save old Appendix A PhL 6.10.15

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

Shilov proves these facts but the proof is a bit spread out over several sections. Here we provide an alternative proof that is perhaps more direct. Recall the definition (1.5) of the rank of a matrix :

Definition: The **rank** r of an m x n matrix A is the dimension of the largest non-vanishing subdeterminant (minor) within A. [Shilov 1.92] Thus, $r \leq \min(m,n)$. *(1.5)* (A.1)

The following well-known linear algebra fact will be used below (theorem and contrapositive):

the columns/rows of a square matrix M are linearly independent $\Leftrightarrow$ $\det(M) \neq 0$
the columns/rows of a square matrix M are linearly dependent $\Leftrightarrow$ $\det(M) = 0$. (A.2)

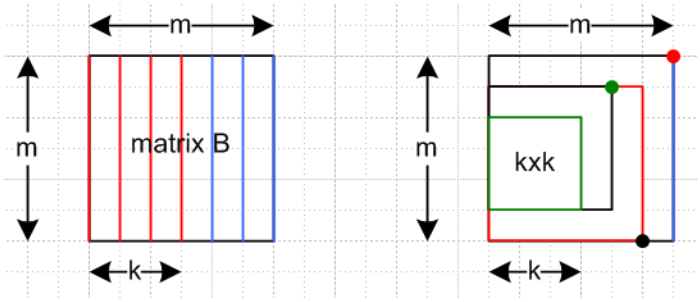
Roughly speaking, if the rows of M are linearly independent, we expect the system of linear equations $\mathbf{y} = M\mathbf{x}$ to be solvable by Cramer's Rule $\mathbf{x} = M^{-1}\mathbf{y} = [\det(M)]^{-1}\,\mathrm{cof}(M^{T})\,\mathbf{y}$ which requires $\det(M) \neq 0$. More obscurely, if a set of n column vectors $\mathbf{c}_i$ is linearly independent, we expect the volume of the n-piped spanned by these vectors in E^n to be non-zero, and that volume is given by $|\det(\mathbf{c}_1,\mathbf{c}_2...\mathbf{c}_n)| = |\det(M)|$. See our Tensor Analysis document (B.6.10).

Once Theorem 1 below is established, the rest is easy.

Theorem 1. If all kxk minors in k columns of matrix A vanish, the k columns are linearly dependent. (A.3)

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

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 in red, and then the arbitrary added columns are shown in blue.



(A.4)

Consider now the process of computing det(B) by going down the rightmost column using the standard cofactor sum formula. 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) ≠ 0. 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. QED.

Theorem 2. Rank(A) = r ⇒ A has r linearly independent columns. (A.5)

Proof: If rank(A) = r, A must have at least one non-vanishing rxr minor. The r columns of this minor are therefore linearly independent (each of these mini-columns has r elements). This means that the r *full* columns containing these mini-columns are also independent (see * below). So matrix A has at least r independent columns. Now let k = r+1. We know that all kxk minors in A vanish. By Theorem 1, 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 r.

* Since the mini-columns $\mathbf{C_i}$ are independent, one cannot write $\Sigma_i k_i \mathbf{C_i} = 0$ (with at least one $k_j \neq 0$). If the corresponding full columns $\mathbf{c_i}$ were dependent, one could write $\Sigma_i k_i \mathbf{c_i} = 0$. But if $\Sigma_i k_i \mathbf{c_i} = 0$, then one must have $\Sigma_i k_i \mathbf{C_i} = 0$ since the latter is just a subset of the former equation set. But $\Sigma_i \lambda_i \mathbf{C_i} = 0$ says that the mini-columns are dependent, which they are not. Thus, one cannot write $\Sigma_i k_i \mathbf{c_i} = 0$ and thus the full columns are independent. (A.6)

Theorem 3. A has r linearly independent columns ⇒ Rank(A) = r (A.7)

Proof: If there are r linearly independent columns, then at least one rxr minor within those columns must be non-zero (contrapositive of Theorem 1). 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 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.

Making trivial replacements in the above three theorems and proofs (column → row, leftmost → topmost and so on), one finds that all three theorems are valid for rows as well as columns. Thus:

Theorem 1' . If all kxk minors in k rows of matrix A vanish, the k rows are linearly dependent. (A.8)

Theorem 2'. Rank(A) = r ⇒ A has r linearly independent rows. (A.9)

Theorem 3'. A has r linearly independent rows ⇒ Rank(A) = r (A.10)

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**. Theorems 2 and 3 show that column rank = rank, while Theorems 1' and 2' show that row rank = rank. Therefore:

column rank = row rank = rank = dimension of largest non-vanishing minor (A.11)

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 2 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 $\mathbf{c_2} = \Sigma_{i=1}^{k} k_i \mathbf{c_i} = 1\ \mathbf{c_2}$. If k = r+1, all kxk minors vanish and by Theorem 1 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.12)

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