in S_{N-1}\text{)}$$

$$\Rightarrow \quad (\mu \circ \pi \circ \eta)(n) = \begin{cases} \pi(\eta(n)) & \pi(\eta(n)) < q \\ \pi(\eta(n))+1 & q \le \pi(\eta(n)) \le N-1 \end{cases}$$

$$= \begin{cases} \pi(n) & 1 \le n < p \quad and \quad 1 \le \pi(n) < q \\ \pi(n)+1 & 1 \le n < p \quad and \quad q \le \pi(n) \le N-1 \\ \pi(n-1) & p \le n \le N \quad and \quad 1 \le \pi(n-1) < q \\ \pi(n-1)+1 & p \le n \le N \quad and \quad q \le \pi(n-1) \le N-1 \end{cases} \qquad \text{(since } \eta(n) = \begin{cases} n & 1 \le n < p \\ n-1 & p \le n \le N \end{cases}\text{)}$$

$$= \sigma(n) \qquad \text{(by equation 3)}$$

Hence Laplace expansion is just a recursive form of determinant, and since $p,q \in [1,N]$, we can expand the determinant into sum of cofactors along any row p or along any column q. Finally, lets do a verification by checking the number of terms in Leibniz formula and in Laplace expansion.

number of terms in Leibniz formula	=	$sizeof(S_N)$	=	$N!$
number of terms in Laplace expansion	=	$f(N)$	=	$Nf(N-1)$
			=	$N(N-1)f(N-2)$
			=	$N!$

Property 1 : Swapping of two rows

Determinant of a matrix will have its sign reversed if two of its **consecutive rows** are swapped. Consider a N×N matrix A, if row p and row p+1 are swapped to generate a new matrix A', determinant of A' equals to negative of determinant of A. We prove by induction. For 2×2 matrix :

$$\begin{aligned} \det(A) &= a_{1,1}a_{2,2} - a_{2,1}a_{1,2} \\ \det(A') &= a_{2,1}a_{1,2} - a_{1,1}a_{2,2} \\ &= -\det(A) \end{aligned}$$

Now, by assuming that (N-1)×(N-1) case is true $\forall p \in [1,N-1]$, lets consider the N×N case, we have Laplace expansion of determinant along the 1st column of matrix A' as :

$$\begin{aligned} \det(A') &= \textstyle\sum_{n=1}^{N} a'_{n,1}(-1)^{n+1}\det(A'_{n,1}) \\ &= +(\textstyle\sum_{n=1}^{p-1} a'_{n,1}(-1)^{n+1}\det(A'_{n,1})) + (a'_{p,1}(-1)^{p+1}\det(A'_{p,1})) + (a'_{p+1,1}(-1)^{p+2}\det(A'_{p+1,1})) + (\textstyle\sum_{n=p+2}^{N} a'_{n,1}(-1)^{n+1}\det(A'_{n,1})) \\ &= -(\textstyle\sum_{n=1}^{p-1} a_{n,1}(-1)^{n+1}\det(A_{n,1})) + (a_{p+1,1}(-1)^{p+1}\det(A_{p+1,1})) + (a_{p,1}(-1)^{p+2}\det(A_{p,1})) - (\textstyle\sum_{n=p+2}^{N} a_{n,1}(-1)^{n+1}\det(A_{n,1})) \\ &= -(\textstyle\sum_{n=1}^{p-1} a_{n,1}(-1)^{n+1}\det(A_{n,1})) - (a_{p,1}(-1)^{p}\det(A_{p,1})) - (a_{p+1,1}(-1)^{p+1}\det(A_{p+1,1})) - (\textstyle\sum_{n=p+2}^{N} a_{n,1}(-1)^{n+1}\det(A_{n,1})) \\ &= -\textstyle\sum_{n=1}^{N} a_{n,1}(-1)^{n+1}\det(A_{n,1}) \\ &= -\det(A) \end{aligned}$$

Remark :

$$a'_{n,m} = \begin{cases} a_{n,m} & n \neq p, p+1 \quad \forall m \in [1,N] \\ a_{p+1,m} & n = p \quad \forall m \in [1,N] \\ a_{p,m} & n = p+1 \quad \forall m \in [1,N] \end{cases}$$

$$\det(A'_{n,1}) = \begin{cases} -\det(A_{n,1}) & n \neq p, p+1 \quad \forall m \in [1,N] \quad \textit{induction-assumption} \quad (N-1)\times(N-1)\_case \\ \det(A_{p+1,1}) & n = p \quad \forall m \in [1,N] \\ \det(A_{p,1}) & n = p+1 \quad \forall m \in [1,N] \end{cases}$$

The above property can be extended to **swapping between any two rows** (p and q). What we need to do is to count the number of consecutive row swappings needed so as to achieve a swapping between row p and row q. Without loss of generality, we assume q > p. There are 2 steps : (1) q–p consecutive row swapping to **sink row p**, until it is lower than row q, followed by (2) q–(p+1) consecutive row swapping to **float row q**, until it reaches the original position of row p. The total number of consecutive row swap is thus q–p + q–(p+1) = 2(q–p)+1. Therefore, determinant of the resulting matrix equals to determinant of the original matrix times $(-1)^{2(q+p)-1} = -1$. Hence, the swapping of any two non consecutive rows does also result in a reversely signed determinant. Furthermore, with the help of this property, we can find out the determinant of **matrix having two identical rows**. Suppose matrix A is a square matrix with row p equivalent to row q, hence we can generate matrix A' by swapping row p and row q of matrix A. We have :

$$\begin{aligned} &\det(A') = \det(A) && \text{(since A' = A as row p and row q are the same)} \\ &\det(A') = -\det(A) && \text{(since determinant becomes negative when 2 rows are swapped)} \\ \Rightarrow\ &\det(A) = -\det(A) = 0 \end{aligned}$$

Property 2 : Multilinear property

Multilinear property of determinant states that if A, B, C are 3 N×N matrices which are different in row p only :

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p,1} & a_{p,2} & \cdots & a_{p,N} \\ \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,N} \end{bmatrix} \quad B = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ b_1 & b_2 & \cdots & b_N \\ \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,N} \end{bmatrix} \quad C = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots \\ a_{p,1}+sb_1 & a_{p,2}+sb_2 & \cdots & a_{p,N}+sb_N \\ \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,N} \end{bmatrix}$$

then we have : $\det(C) = \det(A) + s\det(B)$ and **beware that** : $C \neq A + sB$. Again, we prove by induction. For 2×2 matrix :

$$\begin{aligned} \det(A) &= a_{1,1}a_{2,2} - a_{2,1}a_{1,2} \\ \det(B) &= a_{1,1}b_2 - b_1 a_{1,2} \\ \det(C) &= a_{1,1}(a_{2,2}+sb_2) - (a_{2,1}+sb_1)a_{1,2} \\ &= a_{1,1}a_{2,2} - a_{2,1}a_{1,2} + s(a_{1,1}b_2 - b_1 a_{1,2}) \\ &= \det(A) + s\det(B) \end{aligned}$$

Now, by assuming that (N-1)×(N-1) case is true ∀p∈[1,N-1], lets consider the N×N case, we have Laplace expansion of determinant along the 1st column of matrix C as :

$$\begin{aligned} \det(C) &= \textstyle\sum_{n=1}^{N} c_{n,1}(-1)^{n+1}\det(C_{n,1}) \\ &= (\textstyle\sum_{n=1}^{p-1} c_{n,1}(-1)^{n+1}\det(C_{n,1})) + (c_{p,1}(-1)^{p+1}\det(C_{p,1})) + (\textstyle\sum_{n=p+1}^{N} c_{n,1}(-1)^{n+1}\det(C_{n,1})) \\ &= (\textstyle\sum_{n=1}^{p-1} a_{n,1}(-1)^{n+1}(\det(A_{n,1}) + s\det(B_{n,1}))) + ((a_{p,1}+sb_1)(-1)^{p+1}\det(A_{p,1})) + (\textstyle\sum_{n=p+1}^{N} a_{n,1}(-1)^{n+1}(\det(A_{n,1}) + s\det(B_{n,1}))) \quad \text{(remark)} \\ &= (\textstyle\sum_{n=1}^{p-1} a_{n,1}(-1)^{n+1}\det(A_{n,1})) + (a_{p,1}(-1)^{p+1}\det(A_{p,1})) + (\textstyle\sum_{n=p+1}^{N} a_{n,1}(-1)^{n+1}\det(A_{n,1})) + \\ &\quad s(\textstyle\sum_{n=1}^{p-1} a_{n,1}(-1)^{n+1}\det(B_{n,1})) + s(b_1(-1)^{p+1}\det(A_{p,1})) + s(\textstyle\sum_{n=p+1}^{N} a_{n,1}(-1)^{n+1}\det(B_{n,1})) \\ &= (\textstyle\sum_{n=1}^{p-1} a_{n,1}(-1)^{n+1}\det(A_{n,1})) + (a_{p,1}(-1)^{p+1}\det(A_{p,1})) + (\textstyle\sum_{n=p+1}^{N} a_{n,1}(-1)^{n+1}\det(A_{n,1})) + \\ &\quad s(\textstyle\sum_{n=1}^{p-1} b_{n,1}(-1)^{n+1}\det(B_{n,1})) + s(b_{p,1}(-1)^{p+1}\det(B_{p,1})) + s(\textstyle\sum_{n=p+1}^{N} b_{n,1}(-1)^{n+1}\det(B_{n,1})) \quad \text{(remark)} \\ &= (\textstyle\sum_{n=1}^{N} a_{n,1}(-1)^{n+1}\det(A_{n,1})) + s(\textstyle\sum_{n=1}^{N} b_{n,1}(-1)^{n+1}\det(B_{n,1})) \\ &= \det(A) + s\det(B) \end{aligned}$$

Remark :

$$b_{n,m} = \begin{cases} a_{n,m} & n \neq p \quad \forall m \in [1.N] \\ b_m & n = p \quad \forall m \in [1.N] \end{cases}$$

$$c_{n,m} = \begin{cases} a_{n,m} & n \neq p \quad \forall m \in [1.N] \\ a_{n,m} + sb_m & n = p \quad \forall m \in [1.N] \end{cases}$$

$$\det(B_{n,m}) = \begin{cases} not-needed & n \neq p \quad \forall m \in [1.N] \\ \det(A_{p,m}) & n = p \quad \forall m \in [1.N] \end{cases}$$

$$\det(C_{n,m}) = \begin{cases} \det(A_{p,m}) + s\det(B_{p,m}) & n \neq p \quad \forall m \in [1.N] \quad induction-assumption \quad (N-1)\times(N-1)\_case \\ \det(A_{p,m}) & n = p \quad \forall m \in [1.N] \end{cases}$$

The above property can be extended to **elementary row addition**. Suppose we generate matrix A' from matrix A, by adding row p of matrix A with row q of matrix A scaled by s, then determinant of A' equals to determinant of A, i.e. determinant doesn't change after **elementary row addition** (and also for **elementary column addition**). Here is the proof for elementary row addition.

$$\begin{aligned}
\det(A') &= \det\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N} \\ \dots & \dots & \dots & \dots \\ a_{p,1}+sa_{q,1} & a_{p,2}+sa_{q,2} & \dots & a_{p,N}+sa_{q,N} \\ \dots & \dots & \dots & \dots \\ a_{q,1} & a_{q,2} & \dots & a_{q,N} \\ \dots & \dots & \dots & \dots \\ a_{N,1} & a_{N,2} & \dots & a_{N,N} \end{bmatrix} \\
&= \det\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N} \\ \dots & \dots & \dots & \dots \\ a_{p,1} & a_{p,2} & \dots & a_{p,N} \\ \dots & \dots & \dots & \dots \\ a_{q,1} & a_{q,2} & \dots & a_{q,N} \\ \dots & \dots & \dots & \dots \\ a_{N,1} & a_{N,2} & \dots & a_{N,N} \end{bmatrix} + s\det\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,N} \\ \dots & \dots & \dots & \dots \\ a_{q,1} & a_{q,2} & \dots & a_{q,N} \\ \dots & \dots & \dots & \dots \\ a_{q,1} & a_{q,2} & \dots & a_{q,N} \\ \dots & \dots & \dots & \dots \\ a_{N,1} & a_{N,2} & \dots & a_{N,N} \end{bmatrix} \qquad \text{(det of matrix having two identical rows is zero)} \\
&= \det(A) \qquad \text{(i.e. } \det(EA) = \det(A) \text{ and } \det(AF) = \det(A)\text{)}
\end{aligned}$$

Property 3 : Determinant of upper triangular matrix

The determinant of upper triangular matrix U (or lower triangular matrix L) is the product of the diagonal elements. Lets prove the upper triangular matrix case (lower triangular matrix case is the same, thus omitted). It is easier to prove with the Leibniz formula. Suppose A is a N×N upper triangular matrix, we have :

$$\det(A) = \sum_{\sigma\in S_N} \operatorname{sgn}(\sigma)\prod_{n=1}^{N} a_{\sigma(n),n} \qquad \text{where} \qquad \prod_{n=1}^{N} a_{\sigma(n),n} \neq 0 \quad \text{only if } a_{\sigma(n),n} \neq 0 \;\; \forall n \in [1,N]$$

$$\begin{aligned}
&\text{for n = 1,} && a_{\sigma(1),1} \neq 0 && \text{only if } \sigma(1)=1 \\
&\text{for n = 2,} && a_{\sigma(2),2} \neq 0 && \text{only if } \sigma(2)=1 \text{ or } \sigma(2)=2 \\
&\text{for n = 3,} && a_{\sigma(3),3} \neq 0 && \text{only if } \sigma(3)=1 \text{ or } \sigma(3)=2 \text{ or } \sigma(3)=3
\end{aligned}$$

and so on, hence the only possible choice of σ is : $\sigma' = \begin{pmatrix} 1 & 2 & 3 & \dots & N \\ 1 & 2 & 3 & \dots & N \end{pmatrix}$ with $\operatorname{sgn}(\sigma') = 1$

$$\Rightarrow \det(A) = \sum_{\sigma\in S_N} \operatorname{sgn}(\sigma)\prod_{n=1}^{N} a_{\sigma(n),n} = \sum_{\substack{\sigma\in S_N \\ \sigma\neq\sigma'}} (\operatorname{sgn}(\sigma)\times 0) + \operatorname{sgn}(\sigma')\prod_{n=1}^{N} a_{n,n} = \prod_{n=1}^{N} a_{n,n}$$

Property 4 : Determinant of matrix product

Determinant of matrix product equals to the product of individual matrix's determinant. Since all **square matrices** can be represented as : the result of applying a sequence of <u>elementary row additions</u> (or <u>elementary column additions</u>) to an upper triangular matrix U (or a lower triangular matrix L), i.e. there are 4 different decompositions, we have :

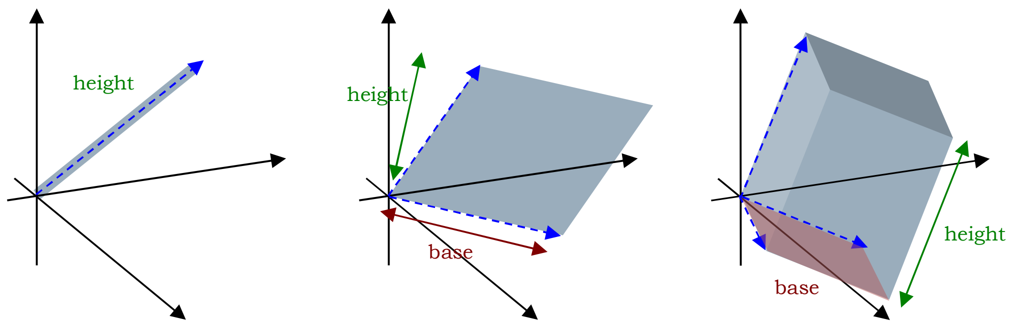
$$\begin{aligned}
A &= E_T E_{T-1}\dots E_3 E_2 E_1 U && \text{(where U is an upper triangular matrix)} \\
B &= V F_1 F_2 F_3 \dots F_{T-1} F_T && \text{(where V is also an upper triangular matrix)} \\
\det(AB) &= \det(E_T E_{T-1}\dots E_3 E_2 E_1 U V F_1 F_2 F_3 \dots F_{T-1} F_T) \\
&= \det(U V F_1 F_2 F_3 \dots F_{T-1} F_T) && \text{(by property 2 for elementary row addition)} \\
&= \det(UV) && \text{(by property 2 for elementary column addition)} \\
&= \textstyle\prod_{n=1}^{N} U_{n,n} V_{n,n} && \text{(since } UV \text{ is also an upper triangular matrix)} \\
&= \textstyle\prod_{n=1}^{N} U_{n,n} \prod_{n=1}^{N} V_{n,n} \\
&= \det(U)\det(V) \\
&= \det(E_T E_{T-1}\dots E_3 E_2 E_1 U)\det(V F_1 F_2 F_3 \dots F_{T-1} F_T) && \text{(reversing property 2)} \\
&= \det(A)\det(B)
\end{aligned}$$

Volume of a parallelepiped

[Remark : In all the above parts, we talk about determinant of ***square matrix*** *A. In here, matrix A is* ***no longer square****, it can be* ***rectangular****.]* First of all, lets define parallelepiped. Given N vectors in M dimensional space, i.e. $x_n \in \Re^M$, s.t. $n \in [1,N]$, then the vector set defines a parallelepiped as the bounded space spanned by the vector set :

$$
\begin{aligned}
parallelepiped &= \{v : v = \textstyle\sum_{n=1}^{N} w_n x_n, w_n \in [0,1], \forall n \in [1,N]\} \\
span\_space &= \{v : v = \textstyle\sum_{n=1}^{N} w_n v_n, w_n \in \Re^1, \forall n \in [1,N]\}
\end{aligned}
$$

where the vectors are edges of the parallelepiped. Please note that parallelepiped defined by a vector set is different from span space of the vector set (see Gram Schmidt process document). Both definitions are listed above for the sake of comparison. Now, we can define volume of parallelepiped by induction on N (not on M). This figure below illustrates the volume of parallelepiped spanned by 1, 2 or 3 vectors in 3D space, i.e. N = 1, 2, 3 and M = 3.



Suppose we pick a vector x_N, then volume is defined as the product of height and base, where base is the volume of the parallelopiped defined by vector set $\{x_1, x_2, \ldots, x_{N-1}\}$, while height is the orthogonal distance from x_N to the span of vector set $\{x_1, x_2, \ldots, x_{N-1}\}$. More specifically, height can be defined by the Gram Schmidt process of $\{x_1, x_2, \ldots, x_{N-1}\}$:

$$
\begin{aligned}
u &= \{u_m : < u_{m1}, u_{m2} > = |u_{m1}||u_{m2}|\delta_{m1,m2}, \forall m1, m2 \in [1, N-1]\} \qquad \text{such that } span(u) = span(\{x_1, x_2, \ldots, x_{N-1}\}) \\
h_N &= x_N - \textstyle\sum_{m=1}^{N-1} proj_{u_m}(x_N)
\end{aligned}
$$

In general, as induction proceeds, height h_n is the orthogonal distance from x_n to the span of remaining vector set.

$$
\begin{aligned}
u &= \{u_m : < u_{m1}, u_{m2} > = |u_{m1}||u_{m2}|\delta_{m1,m2}, \forall m1, m2 \in [1, n-1]\} \qquad \text{such that } span(u) = span(\{x_1, x_2, \ldots, x_{n-1}\}) \\
h_n &= x_n - \textstyle\sum_{m=1}^{n-1} proj_{u_m}(x_n)
\end{aligned}
$$

Please note that : (1) the same set of u can be used in all induction steps, i.e. no need to perform Gram Schmidt process per each induction step, and (2) we can randomly pick one vector in each induction step, the ordering does not matter. Therefore the volume is :

$$
\begin{aligned}
vol_N &= vol_{N-1} h_N \\
&= vol_{N-1}(x_N - \textstyle\sum_{m=1}^{N-1} proj_{u_m}(x_N)) \\
&= vol_{N-2}(x_{N-1} - \textstyle\sum_{m=1}^{N-2} proj_{u_m}(x_{N-1}))(x_N - \textstyle\sum_{m=1}^{N-1} proj_{u_m}(x_N)) \\
&= \ldots \\
&= vol_2(x_3 - \textstyle\sum_{m=1}^{2} proj_{u_m}(x_3))(x_4 - \textstyle\sum_{m=1}^{3} proj_{u_m}(x_4))\ldots\ldots(x_N - \textstyle\sum_{m=1}^{N-1} proj_{u_m}(x_N)) \\
&= vol_1(x_2 - \textstyle\sum_{m=1}^{1} proj_{u_m}(x_2))(x_3 - \textstyle\sum_{m=1}^{2} proj_{u_m}(x_3))(x_4 - \textstyle\sum_{m=1}^{3} proj_{u_m}(x_4))\ldots\ldots(x_N - \textstyle\sum_{m=1}^{N-1} proj_{u_m}(x_N)) \\
&= x_1(x_2 - \textstyle\sum_{m=1}^{1} proj_{u_m}(x_2))(x_3 - \textstyle\sum_{m=1}^{2} proj_{u_m}(x_3))(x_4 - \textstyle\sum_{m=1}^{3} proj_{u_m}(x_4))\ldots\ldots(x_N - \textstyle\sum_{m=1}^{N-1} proj_{u_m}(x_N))
\end{aligned}
$$

Now we pack all the vectors into a matrix A as row data, then we have :

$$
\begin{aligned}
vol_N^2 &= \det(AA^T) && \text{in general} \qquad \text{(equation 1)} \\
vol_N^2 &= \det(A)\det(A^T) = \det(A)^2 && \text{if } N = M \\
vol_N &= \det(A) && \text{if } N = M \\
A &= \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_N \end{pmatrix} && \text{where A is rectangular matrix, i.e. } N \times M
\end{aligned}
$$

Lets prove equation 1 by induction on N. When N equals to 1, then LHS of equation 1 equals to length of x_1, while RHS of equation 1 equals to determinant of a 1×1 matrix, i.e. $x_1^T x_1$, hence LHS = RHS.

Suppose equation 1 holds true for case N–1, lets consider the case N. Lets define the following matrices.

$$A = \begin{pmatrix} B \\ x_N \end{pmatrix} \quad \text{and} \quad A' = \begin{pmatrix} B \\ C \end{pmatrix}$$

where B is a (N–1)×N matrix and C is a 1×N matrix, such that :

$$B = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_{N-2} \\ x_{N-1} \end{pmatrix} \quad \text{and} \quad C = x_N - \sum_{n=1}^{N-1} proj_{u_n}(x_N) = x_N - \sum_{n=1}^{N-1} w_n x_n$$

The expansion of C is always valid, because vector set u = $\{u_n, n \in [1,N-1]\}$ and vector set x = $\{x_n, n \in [1,N-1]\}$ share the same span (according to the definition of Gram Schmidt process), hence it is always true that :

$$\exists w = \{w_n, n \in [1, N-1]\} \quad \text{s.t.} \quad \sum_{n=1}^{N-1} w_n x_n = \sum_{n=1}^{N-1} proj_{u_n}(x_N)$$

The trick lies in the connection between elementary row addition and Gram Schmidt process : matrix A and matrix A' differ by the last row only, and the last row of A' is orthogonal to all the N–1 rows above, we can always find a set of elementary row addition, so that :

$$\begin{aligned}
A &= E_{N-1} \dots E_3 E_2 E_1 A' \\
E_1 &= E_{add}(N,1,-w_1) \\
E_2 &= E_{add}(N,2,-w_2) \\
E_3 &= E_{add}(N,3,-w_3) \\
&\dots
\end{aligned}$$

Therefore we have :

$$\begin{aligned}
\det(AA^T) &= \det(E_{N-1} \dots E_3 E_2 E_1 A' A'^T E_1^T E_2^T E_3^T \dots E_{N-1}^T) \\
&= \det(A' A'^T) && \text{property 2 of determinant} \\
&= \det\left(\begin{bmatrix} B \\ C \end{bmatrix} [B^T \quad C^T]\right) \\
&= \det\left(\begin{bmatrix} BB^T & BC^T \\ CB^T & CC^T \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} BB^T & zeros_{(N-1)\times 1} \\ zeros_{1\times(N-1)} & CC^T \end{bmatrix}\right) && \text{since C is orthogonal to all rows of B} \\
&= \det(BB^T) CC^T && \text{since } CC^T \text{ is } 1\times 1\text{, and by Laplace expansion} \\
&= vol_{N-1}^2 h_N^2 && \text{assumption for case N–1} \\
&= vol_N^2 && \text{definition of volume}
\end{aligned}$$

Remark

Determinant of a matrix is zero if there exists a vector which is a linear combination of the others. The proof is simple, since the ordering of picking x_n does not matter, there exists :

$$\begin{aligned}
x_n &= \sum_{m=1}^{n-1} w_m x_n \\
&= \sum_{m=1}^{n-1} proj_{u_m}(x_n) \\
h_n &= x_n - \sum_{m=1}^{n-1} proj_{u_m}(x_n) = 0
\end{aligned}$$

$$\begin{aligned}
vol_N &= vol_{N-1} h_N \\
&= vol_{N-1}(x_N - \sum_{m=1}^{N-1} proj_{u_m}(x_N)) \\
&= vol_{N-n}(x_n - \sum_{m=1}^{n-1} proj_{u_m}(x_n)) \dots (x_N - \sum_{m=1}^{N-1} proj_{u_m}(x_N)) \\
&= vol_{N-n}(0) \dots (x_N - \sum_{m=1}^{N-1} proj_{u_m}(x_N)) \\
&= 0
\end{aligned}$$

Hence, when a rectangular matrix A degenerates (there exists at least one vector linearly dependent on the others), the volume of parallelepiped is zero, the determinant of $A^T A$ is zero.

Cramer's rule

Cramer's rule is a simple (yet not efficient) method for solving system of linear equation making use of determinant. Cramer's rule states that, for system of equation AX = B, where A is a N×N matrix, while B and X are N×1 matrices :

$$x_n = \frac{\det(A')}{\det(A)}$$

$$A' = (a'_{i,j}) \qquad \text{such that } a'_{i,j} = \begin{cases} a_{i,j} & j \neq n \\ b_i & j = n \end{cases} \qquad \forall n \in [1, N]$$

Lets prove it.

$$\begin{aligned}
\det(A') &= \det\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & b_1 & a_{1,n+1} & \cdots & a_{1,N} \\ a_{2,1} & \cdots & a_{2,n-1} & b_2 & a_{2,n+1} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & \cdots & a_{N,n-1} & b_N & a_{N,n+2} & \cdots & a_{N,N} \end{bmatrix} \\
&= \det\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & \sum_{m=1}^{N} a_{1,m}x_m & a_{1,n+1} & \cdots & a_{1,N} \\ a_{2,1} & \cdots & a_{2,n-1} & \sum_{m=1}^{N} a_{2,m}x_m & a_{2,n+1} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & \cdots & a_{N,n-1} & \sum_{m=1}^{N} a_{N,m}x_m & a_{N,n+2} & \cdots & a_{N,N} \end{bmatrix} \\
&= \det\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & \sum_{m=1}^{N-1} a_{1,m}x_m & a_{1,n+1} & \cdots & a_{1,N} \\ a_{2,1} & \cdots & a_{2,n-1} & \sum_{m=1}^{N-1} a_{2,m}x_m & a_{2,n+1} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & \cdots & a_{N,n-1} & \sum_{m=1}^{N-1} a_{N,m}x_m & a_{N,n+2} & \cdots & a_{N,N} \end{bmatrix} + x_N \det\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & a_{1,N} & a_{1,n+1} & \cdots & a_{1,N} \\ a_{2,1} & \cdots & a_{2,n-1} & a_{2,N} & a_{2,n+1} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & \cdots & a_{N,n-1} & a_{N,N} & a_{N,n+2} & \cdots & a_{N,N} \end{bmatrix} \\
&= \ldots \qquad \text{(repeatedly applying property 2 of determinant)} \\
&= \sum_{m=1}^{N} x_m \det\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & a_{1,m} & a_{1,n+1} & \cdots & a_{1,N} \\ a_{2,1} & \cdots & a_{2,n-1} & a_{2,m} & a_{2,n+1} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & \cdots & a_{N,n-1} & a_{N,m} & a_{N,n+2} & \cdots & a_{N,N} \end{bmatrix} \\
&= \underbrace{\sum_{\substack{m=1 \\ m\neq n}}^{N} x_m \det\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & a_{1,m} & a_{1,n+1} & \cdots & a_{1,N} \\ a_{2,1} & \cdots & a_{2,n-1} & a_{2,m} & a_{2,n+1} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & \cdots & a_{N,n-1} & a_{N,m} & a_{N,n+2} & \cdots & a_{N,N} \end{bmatrix}}_{matrices_having_two_identical_columns} + x_n \det\begin{bmatrix} a_{1,1} & \cdots & a_{1,n-1} & a_{1,n} & a_{1,n+1} & \cdots & a_{1,N} \\ a_{2,1} & \cdots & a_{2,n-1} & a_{2,n} & a_{2,n+1} & \cdots & a_{2,N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{N,1} & \cdots & a_{N,n-1} & a_{N,n} & a_{N,n+2} & \cdots & a_{N,N} \end{bmatrix} \\
&= x_n \det(A)
\end{aligned}$$

$$\Rightarrow \quad x_n = \frac{\det(A')}{\det(A)}$$

Reference

[1] Determinant : A means to calculate volume, Bo Peng, 2007.