

0.9 Matrices

You have probably met matrices as tables of numbers that are multiplied by a strange rule. This section rebuilds them over an arbitrary field F and answers the questions a computational course leaves open. Why is the product defined the way it is? Which rules of school algebra survive, and which fail? What does it mean to divide by a matrix? Everything later in the book, from Gaussian elimination in Chapter 3 to the matrix of a linear map in Chapter 4, rests on the answers.

Throughout, F is a field (Definition 0.7.1). Nothing here uses more about F than the field axioms, so every statement holds over $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{F}_p$ alike.

0.9.1 Matrices and their entries

Definition 0.9.1. Matrix

Let F be a field and $m, n \geq 1$ integers. An $m \times n$ **matrix over** F is a rectangular array

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

with m rows and n columns, whose **entries** a_{ij} lie in F . Formally, it is a function $\{1, \dots, m\} \times \{1, \dots, n\} \rightarrow F$, $(i, j) \mapsto a_{ij}$. The set of all such matrices is $M_{m \times n}(F)$, and we write $M_n(F) = M_{n \times n}(F)$ for the **square** matrices.

The entry a_{ij} sits in row i and column j : **row first, column second**. We write $\mathbf{A} = (a_{ij})$, and $(\mathbf{A})_{ij}$ for the (i, j) -entry when the matrix has no letter of its own. Two matrices are **equal** when they have the **same size** and the same entry in every position; this is equality of functions (Definition 0.4.1).

Two sizes deserve their own names. A matrix in $M_{n \times 1}(F)$ is a **column vector**, and we write

$$F^n = M_{n \times 1}(F), \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

In running text we write $\mathbf{x} = (x_1, \dots, x_n)$ and still mean a column. Columns are printed in bold. A matrix in $M_{1 \times n}(F)$ is a **row vector**. For $j = 1, \dots, n$, the column $\mathbf{e}_j \in F^n$ has 1 in position j and 0 elsewhere.

We will often cut a matrix into its columns. If $\mathbf{a}_j = (a_{1j}, \dots, a_{mj}) \in F^m$ is the j -th column of $\mathbf{A}$, we write

$$\mathbf{A} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n).$$

The i -th row of $\mathbf{A}$ is the row vector $(a_{i1} \quad \cdots \quad a_{in}) \in M_{1 \times n}(F)$.

Example. Matrices over different fields

Identify the size, the field, and a few entries of each matrix:

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 0 \\ 3 & 1 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} i & 1+i \\ 0 & 2-i \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad \mathbf{D} = (5).$$

Here $\mathbf{C}$ is meant over $\mathbb{F}_2$. Also write out the 2×2 real matrix $\mathbf{E}$ with $e_{ij} = i + 2j$.

Solution. $\mathbf{A} \in M_{2 \times 3}(\mathbb{R})$, with $a_{12} = -2$ and $a_{23} = 4$; its second column is $\mathbf{a}_2 = (-2, 1)$. $\mathbf{B} \in M_2(\mathbb{C})$, with $b_{12} = 1 + i$ and $b_{21} = 0$. $\mathbf{C} \in M_3(\mathbb{F}_2)$; every entry is 0 or 1. $\mathbf{D} \in M_1(\mathbb{R})$ is the degenerate case, a 1×1 matrix. It behaves exactly like the number 5, and 1×1 matrices are how numbers will appear inside matrix formulas. Finally $e_{11} = 3$, $e_{12} = 5$, $e_{21} = 4$, $e_{22} = 6$, so $\mathbf{E} = \begin{pmatrix} 3 & 5 \\ 4 & 6 \end{pmatrix}$. Swapping the roles of i and j would give a different matrix, which is why the order “row, column” matters.

0.9.2 Special matrices

A few shapes of matrix occur so often that they have names. The **main diagonal** of a square matrix $\mathbf{A} \in M_n(F)$ consists of the entries $a_{11}, a_{22}, \dots, a_{nn}$.

Definition 0.9.2. Zero matrix

The **zero matrix** $\mathbf{O} = \mathbf{O}_{m \times n} \in M_{m \times n}(F)$ is the matrix all of whose entries are 0.

Definition 0.9.3. Identity matrix

The **identity matrix** $\mathbf{I}_n \in M_n(F)$ has entries

$$(\mathbf{I}_n)_{ij} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

That is, 1 on the main diagonal and 0 elsewhere. We write $\mathbf{I}$ when the size is clear. The symbol δ_{ij} is the **Kronecker delta**.

The columns of $\mathbf{I}_n$ are exactly $\mathbf{e}_1, \dots, \mathbf{e}_n$.

Definition 0.9.4. Diagonal matrix

A square matrix $\mathbf{A} \in M_n(F)$ is **diagonal** if $a_{ij} = 0$ whenever $i \neq j$. We write $\text{diag}(d_1, \dots, d_n)$ for the diagonal matrix with diagonal entries $d_1, \dots, d_n$.

Definition 0.9.5. Upper triangular matrix

A square matrix $\mathbf{A} \in M_n(F)$ is **upper triangular** if $a_{ij} = 0$ whenever $i > j$, that is, every

entry **strictly below** the main diagonal is zero. It is **lower triangular** if $a_{ij} = 0$ whenever $i < j$.

Example. Zero and identity matrices

Write out $\mathbf{O}_{2 \times 3}$, $\mathbf{I}_2$, and $\mathbf{I}_3$. Which of them are diagonal?

Solution.

$$\mathbf{O}_{2 \times 3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$\mathbf{I}_2 = \text{diag}(1, 1)$ and $\mathbf{I}_3 = \text{diag}(1, 1, 1)$ are diagonal. The matrix $\mathbf{O}_{2 \times 3}$ is **not** diagonal, because the definition only applies to square matrices; the square zero matrix $\mathbf{O}_{n \times n} = \text{diag}(0, \dots, 0)$ is diagonal.

Example. Diagonal and triangular

Classify each matrix as diagonal, upper triangular, lower triangular, or none of these:

$$\mathbf{P} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 7 & 6 \end{pmatrix}.$$

Solution. $\mathbf{P} = \text{diag}(3, -1, 0)$ is diagonal. A zero on the diagonal is allowed: the definition only constrains off-diagonal entries. $\mathbf{P}$ is also both upper and lower triangular, and in general a matrix is diagonal exactly when it is both. $\mathbf{Q}$ is upper triangular but not lower triangular, since $q_{12} = 2 \neq 0$. $\mathbf{R}$ differs from $\mathbf{Q}$ in one entry, $r_{32} = 7$. All other entries below the diagonal are still zero, but the clause “ $a_{ij} = 0$ whenever $i > j$ ” fails at $(i, j) = (3, 2)$, so $\mathbf{R}$ is not upper triangular. It is not lower triangular either.

0.9.3 Addition and scalar multiplication

Matrices of the same size are added, and multiplied by scalars, one entry at a time.

Definition 0.9.6. Matrix addition

Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be in $M_{m \times n}(F)$, of the **same size**. Their **sum** is $\mathbf{A} + \mathbf{B} \in M_{m \times n}(F)$ with $(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$.

Definition 0.9.7. Scalar multiplication

Let $\mathbf{A} = (a_{ij}) \in M_{m \times n}(F)$ and $c \in F$. The **scalar multiple** $c\mathbf{A} \in M_{m \times n}(F)$ has entries $(c\mathbf{A})_{ij} = ca_{ij}$.

Definition 0.9.8. Additive inverse

The **additive inverse** of $\mathbf{A} \in M_{m \times n}(F)$ is $-\mathbf{A} = (-1)\mathbf{A}$, with entries $-a_{ij}$. We write $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$.

Matrices of different sizes are never added: $\begin{pmatrix} 1 & 2 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is not defined.

Example. Adding and scaling

Let $\mathbf{A} = \begin{pmatrix} 1 & -2 & 0 \\ 3 & 1 & 4 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 2 & 2 & -1 \\ 0 & -5 & 1 \end{pmatrix}$ in $M_{2 \times 3}(\mathbb{R})$. Compute $\mathbf{A} + \mathbf{B}$, $3\mathbf{A}$, and $\mathbf{A} - 2\mathbf{B}$.

Solution. Entry by entry,

$$\begin{aligned}\mathbf{A} + \mathbf{B} &= \begin{pmatrix} 3 & 0 & -1 \\ 3 & -4 & 5 \end{pmatrix}, \\ 3\mathbf{A} &= \begin{pmatrix} 3 & -6 & 0 \\ 9 & 3 & 12 \end{pmatrix}, \\ \mathbf{A} - 2\mathbf{B} &= \begin{pmatrix} 1-4 & -2-4 & 0+2 \\ 3-0 & 1+10 & 4-2 \end{pmatrix} = \begin{pmatrix} -3 & -6 & 2 \\ 3 & 11 & 2 \end{pmatrix}.\end{aligned}$$

Because both operations work entry by entry, they inherit the rules of F :

Theorem 0.9.1. Properties of addition and scalar multiplication

Let $\mathbf{A}, \mathbf{B}, \mathbf{C} \in M_{m \times n}(F)$ and $c, d \in F$. Then:

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$;
2. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$;
3. $\mathbf{A} + \mathbf{O} = \mathbf{A}$;
4. $\mathbf{A} + (-\mathbf{A}) = \mathbf{O}$;
5. $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$ and $(c + d)\mathbf{A} = c\mathbf{A} + d\mathbf{A}$;
6. $(cd)\mathbf{A} = c(d\mathbf{A})$ and $1\mathbf{A} = \mathbf{A}$.

Proof. Two matrices of the same size are equal when their entries agree, so it suffices to compare (i, j) -entries. For 1, $(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij} = b_{ij} + a_{ij} = (\mathbf{B} + \mathbf{A})_{ij}$, where the middle equality is commutativity of addition in F . For 5, $(c(\mathbf{A} + \mathbf{B}))_{ij} = c(a_{ij} + b_{ij}) = ca_{ij} + cb_{ij} = (c\mathbf{A} + c\mathbf{B})_{ij}$ by distributivity in F . Each remaining identity reduces in the same way to the corresponding field axiom of Definition 0.7.1 in each entry. This proves the theorem. ■

These eight rules will reappear in Chapter 1 as the axioms of a vector space: $M_{m \times n}(F)$ is one of our standard examples.

0.9.4 Matrix multiplication

Addition was entrywise. The obvious guess for multiplication is also entrywise, but that product is almost useless. The product that matters is the one that records **substitution**, and it looks very different.

Suppose new variables y_1, y_2 are given in terms of x_1, x_2 , and z_1, z_2 in terms of y_1, y_2 :

$$\begin{aligned} y_1 &= 2x_1 + x_2 & z_1 &= y_1 + 2y_2 \\ y_2 &= x_1 + 4x_2 & z_2 &= 3y_1 - y_2 \end{aligned}$$

Record each substitution by its table of coefficients: $\mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$ for $\mathbf{y}$ in terms of $\mathbf{x}$, and $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$ for $\mathbf{z}$ in terms of $\mathbf{y}$. Substituting,

$$\begin{aligned} z_1 &= (2x_1 + x_2) + 2(x_1 + 4x_2) \\ &= (1 \cdot 2 + 2 \cdot 1)x_1 + (1 \cdot 1 + 2 \cdot 4)x_2 = 4x_1 + 9x_2, \\ z_2 &= 3(2x_1 + x_2) - (x_1 + 4x_2) \\ &= (3 \cdot 2 + (-1) \cdot 1)x_1 + (3 \cdot 1 + (-1) \cdot 4)x_2 = 5x_1 - x_2. \end{aligned}$$

So $\mathbf{z}$ in terms of $\mathbf{x}$ has coefficient table $\begin{pmatrix} 4 & 9 \\ 5 & -1 \end{pmatrix}$. Look at how each entry arose: the coefficient of x_k in z_i is $a_{i1}b_{1k} + a_{i2}b_{2k}$, row i of $\mathbf{A}$ against column k of $\mathbf{B}$. The entrywise product $\begin{pmatrix} 2 & 2 \\ 3 & -4 \end{pmatrix}$ has nothing to do with the substitution. This forces the definition.

To multiply matrices, run each row of the first against each column of the second: the product of two matrices is the table of the combined substitution.

Definition 0.9.9. Matrix multiplication

Let $\mathbf{A} = (a_{ij}) \in M_{m \times n}(F)$ and $\mathbf{B} = (b_{jk}) \in M_{n \times p}(F)$, so that **the number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$** . Their **product** is the matrix $\mathbf{AB} \in M_{m \times p}(F)$ with entries

$$(\mathbf{AB})_{ik} = \sum_{j=1}^n a_{ij}b_{jk} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}$$

for $1 \leq i \leq m$ and $1 \leq k \leq p$. If the sizes do not match, $\mathbf{AB}$ is **not defined**.

In words: to get the (i, k) -entry of $\mathbf{AB}$, walk along row i of $\mathbf{A}$ and down column k of $\mathbf{B}$ at the same time, multiply the entries you meet in pairs, and add. The size condition says that the row and the column have the same length n , so the pairing makes sense. The product has as many rows as $\mathbf{A}$ and as many columns as $\mathbf{B}$:

$$\begin{matrix} \mathbf{A} & \mathbf{B} \\ m \times n & n \times p \end{matrix} = \begin{matrix} \mathbf{AB} \\ m \times p \end{matrix}.$$

In particular, for $\mathbf{A} \in M_{m \times n}(F)$ and a column $\mathbf{x} \in F^n$, the product $\mathbf{Ax} \in F^m$ is defined, with i -th entry $a_{i1}x_1 + \cdots + a_{in}x_n$. A system of m linear equations in n unknowns is therefore a single equation $\mathbf{Ax} = \mathbf{b}$.

Example. A product of square matrices

Let $\mathbf{A} = \begin{pmatrix} 2 & -1 \\ 0 & 3 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 4 \\ -2 & 1 \end{pmatrix}$ in $M_2(\mathbb{R})$. Compute $\mathbf{AB}$ and $\mathbf{BA}$.

Solution. Row i of $\mathbf{A}$ against column k of $\mathbf{B}$:

$$\begin{aligned}\mathbf{AB} &= \begin{pmatrix} 2 \cdot 1 + (-1)(-2) & 2 \cdot 4 + (-1) \cdot 1 \\ 0 \cdot 1 + 3 \cdot (-2) & 0 \cdot 4 + 3 \cdot 1 \end{pmatrix} = \begin{pmatrix} 4 & 7 \\ -6 & 3 \end{pmatrix}, \\ \mathbf{BA} &= \begin{pmatrix} 1 \cdot 2 + 4 \cdot 0 & 1 \cdot (-1) + 4 \cdot 3 \\ -2 \cdot 2 + 1 \cdot 0 & (-2)(-1) + 1 \cdot 3 \end{pmatrix} = \begin{pmatrix} 2 & 11 \\ -4 & 5 \end{pmatrix}.\end{aligned}$$

Both products are defined, and they are different.

Example. Products of rectangular matrices

Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{pmatrix} \in M_{2 \times 3}(\mathbb{R})$ and $\mathbf{B} = \begin{pmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 0 \end{pmatrix} \in M_{3 \times 2}(\mathbb{R})$. Determine the sizes of $\mathbf{AB}$ and $\mathbf{BA}$ and compute them. Then compute $\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix}$.

Solution. $\mathbf{AB}$ is $(2 \times 3)(3 \times 2)$, so it is 2×2 ; $\mathbf{BA}$ is $(3 \times 2)(2 \times 3)$, so it is 3×3 . Computing row against column,

$$\mathbf{AB} = \begin{pmatrix} 3 + 0 + 2 & 1 + 0 + 0 \\ -3 + 6 + 1 & -1 + 3 + 0 \end{pmatrix} = \begin{pmatrix} 5 & 1 \\ 4 & 2 \end{pmatrix}, \quad \mathbf{BA} = \begin{pmatrix} 2 & 3 & 7 \\ 1 & 3 & 5 \\ 1 & 0 & 2 \end{pmatrix}.$$

For instance $(\mathbf{BA})_{13} = 3 \cdot 2 + 1 \cdot 1 = 7$. The last two products show the extreme cases: a row times a column is 1×1 , namely $\begin{pmatrix} 1 \cdot 3 + 2 \cdot 4 \end{pmatrix} = \begin{pmatrix} 11 \end{pmatrix}$, while a column times a row is 2×2 , namely $\begin{pmatrix} 3 & 6 \\ 4 & 8 \end{pmatrix}$.

Here is a nonexample by minimal change. Take the same six entries as $\mathbf{B}$, but arrange them in two rows: $\mathbf{C} = \begin{pmatrix} 3 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix} \in M_{2 \times 3}(\mathbb{R})$. The entries are still real numbers and $\mathbf{A}$ is unchanged. But $\mathbf{A}$ has 3 columns and $\mathbf{C}$ has 2 rows, so forming $\mathbf{AC}$ would pair a row of length 3 with a column of length 2. The clause “the number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$ ” fails, and $\mathbf{AC}$ is **not defined**.

Why this definition, and not the entrywise one? Because it is the only rule under which “the table of a combined substitution is the product of the tables”. In Chapter 4 the same fact becomes: the matrix of a composition of linear maps is the product of their matrices. Every property of matrix multiplication, good or bad, comes from this.

0.9.5 Three views of a product

The entry formula is how one computes a product by hand, but it hides the structure. Two other descriptions, by columns and by rows, are what we will actually think with. It turns out that multiplying by $\mathbf{A}$ on the left acts on each column of $\mathbf{B}$ separately, and multiplying by $\mathbf{B}$ on the right acts on each row of $\mathbf{A}$ separately:

Theorem 0.9.2. Three views of a product

Let $\mathbf{A} \in M_{m \times n}(F)$ and $\mathbf{B} \in M_{n \times p}(F)$, and write $\mathbf{B} = (\mathbf{b}_1 \ \cdots \ \mathbf{b}_p)$ with columns $\mathbf{b}_k \in F^n$. Then:

1. **(Entries)** $(\mathbf{AB})_{ik} = \sum_{j=1}^n a_{ij}b_{jk}$ for all i, k ;
2. **(Columns)** the k -th column of $\mathbf{AB}$ is $\mathbf{Ab}_k$, that is,

$$\mathbf{AB} = (\mathbf{Ab}_1 \ \mathbf{Ab}_2 \ \cdots \ \mathbf{Ab}_p);$$

3. **(Rows)** the i -th row of $\mathbf{AB}$ is $(\text{row } i \text{ of } \mathbf{A})\mathbf{B}$.

Proof. Part 1 is [Definition 0.9.9](#). For part 2, fix k . The column $\mathbf{b}_k$ has entries $b_{1k}, \dots, b_{nk}$, so by [Definition 0.9.9](#) the i -th entry of $\mathbf{Ab}_k \in F^m$ is $\sum_{j=1}^n a_{ij}b_{jk}$. By part 1 this is $(\mathbf{AB})_{ik}$, the i -th entry of the k -th column of $\mathbf{AB}$. Hence the k -th column of $\mathbf{AB}$ is $\mathbf{Ab}_k$.

For part 3, fix i , and let $\mathbf{r} = (a_{i1} \ \cdots \ a_{in})$ be row i of $\mathbf{A}$. Then $\mathbf{rB}$ is a $1 \times p$ matrix whose k -th entry is $\sum_{j=1}^n a_{ij}b_{jk} = (\mathbf{AB})_{ik}$, again by [Definition 0.9.9](#). This is the k -th entry of row i of $\mathbf{AB}$. This proves the theorem. ■

The column view reduces every matrix product to products $\mathbf{Ax}$ of a matrix with a single column. So it pays to understand $\mathbf{Ax}$ well, and here is the fact to understand:

Theorem 0.9.3. $\mathbf{Ax}$ is a combination of the columns of $\mathbf{A}$

Let $\mathbf{A} = (\mathbf{a}_1 \ \cdots \ \mathbf{a}_n) \in M_{m \times n}(F)$ with columns $\mathbf{a}_j \in F^m$, and let $\mathbf{x} = (x_1, \dots, x_n) \in F^n$. Then

$$\mathbf{Ax} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

In particular $\mathbf{Ae}_j = \mathbf{a}_j$ for each j . Similarly, if $\mathbf{r}_1, \dots, \mathbf{r}_m$ are the rows of $\mathbf{A}$ and $\mathbf{y} = (y_1 \ \cdots \ y_m)$ is a row vector, then $\mathbf{yA} = y_1\mathbf{r}_1 + \cdots + y_m\mathbf{r}_m$.

Proof. Both sides are in F^m , so we compare i -th entries. By [Definition 0.9.9](#), the i -th entry of $\mathbf{Ax}$ is $\sum_{j=1}^n a_{ij}x_j$. On the right, the i -th entry of $x_j\mathbf{a}_j$ is x_ja_{ij} by [Definition 0.9.7](#), so by [Definition 0.9.6](#) the i -th entry of $\sum_j x_j\mathbf{a}_j$ is $\sum_{j=1}^n x_ja_{ij}$. These agree because multiplication in F is commutative. Taking $\mathbf{x} = \mathbf{e}_j$, whose only nonzero entry is $x_j = 1$, gives $\mathbf{Ae}_j = \mathbf{a}_j$. The statement for rows follows in the same way, comparing the k -th entries $\sum_i y_ia_{ik}$. This proves the theorem. ■

Read the formula as a sentence: **the entries of $\mathbf{x}$ are the weights, and the columns of $\mathbf{A}$ are what gets weighted.** So $\mathbf{Ax} = \mathbf{b}$ has a solution exactly when $\mathbf{b}$ can be built from the columns of $\mathbf{A}$. This single observation drives much of Chapters 2 and 3. Combined with [Theorem 0.9.2](#), it also says that **every column of $\mathbf{AB}$ is a combination of the columns of $\mathbf{A}$** , with weights taken from the corresponding column of $\mathbf{B}$, and **every row of $\mathbf{AB}$ is a combination of the rows of $\mathbf{B}$.**

Example. One product, three ways

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & -1 \\ 0 & 4 \end{pmatrix} \in M_{3 \times 2}(\mathbb{R}), \quad \mathbf{B} = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \in M_2(\mathbb{R}), \quad \mathbf{x} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Compute $\mathbf{Ax}$ as a combination of the columns of $\mathbf{A}$. Then compute $\mathbf{AB}$ by columns, check one row by the row view, and check one entry by the entry formula.

Solution. The columns of $\mathbf{A}$ are $\mathbf{a}_1 = (1, 3, 0)$ and $\mathbf{a}_2 = (2, -1, 4)$. By [Theorem 0.9.3](#),

$$\mathbf{Ax} = 2\mathbf{a}_1 - \mathbf{a}_2 = \begin{pmatrix} 2 \\ 6 \\ 0 \end{pmatrix} - \begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 7 \\ -4 \end{pmatrix}.$$

Columns. The first column of $\mathbf{B}$ is $\mathbf{b}_1 = (2, -1) = \mathbf{x}$, so the first column of $\mathbf{AB}$ is $\mathbf{Ab}_1 = (0, 7, -4)$, just computed. The second column is $\mathbf{b}_2 = (1, 1)$, so $\mathbf{Ab}_2 = \mathbf{a}_1 + \mathbf{a}_2 = (3, 2, 4)$. Hence

$$\mathbf{AB} = \begin{pmatrix} 0 & 3 \\ 7 & 2 \\ -4 & 4 \end{pmatrix}.$$

Rows. Row 2 of $\mathbf{A}$ is $(3 \ -1)$, and by [Theorem 0.9.3](#) (rows version) $(3 \ -1)\mathbf{B} = 3(2 \ 1) - (-1 \ 1) = (7 \ 2)$, which is row 2 of $\mathbf{AB}$.

Entries. $(\mathbf{AB})_{32} = a_{31}b_{12} + a_{32}b_{22} = 0 \cdot 1 + 4 \cdot 1 = 4$, as in the answer.

The column view needed nothing beyond scaling and adding columns of $\mathbf{A}$. When the columns of $\mathbf{B}$ are simple, it is usually the fastest route.

Quick check. Without multiplying out, what is the second column of $\begin{pmatrix} 5 & -1 & 7 \\ 2 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 4 & 1 \end{pmatrix}$?

Solution. The second column of the right factor is $\mathbf{e}_3 = (0, 0, 1)$. By [Theorem 0.9.2](#) and [Theorem 0.9.3](#), the second column of the product is $\mathbf{Ae}_3$, the third column of the left factor: $(7, 3)$.

0.9.6 The rules of multiplication

Matrix multiplication keeps the most important rules of arithmetic: associativity, both distributive laws, and an identity.

Theorem 0.9.4. Properties of matrix multiplication

Let $c \in F$, and let $\mathbf{A}, \mathbf{A}' \in M_{m \times n}(F)$, $\mathbf{B}, \mathbf{B}' \in M_{n \times p}(F)$, and $\mathbf{C} \in M_{p \times q}(F)$. Then:

1. **(Associativity)** $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$;
2. **(Distributivity)** $\mathbf{A}(\mathbf{B} + \mathbf{B}') = \mathbf{AB} + \mathbf{AB}'$ and $(\mathbf{A} + \mathbf{A}')\mathbf{B} = \mathbf{AB} + \mathbf{A'B}$;
3. **(Identity)** $\mathbf{I}_m \mathbf{A} = \mathbf{A} = \mathbf{A} \mathbf{I}_n$;
4. **(Scalars)** $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B} = \mathbf{A}(c\mathbf{B})$;
5. **(Zero)** $\mathbf{O}_{\ell \times m} \mathbf{A} = \mathbf{O}_{\ell \times n}$ and $\mathbf{A} \mathbf{O}_{n \times r} = \mathbf{O}_{m \times r}$ for all $\ell, r \geq 1$.

Idea. Each identity is checked entry by entry. The only one with content is associativity. Both sides have (i, l) -entry equal to the double sum $\sum_j \sum_k a_{ij} b_{jk} c_{kl}$, added up in two different orders, so the proof is “expand, swap the order of the finite sums, and fold back up”. In the substitution picture it is no surprise: substituting in two stages gives the same result whichever pair of stages we combine first.

Proof. For 1, both sides lie in $M_{m \times q}(F)$. Fix i and l . By [Definition 0.9.9](#) applied twice,

$$((\mathbf{AB})\mathbf{C})_{il} = \sum_{k=1}^p (\mathbf{AB})_{ik} c_{kl} = \sum_{k=1}^p \left(\sum_{j=1}^n a_{ij} b_{jk} \right) c_{kl} = \sum_{k=1}^p \sum_{j=1}^n a_{ij} b_{jk} c_{kl},$$

where the last step uses distributivity in F . The same computation gives

$$(\mathbf{A}(\mathbf{BC}))_{il} = \sum_{j=1}^n a_{ij} (\mathbf{BC})_{jl} = \sum_{j=1}^n a_{ij} \left(\sum_{k=1}^p b_{jk} c_{kl} \right) = \sum_{j=1}^n \sum_{k=1}^p a_{ij} b_{jk} c_{kl}.$$

The two double sums add the same np terms in different orders, so they are equal, by commutativity and associativity of addition in F . Hence $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

For 2, $(\mathbf{A}(\mathbf{B} + \mathbf{B}'))_{ik} = \sum_j a_{ij} (b_{jk} + b'_{jk}) = \sum_j a_{ij} b_{jk} + \sum_j a_{ij} b'_{jk} = (\mathbf{AB} + \mathbf{AB}')_{ik}$ by distributivity in F ; the other law is the same with the roles of the factors swapped.

For 3, $(\mathbf{I}_m \mathbf{A})_{ik} = \sum_{j=1}^m \delta_{ij} a_{jk} = a_{ik}$, because by [Definition 0.9.3](#) only the term $j = i$ survives. Likewise $(\mathbf{A} \mathbf{I}_n)_{ik} = \sum_j a_{ij} \delta_{jk} = a_{ik}$.

For 4, $(c(\mathbf{AB}))_{ik} = c \sum_j a_{ij} b_{jk} = \sum_j (ca_{ij}) b_{jk} = ((c\mathbf{A})\mathbf{B})_{ik}$, and moving c next to b_{jk} instead, using commutativity in F , gives $(\mathbf{A}(c\mathbf{B}))_{ik}$. For 5, every term $0 \cdot a_{jk}$ or $a_{ij} \cdot 0$ is 0 by Theorem 0.7.1. This proves the theorem. ■

Associativity means we may write $\mathbf{ABC}$ without brackets, and more generally a product of any number of matrices, as long as consecutive sizes match.

Just as important is what is **missing** from the list. There is no commutative law, and no

cancellation law. Each of the following failures will cost you a wrong proof if you forget it.

Warning. **$\mathbf{AB} \neq \mathbf{BA}$ in general.** Take $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Then $\mathbf{AB} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ but $\mathbf{BA} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$. Before swapping two factors in a proof, you must know that they commute. (And if $\mathbf{A}$ is 2×3 and $\mathbf{B}$ is 3×2 , then $\mathbf{AB}$ and $\mathbf{BA}$ do not even have the same size.)

Warning. **A product can be zero with both factors nonzero.** In $M_2(\mathbb{R})$,

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

So $M_n(F)$ is not a field for $n \geq 2$, and “ $\mathbf{AB} = \mathbf{O}$ implies $\mathbf{A} = \mathbf{O}$ or $\mathbf{B} = \mathbf{O}$ ” is false.

Warning. **Cancellation fails: $\mathbf{AC} = \mathbf{BC}$ does not imply $\mathbf{A} = \mathbf{B}$.** With

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 3 & 9 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

we get $\mathbf{AC} = \mathbf{BC} = \begin{pmatrix} 1 & 1 \\ 3 & 3 \end{pmatrix}$ although $\mathbf{A} \neq \mathbf{B}$. By the column view, both columns of $\mathbf{C}$ are $\mathbf{e}_1$, so $\mathbf{AC}$ and $\mathbf{BC}$ only see the first columns of $\mathbf{A}$ and $\mathbf{B}$, which agree. Canceling $\mathbf{C}$ is legal only when $\mathbf{C}$ is invertible, as we will see below.

Quick check. Find a nonzero $\mathbf{A} \in M_2(\mathbb{R})$ with $\mathbf{A}^2 = \mathbf{AA} = \mathbf{O}$.

Solution. Take $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. By the column view, $\mathbf{Ae}_1 = \mathbf{0}$ and $\mathbf{Ae}_2 = \mathbf{e}_1$, so the columns of $\mathbf{A}^2$ are $\mathbf{A0} = \mathbf{0}$ and $\mathbf{Ae}_1 = \mathbf{0}$. Hence $\mathbf{A}^2 = \mathbf{O}$ although $\mathbf{A} \neq \mathbf{O}$. This small matrix is the source of a remarkable number of counterexamples in this book; it will reappear often.

0.9.7 The transpose

Rows and columns play symmetric roles in a matrix, and the transpose exchanges them.

Definition 0.9.10. Transpose

Let $\mathbf{A} = (a_{ij}) \in M_{m \times n}(F)$. The **transpose** of $\mathbf{A}$ is the matrix $\mathbf{A}^\top \in M_{n \times m}(F)$ with entries

$$(\mathbf{A}^\top)_{ij} = a_{ji}.$$

In words: row i of $\mathbf{A}$ becomes column i of $\mathbf{A}^\top$, and the size flips from $m \times n$ to $n \times m$. A square matrix is reflected in its main diagonal.

Definition 0.9.11. Symmetric and skew-symmetric matrices

A square matrix $\mathbf{A} \in M_n(F)$ is **symmetric** if $\mathbf{A}^\top = \mathbf{A}$, that is, $a_{ij} = a_{ji}$ for all i, j . It is **skew-symmetric** if $\mathbf{A}^\top = -\mathbf{A}$, that is, $a_{ij} = -a_{ji}$ for all i, j .

Example. Transposes, symmetric and skew-symmetric matrices

Compute the transpose of $\mathbf{A} = \begin{pmatrix} 1 & -2 & 0 \\ 3 & 1 & 4 \end{pmatrix}$. Decide whether

$$\mathbf{S} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}, \quad \mathbf{K} = \begin{pmatrix} 0 & 2 & -1 \\ -2 & 0 & 3 \\ 1 & -3 & 0 \end{pmatrix}, \quad \mathbf{S}' = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 6 & 6 \end{pmatrix}$$

over $\mathbb{R}$ are symmetric or skew-symmetric.

Solution. The rows of $\mathbf{A}$ become columns: $\mathbf{A}^\top = \begin{pmatrix} 1 & 3 \\ -2 & 1 \\ 0 & 4 \end{pmatrix} \in M_{3 \times 2}(\mathbb{R})$.

$\mathbf{S}$ is symmetric: $s_{12} = s_{21} = 2$, $s_{13} = s_{31} = 3$, $s_{23} = s_{32} = 5$. $\mathbf{K}$ is skew-symmetric: $k_{12} = 2 = -k_{21}$, $k_{13} = -1 = -k_{31}$, $k_{23} = 3 = -k_{32}$, and each diagonal entry satisfies $k_{ii} = -k_{ii}$, so it is 0. The matrix $\mathbf{S}'$ differs from $\mathbf{S}$ only in the entry $(3, 2)$. Every other symmetric pair still matches, but $s'_{23} = 5 \neq 6 = s'_{32}$, so $\mathbf{S}'$ is **not** symmetric. A 1×1 matrix is always symmetric, and the identity $\mathbf{I}_n$ and every diagonal matrix are symmetric.

Over $\mathbb{R}$, a skew-symmetric matrix has zero diagonal, since $a_{ii} = -a_{ii}$ forces $2a_{ii} = 0$. Over $\mathbb{F}_2$, where $-1 = 1$, “skew-symmetric” and “symmetric” mean the same thing, and $\mathbf{I}_2$ is skew-symmetric. The characteristic of the field (Definition 0.7.2) matters.

The transpose is compatible with the operations, except that it **reverses** products:

Theorem 0.9.5. Properties of the transpose

Let $c \in F$. For matrices of sizes for which each expression is defined:

1. $(\mathbf{A}^\top)^\top = \mathbf{A}$;
2. $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$;
3. $(c\mathbf{A})^\top = c\mathbf{A}^\top$;
4. $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$ for $\mathbf{A} \in M_{m \times n}(F)$ and $\mathbf{B} \in M_{n \times p}(F)$.

Idea. The order reverses for a size reason before any computation: $(\mathbf{AB})^\top$ is $p \times m$, and $\mathbf{B}^\top \mathbf{A}^\top$ is $(p \times n)(n \times m) = p \times m$, while $\mathbf{A}^\top \mathbf{B}^\top$ is $(n \times m)(p \times n)$, usually not even defined. Then compare entries.

Proof. For 1, $((\mathbf{A}^\top)^\top)_{ij} = (\mathbf{A}^\top)_{ji} = a_{ij}$ by Definition 0.9.10. For 2 and 3, $((\mathbf{A} + \mathbf{B})^\top)_{ij} = a_{ji} + b_{ji} = (\mathbf{A}^\top + \mathbf{B}^\top)_{ij}$ and $((c\mathbf{A})^\top)_{ij} = ca_{ji} = (c\mathbf{A}^\top)_{ij}$.

For 4, both sides lie in $M_{p \times m}(F)$. Fix $1 \leq i \leq p$ and $1 \leq k \leq m$. By Definition 0.9.10 and Definition 0.9.9,

$$\begin{aligned} ((\mathbf{AB})^\top)_{ik} &= (\mathbf{AB})_{ki} = \sum_{j=1}^n a_{kj} b_{ji}, \\ (\mathbf{B}^\top \mathbf{A}^\top)_{ik} &= \sum_{j=1}^n (\mathbf{B}^\top)_{ij} (\mathbf{A}^\top)_{jk} = \sum_{j=1}^n b_{ji} a_{kj}. \end{aligned}$$

The two sums agree term by term because multiplication in F is commutative. Hence $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$. This proves the theorem. ■

For example, for every $\mathbf{A} \in M_{m \times n}(F)$, the square matrix $\mathbf{A}^\top \mathbf{A}$ is symmetric: by parts 4 and 1, $(\mathbf{A}^\top \mathbf{A})^\top = \mathbf{A}^\top (\mathbf{A}^\top)^\top = \mathbf{A}^\top \mathbf{A}$. Matrices of this form are central to least squares and the singular value decomposition much later on.

0.9.8 The trace

The trace is the simplest number attached to a square matrix, and it has one surprising property.

Definition 0.9.12. Trace

The **trace** of a square matrix $\mathbf{A} = (a_{ij}) \in M_n(F)$ is the sum of its diagonal entries:

$$\operatorname{tr} \mathbf{A} = \sum_{i=1}^n a_{ii} = a_{11} + a_{22} + \cdots + a_{nn} \in F.$$

Example. Computing traces

Compute $\operatorname{tr} \mathbf{A}$ for $\mathbf{A} = \begin{pmatrix} 2 & 7 & -1 \\ 0 & -3 & 5 \\ 4 & 1 & 6 \end{pmatrix} \in M_3(\mathbb{R})$, and $\operatorname{tr} \mathbf{I}_n$ over $\mathbb{R}$ and over $\mathbb{F}_2$.

Solution. $\operatorname{tr} \mathbf{A} = 2 + (-3) + 6 = 5$; the off-diagonal entries play no role. $\operatorname{tr} \mathbf{I}_n = 1 + \cdots + 1 = n \cdot 1$. Over $\mathbb{R}$ this is n . Over $\mathbb{F}_2$ it is 0 when n is even and 1 when n is odd. The trace is **not** defined for a nonsquare matrix.

Theorem 0.9.6. Properties of the trace

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ and $c \in F$. Then:

1. $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr} \mathbf{A} + \text{tr} \mathbf{B}$ and $\text{tr}(c\mathbf{A}) = c \text{tr} \mathbf{A}$;
2. $\text{tr}(\mathbf{A}^\top) = \text{tr} \mathbf{A}$;
3. $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$. More generally, this holds for all $\mathbf{A} \in M_{m \times n}(F)$ and $\mathbf{B} \in M_{n \times m}(F)$, where $\mathbf{AB}$ is $m \times m$ and $\mathbf{BA}$ is $n \times n$.

Idea. Part 3 is surprising because $\mathbf{AB} \neq \mathbf{BA}$ in general. But the trace only sees the diagonal, and $\text{tr}(\mathbf{AB}) = \sum_i \sum_j a_{ij} b_{ji}$ is a sum over all pairs (i, j) of the product of a_{ij} with the “mirror” entry b_{ji} . That description does not care which matrix is written first.

Proof. Parts 1 and 2 follow from the definitions: the diagonal entries of $\mathbf{A} + \mathbf{B}$, $c\mathbf{A}$, and $\mathbf{A}^\top$ are $a_{ii} + b_{ii}$, ca_{ii} , and a_{ii} .

For 3, let $\mathbf{A} \in M_{m \times n}(F)$ and $\mathbf{B} \in M_{n \times m}(F)$. By Definition 0.9.12 and Definition 0.9.9,

$$\text{tr}(\mathbf{AB}) = \sum_{i=1}^m (\mathbf{AB})_{ii} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ji} = \sum_{j=1}^n \sum_{i=1}^m b_{ji} a_{ij} = \sum_{j=1}^n (\mathbf{BA})_{jj} = \text{tr}(\mathbf{BA}),$$

where the middle equality swaps the order of the finite double sum and uses commutativity of multiplication in F . This proves the theorem. ■

A first payoff: for all $\mathbf{A}, \mathbf{B} \in M_n(F)$, the matrix $\mathbf{AB} - \mathbf{BA}$ has trace $\text{tr}(\mathbf{AB}) - \text{tr}(\mathbf{BA}) = 0$. So over $\mathbb{R}$ it can never equal $\mathbf{I}_n$, whose trace is n (Exercise (C1: $\mathbf{AB} - \mathbf{BA} = \mathbf{I}$ is impossible, or is it?)). The trace will also turn out to be one of the quantities that does not change when we change coordinates.

It is tempting to conclude that the trace of a product ignores the order of **all** the factors. It does not.

Warning. $\text{tr}(\mathbf{ABC}) \neq \text{tr}(\mathbf{BAC})$ in general. Take $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, and $\mathbf{C} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Then $\mathbf{ABC} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ has trace 1, but $\mathbf{BA} = \mathbf{O}$, so $\mathbf{BAC} = \mathbf{O}$ has trace 0. What is true is that a **cyclic** rotation preserves the trace: $\text{tr}(\mathbf{ABC}) = \text{tr}((\mathbf{AB})\mathbf{C}) = \text{tr}(\mathbf{C}(\mathbf{AB})) = \text{tr}(\mathbf{CAB})$ by part 3.

0.9.9 Inverse matrices

In a field we solve $ax = b$ by multiplying by a^{-1} , and this works exactly when $a \neq 0$. We would like to solve a system $\mathbf{Ax} = \mathbf{b}$ the same way, by “dividing by $\mathbf{A}$ ”. What should that mean, and for which $\mathbf{A}$ is it possible?

The naive answer, “every nonzero matrix”, fails. Let $\mathbf{N} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \mathbf{O}$. For **any** $\mathbf{B} \in M_2(F)$, the first column of $\mathbf{BN}$ is $\mathbf{B}$ times the first column of $\mathbf{N}$, that is $\mathbf{B}\mathbf{0} = \mathbf{0}$, by [Theorem 0.9.2](#). The first column of $\mathbf{I}_2$ is $\mathbf{e}_1 \neq \mathbf{0}$, so $\mathbf{BN} \neq \mathbf{I}_2$: no matrix undoes $\mathbf{N}$ from the left. Dividing needs a genuine two-sided inverse, and not every nonzero matrix has one.

An invertible matrix is a square matrix whose effect can be undone by multiplying, on either side.

Definition 0.9.13. Invertible matrix

A square matrix $\mathbf{A} \in M_n(F)$ is **invertible** if there exists $\mathbf{B} \in M_n(F)$ such that

$$\mathbf{AB} = \mathbf{I}_n \quad \text{and} \quad \mathbf{BA} = \mathbf{I}_n.$$

Such a $\mathbf{B}$ is called **an inverse** of $\mathbf{A}$. A square matrix that is not invertible is called **singular**.

In words: $\mathbf{A}$ is invertible when some matrix $\mathbf{B}$ of the same size cancels it **from the right** ($\mathbf{AB} = \mathbf{I}_n$) **and** from the left ($\mathbf{BA} = \mathbf{I}_n$). The definition asks for one $\mathbf{B}$ doing both jobs, and it only speaks about square matrices.

Is “an inverse” really “the inverse”? Yes: an inverse, if it exists, is unique. This is part 1 of the next theorem, and it lets us write $\mathbf{A}^{-1}$. The remaining parts are what inverses are for.

Theorem 0.9.7. Properties of inverses

Let $\mathbf{A}, \mathbf{B} \in M_n(F)$.

1. **(Uniqueness)** If $\mathbf{B}$ and $\mathbf{B}'$ are both inverses of $\mathbf{A}$, then $\mathbf{B} = \mathbf{B}'$. We write $\mathbf{A}^{-1}$ for **the** inverse of an invertible $\mathbf{A}$.
2. If $\mathbf{A}$ is invertible, then so is $\mathbf{A}^{-1}$, and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
3. If $\mathbf{A}$ and $\mathbf{B}$ are invertible, then so is $\mathbf{AB}$, and $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
4. If $\mathbf{A}$ is invertible, then so is $\mathbf{A}^\top$, and $(\mathbf{A}^\top)^{-1} = (\mathbf{A}^{-1})^\top$.
5. If $\mathbf{A}$ is invertible and $c \in F$ is **nonzero**, then $c\mathbf{A}$ is invertible, and $(c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}$.
6. **(Cancellation)** If $\mathbf{A}$ is invertible, then $\mathbf{AX} = \mathbf{AY}$ implies $\mathbf{X} = \mathbf{Y}$ for all $\mathbf{X}, \mathbf{Y} \in M_{n \times p}(F)$, and $\mathbf{XA} = \mathbf{YA}$ implies $\mathbf{X} = \mathbf{Y}$ for all $\mathbf{X}, \mathbf{Y} \in M_{p \times n}(F)$.
7. **(Solving)** If $\mathbf{A}$ is invertible, then for every $\mathbf{b} \in F^n$ the equation $\mathbf{Ax} = \mathbf{b}$ has **exactly one** solution $\mathbf{x} \in F^n$, namely $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Idea. Uniqueness is the standard trick for “prove unique by assuming two”: sandwich $\mathbf{B}'$ between $\mathbf{B}$ and $\mathbf{A}$, and use associativity to collapse the product both ways. For the other parts the candidate inverse is written in the statement, so we only multiply it out on **both** sides, as the definition demands. The order in part 3 is the “socks and shoes” rule: to undo “first $\mathbf{B}$, then $\mathbf{A}$ ”, undo $\mathbf{A}$ first.

Proof. For 1, suppose $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$ and $\mathbf{AB}' = \mathbf{B}'\mathbf{A} = \mathbf{I}_n$. By [Theorem 0.9.4](#) (identity and associativity),

$$\mathbf{B} = \mathbf{BI}_n = \mathbf{B}(\mathbf{AB}') = (\mathbf{BA})\mathbf{B}' = \mathbf{I}_n\mathbf{B}' = \mathbf{B}'.$$

For 2, the equations $\mathbf{AA}^{-1} = \mathbf{I}_n$ and $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$ say precisely that $\mathbf{A}$ is an inverse of $\mathbf{A}^{-1}$.

By part 1 it is the inverse.

For 3, by associativity,

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

and similarly $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}_n$. Hence $\mathbf{B}^{-1}\mathbf{A}^{-1}$ is an inverse of $\mathbf{AB}$, and by part 1 it is $(\mathbf{AB})^{-1}$.

For 4, by [Theorem 0.9.5](#), $\mathbf{A}^\top(\mathbf{A}^{-1})^\top = (\mathbf{A}^{-1}\mathbf{A})^\top = \mathbf{I}_n^\top = \mathbf{I}_n$ and $(\mathbf{A}^{-1})^\top\mathbf{A}^\top = (\mathbf{AA}^{-1})^\top = \mathbf{I}_n$, where $\mathbf{I}_n^\top = \mathbf{I}_n$ since $\mathbf{I}_n$ is symmetric. Hence $(\mathbf{A}^{-1})^\top$ is the inverse of $\mathbf{A}^\top$.

For 5, since $c \neq 0$, the inverse c^{-1} exists in F . By [Theorem 0.9.4](#) (scalars), $(c\mathbf{A})(c^{-1}\mathbf{A}^{-1}) = (cc^{-1})(\mathbf{AA}^{-1}) = \mathbf{I}_n$, and similarly in the other order.

For 6, suppose $\mathbf{AX} = \mathbf{AY}$. Multiplying on the left by $\mathbf{A}^{-1}$ and using associativity, $\mathbf{X} = \mathbf{I}_n\mathbf{X} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{X} = \mathbf{A}^{-1}(\mathbf{AX}) = \mathbf{A}^{-1}(\mathbf{AY}) = (\mathbf{A}^{-1}\mathbf{A})\mathbf{Y} = \mathbf{Y}$. The statement for $\mathbf{XA} = \mathbf{YA}$ follows by multiplying on the right by $\mathbf{A}^{-1}$.

For 7, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ is a solution, since $\mathbf{A}(\mathbf{A}^{-1}\mathbf{b}) = (\mathbf{AA}^{-1})\mathbf{b} = \mathbf{b}$. If $\mathbf{x}$ is any solution, then $\mathbf{Ax} = \mathbf{b} = \mathbf{A}(\mathbf{A}^{-1}\mathbf{b})$, and part 6 (with $p = 1$) gives $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. This proves the theorem. ■

Part 6 is the repair of the failed cancellation law from the warnings above: **you may cancel an invertible matrix**, on the side where it stands, and nothing else.

Example. Some invertible matrices

Show that each of the following is invertible and find its inverse: $\mathbf{I}_n$; a diagonal matrix $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$ with **all** $d_i \neq 0$; $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \in M_2(\mathbb{R})$; and $\mathbf{U} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in M_2(\mathbb{F}_2)$.

Solution. $\mathbf{I}_n\mathbf{I}_n = \mathbf{I}_n$, so $\mathbf{I}_n$ is its own inverse. This is the degenerate case: $\mathbf{I}_n$ plays the role that 1 plays among numbers.

For $\mathbf{D}$, let $\mathbf{E} = \text{diag}(d_1^{-1}, \dots, d_n^{-1})$, which exists since every $d_i \neq 0$. Then $(\mathbf{DE})_{ik} = \sum_j d_{ij}e_{jk}$, and only $j = i = k$ can contribute, so $\mathbf{DE} = \text{diag}(d_1d_1^{-1}, \dots, d_nd_n^{-1}) = \mathbf{I}_n$; likewise $\mathbf{ED} = \mathbf{I}_n$. Hence $\mathbf{D}^{-1} = \mathbf{E}$. For $n = 1$ this says the 1×1 matrix (a) is invertible when $a \neq 0$, with inverse (a^{-1}) , exactly as in the field.

For $\mathbf{A}$, let $\mathbf{B} = \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix}$. Then

$$\mathbf{AB} = \begin{pmatrix} 6-5 & -2+2 \\ 15-15 & -5+6 \end{pmatrix} = \mathbf{I}_2, \quad \mathbf{BA} = \begin{pmatrix} 6-5 & 3-3 \\ -10+10 & -5+6 \end{pmatrix} = \mathbf{I}_2,$$

so $\mathbf{A}^{-1} = \mathbf{B}$. (Where $\mathbf{B}$ came from is explained by [Theorem 0.9.8](#) below.)

For $\mathbf{U}$ over $\mathbb{F}_2$, $\mathbf{U}^2 = \begin{pmatrix} 1 & 1+1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, since $1+1=0$. So $\mathbf{U}$ is its own inverse. Over

$\mathbb{R}$ the same array would have inverse $\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$: the answer depends on the field.

Now a nonexample by minimal change. The matrix $\text{diag}(1, 2)$ is invertible, with inverse $\text{diag}(1, \frac{1}{2})$. Change the entry 2 to 0: $\mathbf{A} = \text{diag}(1, 0)$ is still diagonal and still nonzero. But for any $\mathbf{B} \in M_2(F)$, the second column of $\mathbf{BA}$ is $\mathbf{B}$ times the second column of $\mathbf{A}$, which is $\mathbf{B}\mathbf{0} = \mathbf{0}$, while the second column of $\mathbf{I}_2$ is $\mathbf{e}_2$. So $\mathbf{BA} \neq \mathbf{I}_2$ for every $\mathbf{B}$: the clause “**there exists B** with ... $\mathbf{BA} = \mathbf{I}_n$ ” fails, and $\text{diag}(1, 0)$ is singular. The same argument shows that any square matrix with a zero column is singular, and in particular so is every zero matrix $\mathbf{O}_{n \times n}$.

Why does the definition insist on a **square** matrix and on **both** equations? For nonsquare matrices one-sided inverses can exist without the other side: with $\mathbf{C} = \begin{pmatrix} 1 & 0 \end{pmatrix} \in M_{1 \times 2}(\mathbb{R})$ and $\mathbf{D} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, we get $\mathbf{CD} = \begin{pmatrix} 1 \end{pmatrix} = \mathbf{I}_1$ but $\mathbf{DC} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \mathbf{I}_2$. Demanding both equations makes the inverse unique (part 1) and makes both cancellation laws (part 6) hold.

Warning. For square matrices, $\mathbf{AB} = \mathbf{I}_n$ already implies $\mathbf{BA} = \mathbf{I}_n$. This is true, but it is not obvious, and we do not know it yet. Nothing in this chapter proves it: the entries of $\mathbf{AB}$ and $\mathbf{BA}$ are different sums. It is proved in Chapter 3, using row reduction. Until then, to show that $\mathbf{B}$ is the inverse of $\mathbf{A}$, check **both** $\mathbf{AB} = \mathbf{I}_n$ and $\mathbf{BA} = \mathbf{I}_n$. (The nonsquare example above shows that some argument is needed.)

For 2×2 matrices, invertibility is decided by one number.

Theorem 0.9.8. Inverse of a 2×2 matrix

Let $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(F)$. Then $\mathbf{A}$ is invertible **if and only if** $ad - bc \neq 0$. In that case,

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Idea. Where does the formula come from? Try to kill the off-diagonal entries of $\mathbf{A}$ by multiplying with a matrix built from its own entries. Swapping a and d and negating b and c gives $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = (ad - bc)\mathbf{I}_2$. If $ad - bc \neq 0$, divide by it. If $ad - bc = 0$, the same computation shows that $\mathbf{A}$ sends a nonzero column to $\mathbf{0}$, and an invertible matrix cannot do that, since its inverse would have to send $\mathbf{0}$ back to a nonzero column.

Proof. Write $\Delta = ad - bc$ and $\tilde{\mathbf{A}} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. By [Definition 0.9.9](#) and commutativity of

multiplication in F ,

$$\begin{aligned}\mathbf{A}\tilde{\mathbf{A}} &= \begin{pmatrix} ad - bc & -ab + ba \\ cd - dc & -cb + da \end{pmatrix} = \Delta \mathbf{I}_2, \\ \tilde{\mathbf{A}}\mathbf{A} &= \begin{pmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{pmatrix} = \Delta \mathbf{I}_2.\end{aligned}$$

($\Leftarrow$) Suppose $\Delta \neq 0$, so Δ^{-1} exists in F . By [Theorem 0.9.4](#) (scalars), $\mathbf{A}(\Delta^{-1}\tilde{\mathbf{A}}) = \Delta^{-1}(\mathbf{A}\tilde{\mathbf{A}}) = \Delta^{-1}\Delta\mathbf{I}_2 = \mathbf{I}_2$, and likewise $(\Delta^{-1}\tilde{\mathbf{A}})\mathbf{A} = \mathbf{I}_2$. Hence $\mathbf{A}$ is invertible with $\mathbf{A}^{-1} = \Delta^{-1}\tilde{\mathbf{A}}$, which is the stated formula.

($\Rightarrow$) We prove the contrapositive. Suppose $\Delta = 0$. We first find $\mathbf{x} \in F^2$ with $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{A}\mathbf{x} = \mathbf{0}$.

Case 1. $\mathbf{A} = \mathbf{O}$. Take $\mathbf{x} = \mathbf{e}_1$.

Case 2. $c \neq 0$ or $d \neq 0$. Take $\mathbf{x} = (d, -c) \neq \mathbf{0}$. Then $\mathbf{A}\mathbf{x} = (ad - bc, cd - dc) = (\Delta, 0) = \mathbf{0}$.

Case 3. $c = d = 0$ but $a \neq 0$ or $b \neq 0$. Take $\mathbf{x} = (b, -a) \neq \mathbf{0}$. Then $\mathbf{A}\mathbf{x} = (ab - ba, cb - da) = (0, 0) = \mathbf{0}$, since $c = d = 0$.

These cases cover every $\mathbf{A}$. Now suppose, for a contradiction, that $\mathbf{A}$ is invertible. Then by [Theorem 0.9.4](#),

$$\mathbf{x} = \mathbf{I}_2\mathbf{x} = (\mathbf{A}^{-1}\mathbf{A})\mathbf{x} = \mathbf{A}^{-1}(\mathbf{A}\mathbf{x}) = \mathbf{A}^{-1}\mathbf{0} = \mathbf{0},$$

contradicting $\mathbf{x} \neq \mathbf{0}$. Hence $\mathbf{A}$ is not invertible. This proves the theorem. ■

The number $ad - bc$ will return in Chapter 7 as the determinant of $\mathbf{A}$. Here it arose from a single multiplication, and the proof used nothing beyond the field axioms, so the theorem holds over every field. For instance, $\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}$ has $ad - bc = 6 - 5 = 1$, which explains the inverse found in [Example \(Some invertible matrices\)](#); and $\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ has $ad - bc = 0$, so it is singular, with $\mathbf{A}(2, -1) = \mathbf{0}$.

0.9.10 Powers and polynomials of a matrix

A square matrix can be multiplied by itself, so we can form powers, and then plug the matrix into a polynomial (Definition 0.8.1). This is how polynomials act on matrices throughout the book.

Definition 0.9.14. Powers and polynomials of a matrix

Let $\mathbf{A} \in M_n(F)$. The **powers** of $\mathbf{A}$ are defined recursively by

$$\mathbf{A}^0 = \mathbf{I}_n, \quad \mathbf{A}^{k+1} = \mathbf{A}^k \mathbf{A} \quad (k \in \mathbb{N}).$$

For a polynomial $p = a_0 + a_1x + \cdots + a_Nx^N \in F[x]$, the matrix $p(\mathbf{A})$ is

$$p(\mathbf{A}) = a_0\mathbf{I}_n + a_1\mathbf{A} + a_2\mathbf{A}^2 + \cdots + a_N\mathbf{A}^N \in M_n(F).$$

Two points in this definition deserve attention. The **constant term** a_0 becomes $a_0 \mathbf{I}_n$, because a_0 alone is a number and cannot be added to matrices; this matches $x^0 = 1$ with $\mathbf{A}^0 = \mathbf{I}_n$. And $p(\mathbf{A})$ is well defined, because enlarging N only adds terms $0 \cdot \mathbf{A}^k = \mathbf{O}$, exactly as for evaluation at a number (Definition 0.8.6).

The next theorem says that plugging in a matrix respects sums and products, just as plugging in a number does (Theorem 0.8.5).

Theorem 0.9.9. Polynomials in one matrix

Let $\mathbf{A} \in M_n(F)$ and $p, q \in F[x]$. Then:

1. $\mathbf{A}^j \mathbf{A}^k = \mathbf{A}^{j+k}$ for all $j, k \in \mathbb{N}$;
2. $(p + q)(\mathbf{A}) = p(\mathbf{A}) + q(\mathbf{A})$ and $(pq)(\mathbf{A}) = p(\mathbf{A}) q(\mathbf{A})$;
3. $p(\mathbf{A}) q(\mathbf{A}) = q(\mathbf{A}) p(\mathbf{A})$.

Idea. In the proof of Theorem 0.8.5, the only delicate step moved c^i past a coefficient b_j . Here the coefficients are scalars, and scalars can be moved freely through matrix products. The other step, $c^i c^j = c^{i+j}$, becomes part 1, which is associativity plus induction. Part 3 is then free, because $pq = qp$ in $F[x]$.

Proof. For 1, fix j and use induction on k (Theorem 0.2.7). For $k = 0$, $\mathbf{A}^j \mathbf{A}^0 = \mathbf{A}^j \mathbf{I}_n = \mathbf{A}^j$. If $\mathbf{A}^j \mathbf{A}^k = \mathbf{A}^{j+k}$, then by Definition 0.9.14 and associativity (Theorem 0.9.4),

$$\mathbf{A}^j \mathbf{A}^{k+1} = \mathbf{A}^j (\mathbf{A}^k \mathbf{A}) = (\mathbf{A}^j \mathbf{A}^k) \mathbf{A} = \mathbf{A}^{j+k} \mathbf{A} = \mathbf{A}^{j+k+1}.$$

For 2, let $p = \sum_{i=0}^m a_i x^i$ and $q = \sum_{j=0}^l b_j x^j$. The statement for sums follows from Theorem 0.9.1, since $(a_i + b_i) \mathbf{A}^i = a_i \mathbf{A}^i + b_i \mathbf{A}^i$. For products, by Definition 0.8.2 the coefficient of x^k in pq is $\sum_{i+j=k} a_i b_j$, so

$$\begin{aligned} (pq)(\mathbf{A}) &= \sum_k \left(\sum_{i+j=k} a_i b_j \right) \mathbf{A}^k = \sum_{i=0}^m \sum_{j=0}^l (a_i b_j) \mathbf{A}^i \mathbf{A}^j \\ &= \left(\sum_{i=0}^m a_i \mathbf{A}^i \right) \left(\sum_{j=0}^l b_j \mathbf{A}^j \right) = p(\mathbf{A}) q(\mathbf{A}), \end{aligned}$$

where the second equality uses part 1 and Theorem 0.9.1, and the third uses distributivity and the scalar rule $(a_i \mathbf{A}^i)(b_j \mathbf{A}^j) = (a_i b_j) \mathbf{A}^i \mathbf{A}^j$ from Theorem 0.9.4.

For 3, $pq = qp$ in $F[x]$ by Theorem 0.8.1. Hence by part 2, $p(\mathbf{A}) q(\mathbf{A}) = (pq)(\mathbf{A}) = (qp)(\mathbf{A}) = q(\mathbf{A}) p(\mathbf{A})$. This proves the theorem. ■

Example. Plugging a matrix into a polynomial

Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \in M_2(\mathbb{R})$ and $p = x^2 - 4x + 3$. Compute $p(\mathbf{A})$ directly, and again using the

factorization $p = (x - 1)(x - 3)$.

Solution. First $\mathbf{A}^2 = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 8 \\ 0 & 9 \end{pmatrix}$. Hence

$$p(\mathbf{A}) = \mathbf{A}^2 - 4\mathbf{A} + 3\mathbf{I}_2 = \begin{pmatrix} 1 & 8 \\ 0 & 9 \end{pmatrix} - \begin{pmatrix} 4 & 8 \\ 0 & 12 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The constant term contributed $3\mathbf{I}_2$, not “3”. By [Theorem 0.9.9](#), $p(\mathbf{A}) = (\mathbf{A} - \mathbf{I}_2)(\mathbf{A} - 3\mathbf{I}_2)$, and indeed

$$(\mathbf{A} - \mathbf{I}_2)(\mathbf{A} - 3\mathbf{I}_2) = \begin{pmatrix} 0 & 2 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} -2 & 2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that neither factor is zero; this is the second warning of the multiplication part in action. That the diagonal entries 1, 3 of $\mathbf{A}$ are the roots of p is no accident, as we will see with eigenvalues.

[Theorem 0.9.9](#) is about polynomials in **one** matrix. Identities with **two** different matrices are another matter.

Warning. School identities fail for matrices that do not commute. Expanding with distributivity, $(\mathbf{A} + \mathbf{B})^2 = \mathbf{A}^2 + \mathbf{AB} + \mathbf{BA} + \mathbf{B}^2$, and this equals $\mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2$ only when $\mathbf{AB} = \mathbf{BA}$. With $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$: $(\mathbf{A} + \mathbf{B})^2 = \mathbf{I}_2$, but $\mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$. Before using a polynomial identity valid for numbers, check that all the matrices involved commute; polynomials in a single $\mathbf{A}$ always do.

0.9.11 Two equivalence relations on matrices

We close with two relations on matrices that the rest of the book studies in depth. Both are equivalence relations (Definition 0.5.2), and the proofs are short exercises in using [Theorem 0.9.7](#).

Example. Similarity

For $\mathbf{A}, \mathbf{B} \in M_n(F)$, say $\mathbf{A}$ is **similar** to $\mathbf{B}$, written $\mathbf{A} \sim \mathbf{B}$, if there is an **invertible** $\mathbf{P} \in M_n(F)$ with $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. Prove that similarity is an equivalence relation on $M_n(F)$. Then check that $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ is similar to $\text{diag}(1, 3)$ using $\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, and show that $\text{diag}(1, 2)$ is **not** similar to $\mathbf{I}_2$ over $\mathbb{R}$.

Solution. We check the three conditions of Definition 0.5.2.

Reflexive. Let $\mathbf{A} \in M_n(F)$. The identity $\mathbf{I}_n$ is invertible with $\mathbf{I}_n^{-1} = \mathbf{I}_n$ ([Example \(Some invertible matrices\)](#)), and $\mathbf{I}_n^{-1}\mathbf{A}\mathbf{I}_n = \mathbf{A}$. Hence $\mathbf{A} \sim \mathbf{A}$.

Symmetric. Suppose $\mathbf{A} \sim \mathbf{B}$, so $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ with $\mathbf{P}$ invertible. Multiplying on the left

by $\mathbf{P}$ and on the right by $\mathbf{P}^{-1}$, and using associativity, $\mathbf{PBP}^{-1} = (\mathbf{PP}^{-1})\mathbf{A}(\mathbf{PP}^{-1}) = \mathbf{A}$. By [Theorem 0.9.7](#) (part 2), $\mathbf{Q} = \mathbf{P}^{-1}$ is invertible with $\mathbf{Q}^{-1} = \mathbf{P}$, so $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{BQ}$. Hence $\mathbf{B} \sim \mathbf{A}$.

Transitive. Suppose $\mathbf{A} \sim \mathbf{B}$ and $\mathbf{B} \sim \mathbf{C}$, so $\mathbf{B} = \mathbf{P}^{-1}\mathbf{AP}$ and $\mathbf{C} = \mathbf{Q}^{-1}\mathbf{BQ}$ with $\mathbf{P}, \mathbf{Q}$ invertible. Then

$$\mathbf{C} = \mathbf{Q}^{-1}(\mathbf{P}^{-1}\mathbf{AP})\mathbf{Q} = (\mathbf{Q}^{-1}\mathbf{P}^{-1})\mathbf{A}(\mathbf{PQ}) = (\mathbf{PQ})^{-1}\mathbf{A}(\mathbf{PQ}),$$

where the last step is [Theorem 0.9.7](#) (part 3), which also says $\mathbf{PQ}$ is invertible. Hence $\mathbf{A} \sim \mathbf{C}$. This shows that similarity is an equivalence relation.

For the numerical part, $\mathbf{P}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$ by [Theorem 0.9.8](#) (here $ad - bc = 1$), and

$$\mathbf{P}^{-1}\mathbf{AP} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}.$$

Finally, if $\mathbf{P}^{-1}\mathbf{I}_2\mathbf{P} = \text{diag}(1, 2)$ for some invertible $\mathbf{P}$, then $\text{diag}(1, 2) = \mathbf{P}^{-1}\mathbf{P} = \mathbf{I}_2$, which is false. (Alternatively: similar matrices have equal traces by [Theorem 0.9.6](#), and $\text{tr } \text{diag}(1, 2) = 3 \neq 2 = \text{tr } \mathbf{I}_2$.) Hence $\text{diag}(1, 2) \not\sim \mathbf{I}_2$.

In Chapter 4 we will see that similar matrices describe the same linear map in two different coordinate systems. That is why a quantity like the trace, which similar matrices share, is really a property of the map and not of the matrix. Deciding when two matrices are similar is one of the main problems of this book.

Example. Row equivalence

For $\mathbf{A}, \mathbf{B} \in M_{m \times n}(F)$, say $\mathbf{A}$ is **row equivalent** to $\mathbf{B}$ if there is an **invertible** $\mathbf{E} \in M_m(F)$ with $\mathbf{B} = \mathbf{EA}$. Prove that row equivalence is an equivalence relation on $M_{m \times n}(F)$. Then, with $\mathbf{E} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$, compute $\mathbf{EA}$ for $\mathbf{A} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 0 \end{pmatrix}$ and describe what $\mathbf{E}$ did to the rows of $\mathbf{A}$.

Solution. Reflexive. $\mathbf{A} = \mathbf{I}_m\mathbf{A}$ and $\mathbf{I}_m$ is invertible, so $\mathbf{A}$ is row equivalent to itself.

Symmetric. Suppose $\mathbf{B} = \mathbf{EA}$ with $\mathbf{E}$ invertible. Multiplying on the left by $\mathbf{E}^{-1}$, $\mathbf{E}^{-1}\mathbf{B} = (\mathbf{E}^{-1}\mathbf{E})\mathbf{A} = \mathbf{A}$, and $\mathbf{E}^{-1}$ is invertible by [Theorem 0.9.7](#) (part 2). So $\mathbf{B}$ is row equivalent to $\mathbf{A}$.

Transitive. Suppose $\mathbf{B} = \mathbf{EA}$ and $\mathbf{C} = \mathbf{E}'\mathbf{B}$ with $\mathbf{E}, \mathbf{E}'$ invertible. Then $\mathbf{C} = \mathbf{E}'(\mathbf{EA}) = (\mathbf{E}'\mathbf{E})\mathbf{A}$ by associativity, and $\mathbf{E}'\mathbf{E}$ is invertible by [Theorem 0.9.7](#) (part 3). So $\mathbf{A}$ is row equivalent to $\mathbf{C}$. This shows that row equivalence is an equivalence relation.

For the computation, $\mathbf{E}$ is invertible, with inverse $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ (check both products, or use [Theorem 0.9.8](#) with $ad - bc = 1$). By the row view of [Theorem 0.9.2](#) and [Theorem 0.9.3](#), row

1 of $\mathbf{EA}$ is $1 \cdot (\text{row 1 of } \mathbf{A})$ and row 2 is $-2 \cdot (\text{row 1}) + 1 \cdot (\text{row 2})$:

$$\mathbf{EA} = \begin{pmatrix} 1 & 2 & 1 \\ 2-2 & 5-4 & 0-2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -2 \end{pmatrix}.$$

So $\mathbf{E}$ subtracted twice row 1 from row 2, which is the first step of Gaussian elimination on $\mathbf{A}$.

In Chapter 3 we will see that every elementary row operation is multiplication on the left by an invertible matrix, and that row equivalence is exactly “reachable by a sequence of row operations”. This is why Gaussian elimination never changes the solutions of a linear system.

0.9.12 Exercises

0.9.12.1 A. Check your understanding

Exercise. A1

- Let $\mathbf{A} \in M_{2 \times 3}(F)$ and $\mathbf{B} \in M_{3 \times 4}(F)$. Which of $\mathbf{AB}$, $\mathbf{BA}$, $\mathbf{A}^\top \mathbf{A}$, and $\mathbf{B}^\top \mathbf{A}^\top$ are defined? Give the size of each one that is.
- Define what it means for $\mathbf{A} \in M_n(F)$ to be invertible.
- Determine whether the following statement is true: “if $\mathbf{A}, \mathbf{B} \in M_n(F)$ and $\mathbf{AB} = \mathbf{O}$, then $\mathbf{A} = \mathbf{O}$ or $\mathbf{B} = \mathbf{O}$.” Justify your answer.
- Determine whether the following statement is true: “ $(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B}) = \mathbf{A}^2 - \mathbf{B}^2$ for all $\mathbf{A}, \mathbf{B} \in M_2(\mathbb{R})$.” Justify your answer.
- Let $\mathbf{A} \in M_3(F)$ have columns $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$. Express $\mathbf{A}(2, 0, -1)$ and the second column of $\mathbf{A}^2$ in terms of the columns of $\mathbf{A}$ and $\mathbf{A}$ itself.

Solution.

- $\mathbf{AB}$ is $(2 \times 3)(3 \times 4)$, defined, of size 2×4 . $\mathbf{BA}$ would pair rows of length 4 with columns of length 2, so it is not defined. $\mathbf{A}^\top \mathbf{A}$ is $(3 \times 2)(2 \times 3)$, defined, of size 3×3 . $\mathbf{B}^\top \mathbf{A}^\top$ is $(4 \times 3)(3 \times 2)$, defined, of size 4×2 ; it equals $(\mathbf{AB})^\top$ by [Theorem 0.9.5](#).
- $\mathbf{A}$ is invertible if there exists $\mathbf{B} \in M_n(F)$ with $\mathbf{AB} = \mathbf{I}_n$ and $\mathbf{BA} = \mathbf{I}_n$.
- False. $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \mathbf{O}$ with both factors nonzero.
- False. Expanding gives $(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B}) = \mathbf{A}^2 - \mathbf{AB} + \mathbf{BA} - \mathbf{B}^2$, which differs from $\mathbf{A}^2 - \mathbf{B}^2$ unless $\mathbf{AB} = \mathbf{BA}$. For $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, we have $\mathbf{A}^2 = \mathbf{B}^2 = \mathbf{O}$, but $(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \neq \mathbf{O}$.
- By [Theorem 0.9.3](#), $\mathbf{A}(2, 0, -1) = 2\mathbf{a}_1 - \mathbf{a}_3$. By [Theorem 0.9.2](#), the second column of $\mathbf{A}^2 = \mathbf{AA}$ is $\mathbf{Aa}_2$.

0.9.12.2 B. Practice

Exercise. B1: Products and sizes

Let

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 3 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \\ 0 & 4 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

over $\mathbb{R}$.

- Compute $\mathbf{AB}$ and $\mathbf{BA}$.
- Determine which of $\mathbf{AC}$ and $\mathbf{CA}$ is defined, and compute it.
- Compute $\mathbf{Ax}$ for $\mathbf{x} = (2, 0, -1)$ as a combination of the columns of $\mathbf{A}$.

Solution.

- $\mathbf{AB}$ is 2×2 and $\mathbf{BA}$ is 3×3 . Row against column,

$$\mathbf{AB} = \begin{pmatrix} 2 - 1 + 0 & 1 + 0 + 8 \\ 0 + 3 + 0 & 0 + 0 + 4 \end{pmatrix} = \begin{pmatrix} 1 & 9 \\ 3 & 4 \end{pmatrix}, \quad \mathbf{BA} = \begin{pmatrix} 2 & 1 & 5 \\ 1 & -1 & 2 \\ 0 & 12 & 4 \end{pmatrix}.$$

For instance $(\mathbf{BA})_{12} = 2 \cdot (-1) + 1 \cdot 3 = 1$ and $(\mathbf{BA})_{32} = 0 \cdot (-1) + 4 \cdot 3 = 12$.

- $\mathbf{AC}$ would be $(2 \times 3)(2 \times 2)$, and $3 \neq 2$, so it is not defined. $\mathbf{CA}$ is $(2 \times 2)(2 \times 3)$, of size 2×3 :

$$\mathbf{CA} = \begin{pmatrix} 1 & -1 + 6 & 2 + 2 \\ 3 & -3 + 12 & 6 + 4 \end{pmatrix} = \begin{pmatrix} 1 & 5 & 4 \\ 3 & 9 & 10 \end{pmatrix}.$$

- By [Theorem 0.9.3](#), $\mathbf{Ax} = 2\mathbf{a}_1 + 0\mathbf{a}_2 - \mathbf{a}_3 = 2(1, 0) - (2, 1) = (0, -1)$.

Exercise. B2: Traces of products

- With $\mathbf{A}$ and $\mathbf{B}$ as in [Exercise \(B1: Products and sizes\)](#), verify that $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, although $\mathbf{AB}$ and $\mathbf{BA}$ have different sizes.
- Let $\mathbf{A} = \text{diag}(1, 2)$, $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, and $\mathbf{C} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. Show that $\text{tr}(\mathbf{ABC}) \neq \text{tr}(\mathbf{BAC})$, and verify that $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB})$.

Solution.

- From [Exercise \(B1: Products and sizes\)](#), $\text{tr}(\mathbf{AB}) = 1 + 4 = 5$ and $\text{tr}(\mathbf{BA}) = 2 + (-1) + 4 = 5$, as [Theorem 0.9.6](#) predicts.
- By the column view, $\mathbf{BC}$ has columns $\mathbf{Be}_2 = \mathbf{e}_1$ and $\mathbf{B0} = \mathbf{0}$, so $\mathbf{BC} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, and $\mathbf{ABC} = \mathbf{A}(\mathbf{BC}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, with trace 1. Next, $\mathbf{BA} = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$, so $\mathbf{BAC} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$, with trace 2. Hence $\text{tr}(\mathbf{ABC}) = 1 \neq 2 = \text{tr}(\mathbf{BAC})$. Finally $\mathbf{AB} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and

$$\mathbf{CAB} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \text{ with trace } 1 = \text{tr}(\mathbf{ABC}), \text{ as the cyclic rule predicts.}$$

Exercise. B3: Inverses of 2×2 matrices

- (a) Find the inverse of $\begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix} \in M_2(\mathbb{R})$, and check both products.
- (b) Show that $\mathbf{A} = \begin{pmatrix} 2 & 6 \\ 1 & 3 \end{pmatrix}$ is not invertible by finding a nonzero $\mathbf{x}$ with $\mathbf{Ax} = \mathbf{0}$.
- (c) Find the inverse of $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in M_2(\mathbb{F}_5)$.
- (d) For which $t \in \mathbb{R}$ is $\begin{pmatrix} t & 2 \\ 2 & t \end{pmatrix}$ invertible?

Solution.

- (a) Here $ad - bc = 9 - 8 = 1 \neq 0$, so by [Theorem 0.9.8](#) the inverse is $\mathbf{B} = \begin{pmatrix} 3 & -4 \\ -2 & 3 \end{pmatrix}$. Check:
- $$\begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix} \mathbf{B} = \begin{pmatrix} 9 - 8 & -12 + 12 \\ 6 - 6 & -8 + 9 \end{pmatrix} = \mathbf{I}_2 \text{ and } \mathbf{B} \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 9 - 8 & 12 - 12 \\ -6 + 6 & -8 + 9 \end{pmatrix} = \mathbf{I}_2.$$
- (b) Here $ad - bc = 6 - 6 = 0$. Following the proof of [Theorem 0.9.8](#), take $\mathbf{x} = (d, -c) = (3, -1) \neq \mathbf{0}$. Then $\mathbf{Ax} = (6 - 6, 3 - 3) = \mathbf{0}$. If $\mathbf{A}$ were invertible, then $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{Ax}) = \mathbf{0}$, a contradiction. Hence $\mathbf{A}$ is not invertible.
- (c) In $\mathbb{F}_5$, $ad - bc = 4 - 6 = -2 = 3 \neq 0$, and $3^{-1} = 2$ since $3 \cdot 2 = 6 = 1$. By [Theorem 0.9.8](#) the inverse is $2 \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} 8 & -4 \\ -6 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 4 & 2 \end{pmatrix}$. Check: $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 4 & 2 \end{pmatrix} = \begin{pmatrix} 11 & 5 \\ 25 & 11 \end{pmatrix} = \mathbf{I}_2$ and $\begin{pmatrix} 3 & 1 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 6 & 10 \\ 10 & 16 \end{pmatrix} = \mathbf{I}_2$ in $\mathbb{F}_5$.
- (d) By [Theorem 0.9.8](#), it is invertible if and only if $t^2 - 4 \neq 0$, that is, $t \neq 2$ and $t \neq -2$.

Exercise. B4: Transposes of products

- (a) Prove that $(\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_k)^\top = \mathbf{A}_k^\top \cdots \mathbf{A}_2^\top \mathbf{A}_1^\top$ for all $k \geq 1$ and matrices $\mathbf{A}_1, \dots, \mathbf{A}_k$ whose sizes allow the product.
- (b) Let $\mathbf{A} \in M_n(F)$. Prove that $\mathbf{A} + \mathbf{A}^\top$ is symmetric and $\mathbf{A} - \mathbf{A}^\top$ is skew-symmetric.

Solution.

- (a) We use induction on k ([Theorem 0.2.7](#)). For $k = 1$ there is nothing to prove. Suppose the formula holds for products of k matrices, and let $\mathbf{A}_1, \dots, \mathbf{A}_{k+1}$ be given. By associativity, $\mathbf{A}_1 \cdots \mathbf{A}_{k+1} = (\mathbf{A}_1 \cdots \mathbf{A}_k) \mathbf{A}_{k+1}$. By [Theorem 0.9.5](#) (part 4) and then the induction hypothesis,

$$(\mathbf{A}_1 \cdots \mathbf{A}_{k+1})^\top = \mathbf{A}_{k+1}^\top (\mathbf{A}_1 \cdots \mathbf{A}_k)^\top = \mathbf{A}_{k+1}^\top \mathbf{A}_k^\top \cdots \mathbf{A}_1^\top.$$

This proves the formula for all $k \geq 1$.

- (b) By [Theorem 0.9.5](#) (parts 1–3), $(\mathbf{A} + \mathbf{A}^\top)^\top = \mathbf{A}^\top + (\mathbf{A}^\top)^\top = \mathbf{A}^\top + \mathbf{A} = \mathbf{A} + \mathbf{A}^\top$, so $\mathbf{A} + \mathbf{A}^\top$ is symmetric. Similarly $(\mathbf{A} - \mathbf{A}^\top)^\top = \mathbf{A}^\top - \mathbf{A} = -(\mathbf{A} - \mathbf{A}^\top)$, so $\mathbf{A} - \mathbf{A}^\top$ is skew-symmetric.

0.9.12.3 C. Going deeper

Exercise. C1: $\mathbf{AB} - \mathbf{BA} = \mathbf{I}$ is impossible, or is it?

- (a) Prove that there are no $\mathbf{A}, \mathbf{B} \in M_n(\mathbb{R})$, for any $n \geq 1$, with $\mathbf{AB} - \mathbf{BA} = \mathbf{I}_n$.
- (b) Show that the argument of (a) breaks down in $M_2(\mathbb{F}_2)$, and that $\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ satisfy $\mathbf{AB} - \mathbf{BA} = \mathbf{I}_2$ over $\mathbb{F}_2$.

Hint: consider the trace of both sides.

Solution.

- (a) Suppose, for a contradiction, that $\mathbf{AB} - \mathbf{BA} = \mathbf{I}_n$. Taking traces and using [Theorem 0.9.6](#) (parts 1 and 3),

$$n = \text{tr } \mathbf{I}_n = \text{tr}(\mathbf{AB} - \mathbf{BA}) = \text{tr}(\mathbf{AB}) - \text{tr}(\mathbf{BA}) = 0.$$

This contradicts $n \geq 1$. Hence no such $\mathbf{A}, \mathbf{B}$ exist.

- (b) Over $\mathbb{F}_2$ the same computation gives $\text{tr } \mathbf{I}_2 = 1 + 1 = 0$, so taking traces yields $0 = 0$ and no contradiction. For the given matrices, the column view gives $\mathbf{AB} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ and $\mathbf{BA} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, so $\mathbf{AB} - \mathbf{BA} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}_2$, because $-1 = 1$ in $\mathbb{F}_2$.

Exercise. C2: Triangular matrices with nonzero diagonal

Let F be a field.

- (a) Let $\mathbf{U} = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ with $a, d \neq 0$. Show that $\mathbf{U}$ is invertible and that $\mathbf{U}^{-1}$ is upper triangular.

- (b) Let $\mathbf{U} = \begin{pmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{pmatrix}$ with $a, d, f \neq 0$. Show that

$$\mathbf{V} = \begin{pmatrix} a^{-1} & -b(ad)^{-1} & (be - cd)(adf)^{-1} \\ 0 & d^{-1} & -e(df)^{-1} \\ 0 & 0 & f^{-1} \end{pmatrix}$$

is the inverse of $\mathbf{U}$.

- (c) Explain how the entries of $\mathbf{V}$ can be found, column by column, by solving $\mathbf{U}\mathbf{v}_k = \mathbf{e}_k$.

The general case, for all n , follows from the methods of Chapter 3.

Solution.

- (a) Here $ad - b \cdot 0 = ad \neq 0$, since a product of nonzero elements of a field is nonzero (Theorem 0.7.1). By Theorem 0.9.8, $\mathbf{U}$ is invertible with $\mathbf{U}^{-1} = (ad)^{-1} \begin{pmatrix} d & -b \\ 0 & a \end{pmatrix} = \begin{pmatrix} a^{-1} & -b(ad)^{-1} \\ 0 & d^{-1} \end{pmatrix}$, which is upper triangular.
- (b) The inverses a^{-1}, d^{-1}, f^{-1} exist, and $ad, df, adf \neq 0$ by Theorem 0.7.1, so $\mathbf{V}$ is defined. We check $\mathbf{UV} = \mathbf{I}_3$ entry by entry. The diagonal entries are $a \cdot a^{-1} = 1$, $d \cdot d^{-1} = 1$, $f \cdot f^{-1} = 1$. Below the diagonal both matrices are zero, so $(\mathbf{UV})_{21} = (\mathbf{UV})_{31} = (\mathbf{UV})_{32} = 0$. Above the diagonal,

$$\begin{aligned} (\mathbf{UV})_{12} &= a \cdot (-b(ad)^{-1}) + bd^{-1} = -bd^{-1} + bd^{-1} = 0, \\ (\mathbf{UV})_{23} &= d \cdot (-e(df)^{-1}) + ef^{-1} = -ef^{-1} + ef^{-1} = 0, \\ (\mathbf{UV})_{13} &= a(be - cd)(adf)^{-1} + b(-e(df)^{-1}) + cf^{-1} \\ &= (be - cd)(df)^{-1} - be(df)^{-1} + cd(df)^{-1} = 0. \end{aligned}$$

Hence $\mathbf{UV} = \mathbf{I}_3$. For $\mathbf{VU}$, again the diagonal entries are 1 and the entries below the diagonal are 0, and

$$\begin{aligned} (\mathbf{VU})_{12} &= a^{-1}b - b(ad)^{-1}d = a^{-1}b - a^{-1}b = 0, \\ (\mathbf{VU})_{23} &= d^{-1}e - e(df)^{-1}f = 0, \\ (\mathbf{VU})_{13} &= a^{-1}c - b(ad)^{-1}e + (be - cd)(adf)^{-1}f \\ &= a^{-1}c - be(ad)^{-1} + be(ad)^{-1} - cd(ad)^{-1} = 0, \end{aligned}$$

using $cd(ad)^{-1} = ca^{-1}$. Hence $\mathbf{VU} = \mathbf{I}_3$, and $\mathbf{V} = \mathbf{U}^{-1}$. Both products had to be checked, since we may not yet use that one implies the other.

- (c) By Theorem 0.9.2, $\mathbf{UV} = \mathbf{I}_3$ says exactly that the columns $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ of $\mathbf{V}$ satisfy $\mathbf{U}\mathbf{v}_k = \mathbf{e}_k$. Because $\mathbf{U}$ is upper triangular, each system can be solved from the bottom row up. For example, $\mathbf{U}\mathbf{v}_3 = \mathbf{e}_3$ reads $fx_3 = 1$, $dx_2 + ex_3 = 0$, $ax_1 + bx_2 + cx_3 = 0$, giving $x_3 = f^{-1}$, then $x_2 = -e(df)^{-1}$, then $x_1 = (be - cd)(adf)^{-1}$, which is the third column of $\mathbf{V}$. (This only finds a candidate satisfying $\mathbf{UV} = \mathbf{I}_3$; the check $\mathbf{VU} = \mathbf{I}_3$ in (b) is still needed.)

Exercise. C3: Matrices that commute

Let $\mathbf{N} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in M_2(\mathbb{R})$.

- (a) Find all $\mathbf{X} \in M_2(\mathbb{R})$ with $\mathbf{XN} = \mathbf{NX}$.
- (b) Deduce that the only matrices $\mathbf{X} \in M_2(\mathbb{R})$ with $\mathbf{XY} = \mathbf{YX}$ for every $\mathbf{Y} \in M_2(\mathbb{R})$ are the scalar multiples $c\mathbf{I}_2$, $c \in \mathbb{R}$.

Solution.

- (a) Let $\mathbf{X} = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$. By the column view, $\mathbf{X}\mathbf{N} = (\mathbf{X}\mathbf{0} \quad \mathbf{X}\mathbf{e}_1) = \begin{pmatrix} 0 & p \\ 0 & r \end{pmatrix}$, and by the row view, $\mathbf{N}\mathbf{X}$ has rows (row 2 of $\mathbf{X}$) and $\mathbf{0}$: $\mathbf{N}\mathbf{X} = \begin{pmatrix} r & s \\ 0 & 0 \end{pmatrix}$. Comparing entries, $\mathbf{X}\mathbf{N} = \mathbf{N}\mathbf{X}$ if and only if $r = 0$ and $p = s$, that is, $\mathbf{X} = \begin{pmatrix} p & q \\ 0 & p \end{pmatrix} = p\mathbf{I}_2 + q\mathbf{N}$ with $p, q \in \mathbb{R}$ arbitrary.
- (b) Every $c\mathbf{I}_2$ commutes with every $\mathbf{Y}$, by [Theorem 0.9.4](#): $(c\mathbf{I}_2)\mathbf{Y} = c\mathbf{Y} = \mathbf{Y}(c\mathbf{I}_2)$. Conversely, suppose $\mathbf{X}\mathbf{Y} = \mathbf{Y}\mathbf{X}$ for every $\mathbf{Y}$. Taking $\mathbf{Y} = \mathbf{N}$, part (a) gives $\mathbf{X} = \begin{pmatrix} p & q \\ 0 & p \end{pmatrix}$. Taking $\mathbf{Y} = \mathbf{N}^\top = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, we compute $\mathbf{X}\mathbf{N}^\top = \begin{pmatrix} q & 0 \\ p & 0 \end{pmatrix}$ and $\mathbf{N}^\top\mathbf{X} = \begin{pmatrix} 0 & 0 \\ p & q \end{pmatrix}$. Equality forces $q = 0$. Hence $\mathbf{X} = p\mathbf{I}_2$, as claimed.