4.5 MORE ABOUT RANK

Since equivalent matrices have the same rank, it follows that if $\mathbf{P}$ and $\mathbf{Q}$ are nonsingular matrices such that the product $\mathbf{PAQ}$ is defined, then

$$rank\,(\mathbf{A}) = rank\,(\mathbf{PAQ}) = rank\,(\mathbf{PA}) = rank\,(\mathbf{AQ}).$$

In other words, rank is invariant under multiplication by a nonsingular matrix. However, multiplication by rectangular or singular matrices can alter the rank, and the following formula shows exactly how much alteration occurs.

Rank of a Product

If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$, then

$$rank\,(\mathbf{AB}) = rank\,(\mathbf{B}) - \dim N\,(\mathbf{A}) \cap R\,(\mathbf{B}). \tag{4.5.1}$$

Proof. Start with a basis $\mathcal{S} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_s\}$ for $N\,(\mathbf{A}) \cap R\,(\mathbf{B})$, and notice $N\,(\mathbf{A}) \cap R\,(\mathbf{B}) \subseteq R\,(\mathbf{B})$. If $\dim R\,(\mathbf{B}) = s + t$, then, as discussed in Example 4.4.5, there exists an extension set $\mathcal{S}_{ext} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_t\}$ such that $\mathcal{B} = \{\mathbf{x}_1, \ldots, \mathbf{x}_s, \mathbf{z}_1, \ldots, \mathbf{z}_t\}$ is a basis for $R\,(\mathbf{B})$. The goal is to prove that $\dim R\,(\mathbf{AB}) = t$, and this is done by showing $\mathcal{T} = \{\mathbf{Az}_1, \mathbf{Az}_2, \ldots, \mathbf{Az}_t\}$ is a basis for $R\,(\mathbf{AB})$. $\mathcal{T}$ spans $R\,(\mathbf{AB})$ because if $\mathbf{b} \in R\,(\mathbf{AB})$, then $\mathbf{b} = \mathbf{ABy}$ for some $\mathbf{y}$, but $\mathbf{By} \in R\,(\mathbf{B})$ implies $\mathbf{By} = \sum_{i=1}^{s} \xi_i \mathbf{x}_i + \sum_{i=1}^{t} \eta_i \mathbf{z}_i$, so

$$\mathbf{b} = \mathbf{A}\left(\sum_{i=1}^{s} \xi_i \mathbf{x}_i + \sum_{i=1}^{t} \eta_i \mathbf{z}_i\right) = \sum_{i=1}^{s} \xi_i \mathbf{Ax}_i + \sum_{i=1}^{t} \eta_i \mathbf{Az}_i = \sum_{i=1}^{t} \eta_i \mathbf{Az}_i.$$

$\mathcal{T}$ is linearly independent because if $\mathbf{0} = \sum_{i=1}^{t} \alpha_i \mathbf{Az}_i = \mathbf{A} \sum_{i=1}^{t} \alpha_i \mathbf{z}_i$, then $\sum_{i=1}^{t} \alpha_i \mathbf{z}_i \in N\,(\mathbf{A}) \cap R\,(\mathbf{B})$, so there are scalars β_j such that

$$\sum_{i=1}^{t} \alpha_i \mathbf{z}_i = \sum_{j=1}^{s} \beta_j \mathbf{x}_j \quad \text{or, equivalently,} \quad \sum_{i=1}^{t} \alpha_i \mathbf{z}_i - \sum_{j=1}^{s} \beta_j \mathbf{x}_j = \mathbf{0},$$

and hence the only solution for the α_i 's and β_i 's is the trivial solution because $\mathcal{B}$ is an independent set. Thus $\mathcal{T}$ is a basis for $R\,(\mathbf{AB})$, so $t = \dim R\,(\mathbf{AB}) = rank\,(\mathbf{AB})$, and hence

$$rank\,(\mathbf{B}) = \dim R\,(\mathbf{B}) = s + t = \dim N\,(\mathbf{A}) \cap R\,(\mathbf{B}) + rank\,(\mathbf{AB}). \quad \blacksquare$$

It's sometimes necessary to determine an explicit basis for $N\,(\mathbf{A}) \cap R\,(\mathbf{B})$. In particular, such a basis is needed to construct the Jordan chains that are associated with the Jordan form that is discussed on pp. 582 and 594. The following example outlines a procedure for finding such a basis.

Basis for an Intersection

If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$, then a basis for $N(\mathbf{A}) \cap R(\mathbf{B})$ can be constructed by the following procedure.

- ▷ Find a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_r\}$ for $R(\mathbf{B})$.
- ▷ Set $\mathbf{X}_{n\times r} = \big(\mathbf{x}_1 \,|\, \mathbf{x}_2 \,|\, \cdots \,|\, \mathbf{x}_r\big)$.
- ▷ Find a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\}$ for $N(\mathbf{AX})$.
- ▷ $\mathcal{B} = \{\mathbf{Xv}_1, \mathbf{Xv}_2, \ldots, \mathbf{Xv}_s\}$ is a basis for $N(\mathbf{A}) \cap R(\mathbf{B})$.

Proof. The strategy is to argue that $\mathcal{B}$ is a maximal linear independent subset of $N(\mathbf{A}) \cap R(\mathbf{B})$. Since each $\mathbf{Xv}_j$ belongs to $R(\mathbf{X}) = R(\mathbf{B})$, and since $\mathbf{AXv}_j = \mathbf{0}$, it's clear that $\mathcal{B} \subset N(\mathbf{A}) \cap R(\mathbf{B})$. Let $\mathbf{V}_{r\times s} = \big(\mathbf{v}_1 \,|\, \mathbf{v}_2 \,|\, \cdots \,|\, \mathbf{v}_s\big)$, and notice that $\mathbf{V}$ and $\mathbf{X}$ each have full column rank. Consequently, $N(\mathbf{X}) = \mathbf{0}$ so, by (4.5.1),

$$rank\,(\mathbf{XV})_{n\times s} = rank\,(\mathbf{V}) - \dim N(\mathbf{X}) \cap R(\mathbf{V}) = rank\,(\mathbf{V}) = s,$$

which insures that $\mathcal{B}$ is linearly independent. $\mathcal{B}$ is a *maximal* independent subset of $N(\mathbf{A}) \cap R(\mathbf{B})$ because (4.5.1) also guarantees that

$$\begin{aligned} s &= \dim N(\mathbf{AX}) = \dim N(\mathbf{X}) + \dim N(\mathbf{A}) \cap R(\mathbf{X}) \quad \text{(see Exercise 4.5.10)} \\ &= \dim N(\mathbf{A}) \cap R(\mathbf{B}). \quad \blacksquare \end{aligned}$$

The utility of (4.5.1) is mitigated by the fact that although $rank\,(\mathbf{A})$ and $rank\,(\mathbf{B})$ are frequently known or can be estimated, the term $\dim N(\mathbf{A}) \cap R(\mathbf{B})$ can be costly to obtain. In such cases (4.5.1) still provides us with useful upper and lower bounds for $rank\,(\mathbf{AB})$ that depend only on $rank\,(\mathbf{A})$ and $rank\,(\mathbf{B})$.

Bounds on the Rank of a Product

If $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$, then

- $rank\,(\mathbf{AB}) \le \min\{rank\,(\mathbf{A}),\ rank\,(\mathbf{B})\}$, (4.5.2)
- $rank\,(\mathbf{A}) + rank\,(\mathbf{B}) - n \le rank\,(\mathbf{AB})$. (4.5.3)

Proof. In words, (4.5.2) says that the rank of a product cannot exceed the rank of either factor. To prove $rank\,(\mathbf{AB}) \le rank\,(\mathbf{B})$, use (4.5.1) and write

$$rank\,(\mathbf{AB}) = rank\,(\mathbf{B}) - \dim N\,(\mathbf{A}) \cap R\,(\mathbf{B}) \le rank\,(\mathbf{B}).$$

This says that the rank of a product cannot exceed the rank of the right-hand factor. To show that $rank\,(\mathbf{AB}) \le rank\,(\mathbf{A})$, remember that transposition does not alter rank, and use the reverse order law for transposes together with the previous statement to write

$$rank\,(\mathbf{AB}) = rank\,(\mathbf{AB})^T = rank\left(\mathbf{B}^T\mathbf{A}^T\right) \le rank\left(\mathbf{A}^T\right) = rank\,(\mathbf{A}).$$

To prove (4.5.3), notice that $N\,(\mathbf{A}) \cap R\,(\mathbf{B}) \subseteq N\,(\mathbf{A})$, and recall from (4.4.5) that if $\mathcal{M}$ and $\mathcal{N}$ are spaces such that $\mathcal{M} \subseteq \mathcal{N}$, then $\dim \mathcal{M} \le \dim \mathcal{N}$. Therefore,

$$\dim N\,(\mathbf{A}) \cap R\,(\mathbf{B}) \le \dim N\,(\mathbf{A}) = n - rank\,(\mathbf{A}),$$

and the lower bound on $rank\,(\mathbf{AB})$ is obtained from (4.5.1) by writing

$$rank\,(\mathbf{AB}) = rank\,(\mathbf{B}) - \dim N\,(\mathbf{A}) \cap R\,(\mathbf{B}) \ge rank\,(\mathbf{B}) + rank\,(\mathbf{A}) - n. \quad \blacksquare$$

The products $\mathbf{A}^T\mathbf{A}$ and $\mathbf{AA}^T$ and their complex counterparts $\mathbf{A}^*\mathbf{A}$ and $\mathbf{AA}^*$ deserve special attention because they naturally appear in a wide variety of applications.

Products $\mathbf{A}^T\mathbf{A}$ and $\mathbf{AA}^T$

For $\mathbf{A} \in \Re^{m\times n}$, the following statements are true.

- $rank\left(\mathbf{A}^T\mathbf{A}\right) = rank\,(\mathbf{A}) = rank\left(\mathbf{AA}^T\right).$ (4.5.4)
- $R\left(\mathbf{A}^T\mathbf{A}\right) = R\left(\mathbf{A}^T\right)$ and $R\left(\mathbf{AA}^T\right) = R\,(\mathbf{A}).$ (4.5.5)
- $N\left(\mathbf{A}^T\mathbf{A}\right) = N\,(\mathbf{A})$ and $N\left(\mathbf{AA}^T\right) = N\left(\mathbf{A}^T\right).$ (4.5.6)

For $\mathbf{A} \in \mathcal{C}^{m\times n}$, the transpose operation $(\star)^T$ must be replaced by the conjugate transpose operation $(\star)^*$.

Proof. First observe that $N\left(\mathbf{A}^T\right) \cap R(\mathbf{A}) = \{\mathbf{0}\}$ because

$$\begin{aligned}\mathbf{x} \in N\left(\mathbf{A}^T\right) \cap R(\mathbf{A}) &\implies \mathbf{A}^T\mathbf{x} = \mathbf{0} \text{ and } \mathbf{x} = \mathbf{A}\mathbf{y} \text{ for some } \mathbf{y} \\ &\implies \mathbf{x}^T\mathbf{x} = \mathbf{y}^T\mathbf{A}^T\mathbf{x} = 0 \implies \sum x_i^2 = 0 \\ &\implies \mathbf{x} = \mathbf{0}.\end{aligned}$$

Formula (4.5.1) for the rank of a product now guarantees that

$$rank\left(\mathbf{A}^T\mathbf{A}\right) = rank(\mathbf{A}) - \dim N\left(\mathbf{A}^T\right) \cap R(\mathbf{A}) = rank(\mathbf{A}),$$

which is half of (4.5.4)—the other half is obtained by reversing the roles of $\mathbf{A}$ and $\mathbf{A}^T$. To prove (4.5.5) and (4.5.6), use the facts $R(\mathbf{AB}) \subseteq R(\mathbf{A})$ and $N(\mathbf{B}) \subseteq N(\mathbf{AB})$ (see Exercise 4.2.12) to write $R\left(\mathbf{A}^T\mathbf{A}\right) \subseteq R\left(\mathbf{A}^T\right)$ and $N(\mathbf{A}) \subseteq N\left(\mathbf{A}^T\mathbf{A}\right)$. The first half of (4.5.5) and (4.5.6) now follows because

$$\dim R\left(\mathbf{A}^T\mathbf{A}\right) = rank\left(\mathbf{A}^T\mathbf{A}\right) = rank(\mathbf{A}) = rank\left(\mathbf{A}^T\right) = \dim R\left(\mathbf{A}^T\right),$$

$$\dim N(\mathbf{A}) = n - rank(\mathbf{A}) = n - rank\left(\mathbf{A}^T\mathbf{A}\right) = \dim N\left(\mathbf{A}^T\mathbf{A}\right).$$

Reverse the roles of $\mathbf{A}$ and $\mathbf{A}^T$ to get the second half of (4.5.5) and (4.5.6). ■

To see why (4.5.4)—(4.5.6) might be important, consider an $m \times n$ system of equations $\mathbf{Ax} = \mathbf{b}$ that may or may not be consistent. Multiplying on the left-hand side by $\mathbf{A}^T$ produces the $n \times n$ system

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

called ***the associated system of normal equations***, which has some extremely interesting properties. First, notice that the normal equations are always consistent, regardless of whether or not the original system is consistent because (4.5.5) guarantees that $\mathbf{A}^T\mathbf{b} \in R\left(\mathbf{A}^T\right) = R\left(\mathbf{A}^T\mathbf{A}\right)$ (i.e., the right-hand side is in the range of the coefficient matrix), so (4.2.3) insures consistency. However, if $\mathbf{Ax} = \mathbf{b}$ happens to be consistent, then $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ have the same solution set because if $\mathbf{p}$ is a particular solution of the original system, then $\mathbf{Ap} = \mathbf{b}$ implies $\mathbf{A}^T\mathbf{A}\mathbf{p} = \mathbf{A}^T\mathbf{b}$ (i.e., $\mathbf{p}$ is also a particular solution of the normal equations), so the general solution of $\mathbf{Ax} = \mathbf{b}$ is $\mathcal{S} = \mathbf{p} + N(\mathbf{A})$, and the general solution of $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ is

$$\mathbf{p} + N\left(\mathbf{A}^T\mathbf{A}\right) = \mathbf{p} + N(\mathbf{A}) = \mathcal{S}.$$

Furthermore, if $\mathbf{Ax} = \mathbf{b}$ is consistent and has a unique solution, then the same is true for $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, and the unique solution common to both systems is

$$\mathbf{x} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}. \tag{4.5.7}$$

This follows because a unique solution (to either system) exists if and only if $\mathbf{0} = N(\mathbf{A}) = N(\mathbf{A}^T\mathbf{A})$, and this insures $(\mathbf{A}^T\mathbf{A})_{n\times n}$ must be nonsingular (by (4.2.11)), so (4.5.7) is the unique solution to both systems. **Caution!** When $\mathbf{A}$ is not square, $\mathbf{A}^{-1}$ does not exist, and the reverse order law for inversion doesn't apply to $(\mathbf{A}^T\mathbf{A})^{-1}$, so (4.5.7) cannot be further simplified.

There is one outstanding question—what do the solutions of the normal equations $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ represent when the original system $\mathbf{A}\mathbf{x} = \mathbf{b}$ is *not* consistent? The answer, which is of fundamental importance, will have to wait until §4.6, but let's summarize what has been said so far.

Normal Equations

- For an $m \times n$ system $\mathbf{A}\mathbf{x} = \mathbf{b}$, the associated system of ***normal equations*** is defined to be the $n \times n$ system $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$.
- $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ is always consistent, even when $\mathbf{A}\mathbf{x} = \mathbf{b}$ is not consistent.
- When $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent, its solution set agrees with that of $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$. As discussed in §4.6, the normal equations provide least squares solutions to $\mathbf{A}\mathbf{x} = \mathbf{b}$ when $\mathbf{A}\mathbf{x} = \mathbf{b}$ is inconsistent.
- $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ has a unique solution if and only if $rank(\mathbf{A}) = n$, in which case the unique solution is $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.
- When $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent and has a unique solution, then the same is true for $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, and the unique solution to both systems is given by $\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.

Example 4.5.1

Caution! Use of the product $\mathbf{A}^T\mathbf{A}$ or the normal equations is not recommended for numerical computation. Any sensitivity to small perturbations that is present in the underlying matrix $\mathbf{A}$ is magnified by forming the product $\mathbf{A}^T\mathbf{A}$. In other words, if $\mathbf{A}\mathbf{x} = \mathbf{b}$ is somewhat ill-conditioned, then the associated system of normal equations $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ will be ill-conditioned to an even greater extent, and the theoretical properties surrounding $\mathbf{A}^T\mathbf{A}$ and the normal equations may be lost in practical applications. For example, consider the nonsingular system $\mathbf{A}\mathbf{x} = \mathbf{b}$, where

$$\mathbf{A} = \begin{pmatrix} 3 & 6 \\ 1 & 2.01 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 9 \\ 3.01 \end{pmatrix}.$$

If Gaussian elimination with 3-digit floating-point arithmetic is used to solve $\mathbf{A}\mathbf{x} = \mathbf{b}$, then the 3-digit solution is $(1,\ 1)$, and this agrees with the exact solution. However if 3-digit arithmetic is used to form the associated system of normal equations, the result is

$$\begin{pmatrix} 10 & 20 \\ 20 & 40 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 30 \\ 60.1 \end{pmatrix}.$$

The 3-digit representation of $\mathbf{A}^T\mathbf{A}$ is singular, and the associated system of normal equations is inconsistent. For these reasons, the normal equations are often avoided in numerical computations. Nevertheless, the normal equations are an important theoretical idea that leads to practical tools of fundamental importance such as the method of least squares developed in §4.6 and §5.13.

Because the concept of rank is at the heart of our subject, it's important to understand rank from a variety of different viewpoints. The statement below is one more way to think about rank.[29]

Rank and the Largest Nonsingular Submatrix

The rank of a matrix $\mathbf{A}_{m\times n}$ is precisely the order of a maximal square nonsingular submatrix of $\mathbf{A}$. In other words, to say $rank\,(\mathbf{A}) = r$ means that there is at least one $r \times r$ nonsingular submatrix in $\mathbf{A}$, and there are no nonsingular submatrices of larger order.

Proof. First demonstrate that there exists an $r \times r$ nonsingular submatrix in $\mathbf{A}$, and then show there can be no nonsingular submatrix of larger order. Begin with the fact that there must be a maximal linearly independent set of r rows in $\mathbf{A}$ as well as a maximal independent set of r columns, and prove that the submatrix $\mathbf{M}_{r\times r}$ lying on the intersection of these r rows and r columns is nonsingular. The r independent rows can be permuted to the top, and the remaining rows can be annihilated using row operations, so

$$\mathbf{A} \overset{\text{row}}{\sim} \begin{pmatrix} \mathbf{U}_{r\times n} \\ \mathbf{0} \end{pmatrix}.$$

Now permute the r independent columns containing $\mathbf{M}$ to the left-hand side, and use column operations to annihilate the remaining columns to conclude that

$$\mathbf{A} \overset{\text{row}}{\sim} \begin{pmatrix} \mathbf{U}_{r\times n} \\ \mathbf{0} \end{pmatrix} \overset{\text{col}}{\sim} \begin{pmatrix} \mathbf{M}_{r\times r} & \mathbf{N} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \overset{\text{col}}{\sim} \begin{pmatrix} \mathbf{M}_{r\times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

---

[29] This is the last characterization of rank presented in this text, but historically this was the essence of the first definition (p. 44) of rank given by Georg Frobenius (p. 662) in 1879.

Rank isn't changed by row or column operations, so $r = rank\,(\mathbf{A}) = rank\,(\mathbf{M})$, and thus $\mathbf{M}$ is nonsingular. Now suppose that $\mathbf{W}$ is any other nonsingular submatrix of $\mathbf{A}$, and let $\mathbf{P}$ and $\mathbf{Q}$ be permutation matrices such that $\mathbf{PAQ} = \begin{pmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix}$. If

$$\mathbf{E} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{Y}\mathbf{W}^{-1} & \mathbf{I} \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} \mathbf{I} & -\mathbf{W}^{-1}\mathbf{X} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \quad \text{and} \quad \mathbf{S} = \mathbf{Z} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{X},$$

then

$$\mathbf{EPAQF} = \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix} \implies \mathbf{A} \sim \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix}, \tag{4.5.8}$$

and hence $r = rank\,(\mathbf{A}) = rank\,(\mathbf{W}) + rank\,(\mathbf{S}) \geq rank\,(\mathbf{W})$ (recall Example 3.9.3). This guarantees that no nonsingular submatrix of $\mathbf{A}$ can have order greater than $r = rank\,(\mathbf{A})$. ■

Example 4.5.2

Problem: Determine the rank of $\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 1 \\ 3 & 6 & 1 \end{pmatrix}$.

Solution: $rank\,(\mathbf{A}) = 2$ because there is at least one 2×2 nonsingular submatrix (e.g., there is one lying on the intersection of rows 1 and 2 with columns 2 and 3), and there is no larger nonsingular submatrix (the entire matrix is singular). Notice that not all 2×2 matrices are nonsingular (e.g., consider the one lying on the intersection of rows 1 and 2 with columns 1 and 2).

Earlier in this section we saw that it is impossible to *increase* the rank by means of matrix multiplication—i.e., (4.5.2) says $rank\,(\mathbf{AE}) \leq rank\,(\mathbf{A})$. In a certain sense there is a dual statement for matrix addition that says that it is impossible to *decrease* the rank by means of a "small" matrix addition—i.e., $rank\,(\mathbf{A} + \mathbf{E}) \geq rank\,(\mathbf{A})$ whenever $\mathbf{E}$ has entries of small magnitude.

Small Perturbations Can't Reduce Rank

If $\mathbf{A}$ and $\mathbf{E}$ are $m \times n$ matrices such that $\mathbf{E}$ has entries of sufficiently small magnitude, then

$$rank\,(\mathbf{A} + \mathbf{E}) \geq rank\,(\mathbf{A}). \tag{4.5.9}$$

The term "sufficiently small" is further clarified in Exercise 5.12.4.

Proof. Suppose $rank(\mathbf{A}) = r$, and let $\mathbf{P}$ and $\mathbf{Q}$ be nonsingular matrices that reduce $\mathbf{A}$ to rank normal form—i.e., $\mathbf{PAQ} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. If $\mathbf{P}$ and $\mathbf{Q}$ are applied to $\mathbf{E}$ to form $\mathbf{PEQ} = \begin{pmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{pmatrix}$, where $\mathbf{E}_{11}$ is $r \times r$, then

$$\mathbf{P}(\mathbf{A}+\mathbf{E})\mathbf{Q} = \begin{pmatrix} \mathbf{I}_r + \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{pmatrix}. \tag{4.5.10}$$

If the magnitude of the entries in $\mathbf{E}$ are small enough to insure that $\mathbf{E}_{11}^k \to \mathbf{0}$ as $k \to \infty$, then the discussion of the Neumann series on p. 126 insures that $\mathbf{I} + \mathbf{E}_{11}$ is nonsingular. (Exercise 4.5.14 gives another condition on the size of $\mathbf{E}_{11}$ to insure this.) This allows the right-hand side of (4.5.10) to be further reduced by writing

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{E}_{21}(\mathbf{I}+\mathbf{E}_{11})^{-1} & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{I}+\mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{pmatrix}\begin{pmatrix} \mathbf{I} & -(\mathbf{I}+\mathbf{E}_{11})^{-1}\mathbf{E}_{12} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{I}-\mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix},$$

where $\mathbf{S} = \mathbf{E}_{22} - \mathbf{E}_{21}\left(\mathbf{I}+\mathbf{E}_{11}\right)^{-1}\mathbf{E}_{12}$. In other words,

$$\mathbf{A}+\mathbf{E} \sim \begin{pmatrix} \mathbf{I}-\mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix},$$

and therefore

$$\begin{aligned} rank(\mathbf{A}+\mathbf{E}) &= rank(\mathbf{I}_r+\mathbf{E}_{11}) + rank(\mathbf{S}) \quad \text{(recall Example 3.9.3)} \\ &= rank(\mathbf{A}) + rank(\mathbf{S}) \\ &\geq rank(\mathbf{A}). \quad \blacksquare \end{aligned} \tag{4.5.11}$$

Example 4.5.3

A Pitfall in Solving Singular Systems. Solving $\mathbf{Ax} = \mathbf{b}$ with floating-point arithmetic produces the exact solution of a perturbed system whose coefficient matrix is $\mathbf{A}+\mathbf{E}$. If $\mathbf{A}$ is nonsingular, and if we are using a stable algorithm (an algorithm that insures that the entries in $\mathbf{E}$ have small magnitudes), then (4.5.9) guarantees that we are finding the exact solution to a nearby system that is also nonsingular. On the other hand, if $\mathbf{A}$ is singular, then perturbations of even the slightest magnitude can increase the rank, thereby producing a system with fewer free variables than the original system theoretically demands, so even a stable algorithm can result in a significant loss of information. But what are the chances that this will actually occur in practice? To answer this, recall from (4.5.11) that

$$rank(\mathbf{A}+\mathbf{E}) = rank(\mathbf{A}) + rank(\mathbf{S}), \quad \text{where} \quad \mathbf{S} = \mathbf{E}_{22} - \mathbf{E}_{21}\left(\mathbf{I}+\mathbf{E}_{11}\right)^{-1}\mathbf{E}_{12}.$$

If the rank is not to jump, then the perturbation $\mathbf{E}$ must be such that $\mathbf{S} = \mathbf{0}$, which is equivalent to saying $\mathbf{E}_{22} = \mathbf{E}_{21}\left(\mathbf{I} + \mathbf{E}_{11}\right)^{-1}\mathbf{E}_{12}$. Clearly, this requires the existence of a very specific (and quite special) relationship among the entries of $\mathbf{E}$, and a random perturbation will almost never produce such a relationship. Although rounding errors cannot be considered to be truly random, they are random enough so as to make the possibility that $\mathbf{S} = \mathbf{0}$ very unlikely. Consequently, when $\mathbf{A}$ is singular, the small perturbation $\mathbf{E}$ due to roundoff makes the possibility that $rank\,(\mathbf{A} + \mathbf{E}) > rank\,(\mathbf{A})$ very likely. The moral is to avoid floating-point solutions of singular systems. Singular problems can often be distilled down to a nonsingular core or to nonsingular pieces, and these are the components you should be dealing with.

Since no more significant characterizations of rank will be given, it is appropriate to conclude this section with a summary of all of the different ways we have developed to say "rank."

Summary of Rank

For $\mathbf{A} \in \Re^{m\times n}$, each of the following statements is true.

- $rank\,(\mathbf{A}) =$ The number of nonzero rows in any row echelon form that is row equivalent to $\mathbf{A}$.
- $rank\,(\mathbf{A}) =$ The number of pivots obtained in reducing $\mathbf{A}$ to a row echelon form with row operations.
- $rank\,(\mathbf{A}) =$ The number of basic columns in $\mathbf{A}$ (as well as the number of basic columns in any matrix that is row equivalent to $\mathbf{A}$).
- $rank\,(\mathbf{A}) =$ The number of independent columns in $\mathbf{A}$—i.e., the size of a maximal independent set of columns from $\mathbf{A}$.
- $rank\,(\mathbf{A}) =$ The number of independent rows in $\mathbf{A}$—i.e., the size of a maximal independent set of rows from $\mathbf{A}$.
- $rank\,(\mathbf{A}) = \dim R\,(\mathbf{A})$.
- $rank\,(\mathbf{A}) = \dim R\left(\mathbf{A}^T\right)$.
- $rank\,(\mathbf{A}) = n - \dim N\,(\mathbf{A})$.
- $rank\,(\mathbf{A}) = m - \dim N\left(\mathbf{A}^T\right)$.
- $rank\,(\mathbf{A}) =$ The size of the largest nonsingular submatrix in $\mathbf{A}$.

For $\mathbf{A} \in \mathcal{C}^{m\times n}$, replace $(\star)^T$ with $(\star)^*$.

Exercises for section 4.5

4.5.1. Verify that $rank\left(\mathbf{A}^T\mathbf{A}\right) = rank\left(\mathbf{A}\right) = rank\left(\mathbf{A}\mathbf{A}^T\right)$ for

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 1 & -4 \\ -1 & -3 & 1 & 0 \\ 2 & 6 & 2 & -8 \end{pmatrix}.$$

4.5.2. Determine $\dim N(\mathbf{A}) \cap R(\mathbf{B})$ for

$$\mathbf{A} = \begin{pmatrix} -2 & 1 & 1 \\ -4 & 2 & 2 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 3 & 1 & -4 \\ -1 & -3 & 1 & 0 \\ 2 & 6 & 2 & -8 \end{pmatrix}.$$

4.5.3. For the matrices given in Exercise 4.5.2, use the procedure described on p. 211 to determine a basis for $N(\mathbf{A}) \cap R(\mathbf{B})$.

4.5.4. If $\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_k$ is a product of square matrices such that some $\mathbf{A}_i$ is singular, explain why the entire product must be singular.

4.5.5. For $\mathbf{A} \in \Re^{m\times n}$, explain why $\mathbf{A}^T\mathbf{A} = \mathbf{0}$ implies $\mathbf{A} = \mathbf{0}$.

4.5.6. Find $rank\left(\mathbf{A}\right)$ and all nonsingular submatrices of maximal order in

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 1 \\ 4 & -2 & 1 \\ 8 & -4 & 1 \end{pmatrix}.$$

4.5.7. Is it possible that $rank\left(\mathbf{AB}\right) < rank\left(\mathbf{A}\right)$ and $rank\left(\mathbf{AB}\right) < rank\left(\mathbf{B}\right)$ for the same pair of matrices?

4.5.8. Is $rank\left(\mathbf{AB}\right) = rank\left(\mathbf{BA}\right)$ when both products are defined? Why?

4.5.9. Explain why $rank\left(\mathbf{AB}\right) = rank\left(\mathbf{A}\right) - \dim N\left(\mathbf{B}^T\right) \cap R\left(\mathbf{A}^T\right)$.

4.5.10. Explain why $\dim N\left(\mathbf{A}_{m\times n}\mathbf{B}_{n\times p}\right) = \dim N(\mathbf{B}) + \dim R(\mathbf{B}) \cap N(\mathbf{A})$.

4.5.11. ***Sylvester's law of nullity,*** given by James J. Sylvester in 1884, states that for *square matrices* $\mathbf{A}$ and $\mathbf{B}$,

$$\max\{\nu(\mathbf{A}),\ \nu(\mathbf{B})\} \le \nu(\mathbf{AB}) \le \nu(\mathbf{A}) + \nu(\mathbf{B}),$$

where $\nu(\star) = \dim N(\star)$ denotes the nullity.

(a) Establish the validity of Sylvester's law.

(b) Show Sylvester's law is not valid for rectangular matrices because $\nu(\mathbf{A}) > \nu(\mathbf{AB})$ is possible. Is $\nu(\mathbf{B}) > \nu(\mathbf{AB})$ possible?

4.5.12. For matrices $\mathbf{A}_{m\times n}$ and $\mathbf{B}_{n\times p}$, prove each of the following statements:

(a) $rank(\mathbf{AB}) = rank(\mathbf{A})$ and $R(\mathbf{AB}) = R(\mathbf{A})$ if $rank(\mathbf{B}) = n$.

(b) $rank(\mathbf{AB}) = rank(\mathbf{B})$ and $N(\mathbf{AB}) = N(\mathbf{B})$ if $rank(\mathbf{A}) = n$.

4.5.13. Perform the following calculations using the matrices:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 1 & 2.01 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 1.01 \end{pmatrix}.$$

(a) Find $rank(\mathbf{A})$, and solve $\mathbf{Ax} = \mathbf{b}$ using exact arithmetic.

(b) Find $rank\left(\mathbf{A}^T\mathbf{A}\right)$, and solve $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ exactly.

(c) Find $rank(\mathbf{A})$, and solve $\mathbf{Ax} = \mathbf{b}$ with 3-digit arithmetic.

(d) Find $\mathbf{A}^T\mathbf{A}$, $\mathbf{A}^T\mathbf{b}$, and the solution of $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ with 3-digit arithmetic.

4.5.14. Prove that if the entries of $\mathbf{F}_{r\times r}$ satisfy $\sum_{j=1}^{r} |f_{ij}| < 1$ for each i (i.e., each absolute row sum < 1), then $\mathbf{I} + \mathbf{F}$ is nonsingular. **Hint:** Use the triangle inequality for scalars $|\alpha+\beta| \le |\alpha| + |\beta|$ to show $N(\mathbf{I} + \mathbf{F}) = \mathbf{0}$.

4.5.15. If $\mathbf{A} = \begin{pmatrix} \mathbf{W} & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{pmatrix}$, where $rank(\mathbf{A}) = r = rank(\mathbf{W}_{r\times r})$, show that there are matrices $\mathbf{B}$ and $\mathbf{C}$ such that

$$\mathbf{A} = \begin{pmatrix} \mathbf{W} & \mathbf{WC} \\ \mathbf{BW} & \mathbf{BWC} \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \mathbf{B} \end{pmatrix} \mathbf{W} \left(\mathbf{I} \mid \mathbf{C}\right).$$

4.5.16. For a convergent sequence $\{\mathbf{A}_k\}_{k=1}^{\infty}$ of matrices, let $\mathbf{A} = \lim_{k\to\infty} \mathbf{A}_k$.

(a) Prove that if each $\mathbf{A}_k$ is singular, then $\mathbf{A}$ is singular.

(b) If each $\mathbf{A}_k$ is nonsingular, must $\mathbf{A}$ be nonsingular? Why?

4.5.17. The Frobenius Inequality. Establish the validity of Frobenius's 1911 result that states that if $\mathbf{ABC}$ exists, then

$$rank\,(\mathbf{AB}) + rank\,(\mathbf{BC}) \leq rank\,(\mathbf{B}) + rank\,(\mathbf{ABC}).$$

Hint: If $\mathcal{M} = R\,(\mathbf{BC}) \cap N\,(\mathbf{A})$ and $\mathcal{N} = R\,(\mathbf{B}) \cap N\,(\mathbf{A})$, then $\mathcal{M} \subseteq \mathcal{N}$.

4.5.18. If $\mathbf{A}$ is $n \times n$, prove that the following statements are equivalent:

(a) $N\,(\mathbf{A}) = N\left(\mathbf{A}^2\right)$.
(b) $R\,(\mathbf{A}) = R\left(\mathbf{A}^2\right)$.
(c) $R\,(\mathbf{A}) \cap N\,(\mathbf{A}) = \{\mathbf{0}\}$.

4.5.19. Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ matrices such that $\mathbf{A} = \mathbf{A}^2$, $\mathbf{B} = \mathbf{B}^2$, and $\mathbf{AB} = \mathbf{BA} = \mathbf{0}$.

(a) Prove that $rank\,(\mathbf{A} + \mathbf{B}) = rank\,(\mathbf{A}) + rank\,(\mathbf{B})$. **Hint:** Consider $\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} (\mathbf{A} + \mathbf{B})(\mathbf{A} \mid \mathbf{B})$.
(b) Prove that $rank\,(\mathbf{A}) + rank\,(\mathbf{I} - \mathbf{A}) = n$.

4.5.20. Moore–Penrose Inverse. For $\mathbf{A} \in \Re^{m \times n}$ such that $rank\,(\mathbf{A}) = r$, let $\mathbf{A} = \mathbf{BC}$ be the full rank factorization of $\mathbf{A}$ in which $\mathbf{B}_{m \times r}$ is the matrix of basic columns from $\mathbf{A}$ and $\mathbf{C}_{r \times n}$ is the matrix of nonzero rows from $\mathbf{E_A}$ (see Exercise 3.9.8). The matrix defined by

$$\mathbf{A}^\dagger = \mathbf{C}^T \left(\mathbf{B}^T \mathbf{A} \mathbf{C}^T\right)^{-1} \mathbf{B}^T$$

is called the ***Moore–Penrose***[30] ***inverse*** of $\mathbf{A}$. Some authors refer to $\mathbf{A}^\dagger$ as the *pseudoinverse* or the *generalized inverse* of $\mathbf{A}$. A more elegant treatment is given on p. 423, but it's worthwhile to introduce the idea here so that it can be used and viewed from different perspectives.

(a) Explain why the matrix $\mathbf{B}^T \mathbf{A} \mathbf{C}^T$ is nonsingular.
(b) Verify that $\mathbf{x} = \mathbf{A}^\dagger \mathbf{b}$ solves the normal equations $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$ (as well as $\mathbf{Ax} = \mathbf{b}$ when it is consistent).
(c) Show that the general solution for $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$ (as well as $\mathbf{Ax} = \mathbf{b}$ when it is consistent) can be described as

$$\mathbf{x} = \mathbf{A}^\dagger \mathbf{b} + \left(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}\right) \mathbf{h},$$

[30] This is in honor of Eliakim H. Moore (1862–1932) and Roger Penrose (a famous contemporary English mathematical physicist). Each formulated a concept of generalized matrix inversion—Moore's work was published in 1922, and Penrose's work appeared in 1955. E. H. Moore is considered by many to be America's first great mathematician.

where $\mathbf{h}$ is a "free variable" vector in $\Re^{n\times 1}$.
Hint: Verify $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{A}$, and then show $R\left(\mathbf{I} - \mathbf{A}^{\dagger}\mathbf{A}\right) = N(\mathbf{A})$.

(d) If $rank(\mathbf{A}) = n$, explain why $\mathbf{A}^{\dagger} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$.

(e) If $\mathbf{A}$ is square and nonsingular, explain why $\mathbf{A}^{\dagger} = \mathbf{A}^{-1}$.

(f) Verify that $\mathbf{A}^{\dagger} = \mathbf{C}^T\left(\mathbf{B}^T\mathbf{A}\mathbf{C}^T\right)^{-1}\mathbf{B}^T$ satisfies the Penrose equations:

$$\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{A}, \qquad \left(\mathbf{A}\mathbf{A}^{\dagger}\right)^T = \mathbf{A}\mathbf{A}^{\dagger},$$
$$\mathbf{A}^{\dagger}\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{A}^{\dagger}, \qquad \left(\mathbf{A}^{\dagger}\mathbf{A}\right)^T = \mathbf{A}^{\dagger}\mathbf{A}.$$

Penrose originally defined $\mathbf{A}^{\dagger}$ to be the unique solution to these four equations.

columns) in $\mathbf{P}$ are the unit vectors that appear according to the permutation $\boldsymbol{\pi} = \begin{pmatrix} 1 & 2 & \cdots & n \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$, then

$$a_{ij} = \left[\mathbf{PBP}^T\right]_{ij} = \left[\begin{pmatrix} \mathbf{e}_{\pi_1}^T \\ \vdots \\ \mathbf{e}_{\pi_n}^T \end{pmatrix} \mathbf{B} \left(\mathbf{e}_{\pi_1} \quad \cdots \quad \mathbf{e}_{\pi_n} \right) \right]_{ij} = \mathbf{e}_{\pi_i}^T \mathbf{B} \mathbf{e}_{\pi_j} = b_{\pi_i \pi_j}.$$

Consequently, $a_{ij} \neq 0$ if and only if $b_{\pi_i\pi_j} \neq 0$, and thus $\mathcal{G}(\mathbf{A})$ and $\mathcal{G}(\mathbf{B})$ are the same except for the fact that node N_k in $\mathcal{G}(\mathbf{A})$ is node N_{π_k} in $\mathcal{G}(\mathbf{B})$ for each $k = 1, 2, \ldots, n$. Now we can prove $\mathcal{G}(\mathbf{A})$ is *not* strongly connected $\iff$ $\mathbf{A}$ is reducible. If $\mathbf{A}$ is reducible, then there is a permutation matrix such that $\mathbf{P}^T\mathbf{AP} = \mathbf{B} = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{0} & \mathbf{Z} \end{pmatrix}$, where $\mathbf{X}$ is $r \times r$ and $\mathbf{Z}$ is $n-r \times n-r$. The zero pattern in $\mathbf{B}$ indicates that the nodes $\{N_1, N_2, \ldots, N_r\}$ in $\mathcal{G}(\mathbf{B})$ are inaccessible from nodes $\{N_{r+1}, N_{r+2}, \ldots, N_n\}$, and hence $\mathcal{G}(\mathbf{B})$ is not strongly connected—e.g., there is no sequence of edges leading from N_{r+1} to N_1. Since $\mathcal{G}(\mathbf{B})$ is the same as $\mathcal{G}(\mathbf{A})$ except that the nodes have different numbers, we may conclude that $\mathcal{G}(\mathbf{A})$ is also not strongly connected. Conversely, if $\mathcal{G}(\mathbf{A})$ is not strongly connected, then there are two nodes in $\mathcal{G}(\mathbf{A})$ such that one is inaccessible from the other by any sequence of directed edges. Relabel the nodes in $\mathcal{G}(\mathbf{A})$ so that this pair is N_1 and N_n, where N_1 is inaccessible from N_n. If there are additional nodes—excluding N_n itself—which are also inaccessible from N_n, label them $N_2, N_3, \ldots, N_r$ so that the set of all nodes that are inaccessible from N_n—with the possible exception of N_n itself—is $\overline{N_n} = \{N_1, N_2, \ldots, N_r\}$ (inaccessible nodes). Label the remaining nodes—which are all accessible from N_n—as $\underline{N_n} = \{N_{r+1}, N_{r+2}, \ldots, N_{n-1}\}$ (accessible nodes). It follows that no node in $\overline{N_n}$ can be accessible from any node in $\underline{N_n}$, for otherwise nodes in $\overline{N_n}$ would be accessible from N_n through nodes in $\underline{N_n}$. In other words, if $N_{r+k} \in \underline{N_n}$ and $N_{r+k} \to N_i \in \overline{N_n}$, then $N_n \to N_{r+k} \to N_i$, which is impossible. This means that if $\boldsymbol{\pi} = \begin{pmatrix} 1 & 2 & \cdots & n \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$ is the permutation generated by the relabeling process, then $a_{\pi_i\pi_j} = 0$ for each $i = r+1, r+2, \ldots, n-1$ and $j = 1, 2, \ldots, r$. Therefore, if $\mathbf{B} = \mathbf{P}^T\mathbf{AP}$, where $\mathbf{P}$ is the permutation matrix corresponding to the permutation $\boldsymbol{\pi}$, then $b_{ij} = a_{\pi_i\pi_j}$, so $\mathbf{P}^T\mathbf{AP} = \mathbf{B} = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{0} & \mathbf{Z} \end{pmatrix}$, where $\mathbf{X}$ is $r \times r$ and $\mathbf{Z}$ is $n-r \times n-r$, and thus $\mathbf{A}$ is reducible.

Solutions for exercises in section 4. 5

4.5.1. $rank\left(\mathbf{A}^T\mathbf{A}\right) = rank\,(\mathbf{A}) = rank\left(\mathbf{A}\mathbf{A}^T\right) = 2$

4.5.2. $\dim N(\mathbf{A}) \cap R(\mathbf{B}) = rank\,(\mathbf{B}) - rank\,(\mathbf{AB}) = 2 - 1 = 1.$

4.5.3. Gaussian elimination yields $\mathbf{X} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 2 & 2 \end{pmatrix}$, $\mathbf{V} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and $\mathbf{XV} = \begin{pmatrix} 2 \\ 0 \\ 4 \end{pmatrix}$.

4.5.4. Statement (4.5.2) says that the rank of a product cannot exceed the rank of any factor.

4.5.5. $rank\,(\mathbf{A}) = rank\,(\mathbf{A}^T\mathbf{A}) = 0 \implies \mathbf{A} = \mathbf{0}$.

4.5.6. $rank\,(\mathbf{A}) = 2$, and there are *six* 2×2 nonsingular submatrices in $\mathbf{A}$.

4.5.7. Yes. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$ is one of many examples.

4.5.8. No—it is not difficult to construct a counterexample using two singular matrices. If either matrix is nonsingular, then the statement is true.

4.5.9. Transposition does not alter rank, so (4.5.1) says

$$\begin{aligned} rank\,(\mathbf{AB}) &= rank(\mathbf{AB})^T = rank\,(\mathbf{B}^T\mathbf{A}^T) \\ &= rank\,(\mathbf{A}^T) - \dim N\,(\mathbf{B}^T) \cap R\,(\mathbf{A}^T) \\ &= rank\,(\mathbf{A}) - \dim N\,(\mathbf{B}^T) \cap R\,(\mathbf{A}^T). \end{aligned}$$

4.5.10. This follows immediately from (4.5.1) because $\dim N\,(\mathbf{AB}) = p - rank\,(\mathbf{AB})$ and $\dim N\,(\mathbf{B}) = p - rank\,(\mathbf{B})$.

4.5.11. (a) First notice that $N\,(\mathbf{B}) \subseteq N\,(\mathbf{AB})$ (Exercise 4.2.12) for all conformable $\mathbf{A}$ and $\mathbf{B}$, so, by (4.4.5), $\dim N\,(\mathbf{B}) \le \dim N\,(\mathbf{AB})$, or $\nu(\mathbf{B}) \le \nu(\mathbf{AB})$, is always true—this also answers the second half of part (b). If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times n$, then the rank-plus-nullity theorem together with (4.5.2) produces

$$\nu(\mathbf{A}) = \dim N\,(\mathbf{A}) = n - rank\,(\mathbf{A}) \le n - rank\,(\mathbf{AB}) = \dim N\,(\mathbf{AB}) = \nu(\mathbf{AB}),$$

so, together with the first observation, we have $\max\{\nu(\mathbf{A}),\ \nu(\mathbf{B})\} \le \nu(\mathbf{AB})$. The rank-plus-nullity theorem applied to (4.5.3) yields $\nu(\mathbf{AB}) \le \nu(\mathbf{A}) + \nu(\mathbf{B})$.

(b) To see that $\nu(\mathbf{A}) > \nu(\mathbf{AB})$ is possible for rectangular matrices, consider $\mathbf{A} = (\,1 \quad 1\,)$ and $\mathbf{B} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

4.5.12. (a) $rank\,(\mathbf{B}_{n\times p}) = n \implies R\,(\mathbf{B}) = \Re^n \implies N\,(\mathbf{A}) \cap R\,(\mathbf{B}) = N\,(\mathbf{A}) \implies$

$$\begin{aligned} rank\,(\mathbf{AB}) &= rank\,(\mathbf{B}) - \dim N\,(\mathbf{A}) \cap R\,(\mathbf{B}) = n - \dim N\,(\mathbf{A}) \\ &= n - (n - rank\,(\mathbf{A})) = rank\,(\mathbf{A}) \end{aligned}$$

It's always true that $R\,(\mathbf{AB}) \subseteq R\,(\mathbf{A})$. When $\dim R\,(\mathbf{AB}) = \dim R\,(\mathbf{A})$ (i.e., when $rank\,(\mathbf{AB}) = rank\,(\mathbf{A})$), (4.4.6) implies $R\,(\mathbf{AB}) = R\,(\mathbf{A})$.

(b) $rank\,(\mathbf{A}_{m\times n}) = n \implies N\,(\mathbf{A}) = \{\mathbf{0}\} \implies N\,(\mathbf{A}) \cap R\,(\mathbf{B}) = \{\mathbf{0}\} \implies$

$$rank\,(\mathbf{AB}) = rank\,(\mathbf{B}) - \dim N\,(\mathbf{A}) \cap R\,(\mathbf{B}) = rank\,(\mathbf{B})$$

Assuming the product exists, it is always the case that $N\,(\mathbf{B}) \subseteq N\,(\mathbf{AB})$. Use $rank\,(\mathbf{B}) = rank\,(\mathbf{AB}) \implies p - rank\,(\mathbf{B}) = p - rank\,(\mathbf{AB}) \implies \dim N\,(\mathbf{B}) = \dim N\,(\mathbf{AB})$ together with (4.4.6) to conclude that $N\,(\mathbf{B}) = N\,(\mathbf{AB})$.

4.5.13. (a) $rank\,(\mathbf{A}) = 2$, and the unique exact solution is $(-1,\ 1)$.

(b) Same as part (a).

(c) The 3-digit rank is 2, and the unique 3-digit solution is $(-1,\ 1)$.

(d) The 3-digit normal equations $\begin{pmatrix} 6 & 12 \\ 12 & 24 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6.01 \\ 12 \end{pmatrix}$ have infinitely many 3-digit solutions.

4.5.14. Use an indirect argument. Suppose $\mathbf{x} \in N(\mathbf{I}+\mathbf{F})$ in which $x_i \neq 0$ is a component of maximal magnitude. Use the triangle inequality together with $\mathbf{x} = -\mathbf{F}\mathbf{x}$ to conclude

$$|x_i| = \Big|\sum_{j=1}^{r} f_{ij}x_j\Big| \le \sum_{j=1}^{r} |f_{ij}x_j| = \sum_{j=1}^{r} |f_{ij}|\,|x_j| \le \Big(\sum_{j=1}^{r} |f_{ij}|\Big)|x_i| < |x_i|,$$

which is impossible. Therefore, $N(\mathbf{I}+\mathbf{F}) = \mathbf{0}$, and hence $\mathbf{I}+\mathbf{F}$ is nonsingular.

4.5.15. Follow the approach used in (4.5.8) to write

$$\mathbf{A} \sim \begin{pmatrix} \mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{pmatrix}, \quad \text{where} \quad \mathbf{S} = \mathbf{Z} - \mathbf{Y}\mathbf{W}^{-1}\mathbf{X}.$$

$rank\,(\mathbf{A}) = rank\,(\mathbf{W}) \implies rank\,(\mathbf{S}) = 0 \implies \mathbf{S} = \mathbf{0}$, so $\mathbf{Z} = \mathbf{Y}\mathbf{W}^{-1}\mathbf{X}$. The desired conclusion now follows by taking $\mathbf{B} = \mathbf{Y}\mathbf{W}^{-1}$ and $\mathbf{C} = \mathbf{W}^{-1}\mathbf{X}$.

4.5.16. (a) Suppose that $\mathbf{A}$ is nonsingular, and let $\mathbf{E}_k = \mathbf{A}_k - \mathbf{A}$ so that $\lim_{k\to\infty} \mathbf{E}_k = \mathbf{0}$. This together with (4.5.9) implies there exists a sufficiently large value of k such that

$$rank\,(\mathbf{A}_k) = rank\,(\mathbf{A}+\mathbf{E}_k) \ge rank\,(\mathbf{A}) = n,$$

which is impossible because each $\mathbf{A}_k$ is singular. Therefore, the supposition that $\mathbf{A}$ is nonsingular must be false.

(b) No!—consider $\left[\frac{1}{k}\right]_{1\times 1} \to [0]$.

4.5.17. $\mathcal{M} \subseteq \mathcal{N}$ because $R(\mathbf{BC}) \subseteq R(\mathbf{B})$, and therefore $\dim \mathcal{M} \le \dim \mathcal{N}$. Formula (4.5.1) guarantees $\dim \mathcal{M} = rank\,(\mathbf{BC}) - rank\,(\mathbf{ABC})$ and $\dim \mathcal{N} = rank\,(\mathbf{B}) - rank\,(\mathbf{AB})$, so the desired conclusion now follows.

4.5.18. $N(\mathbf{A}) \subseteq N(\mathbf{A}^2)$ and $R(\mathbf{A}^2) \subseteq R(\mathbf{A})$ always hold, so (4.4.6) insures

$$\begin{aligned} N(\mathbf{A}) = N(\mathbf{A}^2) &\iff \dim N(\mathbf{A}) = \dim N(\mathbf{A}^2) \\ &\iff n - rank\,(\mathbf{A}) = n - rank\,(\mathbf{A}^2) \\ &\iff rank\,(\mathbf{A}) = rank\,(\mathbf{A}^2) \\ &\iff R(\mathbf{A}) = R(\mathbf{A}^2). \end{aligned}$$

Formula (4.5.1) says $rank\,(\mathbf{A}^2) = rank\,(\mathbf{A}) - \dim R(\mathbf{A}) \cap N(\mathbf{A})$, so
$R(\mathbf{A}^2) = R(\mathbf{A}) \iff rank\,(\mathbf{A}^2) = rank\,(\mathbf{A}) \iff \dim R(\mathbf{A}) \cap N(\mathbf{A}) = 0.$

4.5.19. (a) Since

$$\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} (\mathbf{A}+\mathbf{B})(\mathbf{A} \mid \mathbf{B}) = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} (\mathbf{A} \mid \mathbf{B}) = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix},$$

the result of Example 3.9.3 together with (4.5.2) insures

$$rank\,(\mathbf{A}) + rank\,(\mathbf{B}) \le rank\,(\mathbf{A}+\mathbf{B}).$$

Couple this with the fact that $rank\,(\mathbf{A}+\mathbf{B}) \le rank\,(\mathbf{A}) + rank\,(\mathbf{B})$ (see Example 4.4.8) to conclude $rank\,(\mathbf{A}+\mathbf{B}) = rank\,(\mathbf{A}) + rank\,(\mathbf{B})$.
(b) Verify that if $\mathbf{B} = \mathbf{I} - \mathbf{A}$, then $\mathbf{B}^2 = \mathbf{B}$ and $\mathbf{AB} = \mathbf{BA} = \mathbf{0}$, and apply the result of part (a).

4.5.20. (a) $\mathbf{B}^T\mathbf{A}\mathbf{C}^T = \mathbf{B}^T\mathbf{B}\mathbf{C}\mathbf{C}^T$. The products $\mathbf{B}^T\mathbf{B}$ and $\mathbf{C}\mathbf{C}^T$ are each nonsingular because they are $r \times r$ with

$$rank\,\left(\mathbf{B}^T\mathbf{B}\right) = rank\,(\mathbf{B}) = r \quad \text{and} \quad rank\,\left(\mathbf{C}\mathbf{C}^T\right) = rank\,(\mathbf{C}) = r.$$

(b) Notice that $\mathbf{A}^\dagger = \mathbf{C}^T\left(\mathbf{B}^T\mathbf{B}\mathbf{C}\mathbf{C}^T\right)^{-1}\mathbf{B}^T = \mathbf{C}^T\left(\mathbf{C}\mathbf{C}^T\right)^{-1}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T$, so

$$\mathbf{A}^T\mathbf{A}\mathbf{A}^\dagger\mathbf{b} = \mathbf{C}^T\mathbf{B}^T\mathbf{B}\mathbf{C}\mathbf{C}^T\left(\mathbf{C}\mathbf{C}^T\right)^{-1}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{b} = \mathbf{C}^T\mathbf{B}^T\mathbf{b} = \mathbf{A}^T\mathbf{b}.$$

If $\mathbf{Ax} = \mathbf{b}$ is consistent, then its solution set agrees with the solution set for the normal equations.
(c) $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{B}\mathbf{C}\mathbf{C}^T\left(\mathbf{C}\mathbf{C}^T\right)^{-1}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{B}\mathbf{C} = \mathbf{B}\mathbf{C} = \mathbf{A}$. Now,

$$\begin{aligned} \mathbf{x} \in R\left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right) &\implies \mathbf{x} = \left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right)\mathbf{y} \quad \text{for some } \mathbf{y} \\ &\implies \mathbf{Ax} = \left(\mathbf{A} - \mathbf{A}\mathbf{A}^\dagger\mathbf{A}\right)\mathbf{y} = \mathbf{0} \implies \mathbf{x} \in N\,(\mathbf{A}). \end{aligned}$$

Conversely,

$$\mathbf{x} \in N\,(\mathbf{A}) \implies \mathbf{Ax} = \mathbf{0} \implies \mathbf{x} = \left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right)\mathbf{x} \implies \mathbf{x} \in R\left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right),$$

so $R\left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right) = N\,(\mathbf{A})$. As $\mathbf{h}$ ranges over all of $\Re^{n\times 1}$, the expression $\left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right)\mathbf{h}$ generates $R\left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right) = N\,(\mathbf{A})$. Since $\mathbf{A}^\dagger\mathbf{b}$ is a particular solution of $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$, the general solution is

$$\mathbf{x} = \mathbf{A}^\dagger\mathbf{b} + N\,(\mathbf{A}) = \mathbf{A}^\dagger\mathbf{b} + \left(\mathbf{I} - \mathbf{A}^\dagger\mathbf{A}\right)\mathbf{h}.$$

(d) If $r = n$, then $\mathbf{B} = \mathbf{A}$ and $\mathbf{C} = \mathbf{I}_n$.
(e) If $\mathbf{A}$ is nonsingular, then so is $\mathbf{A}^T$, and

$$\mathbf{A}^\dagger = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T = \mathbf{A}^{-1}\left(\mathbf{A}^T\right)^{-1}\mathbf{A}^T = \mathbf{A}^{-1}.$$

(f) Follow along the same line as indicated in the solution to part (c) for the case $\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}$.