

3.4 Elementary Matrices and Inverses

Gaussian elimination is a list of instructions: swap these rows, scale that one, add a multiple of one row to another. So far it lives outside algebra, as something we do to a matrix. This section turns every row operation into a matrix, so that “apply an operation” becomes “multiply on the left”. With that one observation the algorithm becomes a tool for proofs. We use it to prove the main theorem about invertible matrices, to show that a one-sided inverse of a square matrix is automatically two-sided, which Chapter 0 stated after Definition 0.9.13 but could not prove, and to compute inverses.

Throughout, F is a field, and we use the three elementary row operations of Definition 3.2.1 with their notation: the swap $R_i \leftrightarrow R_j$ ($i \neq j$), the scaling $R_i \rightarrow cR_i$ (c **nonzero**), and the replacement $R_i \rightarrow R_i + cR_j$ ($i \neq j$, any $c \in F$).

3.4.1 Row operations are matrix multiplications

We have already seen one row operation disguised as a product. In Example (Row equivalence) of Chapter 0, multiplying $\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 0 \end{pmatrix}$ on the left by $\mathbf{E} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$ replaced row 2 by row 2 minus twice row 1, that is, it performed $R_2 \rightarrow R_2 - 2R_1$. Look at $\mathbf{E}$ itself: it is what that same operation does to the identity matrix $\mathbf{I}_2$. This is no accident, and it deserves a name.

An elementary matrix is the identity matrix after one row operation.

Definition 3.4.1. Elementary matrix

Let $m \geq 1$. A matrix $\mathbf{E} \in M_m(F)$ is an **elementary matrix** if it is obtained from $\mathbf{I}_m$ by performing **exactly one** elementary row operation. Writing $\mathbf{E}_{ij}$ for the matrix unit (1 in entry (i, j) , 0 elsewhere; see Example (Standard Bases)), the three kinds are:

1. $\mathbf{P}_{ij}$, obtained by $R_i \leftrightarrow R_j$ with $i \neq j$: the identity with rows i and j swapped;
2. $\mathbf{D}_i(c) = \text{diag}(1, \dots, 1, c, 1, \dots, 1)$, with c in position i , obtained by $R_i \rightarrow cR_i$ with $c \neq 0$;
3. $\mathbf{I}_m + c\mathbf{E}_{ij}$, obtained by $R_i \rightarrow R_i + cR_j$ with $i \neq j$ and $c \in F$: the identity with an extra c in entry (i, j) .

In words: start from $\mathbf{I}_m$, do **one** operation, and record the result. The size m is the number of **rows** of the matrices we plan to operate on. Type 3 puts c in row i , column j : the row that changes is i , and the row that gets copied is j .

Examples. In $M_3(F)$,

$$\mathbf{P}_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{D}_2(5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{I}_3 - 3\mathbf{E}_{31} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}$$

come from $R_1 \leftrightarrow R_3$, $R_2 \rightarrow 5R_2$ (over a field where $5 \neq 0$), and $R_3 \rightarrow R_3 - 3R_1$. The degenerate case is worth a look: $\mathbf{D}_1(1) = \mathbf{I}_m$ is elementary, since “multiply row 1 by 1” is a legal operation, and so is $\mathbf{I}_m + 0\mathbf{E}_{12}$. The identity is an elementary matrix. For $m = 1$ the elementary matrices are exactly the 1×1 matrices (c) with $c \neq 0$, since there are no two distinct rows to swap or combine.

Nonexample by minimal change. Change $c = 5$ in $\mathbf{D}_2(5)$ to $c = 0$. The matrix $\text{diag}(1, 0, 1)$

is still diagonal and still differs from $\mathbf{I}_3$ in one entry, but “multiply row 2 by 0” is not an elementary row operation: the clause $c \neq 0$ fails. Likewise $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ is a rearrangement of the rows of $\mathbf{I}_3$, but every one of its rows has moved, and a single swap moves only two rows. So it is not elementary. It is, however, the **product** $\mathbf{P}_{12}\mathbf{P}_{23}$ of two elementary matrices, as Exercise A1 at the end of this section checks.

The whole point of the definition is the next theorem. To perform a row operation on **any** matrix, multiply on the left by what that operation does to the identity.

Theorem 3.4.1. Row operations are left multiplications

Let ρ be an elementary row operation on matrices with m rows, and let $\mathbf{E} = \rho(\mathbf{I}_m)$ be the corresponding elementary matrix. Then for **every** $n \geq 1$ and every $\mathbf{A} \in M_{m \times n}(F)$,

$$\rho(\mathbf{A}) = \mathbf{E}\mathbf{A}.$$

Idea. Each operation builds every new row as a combination of the old rows, with coefficients that do not depend on the matrix: for $R_i \rightarrow R_i + cR_j$, new row i is $1 \cdot (\text{row } i) + c \cdot (\text{row } j)$, and every other row is copied. Feed in $\mathbf{I}_m$, whose rows are the standard row vectors, and the coefficients appear as the entries of $\mathbf{E}$. Then the row view of a product says $\mathbf{E}\mathbf{A}$ forms exactly those combinations of the rows of $\mathbf{A}$.

Proof. Let $\mathbf{r}_1, \dots, \mathbf{r}_m$ be the rows of $\mathbf{A}$, and write $\mathbf{e}_l^\top$ for the l -th row of $\mathbf{I}_m$. By Definition 3.2.1, for each k there are scalars $c_{k1}, \dots, c_{km}$, depending on ρ but **not** on $\mathbf{A}$, such that row k of $\rho(\mathbf{A})$ is $c_{k1}\mathbf{r}_1 + \dots + c_{km}\mathbf{r}_m$. Explicitly, every row that the operation does not touch has $c_{kk} = 1$ and all other $c_{kl} = 0$, and the touched rows are:

- for $R_i \leftrightarrow R_j$: row i is $\mathbf{r}_j$ and row j is $\mathbf{r}_i$;
- for $R_i \rightarrow cR_i$: row i is $c\mathbf{r}_i$;
- for $R_i \rightarrow R_i + cR_j$: row i is $\mathbf{r}_i + c\mathbf{r}_j$.

Applying ρ to $\mathbf{I}_m$, row k of $\mathbf{E}$ is $c_{k1}\mathbf{e}_1^\top + \dots + c_{km}\mathbf{e}_m^\top = (c_{k1} \ \dots \ c_{km})$. So the (k, l) -entry of $\mathbf{E}$ is c_{kl} . By Theorem 0.9.2 (rows), row k of $\mathbf{E}\mathbf{A}$ is (row k of $\mathbf{E}$) $\mathbf{A}$, and by Theorem 0.9.3 (rows version) this equals $c_{k1}\mathbf{r}_1 + \dots + c_{km}\mathbf{r}_m$, which is row k of $\rho(\mathbf{A})$. Since every row agrees, $\rho(\mathbf{A}) = \mathbf{E}\mathbf{A}$. This proves the theorem. ■

So a sequence of operations $\rho_1, \rho_2, \dots, \rho_k$, performed in that order, turns $\mathbf{A}$ into

$$\rho_k(\dots \rho_2(\rho_1(\mathbf{A})) \dots) = \mathbf{E}_k \dots \mathbf{E}_2 \mathbf{E}_1 \mathbf{A}, \quad \mathbf{E}_i = \rho_i(\mathbf{I}_m),$$

by applying the theorem k times and using associativity. The whole of Gaussian elimination is a single matrix $\mathbf{E}_k \dots \mathbf{E}_1$ standing on the left.

Warning. The **first** operation stands **rightmost**. Doing $R_1 \leftrightarrow R_2$ and then $R_1 \rightarrow R_1 + R_2$

on $\mathbf{I}_2$ gives $(\mathbf{I}_2 + \mathbf{E}_{12})\mathbf{P}_{12} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, while the product in the other order is $\mathbf{P}_{12}(\mathbf{I}_2 + \mathbf{E}_{12}) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Also, multiplying by $\mathbf{E}$ on the **right** performs a **column** operation: $\mathbf{A}(\mathbf{I}_2 + \mathbf{E}_{12})$ adds column 1 to column 2, and column operations change solution sets.

Example. Three operations, three products

Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 4 \end{pmatrix} \in M_{3 \times 2}(\mathbb{R})$. Compute $\mathbf{P}_{13}\mathbf{A}$, $\mathbf{D}_2(5)\mathbf{A}$ and $(\mathbf{I}_3 - 3\mathbf{E}_{31})\mathbf{A}$, and check that each is the result of the corresponding row operation.

Solution. By the row view, row k of $\mathbf{EA}$ is the combination of the rows of $\mathbf{A}$ whose weights are row k of $\mathbf{E}$.

- Rows of $\mathbf{P}_{13}$ are $\mathbf{e}_3^\top, \mathbf{e}_2^\top, \mathbf{e}_1^\top$, so $\mathbf{P}_{13}\mathbf{A} = \begin{pmatrix} 3 & 4 \\ 0 & 1 \\ 1 & 2 \end{pmatrix}$: rows 1 and 3 are swapped.
- Row 2 of $\mathbf{D}_2(5)$ is $5\mathbf{e}_2^\top$, so $\mathbf{D}_2(5)\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 5 \\ 3 & 4 \end{pmatrix}$: row 2 is multiplied by 5.
- Row 3 of $\mathbf{I}_3 - 3\mathbf{E}_{31}$ is $\begin{pmatrix} -3 & 0 & 1 \end{pmatrix}$, so row 3 of the product is $-3(1, 2) + (3, 4) = (0, -2)$, and $(\mathbf{I}_3 - 3\mathbf{E}_{31})\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 0 & -2 \end{pmatrix}$: this is $R_3 \rightarrow R_3 - 3R_1$.

In each case the product agrees with the operation, as [Theorem 3.4.1](#) predicts.

Quick check. Which 3×3 elementary matrix performs $R_2 \rightarrow R_2 + 4R_3$? Which operation undoes it, and what is its matrix?

Solution. Apply the operation to $\mathbf{I}_3$: row 2 becomes $\mathbf{e}_2^\top + 4\mathbf{e}_3^\top$, so the matrix is $\mathbf{I}_3 + 4\mathbf{E}_{23}$, with 4 in entry $(2, 3)$. The operation $R_2 \rightarrow R_2 - 4R_3$ undoes it, since row 3 is unchanged by both; its matrix is $\mathbf{I}_3 - 4\mathbf{E}_{23}$.

Every row operation can be undone by another row operation, and the matching statement for matrices is:

Theorem 3.4.2. Elementary matrices are invertible

Every elementary matrix is invertible, and its inverse is an elementary matrix of the same

kind:

$$\mathbf{P}_{ij}^{-1} = \mathbf{P}_{ij}, \quad \mathbf{D}_i(c)^{-1} = \mathbf{D}_i(c^{-1}), \quad (\mathbf{I}_m + c\mathbf{E}_{ij})^{-1} = \mathbf{I}_m - c\mathbf{E}_{ij} \quad (i \neq j).$$

Proof. Let $\mathbf{E} = \rho(\mathbf{I}_m)$, and let ρ' be the operation on the same line of the display: $\rho' = \rho$ for a swap, $R_i \rightarrow c^{-1}R_i$ for $R_i \rightarrow cR_i$ (legal, since $c \neq 0$ gives $c^{-1} \neq 0$), and $R_i \rightarrow R_i - cR_j$ for $R_i \rightarrow R_i + cR_j$. In each case $\rho'(\rho(\mathbf{A})) = \mathbf{A}$ and $\rho(\rho'(\mathbf{A})) = \mathbf{A}$ for every $\mathbf{A}$, as in Theorem 3.2.2 (a): swapping twice restores the order; scaling by c and c^{-1} multiplies row i by $cc^{-1} = 1$; and adding $c\mathbf{r}_j$ and then $-c\mathbf{r}_j$ to row i restores row i , because $i \neq j$ means row j is not changed in between. Let $\mathbf{E}' = \rho'(\mathbf{I}_m)$. By Theorem 3.4.1,

$$\mathbf{E}'\mathbf{E} = \rho'(\mathbf{E}) = \rho'(\rho(\mathbf{I}_m)) = \mathbf{I}_m \quad \text{and} \quad \mathbf{E}\mathbf{E}' = \rho(\mathbf{E}') = \rho(\rho'(\mathbf{I}_m)) = \mathbf{I}_m.$$

Hence $\mathbf{E}$ is invertible with $\mathbf{E}^{-1} = \mathbf{E}'$, which is the matrix in the display. This proves the theorem. ■

The hypothesis $i \neq j$ did real work: “add c times row i to row i ” is scaling by $1 + c$, which destroys row i when $c = -1$ and cannot be undone. That is why Definition 3.2.1 excludes it.

3.4.2 The invertible matrix theorem

Chapter 0 defined an invertible matrix by the existence of a two-sided inverse, and gave only one practical test, $ad - bc \neq 0$, for 2×2 matrices (Theorem 0.9.8). For larger matrices, searching for $\mathbf{B}$ with $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$ directly is hopeless. What we want is a list of conditions, each of which we can check by a method we already have, that are all equivalent to invertibility. Then we may test whichever is cheapest.

We need one small fact about square matrices in reduced form.

Lemma 3.4.3. Square reduced matrices

Let $\mathbf{R} \in M_n(F)$ be in reduced row echelon form. Then either $\mathbf{R} = \mathbf{I}_n$, or the last row of $\mathbf{R}$ is zero.

Proof. Suppose the last row of $\mathbf{R}$ is not zero. Since zero rows of a matrix in echelon form come after all nonzero rows (Definition 3.2.3), no row of $\mathbf{R}$ is zero, so each of the n rows has a pivot. Let the pivot of row k be in column c_k . Pivots move strictly to the right as we go down, so $1 \leq c_1 < c_2 < \dots < c_n \leq n$. A strictly increasing list of n integers in $\{1, \dots, n\}$ must be $1, 2, \dots, n$, so $c_k = k$. By Definition 3.2.4, each pivot equals 1 and is the only nonzero entry of its column, so column k of $\mathbf{R}$ is $\mathbf{e}_k$ for every k . Hence $\mathbf{R} = \mathbf{I}_n$. ■

Here is the theorem. It collects, for a square matrix, facts about solving systems, about row reduction, and about the columns as vectors in the sense of Chapter 1.

Theorem 3.4.4. Invertible Matrix Theorem

Let $\mathbf{A} \in M_n(F)$, with columns $\mathbf{a}_1, \dots, \mathbf{a}_n \in F^n$. The following are equivalent.

- (a) $\mathbf{A}$ is invertible.
- (b) The homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.
- (c) The reduced row echelon form of $\mathbf{A}$ is $\mathbf{I}_n$.
- (d) $\mathbf{A}$ is a product of elementary matrices.
- (e) For **every** $\mathbf{b} \in F^n$, the system $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution.
- (f) The columns $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ are linearly independent.
- (g) The columns $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ span F^n .
- (h) The columns $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ form a basis of F^n .

Idea. Eight conditions would need 56 implications if proved pairwise; a cycle through them needs far fewer. The plan:

- ① the core cycle $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (a)$, where each arrow uses one earlier tool (cancel by $\mathbf{A}^{-1}$; row reduction; elementary matrices; products of invertible matrices);
- ② $(b) \Leftrightarrow (f)$, which is the column view $\mathbf{A}\mathbf{x} = x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n$;
- ③ $(f) \Leftrightarrow (h) \Leftrightarrow (g)$, which is “count instead of check” in the n -dimensional space F^n ;
- ④ $(g) \Leftrightarrow (e)$, which is “consistent means $\mathbf{b}$ is in the column span”.

The only arrow that needs a new argument is $(b) \Rightarrow (c)$; the others quote a result we already have. An RREF that is not $\mathbf{I}_n$ has a zero row, so it really has fewer equations than unknowns, and such a homogeneous system has a nontrivial solution.

Proof. **(a) $\Rightarrow$ (b).** Suppose $\mathbf{A}$ is invertible and $\mathbf{A}\mathbf{x} = \mathbf{0}$. Multiplying on the left by $\mathbf{A}^{-1}$ and using Theorem 0.9.4, $\mathbf{x} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{x} = \mathbf{A}^{-1}(\mathbf{A}\mathbf{x}) = \mathbf{A}^{-1}\mathbf{0} = \mathbf{0}$.

(b) $\Rightarrow$ (c). Suppose (b) holds. By Theorem 3.2.3, $\mathbf{A}$ is row equivalent to a matrix $\mathbf{R}$ in reduced row echelon form, and $\mathbf{R}$ is the RREF of $\mathbf{A}$ by Theorem 3.3.4. By Lemma 3.3.1, $\mathbf{R}\mathbf{x} = \mathbf{0}$ has the same solutions as $\mathbf{A}\mathbf{x} = \mathbf{0}$, so it too has only the trivial solution. Suppose, for a contradiction, that $\mathbf{R} \neq \mathbf{I}_n$. By Lemma 3.4.3, the last row of $\mathbf{R}$ is zero. If $n = 1$, then $\mathbf{R} = \mathbf{O}_1$ and $\mathbf{x} = (1)$ is a nontrivial solution. If $n \geq 2$, the last equation of $\mathbf{R}\mathbf{x} = \mathbf{0}$ reads $0 = 0$, so $\mathbf{R}\mathbf{x} = \mathbf{0}$ has the same solutions as the homogeneous system $\mathbf{R}'\mathbf{x} = \mathbf{0}$ formed by the first $n - 1$ rows. That system has $n - 1$ equations in $n > n - 1$ unknowns, so by Corollary 3.2.5 it has a nontrivial solution. Either way we contradict (b). Hence $\mathbf{R} = \mathbf{I}_n$.

(c) $\Rightarrow$ (d). Suppose the RREF of $\mathbf{A}$ is $\mathbf{I}_n$. By Definition 3.2.2 there are row operations $\rho_1, \dots, \rho_k$ taking $\mathbf{A}$ to $\mathbf{I}_n$. If $k = 0$, then $\mathbf{A} = \mathbf{I}_n = \mathbf{D}_1(1)$ is elementary. Otherwise, with $\mathbf{E}_i = \rho_i(\mathbf{I}_n)$, Theorem 3.4.1 gives $\mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A} = \mathbf{I}_n$. Each $\mathbf{E}_i$ is invertible by Theorem 3.4.2, so multiplying on the left by $\mathbf{E}_k^{-1}$, then $\mathbf{E}_{k-1}^{-1}$, and so on, gives

$$\mathbf{A} = \mathbf{E}_1^{-1} \mathbf{E}_2^{-1} \cdots \mathbf{E}_k^{-1},$$

and each $\mathbf{E}_i^{-1}$ is elementary by Theorem 3.4.2.

(d) $\Rightarrow$ (a). Suppose $\mathbf{A} = \mathbf{E}_1 \cdots \mathbf{E}_k$ with each $\mathbf{E}_i$ elementary and $k \geq 1$. Each $\mathbf{E}_i$ is invertible by Theorem 3.4.2, and a product of two invertible matrices is invertible by Theorem 0.9.7 (part 3). By induction on k , $\mathbf{A}$ is invertible.

So far (a), (b), (c), (d) are equivalent.

(b) $\Leftrightarrow$ (f). By Theorem 0.9.3, $\mathbf{A}\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n$. So “ $\mathbf{A}\mathbf{x} = \mathbf{0}$ only for $\mathbf{x} = \mathbf{0}$ ” says exactly that $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$ forces $x_1 = \cdots = x_n = 0$, which is Definition 1.4.1 for the columns.

(f) $\Leftrightarrow$ (h) $\Leftrightarrow$ (g). The columns form a list of exactly n vectors in F^n , and $\dim F^n = n$ (Example (Dimensions of the Running Examples)). By Theorem 1.6.4, such a list is a basis if it is linearly independent, and also if it spans. Conversely a basis is independent and spanning by Definition 1.5.1. Hence (f) $\Rightarrow$ (h) $\Rightarrow$ (g) $\Rightarrow$ (h) $\Rightarrow$ (f).

(g) $\Leftrightarrow$ (e). By Theorem 3.1.1, $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b} \in \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$. So (e) says every $\mathbf{b} \in F^n$ lies in this span, that is, $F^n \subseteq \text{span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$. The reverse inclusion always holds, so (e) is equivalent to (g).

Combining the four parts, all eight conditions are equivalent. This proves the theorem. ■

In practice the theorem is a menu. To show a matrix **is** invertible, pick the most convenient item, very often (b): take a solution of $\mathbf{A}\mathbf{x} = \mathbf{0}$ and show it is $\mathbf{0}$. To show a matrix is **not** invertible, exhibit one nonzero solution of $\mathbf{A}\mathbf{x} = \mathbf{0}$, or one linear relation among the columns. The list grows in both directions. Chapter 2 has already given the operator form of it, in terms of kernels and images (Theorem 2.5.7), and later chapters add determinants in Chapter 7, eigenvalues in Chapter 9, and positive definiteness and singular values in Chapter 13.

Quick check. Is $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{pmatrix} \in M_3(\mathbb{R})$ invertible? Use whichever item of Theorem 3.4.4 is quickest.

Solution. No. The third column is the sum of the first two: $(1, 0, 1) + (2, 1, 3) = (3, 1, 4)$. So $\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_3 = \mathbf{0}$ is a linear relation with coefficients not all zero, item (f) fails, and hence (a) fails. Equivalently, $\mathbf{x} = (1, 1, -1)$ is a nontrivial solution of $\mathbf{A}\mathbf{x} = \mathbf{0}$.

Warning. The theorem is about **square** matrices. For $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \in M_{2 \times 3}(\mathbb{R})$, the columns span $\mathbb{R}^2$ (the first two are $\mathbf{e}_1, \mathbf{e}_2$), so (g) holds, but $\mathbf{x} = (1, 1, -1)$ solves $\mathbf{A}\mathbf{x} = \mathbf{0}$, so (b) fails. The proof of (f) $\Leftrightarrow$ (g) used that the number of columns equals $\dim F^n$, and that is exactly what breaks.

3.4.3 One-sided inverses are two-sided

Chapter 0 left a gap on purpose. The definition of invertibility (Definition 0.9.13) asks for **both** $\mathbf{AB} = \mathbf{I}_n$ and $\mathbf{BA} = \mathbf{I}_n$. Chapter 0 then stated that, for square matrices, the first equation already implies the second, but that nothing there could prove it: the entries of $\mathbf{AB}$ and $\mathbf{BA}$ are different sums. Now we can prove it, and the proof uses the menu above instead of entries.

Theorem 3.4.5. One-sided inverses of square matrices

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$. If $\mathbf{AB} = \mathbf{I}_n$, then $\mathbf{BA} = \mathbf{I}_n$. Hence $\mathbf{A}$ and $\mathbf{B}$ are both invertible, $\mathbf{B} = \mathbf{A}^{-1}$ and $\mathbf{A} = \mathbf{B}^{-1}$.

Idea. We cannot touch $\mathbf{BA}$ directly, so we aim at invertibility of one factor instead. Which one? From $\mathbf{AB} = \mathbf{I}_n$, a vector killed by $\mathbf{B}$ is killed by $\mathbf{AB} = \mathbf{I}_n$, so it is $\mathbf{0}$. That is item (b) for $\mathbf{B}$. Once $\mathbf{B}^{-1}$ exists, we may cancel it: $\mathbf{A} = (\mathbf{AB})\mathbf{B}^{-1} = \mathbf{B}^{-1}$.

Proof. Suppose $\mathbf{AB} = \mathbf{I}_n$, and let $\mathbf{x} \in F^n$ with $\mathbf{Bx} = \mathbf{0}$. Then, by Theorem 0.9.4,

$$\mathbf{x} = \mathbf{I}_n \mathbf{x} = (\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{Bx}) = \mathbf{A}\mathbf{0} = \mathbf{0}.$$

So $\mathbf{Bx} = \mathbf{0}$ has only the trivial solution, and $\mathbf{B}$ is invertible by Theorem 3.4.4 ((b) $\Rightarrow$ (a)). Multiplying $\mathbf{AB} = \mathbf{I}_n$ on the right by $\mathbf{B}^{-1}$ and using associativity,

$$\mathbf{A} = \mathbf{A}(\mathbf{BB}^{-1}) = (\mathbf{AB})\mathbf{B}^{-1} = \mathbf{I}_n \mathbf{B}^{-1} = \mathbf{B}^{-1}.$$

Therefore $\mathbf{BA} = \mathbf{BB}^{-1} = \mathbf{I}_n$. Since $\mathbf{A} = \mathbf{B}^{-1}$ is invertible with inverse $\mathbf{B}$ by Theorem 0.9.7 (part 2), we get $\mathbf{A}^{-1} = \mathbf{B}$. This proves the theorem. ■

By symmetry, $\mathbf{BA} = \mathbf{I}_n$ also implies $\mathbf{AB} = \mathbf{I}_n$: apply the theorem with the roles of $\mathbf{A}$ and $\mathbf{B}$ swapped. So from now on, **to verify that $\mathbf{B}$ is the inverse of a square matrix $\mathbf{A}$, one product suffices**. We also get a useful consequence of the same kind: if $\mathbf{A}, \mathbf{B} \in M_n(F)$ and $\mathbf{AB}$ is invertible, then $\mathbf{A}$ and $\mathbf{B}$ are invertible (Exercise B3 at the end of this section).

Warning. For **nonsquare** matrices a one-sided inverse need not be two-sided. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then $\mathbf{AB} = \mathbf{I}_2$, but $\mathbf{BA} = \text{diag}(1, 1, 0) \neq \mathbf{I}_3$. In fact no $\mathbf{C} \in M_{3 \times 2}(F)$ has $\mathbf{CA} = \mathbf{I}_3$: $\mathbf{Ae}_3 = \mathbf{0}$, so $\mathbf{CAe}_3 = \mathbf{0} \neq \mathbf{e}_3 = \mathbf{I}_3 \mathbf{e}_3$. The proof above breaks because $\mathbf{Bx} = \mathbf{0} \Rightarrow \mathbf{x} = \mathbf{0}$ no longer gives invertibility: $\mathbf{B}$ is not square, so Theorem 3.4.4 does not apply.

3.4.4 Computing the inverse

Theorem 3.4.4 (c) says $\mathbf{A}$ is invertible exactly when it row reduces to $\mathbf{I}_n$. The proof of (c) $\Rightarrow$ (d) even produced a matrix: if $\mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A} = \mathbf{I}_n$, then $\mathbf{E}_k \cdots \mathbf{E}_1$ is a left inverse, hence **the** inverse by **Theorem 3.4.5**. We do not want to multiply out the $\mathbf{E}_i$ by hand. The trick is to let the operations act on $\mathbf{I}_n$ at the same time as on $\mathbf{A}$: since $\mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{I}_n = \mathbf{E}_k \cdots \mathbf{E}_1$, the operations that turn $\mathbf{A}$ into $\mathbf{I}_n$ turn $\mathbf{I}_n$ into $\mathbf{A}^{-1}$. We write $[\mathbf{A} \mid \mathbf{I}_n]$ for the $n \times 2n$ matrix whose first n columns are those of $\mathbf{A}$ and whose last n columns are those of $\mathbf{I}_n$.

Algorithm. **Input:** $\mathbf{A} \in M_n(F)$.

1. Form the $n \times 2n$ matrix $[\mathbf{A} \mid \mathbf{I}_n]$.
2. Row reduce it to reduced row echelon form $[\mathbf{R} \mid \mathbf{C}]$, with $\mathbf{R}, \mathbf{C} \in M_n(F)$.
3. If $\mathbf{R} = \mathbf{I}_n$, output $\mathbf{A}^{-1} = \mathbf{C}$. Otherwise, output “ $\mathbf{A}$ is not invertible”.

In practice, stop as soon as a zero row appears in the left half: $\mathbf{A}$ is then not invertible (see the remark after the theorem).

Theorem 3.4.6. Inverse by row reduction

Let $\mathbf{A} \in M_n(F)$, and let $[\mathbf{R} \mid \mathbf{C}]$ be the reduced row echelon form of $[\mathbf{A} \mid \mathbf{I}_n]$, with $\mathbf{R}, \mathbf{C} \in M_n(F)$. Then $\mathbf{R}$ is the reduced row echelon form of $\mathbf{A}$. Consequently, $\mathbf{A}$ is invertible if and only if $\mathbf{R} = \mathbf{I}_n$, and in that case $\mathbf{A}^{-1} = \mathbf{C}$.

Idea. Two things need checking. First, the operations act on the two halves independently, so the left half is transformed exactly as $\mathbf{A}$ alone would be. Second, the left half of a reduced matrix is itself reduced, so it really is the RREF of $\mathbf{A}$. The first is the column view of a product; the second is a small check on the staircase shape. After that, $\mathbf{R} = \mathbf{E}\mathbf{A}$ and $\mathbf{C} = \mathbf{E}\mathbf{I}_n = \mathbf{E}$, and **Theorem 3.4.5** finishes.

Proof. Let $\rho_1, \dots, \rho_k$ be row operations taking $[\mathbf{A} \mid \mathbf{I}_n]$ to $[\mathbf{R} \mid \mathbf{C}]$, and put $\mathbf{E} = \mathbf{E}_k \cdots \mathbf{E}_1$ with $\mathbf{E}_i = \rho_i(\mathbf{I}_n)$ (and $\mathbf{E} = \mathbf{I}_n$ if $k = 0$). By **Theorem 3.4.1**, $[\mathbf{R} \mid \mathbf{C}] = \mathbf{E}[\mathbf{A} \mid \mathbf{I}_n]$. By **Theorem 0.9.2** (columns), each column of $\mathbf{E}[\mathbf{A} \mid \mathbf{I}_n]$ is $\mathbf{E}$ times the corresponding column, so

$$\mathbf{R} = \mathbf{E}\mathbf{A} \quad \text{and} \quad \mathbf{C} = \mathbf{E}\mathbf{I}_n = \mathbf{E}.$$

Claim. $\mathbf{R}$ is in reduced row echelon form.

Proof. If row k of $\mathbf{R}$ is nonzero, its first nonzero entry is also the first nonzero entry of row k of $[\mathbf{R} \mid \mathbf{C}]$, so the pivots of the nonzero rows of $\mathbf{R}$ are pivots of $[\mathbf{R} \mid \mathbf{C}]$, in the same positions. Suppose row k of $\mathbf{R}$ is zero but some later row $l > k$ of $\mathbf{R}$ is nonzero. Row k of $[\mathbf{R} \mid \mathbf{C}]$ cannot be zero, since zero rows of $[\mathbf{R} \mid \mathbf{C}]$ come after its nonzero rows and row l is nonzero. So its pivot lies in a column $> n$, while the pivot of row l lies in a column $\leq n$, to the left: this contradicts the staircase condition of Definition 3.2.3. Hence the zero rows of $\mathbf{R}$ come last, the pivots of $\mathbf{R}$ move strictly right, and each pivot of $\mathbf{R}$ is 1 and the only nonzero entry in its column, because this holds in $[\mathbf{R} \mid \mathbf{C}]$. So $\mathbf{R}$ satisfies Definition 3.2.4. ■

Now $\mathbf{R} = \mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A}$ is obtained from $\mathbf{A}$ by the operations $\rho_1, \dots, \rho_k$ (Theorem 3.4.1), so $\mathbf{A}$ is row equivalent to $\mathbf{R}$, and $\mathbf{R}$ is in reduced form. By Theorem 3.3.4, $\mathbf{R}$ is the RREF of $\mathbf{A}$. By Theorem 3.4.4 ((a) $\Leftrightarrow$ (c)), $\mathbf{A}$ is invertible if and only if $\mathbf{R} = \mathbf{I}_n$. If $\mathbf{R} = \mathbf{I}_n$, then $\mathbf{E}\mathbf{A} = \mathbf{I}_n$, and Theorem 3.4.5 gives $\mathbf{A}^{-1} = \mathbf{E} = \mathbf{C}$. This proves the theorem. ■

Remark. The early stop is justified the same way. If, after some operations, $[\mathbf{A} \mid \mathbf{I}_n]$ has become $[\mathbf{A}' \mid \mathbf{E}']$ with a zero row in $\mathbf{A}'$, then, exactly as in the proof, $\mathbf{E}'$ is the product of the elementary matrices of the operations used so far and $\mathbf{A}' = \mathbf{E}'\mathbf{A}$. So $\mathbf{E}'$ is invertible by Theorem 3.4.2 and Theorem 0.9.7 (part 3). A matrix with a zero row k is not invertible, since row k of $\mathbf{A}'\mathbf{X}$ is zero for every $\mathbf{X}$, so $\mathbf{A}'\mathbf{X} \neq \mathbf{I}_n$. If $\mathbf{A}$ were invertible, $\mathbf{A}' = \mathbf{E}'\mathbf{A}$ would be invertible by Theorem 0.9.7 (part 3). Hence $\mathbf{A}$ is not invertible.

Example. $\mathbf{A}$ 3×3 inverse

Find the inverse of $\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 3 \\ 1 & 3 & 3 \end{pmatrix} \in M_3(\mathbb{R})$, or show that it does not exist.

Solution. Row reduce $[\mathbf{A} \mid \mathbf{I}_3]$, naming each operation.

$$\begin{aligned} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 3 & 3 & 0 & 0 & 1 \end{array} \right] & \xrightarrow[\substack{R_2 \rightarrow R_2 - 2R_1 \\ R_3 \rightarrow R_3 - R_1}]{} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 1 & 2 & -1 & 0 & 1 \end{array} \right] \\ & \xrightarrow{R_3 \rightarrow R_3 - R_2} \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array} \right] \end{aligned}$$

The left half now has three pivots, so no zero row will appear. Clear above the pivots, from the bottom up:

$$\xrightarrow[\substack{R_2 \rightarrow R_2 - R_3 \\ R_1 \rightarrow R_1 - R_3}]{} \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & -3 & 2 & -1 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array} \right] \xrightarrow{R_1 \rightarrow R_1 - 2R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 6 & -3 & 1 \\ 0 & 1 & 0 & -3 & 2 & -1 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{array} \right].$$

The left half is $\mathbf{I}_3$, so by Theorem 3.4.6 $\mathbf{A}$ is invertible and

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

Check one product: row 1 of $\mathbf{A}$ times the columns of $\mathbf{A}^{-1}$ gives $6 - 6 + 1 = 1$, $-3 + 4 - 1 = 0$, $1 - 2 + 1 = 0$; rows 2 and 3 give $(0, 1, 0)$ and $(0, 0, 1)$ in the same way. So $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_3$, and by Theorem 3.4.5 this single product is enough.

Example. When the algorithm stops

Decide whether $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 1 & 3 & 4 \end{pmatrix} \in M_3(\mathbb{R})$ is invertible.

Solution. Row reduce $[\mathbf{A} \mid \mathbf{I}_3]$:

$$\begin{aligned} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 7 & 0 & 1 & 0 \\ 1 & 3 & 4 & 0 & 0 & 1 \end{array} \right] &\xrightarrow[\substack{R_2 \rightarrow R_2 - 2R_1 \\ R_3 \rightarrow R_3 - R_1}]{} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 1 & 1 & -1 & 0 & 1 \end{array} \right] \\ &\xrightarrow{R_3 \rightarrow R_3 - R_2} \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{array} \right]. \end{aligned}$$

A zero row has appeared in the left half, so by the remark after [Theorem 3.4.6](#), $\mathbf{A}$ is not invertible. The right half of that row is not wasted: it records the combination of rows that produced zero. Row 3 of the left half is $1 \cdot (\text{row } 1) - 1 \cdot (\text{row } 2) + 1 \cdot (\text{row } 3)$ of $\mathbf{A}$, and indeed $(1, 2, 3) - (2, 5, 7) + (1, 3, 4) = (0, 0, 0)$.

Finally, the method writes an invertible matrix as a product of elementary matrices, as in [Theorem 3.4.4](#) (d).

Example. Factoring into elementary matrices

Write $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \in M_2(\mathbb{R})$ as a product of elementary matrices, and read off $\mathbf{A}^{-1}$.

Solution. Reduce $\mathbf{A}$ to $\mathbf{I}_2$, keeping integers:

$$\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \xrightarrow{R_2 \rightarrow R_2 - 2R_1} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \xrightarrow{R_1 \rightarrow R_1 - R_2} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \xrightarrow{R_2 \rightarrow R_2 - R_1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The elementary matrices of these operations are $\mathbf{E}_1 = \mathbf{I}_2 - 2\mathbf{E}_{21}$, $\mathbf{E}_2 = \mathbf{I}_2 - \mathbf{E}_{12}$, $\mathbf{E}_3 = \mathbf{I}_2 - \mathbf{E}_{21}$, and by [Theorem 3.4.1](#), $\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{I}_2$. Multiplying on the left by $\mathbf{E}_3^{-1}$, then $\mathbf{E}_2^{-1}$, then $\mathbf{E}_1^{-1}$, and using [Theorem 3.4.2](#),

$$\mathbf{A} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1}\mathbf{E}_3^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Multiplying out confirms it: the last two factors give $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$, and then the first factor adds twice row 1 to row 2, giving $\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}$. Also $\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1$ is a left inverse of $\mathbf{A}$, hence $\mathbf{A}^{-1}$ by

Theorem 3.4.5:

$$\mathbf{A}^{-1} = \mathbf{E}_3 \mathbf{E}_2 \mathbf{E}_1 = \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix},$$

which agrees with Theorem 0.9.8, since $ad - bc = 1$. The factorization is not unique: a different sequence of operations gives different factors.

3.4.5 Permutation matrices

A row swap is one elementary matrix. Doing several of them in a row is not, but the product is still a matrix of a single recognizable kind, and it is worth naming here because it turns up far beyond this chapter: it is how a relabeling of coordinates becomes a matrix.

Definition 3.4.2. Permutation matrix

Let $\sigma \in S_n$ be a permutation of $\{1, \dots, n\}$. The **permutation matrix** of σ is

$$\mathbf{P}_\sigma = \begin{pmatrix} \mathbf{e}_{\sigma(1)} & \mathbf{e}_{\sigma(2)} & \cdots & \mathbf{e}_{\sigma(n)} \end{pmatrix} \in M_n(F),$$

the matrix whose j -th column is $\mathbf{e}_{\sigma(j)}$. A matrix is a **permutation matrix** if it equals $\mathbf{P}_\sigma$ for some $\sigma \in S_n$.

Equivalently, a permutation matrix has exactly one 1 in each row and in each column, and zeros elsewhere. The identity is $\mathbf{P}_{\text{id}}$, and $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is $\mathbf{P}_\tau$ for the transposition $\tau = (1\ 2)$, which is also the elementary matrix $\mathbf{P}_{12}$. By contrast $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is **not** a permutation matrix: each column is still nonzero, but its first row contains two 1's, and its second column $(1, 1)$ is not a standard basis vector.

Lemma 3.4.7. Permutation matrices

Let $\sigma, \tau \in S_n$.

- (a) $\mathbf{P}_\sigma \mathbf{P}_\tau = \mathbf{P}_{\sigma \circ \tau}$. In particular a product of permutation matrices is a permutation matrix.
- (b) $\mathbf{P}_\sigma$ is invertible and $\mathbf{P}_\sigma^{-1} = \mathbf{P}_\sigma^\top$.
- (c) If $\tau = (r\ s)$ is a transposition, then $\mathbf{P}_\tau$ is the elementary matrix $\mathbf{P}_{rs}$ that swaps rows r and s , and $\mathbf{P}_\tau^2 = \mathbf{I}$.

Proof.

- (a) By Theorem 0.9.3, $\mathbf{P}_\tau \mathbf{e}_j = \mathbf{e}_{\tau(j)}$ and so $\mathbf{P}_\sigma \mathbf{P}_\tau \mathbf{e}_j = \mathbf{P}_\sigma \mathbf{e}_{\tau(j)} = \mathbf{e}_{\sigma(\tau(j))}$. So the j -th column of $\mathbf{P}_\sigma \mathbf{P}_\tau$ is $\mathbf{e}_{(\sigma \circ \tau)(j)}$, for every j , which is the claim.
- (b) By Theorem 0.9.2, $(\mathbf{P}_\sigma^\top \mathbf{P}_\sigma)_{ij} = \mathbf{e}_{\sigma(i)}^\top \mathbf{e}_{\sigma(j)}$, which is 1 if $\sigma(i) = \sigma(j)$ and 0 otherwise. Since σ is injective, $\sigma(i) = \sigma(j)$ exactly when $i = j$. Hence $\mathbf{P}_\sigma^\top \mathbf{P}_\sigma = \mathbf{I}$, and by Theorem 3.4.5, $\mathbf{P}_\sigma$ is invertible with inverse $\mathbf{P}_\sigma^\top$.
- (c) Swapping rows r and s of $\mathbf{I}$ produces a matrix whose column r is $\mathbf{e}_s$, whose column s is $\mathbf{e}_r$, and whose other columns j are $\mathbf{e}_j$. These are the columns $\mathbf{e}_{\tau(j)}$, so the swap

matrix is $\mathbf{P}_\tau$. Since $\tau \circ \tau = \text{id}$, part (a) gives $\mathbf{P}_\tau^2 = \mathbf{P}_{\text{id}} = \mathbf{I}$. This proves the lemma. ■

By (c) and Theorem 3.4.1, $\mathbf{P}_\tau \mathbf{A}$ is $\mathbf{A}$ with rows r and s swapped, and by (a) every permutation matrix is a product of such swaps applied in turn, since for $n \geq 2$ every permutation is a product of transpositions (Theorem 0.10.4). So $\mathbf{P}_\sigma \mathbf{A}$ is always $\mathbf{A}$ with its rows rearranged.

3.4.6 Row equivalence, revisited

Two descriptions of “row equivalent” are now on the table. Example (Row equivalence) of Chapter 0 said $\mathbf{B} = \mathbf{EA}$ for some invertible $\mathbf{E}$. Definition 3.2.2 says $\mathbf{B}$ is reachable from $\mathbf{A}$ by finitely many row operations. Chapter 0 promised that they agree, and now it is one line each way.

Theorem 3.4.8. Row equivalence is left multiplication by an invertible matrix

Let $\mathbf{A}, \mathbf{B} \in M_{m \times n}(F)$. Then $\mathbf{A}$ is row equivalent to $\mathbf{B}$ (that is, $\mathbf{B}$ can be reached from $\mathbf{A}$ by row operations) if and only if $\mathbf{B} = \mathbf{EA}$ for some invertible $\mathbf{E} \in M_m(F)$.

Proof. ($\Rightarrow$) Suppose row operations $\rho_1, \dots, \rho_k$ take $\mathbf{A}$ to $\mathbf{B}$. By Theorem 3.4.1, $\mathbf{B} = \mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A}$ with $\mathbf{E}_i = \rho_i(\mathbf{I}_m)$, and $\mathbf{E} = \mathbf{E}_k \cdots \mathbf{E}_1$ is invertible by Theorem 3.4.2 and Theorem 0.9.7 (part 3). (If $k = 0$, take $\mathbf{E} = \mathbf{I}_m$.)

($\Leftarrow$) Suppose $\mathbf{B} = \mathbf{EA}$ with $\mathbf{E}$ invertible. By Theorem 3.4.4 ((a) $\Rightarrow$ (d)), $\mathbf{E} = \mathbf{E}_k \cdots \mathbf{E}_1$ for some elementary matrices $\mathbf{E}_i$. Let ρ_i be the operation with $\rho_i(\mathbf{I}_m) = \mathbf{E}_i$. By Theorem 3.4.1, performing ρ_1 , then ρ_2 , ..., then ρ_k on $\mathbf{A}$ produces $\mathbf{E}_k \cdots \mathbf{E}_1 \mathbf{A} = \mathbf{B}$. Hence $\mathbf{A}$ is row equivalent to $\mathbf{B}$. ■

So the equivalence relation of Example (Row equivalence) is the one Gaussian elimination moves along, and “two matrices have the same RREF” is the same as “each is an invertible matrix times the other”. In the next section this form, $\mathbf{B} = \mathbf{EA}$, is what makes row spaces and ranks easy to compare.

3.4.7 Exercises

3.4.7.1 A. Check your understanding

Exercise. A1

- Define an elementary matrix, and write down the 3×3 elementary matrix that performs $R_3 \rightarrow R_3 + 7R_1$.
- State four of the conditions in Theorem 3.4.4.
- True or false: the product of two elementary matrices is always an elementary matrix. Justify your answer.
- True or false: if $\mathbf{A} \in M_{2 \times 3}(F)$, $\mathbf{B} \in M_{3 \times 2}(F)$ and $\mathbf{AB} = \mathbf{I}_2$, then $\mathbf{BA} = \mathbf{I}_3$. Justify your answer.

- (e) True or false: if $\mathbf{A} \in M_n(F)$ and $\mathbf{A}\mathbf{x} = \mathbf{e}_j$ has a solution for each $j = 1, \dots, n$, then $\mathbf{A}$ is invertible. Justify your answer.
- (f) Describe the method for computing the inverse of a square matrix, and how it detects a matrix that is not invertible.

Solution.

- (a) An elementary matrix in $M_m(F)$ is one obtained from $\mathbf{I}_m$ by exactly one elementary row operation (Definition 3.4.1). Applying $R_3 \rightarrow R_3 + 7R_1$ to $\mathbf{I}_3$ gives $\mathbf{I}_3 + 7\mathbf{E}_{31} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 7 & 0 & 1 \end{pmatrix}$.
- (b) For $\mathbf{A} \in M_n(F)$, each of the following is equivalent to $\mathbf{A}$ being invertible: $\mathbf{A}\mathbf{x} = \mathbf{0}$ has only the trivial solution; the RREF of $\mathbf{A}$ is $\mathbf{I}_n$; $\mathbf{A}$ is a product of elementary matrices; the columns of $\mathbf{A}$ form a basis of F^n .
- (c) False. A single row operation changes at most two rows of $\mathbf{I}_3$ (a swap changes two, a scaling or an addition changes one), so every elementary 3×3 matrix has at least one row equal to the corresponding row of $\mathbf{I}_3$. But $\mathbf{P}_{12}\mathbf{P}_{23} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ (by the row view, its rows are rows 2, 1, 3 of $\mathbf{P}_{23}$) differs from $\mathbf{I}_3$ in every row. Hence it is a product of two elementary matrices that is not elementary.
- (d) False. With $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$, $\mathbf{AB} = \mathbf{I}_2$ but $\mathbf{BA} = \text{diag}(1, 1, 0) \neq \mathbf{I}_3$.
Theorem 3.4.5 needs square matrices.
- (e) True. Let $\mathbf{x}_j$ solve $\mathbf{A}\mathbf{x}_j = \mathbf{e}_j$, and let $\mathbf{B}$ have columns $\mathbf{x}_1, \dots, \mathbf{x}_n$. By Theorem 0.9.2, $\mathbf{AB}$ has columns $\mathbf{e}_1, \dots, \mathbf{e}_n$, so $\mathbf{AB} = \mathbf{I}_n$, and $\mathbf{A}$ is invertible by Theorem 3.4.5.
- (f) Row reduce $[\mathbf{A} \mid \mathbf{I}_n]$ to reduced form $[\mathbf{R} \mid \mathbf{C}]$. If $\mathbf{R} = \mathbf{I}_n$, then $\mathbf{A}^{-1} = \mathbf{C}$; if $\mathbf{R} \neq \mathbf{I}_n$, in particular if a zero row appears in the left half at any stage, $\mathbf{A}$ is not invertible (Theorem 3.4.6).

3.4.7.2 B. Practice

Exercise. B1: Inverses by row reduction

For each matrix, find the inverse or show that it is not invertible.

- (a) $\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 3 \\ 4 & 1 & 8 \end{pmatrix} \in M_3(\mathbb{R})$.
- (b) $\mathbf{B} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 2 & 5 & -1 \end{pmatrix} \in M_3(\mathbb{R})$.

$$(c) \mathbf{C} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \in M_3(\mathbb{F}_2).$$

Solution.

(a) Row reduce $[\mathbf{A} \mid \mathbf{I}_3]$:

$$\begin{aligned} \left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 2 & 1 & 3 & 0 & 1 & 0 \\ 4 & 1 & 8 & 0 & 0 & 1 \end{array} \right] &\xrightarrow{\substack{R_2 \rightarrow R_2 - 2R_1 \\ R_3 \rightarrow R_3 - 4R_1}} \left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{array} \right] \\ &\xrightarrow{R_3 \rightarrow R_3 - R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & -1 & 1 \end{array} \right] \\ &\xrightarrow{\substack{R_2 \rightarrow R_2 + R_3 \\ R_1 \rightarrow R_1 - 2R_3}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 5 & 2 & -2 \\ 0 & 1 & 0 & -4 & 0 & 1 \\ 0 & 0 & 1 & -2 & -1 & 1 \end{array} \right]. \end{aligned}$$

The left half is $\mathbf{I}_3$, so by [Theorem 3.4.6](#), $\mathbf{A}^{-1} = \begin{pmatrix} 5 & 2 & -2 \\ -4 & 0 & 1 \\ -2 & -1 & 1 \end{pmatrix}$. Check: row 1 of $\mathbf{A}$

against the columns of $\mathbf{A}^{-1}$ gives $5 - 4 = 1$, $2 - 2 = 0$, $-2 + 2 = 0$, and rows 2 and 3 give $(0, 1, 0)$ and $(0, 0, 1)$, so $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_3$, which suffices by [Theorem 3.4.5](#).

(b) Row reduce $[\mathbf{B} \mid \mathbf{I}_3]$:

$$\begin{aligned} \left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 2 & 5 & -1 & 0 & 0 & 1 \end{array} \right] &\xrightarrow{R_3 \rightarrow R_3 - 2R_1} \left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & -2 & 0 & 1 \end{array} \right] \\ &\xrightarrow{R_3 \rightarrow R_3 - R_2} \left[\begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -2 & -1 & 1 \end{array} \right]. \end{aligned}$$

A zero row appears in the left half, so $\mathbf{B}$ is not invertible (remark after [Theorem 3.4.6](#)).

The right half of that row says $-2(\text{row } 1) - (\text{row } 2) + (\text{row } 3) = \mathbf{0}$, that is, row 3 of $\mathbf{B}$ is $2(\text{row } 1) + (\text{row } 2)$.

(c) Over $\mathbb{F}_2$ we have $-1 = 1$, so subtracting a row is adding it. Row reduce $[\mathbf{C} \mid \mathbf{I}_3]$:

$$\begin{aligned} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{array} \right] &\xrightarrow{R_3 \rightarrow R_3 + R_1} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array} \right] \\ &\xrightarrow{R_2 \rightarrow R_2 + R_3} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array} \right] \\ &\xrightarrow{R_1 \rightarrow R_1 + R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{array} \right]. \end{aligned}$$

Hence $\mathbf{C}^{-1} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$. Check in $\mathbb{F}_2$: row 1 of $\mathbf{C}$ is $(1, 1, 0)$, and against the columns of $\mathbf{C}^{-1}$ it gives $0 + 1 = 1$, $1 + 1 = 0$, $1 + 1 = 0$; rows 2 and 3 give $(0, 1, 0)$ and $(0, 0, 1)$ similarly.

Exercise. B2: A product of elementary matrices

Let $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix} \in M_2(\mathbb{R})$.

- (a) Write $\mathbf{A}$ as a product of elementary matrices.
- (b) Hence write $\mathbf{A}^{-1}$ as a product of elementary matrices, and compute it.

Solution.

- (a) Reduce $\mathbf{A}$ to $\mathbf{I}_2$:

$$\begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix} \xrightarrow{R_2 \rightarrow R_2 - 2R_1} \begin{pmatrix} 1 & 3 \\ 0 & -1 \end{pmatrix} \xrightarrow{R_2 \rightarrow -R_2} \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} \xrightarrow{R_1 \rightarrow R_1 - 3R_2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

With $\mathbf{E}_1 = \mathbf{I}_2 - 2\mathbf{E}_{21}$, $\mathbf{E}_2 = \mathbf{D}_2(-1)$, $\mathbf{E}_3 = \mathbf{I}_2 - 3\mathbf{E}_{12}$, [Theorem 3.4.1](#) gives $\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{I}_2$. Multiplying on the left by $\mathbf{E}_3^{-1}, \mathbf{E}_2^{-1}, \mathbf{E}_1^{-1}$ in turn and using [Theorem 3.4.2](#),

$$\mathbf{A} = \mathbf{E}_1^{-1}\mathbf{E}_2^{-1}\mathbf{E}_3^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}.$$

Check: the last two factors give $\begin{pmatrix} 1 & 3 \\ 0 & -1 \end{pmatrix}$, and adding twice row 1 to row 2 gives $\begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}$.

- (b) Since $\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{I}_2$, [Theorem 3.4.5](#) gives

$$\mathbf{A}^{-1} = \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1 = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -5 & 3 \\ 2 & -1 \end{pmatrix}.$$

(The last two factors give $\begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}$, and $R_1 \rightarrow R_1 - 3R_2$ then gives the answer.) This agrees with [Theorem 0.9.8](#), since $ad - bc = -1$.

Exercise. B3: Invertible products

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$, and suppose $\mathbf{AB}$ is invertible. Prove that $\mathbf{A}$ and $\mathbf{B}$ are both invertible.

Solution. Let $\mathbf{C} = (\mathbf{AB})^{-1}$. Then $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} = \mathbf{I}_n$ by associativity, so $\mathbf{A}$ has the right inverse $\mathbf{BC}$, and by [Theorem 3.4.5](#) $\mathbf{A}$ is invertible. Similarly $(\mathbf{CA})\mathbf{B} = \mathbf{C}(\mathbf{AB}) = \mathbf{I}_n$,

so $\mathbf{B}$ has the left inverse $\mathbf{CA}$. By [Theorem 3.4.5](#) with the roles of the two matrices swapped ($\mathbf{CA}$ times $\mathbf{B}$ is $\mathbf{I}_n$), $\mathbf{B}$ is invertible. This proves the claim.

3.4.7.3 C. Going deeper

Exercise. C1: Triangular matrices

Let $\mathbf{U} = (u_{ij}) \in M_n(F)$ be upper triangular (Definition 0.9.5).

- (a) Prove that $\mathbf{U}$ is invertible if and only if $u_{ii} \neq 0$ for every i .
- (b) Prove that if $\mathbf{U}$ is invertible, then $\mathbf{U}^{-1}$ is upper triangular.

Hint: for (a), use item (b) or item (f) of [Theorem 3.4.4](#); for (b), consider the systems $\mathbf{U}\mathbf{x} = \mathbf{e}_k$.

Solution.

- (a) ($\Leftarrow$) Suppose every $u_{ii} \neq 0$, and let $\mathbf{U}\mathbf{x} = \mathbf{0}$. Row i of this system reads $u_{ii}x_i + u_{i,i+1}x_{i+1} + \cdots + u_{in}x_n = 0$, since $u_{ij} = 0$ for $j < i$. We show $x_n = x_{n-1} = \cdots = x_1 = 0$ by downward induction. Row n is $u_{nn}x_n = 0$, and $u_{nn} \neq 0$, so $x_n = 0$. If $x_{i+1} = \cdots = x_n = 0$, row i becomes $u_{ii}x_i = 0$, so $x_i = 0$ since $u_{ii} \neq 0$. Hence $\mathbf{x} = \mathbf{0}$, and $\mathbf{U}$ is invertible by [Theorem 3.4.4](#) ((b) $\Rightarrow$ (a)).
- ($\Rightarrow$) Suppose $u_{kk} = 0$ for some k . Every column $j \leq k$ of $\mathbf{U}$ has zero entries in rows $k, k+1, \dots, n$: below the diagonal by triangularity, and in row k of column k by assumption. So the k columns $\mathbf{u}_1, \dots, \mathbf{u}_k$ all lie in $\text{span}(\mathbf{e}_1, \dots, \mathbf{e}_{k-1})$, a subspace of dimension $k-1$ (its spanning list is a sub-list of the standard basis, hence independent). By [Theorem 1.6.3](#) (a), a list of $k > k-1$ vectors there is linearly dependent. A nontrivial relation among $\mathbf{u}_1, \dots, \mathbf{u}_k$, extended by zero coefficients on $\mathbf{u}_{k+1}, \dots, \mathbf{u}_n$, is a nontrivial relation among all the columns. So the columns of $\mathbf{U}$ are dependent, and $\mathbf{U}$ is not invertible by [Theorem 3.4.4](#) ((a) $\Rightarrow$ (f)). (For $k = 1$, the span is $\{\mathbf{0}\}$ and $\mathbf{u}_1 = \mathbf{0}$, so the same conclusion holds.)
- (b) Suppose $\mathbf{U}$ is invertible, so every $u_{ii} \neq 0$ by (a). Fix k , and let $\mathbf{x}$ be the k -th column of $\mathbf{U}^{-1}$. By [Theorem 0.9.2](#), $\mathbf{U}\mathbf{x}$ is the k -th column of $\mathbf{U}\mathbf{U}^{-1} = \mathbf{I}_n$, so $\mathbf{U}\mathbf{x} = \mathbf{e}_k$. For $i > k$, row i of this system reads $u_{ii}x_i + \cdots + u_{in}x_n = 0$. By the same downward induction as in (a), running only over $i = n, n-1, \dots, k+1$, we get $x_n = \cdots = x_{k+1} = 0$. So the k -th column of $\mathbf{U}^{-1}$ has zeros below position k . As k was arbitrary, $\mathbf{U}^{-1}$ is upper triangular.

Exercise. C2: All right inverses

Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \in M_{2 \times 3}(\mathbb{R})$.

- (a) Find all $\mathbf{B} \in M_{3 \times 2}(\mathbb{R})$ with $\mathbf{AB} = \mathbf{I}_2$.
- (b) Show that there is no $\mathbf{C} \in M_{3 \times 2}(\mathbb{R})$ with $\mathbf{CA} = \mathbf{I}_3$.

Solution.

- (a) By Theorem 0.9.2, $\mathbf{AB} = \mathbf{I}_2$ means that the columns $\mathbf{b}_1, \mathbf{b}_2$ of $\mathbf{B}$ satisfy $\mathbf{Ab}_1 = \mathbf{e}_1$ and $\mathbf{Ab}_2 = \mathbf{e}_2$. The matrix $\mathbf{A}$ is already in reduced form with pivots in columns 1 and 2 and x_3 free, so $\mathbf{Ax} = \mathbf{0}$ means $x_1 = -x_3$, $x_2 = -x_3$, and its solutions are the multiples of $(-1, -1, 1)$. One solution of $\mathbf{Ax} = \mathbf{e}_1$ is $(1, 0, 0)$, and one solution of $\mathbf{Ax} = \mathbf{e}_2$ is $(0, 1, 0)$. By Theorem 3.1.3,

$$\mathbf{b}_1 = (1, 0, 0) + s(-1, -1, 1), \quad \mathbf{b}_2 = (0, 1, 0) + t(-1, -1, 1) \quad (s, t \in \mathbb{R}).$$

Hence the right inverses of $\mathbf{A}$ are exactly

$$\mathbf{B} = \begin{pmatrix} 1-s & -t \\ -s & 1-t \\ s & t \end{pmatrix}, \quad s, t \in \mathbb{R},$$

infinitely many of them.

- (b) Suppose $\mathbf{CA} = \mathbf{I}_3$, and let $\mathbf{x} = (-1, -1, 1)$, so $\mathbf{Ax} = \mathbf{0}$. Then $\mathbf{x} = \mathbf{I}_3\mathbf{x} = \mathbf{C}(\mathbf{Ax}) = \mathbf{C}\mathbf{0} = \mathbf{0}$, a contradiction since $\mathbf{x} \neq \mathbf{0}$. Hence $\mathbf{A}$ has no left inverse.

Exercise. C3: Nilpotent perturbations of the identity

Let $\mathbf{A} \in M_n(F)$ and suppose $\mathbf{A}^k = \mathbf{0}$ for some integer $k \geq 1$.

- (a) Prove that $\mathbf{I}_n + \mathbf{A}$ is invertible, by using item (b) of Theorem 3.4.4.
 (b) Verify that $(\mathbf{I}_n + \mathbf{A})(\mathbf{I}_n - \mathbf{A} + \mathbf{A}^2 - \cdots + (-1)^{k-1}\mathbf{A}^{k-1}) = \mathbf{I}_n$, and deduce a formula for $(\mathbf{I}_n + \mathbf{A})^{-1}$.

Solution.

- (a) Let $\mathbf{x} \in F^n$ with $(\mathbf{I}_n + \mathbf{A})\mathbf{x} = \mathbf{0}$. Then $\mathbf{Ax} = -\mathbf{x}$. Multiplying on the left by $\mathbf{A}$ and using this again, $\mathbf{A}^2\mathbf{x} = \mathbf{A}(-\mathbf{x}) = -\mathbf{Ax} = (-1)^2\mathbf{x}$, and inductively $\mathbf{A}^j\mathbf{x} = (-1)^j\mathbf{x}$ for every $j \geq 1$. Taking $j = k$, $\mathbf{0} = \mathbf{A}^k\mathbf{x} = (-1)^k\mathbf{x}$, so $\mathbf{x} = \mathbf{0}$ (multiply by $(-1)^k$). By Theorem 3.4.4 ((b) $\Rightarrow$ (a)), $\mathbf{I}_n + \mathbf{A}$ is invertible.
 (b) Expanding with Theorem 0.9.4,

$$(\mathbf{I}_n + \mathbf{A}) \sum_{j=0}^{k-1} (-1)^j \mathbf{A}^j = \sum_{j=0}^{k-1} (-1)^j \mathbf{A}^j + \sum_{j=0}^{k-1} (-1)^j \mathbf{A}^{j+1} = \mathbf{I}_n + (-1)^{k-1} \mathbf{A}^k = \mathbf{I}_n,$$

since the terms $(-1)^j \mathbf{A}^j$ for $1 \leq j \leq k-1$ in the first sum cancel against the terms $(-1)^{j-1} \mathbf{A}^j$ of the second, and $\mathbf{A}^k = \mathbf{0}$. By Theorem 3.4.5, $(\mathbf{I}_n + \mathbf{A})^{-1} = \mathbf{I}_n - \mathbf{A} + \mathbf{A}^2 - \cdots + (-1)^{k-1} \mathbf{A}^{k-1}$.