Section 3.3. Matrix Rank and the Inverse of a Full Rank Matrix

Note. The lengthy section (21 pages in the text) gives a thorough study of the rank of a matrix (and matrix products) and considers inverses of matrices briefly at the end.

Note. Recall that the row space of a matrix A is the span of the row vectors of A and the row rank of A is the dimension of this row space. Similarly, the column space of A is the span of the column vectors of A and the column rank is the dimension of this column space. You will recall that the dimension of the column space and the dimension of the row space of a given matrix are the same (see Theorem 2.4 of Fraleigh and Beauregard's *Linear Algebra*, 3rd Edition, Addison-Wesley Publishing Company, 1995, in 2.2. The Rank of a Matrix). We now give a proof of this based in part on Gentle's argument and on Peter Lancaster's *Theory of Matrices*, Academic Press (1969), page 42. First, we need a lemma.

Lemma 3.3.1. Let $\{a^i\}_{i=1}^k = \{[a_1^i, a_2^i, \ldots, a_n^i]\}_{i=1}^k$ be a set of vectors in $\mathbb{R}^n$ and let $\pi \in S_n$. Then the set of vectors $\{a^i\}_{i=1}^k$ is linearly independent if and only if the set of vectors $\{[a_{\pi(1)}^i, a_{\pi(2)}^i, \ldots, a_{\pi(n)}^i]\}_{i=1}^k$ is linearly independent. That is, permuting all the entries in a set of vectors by the same permutation preserves the linear dependence/independence of the set.

Theorem 3.3.2. Let A be an $n \times m$ matrix. Then the row rank of A equals the column rank of A. This common quantity is called the *rank* of A.

Note. Recall that $\mathcal{V}(A)$ denotes the column space of matrix A (see page 41 of the text) and so $\mathcal{V}(A^T)$ is the row space of A. So from the definition of rank and Theorem 3.3.2, we can conclude that $\text{rank}(A) = \dim(\mathcal{V}(A))$, $\text{rank}(A) = \text{rank}(A^T)$, and $\dim(\mathcal{V}(A)) = \dim(\mathcal{V}(A^T))$.

Definition. A matrix is of *full rank* if its rank is the same as its smaller dimension. A matrix that is not full rank is *rank deficient* and the *rank deficiency* is the difference between its smaller dimension and the rank. A full rank matrix which is square is *nonsingular*. A square matrix which is not nonsingular is *singular*.

Note. The previous definition of singular/nonsingular may be new to you. Later we will relate nonsingular and invertibility, as you expect (see Note 3.3.A below).

Theorem 3.3.3. If P and Q are products of elementary matrices then $\text{rank}(PAQ) = \text{rank}(A)$.

Note. Since the rank of I_n is n, then Theorem 3.3.3 implies with $A = I_n$ that each elementary matrix is full rank.

Note. Recall that Exercise 2.1.G states: Let A be a set of n-vectors.

(i) If $B \subset A$ then $\dim(\text{span}(B)) \leq \dim(\text{span}(A))$.

(ii) If $A = A_1 \cup A_2$ then $\dim(\text{span}(A)) \leq \dim(\text{span}(A_1)) + \dim(\text{span}(A_2))$.

(iii) If $A = A_1 \oplus A_2$ (so that $A_1 \perp A_2$) for vector spaces A, A_1, and A_2, then $\dim(A) = \dim(A_1) + \dim(A_2)$.

These observations are useful in proving the next theorem.

Theorem 3.3.4. Let A be a matrix partitioned as $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$. Then

(i) $\text{rank}(A_{ij}) \leq \text{rank}(A)$ for $i, j \in \{1, 2\}$.

(ii) $\text{rank}(A) \leq \text{rank}([A_{11}|A_{12}]) + \text{rank}([A_{21}|A_{22}])$.

(iii) $\text{rank}(A) \leq \text{rank}\left(\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}\right) + \text{rank}\left(\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}\right)$.

(iv) If $\mathcal{V}([A_{11}|A_{12}]^T) \perp \mathcal{V}([A_{21}|A_{22}]^T)$ then $\text{rank}(A) = \text{rank}([A_{11}|A_{12}]) + \text{rank}([A_{21}|A_{22}])$ and if $\mathcal{V}\left(\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}\right) \perp \mathcal{V}\left(\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}\right)$ then

$$\text{rank}(A) = \text{rank}\left(\begin{bmatrix} A_{11} \\ A_{21} \end{bmatrix}\right) + \text{rank}\left(\begin{bmatrix} A_{12} \\ A_{22} \end{bmatrix}\right).$$

Note. In Exercise 3.3.C, it is to be shown using Theorem 3.3.4(iv) that for a block diagonal matrix $A = \text{diag}(A_{11}, A_{22}, \ldots, A_{kk})$ (see Section 3.1), we have $\text{rank}(A) = \text{rank}(A_{11}) + \text{rank}(A_{22}) + \cdots + \text{rank}(A_{kk})$.

Theorem 3.3.5. Let A be an $n \times k$ matrix and B be a $k \times m$ matrix. Then $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$.

Note. By Theorem 3.3.5, for $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, the outer product xy^T satisfies

$$\text{rank}(xy^T) \leq \min\{\text{rank}(x), \text{rank}(y^T)\} = 1.$$

Theorem 3.3.6. Let A and B be $n \times m$ matrices. Then

$$|\text{rank}(A) - \text{rank}(B)| \leq \text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B).$$

Note. If $n \times m$ matrix A is of rank r, then it has r linearly independent rows. So there is a permutation matrix E_{π_1} such that $E_{\pi_1}A$ is a matrix whose first r rows are linearly independent (and certainly the choice of E_{π_1} is not unique). Since $E_{\pi_1}A$ is rank r (by Theorem 3.3.3), it has r linearly independent columns (by Theorem 3.3.2) and there is permutation matrix E_{π_2} such that $E_{\pi_1}AE_{\pi_2}$ is a matrix whose first r columns are linearly independent. The matrix $B = E_{\pi_1}AE_{\pi_2}$ can then be partitioned in a way that isolates linearly independent "sub-rows" and "sub-columns."

Definition. Let B be an $n \times m$ matrix of rank r whose first r rows are linearly independent and whose first r columns are linearly independent. Then the partitioning of B as $B = \begin{bmatrix} W & X \\ Y & Z \end{bmatrix}$, where W is a $r \times r$ full rank submatrix, X is $r \times (m-r)$, Y is $(n-r) \times r$, and Z is $(n-r) \times (m-r)$, is a *full rank partitioning of* B.

Note. If $n \times m$ matrix A is of rank r then for any $q \leq r$ (with E_{π_1} and E_{π_2} as described in the previous note) we have $E_{\pi_1} A E_{\pi_2} = \begin{bmatrix} S & T \\ U & V \end{bmatrix}$ where S is a $q \times q$ full rank matrix, T is $q \times (m-q)$, U is $(n-q) \times q$, and V is $(n-q) \times (m-q)$.

Note. A *system of n linear equations* in m unknowns $x_1, x_2, \ldots, x_m$ is a system of the form

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1m}x_m & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2m}x_m & = & b_2 \\ \vdots & & & & & & & & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & a_{nm}x_m & = & b_n. \end{array}$$

With $A = [a_{ij}]$, $b \in \mathbb{R}^n$ with entries b_i, and $x \in \mathbb{R}^m$ with entries x_j (so that b and x are column vectors by our convention, see Section 3.2), the system can be written as $Ax = b$. A is the *coefficient matrix*. For a given A and b, a vector $x \in \mathbb{R}^m$ for which $Ax = b$ is a *solution* to the system of equations. If a solution exists, the system of equations is *consistent*. If a solution does not exists, the system of equations is *inconsistent*.

Note. In sophomore linear algebra, you used elementary row operations to explore solutions to systems of equations. Here, we use the rank of the coefficient matrix and the existence of an inverse of the coefficient matrix to explore solutions to systems of equations.

Note. Recall from Section 3.2 (see page 5 of the class notes) that with $a_1, a_2, \ldots, a_m$ as the columns of A and $x \in \mathbb{R}^m$ a vector of scalars $x_1, x_2, \ldots, x_m$ we have $Ax = \sum_{i=1}^m x_i a_i$. So for any given $x \in \mathbb{R}^m$, the vector Ax is a linear combination of the columns of A and so having a solution x to the system $Ax = b$ is equivalent to saying that b is in the column space, $\mathcal{V}(A)$, of A.

Note. A system $Ax = b$ is consistent if and only if $\text{rank}([A \mid b]) = \text{rank}(A)$. This holds because of $b \in \mathcal{V}(A)$ implies $\text{rank}(A) = \text{rank}([A \mid b])$ and $\text{rank}([A \mid b]) = \text{rank}(A)$ implies $b \in \mathcal{V}(A)$. Also, if for $n \times m$ matrix A, where $n \leq m$, we have $\text{rank}(A) = n$ (so that A is of full row rank), then $\text{rank}([A \mid b]) = n$ (since $[A \mid b]$ is $n \times (m+1)$) and so, in this case, the system $Ax = b$ is consistent for any $b \in \mathbb{R}^n$. The matrix $[A \mid b]$ is called the *augmented matrix* for the system $Ax = b$.

Note. Let A be a $n \times n$ nonsingular matrix (that is, A is square and full rank). Then for the ith unit vector $e_i \in \mathbb{R}^n$, $e_i \in \mathcal{V}(A)$ and so $Ax_i = e_i$ has a solution x_i for $i = 1, 2, \ldots, n$. Creating $n \times n$ matrix X with columns x_i and I_n (with columns e_i), we can write these n systems of equations as the matrix equation $AX = I_n$. Since $AX = I_n$, X is a "right inverse" of A. Since $\text{rank}(I_n) \leq \min\{\text{rank}(A), \text{rank}(X)\}$ by Theorem 3.3.5, and $n = \text{rank}(I_n) = \text{rank}(A)$, then it must be that $\text{rank}(X) = n$. So a similar argument shows that X has a right inverse, say $XY = I_n$. But then $A = AI_n = A(XY) = (AX)Y = I_nY = Y$ and so $XA = XY = I_n$, and X is also a "left inverse" of A.

Definition. For $n \times n$ full rank matrix A, the matrix B such that $BA = AB = I_n$ is the *inverse* of matrix A, denoted $B = A^{-1}$. (Of course A^{-1} is unique for a given matrix A.)

Theorem 3.3.7. Let A be an $n \times n$ full rank matrix. Then $(A^{-1})^T = (A^T)^{-1}$.

Note. Gentle uses some unusual notation. He denotes $(A^T)^{-1} = (A^T)^{-1} = A^{-T}$. He "sometimes" denotes AB^{-1} as A/B (UGH!) and $B^{-1}A$ as $B\backslash A$. I will avoid this notation.

Theorem 3.3.8. $n \times m$ matrix A, where $n \le m$, has a right inverse if and only if A is of full row rank n. $n \times m$ matrix A, where $m \le n$, has a left inverse if and only if A has full column rank m.

Note 3.3.A. Theorem 3.3.8 shows that a square matrix is nonsingular if and only if it is invertible.

Note. With A an $n \times m$ matrix, if $n \times n$ matrix AA^T is of full rank, then $(AA^T)^{-1}$ exists and the right inverse of A is $A^T(AA^T)^{-1}$ since $AA^T(AA^T)^{-1} = I_n$. Similarly, if A^TA is of full rank, then the left inverse of A is $(A^TA)^{-1}A^T$ since $(A^TA)^{-1}A^TA = I_m$.

Definition. Matrices A and B of the same size that have the same rank are *equivalent*, denoted $A \sim B$. For $m \times n$ matrix A with rank r where $0 < r \leq \min\{n, m\}$, then the *equivalent canonical form of* A is the $n \times m$ matrix $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$.

Theorem 3.3.9. If A is an $n \times m$ matrix of rank $r > 0$ then there are matrices P and Q, both products of elementary matrices, such that PAQ is the equivalent canonical form of A, $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$.

Note. If A is symmetric, then the same operations in the proof of Theorem 3.3.9 are performed on the rows which are performed on the columns so that we have $PAP^T = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ for P a product of elementary matrices.

Note. You dealt with row equivalence in Linear Algebra. This "equivalence" is a combination of row equivalence and column equivalence so that $A \sim B$ if and only if $B = PAQ$ where P and Q are products of some elementary matrices.

Definition. A matrix R is in *row echelon form* ("REF") if

(1) $r_{ij} = 0$ for $i > j$, and

(2) if k is such that $r_{ik} \neq 0$ and $r_{i\ell} = 0$ for $\ell < k$ then $r_{i+1,j} = 0$ for $j \leq k$.

Note. To see that this definition of row echelon form is consistent with your sophomore Linear Algebra experience, notice that Condition (1) implies that there are only 0's below the main diagonal. Condition (2) implies that r_{ik} is the first nonzero entry in row i (called the "pivot") and that the first nonzero entry in row $i+1$ lies to the right of pivot r_{ik} (that is, in a column of index greater than k).

Theorem 3.3.10. For any matrix A there is a matrix P a product of elementary matrices such that PA is in row echelon form.

Note. The proof of Theorem 3.3.10 follows by applying the technique of *Gauss-Jordan elimination.* An algorithmic explanation of Gauss-Jordan elimination can be found in my online notes for Linear Algebra (MATH 2010) for 1.4. Solving Systems of Linear Equations.

Note. Gentle does not define reduced row echelon form of a matrix in which the matrix is in row echelon form where each pivot is 1 and all entries above the pivots are 0. We can use Gentle's approach to define this as follows.

Definition. A matrix R is *reduced row echelon form* ("RREF") if it is in row echelon form and

(3) if k is such that $r_{ik} \neq 0$ and $r_{i\ell} = 0$ for $\ell < k$ then $r_{ik} = 1$ and $r_{jk} = 0$ for $j < i$.

Definition. A (square) upper triangular matrix H is in *Hermite form* if

(1) $h_{ii} = 0$ or 1,

(2) if $h_{ii} = 0$ then $h_{ij} = 0$ for all j, and

(3) if $h_{ii} = 1$, then $h_{ki} = 0$ for all $k \neq i$.

Note. If H is in Hermite form the Condition (1) implies that the main diagonal entries are 0's and 1's. Condition (2) implies that the rows containing a 0 diagonal entry are all 0's. Condition (3) implies that the columns containing 1 diagonal entry has all other entries 0. Notice that a diagonal entry $h_{ii} = 0$ may have nonzero entries above it in column i. For example, $H = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ is in Hermite form. In Exercise 3.3.B it is shown that if H is in Hermite form then $H^2 = H$.

Theorem 3.3.11. If A is a square full rank matrix (that is, nonsingular) and if B and C are conformable matrices for the multiplications AB and CA then $\text{rank}(AB) = \text{rank}(B)$ and $\text{rank}(CA) = \text{rank}(C)$.

Note. In fact, Theorem 3.3.11 can be extended to nonsquare matrices as follows.

Theorem 3.3.12. If A is a full column rank matrix and B is conformable for the multiplication AB, then $\text{rank}(AB) = \text{rank}(B)$. If A is a full row rank matrix and C is conformable for the multiplication CA, then $\text{rank}(CA) = \text{rank}(C)$.

Note. Recall the $n \times n$ symmetric matrix A is positive definite if for each $x \in \mathbb{R}^n$ with $x \neq 0$ we have that the quadratic form satisfies $x^T A x > 0$. The next result shows that positive definiteness is preserved under a particular type of multiplication by a full rank matrix.

Theorem 3.3.13. Let C be $n \times n$ and positive definite and let A be $n \times m$.

(1) If C is positive definite and A is of full column rank $m \leq n$ then $A^T C A$ is positive definite.

(2) If $A^T C A$ is positive definite then A is of full column rank $m \leq n$.

Theorem 3.3.14. Properties of $A^T A$.

Let A be an $n \times m$ matrix.

(1) $A^T A = 0$ if and only if $A = 0$.

(2) $A^T A$ is nonnegative definite.

(3) $A^T A$ is positive definite if and only if A is of full column rank.

(4) $(A^T A)B = (A^T A)C$ if and only if $AB = AC$, and $B(A^T A) = C(A^T A)$ if and only if $BA^T = CA^T$.

(5) $A^T A$ is of full rank if and only if A is of full column rank.

(6) $\text{rank}(A^T A) = \text{rank}(A)$.

The product $A^T A$ is called a *Gramian matrix*.

Note. From Theorem 3.3.5, we have an upper bound on the rank of a product of two matrices: $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$. We now put a lower bound on the rank of a product.

Theorem 3.3.15. Sylvester's Law of Nullity

If A is a $n \times n$ matrix and B is $n \times \ell$ then $\text{rank}(AB) \geq \text{rank}(A) + \text{rank}(B) - n$.

Note. The following result relates the value of $\det(A)$ and the invertibility of A.

Theorem 3.3.16. $n \times n$ matrix A is invertible if and only if $\det(A) \neq 0$.

Note. Gentle motivates inverting sums and differences of matrices by referring to regression analysis and adding or omitting data (see page 93). Thus we consider the following, the proof of which is to be given in Exercise 3.12.

Theorem 3.3.17. Let A and B be $n \times n$ full rank matrices. Then:

(1) $A(I + A)^{-1} = (I + A^{-1})^{-1}$,

(2) $(A + BB^T)^{-1}B = A^{-1}B(I + B^T A^{-1}B)^{-1}$,

(3) $(A^{-1} + B^{-1})^{-1} = A(A + B)^{-1}B$,

(4) $A - A(A + B)^{-1}A = B - B(A + B)^{-1}B$,

(5) $A^{-1} + B^{-1} = A^{-1}(A + B)B^{-1}$,

(6) $(I + AB)^{-1} = I - A(I + BA)^{-1}B$,

(7) $(I + AB)^{-1}A = A(I + BA)^{-1}$,

where we require the invertibility of relevant sums and differences.

Theorem 3.3.18. If A and B are $n \times n$ full rank matrices then the Kronecker product satisfies $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

Note. Now, we summarize our vocabulary and results on invertible matrices. First, A^{-1} exists then A must be square. A square $n \times n$ matrix A is *full rank* if $\text{rank}(A) = n$ and this value is the dimension of both the row space and column space of A (see Theorem 3.3.2). A full rank square matrix is *nonsingular* (by definition). We argued that a full rank square matrix has an inverse by considering associated system of equations. So a nonsingular matrix is invertible. In Theorem 3.3.16 we showed that square matrix A is invertible (that is, nonsingular) if and only if $\det(A) \neq 0$.

Revised: 6/23/2020