Appendix A : Proof that matrix rank equals the number of independent columns and rows

Shilov proves this fact but the proof is a bit spread out over several sections. Here we provide a self-contained alternate proof.

We first quote a series of Theorems that are proved from scratch in Appendix B. Most of these will be very familiar to the reader, but the derivations might be less familiar,

Theorem 1: $\det(M^{\mathbf{T}}) = \det(M)$. *(B.1.10)*

Theorem 2: det(M) can be represented in these two ways, where ε is the permutation tensor (B.2.1) :

$$\det(M) = \Sigma_{a_1 a_2 \cdots a_n} \varepsilon_{a_1 a_2 \cdots a_n} M_{1a_1} M_{2a_2} \cdots\cdots M_{na_n} \qquad \textit{(B.2.10a)}$$

$$\det(M) = \Sigma_{a_1 a_2 \cdots a_n} \varepsilon_{a_1 a_2 \cdots a_n} M_{a_1 1} M_{a_2 2} \cdots\cdots M_{a_n n} \ . \qquad \textit{(B.2.10b)}$$

Theorem 3: For a square matrix M, adding a multiple of one row to another does not change det(M). The same is true for adding a multiple of one column to another. *(B.2.12)*

Theorem 4: Swapping two rows (columns) of square matrix M causes $\det(M) \rightarrow -\det(M)$. *(B.2.13)*

Corollary 4: If two rows (columns) of square matrix M are the same, then det(M) = 0. *(B.2.14)*

Theorem 5 (Cofactor Expansions):

$$\det(M) = \Sigma_n M_{sn} \operatorname{cof}(M_{sn}) = \Sigma_n (\mathbf{r}_s)_n \operatorname{cof}(M_{sn}) \quad \text{work across row s} \quad s = 1,2....n \qquad \textit{(B.3.16a)}$$

$$\det(M) = \Sigma_n M_{ns} \operatorname{cof}(M_{ns}) = \Sigma_n (\mathbf{c}_s)_n \operatorname{cof}(M_{ns}) \quad \text{work down column s} \quad s = 1,2....n \qquad \textit{(B.3.16b)}$$

Theorem 6: $$M^{-1} = \frac{C^{\mathbf{T}}}{\det(M)} = \frac{[\operatorname{cof}(M)]^{\mathbf{T}}}{\det(M)} = \frac{\operatorname{cof}(M^{\mathbf{T}})}{\det(M)} \qquad \textit{(B.4.8)}$$

Theorem 7 (Cramer's Rule):

$$\mathbf{y} = M\mathbf{x} \quad \Rightarrow \quad x_s = \frac{\det(M[\mathbf{c}_s \rightarrow \mathbf{y}])}{\det(M)} \qquad \textit{(B.4.14)}$$

To this list, we shall now add four new Theorems which will be proven right here. The definition of **linear independence** and **linear dependence** is given in the discussion surrounding (1.6) and won't be repeated here.

Theorem 8: Columns of square matrix M are linearly dependent $\Leftrightarrow$ det(M) = 0. (A.1)
Contrapositive: Columns of square matrix M are linearly independent $\Leftrightarrow$ $\det(M) \neq 0$

If we can prove Theorem 8 as stated, then the theorem is also true for rows. The reason is that swapping rows and columns corresponds to $M \leftrightarrow M^{\mathbf{T}}$ and Theorem 1 says $\det(M^{\mathbf{T}}) = \det(M)$.

Proof of $\Rightarrow$: According to (1.6) and nearby discussion, our premise of linear dependence says that *either* one or more columns of M are zero, *or* at least one column can be written as a linear combination of the other columns, so perhaps $\mathbf{c}_j = \Sigma_i k_i \mathbf{c}_i$. If a column is zero, det(M) = 0 from Theorem 5 going down that column. If $\mathbf{c}_j = \Sigma_{i\neq j} k_i \mathbf{c}_i$, one can add $-\Sigma_{i\neq j} k_i \mathbf{c}_i$ to $\mathbf{c}_j$ without changing det(M) according to Theorem 3. But this makes the new column j vanish, so again det(M) = 0.

Proof of $\Leftarrow$: Here we shall prove the contrapositive instead of the claim:

claim:	det(M) = 0 $\Rightarrow$ the columns of M are linearly dependent
contrapositive:	the columns of M are linearly independent $\Rightarrow$ det(M) $\neq$ 0

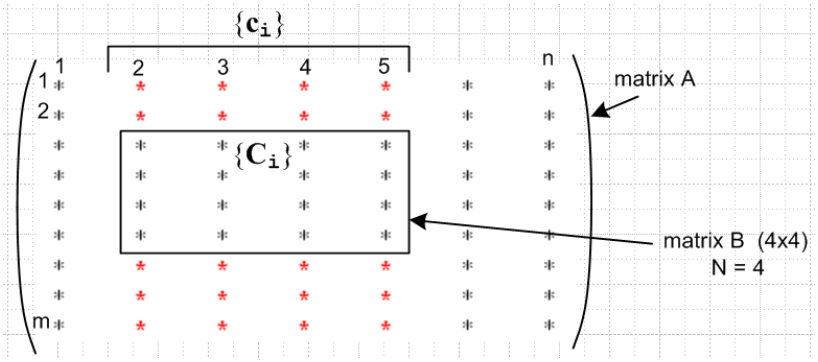
If the column vectors $\mathbf{c}_i$ of matrix M are linearly independent, then (1.6) says $\Sigma x_i \mathbf{c}_i = 0 \Rightarrow \mathbf{x} = 0$. Therefore we know that $\Sigma x_i M_{ji} = 0 \Rightarrow x_j = 0$ which is the same as $M\mathbf{x} = 0 \Rightarrow \mathbf{x} = 0$. Thinking of $M:E^n \rightarrow E^n$, we claim that mapping $M\mathbf{x} = \mathbf{y}$ is one-to-one. Certainly each $\mathbf{x}$ goes into a single $\mathbf{y}$, Could a $\mathbf{y}$ map back into two different values of $\mathbf{x}$, call them $\mathbf{x}$ and $\mathbf{x}'$ with $\mathbf{x} \neq \mathbf{x}'$? Then $M\mathbf{x} = \mathbf{y}$ and $M\mathbf{x}' = \mathbf{y}$. Subtract to get $M(\mathbf{x}-\mathbf{x}') = 0$. But since $M\mathbf{z} = 0 \Rightarrow \mathbf{z} = 0$ (as just shown), we find $\mathbf{x} = \mathbf{x}'$. Thus, the mapping $M\mathbf{x} = \mathbf{y}$ really is one-to-one, and that means it is invertible and the inverse is unique. From Theorem 6 the inverse is in fact given by $M^{-1} = \text{cof}(M^T)/\det(M)$. For M^{-1} to exist, one must have det(M) $\neq$ 0.

Another proof is very simple but perhaps less convincing. One can show (see our Tensor Analysis document (B.6.10) in Refs) that if a set of column vectors $\mathbf{c}_i$ of matrix M spans an n-piped in E^n, the "volume" of that n-piped is given by V = |det(M)|. If those vectors are linearly independent, then V $\neq$ 0. This is obvious in 2D (volume = area) and 3D but less obvious for general E^n. For example, in 3D if a third vector lies in the plane of the other two (is dependent), there is no volume.

Theorem 9 (mini-columns theorem). Let R be a subset of N rows of nxm matrix A, and let S be a subset of columns $\{\mathbf{c}_i\}$ in A. The matrix elements of A included in the intersection of these two sets form a set of "mini-columns" $\{\mathbf{C}_i\}$. This set of mini-columns forms a matrix B within A. Matrix B could be contiguous, or it could be non-contiguous in one or both directions. The claim and its contrapositive are:

(a) $\{\mathbf{C}_i\}$ linearly independent in B $\Rightarrow$ $\{\mathbf{c}_i\}$ linearly independent in A
(b) $\{\mathbf{c}_i\}$ linearly dependent in A $\Rightarrow$ $\{\mathbf{C}_i\}$ linearly dependent in B (A.2)

A picture is worth a thousand words. Here matrix B happens to be contiguous in both directions.



(A.3)

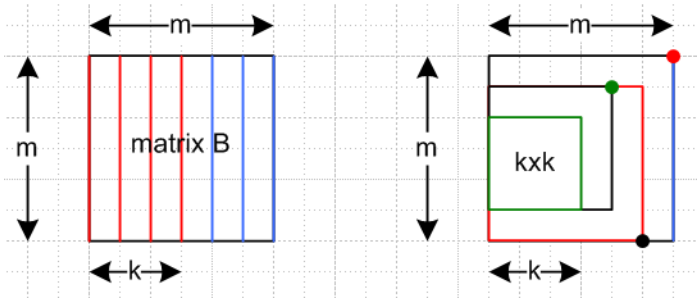
Proof of (b): If the set $\{\mathbf{c_i}\}$ is linearly dependent in A, then one can write $\Sigma_{i\in S}\, k_i\mathbf{c_i} = 0$ with $\mathbf{k} \neq 0$. This is a set of m equations which contains as a subset the set of N equations $\Sigma_{i\in S}\, k_i\mathbf{C_i} = 0$. This last equation then says that the $\{\mathbf{C_i}\}$ are linearly dependent in B. There is of course a similar "mini rows theorem".

Theorem 10. If all kxk minors in k columns of matrix A vanish, the k columns are linearly dependent. (A.4)

Contrapositive: If k columns are linearly independent, they must contain at least one non-zero kxk minor.

Once proven for columns, replacing $A \rightarrow A^{\mathbf{T}}$ gives the same theorem for rows.

Proof: Assume matrix A has m rows and n columns. In order for kxk minors to exist, one must have $k \leq \min(m,n)$. Gather up the k columns and make them be the leftmost k columns of a new matrix B. Since these columns have m elements, and since $k \leq m$, add m-k *arbitrary* new columns to the right of the k columns so that matrix B is then a square matrix. On the left below we show the k columns taken from matrix A (not shown) in red, and then the arbitrary added columns are shown in blue.



(A.5)

Consider now the process of computing det(B) by going down the rightmost column using the standard cofactor sum formula of Theorem 5. This det(B) is a linear combination of (m-1)x(m-1) minors all in the left m-1 columns. Consider one of these minors. If we evaluate *it* using the same cofactor formula (of one less dimension), it will be a linear combination of (m-2)x(m-2) minors in the left m-2 columns. We keep going until we are evaluating a set of kxk cofactors in the left k columns. But these all vanish by the theorem premise. Thus, reversing this logic we conclude that det(B) = 0. The picture on the right above shows one minor at each level of the descent just described (red dot goes with red minor, etc).

Now suppose it were possible that the k red columns were independent. Since the added blue columns are arbitrary, we could certainly find a set of added blue columns so that all m columns were independent. But then we would have $\det(B) \neq 0$ by Theorem 8. But we just showed that det(B) = 0, so it must not be possible to have the k red columns be independent, so they must be dependent.

We are finally ready for the Main Act. Recall first that :

Definition: The **rank** r of an m x n matrix A is the dimension of the largest non-vanishing minor within A. [Shilov 1.92] Thus, $r \leq \min(m,n)$. *(1.5)* (A.6)

Theorem 11. For a general nxm matrix A, (A.7)

rank(A) = r $\Leftrightarrow$ A has exactly r linearly independent columns

Once this theorem is proved for columns, it is also true for rows since rank(A^T) = rank(A).

Proof of $\Rightarrow$: If rank(A) = r, A must have at least one non-vanishing rxr minor. Think of this minor being a set of mini-columns as in Theorem 9. Since then det(minor) $\neq$ 0, this set of mini-columns is linearly independent according to Theorem 8. Thus, the set of corresponding full columns (those that pass down through the minor) are linearly independent from Theorem 9(a). So matrix A has at least these r independent columns.

Now let k = r+1. Since rank(A) = r we know that all kxk minors in A vanish. By Theorem 10, any set of k columns must be linearly dependent. Since k = r+1, *any* r+1 columns are linearly dependent, so the number of independent columns is just r, the number of columns passing down through the minor.

Proof of $\Leftarrow$: If there are r linearly independent columns, then at least one rxr minor within those columns must be non-zero (contrapositive of Theorem 10). Suppose there were an (r+k)x(r+k) non-vanishing minor. Then the r+k mini-columns of that minor would be independent, and thus by Theorem 9 there would be r+k independent full columns. But the theorem premise says there are only r independent columns, so all minors larger than rxr must vanish. Therefore rank(A) = r.

Comment: Sometimes the number of linearly independent rows of a matrix is called the **row rank**, and the number of independent columns is called the **column rank**. Theorem 11 and its row version show that

column rank = row rank = rank = dimension of largest non-vanishing minor (A.8)

Basis Minors and Basis Columns. If rank(A) = r, we know there must exist at least one non-vanishing rxr minor in A. Shilov refers to such a minor as a **basis minor** and the columns passing through this minor are called **basis columns**. We showed in Theorem 9 that the set of such basis columns is linearly independent (which is why they are called basis columns). Obviously any of these columns can be written as a linear combination of the basis columns, such as $\mathbf{c}_2 = \sum_{i=1}^{k} k_i \mathbf{c}_i = 1\ \mathbf{c}_2$. If k = r+1, all kxk minors vanish and by Theorem 10 all sets of k = r+1 columns are linearly dependent. Thus, every non-basis column is a linear combination of the basis columns. We have thus proved Shilov's Basis Minor Theorem which we quote from section 1.93 of his book (p 25)

THEOREM (***Basis minor theorem***). *Any column of the matrix A is a linear combination of its basis columns.* (A.9)

In his proof that column rank = rank, Shilov uses the above theorem as a starting point.