

Chapter 3

Matrices and linear transformations

Learning goals

After this chapter you should be able to:

- read $A\mathbf{x}$ as a combination of the columns of A and use it to restate $A\mathbf{x} = \mathbf{b}$;
- recognise linear transformations of $\mathbb{R}^n$, find their standard matrices, and describe rotations, shears, projections and reflections;
- multiply matrices, interpret the product as composition, and use the rules of matrix algebra;
- work with transposes, symmetric, triangular and block matrices and the trace;
- decide invertibility, compute inverses by Gauss–Jordan elimination, and state the invertible matrix theorem;
- factor a matrix as $A = LU$ and use the factorisation to solve systems.

3.1 The matrix-vector product

Lecture 12

In Section 1.1 we solved the system $2x - y = 1$, $x + y = 5$ in three ways: by rows, by columns, and in the compact form $A\mathbf{x} = \mathbf{b}$, where the matrix A holds the coefficients, $\mathbf{x}$ the unknowns and $\mathbf{b}$ the right-hand side. This section makes the compact form precise. We define the product of a matrix and a vector, give two ways to compute it (by columns and by rows), and prove that it is *linear*. With these tools we look at linear systems again: $A\mathbf{x} = \mathbf{b}$ becomes a statement we can reason with, not only shorthand. We find exactly when it can be solved for every right-hand side, and we see that all of its solutions are one solution plus the solutions of the homogeneous system $A\mathbf{x} = \mathbf{0}$.

Throughout, $\mathbb{F}$ is a field (Section 2.1) and $A = (a_{ij})$ is an $m \times n$ matrix over $\mathbb{F}$ with columns $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{F}^m$; the column $\mathbf{a}_j = (a_{1j}, \dots, a_{mj})$ holds the entries of column j .

3.1.1 A combination of the columns

In the column picture (Section 2.2) a system says that the unknowns are the weights of a linear combination of the columns that equals the right-hand side. We take this as the definition of the product.

Definition 3.1.1 (Matrix-vector product). Let A be an $m \times n$ matrix over $\mathbb{F}$ with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, and let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{F}^n$. The *product* of A and $\mathbf{x}$ is the vector

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n \in \mathbb{F}^m,$$

the linear combination of the columns of A with the entries of $\mathbf{x}$ as weights. The product is defined only when $\mathbf{x}$ has as many entries as A has columns.

The first entry of $\mathbf{x}$ says how much of the first column to take, the second how much of the second column, and so on; then the scaled columns are added.

Example 3.1.2. Let

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

The columns of A are $\mathbf{a}_1 = (2, 1)$ and $\mathbf{a}_2 = (-1, 1)$, so by the definition

$$A\mathbf{x} = 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 + (-3) \\ 2 + 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}.$$

This is the right-hand side $\mathbf{b} = (1, 5)$ of the system $2x - y = 1$, $x + y = 5$, so $\mathbf{x} = (2, 3)$ solves it, as found in Example 1.1.5. In the plane, two copies of $\mathbf{a}_1$ followed by three copies of $\mathbf{a}_2$ end at the point $(1, 5)$ (Figure 3.1).

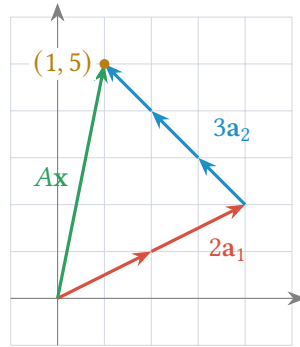


Figure 3.1: Example 3.1.2: $A\mathbf{x} = 2\mathbf{a}_1 + 3\mathbf{a}_2$ walks two copies of the first column and three of the second, and lands on $(1, 5)$.

Sizes. The definition needs one weight for each column, and the result is a combination of columns with m entries each. So an $m \times n$ matrix multiplies vectors in $\mathbb{F}^n$ and produces vectors in $\mathbb{F}^m$: one entry of $A\mathbf{x}$ for each row of A . In shorthand, $(m \times n)(n \times 1) = (m \times 1)$: the inner sizes must match, and the outer sizes give the size of the result.

Example 3.1.3. Let

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}.$$

A is 2×3 and $\mathbf{x} \in \mathbb{F}^3$, so $A\mathbf{x}$ is defined and lies in $\mathbb{F}^2$:

$$A\mathbf{x} = 3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 3 + 2 + (-2) \\ 0 + 1 + 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

On the other hand, $A(1, 4)$ is not defined: A has three columns and $(1, 4)$ supplies only two weights.

Taking $\mathbf{x} = (0, 1, 0)$, with a single 1 in position 2, gives $A\mathbf{x} = 0\mathbf{a}_1 + 1\mathbf{a}_2 + 0\mathbf{a}_3 = \mathbf{a}_2 = (2, 1)$, the second column of A .

The last observation holds in general: if $\mathbf{x} \in \mathbb{F}^n$ has a 1 in position j and zeros elsewhere, then $A\mathbf{x} = \mathbf{a}_j$, the j -th column of A . So the product of A with these n vectors recovers the whole matrix. This is the key to Section 3.2, where these vectors are called the standard basis vectors.

3.1.2 The row rule

The definition works with columns. To compute by hand, it is usually quicker to produce $A\mathbf{x}$ one entry at a time.

Proposition 3.1.4 (Row rule). For $i = 1, \dots, m$, the i -th entry of $A\mathbf{x}$ is

$$(A\mathbf{x})_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n.$$

Equivalently, $(A\mathbf{x})_i = \mathbf{r}_i \cdot \mathbf{x}$, where $\mathbf{r}_i = (a_{i1}, \dots, a_{in})$ is the i -th row of A and $\cdot$ is the dot product on $\mathbb{F}^n$ (Sections 1.4 and 2.2).

Proof. Vectors in $\mathbb{F}^m$ are added and scaled entry by entry (Definition 2.1.18), so the i -th entry of a linear combination is the same combination of the i -th entries:

$$(A\mathbf{x})_i = (x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n)_i = x_1(\mathbf{a}_1)_i + \cdots + x_n(\mathbf{a}_n)_i.$$

The i -th entry of the column $\mathbf{a}_j$ is $(\mathbf{a}_j)_i = a_{ij}$, so this is $x_1a_{i1} + \cdots + x_na_{in}$. Finally, multiplication in $\mathbb{F}$ is commutative, so $x_ja_{ij} = a_{ij}x_j$ for each j , and the sum is $a_{i1}x_1 + \cdots + a_{in}x_n = \mathbf{r}_i \cdot \mathbf{x}$. $\square$

For Example 3.1.3 the row rule gives, row by row,

$$(A\mathbf{x})_1 = 1 \cdot 3 + 2 \cdot 1 + (-1) \cdot 2 = 3 + 2 - 2 = 3, \quad (A\mathbf{x})_2 = 0 \cdot 3 + 1 \cdot 1 + 3 \cdot 2 = 0 + 1 + 6 = 7,$$

the same vector $(3, 7)$. The same six products appear in both computations, grouped by columns in the first and by rows in the second.

The two ways of computing $A\mathbf{x}$ match the two pictures of a system. The *column rule* (Definition 3.1.1) is the column picture: a combination of the columns. The *row rule* is the row picture: one equation $\mathbf{r}_i \cdot \mathbf{x} = b_i$ for each row. Use the row rule to compute, and the column rule to understand what $A\mathbf{x}$ means.

3.1.3 The identity matrix

Definition 3.1.5 (Identity matrix). The *identity matrix* I_n is the $n \times n$ matrix with 1 in every diagonal position (i, i) and 0 in every other position. For example,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We write I when the size is clear.

The j -th column of I_n has a 1 in position j and zeros elsewhere. By the column rule,

$$I_3 \mathbf{x} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{x},$$

and in general the row rule gives $(I_n \mathbf{x})_i = 0 \cdot x_1 + \cdots + 1 \cdot x_i + \cdots + 0 \cdot x_n = x_i$, so

$$I_n \mathbf{x} = \mathbf{x} \quad \text{for every } \mathbf{x} \in \mathbb{F}^n.$$

The identity matrix does to vectors what the number 1 does to numbers.

3.1.4 Linearity

Theorem 3.1.6 (Linearity of the product). Let A be an $m \times n$ matrix over $\mathbb{F}$. For all $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$ and all $c, d \in \mathbb{F}$:

- (a) $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$;
- (b) $A(c\mathbf{u}) = c(A\mathbf{u})$;
- (c) $A(c\mathbf{u} + d\mathbf{v}) = cA\mathbf{u} + dA\mathbf{v}$.

Part (a) says that A respects addition and part (b) that it respects scaling; together they are called *linearity*. We use the rules of vector algebra in $\mathbb{F}^m$ (Theorem 1.2.7, which holds over every field, Section 2.1), numbered as there.

Proof. (a) The entries of $\mathbf{u} + \mathbf{v}$ are $u_j + v_j$. By Definition 3.1.1,

$$\begin{aligned} A(\mathbf{u} + \mathbf{v}) &= (u_1 + v_1)\mathbf{a}_1 + \cdots + (u_n + v_n)\mathbf{a}_n \\ &= (u_1\mathbf{a}_1 + v_1\mathbf{a}_1) + \cdots + (u_n\mathbf{a}_n + v_n\mathbf{a}_n) && \text{rule (8), distributive law over scalar addition} \\ &= (u_1\mathbf{a}_1 + \cdots + u_n\mathbf{a}_n) + (v_1\mathbf{a}_1 + \cdots + v_n\mathbf{a}_n) && \text{rules (1) and (2), commutative and associative} \\ &= A\mathbf{u} + A\mathbf{v} && \text{Definition 3.1.1.} \end{aligned}$$

In the third line, rules (1) and (2) allow the $2n$ vectors of the sum to be added in any order and grouped in any way; we put the u -terms first and the v -terms after them.

(b) The entries of $c\mathbf{u}$ are cu_j . By Definition 3.1.1,

$$\begin{aligned} A(c\mathbf{u}) &= (cu_1)\mathbf{a}_1 + \cdots + (cu_n)\mathbf{a}_n \\ &= c(u_1\mathbf{a}_1) + \cdots + c(u_n\mathbf{a}_n) && \text{rule (6), scaling twice: } (cu_j)\mathbf{a}_j = c(u_j\mathbf{a}_j) \\ &= c(u_1\mathbf{a}_1 + \cdots + u_n\mathbf{a}_n) && \text{rule (7), distributive law over vector addition} \\ &= c(A\mathbf{u}) && \text{Definition 3.1.1.} \end{aligned}$$

Rule (7) is stated for two vectors; for n vectors it follows by applying it $n - 1$ times.

(c) By (a), then (b) twice: $A(c\mathbf{u} + d\mathbf{v}) = A(c\mathbf{u}) + A(d\mathbf{v}) = cA\mathbf{u} + dA\mathbf{v}$. □

For example, with A as in Example 3.1.2, $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (1, 3)$: $\mathbf{u} + \mathbf{v} = (2, 3)$ and $A(\mathbf{u} + \mathbf{v}) = (1, 5)$ by that example, while $A\mathbf{u} = 1\mathbf{a}_1 = (2, 1)$ and $A\mathbf{v} = (2, 1) + 3(-1, 1) = (-1, 4)$, so $A\mathbf{u} + A\mathbf{v} = (2 + (-1), 1 + 4) = (1, 5)$ as well.

By induction, (c) extends to any number of vectors: $A(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = c_1A\mathbf{v}_1 + \cdots + c_kA\mathbf{v}_k$. A function from $\mathbb{F}^n$ to $\mathbb{F}^m$ with properties (a) and (b) is called a *linear transformation*; Section 3.2 shows that every matrix defines one and that every linear transformation of $\mathbb{F}^n$ into $\mathbb{F}^m$ comes from a matrix.

3.1.5 Linear systems revisited

Consider a system of m linear equations in n unknowns,

$$a_{i1}x_1 + \cdots + a_{in}x_n = b_i \quad (i = 1, \dots, m),$$

with coefficient matrix $A = (a_{ij})$ and right-hand side $\mathbf{b} = (b_1, \dots, b_m) \in \mathbb{F}^m$ (Section 2.2).

Theorem 3.1.7 (Three forms of a linear system). For $\mathbf{s} \in \mathbb{F}^n$ the following are equivalent:

- (1) $\mathbf{s}$ is a solution of the system above;
- (2) $s_1\mathbf{a}_1 + \cdots + s_n\mathbf{a}_n = \mathbf{b}$;
- (3) $A\mathbf{s} = \mathbf{b}$.

In particular, the system, the vector equation $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ and the matrix equation $A\mathbf{x} = \mathbf{b}$ have the same solution set.

Proof. (2) and (3) say the same thing, since $A\mathbf{s} = s_1\mathbf{a}_1 + \cdots + s_n\mathbf{a}_n$ by Definition 3.1.1. For (1) and (3): two vectors of $\mathbb{F}^m$ are equal exactly when their entries are equal, and by the row rule the i -th entry of $A\mathbf{s}$ is $a_{i1}s_1 + \cdots + a_{in}s_n$. So $A\mathbf{s} = \mathbf{b}$ holds exactly when $a_{i1}s_1 + \cdots + a_{in}s_n = b_i$ for every i , which is (1). $\square$

From now on we write a linear system as $A\mathbf{x} = \mathbf{b}$ and solve it, as before, by row reducing the augmented matrix $[A \mid \mathbf{b}]$.

Example 3.1.8. In Example 2.3.8 elimination solved the system with

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 3 \\ -1 & 0 & 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 8 \\ 4 \end{bmatrix},$$

and found $\mathbf{x} = (2, -1, 3)$. By the row rule,

$$A\mathbf{x} = \begin{bmatrix} 1 \cdot 2 + 2 \cdot (-1) + 1 \cdot 3 \\ 2 \cdot 2 + 5 \cdot (-1) + 3 \cdot 3 \\ (-1) \cdot 2 + 0 \cdot (-1) + 2 \cdot 3 \end{bmatrix} = \begin{bmatrix} 2 - 2 + 3 \\ 4 - 5 + 9 \\ -2 + 0 + 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \\ 4 \end{bmatrix} = \mathbf{b},$$

and by the column rule this says $2\mathbf{a}_1 - \mathbf{a}_2 + 3\mathbf{a}_3 = \mathbf{b}$.

By Theorem 3.1.7 and Proposition 2.2.9, $A\mathbf{x} = \mathbf{b}$ has a solution exactly when $\mathbf{b}$ is a linear combination of the columns of A , that is, when $\mathbf{b} \in \text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.

3.1.6 Solvable for every right-hand side

Often we need more than one right-hand side: when does $A\mathbf{x} = \mathbf{b}$ have a solution for *every* $\mathbf{b} \in \mathbb{F}^m$? The answer can be read off from pivots (Sections 2.3 and 2.4).

Theorem 3.1.9. For an $m \times n$ matrix A over $\mathbb{F}$ the following are equivalent:

- (a) $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{F}^m$;
- (b) every $\mathbf{b} \in \mathbb{F}^m$ is a linear combination of the columns of A ;

- (c) the columns of A span $\mathbb{F}^m$, that is, $\text{Span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} = \mathbb{F}^m$;
- (d) an echelon form of A has a pivot in every row.

Proof. (a) $\iff$ (b) by Theorem 3.1.7: $A\mathbf{x} = \mathbf{b}$ has a solution exactly when $\mathbf{b}$ is a combination of the columns. (b) $\iff$ (c) by the definition of span: the span is the set of all combinations of the columns, and it always lies in $\mathbb{F}^m$, so it equals $\mathbb{F}^m$ exactly when every $\mathbf{b} \in \mathbb{F}^m$ is such a combination.

It remains to link (a) with (d). Let U be an echelon form of A , reached from A by a sequence of elementary row operations (Theorem 2.3.7). For any $\mathbf{b}$, the same operations, applied to $[A \mid \mathbf{b}]$, produce $[U \mid \mathbf{d}]$ for some $\mathbf{d} \in \mathbb{F}^m$, because a row operation acts on the coefficient part and on the right-hand side independently. By Theorem 2.3.4 the systems $A\mathbf{x} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{d}$ have the same solutions.

(d) $\implies$ (a). Suppose every row of U has a pivot, and let $\mathbf{b} \in \mathbb{F}^m$. Then $[U \mid \mathbf{d}]$ is in row echelon form, and none of its rows has the form $[0 \cdots 0 \mid c]$ with $c \neq 0$, since every row has a nonzero entry to the left of the bar; so the last column is not a pivot column. By Theorem 2.5.3(a) the system $U\mathbf{x} = \mathbf{d}$ is consistent, and therefore so is $A\mathbf{x} = \mathbf{b}$.

(a) $\implies$ (d). We prove the contrapositive. Suppose some row of U has no pivot. That row has no leading entry, so it is a zero row, and zero rows are at the bottom of an echelon form; hence the last row of U is zero. Let $\mathbf{d} = (0, \dots, 0, 1)$. The last row of $[U \mid \mathbf{d}]$ reads $0 = 1$, so $U\mathbf{x} = \mathbf{d}$ has no solution. Every elementary row operation can be undone by an elementary row operation (proof of Theorem 2.3.4), so applying the inverses of the operations that took A to U , in reverse order, to $[U \mid \mathbf{d}]$ gives a matrix $[A \mid \mathbf{b}]$ for some $\mathbf{b} \in \mathbb{F}^m$; its coefficient part is A because these inverse operations take U back to A . The two matrices are row equivalent, so $A\mathbf{x} = \mathbf{b}$ has no solution either, and (a) fails. $\square$

Example 3.1.10. For $A = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$,

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \xrightarrow{R_2 \rightarrow R_2 - 2R_1} \begin{bmatrix} 1 & 1 \\ 0 & -3 \end{bmatrix},$$

since $2 - 2 \cdot 1 = 0$ and $-1 - 2 \cdot 1 = -3$. There is a pivot in both rows, so $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^2$: the columns $(2, 1)$ and $(-1, 1)$ span $\mathbb{R}^2$ (as Theorem 1.3.8 also shows).

For $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, the replacement $R_2 \rightarrow R_2 - 2R_1$ gives $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$, since $2 - 2 \cdot 1 = 0$ and $4 - 2 \cdot 2 = 0$. The second row has no pivot, so some $\mathbf{b}$ is out of reach. In the column picture: $(2, 4) = 2(1, 2)$, so every combination $x(1, 2) + y(2, 4) = (x + 2y)(1, 2)$ lies on the line through the origin spanned by $(1, 2)$, and $\mathbf{b} = (1, 3)$, which is not a multiple of $(1, 2)$, is never reached (Figure 3.2).

More rows than columns. An echelon form has at most one pivot in each column, so an $m \times n$ matrix has at most n pivots. If $m > n$, some row has no pivot, and by Theorem 3.1.9 the columns do not span $\mathbb{F}^m$. For instance, two vectors never span $\mathbb{R}^3$, and a system of three equations in two unknowns cannot be solvable for every right-hand side.

3.1.7 The shape of the solution set

Recall that a system $A\mathbf{x} = \mathbf{0}$ is *homogeneous* (Definition 2.5.7). In the new notation the trivial solution is the statement $A\mathbf{0} = 0\mathbf{a}_1 + \cdots + 0\mathbf{a}_n = \mathbf{0}$: each term is $\mathbf{0}$ by Proposition 1.2.8(d), which holds over every field with the same proof, and a sum of zero vectors is $\mathbf{0}$ by rule (3) of Theorem 1.2.7. Section 2.5

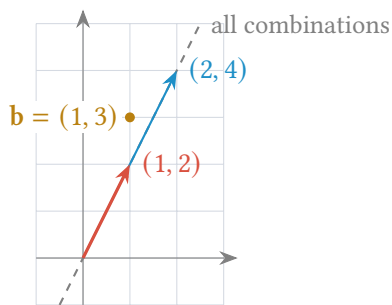


Figure 3.2: The columns $(1, 2)$ and $(2, 4)$ of the second matrix in Example 3.1.10 lie on one line through the origin; every combination of them stays on that line, and $\mathbf{b} = (1, 3)$ is never reached.

showed that the solutions of a consistent system are one particular solution plus the solutions of the homogeneous system (Theorem 2.5.11). In matrix language the statement is short, and linearity gives a two-line proof.

Theorem 3.1.11 (Structure of the solution set, matrix form). Let $\mathbf{p} \in \mathbb{F}^n$ satisfy $A\mathbf{p} = \mathbf{b}$. Then the solutions of $A\mathbf{x} = \mathbf{b}$ are exactly the vectors

$$\mathbf{x} = \mathbf{p} + \mathbf{h}, \quad \text{where } A\mathbf{h} = \mathbf{0}.$$

Proof. Every such vector is a solution. If $A\mathbf{h} = \mathbf{0}$, then by Theorem 3.1.6(a) $A(\mathbf{p} + \mathbf{h}) = A\mathbf{p} + A\mathbf{h} = \mathbf{b} + \mathbf{0} = \mathbf{b}$.

Every solution has this form. Let $A\mathbf{x} = \mathbf{b}$ and put $\mathbf{h} = \mathbf{x} - \mathbf{p}$, which is $\mathbf{x} + (-1)\mathbf{p}$ by Proposition 1.2.8(f). By Theorem 3.1.6(a) and (b) with $c = -1$, $A\mathbf{h} = A\mathbf{x} + (-1)A\mathbf{p} = \mathbf{b} + (-1)\mathbf{b} = \mathbf{b} + (-\mathbf{b}) = \mathbf{0}$, using Proposition 1.2.8(f) again and rule (4) of Theorem 1.2.7. So $\mathbf{h}$ solves the homogeneous system, and $\mathbf{x} = \mathbf{p} + \mathbf{h}$ (rules (2), (4) and (3) of Theorem 1.2.7: $\mathbf{p} + (\mathbf{x} - \mathbf{p}) = \mathbf{x} + (\mathbf{p} + (-\mathbf{p})) = \mathbf{x}$, after reordering by rule (1)). $\square$

Example 3.1.12. Take the second matrix of Example 3.1.10, $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, whose columns reach only the line through $(1, 2)$, and $\mathbf{b} = (3, 6) = 3(1, 2)$, which lies on that line. The vector $\mathbf{p} = (3, 0)$ is a solution: by the row rule $A\mathbf{p} = (3 + 0, 6 + 0) = (3, 6)$. The homogeneous system is $x + 2y = 0$, $2x + 4y = 0$, the same equation twice; its solutions are $y = t$, $x = -2t$, that is $\mathbf{h} = t(-2, 1)$ for $t \in \mathbb{R}$, a line through the origin. By Theorem 3.1.11 the solutions of $A\mathbf{x} = \mathbf{b}$ are

$$\mathbf{x} = \begin{bmatrix} 3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \quad t \in \mathbb{R},$$

the line of homogeneous solutions shifted to pass through $\mathbf{p}$ (Figure 3.3). For $t = 1$ we get $(1, 1)$, and indeed $1 + 2 = 3$ and $2 + 4 = 6$.

Proposition 3.1.13 (Uniqueness). For an $m \times n$ matrix A with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$, the following are equivalent:

- (a) whenever $A\mathbf{x} = \mathbf{b}$ has a solution, it has only one;
- (b) the homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution;
- (c) the columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ are linearly independent;
- (d) an echelon form of A has a pivot in every column.

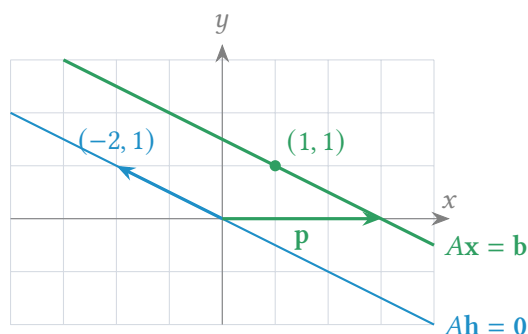


Figure 3.3: Example 3.1.12: the solutions of $Ax = 0$ form a line through the origin (blue); the solutions of $Ax = b$ form the parallel line through $p = (3, 0)$ (green).

Proof. (a) $\Rightarrow$ (b): $Ax = 0$ has the solution 0 , so by (a) it has no other. (b) $\Rightarrow$ (a): if p solves $Ax = b$, then by Theorem 3.1.11 every solution is $p + h$ with $Ah = 0$, so $h = 0$ and the solution is p . (b) $\iff$ (c) $\iff$ (d) is the elimination test, Theorem 2.6.8: since $Ax = x_1a_1 + \cdots + x_na_n$, statement (b) says that the only combination of the columns equal to 0 is the one with all weights 0 . $\square$

Theorem 3.1.9 and Proposition 3.1.13 form a pair: a pivot in every row of an echelon form means that $Ax = b$ has *at least* one solution for every b , and a pivot in every column means that it has *at most* one.

Check yourself

Let $A = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{bmatrix}$ and $x = (2, 1, -1)$. (a) Compute Ax by the column rule and again by the row rule. (b) Do the columns of A span $\mathbb{R}^2$? Use Theorem 3.1.9. (c) Can $Ax = 0$ have only the trivial solution? Explain with Proposition 3.1.13.

Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

A matrix is determined by what it does. If A and B are $m \times n$ matrices with $Ax = Bx$ for every $x \in \mathbb{F}^n$, then $A = B$: taking x with a single 1 in position j gives $a_j = b_j$ for every j , so the two matrices have the same columns.

Over a finite field. Everything in this section uses only the field operations, so it holds over $\mathbb{Z}_p$. Over $\mathbb{Z}_2$, for instance,

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1+1+0 \\ 0+1+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

because $1 + 1 = 0$ in $\mathbb{Z}_2$. So this homogeneous system has the nontrivial solution $(1, 1, 1)$, and by Proposition 3.1.13 the three columns are linearly dependent over $\mathbb{Z}_2$. Products of this kind are the basis of error-correcting codes, where $Ax = 0$ is the test that a received word of bits is a valid codeword.

How much work is Ax ? By the row rule each of the m entries needs n multiplications and $n - 1$ additions, so Ax costs mn multiplications and $m(n - 1)$ additions. When many entries of A are zero, as

in the large matrices of applications, only the nonzero entries need to be multiplied, and computers store such *sparse* matrices by listing their nonzero entries only.

3.2 Linear transformations of $\mathbb{R}^n$

Lecture 13

In Section 3.1 we multiplied a matrix by a vector and proved that the product is linear: $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u}+A\mathbf{v}$ and $A(c\mathbf{u}) = c(A\mathbf{u})$. In this section we change the point of view. Instead of a table of numbers, we think of a matrix as a machine that takes a vector and returns a new one, and we ask which machines of this kind exist. The answer is the main theorem of the section: the functions that respect addition and scaling, the *linear transformations*, are exactly the functions $\mathbf{x} \mapsto A\mathbf{x}$, and the matrix can be read off from what the function does to the standard basis vectors. We then find the matrices of the basic moves of the plane: rotations, scalings, shears, projections and reflections.

Throughout, $\mathbb{F}$ is a field (Section 2.1); the geometric examples are in $\mathbb{R}^2$.

3.2.1 Transformations and linearity

A *function* T from $\mathbb{F}^n$ to $\mathbb{F}^m$, written $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$, is a rule that assigns to each vector $\mathbf{x} \in \mathbb{F}^n$ exactly one vector $T(\mathbf{x}) \in \mathbb{F}^m$. The set $\mathbb{F}^n$ is the *domain* of T and $\mathbb{F}^m$ is its *codomain*; the vector $T(\mathbf{x})$ is the *image* of $\mathbf{x}$ under T . In linear algebra such a function is also called a *transformation* or a *map*. Two functions $T, S: \mathbb{F}^n \rightarrow \mathbb{F}^m$ are equal when $T(\mathbf{x}) = S(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{F}^n$.

Every $m \times n$ matrix A over $\mathbb{F}$ gives a transformation

$$T_A: \mathbb{F}^n \rightarrow \mathbb{F}^m, \quad T_A(\mathbf{x}) = A\mathbf{x},$$

called a *matrix transformation*. The sizes are forced by Section 3.1: $A\mathbf{x}$ is defined for $\mathbf{x} \in \mathbb{F}^n$, one entry for each column, and lies in $\mathbb{F}^m$, one entry for each row.

Definition 3.2.1 (Linear transformation). A transformation $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$ is *linear* if

- (i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$, and
- (ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{F}^n$ and all $c \in \mathbb{F}$.

Rule (i) says that adding first and then transforming gives the same as transforming first and then adding; rule (ii) says the same for scaling. Geometrically, in $\mathbb{R}^2$: the vectors $\mathbf{u}, \mathbf{v}$ and $\mathbf{u} + \mathbf{v}$ are three corners of a parallelogram with its fourth corner at the origin, and a linear transformation carries them to $T(\mathbf{u}), T(\mathbf{v})$ and $T(\mathbf{u}) + T(\mathbf{v})$, again three corners of a parallelogram (possibly flattened into a segment).

Proposition 3.2.2 (Matrix transformations are linear). For every $m \times n$ matrix A over $\mathbb{F}$, the matrix transformation T_A is linear.

Proof. Rules (i) and (ii) for T_A are parts (a) and (b) of Theorem 3.1.6. □

Proposition 3.2.3 (First consequences). Let $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$ be linear. Then

(a) $T(\mathbf{0}) = \mathbf{0}$;

(b) $T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + \cdots + c_kT(\mathbf{v}_k)$ for all $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{F}^n$ and $c_1, \dots, c_k \in \mathbb{F}$.

Conversely, a transformation T with $T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$ and $c, d \in \mathbb{F}$ is linear.

We use Proposition 1.2.8, which holds in $\mathbb{F}^n$ over every field with the same proof (Section 2.1).

Proof. (a) By Proposition 1.2.8(d), $\mathbf{0} = 0\mathbf{0}$ in $\mathbb{F}^n$. By rule (ii) with $c = 0$,

$$T(\mathbf{0}) = T(0\mathbf{0}) = 0T(\mathbf{0}) = \mathbf{0},$$

the last step by Proposition 1.2.8(d) again, now in $\mathbb{F}^m$.

(b) By induction on k . For $k = 1$ this is rule (ii). If it holds for $k - 1$ vectors, then by rule (i), the case $k - 1$, and rule (ii),

$$\begin{aligned} T((c_1\mathbf{v}_1 + \cdots + c_{k-1}\mathbf{v}_{k-1}) + c_k\mathbf{v}_k) \\ &= T(c_1\mathbf{v}_1 + \cdots + c_{k-1}\mathbf{v}_{k-1}) + T(c_k\mathbf{v}_k) \\ &= c_1T(\mathbf{v}_1) + \cdots + c_{k-1}T(\mathbf{v}_{k-1}) + c_kT(\mathbf{v}_k). \end{aligned}$$

For the converse, $c = d = 1$ gives rule (i), and $d = 0$ gives $T(c\mathbf{u}) = T(c\mathbf{u} + 0\mathbf{u}) = cT(\mathbf{u}) + 0T(\mathbf{u}) = cT(\mathbf{u})$, which is rule (ii) (Proposition 1.2.8(d) and rule (3) of Theorem 1.2.7). $\square$

Part (b) is what makes linear transformations manageable: once we know T on a few vectors, we know it on every linear combination of them. Part (a) gives a quick test for non-linearity: a map that moves the origin is not linear.

Example 3.2.4 (Checking linearity from the definition). Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $T(x_1, x_2) = (x_1 + 2x_2, 3x_1)$. For $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$, the sum $\mathbf{u} + \mathbf{v}$ has entries $u_1 + v_1$ and $u_2 + v_2$, so

$$T(\mathbf{u} + \mathbf{v}) = \begin{bmatrix} (u_1 + v_1) + 2(u_2 + v_2) \\ 3(u_1 + v_1) \end{bmatrix} = \begin{bmatrix} (u_1 + 2u_2) + (v_1 + 2v_2) \\ 3u_1 + 3v_1 \end{bmatrix} = T(\mathbf{u}) + T(\mathbf{v}),$$

and for $c \in \mathbb{R}$,

$$T(c\mathbf{u}) = \begin{bmatrix} cu_1 + 2cu_2 \\ 3cu_1 \end{bmatrix} = c \begin{bmatrix} u_1 + 2u_2 \\ 3u_1 \end{bmatrix} = cT(\mathbf{u}).$$

Both rules hold, so T is linear. (Each step uses only the field axioms, so the same formula defines a linear map of $\mathbb{F}^2$ over every field.)

Example 3.2.5 (Two maps that are not linear). (a) The *translation* $T(\mathbf{x}) = \mathbf{x} + (1, 0)$ of $\mathbb{R}^2$ moves every point one unit to the right. It sends $\mathbf{0}$ to $(1, 0) \neq \mathbf{0}$, so by Proposition 3.2.3(a) it is not linear. More generally, a translation $\mathbf{x} \mapsto \mathbf{x} + \mathbf{b}$ is linear only when $\mathbf{b} = \mathbf{0}$.

(b) Let $S: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $S(x_1, x_2) = (x_1^2, x_2)$. Here $S(\mathbf{0}) = \mathbf{0}$, so test (a) does not decide. Take $\mathbf{u} = (1, 0)$ and $c = 2$:

$$S(2\mathbf{u}) = S(2, 0) = (4, 0), \quad 2S(\mathbf{u}) = 2(1, 0) = (2, 0).$$

These differ, so rule (ii) fails and S is not linear.

One counterexample shows that a map is not linear; to show that a map is linear, the rules must be checked for *all* vectors and scalars, as in Example 3.2.4.

3.2.2 The standard matrix

Every matrix gives a linear transformation (Proposition 3.2.2). The converse needs the vectors already met at the end of Example 3.1.3.

Definition 3.2.6 (Standard basis vectors). For $j = 1, \dots, n$, the *standard basis vector* $\mathbf{e}_j \in \mathbb{F}^n$ has 1 in position j and 0 in every other position:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

They are the columns of the identity matrix I_n . In $\mathbb{R}^2$, $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ are one step along the first and the second axis. Every $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{F}^n$ is a combination of them with its own entries as weights:

$$\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n, \quad (3.1)$$

since the right-hand side has x_i in position i (it is $I_n \mathbf{x}$, Section 3.1). For instance $(3, 2) = 3\mathbf{e}_1 + 2\mathbf{e}_2$.

Theorem 3.2.7 (The standard matrix). Let $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$ be linear. There is exactly one $m \times n$ matrix A over $\mathbb{F}$ with

$$T(\mathbf{x}) = A\mathbf{x} \quad \text{for every } \mathbf{x} \in \mathbb{F}^n,$$

namely the matrix whose columns are the images of the standard basis vectors:

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \end{bmatrix}.$$

It is called the *standard matrix* of T . In particular, the linear transformations $\mathbb{F}^n \rightarrow \mathbb{F}^m$ are exactly the matrix transformations T_A .

Proof. Existence. Let A be the $m \times n$ matrix with columns $\mathbf{a}_j = T(\mathbf{e}_j)$; each lies in $\mathbb{F}^m$. For $\mathbf{x} \in \mathbb{F}^n$, by (3.1), Proposition 3.2.3(b) and Definition 3.1.1,

$$T(\mathbf{x}) = T(x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_n) = x_1 T(\mathbf{e}_1) + \dots + x_n T(\mathbf{e}_n) = x_1 \mathbf{a}_1 + \dots + x_n \mathbf{a}_n = A\mathbf{x}.$$

Uniqueness. Suppose B is an $m \times n$ matrix with $B\mathbf{x} = T(\mathbf{x})$ for every $\mathbf{x}$. Taking $\mathbf{x} = \mathbf{e}_j$, and recalling that $B\mathbf{e}_j$ is the j -th column $\mathbf{b}_j$ of B (Section 3.1), we get $\mathbf{b}_j = T(\mathbf{e}_j) = \mathbf{a}_j$ for every j . So B and A have the same columns, and $B = A$.

The last sentence combines existence with Proposition 3.2.2. □

Geometrically, in $\mathbb{R}^2$: a linear transformation is decided by the two vectors $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$. The grid of the plane, made of the points $x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2$ with x_1 or x_2 an integer, is carried to the lines through the points $x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2)$: grid lines go to parallel, evenly spaced lines (or, when $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ lie on one line, all collapse onto it). Figure 3.4 shows this for the map of Example 3.2.4.

Example 3.2.8 (Finding a standard matrix). (a) For $T(x_1, x_2) = (x_1 + 2x_2, 3x_1)$ of Example 3.2.4,

$$T(\mathbf{e}_1) = T(1, 0) = (1 + 0, 3 \cdot 1) = (1, 3), \quad T(\mathbf{e}_2) = T(0, 1) = (0 + 2, 3 \cdot 0) = (2, 0),$$

so the standard matrix is

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}.$$

Check by the row rule: $A\mathbf{x} = (1 \cdot x_1 + 2x_2, 3x_1 + 0 \cdot x_2) = (x_1 + 2x_2, 3x_1) = T(\mathbf{x})$.

(b) The map $T: \mathbb{R}^3 \rightarrow \mathbb{R}^2$, $T(x_1, x_2, x_3) = (x_1 - x_3, 2x_2 + x_3)$ is linear (check the rules as in Example 3.2.4). Its standard matrix is 2×3 , with columns $T(\mathbf{e}_1) = (1, 0)$, $T(\mathbf{e}_2) = (0, 2)$ and $T(\mathbf{e}_3) = (-1, 1)$:

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & 1 \end{bmatrix}.$$

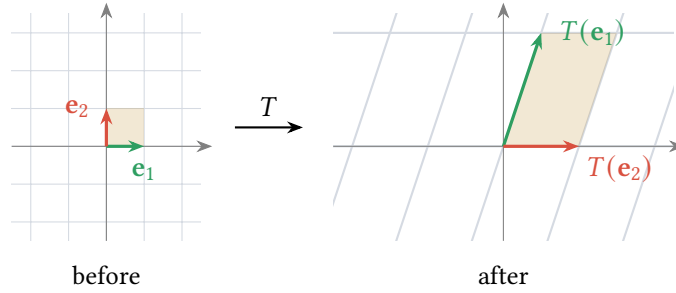


Figure 3.4: The map of Examples 3.2.4 and 3.2.8(a). Left: the grid with $\mathbf{e}_1$ (green) and $\mathbf{e}_2$ (coral). Right: its image, the lines through the points $x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2)$, with $T(\mathbf{e}_1) = (1, 0)$ and $T(\mathbf{e}_2) = (0, 2)$ the columns of A ; the shaded unit square goes to the shaded parallelogram.

3.2.3 Rotations

Every point $\mathbf{x} = (x, y) \neq \mathbf{0}$ of the plane can be described by its distance $r = \|\mathbf{x}\| > 0$ from the origin and an angle α , measured counterclockwise from the positive first axis, with

$$x = r \cos \alpha, \quad y = r \sin \alpha$$

(polar coordinates). The *rotation through the angle θ* (counterclockwise, about the origin) is the map $R_\theta: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ that keeps the distance and adds θ to the angle:

$$R_\theta(r \cos \alpha, r \sin \alpha) = (r \cos(\alpha + \theta), r \sin(\alpha + \theta)), \quad R_\theta(\mathbf{0}) = \mathbf{0}.$$

This does not depend on which angle α is chosen for the point, since angles differing by a multiple of 2π give the same cosine and sine. A rotation clockwise through θ is $R_{-\theta}$.

Theorem 3.2.9 (Rotations). The rotation R_θ is linear, with standard matrix

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$

Proof. Let $\mathbf{x} = (r \cos \alpha, r \sin \alpha)$. The angle addition formulas of trigonometry,

$$\cos(\alpha + \theta) = \cos \alpha \cos \theta - \sin \alpha \sin \theta, \quad \sin(\alpha + \theta) = \sin \alpha \cos \theta + \cos \alpha \sin \theta,$$

multiplied by r , give, with $x = r \cos \alpha$ and $y = r \sin \alpha$,

$$R_\theta(\mathbf{x}) = \begin{bmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

by the row rule. The same formula holds for $\mathbf{x} = \mathbf{0}$. So $R_\theta = T_A$ for the matrix A above, which is linear by Proposition 3.2.2, and A is its standard matrix by Theorem 3.2.7. $\square$

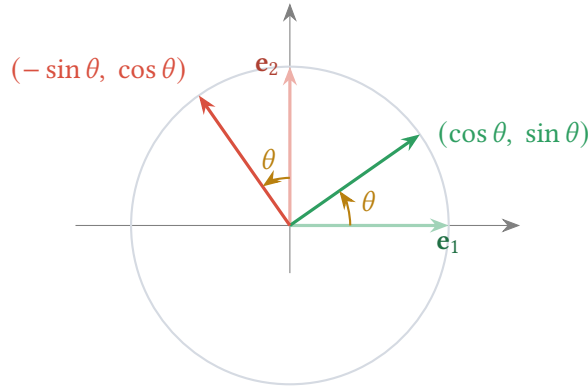


Figure 3.5: Theorem 3.2.9: the rotation through θ sends $\mathbf{e}_1$ to $(\cos \theta, \sin \theta)$ and $\mathbf{e}_2$ to $(-\sin \theta, \cos \theta)$, the columns of its standard matrix.

The columns can also be seen directly (Figure 3.5): $\mathbf{e}_1$ has $r = 1$, $\alpha = 0$, so $R_\theta(\mathbf{e}_1) = (\cos \theta, \sin \theta)$, a point of the unit circle; $\mathbf{e}_2$ has $r = 1$, $\alpha = \pi/2$, so $R_\theta(\mathbf{e}_2) = (\cos(\theta + \pi/2), \sin(\theta + \pi/2)) = (-\sin \theta, \cos \theta)$, a quarter turn further on.

Example 3.2.10 (A quarter turn and a sixth of a turn). (a) For $\theta = 90^\circ$ ($\pi/2$), $\cos \theta = 0$ and $\sin \theta = 1$, so

$$R_{90^\circ} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad R_{90^\circ} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \cdot 2 + (-1) \cdot 1 \\ 1 \cdot 2 + 0 \cdot 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

In general $R_{90^\circ}(x, y) = (-y, x)$; note that $(x, y) \cdot (-y, x) = 0$, so the image is perpendicular to the original vector, as a quarter turn should be.

(b) For $\theta = 60^\circ$, $\cos \theta = \frac{1}{2}$ and $\sin \theta = \frac{\sqrt{3}}{2}$, so $R_{60^\circ} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}$ and $R_{60^\circ}(2, 0) = (1, \sqrt{3})$, a vector of length $\sqrt{1+3} = 2$ at 60° to the first axis.

3.2.4 Scalings and shears

Example 3.2.11 (Scalings and shears). (a) For $k \in \mathbb{R}$, the matrix $kI_2 = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$ sends $\mathbf{x}$ to $k\mathbf{x}$. For $k > 1$ it stretches the plane away from the origin (a *dilation*); for $0 < k < 1$ it shrinks it (a *contraction*).

(b) For $a, d \in \mathbb{R}$, the matrix $\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$ sends (x, y) to (ax, dy) : it stretches by the factor a along the first axis and by d along the second. With $a = 2$ and $d = \frac{1}{2}$ the plane becomes twice as wide and half as tall.

(c) For $k \in \mathbb{R}$, the *shear* $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$ sends (x, y) to $(x + ky, y)$. It fixes $\mathbf{e}_1$ and sends $\mathbf{e}_2$ to $(k, 1)$. Every point slides parallel to the first axis by k times its height y : points of the first axis do not move, and points higher up move further. The unit square, with corners $\mathbf{0}$, $\mathbf{e}_1$, $\mathbf{e}_1 + \mathbf{e}_2$, $\mathbf{e}_2$, goes to the parallelogram with corners $\mathbf{0}$, $(1, 0)$, $(1 + k, 1)$, $(k, 1)$, which has the same base 1 and the same height 1, hence the same area. How a linear transformation changes area is measured by the determinant (Chapter 7).

3.2.5 Projections

Let $\mathbf{u} \in \mathbb{R}^n$ be a unit vector (Definition 1.4.8) and let $L = \text{Span}\{\mathbf{u}\} = \{t\mathbf{u} : t \in \mathbb{R}\}$, the line through the origin in the direction of $\mathbf{u}$.

Proposition 3.2.12 (Projection onto a line). For every $\mathbf{x} \in \mathbb{R}^n$ there is exactly one point $\mathbf{p} \in L$ with $\mathbf{x} - \mathbf{p}$ orthogonal to $\mathbf{u}$, namely

$$\mathbf{p} = P(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{u}) \mathbf{u},$$

and it is the point of L closest to $\mathbf{x}$. The map $P: \mathbb{R}^n \rightarrow \mathbb{R}^n$, the *orthogonal projection onto L* , is linear, and its standard matrix has entry $u_i u_j$ in row i , column j . In $\mathbb{R}^2$, with $\mathbf{u} = (u_1, u_2)$,

$$P = \begin{bmatrix} u_1^2 & u_1 u_2 \\ u_1 u_2 & u_2^2 \end{bmatrix}.$$

Proof. A point of L is $t\mathbf{u}$ for some $t \in \mathbb{R}$. By Theorem 1.4.6,

$$(\mathbf{x} - t\mathbf{u}) \cdot \mathbf{u} = \mathbf{x} \cdot \mathbf{u} - t(\mathbf{u} \cdot \mathbf{u}) = \mathbf{x} \cdot \mathbf{u} - t,$$

since $\mathbf{u} \cdot \mathbf{u} = \|\mathbf{u}\|^2 = 1$. This is 0 exactly when $t = \mathbf{x} \cdot \mathbf{u}$, so $\mathbf{p}$ exists and is unique.

Closest point. For any $s \in \mathbb{R}$, the vectors $\mathbf{x} - \mathbf{p}$ and $\mathbf{p} - s\mathbf{u} = (t - s)\mathbf{u}$ are orthogonal, so by Pythagoras (Theorem 1.4.14)

$$\|\mathbf{x} - s\mathbf{u}\|^2 = \|\mathbf{x} - \mathbf{p}\|^2 + \|\mathbf{p} - s\mathbf{u}\|^2 \geq \|\mathbf{x} - \mathbf{p}\|^2,$$

with equality only when $s\mathbf{u} = \mathbf{p}$.

Linearity. By Theorem 1.4.6, $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{u} = \mathbf{x} \cdot \mathbf{u} + \mathbf{y} \cdot \mathbf{u}$ and $(c\mathbf{x}) \cdot \mathbf{u} = c(\mathbf{x} \cdot \mathbf{u})$. Hence, by rules (8) and (6) of Theorem 1.2.7,

$$P(\mathbf{x} + \mathbf{y}) = (\mathbf{x} \cdot \mathbf{u} + \mathbf{y} \cdot \mathbf{u}) \mathbf{u} = P(\mathbf{x}) + P(\mathbf{y}), \quad P(c\mathbf{x}) = (c(\mathbf{x} \cdot \mathbf{u})) \mathbf{u} = cP(\mathbf{x}).$$

Matrix. By Theorem 3.2.7 the j -th column is $P(\mathbf{e}_j) = (\mathbf{e}_j \cdot \mathbf{u}) \mathbf{u} = u_j \mathbf{u}$, whose i -th entry is $u_j u_i = u_i u_j$. $\square$

If the line is given by a direction $\mathbf{w} \neq \mathbf{0}$ that is not a unit vector, use $\mathbf{u} = \mathbf{w}/\|\mathbf{w}\|$ (Section 1.4); then

$$P(\mathbf{x}) = \frac{\mathbf{x} \cdot \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}} \mathbf{w}.$$

Example 3.2.13 (Two projections in the plane). (a) Onto the first axis: $\mathbf{u} = \mathbf{e}_1$, so $P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $P(x, y) = (x, 0)$; the first entry is kept and the second replaced by 0.

(b) Onto the line $y = x$: $\mathbf{u} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, and every product $u_i u_j$ is $\frac{1}{2}$, so

$$P = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad P \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 1 \\ \frac{1}{2} \cdot 3 + \frac{1}{2} \cdot 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

The difference $(3, 1) - (2, 2) = (1, -1)$ satisfies $(1, -1) \cdot \mathbf{u} = 0$, as it should (Figure 3.6).

A projection flattens the plane onto the line: every point of the line through $\mathbf{x}$ perpendicular to L lands on the same point $P(\mathbf{x})$. So different vectors have the same image, and a projection cannot be undone; this is the first example of a matrix with no inverse (Section 3.5). Orthogonal projection onto planes and other subspaces is the subject of Chapter 10.

3.2.6 Reflections

The *reflection across L* sends $\mathbf{x}$ to its mirror image $Q(\mathbf{x})$ on the other side of L : the point for which $P(\mathbf{x})$ is the midpoint of the segment from $\mathbf{x}$ to $Q(\mathbf{x})$. So $Q(\mathbf{x}) - P(\mathbf{x}) = P(\mathbf{x}) - \mathbf{x}$, that is,

$$Q(\mathbf{x}) = 2P(\mathbf{x}) - \mathbf{x}.$$

Proposition 3.2.14 (Reflection across a line). The reflection $Q: \mathbb{R}^n \rightarrow \mathbb{R}^n$ across $L = \text{Span}\{\mathbf{u}\}$ is linear, and its standard matrix has entry $2u_i^2 - 1$ in each diagonal position (i, i) and $2u_i u_j$ in position (i, j) for $i \neq j$. In $\mathbb{R}^2$,

$$Q = \begin{bmatrix} 2u_1^2 - 1 & 2u_1 u_2 \\ 2u_1 u_2 & 2u_2^2 - 1 \end{bmatrix}.$$

Proof. By the linearity of P (Proposition 3.2.12) and the rules of vector algebra,

$$Q(\mathbf{x} + \mathbf{y}) = 2P(\mathbf{x}) + 2P(\mathbf{y}) - \mathbf{x} - \mathbf{y} = (2P(\mathbf{x}) - \mathbf{x}) + (2P(\mathbf{y}) - \mathbf{y}) = Q(\mathbf{x}) + Q(\mathbf{y}),$$

and $Q(c\mathbf{x}) = 2cP(\mathbf{x}) - c\mathbf{x} = cQ(\mathbf{x})$. The j -th column of the standard matrix is $Q(\mathbf{e}_j) = 2u_j\mathbf{u} - \mathbf{e}_j$, whose i -th entry is $2u_i u_j$, minus 1 when $i = j$. $\square$

Example 3.2.15 (Two reflections in the plane). (a) Across the line $y = x$, with P as in Example 3.2.13(b):

$$Q(\mathbf{e}_1) = 2 \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad Q(\mathbf{e}_2) = 2 \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

So $Q(x, y) = (y, x)$: the reflection swaps the two entries, and $Q(3, 1) = (1, 3)$. The projection $(2, 2)$ is the midpoint of $(3, 1)$ and $(1, 3)$ (Figure 3.6).

(b) Across the first axis, $\mathbf{u} = \mathbf{e}_1$ and $Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, so $Q(x, y) = (x, -y)$.

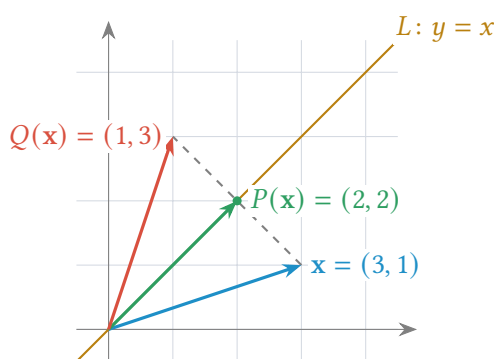


Figure 3.6: Examples 3.2.13(b) and 3.2.15(a): the projection of $\mathbf{x} = (3, 1)$ onto the line $y = x$ is $P(\mathbf{x}) = (2, 2)$, the foot of the perpendicular, and the reflection is $Q(\mathbf{x}) = (1, 3)$; $P(\mathbf{x})$ is the midpoint of $\mathbf{x}$ and $Q(\mathbf{x})$.

A reflection turns the plane over. Going counterclockwise around the origin, $\mathbf{e}_1$ comes before $\mathbf{e}_2$; after a reflection the images come in the opposite order (across $y = x$ they simply swap). A rotation keeps the order, since it turns both vectors by the same angle. This change of *orientation* is recorded by the sign of the determinant, which makes the argument precise (Chapter 7).

Check yourself

(a) Show that $T(x_1, x_2) = (x_2, x_1 - 3x_2)$ is linear and find its standard matrix. (b) Find the standard matrix of the rotation through 180° , and describe the map in words. (c) Show that $T(x_1, x_2) = (x_1 x_2, 0)$ is not linear. (d) Find the standard matrix of the projection onto the second axis, and of the reflection across it.

Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

Rotations keep lengths and angles. For $\mathbf{x} = (x, y)$, $R_\theta(\mathbf{x}) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$, and expanding,

$$\|R_\theta(\mathbf{x})\|^2 = x^2(\cos^2 \theta + \sin^2 \theta) + y^2(\sin^2 \theta + \cos^2 \theta) = x^2 + y^2 = \|\mathbf{x}\|^2,$$

the cross terms $\mp 2xy \cos \theta \sin \theta$ cancelling. The same computation for two vectors gives $R_\theta(\mathbf{x}) \cdot R_\theta(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$, so rotations also keep angles. Maps with this property are studied in Chapter 11 (orthogonal matrices).

Reflections in angle form. If L makes the angle φ with the first axis, then $\mathbf{u} = (\cos \varphi, \sin \varphi)$, and Proposition 3.2.14 with the double-angle formulas $2 \cos^2 \varphi - 1 = \cos 2\varphi$, $2 \sin \varphi \cos \varphi = \sin 2\varphi$, $2 \sin^2 \varphi - 1 = -\cos 2\varphi$ gives

$$Q = \begin{bmatrix} \cos 2\varphi & \sin 2\varphi \\ \sin 2\varphi & -\cos 2\varphi \end{bmatrix}.$$

For $\varphi = 45^\circ$ this is the swap matrix of Example 3.2.15(a).

Translations in computer graphics. A translation $\mathbf{x} \mapsto \mathbf{x} + \mathbf{b}$ with $\mathbf{b} \neq \mathbf{0}$ is not linear (Example 3.2.5), yet graphics software moves objects with matrices. The trick is to write the point (x, y) as $(x, y, 1)$ in $\mathbb{R}^3$ (*homogeneous coordinates*). Then

$$\begin{bmatrix} 1 & 0 & b_1 \\ 0 & 1 & b_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x + b_1 \\ y + b_2 \\ 1 \end{bmatrix},$$

so the translation of the plane becomes part of a linear map of $\mathbb{R}^3$, and rotations, scalings and translations can all be combined by the matrix multiplication of Section 3.3.

3.3 Matrix multiplication

Lecture 14

An $m \times n$ matrix A turns every vector $\mathbf{x} \in \mathbb{F}^n$ into the vector $A\mathbf{x} \in \mathbb{F}^m$, and this assignment is linear (Theorem 3.1.6); Section 3.2 calls it the linear transformation T_A of A (Proposition 3.2.2). In this section we run two such transformations one after the other: first a matrix B acts on $\mathbf{x}$, then A acts on the result $B\mathbf{x}$. We show that the combined transformation is again multiplication by a single matrix, which we call the product AB . We then give a quick rule for computing AB entry by entry, prove the rules the product obeys, and see where it differs from the multiplication of numbers: in general $AB \neq BA$, and a product of two nonzero matrices can be zero.

Throughout, $\mathbb{F}$ is a field (Section 2.1) and all matrices have entries in $\mathbb{F}$.

3.3.1 One matrix after another

Let B be an $n \times p$ matrix and A an $m \times n$ matrix. Then B sends $\mathbb{F}^p$ to $\mathbb{F}^n$ and A sends $\mathbb{F}^n$ to $\mathbb{F}^m$, so the output of B is a valid input for A :

$$\mathbb{F}^p \xrightarrow{B} \mathbb{F}^n \xrightarrow{A} \mathbb{F}^m, \quad \mathbf{x} \mapsto B\mathbf{x} \mapsto A(B\mathbf{x}).$$

To see which matrix, if any, does both steps at once, let $\mathbf{b}_1, \dots, \mathbf{b}_p \in \mathbb{F}^n$ be the columns of B and $\mathbf{x} = (x_1, \dots, x_p)$. By the column rule (Definition 3.1.1), $B\mathbf{x} = x_1\mathbf{b}_1 + \dots + x_p\mathbf{b}_p$, and since A respects linear combinations (Theorem 3.1.6, extended to p terms),

$$A(B\mathbf{x}) = A(x_1\mathbf{b}_1 + \dots + x_p\mathbf{b}_p) = x_1(A\mathbf{b}_1) + \dots + x_p(A\mathbf{b}_p).$$

The right-hand side is a linear combination of the p vectors $A\mathbf{b}_1, \dots, A\mathbf{b}_p \in \mathbb{F}^m$ with the entries of $\mathbf{x}$ as weights. By the column rule again, it is the product of $\mathbf{x}$ with the $m \times p$ matrix whose columns are $A\mathbf{b}_1, \dots, A\mathbf{b}_p$. This matrix is the product of A and B .

Definition 3.3.1 (Matrix product). Let A be an $m \times n$ matrix and B an $n \times p$ matrix with columns $\mathbf{b}_1, \dots, \mathbf{b}_p$. The *product* AB is the $m \times p$ matrix

$$AB = [A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \dots \quad A\mathbf{b}_p],$$

whose j -th column is A times the j -th column of B . The product is defined only when A has as many columns as B has rows.

The computation before the definition proves the property that the definition was built to have.

Theorem 3.3.2 (The product is composition). Let A be $m \times n$ and B be $n \times p$. Then for every $\mathbf{x} \in \mathbb{F}^p$,

$$(AB)\mathbf{x} = A(B\mathbf{x}).$$

Proof. Write $\mathbf{x} = (x_1, \dots, x_p)$. By Definition 3.1.1 applied to the matrix AB , whose columns are $A\mathbf{b}_1, \dots, A\mathbf{b}_p$, we have $(AB)\mathbf{x} = x_1(A\mathbf{b}_1) + \dots + x_p(A\mathbf{b}_p)$. By Definition 3.1.1 applied to B , and Theorem 3.1.6 (part (c), extended to p terms by induction), $A(B\mathbf{x}) = A(x_1\mathbf{b}_1 + \dots + x_p\mathbf{b}_p) = x_1(A\mathbf{b}_1) + \dots + x_p(A\mathbf{b}_p)$. The two vectors are equal. $\square$

Notice the order: in AB the matrix on the *right* acts first. This is the same convention as for composing functions, where $(f \circ g)(x) = f(g(x))$ means “apply g , then f ”.

Proposition 3.3.3 (Composition of linear transformations). Let $S: \mathbb{F}^p \rightarrow \mathbb{F}^n$ and $T: \mathbb{F}^n \rightarrow \mathbb{F}^m$ be linear transformations with standard matrices B and A (Theorem 3.2.7). Then the composite $T \circ S: \mathbb{F}^p \rightarrow \mathbb{F}^m$, $\mathbf{x} \mapsto T(S(\mathbf{x}))$, is a linear transformation, and its standard matrix is AB .

Proof. For every $\mathbf{x} \in \mathbb{F}^p$ we have $S(\mathbf{x}) = B\mathbf{x}$ and $T(\mathbf{y}) = A\mathbf{y}$ for $\mathbf{y} \in \mathbb{F}^n$, so by Theorem 3.3.2, $(T \circ S)(\mathbf{x}) = A(B\mathbf{x}) = (AB)\mathbf{x}$. Thus $T \circ S$ is multiplication by the matrix AB , which is linear by Theorem 3.1.6. Its standard matrix has j -th column $(T \circ S)(\mathbf{e}_j) = (AB)\mathbf{e}_j$, which is the j -th column of AB (Section 3.1: multiplying by $\mathbf{e}_j$ picks out column j). So the standard matrix is AB . $\square$

Sizes. Each column $\mathbf{b}_j$ of B must be a valid input for A , so AB needs B to have as many rows as A has columns. The product has one column for each column of B , and each column $A\mathbf{b}_j$ has one entry for each row of A . In shorthand,

$$(m \times n)(n \times p) = (m \times p) :$$

the inner sizes must match, and the outer sizes give the size of the product. When $p = 1$, B is a single column $\mathbf{x} \in \mathbb{F}^n$ and AB is the matrix-vector product $A\mathbf{x}$ of Section 3.1, so the new product extends the old one.

Example 3.3.4 (Rotate, then shear). Let

$$B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

B rotates the plane a quarter turn counterclockwise, and A is a shear (Examples 3.2.10 and 3.2.11). The columns of B are $(0, 1)$ and $(-1, 0)$, so by Definition 3.3.1 and the column rule

$$A \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A \begin{bmatrix} -1 \\ 0 \end{bmatrix} = (-1) \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix},$$

and therefore

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

By Theorem 3.3.2, multiplying by AB rotates a quarter turn and then shears. Figure 3.7 follows the unit square, whose sides are $\mathbf{e}_1$ and $\mathbf{e}_2$: the rotation turns it into the square with sides $(0, 1)$ and $(-1, 0)$, and the shear then slides it to the parallelogram with sides $(1, 1)$ and $(-1, 0)$, the columns of AB . For a single vector, $\mathbf{x} = (1, 2)$: $B\mathbf{x} = (-2, 1)$ and $A(B\mathbf{x}) = (-2 + 1, 1) = (-1, 1)$, while $(AB)\mathbf{x} = (1 \cdot 1 + (-1) \cdot 2, 1 \cdot 1 + 0 \cdot 2) = (-1, 1)$.

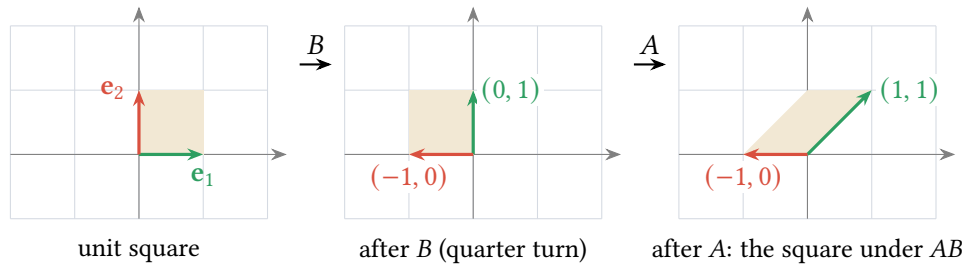


Figure 3.7: Example 3.3.4: the unit square (left), after the quarter turn B (middle), and after the shear A (right). The final sides $(1, 1)$ and $(-1, 0)$ are the columns of AB ; green marks where $\mathbf{e}_1$ goes, coral where $\mathbf{e}_2$ goes.

3.3.2 The row-column rule

Definition 3.3.1 builds AB one column at a time. By hand it is usually quicker to compute one entry at a time. We write $(AB)_{ij}$ for the entry of AB in row i and column j .

Theorem 3.3.5 (Row-column rule). Let $A = (a_{ik})$ be $m \times n$ and $B = (b_{kj})$ be $n \times p$. For $i = 1, \dots, m$ and $j = 1, \dots, p$,

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

In words: entry (i, j) of AB is row i of A times column j of B , entry by entry and added up; it is the dot product $\mathbf{r}_i \cdot \mathbf{b}_j$ of the i -th row of A with the j -th column of B .

Proof. Entry (i, j) of AB is entry i of its j -th column, which is $A\mathbf{b}_j$ with $\mathbf{b}_j = (b_{1j}, \dots, b_{nj})$. By the row rule (Proposition 3.1.4), entry i of $A\mathbf{b}_j$ is $a_{i1}b_{1j} + \cdots + a_{in}b_{nj}$. $\square$

Example 3.3.6. Let

$$A = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 0 & -1 \\ 1 & 2 \end{bmatrix}.$$

A is 2×3 and B is 3×2 , so AB is 2×2 . Row by column:

$$\begin{aligned}(AB)_{11} &= 1 \cdot 2 + 2 \cdot 0 + 0 \cdot 1 = 2, & (AB)_{12} &= 1 \cdot 1 + 2 \cdot (-1) + 0 \cdot 2 = 1 - 2 = -1, \\ (AB)_{21} &= (-1) \cdot 2 + 1 \cdot 0 + 3 \cdot 1 = -2 + 3 = 1, & (AB)_{22} &= (-1) \cdot 1 + 1 \cdot (-1) + 3 \cdot 2 = -1 - 1 + 6 = 4,\end{aligned}$$

so

$$AB = \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix}.$$

The product BA is defined as well, since B has two columns and A has two rows, but it is 3×3 . Row i of B times column j of A gives

$$BA = \begin{bmatrix} 2 \cdot 1 + 1 \cdot (-1) & 2 \cdot 2 + 1 \cdot 1 & 2 \cdot 0 + 1 \cdot 3 \\ 0 \cdot 1 + (-1) \cdot (-1) & 0 \cdot 2 + (-1) \cdot 1 & 0 \cdot 0 + (-1) \cdot 3 \\ 1 \cdot 1 + 2 \cdot (-1) & 1 \cdot 2 + 2 \cdot 1 & 1 \cdot 0 + 2 \cdot 3 \end{bmatrix} = \begin{bmatrix} 1 & 5 & 3 \\ 1 & -1 & -3 \\ -1 & 4 & 6 \end{bmatrix}.$$

AB and BA do not even have the same size.

Rows of a product. Fix i and let j run from 1 to p : by the row-column rule, row i of AB is $(\sum_k a_{ik}b_{k1}, \dots, \sum_k a_{ik}b_{kp}) = a_{i1}(\text{row 1 of } B) + \dots + a_{in}(\text{row } n \text{ of } B)$. So each row of AB is a combination of the rows of B , with weights from the matching row of A , just as each column of AB is a combination of the columns of A , with weights from the matching column of B .

3.3.3 Adding and scaling matrices

The rules of the product involve two simpler operations, which act on matrices exactly as the vector operations act on vectors.

Definition 3.3.7 (Sum and scalar multiple of matrices). Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ matrices and $c \in \mathbb{F}$. The *sum* $A + B$ and the *scalar multiple* cA are the $m \times n$ matrices with entries

$$(A + B)_{ij} = a_{ij} + b_{ij}, \quad (cA)_{ij} = c a_{ij}.$$

The *zero matrix* O (or $O_{m \times n}$) has every entry 0, and $-A = (-1)A$. The sum is defined only for matrices of the same size.

For example,

$$\begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} 3 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 1 & 0 \end{bmatrix}, \quad 3 \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 0 & -3 \end{bmatrix}.$$

An $m \times n$ matrix is a list of mn entries arranged in a rectangle, and both operations act entry by entry. So the eight rules of vector algebra (Theorem 1.2.7) hold for $m \times n$ matrices, with O in the role of $\mathbf{0}$, and with the same proofs. In terms of columns: the columns of $A + B$ are $\mathbf{a}_j + \mathbf{b}_j$ and those of cA are $c \mathbf{a}_j$.

Proposition 3.3.8. Let A and B be $m \times n$ matrices, $c \in \mathbb{F}$ and $\mathbf{x} \in \mathbb{F}^n$. Then

$$(A + B)\mathbf{x} = A\mathbf{x} + B\mathbf{x}, \quad (cA)\mathbf{x} = c(A\mathbf{x}).$$

Proof. By the row rule and Definition 3.3.7, entry i of $(A + B)\mathbf{x}$ is $\sum_k (a_{ik} + b_{ik})x_k = \sum_k a_{ik}x_k + \sum_k b_{ik}x_k$ (distributivity, commutativity and associativity in $\mathbb{F}$), which is entry i of $A\mathbf{x} + B\mathbf{x}$. Likewise entry i of $(cA)\mathbf{x}$ is $\sum_k (c a_{ik})x_k = c \sum_k a_{ik}x_k$, entry i of $c(A\mathbf{x})$. $\square$

3.3.4 The rules of the product

Theorem 3.3.9 (Rules of matrix multiplication). Let $c \in \mathbb{F}$, and let A, B, C be matrices of sizes for which each side below is defined. Then:

- (a) $(AB)C = A(BC)$ (associativity)
- (b) $A(B + C) = AB + AC$ (distributive law, left)
- (c) $(A + B)C = AC + BC$ (distributive law, right)
- (d) $c(AB) = (cA)B = A(cB)$ (scalars)
- (e) $I_m A = A = A I_n$ for every $m \times n$ matrix A (identity)

Each rule is an equation between two matrices of the same size, and two matrices are equal when their columns are equal. So each proof compares the j -th columns of the two sides, using Definition 3.3.1 (column j of XY is X times column j of Y) and Definition 3.3.7 (column j of $Y + Z$ is the sum of the j -th columns, column j of cY is c times the j -th column).

Proof. Write $\mathbf{b}_j$ and $\mathbf{c}_j$ for the j -th columns of B and C .

- (a) Let A be $m \times n$, B be $n \times p$ and C be $p \times q$. Column j of $(AB)C$ is $(AB)\mathbf{c}_j$, which equals $A(B\mathbf{c}_j)$ by Theorem 3.3.2. Column j of BC is $B\mathbf{c}_j$, so column j of $A(BC)$ is $A(B\mathbf{c}_j)$ as well. All q columns agree.
- (b) Column j of $A(B + C)$ is $A(\mathbf{b}_j + \mathbf{c}_j) = A\mathbf{b}_j + A\mathbf{c}_j$ by Theorem 3.1.6(a), which is column j of AB plus column j of AC , that is, column j of $AB + AC$.
- (c) Column j of $(A + B)C$ is $(A + B)\mathbf{c}_j = A\mathbf{c}_j + B\mathbf{c}_j$ by Proposition 3.3.8, which is column j of $AC + BC$.
- (d) Column j of $c(AB)$ is $c(A\mathbf{b}_j)$. Column j of $(cA)B$ is $(cA)\mathbf{b}_j = c(A\mathbf{b}_j)$ by Proposition 3.3.8, and column j of $A(cB)$ is $A(c\mathbf{b}_j) = c(A\mathbf{b}_j)$ by Theorem 3.1.6(b).
- (e) Column j of $I_m A$ is $I_m \mathbf{a}_j = \mathbf{a}_j$, since the identity leaves every vector unchanged (Section 3.1). Column j of $A I_n$ is $A\mathbf{e}_j$, where $\mathbf{e}_j$, the j -th column of I_n , has a 1 in position j and zeros elsewhere; so $A\mathbf{e}_j = \mathbf{a}_j$ (Section 3.1). $\square$

Associativity has a short meaning: doing C , then B , then A gives the same result however the steps are grouped. It lets us write ABC , and longer products, without brackets. The rules also give, for instance, $AO = O$ and $OA = O$ whenever the sizes allow, since column j of AO is $A\mathbf{0} = \mathbf{0}$, and every column of OA is $O\mathbf{a}_j = \mathbf{0}$.

Definition 3.3.10 (Powers). Let A be an $n \times n$ matrix. Its powers are $A^0 = I_n$, $A^1 = A$, $A^2 = AA$, and in general $A^{k+1} = A^k A$ for $k \geq 0$.

By associativity, A^k is the product of k copies of A grouped in any way, and $A^j A^k = A^{j+k}$ for all $j, k \geq 0$ (by induction on k : $A^j A^0 = A^j I_n = A^j$, and $A^j A^{k+1} = A^j (A^k A) = (A^j A^k) A = A^{j+k} A = A^{j+k+1}$). Only square matrices have powers: AA is defined only when A has as many columns as rows.

3.3.5 The order matters

For numbers, $ab = ba$. For matrices this fails even when both products are defined and have the same size.

Example 3.3.11 (Shear, then rotate). Take A and B from Example 3.3.4. The columns of A are $(1, 0)$ and $(1, 1)$, so

$$B \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad B \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad \text{so} \quad BA = \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \neq \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} = AB.$$

Shearing and then rotating sends the unit square to the parallelogram with sides $(0, 1)$ and $(-1, 1)$; rotating and then shearing sends it to the one with sides $(1, 1)$ and $(-1, 0)$ (Figure 3.8).

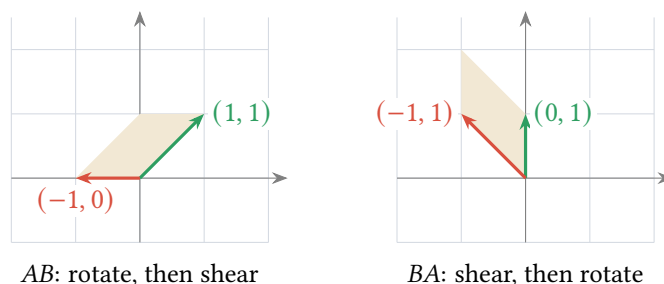


Figure 3.8: Example 3.3.11: the unit square under AB (rotate, then shear; left) and under BA (shear, then rotate; right). The two results differ, so $AB \neq BA$.

So matrix multiplication is not *commutative*: we must keep track of the order, and we speak of multiplying B by A *on the left* (AB) or *on the right* (BA). When $AB = BA$ we say that A and B *commute*. Some pairs do, for example every $n \times n$ matrix with I_n (Theorem 3.3.9(e)), every square matrix with its own powers ($A A^k = A^{k+1} = A^k A$), and any two rotations of the plane: by the angle addition formulas, $R_\alpha R_\beta = \begin{bmatrix} \cos \alpha \cos \beta - \sin \alpha \sin \beta & -(\sin \alpha \cos \beta + \cos \alpha \sin \beta) \\ \sin \alpha \cos \beta + \cos \alpha \sin \beta & \cos \alpha \cos \beta - \sin \alpha \sin \beta \end{bmatrix} = R_{\alpha+\beta}$, which is symmetric in α and β , so $R_\alpha R_\beta = R_\beta R_\alpha$. But commuting is the exception.

A consequence is that familiar algebra must be redone with care. For $n \times n$ matrices, the rules give

$$(A + B)^2 = (A + B)(A + B) = A^2 + AB + BA + B^2,$$

which equals $A^2 + 2AB + B^2$ only when $AB = BA$. Likewise $(AB)^2 = ABAB$ need not equal $A^2 B^2 = A A B B$.

Example 3.3.12 (A nonzero matrix with zero square). Let

$$N = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Its columns are $\mathbf{0}$ and $\mathbf{e}_1 = (1, 0)$, so $N\mathbf{e}_1 = \mathbf{0}$ and $N\mathbf{e}_2 = \mathbf{e}_1$. The columns of N^2 are

$$N \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad N \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ (the first column of } N), \quad \text{so} \quad N^2 = O,$$

although $N \neq O$. Geometrically, N sends $\mathbf{e}_2$ to $\mathbf{e}_1$ and $\mathbf{e}_1$ to $\mathbf{0}$; applied twice, it sends every vector to $\mathbf{0}$.

For numbers, $ab = 0$ forces $a = 0$ or $b = 0$; Example 3.3.12 shows that this fails for matrices. It follows that matrices cannot be cancelled in general: $NN = O = NO$, yet $N \neq O$. So from $AB = AC$ we may not conclude $B = C$. Section 3.5 shows that an *invertible* matrix can always be cancelled.

Check yourself

Let $A = \begin{bmatrix} 2 & 0 \\ 1 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$. (a) Compute AB column by column, using Definition 3.3.1. (b) Compute BA with the row-column rule. Do A and B commute? (c) Find a 2×2 matrix $M \neq O$ with $M^2 = O$, other than N of Example 3.3.12.

Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

Columns times rows. A column $\mathbf{u} \in \mathbb{F}^m$ is an $m \times 1$ matrix and a row $\mathbf{r}$ with p entries is a $1 \times p$ matrix, so $\mathbf{u} \mathbf{r}$ is an $m \times p$ matrix, with entry (i, j) equal to $u_i r_j$. Grouping the sum in the row-column rule by k instead of by (i, j) gives

$$AB = \mathbf{a}_1 \mathbf{r}_1 + \mathbf{a}_2 \mathbf{r}_2 + \cdots + \mathbf{a}_n \mathbf{r}_n,$$

where $\mathbf{a}_k$ is the k -th column of A and $\mathbf{r}_k$ the k -th row of B : entry (i, j) of the right-hand side is $\sum_k a_{ik} b_{kj}$. For Example 3.3.6,

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix} + \begin{bmatrix} 0 & -2 \\ 0 & -1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix}.$$

Writing a matrix as a sum of such column-times-row pieces is the starting point of the singular value decomposition in Chapter 11.

How much work is AB ? By the row-column rule each of the mp entries needs n multiplications, so AB costs mnp multiplications; for two $n \times n$ matrices that is n^3 . In 1969 Volker Strassen found a way to multiply 2×2 matrices, even when their entries are themselves matrices, with 7 multiplications instead of 8. Applied over and over to blocks, it multiplies $n \times n$ matrices with roughly $n^{2.81}$ multiplications, and it was the first method to beat n^3 . How small the exponent can be made is still an open question.

Why not multiply entry by entry? One could define a product of two $m \times n$ matrices entry by entry, as for the sum. That product exists and has uses, but it does not describe doing one transformation after another, so it does not satisfy Theorem 3.3.2. The row-column rule may look strange at first; it is exactly what composition requires.

3.4 Matrix algebra

Lecture 15

Section 3.3 gave us three operations on matrices: the sum $A + B$ of two matrices of the same size and the scalar multiple cA (Definition 3.3.7), and the product AB (Definition 3.3.1) when the number of columns of A equals the number of rows of B . This section adds a fourth operation, the *transpose*, which turns rows into columns, and studies the special shapes of matrix that recur throughout the course: symmetric, triangular and diagonal matrices. We then learn to cut a matrix into blocks and multiply block by block, and we finish with the *trace*, a single number attached to every square matrix.

Throughout, $\mathbb{F}$ is a field, and all matrices have entries in $\mathbb{F}$. We write a_{ij} for the entry of A in row i and column j , and also $(A)_{ij}$ when A is itself an expression such as AB . By Theorem 3.3.5 the product of an $m \times n$ matrix A and an $n \times p$ matrix B is the $m \times p$ matrix with entries

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

row i of A times column j of B , and its j -th column is $A\mathbf{b}_j$, where $\mathbf{b}_j$ is the j -th column of B .

3.4.1 The transpose

Definition 3.4.1 (Transpose). The *transpose* of an $m \times n$ matrix A is the $n \times m$ matrix A^T whose rows are the columns of A :

$$(A^T)_{ij} = a_{ji} \quad (1 \leq i \leq n, 1 \leq j \leq m).$$

So the entry of A^T in row i and column j is the entry of A in row j and column i : transposing swaps the two indices. Row i of A^T is column i of A , and column j of A^T is row j of A .

Example 3.4.2.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

A is 2×3 and A^T is 3×2 . The entry $2 = a_{12}$ of A is the entry $(A^T)_{21}$ of A^T . A column vector $\mathbf{v} = (v_1, \dots, v_n)$ is an $n \times 1$ matrix, and its transpose $\mathbf{v}^T = [v_1 \ \cdots \ v_n]$ is a $1 \times n$ matrix, a row vector.

For a square matrix the *main diagonal* is formed by the entries $a_{11}, a_{22}, \dots, a_{nn}$, from the top left to the bottom right. Transposing a square matrix reflects it across its main diagonal: the diagonal entries stay in place, and the entry in position (i, j) changes places with the entry in position (j, i) (Figure 3.9).

The figure shows two 3x3 matrices, A and A^T, connected by a horizontal arrow labeled 'transpose'. Matrix A has entries: row 1 [2, 1, 7], row 2 [0, 3, 4], row 3 [6, 8, 5]. Matrix A^T has entries: row 1 [2, 0, 6], row 2 [1, 3, 8], row 3 [7, 4, 5]. A gold diagonal line runs from the top-left to the bottom-right in both matrices. The entries 2, 3, and 5 are on the diagonal and remain in their positions. The entry 1 at (1,2) moves to 2 at (2,1). The entry 7 at (1,3) moves to 6 at (3,1). The entry 4 at (2,3) moves to 8 at (3,2). The entries 6 and 8 in the third row of A are highlighted in blue, and their corresponding entries 6 and 8 in the first and second rows of A^T are also highlighted in blue. The entries 2, 1, 7 in the first row of A are highlighted in red, and their corresponding entries 2, 1, 7 in the first, second, and third columns of A^T are also highlighted in red.

Figure 3.9: Transposing a square matrix reflects it across the main diagonal (gold). The entry in position $(1, 3)$ and the entry in position $(3, 1)$ change places; the diagonal entries stay where they are.

Theorem 3.4.3 (Rules for the transpose). Let $c \in \mathbb{F}$. Whenever the sizes allow the operations on the left:

- (a) $(A^T)^T = A$;
- (b) $(A + B)^T = A^T + B^T$;
- (c) $(cA)^T = cA^T$;
- (d) $(AB)^T = B^T A^T$.

Proof. In each part both sides are matrices of the same size, so it is enough to compare entries.

(a) $((A^T)^T)_{ij} = (A^T)_{ji} = a_{ij}.$

(b) $((A + B)^T)_{ij} = (A + B)_{ji} = a_{ji} + b_{ji} = (A^T)_{ij} + (B^T)_{ij} = (A^T + B^T)_{ij}.$

(c) $((cA)^T)_{ij} = (cA)_{ji} = c a_{ji} = c (A^T)_{ij}.$

(d) Let A be $m \times n$ and B be $n \times p$. Then AB is $m \times p$ and $(AB)^T$ is $p \times m$; also B^T is $p \times n$ and A^T is $n \times m$, so $B^T A^T$ is defined and is $p \times m$. For $1 \leq i \leq p$ and $1 \leq j \leq m$,

$$((AB)^T)_{ij} = (AB)_{ji} = \sum_{k=1}^n a_{jk} b_{ki}, \quad (B^T A^T)_{ij} = \sum_{k=1}^n (B^T)_{ik} (A^T)_{kj} = \sum_{k=1}^n b_{ki} a_{jk}.$$

Multiplication in $\mathbb{F}$ is commutative, so $b_{ki} a_{jk} = a_{jk} b_{ki}$ for every k , and the two sums are equal. $\square$

The order in (d) must be reversed: if A is 2×3 and B is 3×4 , then $(AB)^T$ is 4×2 , while $A^T B^T$ would be a $(3 \times 2)(4 \times 3)$ product, which is not even defined. Even for square matrices, where both orders are defined, $A^T B^T$ is usually wrong.

Example 3.4.4. Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 3 & 1 \cdot 0 + 2 \cdot 1 \\ 0 \cdot 1 + 1 \cdot 3 & 0 \cdot 0 + 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 3 & 1 \end{bmatrix}, \quad (AB)^T = \begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix},$$

and

$$B^T A^T = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1+6 & 0+3 \\ 0+2 & 0+1 \end{bmatrix} = \begin{bmatrix} 7 & 3 \\ 2 & 1 \end{bmatrix} = (AB)^T,$$

while

$$A^T B^T = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} \neq (AB)^T.$$

Applying (d) twice, $(ABC)^T = ((AB)C)^T = C^T(AB)^T = C^T B^T A^T$, and by induction $(A_1 A_2 \cdots A_k)^T = A_k^T \cdots A_2^T A_1^T$.

Dot products as matrix products. We identify a 1×1 matrix $[c]$ with its single entry c . For $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$, the product $\mathbf{u}^T \mathbf{v}$ of the $1 \times n$ matrix $\mathbf{u}^T$ and the $n \times 1$ matrix $\mathbf{v}$ is the 1×1 matrix with entry $u_1 v_1 + \cdots + u_n v_n$. So

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}.$$

Proposition 3.4.5. Let A be an $m \times n$ matrix, $\mathbf{x} \in \mathbb{F}^n$ and $\mathbf{y} \in \mathbb{F}^m$. Then $(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A^T \mathbf{y})$.

Proof. By Theorem 3.4.3(d) and associativity of the product (Theorem 3.3.9(a)), $(A\mathbf{x}) \cdot \mathbf{y} = (A\mathbf{x})^T \mathbf{y} = (\mathbf{x}^T A^T) \mathbf{y} = \mathbf{x}^T (A^T \mathbf{y}) = \mathbf{x} \cdot (A^T \mathbf{y})$. $\square$

This identity is how the transpose enters geometry: in Chapter 10 it leads to orthogonal projections and least squares.

3.4.2 Symmetric matrices

Definition 3.4.6 (Symmetric and skew-symmetric matrices). A square matrix A is *symmetric* if $A^T = A$, that is, $a_{ij} = a_{ji}$ for all i, j . It is *skew-symmetric* if $A^T = -A$, that is, $a_{ji} = -a_{ij}$ for all i, j .

A non-square matrix cannot be symmetric, since A^T and A have different sizes. In a symmetric matrix each entry above the diagonal equals its mirror image below the diagonal.

Example 3.4.7.

$$S = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 4 \\ 0 & 4 & 5 \end{bmatrix} \text{ is symmetric,} \quad K = \begin{bmatrix} 0 & 2 & -1 \\ -2 & 0 & 3 \\ 1 & -3 & 0 \end{bmatrix} \text{ is skew-symmetric.}$$

For a skew-symmetric matrix, $a_{ii} = -a_{ii}$, so $(1+1)a_{ii} = 0$. Over $\mathbb{R}$ (indeed over any field in which $1+1 \neq 0$) this forces $a_{ii} = 0$: the diagonal of a skew-symmetric matrix is zero. Over $\mathbb{Z}_2$, where $-1 = 1$, skew-symmetric and symmetric mean the same thing.

Proposition 3.4.8. Symmetric matrices can be built from any matrix:

- (a) For every $m \times n$ matrix A , the $n \times n$ matrix $A^T A$ and the $m \times m$ matrix AA^T are symmetric. Moreover $(A^T A)_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j$, where $\mathbf{a}_1, \dots, \mathbf{a}_n$ are the columns of A .
- (b) For every square matrix A , the matrix $A + A^T$ is symmetric and $A - A^T$ is skew-symmetric.
- (c) If $1 + 1 \neq 0$ in $\mathbb{F}$ (for instance over $\mathbb{R}$ or $\mathbb{C}$), every square matrix is, in exactly one way, the sum of a symmetric and a skew-symmetric matrix:

$$A = \frac{1}{2} (A + A^T) + \frac{1}{2} (A - A^T).$$

Proof. (a) By Theorem 3.4.3(d) and (a), $(A^T A)^T = A^T (A^T)^T = A^T A$, and likewise $(AA^T)^T = (A^T)^T A^T = AA^T$. For the entries: row i of A^T is column i of A , so $(A^T A)_{ij} = \sum_k (A^T)_{ik} a_{kj} = \sum_k a_{ki} a_{kj} = \mathbf{a}_i \cdot \mathbf{a}_j$.

(b) By Theorem 3.4.3(a)–(c), $(A + A^T)^T = A^T + A = A + A^T$ and $(A - A^T)^T = A^T - A = -(A - A^T)$.

(c) Here $\frac{1}{2}$ denotes the inverse of $1 + 1$ in $\mathbb{F}$. The two matrices in the displayed sum are symmetric and skew-symmetric by (b) and Theorem 3.4.3(c), and they add up to $\frac{1}{2}A + \frac{1}{2}A = A$. For uniqueness, suppose $A = S + K = S' + K'$ with S, S' symmetric and K, K' skew-symmetric. Then $M = S - S' = K' - K$ is both symmetric and skew-symmetric, so $M = M^T = -M$ and $(1 + 1)M = O$, hence $M = O$ (multiply by $\frac{1}{2}$). So $S = S'$ and $K = K'$. $\square$

Example 3.4.9. Let A have columns $\mathbf{a}_1 = (1, 0, 1)$ and $\mathbf{a}_2 = (2, 1, 0)$. Then $\mathbf{a}_1 \cdot \mathbf{a}_1 = 1 + 0 + 1 = 2$, $\mathbf{a}_1 \cdot \mathbf{a}_2 = 2 + 0 + 0 = 2$ and $\mathbf{a}_2 \cdot \mathbf{a}_2 = 4 + 1 + 0 = 5$, so

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad A^T A = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}, \quad AA^T = \begin{bmatrix} 5 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

both symmetric (the entries of AA^T are the dot products of the rows of A).

Matrices of the form $A^T A$ return in least squares (Section 10.4) and in the singular value decomposition (Chapter 11), and symmetric matrices are the subject of the spectral theorem (Section 11.1).

3.4.3 Triangular and diagonal matrices

Definition 3.4.10 (Triangular and diagonal matrices). A square matrix $A = (a_{ij})$ is

- *upper triangular* if $a_{ij} = 0$ whenever $i > j$ (all entries below the main diagonal are zero);
- *lower triangular* if $a_{ij} = 0$ whenever $i < j$ (all entries above the main diagonal are zero);
- *diagonal* if $a_{ij} = 0$ whenever $i \neq j$, that is, if it is both upper and lower triangular.

The diagonal matrix with diagonal entries $d_1, \dots, d_n$ is written $\text{diag}(d_1, \dots, d_n)$.

The identity matrix $I_n = \text{diag}(1, \dots, 1)$ and the zero matrix are diagonal (Figure 3.10). An echelon form U of a square matrix is upper triangular: the leading entry of each nonzero row lies strictly to the right of the leading entry of the row above, so by induction the leading entry of row i lies in a column $j \geq i$, and every entry of row i to the left of column i is zero; zero rows have no nonzero entries at all. Since A^T has $(A^T)_{ij} = a_{ji}$, the transpose of an upper triangular matrix is lower triangular and vice versa.

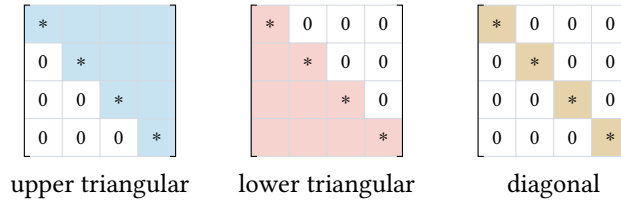


Figure 3.10: Upper triangular, lower triangular and diagonal 4×4 matrices. Entries in the white regions are zero; the shaded entries may take any value.

Theorem 3.4.11 (Products of triangular matrices). Let A and B be $n \times n$ matrices.

- (a) If A and B are upper triangular, then AB is upper triangular and $(AB)_{ii} = a_{ii}b_{ii}$ for every i .
- (b) If A and B are lower triangular, then AB is lower triangular and $(AB)_{ii} = a_{ii}b_{ii}$ for every i .
- (c) $\text{diag}(d_1, \dots, d_n) \text{diag}(e_1, \dots, e_n) = \text{diag}(d_1e_1, \dots, d_ne_n)$. In particular, diagonal matrices of the same size commute.

Proof. (a) We have $(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$. Since A is upper triangular, $a_{ik} = 0$ unless $k \geq i$; since B is upper triangular, $b_{kj} = 0$ unless $k \leq j$. So the k -th term can be nonzero only if $i \leq k \leq j$. If $i > j$ there is no such k , every term is zero, and $(AB)_{ij} = 0$; so AB is upper triangular. If $i = j$ the only possible k is $k = i$, and $(AB)_{ii} = a_{ii}b_{ii}$.

(b) A^T and B^T are upper triangular, so by Theorem 3.4.3(d) and part (a), $(AB)^T = B^T A^T$ is upper triangular, hence $AB = ((AB)^T)^T$ is lower triangular. Its diagonal entries are those of $B^T A^T$, namely $b_{ii}a_{ii} = a_{ii}b_{ii}$.

(c) A diagonal matrix is both upper and lower triangular, so by (a) and (b) the product is both, hence diagonal, with diagonal entries $d_i e_i$. Since $d_i e_i = e_i d_i$, the product in the other order is the same matrix. $\square$

Example 3.4.12. Row by column,

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 4 & 5 \\ 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 \cdot 4 + 2 \cdot 0 & 1 \cdot 5 + 2 \cdot 6 \\ 0 \cdot 4 + 3 \cdot 0 & 0 \cdot 5 + 3 \cdot 6 \end{bmatrix} = \begin{bmatrix} 4 & 17 \\ 0 & 18 \end{bmatrix};$$

upper triangular, with diagonal entries $4 = 1 \cdot 4$ and $18 = 3 \cdot 6$. The entry above the diagonal, 17, is not simply $2 \cdot 5$; only the diagonal entries multiply in this simple way.

Triangular matrices are the reason elimination works so well: in Section 3.7 (Lecture 18) we write a matrix as a product $A = LU$ of a lower triangular and an upper triangular matrix.

3.4.4 Block matrices

Large matrices often have structure that is easier to see in pieces. Cutting a matrix by horizontal lines between some of its rows and vertical lines between some of its columns divides it into smaller matrices, its *blocks*; a matrix viewed this way is a *block matrix* (or *partitioned matrix*). For example,

$$M = \left[\begin{array}{cc|cc} 1 & 0 & 1 & 2 \\ 0 & 1 & 3 & 4 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] = \begin{bmatrix} I & X \\ O & I \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},$$

where $I = I_2$ and O is the 2×2 zero matrix. In general we write $A = (A_{rs})$, where A_{rs} is the block in the r -th row of blocks and the s -th column of blocks.

Theorem 3.4.13 (Block multiplication). Let A be $m \times n$ and B be $n \times p$. Cut the rows of A into groups of sizes $m_1, \dots, m_R$, the columns of A into groups of sizes $n_1, \dots, n_S$, the rows of B into the same groups $n_1, \dots, n_S$, and the columns of B into groups of sizes $p_1, \dots, p_T$. Then AB , cut into rows $m_1, \dots, m_R$ and columns $p_1, \dots, p_T$, has blocks

$$(AB)_{rt} = \sum_{s=1}^S A_{rs} B_{st} = A_{r1} B_{1t} + A_{r2} B_{2t} + \dots + A_{rS} B_{St}.$$

In particular, for 2×2 block matrices with matching cuts,

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}.$$

The condition on the cuts makes every product $A_{rs}B_{st}$ defined: A_{rs} has n_s columns and B_{st} has n_s rows. Within each product the order must be kept, since matrices do not commute.

Proof. Let K_s be the set of column indices of A in the s -th group; it is also the set of row indices of B in the s -th group, and $\{1, \dots, n\}$ is the disjoint union of $K_1, \dots, K_S$. Take an entry of AB in row i and column j , where i lies in the r -th row group and j in the t -th column group. Splitting the sum over k according to the groups,

$$(AB)_{ij} = \sum_{k=1}^n a_{ik} b_{kj} = \sum_{s=1}^S \sum_{k \in K_s} a_{ik} b_{kj}.$$

For fixed s , the numbers a_{ik} with $k \in K_s$ form the row of A_{rs} that contains row i of A , and the numbers b_{kj} with $k \in K_s$ form the column of B_{st} that contains column j of B . So the inner sum is the corresponding entry of $A_{rs}B_{st}$, and $(AB)_{ij}$ is the corresponding entry of $\sum_s A_{rs}B_{st}$. Since matrices are added entry by entry, this is the statement of the theorem. $\square$

Example 3.4.14. For the matrix M above, block multiplication gives

$$M^2 = \begin{bmatrix} I & X \\ O & I \end{bmatrix} \begin{bmatrix} I & X \\ O & I \end{bmatrix} = \begin{bmatrix} II + XO & IX + XI \\ OI + IO & OX + II \end{bmatrix} = \begin{bmatrix} I & 2X \\ O & I \end{bmatrix} = \left[\begin{array}{cc|cc} 1 & 0 & 2 & 4 \\ 0 & 1 & 6 & 8 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right],$$

using $IY = YI = Y$ and $OY = YO = O$ for 2×2 matrices Y . Entry by entry, the product of two 4×4 matrices needs 64 products of entries; with blocks we needed none. By induction, $M^k = \begin{bmatrix} I & kX \\ O & I \end{bmatrix}$ for every $k \geq 1$.

Two special cuts are rules we already know. Cutting B into its columns $\mathbf{b}_1, \dots, \mathbf{b}_p$ (and not cutting A at all) gives the column form of the product (Definition 3.3.1), $AB = A [\mathbf{b}_1 \cdots \mathbf{b}_p] = [A\mathbf{b}_1 \cdots A\mathbf{b}_p]$. Cutting A into its columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ and a column vector $\mathbf{x}$ into its entries gives $A\mathbf{x} = x_1\mathbf{a}_1 + \dots + x_n\mathbf{a}_n$, the definition of the matrix-vector product (Definition 3.1.1).

3.4.5 The trace

Definition 3.4.15 (Trace). The *trace* of an $n \times n$ matrix A is the sum of its diagonal entries:

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

For example, $\operatorname{tr} I_n = n$ (more precisely, $1 + 1 + \cdots + 1$ with n terms, which is n in $\mathbb{R}$), and for the symmetric matrix S of Example 3.4.7, $\operatorname{tr} S = 2 + 3 + 5 = 10$.

Theorem 3.4.16 (Properties of the trace). Let A and B be $n \times n$ matrices and $c \in \mathbb{F}$.

- (a) $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$ and $\operatorname{tr}(cA) = c \operatorname{tr} A$;
- (b) $\operatorname{tr}(A^T) = \operatorname{tr} A$;
- (c) more generally, if A is $m \times n$ and B is $n \times m$, then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof. (a) $\operatorname{tr}(A + B) = \sum_i (a_{ii} + b_{ii}) = \sum_i a_{ii} + \sum_i b_{ii}$, using the commutative and associative laws of addition in $\mathbb{F}$; and $\operatorname{tr}(cA) = \sum_i ca_{ii} = c \sum_i a_{ii}$ by the distributive law.

(b) $(A^T)_{ii} = a_{ii}$, so the two diagonals are the same.

(c) AB is $m \times m$ and BA is $n \times n$, so both traces are defined. By the definition of the product,

$$\operatorname{tr}(AB) = \sum_{i=1}^m (AB)_{ii} = \sum_{i=1}^m \sum_{k=1}^n a_{ik} b_{ki}, \quad \operatorname{tr}(BA) = \sum_{k=1}^n (BA)_{kk} = \sum_{k=1}^n \sum_{i=1}^m b_{ki} a_{ik}.$$

Both double sums add the mn products $a_{ik} b_{ki} = b_{ki} a_{ik}$, one for each pair (i, k) , only in a different order. Since addition in $\mathbb{F}$ is commutative and associative, they are equal. $\square$

Example 3.4.17. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. By the column rule, the first column of AB is $A(0, 1) = (2, 4)$, the second column of A , and the second column of AB is $A(1, 0) = (1, 3)$, the first column of A . Similarly, row i of BA is row i of B times A , so the first row of BA is $(0, 1)A = (3, 4)$ and the second is $(1, 0)A = (1, 2)$:

$$AB = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}, \quad BA = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}.$$

So $AB \neq BA$, but $\operatorname{tr}(AB) = 2 + 3 = 5 = 3 + 2 = \operatorname{tr}(BA)$.

Three or more factors. Applying Theorem 3.4.16(c) to the two factors A and BC gives $\operatorname{tr}(ABC) = \operatorname{tr}(BCA)$, and again $= \operatorname{tr}(CAB)$: the trace is unchanged by *cyclic* moves. Other rearrangements can change it. With

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

we get $ABC = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, of trace 1, but $AC = O$, so $ACB = O$, of trace 0.

Proposition 3.4.18. There are no $n \times n$ real matrices A and B with $AB - BA = I_n$.

Proof. By Theorem 3.4.16(a) and (c), $\operatorname{tr}(AB - BA) = \operatorname{tr}(AB) - \operatorname{tr}(BA) = 0$, while $\operatorname{tr} I_n = n \neq 0$. $\square$

The proof works over every field in which $n = 1 + \cdots + 1$ (n terms) is not zero. Over $\mathbb{Z}_2$ and $n = 2$ it fails, and so does the statement: $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$ in $\mathbb{Z}_2$, since $-1 = 1$ there.

Preview. In Chapter 8 (Lecture 38) we will see that, over $\mathbb{C}$, the trace of an $n \times n$ matrix is the sum of its eigenvalues, counted with multiplicity. We do not prove this here; it is stated only as a look ahead.

Check yourself

Let $A = \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 0 & 3 \end{bmatrix}$. (a) Write down A^T and compute $A^T A$. Is it symmetric? (b) Compute $\text{tr}(A^T A)$ and $\text{tr}(AA^T)$ without computing AA^T . (c) True or false: if A and B are $n \times n$ and symmetric, then AB is symmetric.

Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

Strictly triangular matrices are nilpotent. An upper triangular matrix with zeros on the diagonal is called *strictly* upper triangular. If N is $n \times n$ and strictly upper triangular, then $N^n = O$. Indeed, $(N)_{ij} = 0$ unless $j \geq i + 1$, and the argument in the proof of Theorem 3.4.11 shows by induction that $(N^k)_{ij} = 0$ unless $j \geq i + k$; for $k = n$ no pair of indices between 1 and n satisfies $j \geq i + n$. For example,

$$\begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and the cube is } O.$$

The trace is the only such function. Up to a constant factor, the trace is the only function f from $n \times n$ matrices to $\mathbb{F}$ with $f(A + B) = f(A) + f(B)$, $f(cA) = cf(A)$ and $f(AB) = f(BA)$. To see the idea, let E_{ij} be the matrix with a single 1 in position (i, j) . Then $E_{ij}E_{ji} = E_{ii}$ and $E_{ji}E_{ij} = E_{jj}$, so $f(E_{ii}) = f(E_{jj})$; and for $i \neq j$, $E_{ij} = E_{ii}E_{ij}$ while $E_{ij}E_{ii} = O$, so $f(E_{ij}) = f(O) = 0$. Writing $A = \sum a_{ij}E_{ij}$ gives $f(A) = f(E_{11}) \text{tr } A$.

The conjugate transpose. Over $\mathbb{C}$ the transpose is usually combined with complex conjugation: the *conjugate transpose* of A is $A^* = \overline{A}^T$, with $(A^*)_{ij} = \overline{a_{ji}}$. It obeys the same rules as the transpose, except that $(cA)^* = \bar{c}A^*$, and matrices with $A^* = A$ (called *Hermitian*) play the role of symmetric matrices in Chapters 10 and 11.

3.5 Inverse matrices

Lecture 16

For numbers, undoing a multiplication is easy. If $3x = 6$, we multiply both sides by $\frac{1}{3}$, the inverse of 3, and get $x = 2$. For matrices we now have the same kind of equation, $Ax = b$, and Section 3.2 showed that a matrix moves the plane. So we ask: can we undo what A does? A shear slides every point sideways, by an amount that grows with its height, and clearly we can slide everything back. A projection flattens the whole plane onto a line, and nothing can unflatten it.

This section makes the question precise. We define the inverse of a matrix, find a formula for 2×2 matrices, and prove the rules that inverses obey. Then we learn *Gauss–Jordan inversion*, a method that computes the inverse of any invertible matrix, or tells us that there is none.

Throughout, $\mathbb{F}$ is a field (Section 2.1) and all matrices have entries in $\mathbb{F}$. We use these facts about the matrix product from Section 3.3: column j of AB is A times column j of B ; $(AB)\mathbf{x} = A(B\mathbf{x})$; the product is associative, $(AB)C = A(BC)$; $I_m A = A = A I_n$ for an $m \times n$ matrix A ; $c(AB) = (cA)B = A(cB)$ for scalars c ; and a product with a zero matrix O is a zero matrix. From Section 3.4 we use $(AB)^T = B^T A^T$ and $I^T = I$.

3.5.1 Invertible matrices

Definition 3.5.1 (Invertible matrix, inverse). Let A be an $n \times n$ matrix over $\mathbb{F}$. We say that A is *invertible* if there is an $n \times n$ matrix B such that

$$AB = I_n \quad \text{and} \quad BA = I_n.$$

Such a matrix B is called an *inverse* of A .

Only square matrices are called invertible here. A matrix that is not square can still have a one-sided inverse (see the supplement at the end of this section), but never a two-sided one (this is proved with the rank in Corollary 5.2.14). Both products in the definition are the matrix product of Section 3.3, and both must equal I_n (Definition 3.1.5).

Example 3.5.2. Let

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

A is a shear: it slides every point to the right by its height. To undo it we slide back the other way, which is what B does. We compute AB column by column. Column 1 of AB is A times the first column $(1, 0)$ of B , that is, one copy of the first column of A :

$$A \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Column 2 of AB is

$$A \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1+1 \\ 0+1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

So $AB = I_2$. In the other order, column 1 of BA is $B(1, 0) = (1, 0)$, the first column of B , and column 2 is

$$B \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

So $BA = I_2$ too, and the shear A is invertible with inverse B . In pictures: A slides the plane to the right by the height of each point, and B slides it back (Figure 3.11). In either order, every point ends where it started.

Example 3.5.3. The projection

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

flattens the plane onto the x -axis. It is not invertible. Indeed, $P(0, 1) = 0(1, 0) + 1(0, 0) = \mathbf{0}$. Suppose some 2×2 matrix B had $BP = I$. Then

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = I \begin{bmatrix} 0 \\ 1 \end{bmatrix} = (BP) \begin{bmatrix} 0 \\ 1 \end{bmatrix} = B \left(P \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) = B\mathbf{0} = \mathbf{0},$$

which is false. So no B satisfies $BP = I$, and in particular P has no inverse.

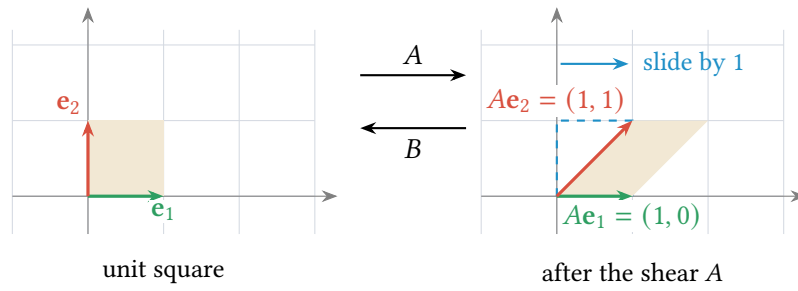


Figure 3.11: Example 3.5.2: the shear A turns the unit square (left) into a parallelogram (right), sliding the top edge one unit to the right; e_1 (green) stays, e_2 (coral) goes to $(1, 1)$. The matrix B slides it back.

The same argument works in general: a square matrix A with $Ax = 0$ for some $x \neq 0$ is not invertible. For if $BA = I$, then $x = Ix = (BA)x = B(Ax) = B0 = 0$, a contradiction. A matrix that sends a nonzero vector to zero can never be undone.

3.5.2 One inverse, and solving $Ax = b$

Proposition 3.5.4 (Uniqueness of the inverse). An $n \times n$ matrix has at most one inverse: if B and C are both inverses of A , then $B = C$.

Proof. We have $AC = I$ (since C is an inverse) and $BA = I$ (since B is an inverse). Hence

$$B = BI = B(AC) = (BA)C = IC = C,$$

using $BI = B$, then $I = AC$, then associativity, then $BA = I$, and finally $IC = C$. $\square$

So an invertible matrix has exactly one inverse, and we give it a name.

Notation. The inverse of an invertible matrix A is written A^{-1} (read “ A inverse”). It is defined by the two equations

$$AA^{-1} = I \quad \text{and} \quad A^{-1}A = I.$$

A warning: A^{-1} is *not* $1/A$. There is no division of matrices. The inverse is a matrix in its own right, and we always multiply by it. In Example 3.5.2, $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$.

Here is the reason inverses matter.

Theorem 3.5.5 (Solving with the inverse). Let A be an invertible $n \times n$ matrix. Then for every $b \in \mathbb{R}^n$ the system $Ax = b$ has exactly one solution, namely

$$x = A^{-1}b.$$

Proof. It is a solution. $A(A^{-1}b) = (AA^{-1})b = Ib = b$.

It is the only one. Suppose $Ax = b$. Multiply both sides on the left by A^{-1} : $A^{-1}(Ax) = A^{-1}b$. The left side is $(A^{-1}A)x = Ix = x$. So $x = A^{-1}b$. $\square$

This is exactly how we solved $3x = 6$: we multiplied by the inverse. With matrices, though, the side matters. We multiply on the *left*, because Ax has A on the left; the product xA^{-1} is not even defined when $n \geq 2$.

3.5.3 Two by two inverses

For 2×2 matrices there is a formula, and a single number decides everything.

Theorem 3.5.6 (Inverse of a 2×2 matrix). Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ over $\mathbb{F}$.

(a) If $ad - bc \neq 0$, then A is invertible and

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

(b) If $ad - bc = 0$, then A is not invertible.

In words: swap the two diagonal entries, change the signs of the other two, and divide by $ad - bc$.

Proof. Let $N = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. We compute AN one column at a time. The first column is

$$A \begin{bmatrix} d \\ -c \end{bmatrix} = d \begin{bmatrix} a \\ c \end{bmatrix} - c \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} ad - bc \\ cd - cd \end{bmatrix} = \begin{bmatrix} ad - bc \\ 0 \end{bmatrix},$$

and the second is

$$A \begin{bmatrix} -b \\ a \end{bmatrix} = -b \begin{bmatrix} a \\ c \end{bmatrix} + a \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} -ab + ab \\ -bc + ad \end{bmatrix} = \begin{bmatrix} 0 \\ ad - bc \end{bmatrix}.$$

So $AN = (ad - bc)I$. In the other order, the columns of NA are

$$N \begin{bmatrix} a \\ c \end{bmatrix} = a \begin{bmatrix} d \\ -c \end{bmatrix} + c \begin{bmatrix} -b \\ a \end{bmatrix} = \begin{bmatrix} ad - bc \\ -ac + ac \end{bmatrix}, \quad N \begin{bmatrix} b \\ d \end{bmatrix} = b \begin{bmatrix} d \\ -c \end{bmatrix} + d \begin{bmatrix} -b \\ a \end{bmatrix} = \begin{bmatrix} bd - bd \\ -bc + ad \end{bmatrix},$$

so $NA = (ad - bc)I$ as well. (We used commutativity of multiplication in $\mathbb{F}$, for instance $cb = bc$.)

(a) Let $\delta = ad - bc \neq 0$ and $B = \delta^{-1}N$. Then $AB = \delta^{-1}(AN) = \delta^{-1}\delta I = I$, and in the same way $BA = \delta^{-1}(NA) = I$. So A is invertible with $A^{-1} = \delta^{-1}N$.

(b) Suppose $ad - bc = 0$. Then $AN = 0I = O$, the zero matrix. If A had an inverse, then

$$N = IN = (A^{-1}A)N = A^{-1}(AN) = A^{-1}O = O.$$

The entries of N are $d, -b, -c, a$, so $a = b = c = d = 0$ and $A = O$. But $OB = O \neq I$ for every B , so the zero matrix has no inverse. This contradiction shows that A is not invertible. $\square$

The number $ad - bc$ is called the *determinant* of A . We meet it properly in Chapter 7, where it measures how A scales areas.

Example 3.5.7. Take the matrix of the system $2x - y = 1$, $x + y = 5$ from Example 1.1.5,

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}.$$

Here $ad - bc = 2 \cdot 1 - (-1) \cdot 1 = 2 + 1 = 3 \neq 0$. Swapping the diagonal entries 2 and 1 and changing the signs of -1 and 1 gives, by Theorem 3.5.6,

$$A^{-1} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}.$$

By Theorem 3.5.5 the system $A\mathbf{x} = (1, 5)$ is solved in one step. By the row rule (Proposition 3.1.4),

$$\mathbf{x} = A^{-1} \begin{bmatrix} 1 \\ 5 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \cdot 1 + 1 \cdot 5 \\ (-1) \cdot 1 + 2 \cdot 5 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

the solution found in Example 1.1.5. For any other right-hand side, the same A^{-1} solves the system at once.

The second matrix of Example 3.1.10, $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, has $ad - bc = 1 \cdot 4 - 2 \cdot 2 = 4 - 4 = 0$. Its columns lie on one line, and by Theorem 3.5.6(b) it has no inverse.

3.5.4 Rules for inverses

Theorem 3.5.8 (Rules for inverses). Let A and B be invertible $n \times n$ matrices.

- (a) A^{-1} is invertible, and $(A^{-1})^{-1} = A$.
- (b) AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.
- (c) A^T is invertible, and $(A^T)^{-1} = (A^{-1})^T$.

Proof. In each part we exhibit a matrix and check both products of Definition 3.5.1; by Proposition 3.5.4 that matrix is then *the* inverse.

(a) The equations $A^{-1}A = I$ and $AA^{-1} = I$, which define A^{-1} , say exactly that A is an inverse of A^{-1} , read from the other side.

(b) By associativity the middle pair comes together:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I,$$

and in the other order

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I.$$

(c) By the rule $(XY)^T = Y^TX^T$ of Section 3.4 and $I^T = I$,

$$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I \quad \text{and} \quad (A^{-1})^TA^T = (AA^{-1})^T = I^T = I. \quad \square$$

Notice the order in (b): it is reversed. Think of putting on socks and then shoes. To undo it, you take off the shoes first and then the socks. The last step done is the first step undone.

Some consequences and warnings.

- *Scalar multiples.* If $c \in \mathbb{F}$ and $c \neq 0$, then cA is invertible with $(cA)^{-1} = c^{-1}A^{-1}$, since $(cA)(c^{-1}A^{-1}) = (cc^{-1})(AA^{-1}) = 1 \cdot I = I$ and in the same way $(c^{-1}A^{-1})(cA) = (c^{-1}c)(A^{-1}A) = I$.

- *Longer products.* If $A_1, \dots, A_k$ are invertible $n \times n$ matrices, then so is $A_1 A_2 \cdots A_k$, and

$$(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} \cdots A_2^{-1} A_1^{-1}.$$

This follows by induction on k : the case $k = 1$ is clear, and if it holds for $k - 1$ factors, then (b) applied to $A_1 \cdots A_{k-1}$ and A_k gives $(A_1 \cdots A_k)^{-1} = A_k^{-1} (A_1 \cdots A_{k-1})^{-1} = A_k^{-1} A_{k-1}^{-1} \cdots A_1^{-1}$.

- *The order matters.* In general $(AB)^{-1} \neq A^{-1} B^{-1}$. For $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ we have $A^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, $B^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ (both by Theorem 3.5.6, with $ad - bc = 1$), and

$$AB = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad (AB)^{-1} = B^{-1} A^{-1} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}, \quad \text{but} \quad A^{-1} B^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}.$$

- *Sums.* A sum of invertible matrices need not be invertible: I and $-I$ are invertible (each is its own inverse), but $I + (-I) = O$ is not.

3.5.5 One product is enough

So far, to show that B is the inverse of A , we have checked two products, AB and BA . Matrix multiplication is not commutative, so it seems that we need both. For square matrices, one is enough. The proof is demanding, so read it slowly. It uses the two pivot theorems of Section 3.1.

Proposition 3.5.9 (One product is enough). Let A and X be $n \times n$ matrices with $AX = I$. Then $XA = I$ as well, so A is invertible and $X = A^{-1}$.

Proof. *Step 1:* $Xz = \mathbf{b}$ has a solution for every $\mathbf{b}$. First, $X\mathbf{y} = \mathbf{0}$ has only the trivial solution: if $X\mathbf{y} = \mathbf{0}$, then

$$\mathbf{y} = I\mathbf{y} = (AX)\mathbf{y} = A(X\mathbf{y}) = A\mathbf{0} = \mathbf{0}.$$

By Proposition 3.1.13, an echelon form of X has a pivot in every column. That makes n pivots. Each pivot is the leading entry of its row, so each row holds at most one pivot, and the n pivots lie in n different rows. X has only n rows, so every row has a pivot. By Theorem 3.1.9, $Xz = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{F}^n$.

Step 2: $XA = I$. Let $\mathbf{b} \in \mathbb{F}^n$, and choose $\mathbf{z}$ with $Xz = \mathbf{b}$ (Step 1). Then

$$(XA)\mathbf{b} = (XA)(Xz) = X(A(Xz)) = X((AX)\mathbf{z}) = X(I\mathbf{z}) = X\mathbf{z} = \mathbf{b}.$$

So XA sends every vector to itself, just as I does. Take $\mathbf{b} = \mathbf{e}_j$, the j -th column of I . Column j of XA is $(XA)\mathbf{e}_j = \mathbf{e}_j$, the j -th column of I (Section 3.1). This holds for every j , so $XA = I$. Together with $AX = I$ this says that X is an inverse of A . $\square$

Likewise, if $XA = I$, then $AX = I$: apply Proposition 3.5.9 to the pair (X, A) in place of (A, X) . Its hypothesis is now $XA = I$, and its conclusion is $AX = I$. So for square matrices, to check that B is the inverse of A , it is enough to check *one* of $AB = I$ and $BA = I$.

This may look like a technicality, but it is what makes the method of the next subsection work.

Only for square matrices. Proposition 3.5.9 is special to square matrices. For

$$A = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad X = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

we get $AX = \begin{bmatrix} 1 \end{bmatrix} = I_1$, but $XA = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \neq I_2$. The proof breaks down in Step 1: for the 2×1 matrix X , a pivot in every column does not give a pivot in every row.

3.5.6 Gauss–Jordan inversion

How do we find the inverse of a bigger matrix? Write the unknown inverse X by its columns $\mathbf{x}_1, \dots, \mathbf{x}_n$, and let $\mathbf{e}_1, \dots, \mathbf{e}_n$ be the columns of I_n , the standard basis vectors of Section 3.2: $\mathbf{e}_j$ has a 1 in position j and zeros elsewhere. Column j of AX is $A\mathbf{x}_j$, so the equation $AX = I$ says, column by column,

$$A\mathbf{x}_1 = \mathbf{e}_1, \quad A\mathbf{x}_2 = \mathbf{e}_2, \quad \dots, \quad A\mathbf{x}_n = \mathbf{e}_n.$$

These are n systems, all with the same matrix A . Row reduction treats each column on its own, so we can solve them all at once: we write A with the whole identity beside it, $[A \mid I_n]$, and row reduce.

Theorem 3.5.10 (Gauss–Jordan inversion). Let A be an $n \times n$ matrix over $\mathbb{F}$.

- (a) If $[A \mid I_n]$ row reduces to $[I_n \mid X]$, then A is invertible and $X = A^{-1}$.
- (b) If the reduced row echelon form of A is not I_n , then A is not invertible.
- (c) Hence A is invertible if and only if its reduced row echelon form is I_n , and in that case $[A \mid I_n]$ row reduces to $[I_n \mid A^{-1}]$.

Proof. (a) Suppose a sequence of row operations takes $[A \mid I_n]$ to $[I_n \mid X]$, and let $\mathbf{x}_j$ be column j of X . A row operation changes each column using only the entries of that column (it adds a multiple of one entry to another in the same column, swaps two entries of the column, or scales one of them), so it acts on each column independently. Keeping only the first n columns and column $n + j$, the same operations therefore take $[A \mid \mathbf{e}_j]$ to $[I_n \mid \mathbf{x}_j]$. By Theorem 2.3.4 the system $A\mathbf{x} = \mathbf{e}_j$ has the same solutions as $I_n\mathbf{x} = \mathbf{x}_j$. Since $I_n\mathbf{x} = \mathbf{x}$, the only solution of the latter is $\mathbf{x} = \mathbf{x}_j$. Hence $A\mathbf{x}_j = \mathbf{e}_j$ for every j , that is, every column of AX equals the matching column of I_n , and $AX = I_n$. By Proposition 3.5.9, $XA = I_n$ too, so $X = A^{-1}$.

(b) First we show that a square matrix R in reduced row echelon form with a pivot in every column is I_n . Such an R has n pivots, at most one in each row, so (as in the proof of Proposition 3.5.9) every one of the n rows has a pivot. Let row i have its pivot in column j_i . Leading entries move strictly to the right as we go down (Definition 2.4.1(ii)), so $1 \leq j_1 < j_2 < \dots < j_n \leq n$. These are n different integers between 1 and n in increasing order, so $j_i = i$ for every i . By conditions (iii) and (iv) of Definition 2.4.1, each pivot is 1 and is the only nonzero entry of its column. So column i of R is $\mathbf{e}_i$ for every i , and $R = I_n$.

Now let R be the reduced row echelon form of A and suppose $R \neq I_n$. By what we just showed, some column of R has no pivot, so the system $R\mathbf{x} = \mathbf{0}$ has a free variable (Definition 2.4.6). By Proposition 2.5.8 it has a nontrivial solution $\mathbf{x} \neq \mathbf{0}$, and by Theorem 2.3.4 (applied to $[A \mid \mathbf{0}]$, which row reduces to $[R \mid \mathbf{0}]$) this $\mathbf{x}$ also solves $A\mathbf{x} = \mathbf{0}$. But if A were invertible, Theorem 3.5.5 with $\mathbf{b} = \mathbf{0}$ would say that $A\mathbf{x} = \mathbf{0}$ has only the solution $\mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$. So A is not invertible.

(c) By Theorems 2.4.4 and 2.4.9, A has exactly one reduced row echelon form R . If A is invertible, then $R = I_n$ by (b). Conversely, suppose $R = I_n$, and apply the row operations that take A to I_n to the whole matrix $[A \mid I_n]$. Since they act on each column independently, the left block becomes I_n , and the

right block becomes some matrix X . By (a), A is invertible and $X = A^{-1}$; so $[A \mid I_n]$ row reduces to $[I_n \mid A^{-1}]$. $\square$

In practice: row reduce $[A \mid I]$ by Gauss–Jordan elimination (Section 2.4). If the left block becomes I , the right block is A^{-1} . If the left block produces a row of zeros, or otherwise cannot become I , then A is not invertible. Section 3.6 collects many more conditions that are equivalent to invertibility in the invertible matrix theorem.

Example 3.5.11. Let

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}.$$

We write A with I_3 beside it. *Forward phase.* The first column needs a zero in row 3. The replacement $R_3 \rightarrow R_3 - R_1$ gives $1 - 1 = 0$, $2 - 2 = 0$, $1 - 0 = 1$ and, after the bar, $0 - 1 = -1$, $0 - 0 = 0$, $1 - 0 = 1$:

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

The left block is now in echelon form, with pivots 1, 1, 1. *Backward phase.* The last pivot, in row 3, has a 1 above it in row 2. The replacement $R_2 \rightarrow R_2 - R_3$ gives $0 - 0 = 0$, $1 - 0 = 1$, $1 - 1 = 0$ and, after the bar, $0 - (-1) = 1$, $1 - 0 = 1$, $0 - 1 = -1$:

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

Row 1 already has a 0 in column 3. The pivot in row 2 has a 2 above it, so we use $R_1 \rightarrow R_1 - 2R_2$: $1 - 2 \cdot 0 = 1$, $2 - 2 \cdot 1 = 0$, $0 - 2 \cdot 0 = 0$ and, after the bar, $1 - 2 \cdot 1 = -1$, $0 - 2 \cdot 1 = -2$, $0 - 2 \cdot (-1) = 2$:

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

The left block is I_3 , so by Theorem 3.5.10(a) A is invertible and

$$A^{-1} = \begin{bmatrix} -1 & -2 & 2 \\ 1 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}.$$

Check. By the row rule, A times the first column $(-1, 1, -1)$ of A^{-1} is

$$\begin{bmatrix} 1 \cdot (-1) + 2 \cdot 1 + 0 \cdot (-1) \\ 0 \cdot (-1) + 1 \cdot 1 + 1 \cdot (-1) \\ 1 \cdot (-1) + 2 \cdot 1 + 1 \cdot (-1) \end{bmatrix} = \begin{bmatrix} -1 + 2 \\ 1 - 1 \\ -1 + 2 - 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{e}_1,$$

as it should be. The other two columns give $A(-2, 1, 0) = (-2 + 2, 1, -2 + 2) = (0, 1, 0)$ and $A(2, -1, 1) = (2 - 2, -1 + 1, 2 - 2 + 1) = (0, 0, 1)$, so $AA^{-1} = I_3$; by Proposition 3.5.9 the other product $A^{-1}A = I_3$ needs no separate check.

Example 3.5.12. When the method fails, it fails visibly. Take $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ and write I_2 beside it. The replacement $R_2 \rightarrow R_2 - 2R_1$ gives $2 - 2 \cdot 1 = 0$, $4 - 2 \cdot 2 = 0$ and, after the bar, $0 - 2 \cdot 1 = -2$, $1 - 2 \cdot 0 = 1$:

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array} \right] \xrightarrow{R_2 \rightarrow R_2 - 2R_1} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 0 & -2 & 1 \end{array} \right].$$

The left block has a row of zeros, and no further row operation can create a pivot in it: the reduced row echelon form of A is $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \neq I_2$. By Theorem 3.5.10(b), A is not invertible, as Theorem 3.5.6(b) also said. Indeed, the second column has no pivot, so there is a free variable, and by the column rule

$$A \begin{bmatrix} -2 \\ 1 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

So A flattens the plane onto the line spanned by $(1, 2)$, like the projection of Example 3.5.3.

Remark 3.5.13. To solve a *single* system $Ax = b$, it is less work to row reduce $[A \mid b]$ than to compute A^{-1} first and then multiply. For a large $n \times n$ matrix, elimination on $[A \mid b]$ takes roughly $n^3/3$ multiplications, while computing A^{-1} takes roughly n^3 . The inverse pays off when we need many right-hand sides, or when we reason about A itself, as in Theorem 3.5.5.

Check yourself

- (a) Use Theorem 3.5.6 to find the inverse of $\begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix}$, and check one of the two products.
- (b) Use Gauss–Jordan inversion to find the inverse of $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.
- (c) Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix}$. Is A invertible? Decide with Theorem 3.5.10, and find a nonzero x with $Ax = 0$ if there is one.

Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

Over a finite field. Everything in this section uses only the field operations, so it holds over $\mathbb{Z}_p$. In Theorem 3.5.6, dividing by $ad - bc$ means multiplying by its inverse in the field. Over $\mathbb{Z}_5$, let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Then $ad - bc = 4 - 6 = -2 = 3$, and $3^{-1} = 2$ in $\mathbb{Z}_5$ because $3 \cdot 2 = 6 = 1$. So

$$A^{-1} = 2 \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = 2 \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 6 \\ 4 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 4 & 2 \end{bmatrix},$$

and indeed, computing in $\mathbb{Z}_5$, $AA^{-1} = \begin{bmatrix} 3+8 & 1+4 \\ 9+16 & 3+8 \end{bmatrix} = \begin{bmatrix} 11 & 5 \\ 25 & 11 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Whether a matrix is invertible can depend on the field: $\begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$ has $ad - bc = 1 - 6 = -5$, so it is invertible over $\mathbb{R}$ but not over $\mathbb{Z}_5$, where $-5 = 0$.

One-sided inverses. An $m \times n$ matrix A with $m \neq n$ can have a *right inverse* X ($AX = I_m$) or a *left inverse* Y ($YA = I_n$), but not both. The 1×2 matrix $A = \begin{bmatrix} 1 & 0 \end{bmatrix}$ of the paragraph after Proposition 3.5.9 has the right inverse $(1, 0)$, and also $(1, t)$ for every t , so a one-sided inverse need not be unique; it has no left inverse, since YA is a 2×2 matrix whose second column is Y times the second column (0) of A , which is 0 . The general statement is proved with the rank of a matrix in Corollary 5.2.14: a right inverse exists exactly when there is a pivot in every row, a left inverse exactly when there is a pivot in every column, and both happen only when $m = n$.

Negative powers. For an invertible $n \times n$ matrix A and $k \geq 1$, the power A^k is invertible with

$$(A^k)^{-1} = (A^{-1})^k,$$

by the rule for longer products with all k factors equal to A . We define $A^{-k} = (A^{-1})^k$ and $A^0 = I$. With these definitions the familiar laws

$$A^j A^k = A^{j+k} \quad \text{and} \quad (A^j)^k = A^{jk}$$

hold for all integers j and k .

Proof. Write $B = A^{-1}$, so $A^{-m} = B^m$ for every $m \geq 0$, and $A^m B^m = B^m A^m = I$ because $B^m = (A^m)^{-1}$.

The product law. If $j, k \geq 0$, this is the law for non-negative powers after Definition 3.3.10. If $j, k \leq 0$, then $A^j A^k = B^{-j} B^{-k} = B^{-j-k} = A^{j+k}$ by the same law applied to B . Suppose now $j \geq 0$ and $k = -m$ with $m > 0$. If $j \geq m$, then $A^j B^m = A^{j-m} (A^m B^m) = A^{j-m}$. If $j < m$, then $A^j B^m = (A^j B^j) B^{m-j} = B^{m-j} = A^{j-m}$. In both cases $A^j A^k = A^{j+k}$. The case $j < 0 \leq k$ is the same argument with $B^m A^m = I$ in place of $A^m B^m = I$.

The power law. First let $k \geq 0$ and j be any integer. We induct on k . For $k = 0$ both sides are I , and $(A^j)^{k+1} = (A^j)^k A^j = A^{jk} A^j = A^{j(k+1)}$ by the product law. Next, the product law gives $A^j A^{-j} = A^{-j} A^j = A^0 = I$, so A^j is invertible with $(A^j)^{-1} = A^{-j}$. Hence for $k = -m$ with $m > 0$, $(A^j)^k = ((A^j)^{-1})^m = (A^{-j})^m = A^{-jm} = A^{jk}$, using the case already proved. $\square$

For the shear of Example 3.5.2, $A^{-k} = \begin{bmatrix} 1 & -k \\ 0 & 1 \end{bmatrix}$: sliding back k times.

Diagonal matrices. A diagonal matrix $D = \text{diag}(d_1, \dots, d_n)$ with every $d_i \neq 0$ is invertible, with $D^{-1} = \text{diag}(d_1^{-1}, \dots, d_n^{-1})$. Indeed, column j of $D \text{diag}(d_1^{-1}, \dots, d_n^{-1})$ is $D(d_j^{-1} \mathbf{e}_j) = d_j^{-1} (d_j \mathbf{e}_j) = \mathbf{e}_j$, so the product is I , and Proposition 3.5.9 gives the other order. If some $d_j = 0$, then $D \mathbf{e}_j = \mathbf{0}$, and D is not invertible by the remark after Example 3.5.3.

3.6 The invertible matrix theorem

Lecture 17

In Section 3.5 we learned Gauss–Jordan inversion and proved that a square matrix is invertible exactly when its reduced row echelon form is the identity. Along the way we met other properties of a square matrix A : the equation $A\mathbf{x} = \mathbf{0}$ may have only the trivial solution; the columns of A may be linearly independent, or they may span the whole space; and $A\mathbf{x} = \mathbf{b}$ may have a solution for every $\mathbf{b}$. These look like different questions. In this section we prove that, for a square matrix, they all have the same answer. This is the *invertible matrix theorem*, one of the central results of the course.

First we need one more tool: every row operation is a multiplication by a matrix, called an *elementary matrix*. This turns row reduction into matrix algebra, and it explains why Gauss–Jordan inversion works.

Throughout, $\mathbb{F}$ is a field and all matrices have entries in $\mathbb{F}$. Row r of the identity I_m has a 1 in position r and zeros elsewhere.

3.6.1 Elementary matrices

Definition 3.6.1 (Elementary matrix). An $m \times m$ matrix E is an *elementary matrix* if it is obtained from the identity I_m by a single elementary row operation (Definition 2.3.1). We call E the elementary matrix of that operation.

There are three kinds, one for each kind of row operation. The identity itself is elementary: it is the elementary matrix of the scaling $R_1 \rightarrow 1 \cdot R_1$.

Example 3.6.2. Over $\mathbb{R}$, from I_3 we get, by the swap $R_1 \leftrightarrow R_2$, the scaling $R_3 \rightarrow 5R_3$ and the replacement $R_3 \rightarrow R_3 - 2R_1$,

$$E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}.$$

(E_1 and E_3 are elementary over every field. Over $\mathbb{Z}_5$, however, $5 = 0$, the scaling by 5 is not allowed, and E_2 is not elementary.)

In the plane the 2×2 elementary matrices are simple moves (Figure 3.12). Swapping the two rows of I_2 gives $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which exchanges the two coordinates: it reflects the plane across the line $y = x$.

Scaling the first row by 2 gives $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, which stretches the plane along the first axis. The replacement $R_2 \rightarrow R_2 + R_1$ gives $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, a shear: it sends (x, y) to $(x, x + y)$.

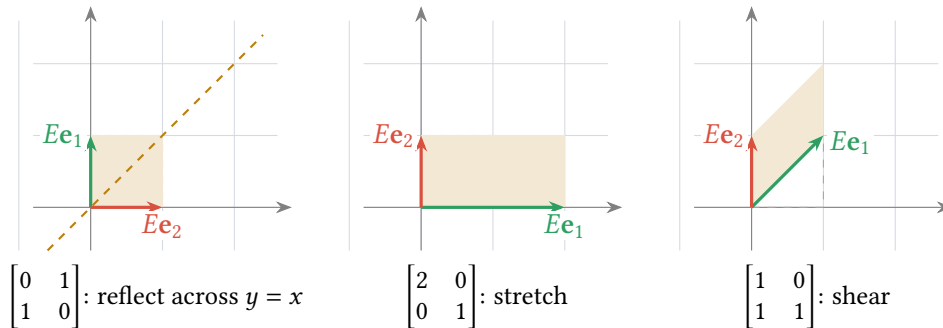


Figure 3.12: Example 3.6.2: the 2×2 elementary matrices as moves of the plane, acting on the unit square. Left: the swap reflects across $y = x$. Middle: scaling the first row by 2 stretches along the first axis. Right: the replacement $R_2 \rightarrow R_2 + R_1$ is a shear. Where e_1 lands is green, where e_2 lands is coral.

The point of elementary matrices is the next theorem.

Theorem 3.6.3 (Row operations are multiplications). Let E be the $m \times m$ elementary matrix of a row operation. Then for every $m \times n$ matrix A over $\mathbb{F}$, the product EA is the matrix obtained by performing that row operation on A .

In words: to perform a row operation, multiply on the left by its elementary matrix.

Proof. By the row-column rule (Theorem 3.3.5), entry (r, k) of EA is row r of E times column k of A :

$$(EA)_{rk} = e_{r1}a_{1k} + e_{r2}a_{2k} + \cdots + e_{rm}a_{mk}.$$

Rows that the operation does not change. If row r of E is row r of I_m , then $e_{rr} = 1$ and every other $e_{rs} = 0$, so $(EA)_{rk} = a_{rk}$ for every k : row r of EA is row r of A .

Swap $R_i \leftrightarrow R_j$. Row i of E is row j of I_m , so $(EA)_{ik} = a_{jk}$, and row i of EA is row j of A . In the same way row j of EA is row i of A .

Scaling $R_i \rightarrow cR_i$. Row i of E has c in position i and zeros elsewhere, so $(EA)_{ik} = c a_{ik}$, and row i of EA is c times row i of A .

Replacement $R_i \rightarrow R_i + cR_j$, $i \neq j$. Row i of E has 1 in position i , c in position j and zeros elsewhere, so $(EA)_{ik} = a_{ik} + c a_{jk}$, and row i of EA is row i of A plus c times row j of A .

In each case every other row of E is the matching row of I_m , so the other rows of A are unchanged. Hence EA is A after the row operation. $\square$

Example 3.6.4. Let E_1, E_2, E_3 be as in Example 3.6.2 and

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 1 & 4 \\ 2 & 5 & 1 \end{bmatrix}.$$

Rows 1 and 2 of E_3 are rows of I_3 , so rows 1 and 2 of E_3A are those of A . Row 3 of E_3 is $(-2, 0, 1)$; times the three columns of A it gives

$$(-2) \cdot 1 + 0 \cdot 3 + 1 \cdot 2 = 0, \quad (-2) \cdot 2 + 0 \cdot 1 + 1 \cdot 5 = 1, \quad (-2) \cdot 0 + 0 \cdot 4 + 1 \cdot 1 = 1.$$

So

$$E_3A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 1 & 4 \\ 0 & 1 & 1 \end{bmatrix},$$

and the new row 3 is indeed $R_3 - 2R_1 = (2 - 2, 5 - 4, 1 - 0) = (0, 1, 1)$. Likewise

$$E_1A = \begin{bmatrix} 3 & 1 & 4 \\ 1 & 2 & 0 \\ 2 & 5 & 1 \end{bmatrix} \quad (\text{rows 1, 2 swapped}), \quad E_2A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 1 & 4 \\ 10 & 25 & 5 \end{bmatrix} \quad (\text{row 3 times 5}).$$

A whole elimination is a sequence of row operations, so it is a sequence of multiplications.

Proposition 3.6.5 (Sequences of row operations). Suppose row operations with elementary matrices $E_1, E_2, \dots, E_k$, used in this order, take the $m \times n$ matrix A to U . Then

$$U = E_k \cdots E_2 E_1 A.$$

In particular, if A and U are row equivalent, then $U = PA$ for a product P of elementary matrices.

Proof. We induct on k . For $k = 1$ this is Theorem 3.6.3. If the first $k - 1$ operations take A to $U' = E_{k-1} \cdots E_1 A$, then the last one takes U' to $E_k U'$ by Theorem 3.6.3, and $E_k U' = E_k (E_{k-1} \cdots E_1 A) = E_k E_{k-1} \cdots E_1 A$ by associativity. If A and U are row equivalent (Definition 2.3.3), a finite sequence of row operations takes A to U ; if it is empty, then $U = A = I_m A$, and I_m is elementary. $\square$

3.6.2 Undoing an elementary matrix

Every row operation can be undone by a row operation of the same kind:

operation	undone by
$R_i \leftrightarrow R_j$	$R_i \leftrightarrow R_j$ (the same swap)
$R_i \rightarrow cR_i, c \neq 0$	$R_i \rightarrow c^{-1}R_i$ (this is why $c \neq 0$)
$R_i \rightarrow R_i + cR_j$	$R_i \rightarrow R_i - cR_j$

Indeed, swapping twice restores the rows; scaling row i by c and then by c^{-1} multiplies it by $c^{-1}c = 1$; and adding c times row j to row i and then subtracting it again leaves row i as it was, because row j itself does not change ($i \neq j$). The same holds in the other order, since the reverse of the reverse operation is the original one. We call these the *reverse* operations.

Theorem 3.6.6 (Elementary matrices are invertible). Every elementary matrix E is invertible, and E^{-1} is the elementary matrix of the reverse row operation.

Proof. Let E be the $m \times m$ elementary matrix of a row operation, and let G be the elementary matrix of the reverse operation. By Theorem 3.6.3, GE is E after the reverse operation. But E is I_m after the original operation, so GE is I_m after the operation followed by its reverse, which is I_m . Exchanging the roles of the two operations, EG is I_m after the reverse operation followed by the original one, which is I_m too. So $GE = EG = I_m$, and $G = E^{-1}$. $\square$

Example 3.6.7. For the matrices of Example 3.6.2: $E_1^{-1} = E_1$ (swap again); E_2^{-1} scales row 3 by $\frac{1}{5}$; and E_3 subtracts 2 times row 1 from row 3, so E_3^{-1} adds it back:

$$E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}, \quad E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}.$$

As a check, $E_3^{-1}E_3$ is E_3 after $R_3 \rightarrow R_3 + 2R_1$. Row 3 of E_3 is $(-2, 0, 1)$ and two times row 1 is $(2, 0, 0)$, so the new row 3 is $(-2 + 2, 0 + 0, 1 + 0) = (0, 0, 1)$, and $E_3^{-1}E_3 = I_3$.

3.6.3 Why Gauss–Jordan inversion works

Suppose row operations take the $n \times n$ matrix A to I_n , and let $E_1, \dots, E_k$ be their elementary matrices, in the order we use them. By Proposition 3.6.5,

$$E_k \cdots E_2 E_1 A = I_n.$$

Put $B = E_k \cdots E_1$. Then $BA = I_n$, and by Proposition 3.5.9 (applied to the pair (B, A)) also $AB = I_n$, so $B = A^{-1}$. The same operations, performed on the whole matrix $[A \mid I_n]$, act on each column separately (column j of $E[A \mid I_n]$ is E times column j), so they take it to

$$[E_k \cdots E_1 A \mid E_k \cdots E_1 I_n] = [I_n \mid B] = [I_n \mid A^{-1}].$$

This is why the right block of Gauss–Jordan inversion ends up as A^{-1} (Theorem 3.5.10): it records the product of the elementary matrices used. We also learn something new.

Theorem 3.6.8 (Products of elementary matrices). Let A be an $n \times n$ matrix over $\mathbb{F}$.

- (a) If $E_k \cdots E_1 A = I_n$ for elementary matrices $E_1, \dots, E_k$, then A is invertible and $A^{-1} = E_k \cdots E_1$.
- (b) A is invertible if and only if A is a product of elementary matrices.

Proof. (a) With $B = E_k \cdots E_1$ we have $BA = I_n$, so $AB = I_n$ by Proposition 3.5.9, and $B = A^{-1}$.

(b) Suppose A is invertible. By Theorem 3.5.10(c) its reduced row echelon form is I_n , so row operations take A to I_n . With $E_1, \dots, E_k$ their elementary matrices, $E_k \cdots E_1 A = I_n$ (Proposition 3.6.5), and $B = E_k \cdots E_1 = A^{-1}$ by (a). By Theorem 3.5.8(a) and the rule for longer products after Theorem 3.5.8,

$$A = (A^{-1})^{-1} = B^{-1} = (E_k \cdots E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1}.$$

Each E_i^{-1} is elementary by Theorem 3.6.6, so A is a product of elementary matrices. (If $k = 0$, then $A = I_n$, which is elementary.) Conversely, every elementary matrix is invertible (Theorem 3.6.6), and a product of invertible $n \times n$ matrices is invertible (the rule for longer products), so a product of elementary matrices is invertible. $\square$

Example 3.6.9. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$. The replacement $R_2 \rightarrow R_2 - 3R_1$ gives $3 - 3 \cdot 1 = 0$ and $7 - 3 \cdot 2 = 1$, and then $R_1 \rightarrow R_1 - 2R_2$ gives $1 - 2 \cdot 0 = 1$ and $2 - 2 \cdot 1 = 0$:

$$\begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix} \xrightarrow{R_2 \rightarrow R_2 - 3R_1} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \xrightarrow{R_1 \rightarrow R_1 - 2R_2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The elementary matrices of the two steps are E_2 after the same operations:

$$E_1 = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}.$$

By Theorem 3.6.8(a) and the row-column rule,

$$A^{-1} = E_2 E_1 = \begin{bmatrix} 1 \cdot 1 + (-2)(-3) & 1 \cdot 0 + (-2) \cdot 1 \\ 0 \cdot 1 + 1 \cdot (-3) & 0 \cdot 0 + 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 7 & -2 \\ -3 & 1 \end{bmatrix}.$$

The 2×2 formula (Theorem 3.5.6) agrees: $ad - bc = 7 - 6 = 1$, so we just swap 1 and 7 and change the signs of 2 and 3. Read the other way,

$$A = E_1^{-1} E_2^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 0 \cdot 0 & 1 \cdot 2 + 0 \cdot 1 \\ 3 \cdot 1 + 1 \cdot 0 & 3 \cdot 2 + 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}.$$

E_1^{-1} adds three times row 1 back to row 2, and E_2^{-1} adds two times row 2 back to row 1. In the plane, A is a shear followed by another shear.

3.6.4 The invertible matrix theorem

Theorem 3.6.10 (The invertible matrix theorem). Let A be an $n \times n$ matrix over $\mathbb{F}$. The following statements are equivalent; that is, either they are all true, or they are all false.

- (1) A is invertible.
- (2) There is an $n \times n$ matrix C with $CA = I_n$.
- (3) The equation $Ax = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.

- (4) A has n pivot positions.
- (5) The reduced row echelon form of A is I_n .
- (6) The columns of A are linearly independent.
- (7) For every $\mathbf{b} \in \mathbb{F}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution.
- (8) The columns of A span $\mathbb{F}^n$.
- (9) For every $\mathbf{b} \in \mathbb{F}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution.
- (10) There is an $n \times n$ matrix D with $AD = I_n$.
- (11) A^T is invertible.
- (12) A is a product of elementary matrices.

So to show that a square matrix is invertible, it is enough to check any one of these statements; and if any one of them fails, they all fail. Figure 3.13 shows the plan of the proof: a loop through (1) to (5), and every other statement attached to the loop. The proof has many steps, but each one is short; read it one step at a time.

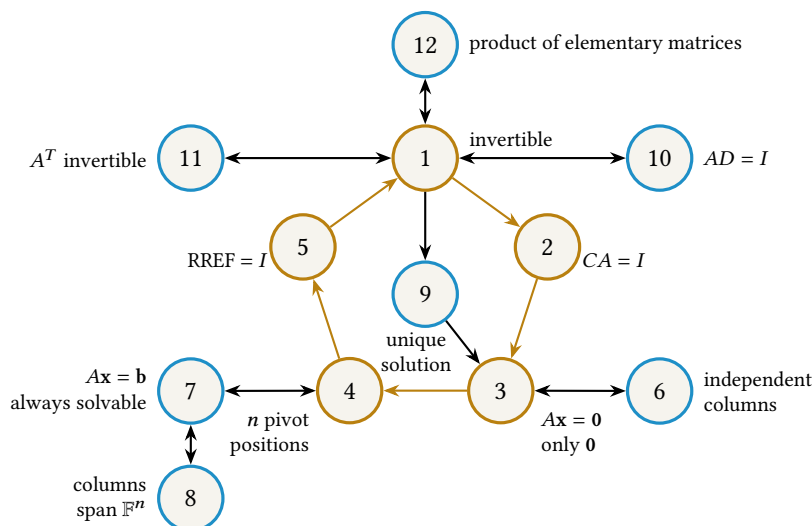


Figure 3.13: The proof of Theorem 3.6.10. Single arrows are implications, double arrows equivalences. The loop $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5) \Rightarrow (1)$ makes (1) to (5) equivalent, and every other statement is joined to the loop.

Proof. We first prove the loop $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5) \Rightarrow (1)$; then each of (1) to (5) implies all the others. Then we attach (6) to (12).

$(1) \Rightarrow (2)$. Take $C = A^{-1}$.

$(2) \Rightarrow (3)$. Suppose $CA = I_n$ and $A\mathbf{x} = \mathbf{0}$. Then $\mathbf{x} = I_n\mathbf{x} = (CA)\mathbf{x} = C(A\mathbf{x}) = C\mathbf{0} = \mathbf{0}$.

$(3) \Rightarrow (4)$. By Proposition 3.1.13, an echelon form of A has a pivot in every column, so it has n pivots. By Proposition 2.4.14 its pivots sit exactly in the pivot positions of A , so A has n pivot positions.

$(4) \Rightarrow (5)$. Let R be the reduced row echelon form of A . It has a pivot in each of the n pivot positions. Leading entries move strictly to the right as we go down (Definition 2.4.1(ii)), so no column holds more than one pivot, and the n pivots lie in n different columns. R has only n columns, so every column of R has a pivot. As shown in the proof of Theorem 3.5.10(b), a square matrix in reduced row echelon form

with a pivot in every column is I_n . So $R = I_n$.

(5) $\Rightarrow$ (1). This is Theorem 3.5.10(c).

(3) $\Leftrightarrow$ (6). This is the equivalence (b) $\Leftrightarrow$ (c) of Proposition 3.1.13.

(4) $\Leftrightarrow$ (7) $\Leftrightarrow$ (8). Let U be an echelon form of A ; by Proposition 2.4.14, A has n pivot positions exactly when U has n pivots. Each pivot is the leading entry of its row, so each row of U holds at most one pivot. Hence U has n pivots if and only if each of its n rows has a pivot. By Theorem 3.1.9, a pivot in every row is equivalent to (7) and to (8).

(1) $\Rightarrow$ (9) $\Rightarrow$ (3). If A is invertible, then for every $\mathbf{b}$ the equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution, $\mathbf{x} = A^{-1}\mathbf{b}$ (Theorem 3.5.5). If (9) holds, then in particular $A\mathbf{x} = \mathbf{0}$ has exactly one solution, and since $\mathbf{x} = \mathbf{0}$ is a solution, it is the only one.

(1) $\Leftrightarrow$ (10). If A is invertible, take $D = A^{-1}$. Conversely, if $AD = I_n$, then A is invertible (with $A^{-1} = D$) by Proposition 3.5.9.

(1) $\Leftrightarrow$ (11). If A is invertible, so is A^T , by Theorem 3.5.8(c). Conversely, if A^T is invertible, then by the same rule so is $(A^T)^T$, which is A (Theorem 3.4.3(a)).

(1) $\Leftrightarrow$ (12). This is Theorem 3.6.8(b).

Every statement is now joined to the loop, so all twelve are equivalent. $\square$

Some of these equivalences we had already met: (1) $\Leftrightarrow$ (5) is Gauss–Jordan inversion, and (1) $\Leftrightarrow$ (10) is Proposition 3.5.9. What is new is how much the single condition “ A is invertible” carries. For instance, (3) is about uniqueness of solutions and (7) about their existence: for a square matrix, uniqueness for $\mathbf{b} = \mathbf{0}$ alone already guarantees existence for every $\mathbf{b}$.

Example 3.6.11. Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

over any field. It is already in echelon form, with pivots 1, -1 and 1, all nonzero in every field. So A has three pivot positions, and statement (4) holds. By Theorem 3.6.10, all twelve statements hold: without any further work we know that A is invertible, that its columns are linearly independent and span $\mathbb{F}^3$, and that $A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $\mathbf{b} \in \mathbb{F}^3$.

Example 3.6.12. Let A be the 3×3 matrix with columns

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \\ 0 & 1 & 1 \end{bmatrix}.$$

The third column is the sum of the first two: $1 + 0 = 1$, $2 + 1 = 3$, $0 + 1 = 1$. So $\mathbf{a}_1 + \mathbf{a}_2 - \mathbf{a}_3 = \mathbf{0}$ is a linear dependence relation, and statement (6) fails. By Theorem 3.6.10, all twelve fail. In particular A is not invertible, and by (7) there is some $\mathbf{b} \in \mathbb{F}^3$ for which $A\mathbf{x} = \mathbf{b}$ has no solution. We know this without solving a single system. To find such a $\mathbf{b}$, row reduce: $R_2 \rightarrow R_2 - 2R_1$ and then $R_3 \rightarrow R_3 - R_2$ give

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & b_1 \\ 2 & 1 & 3 & b_2 \\ 0 & 1 & 1 & b_3 \end{array} \right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & b_1 \\ 0 & 1 & 1 & b_2 - 2b_1 \\ 0 & 0 & 0 & b_3 - b_2 + 2b_1 \end{array} \right],$$

so by Theorem 2.5.3(a) the system is inconsistent whenever $2b_1 - b_2 + b_3 \neq 0$; for example $\mathbf{b} = \mathbf{e}_3$.

Remark 3.6.13 (Only for square matrices). The theorem needs A to be square. The 2×3 matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

is in echelon form with a pivot in every row, so its columns span $\mathbb{F}^2$ and $A\mathbf{x} = \mathbf{b}$ is solvable for every $\mathbf{b}$ (Theorem 3.1.9). But its third column is zero, so $0\mathbf{a}_1 + 0\mathbf{a}_2 + 1\mathbf{a}_3 = \mathbf{0}$ and its columns are linearly dependent. Its transpose, the 3×2 matrix A^T , has a pivot in every column, so its columns are independent (Proposition 3.1.13), but its third row is zero, so it has no pivot there and its columns do not span $\mathbb{F}^3$. For matrices that are not square, the statements of the theorem split apart.

Here is one consequence that is used often.

Proposition 3.6.14. Let A and B be $n \times n$ matrices. If AB is invertible, then A and B are both invertible.

Proof. Let $M = (AB)^{-1}$. Then $A(BM) = (AB)M = I_n$, so statement (10) holds for A (with $D = BM$), and A is invertible. Also $(MA)B = M(AB) = I_n$, so statement (2) holds for B (with $C = MA$), and B is invertible. $\square$

Together with Theorem 3.5.8(b), this says that AB is invertible exactly when both A and B are. For matrices that are not square this fails: $\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = I_1$.

Remark 3.6.15 (The theorem for transformations). Let A be $n \times n$ and $T = T_A: \mathbb{F}^n \rightarrow \mathbb{F}^n$, $T(\mathbf{x}) = A\mathbf{x}$ (Section 3.2). Call T *one-to-one* if different inputs always have different outputs, and *onto* if every $\mathbf{b} \in \mathbb{F}^n$ is an output $T(\mathbf{x})$. Then each of the following is equivalent to (1):

- (a) T is one-to-one;
- (b) T is onto;
- (c) there is a linear transformation $S: \mathbb{F}^n \rightarrow \mathbb{F}^n$ with $S(T(\mathbf{x})) = \mathbf{x}$ and $T(S(\mathbf{x})) = \mathbf{x}$ for every $\mathbf{x}$; then $S = T_{A^{-1}}$.

Proof. (a) $\Leftrightarrow$ (3): $T(\mathbf{x}) = T(\mathbf{y})$ means $A\mathbf{x} - A\mathbf{y} = A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$ (Theorem 3.1.6). If (3) holds, this forces $\mathbf{x} - \mathbf{y} = \mathbf{0}$, so T is one-to-one. Conversely, if T is one-to-one and $A\mathbf{x} = \mathbf{0} = A\mathbf{0}$, then $\mathbf{x} = \mathbf{0}$. (b) $\Leftrightarrow$ (7): this is the definition of onto, since $T(\mathbf{x}) = A\mathbf{x}$. (c): If A is invertible, $S = T_{A^{-1}}$ works, because $A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = \mathbf{x}$ and $A(A^{-1}\mathbf{x}) = \mathbf{x}$. Conversely, let S be such a map and C its standard matrix (Theorem 3.2.7). Then $(CA)\mathbf{x} = C(A\mathbf{x}) = S(T(\mathbf{x})) = \mathbf{x}$ for every $\mathbf{x}$; for $\mathbf{x} = \mathbf{e}_j$ this says that column j of CA is $\mathbf{e}_j$, so $CA = I_n$, and (2) holds. Then A is invertible, $C = A^{-1}$ by Proposition 3.5.4, and $S = T_{A^{-1}}$. $\square$

So the answer to the question that opened Section 3.5, “can we undo what A does?”, is yes exactly when A is invertible, and then A^{-1} does the undoing.

Remark 3.6.16 (Preview: the list keeps growing). Let A be an $n \times n$ matrix over $\mathbb{F}$. Later chapters add further statements that are equivalent to (1) to (12), and prove them there:

- the column space of A is $\mathbb{F}^n$, and the null space of A is $\{\mathbf{0}\}$ (Section 4.1, Lecture 19);
- the columns of A form a basis of $\mathbb{F}^n$ (Section 4.2, Lecture 20);
- the rank of A is n (Section 5.2);
- $\det A \neq 0$ (Chapter 7);
- 0 is not an eigenvalue of A (Chapter 8).

Check yourself

- (a) Write down the 3×3 elementary matrix E of $R_2 \rightarrow R_2 + 4R_3$ and its inverse, and check that $EE^{-1} = I_3$ with Theorem 3.6.3.
- (b) Let $A = \begin{bmatrix} 2 & 1 \\ 6 & 4 \end{bmatrix}$ over $\mathbb{R}$. Find elementary matrices E_1, E_2, E_3 with $E_3E_2E_1A = I_2$, and write both A^{-1} and A as products of elementary matrices.
- (c) Let A be $n \times n$. True or false? (i) If $Ax = \mathbf{e}_1$ has two different solutions, then the columns of A do not span $\mathbb{F}^n$. (ii) If $Ax = \mathbf{e}_1$ has a solution, then A is invertible.

Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

Column operations. Multiplying by an elementary matrix on the *right* performs a column operation. Let E be the $n \times n$ elementary matrix of $R_i \rightarrow R_i + cR_j$, so E is I_n with an extra c in position (i, j) . Its transpose E^T has the extra c in position (j, i) , so it is the elementary matrix of $R_j \rightarrow R_j + cR_i$. For an $m \times n$ matrix A , Theorem 3.4.3 gives $AE = ((AE)^T)^T = (E^T A^T)^T$. By Theorem 3.6.3, $E^T A^T$ is A^T with c times row i added to row j , and the rows of A^T are the columns of A . So AE is A with c times column i added to column j . Note the exchange of i and j . In the same way, for a swap or a scaling $E^T = E$, and AE swaps or scales the corresponding columns of A . For example, with $E = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$ (the matrix of $R_2 \rightarrow R_2 - 3R_1$),

$$\begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 1-6 & 2 \\ 3-21 & 7 \end{bmatrix} = \begin{bmatrix} -5 & 2 \\ -18 & 7 \end{bmatrix} :$$

-3 times column 2 was added to column 1.

The field matters. Theorem 3.6.10 holds over every field, but whether a given matrix satisfies its statements can depend on the field. The matrix $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ has $ad - bc = -1 - 1 = -2$. Over $\mathbb{Q}$ or $\mathbb{R}$ this is nonzero, so the matrix is invertible and all twelve statements hold. Over $\mathbb{Z}_2$, where $-2 = 0$, it equals $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$; its two columns are equal, so statement (6) fails and all twelve fail.

3.7 LU factorisation**Lecture 18**

In practice we often solve many systems with the same matrix A and different right-hand sides: the same bridge under different loads, or the same circuit with different inputs. Each time, elimination performs the same row operations on A , and only the right-hand side changes. It makes sense to do the work on A once and to keep a record of it.

That record is the *LU factorisation* $A = LU$. Here U is the echelon form that elimination produces, and L stores the multipliers that were used to get there. In this section we see how elimination becomes a factorisation and why the multipliers land exactly in L ; then we solve systems with L and U , count how much work this saves, and see what to do when a row swap is needed.

Throughout, $\mathbb{F}$ is a field and all matrices have entries in $\mathbb{F}$. We write $\text{row}_i(A)$ for the i -th row of a matrix A , a $1 \times n$ matrix. We use one consequence of the row-column rule (Theorem 3.3.5) again and again.

Rows of a product. Let $C = (c_{ik})$ be $m \times k$ and B be $k \times n$. Then

$$\text{row}_i(CB) = c_{i1} \text{row}_1(B) + c_{i2} \text{row}_2(B) + \cdots + c_{ik} \text{row}_k(B),$$

a combination of the rows of B with the entries of row i of C as weights. Indeed, entry j of the left side is $(CB)_{ij} = \sum_q c_{iq} b_{qj}$ by Theorem 3.3.5, and entry j of the right side is the same sum.

Elementary matrices (Section 3.6). The only row operation that elimination without swaps uses is the replacement $R_i \rightarrow R_i - lR_j$ with $i \neq j$ and $l \in \mathbb{F}$. Its elementary matrix E (Definition 3.6.1) is I_m with the entry in row i , column j changed from 0 to $-l$. By Theorem 3.6.3, EA is A after the replacement, for every $m \times n$ matrix A . By Theorem 3.6.6, E is invertible and E^{-1} is the elementary matrix of the reverse operation $R_i \rightarrow R_i + lR_j$, that is, I_m with $+l$ in row i , column j : *the inverse flips the sign of the multiplier entry*.

3.7.1 Elimination as a product of matrices

Example 3.7.1. Let

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 5 & 3 \\ -2 & 8 & 4 \end{bmatrix}.$$

The first pivot is 2. To clear the 4 below it, the *multiplier* is $l_{21} = 4/2 = 2$, the entry we clear divided by the pivot, and we apply $R_2 \rightarrow R_2 - 2R_1$: row 2 becomes $(4 - 2 \cdot 2, 5 - 2 \cdot 1, 3 - 2 \cdot 1) = (0, 3, 1)$. To clear the -2 in row 3, the multiplier is $l_{31} = -2/2 = -1$, and $R_3 \rightarrow R_3 - (-1)R_1$ adds row 1: row 3 becomes $(-2 + 2, 8 + 1, 4 + 1) = (0, 9, 5)$. The second pivot is 3; below it is 9, so $l_{32} = 9/3 = 3$, and $R_3 \rightarrow R_3 - 3R_2$ gives $(0 - 3 \cdot 0, 9 - 3 \cdot 3, 5 - 3 \cdot 1) = (0, 0, 2)$. The result

$$U = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

is in row echelon form, with pivots 2, 3, 2. The three steps are multiplications by

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix},$$

so $E_3 E_2 E_1 A = U$.

Example 3.7.2 (From the steps to L). To get A back from Example 3.7.1 we undo the steps in reverse order. Multiplying $E_3 E_2 E_1 A = U$ on the left by E_3^{-1} , then E_2^{-1} , then E_1^{-1} gives

$$A = E_1^{-1} E_2^{-1} E_3^{-1} U, \quad E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}.$$

We multiply from the right. $E_2^{-1} E_3^{-1}$ is E_3^{-1} after $R_3 \rightarrow R_3 - R_1$, which makes row 3 equal to $(0, 3, 1) - (1, 0, 0) = (-1, 3, 1)$. Then E_1^{-1} applies $R_2 \rightarrow R_2 + 2R_1$, which makes row 2 equal to $(0, 1, 0) + 2(1, 0, 0) = (2, 1, 0)$. So

$$L = E_1^{-1} E_2^{-1} E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{bmatrix}, \quad A = LU.$$

Below the diagonal of L sit exactly the multipliers $l_{21} = 2$, $l_{31} = -1$, $l_{32} = 3$, each in the position of the entry it cleared (Figure 3.14).

The product in the other order, which performs the elimination, is not so tidy. E_2E_1 is E_1 after $R_3 \rightarrow R_3 + R_1$, with rows $(1, 0, 0)$, $(-2, 1, 0)$, $(1, 0, 1)$; then E_3 subtracts three times row 2 from row 3:

$$E_3E_2E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 7 & -3 & 1 \end{bmatrix},$$

since $1 - 3 \cdot (-2) = 7$. The entry 7 is not one of the multipliers: in $E_3E_2E_1$, the step E_3 acts on a row that E_1 has already changed. In $E_1^{-1}E_2^{-1}E_3^{-1}$, each step acts on row 1 or on an identity row, and the multipliers stay apart.

As a check of $A = LU$ by the rule for rows of a product: row 1 of LU is $\text{row}_1(U) = (2, 1, 1)$; row 2 is $2(2, 1, 1) + (0, 3, 1) = (4, 5, 3)$; row 3 is $-(2, 1, 1) + 3(0, 3, 1) + (0, 0, 2) = (-2, -1+9, -1+3+2) = (-2, 8, 4)$.

$$\begin{array}{ccc}
 \begin{array}{c} A \\ \left[\begin{array}{ccc} \boxed{2} & 1 & 1 \\ 4 & 5 & 3 \\ -2 & 8 & 4 \end{array} \right] \end{array} & \xrightarrow[\substack{l_{21} = \frac{4}{2} = 2 \\ l_{31} = \frac{-2}{2} = -1}]{\text{stage 1}} & \begin{array}{c} \left[\begin{array}{ccc} 2 & 1 & 1 \\ 0 & \boxed{3} & 1 \\ 0 & 9 & 5 \end{array} \right] \end{array} \xrightarrow[\substack{l_{32} = \frac{9}{3} = 3}]{\text{stage 2}} \begin{array}{c} U \\ \left[\begin{array}{ccc} \boxed{2} & 1 & 1 \\ 0 & \boxed{3} & 1 \\ 0 & 0 & \boxed{2} \end{array} \right] \end{array} \\
 & & \searrow \quad \quad \quad \swarrow \\
 & & L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 3 & 1 \end{bmatrix}
 \end{array}$$

Figure 3.14: Examples 3.7.1 and 3.7.2: each multiplier l_{ip} (blue) is the entry it clears divided by the pivot (gold), and it goes into L in the position of that entry.

Definition 3.7.3 (Unit lower triangular, LU factorisation). A square matrix is *unit lower triangular* if it is lower triangular (Definition 3.4.10) and every entry on its main diagonal is 1. An *LU factorisation* of an $m \times n$ matrix A over $\mathbb{F}$ is a product

$$A = LU,$$

where L is an $m \times m$ unit lower triangular matrix and U is an $m \times n$ matrix in row echelon form.

A product of unit lower triangular matrices of the same size is unit lower triangular, by Theorem 3.4.11(b): it is lower triangular with diagonal entries $1 \cdot 1 = 1$.

3.7.2 Why the multipliers land in L

Forward elimination (Section 2.3) works in *stages*. In stage p the current row p is the *pivot row*; every row below it is replaced by itself minus a multiple of row p , to clear the entries below the pivot; then row p is set aside. Sometimes a stage needs a swap first, when the entry in the pivot position is zero but an entry below it is not. When no stage needs a swap, the multipliers alone describe the whole elimination.

Theorem 3.7.4 (Elimination without swaps is a factorisation). Let A be an $m \times n$ matrix over $\mathbb{F}$. Suppose that forward elimination brings A to a row echelon form U using only replacements, with no row swaps: in stage p ($p = 1, \dots, m-1$), each row $i > p$ is replaced by

$$R_i \rightarrow R_i - l_{ip}R_p \quad (l_{ip} \in \mathbb{F}),$$

with $l_{ip} = 0$ wherever no step is needed. Let L be the $m \times m$ unit lower triangular matrix whose entry in row i , column p is l_{ip} for $i > p$. Then

$$A = LU.$$

Proof. Stage p changes only the rows below row p . So once row p becomes the pivot row, it never changes again, and at that moment it is already $\text{row}_p(U)$. In particular, every step $R_i \rightarrow R_i - l_{ip}R_p$ subtracts $l_{ip} \text{row}_p(U)$.

Row i is changed only in stages $p = 1, \dots, i-1$, once in each. Starting from $\text{row}_i(A)$, it ends as

$$\text{row}_i(U) = \text{row}_i(A) - l_{i1} \text{row}_1(U) - \dots - l_{i,i-1} \text{row}_{i-1}(U).$$

Moving the subtracted rows to the other side,

$$\text{row}_i(A) = l_{i1} \text{row}_1(U) + \dots + l_{i,i-1} \text{row}_{i-1}(U) + 1 \cdot \text{row}_i(U) + 0 \cdot \text{row}_{i+1}(U) + \dots + 0 \cdot \text{row}_m(U).$$

The weights here are the entries of row i of L , so by the rule for rows of a product the right side is $\text{row}_i(LU)$. This holds for every i (for $i = 1$ it reads $\text{row}_1(A) = \text{row}_1(U)$), so $A = LU$. $\square$

So to find L we never multiply matrices: we write down each multiplier in the position of the entry it cleared. The theorem applies to any $m \times n$ matrix, square or not, invertible or not. Elimination may even stop early when the remaining rows are zero; the later multipliers are then 0.

3.7.3 Solving with L and U

If $A = LU$, then $Ax = L(Ux)$. Naming $y = Ux$ splits the system in two triangular ones.

Theorem 3.7.5 (Solving with an LU factorisation). Let $A = LU$, where A is $m \times n$, L is $m \times m$ unit lower triangular and U is $m \times n$. Let $\mathbf{b} \in \mathbb{F}^m$.

- (a) The system $Ly = \mathbf{b}$ has exactly one solution $\mathbf{y}$, found from the top down by *forward substitution*:

$$y_1 = b_1, \quad y_i = b_i - l_{i1}y_1 - \dots - l_{i,i-1}y_{i-1} \quad (i = 2, \dots, m).$$

- (b) For this $\mathbf{y}$, a vector $\mathbf{x} \in \mathbb{F}^n$ satisfies $Ax = \mathbf{b}$ if and only if $Ux = \mathbf{y}$.

Proof. (a) Row i of $Ly = \mathbf{b}$ reads $l_{i1}y_1 + \dots + l_{i,i-1}y_{i-1} + y_i = b_i$, since $l_{ii} = 1$ and the entries of L to the right of the diagonal are zero. Row 1 forces $y_1 = b_1$; once $y_1, \dots, y_{i-1}$ are known, row i forces the stated value of y_i . So there is at most one solution, and the vector built this way satisfies every row, so it is a solution.

(b) If $Ux = \mathbf{y}$, then $Ax = (LU)x = L(Ux) = Ly = \mathbf{b}$. Conversely, if $Ax = \mathbf{b}$, then $L(Ux) = \mathbf{b}$, so Ux is a solution of $Ly = \mathbf{b}$; by (a) it equals $\mathbf{y}$. $\square$

The system $U\mathbf{x} = \mathbf{y}$ is in echelon form, so it is solved by back substitution (Section 2.3); it may have one solution, none, or infinitely many, just as $A\mathbf{x} = \mathbf{b}$ does.

Example 3.7.6. We solve $A\mathbf{x} = (3, 5, -2)$ with L and U from Example 3.7.2.

Forward substitution, $L\mathbf{y} = \mathbf{b}$:

$$\begin{aligned} y_1 &= 3, \\ 2y_1 + y_2 &= 5 & \Rightarrow y_2 = 5 - 6 = -1, \\ -y_1 + 3y_2 + y_3 &= -2 & \Rightarrow y_3 = -2 + 3 + 3 = 4. \end{aligned}$$

Back substitution, $U\mathbf{x} = \mathbf{y} = (3, -1, 4)$:

$$\begin{aligned} 2x_3 &= 4 & \Rightarrow x_3 = 2, \\ 3x_2 + x_3 &= -1 & \Rightarrow 3x_2 = -3, \quad x_2 = -1, \\ 2x_1 + x_2 + x_3 &= 3 & \Rightarrow 2x_1 = 3 + 1 - 2 = 2, \quad x_1 = 1. \end{aligned}$$

So $\mathbf{x} = (1, -1, 2)$. Check by the row rule: $A\mathbf{x} = (2 - 1 + 2, 4 - 5 + 6, -2 - 8 + 8) = (3, 5, -2) = \mathbf{b}$. For another right-hand side we repeat only the two substitutions.

Preview (Section 7.3). Let A be an $n \times n$ matrix with $A = LU$, where L is unit lower triangular and U is in row echelon form. Then $\det A = u_{11}u_{22} \cdots u_{nn}$, the product of the diagonal entries of U . Determinants are defined in Chapter 7, where this is proved; for Example 3.7.1 it gives $\det A = 2 \cdot 3 \cdot 2 = 12$.

3.7.4 How much work it saves

We count multiplications and divisions, the usual measure of the cost of elimination.

Proposition 3.7.7 (Operation counts). Let A be an $n \times n$ matrix whose forward elimination needs no row swaps, as in Theorem 3.7.4.

- (a) Computing U and the multipliers takes at most $\frac{n^3 - n}{3}$ multiplications and divisions.
- (b) If the diagonal entries of U are all nonzero, then for each $\mathbf{b} \in \mathbb{F}^n$, forward substitution takes $\frac{n(n-1)}{2}$ multiplications, and back substitution takes $\frac{n(n-1)}{2}$ multiplications and n divisions: n^2 operations in all.

Proof. (a) In stage p there are $n - p$ rows below the pivot row. Let the pivot lie in column $c \geq p$ (each pivot lies at least one column to the right of the one above). For each row below, the multiplier takes one division, the entry in column c becomes 0 without arithmetic, the entries in columns before c are already 0 and stay so, and each of the $n - c \leq n - p$ entries to the right of column c takes one multiplication (by the multiplier) before the subtraction. So stage p takes at most $(n - p)(n - p + 1)$ operations, and a stage with no pivot (only zero rows left) takes none. With $k = n - p$, the total is at most

$$\sum_{k=1}^{n-1} k(k+1) = \sum_{k=1}^{n-1} k^2 + \sum_{k=1}^{n-1} k = \frac{(n-1)n(2n-1)}{6} + \frac{(n-1)n}{2} = \frac{(n-1)n(2n+2)}{6} = \frac{n^3 - n}{3}.$$

(b) In forward substitution, $y_i = b_i - l_{i1}y_1 - \cdots - l_{i,i-1}y_{i-1}$ takes $i - 1$ multiplications and no division, so $\sum_{i=1}^n (i - 1) = \frac{n(n-1)}{2}$ in all. Since the diagonal entries u_{ii} are nonzero, back substitution computes $x_i = (y_i - u_{i,i+1}x_{i+1} - \cdots - u_{in}x_n)/u_{ii}$ for $i = n, \dots, 1$, with $n - i$ multiplications and one division, so $\sum_{i=1}^n (n - i) = \frac{n(n-1)}{2}$ multiplications and n divisions. The total is $n(n - 1) + n = n^2$. $\square$

For $n = 1000$, elimination costs at most $(10^9 - 10^3)/3 = 333\,333\,000$ operations, about a third of a billion, while each new right-hand side costs 10^6 , more than three hundred times less (Figure 3.15). This is why software factors a matrix once and reuses L and U for every new $\mathbf{b}$. The factors also take no extra memory: U fills the upper part of an $n \times n$ array, and the multipliers fill the places below the diagonal, which are exactly the entries that elimination turns into zeros.

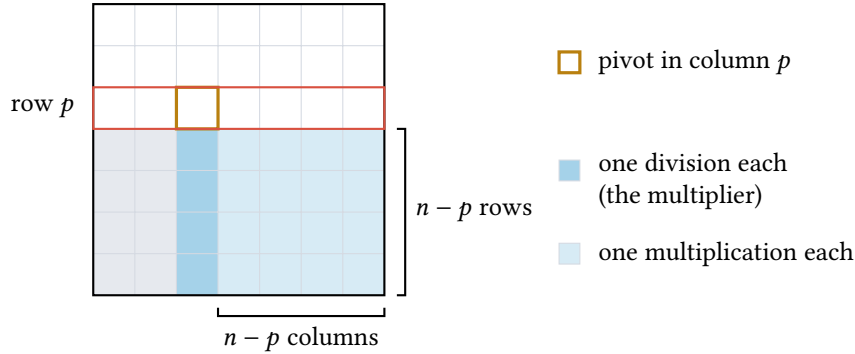


Figure 3.15: Stage p of elimination on an $n \times n$ matrix with its pivot in column p : each of the $n - p$ rows below the pivot row needs one division for its multiplier and one multiplication for each of the $n - p$ entries to the right of the pivot column (shaded), that is, $n - p + 1$ operations per row.

3.7.5 When a row swap is needed

Example 3.7.8 (A matrix with no LU factorisation). Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. Its first pivot position holds a 0, so elimination must start with a swap. In fact A has no LU factorisation. Suppose $A = LU$ with $L = \begin{bmatrix} 1 & 0 \\ l & 1 \end{bmatrix}$ and U in row echelon form. Row 1 of LU is $\text{row}_1(U)$, so $\text{row}_1(U) = (0, 1)$, with leading entry in column 2. In an echelon form row 2 must then be zero, so $u_{21} = 0$. Now the entry of LU in row 2, column 1 is $l u_{11} + 1 \cdot u_{21} = l \cdot 0 + 0 = 0$, but the entry of A there is 1. This contradiction shows that no such L and U exist.

The fix is to make the swaps first.

Definition 3.7.9 (Permutation matrix). An $m \times m$ matrix P is a *permutation matrix* if its rows are the rows of I_m in some order, each used once: $\text{row}_i(P) = \mathbf{e}_{\sigma(i)}^T$ for $i = 1, \dots, m$, where $\sigma(1), \dots, \sigma(m)$ are the numbers $1, \dots, m$ in some order.

For example, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and I_m itself, are permutation matrices. The elementary matrix of a swap $R_i \leftrightarrow R_j$ is a permutation matrix.

Proposition 3.7.10 (Properties of permutation matrices). Let P be an $m \times m$ permutation matrix with $\text{row}_i(P) = \mathbf{e}_{\sigma(i)}^T$.

- (a) For every $m \times n$ matrix A , $\text{row}_i(PA) = \text{row}_{\sigma(i)}(A)$: multiplying by P on the left puts the rows of A in the same order as the rows of I_m in P .
- (b) The product of two $m \times m$ permutation matrices is a permutation matrix.
- (c) P is invertible, and $P^{-1} = P^T$.

Proof. (a) Row i of P has a 1 in column $\sigma(i)$ and zeros elsewhere, so by the rule for rows of a product, $\text{row}_i(PA) = 1 \cdot \text{row}_{\sigma(i)}(A)$.

(b) Let Q be a permutation matrix with $\text{row}_k(Q) = \mathbf{e}_{\tau(k)}^T$. By (a), $\text{row}_i(PQ) = \text{row}_{\sigma(i)}(Q) = \mathbf{e}_{\tau(\sigma(i))}^T$. As i runs through $1, \dots, m$, so does $\sigma(i)$, each value once, and hence so does $\tau(\sigma(i))$. So the rows of PQ are the rows of I_m , each used once.

(c) Column j of P^T is $\text{row}_j(P)$ written as a column, so by the row-column rule $(PP^T)_{ij} = \sum_k p_{ik}p_{jk}$. Row i of P has its only nonzero entry, a 1, in column $\sigma(i)$, and row j in column $\sigma(j)$. So the sum is 1 if $\sigma(i) = \sigma(j)$, that is, if $i = j$, and 0 otherwise. So $PP^T = I_m$, and Proposition 3.5.9 gives $P^TP = I_m$. $\square$

Theorem 3.7.11 ($PA = LU$). Every $m \times n$ matrix A over $\mathbb{F}$ has a factorisation

$$PA = LU,$$

where P is an $m \times m$ permutation matrix, L is an $m \times m$ unit lower triangular matrix and U is an $m \times n$ matrix in row echelon form.

The idea: if we knew in advance which swaps elimination will make, we could make them all at the start and then eliminate with no swaps. The proof does this one stage at a time, carrying the multipliers along when later swaps reorder the rows.

Proof. By induction on m . If $m = 1$, take $P = L = [1]$ and $U = A$: a single row is in row echelon form.

Let $m \geq 2$. If $A = O$, take $P = L = I_m$ and $U = O$. Otherwise let column c be the leftmost column of A with a nonzero entry, say in row r , and let P_1 be the permutation matrix that swaps rows 1 and r ($P_1 = I_m$ if $r = 1$). By Proposition 3.7.10(a), P_1A has a nonzero entry a in row 1, column c , and its first $c - 1$ columns are zero. Cut P_1A after row 1 and after columns $c - 1$ and c :

$$P_1A = \begin{bmatrix} O & a & \mathbf{w} \\ O & \mathbf{v} & B \end{bmatrix},$$

where $\mathbf{v} \in \mathbb{F}^{m-1}$, $\mathbf{w}$ is $1 \times (n - c)$ and B is $(m - 1) \times (n - c)$; blocks with no columns (when $c = 1$ or $c = n$) are simply left out. Put $\mathbf{l} = a^{-1}\mathbf{v}$ and $B' = B - \mathbf{l}\mathbf{w}$. By block multiplication (Theorem 3.4.13),

$$P_1A = \begin{bmatrix} 1 & O \\ \mathbf{l} & I_{m-1} \end{bmatrix} \begin{bmatrix} O & a & \mathbf{w} \\ O & \mathbf{0} & B' \end{bmatrix},$$

since the second block row of the product is $[O \quad \mathbf{l}a \quad \mathbf{l}\mathbf{w} + B'] = [O \quad \mathbf{v} \quad B]$. (This is stage 1 of elimination: $\mathbf{l}$ holds its multipliers.)

If $c = n$, the second factor is already in row echelon form (its row 1 has leading entry a in the last column and its other rows are zero), and $P = P_1$ works, with L the first factor and U the second. If $c < n$, the induction hypothesis gives an $(m - 1) \times (m - 1)$ permutation matrix P' , a unit lower triangular L' and a row echelon form U' with $P'B' = L'U'$. Let

$$Q = \begin{bmatrix} 1 & O \\ O & P' \end{bmatrix}, \quad P = QP_1.$$

The rows of Q are $\mathbf{e}_1^T$ followed by the rows of P' with a 0 put in front, which are $\mathbf{e}_2^T, \dots, \mathbf{e}_m^T$ in some order; so Q is a permutation matrix, and so is P by Proposition 3.7.10(b). By block multiplication,

$$\begin{bmatrix} 1 & O \\ O & P' \end{bmatrix} \begin{bmatrix} 1 & O \\ \mathbf{l} & I_{m-1} \end{bmatrix} = \begin{bmatrix} 1 & O \\ P'\mathbf{l} & P' \end{bmatrix} = \begin{bmatrix} 1 & O \\ P'\mathbf{l} & I_{m-1} \end{bmatrix} \begin{bmatrix} 1 & O \\ O & P' \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & O \\ O & P' \end{bmatrix} \begin{bmatrix} O & a & \mathbf{w} \\ O & \mathbf{0} & B' \end{bmatrix} = \begin{bmatrix} O & a & \mathbf{w} \\ O & \mathbf{0} & P'B' \end{bmatrix} = \begin{bmatrix} O & a & \mathbf{w} \\ O & \mathbf{0} & L'U' \end{bmatrix} = \begin{bmatrix} 1 & O \\ O & L' \end{bmatrix} \begin{bmatrix} O & a & \mathbf{w} \\ O & \mathbf{0} & U' \end{bmatrix}.$$

Hence

$$PA = QP_1A = \begin{bmatrix} 1 & O \\ P'\mathbf{1} & I_{m-1} \end{bmatrix} \begin{bmatrix} 1 & O \\ O & L' \end{bmatrix} \begin{bmatrix} O & a & \mathbf{w} \\ O & \mathbf{0} & U' \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & O \\ P'\mathbf{1} & L' \end{bmatrix}}_L \underbrace{\begin{bmatrix} O & a & \mathbf{w} \\ O & \mathbf{0} & U' \end{bmatrix}}_U.$$

L is unit lower triangular, because L' is. In U , row 1 has its leading entry a in column c ; the other rows are zero in columns $1, \dots, c$, followed by the rows of U' , so their leading entries lie to the right of column c , step to the right, and the zero rows are at the bottom, because U' is in row echelon form. So U is in row echelon form, and $PA = LU$. $\square$

The vector $P'\mathbf{1}$ in the proof shows what happens to the multipliers when a later stage swaps two rows: they are swapped too. Once P is known, Theorem 3.7.4 applied to PA gives the same kind of factors by plain elimination, as in the next example.

Example 3.7.12. Let

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 3 \\ 1 & 3 & 2 \end{bmatrix}.$$

Elimination starts with $R_2 \rightarrow R_2 - 2R_1$, giving $(2 - 2, 4 - 4, 3 - 2) = (0, 0, 1)$, and $R_3 \rightarrow R_3 - R_1$, giving $(1 - 1, 3 - 2, 2 - 1) = (0, 1, 1)$. Now the second pivot position, row 2 and column 2, holds a 0, and the entry below it is 1, so elimination would swap rows 2 and 3. We make that swap first:

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad PA = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 2 \\ 2 & 4 & 3 \end{bmatrix}.$$

Eliminating PA needs no swap: $R_2 \rightarrow R_2 - R_1$ gives $(0, 1, 1)$ and $R_3 \rightarrow R_3 - 2R_1$ gives $(0, 0, 1)$, which is already in row echelon form, so $l_{32} = 0$. By Theorem 3.7.4,

$$PA = LU, \quad L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Check: row 2 of LU is $(1, 2, 1) + (0, 1, 1) = (1, 3, 2)$ and row 3 is $2(1, 2, 1) + (0, 0, 1) = (2, 4, 3)$.

Remark 3.7.13 (Solving with $PA = LU$). If $PA = LU$, then $A\mathbf{x} = \mathbf{b}$ holds if and only if $LU\mathbf{x} = P\mathbf{b}$. Indeed, multiplying $A\mathbf{x} = \mathbf{b}$ on the left by P gives $PA\mathbf{x} = P\mathbf{b}$, and multiplying that by P^{-1} (Proposition 3.7.10(c)) gives back $A\mathbf{x} = \mathbf{b}$. So we reorder the entries of $\mathbf{b}$ as P reorders rows (Proposition 3.7.10(a)) and then substitute as in Theorem 3.7.5.

Check yourself

- (a) Find an LU factorisation of $A = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix}$.
- (b) Use it to solve $A\mathbf{x} = (1, 7)$ by forward and back substitution, and check your answer.
- (c) Write down L and U for $\begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 1 \\ 0 & 4 & 5 \end{bmatrix}$ without multiplying any matrices.

Supplement: beyond the lecture

This material is not needed later in the course, but it rounds out the picture.

Uniqueness. For an invertible matrix the factors are unique. Without invertibility they need not be: for $A = O$, every unit lower triangular L gives $A = LO$.

Proposition 3.7.14. Every unit lower triangular $n \times n$ matrix L is invertible, and L^{-1} is unit lower triangular.

Proof. We row reduce $[L \mid I_n]$ by Gauss–Jordan inversion (Theorem 3.5.10), using only replacements $R_i \rightarrow R_i - cR_p$ with $p < i$: in stage $p = 1, \dots, n-1$ we clear the entries below the diagonal in column p . We claim that both blocks stay unit lower triangular. Suppose they are before such a step. Row p of each block has zeros to the right of column p , so the step changes only entries of row i in columns $1, \dots, p$, all of which lie strictly to the left of the diagonal entry (i, i) ; the diagonal and everything above it stay as they were. In the left block, stage p changes columns $1, \dots, p-1$ of no row, because row p is already zero there (those entries were cleared in earlier stages); so the entries cleared earlier stay zero. After the last stage the left block is I_n , and by Theorem 3.5.10(a) the right block, which is unit lower triangular, is L^{-1} . $\square$

Theorem 3.7.15 (Uniqueness of the LU factorisation). Let A be an invertible $n \times n$ matrix. If $A = L_1U_1 = L_2U_2$, where L_1, L_2 are unit lower triangular and U_1, U_2 are in row echelon form, then $L_1 = L_2$ and $U_1 = U_2$.

Proof. By Proposition 3.7.14 and Theorem 3.5.8, $U_1 = L_1^{-1}A$ is invertible. A square matrix in row echelon form is upper triangular (Section 3.4). If U_1 had fewer than n pivots, some column would have no pivot, the system $U_1\mathbf{x} = \mathbf{0}$ would have a free variable and so a nonzero solution, and U_1 would not be invertible (the remark after Example 3.5.3). So U_1 has a pivot in each of its n columns; as the pivots step to the right, the pivot of row i is in column i , and every diagonal entry u_{ii} of U_1 is nonzero. The same holds for U_2 , which in particular is upper triangular.

Let $M = L_2^{-1}L_1$. It is unit lower triangular (Proposition 3.7.14 and Theorem 3.4.11(b)), and $MU_1 = L_2^{-1}A = U_2$. Write $\mathbf{m}_k$ for column k of M . Column j of U_1 is $u_{1j}\mathbf{e}_1 + \dots + u_{jj}\mathbf{e}_j$, so column j of U_2 is $u_{1j}\mathbf{m}_1 + \dots + u_{jj}\mathbf{m}_j$. We show $\mathbf{m}_j = \mathbf{e}_j$ by induction on j . If $\mathbf{m}_k = \mathbf{e}_k$ for all $k < j$, then column j of U_2 is $u_{1j}\mathbf{e}_1 + \dots + u_{j-1,j}\mathbf{e}_{j-1} + u_{jj}\mathbf{m}_j$. Its entries below row j are 0, because U_2 is upper triangular, and the $\mathbf{e}_k$ with $k < j$ are 0 there, so u_{jj} times each entry of $\mathbf{m}_j$ below row j is 0. As $u_{jj} \neq 0$, those entries are 0; and $\mathbf{m}_j$ has 1 in row j and 0 above it, since M is unit lower triangular. So $\mathbf{m}_j = \mathbf{e}_j$. Hence $M = I$, $L_1 = L_2M = L_2$, and $U_1 = L_1^{-1}A = L_2^{-1}A = U_2$. $\square$

The inverse of L . For L from Example 3.7.2, $L = E_1^{-1}E_2^{-1}E_3^{-1}$, so $L^{-1} = E_3E_2E_1$ by Theorem 3.5.8: the matrix with the entry 7 computed there. So L^{-1} , unlike L , does not simply display the multipliers. In practice nobody forms L^{-1} : forward substitution solves $L\mathbf{y} = \mathbf{b}$ with fewer operations.

Answers to Check yourself

Section 3.1

(a) Column rule: $2(1, -1) + 1(0, 3) + (-1)(2, 1) = (2 + 0 - 2, -2 + 3 - 1) = (0, 0)$. Row rule:

$1 \cdot 2 + 0 \cdot 1 + 2 \cdot (-1) = 0$ and $(-1) \cdot 2 + 3 \cdot 1 + 1 \cdot (-1) = 0$. So $A\mathbf{x} = \mathbf{0}$. (b) Yes. $R_2 \rightarrow R_2 + R_1$ gives $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & 3 \end{bmatrix}$, with a pivot in both rows, so the columns span $\mathbb{R}^2$ (Theorem 3.1.9). (c) No. A 2×3 matrix has at most 2 pivots for 3 columns, so some column has no pivot and $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution (Proposition 3.1.13); $\mathbf{x} = (2, 1, -1)$ from (a) is one, so the columns are dependent.

Section 3.2

(a) $T(\mathbf{x}) = A\mathbf{x}$ with $A = \begin{bmatrix} 0 & 1 \\ 1 & -3 \end{bmatrix}$, so $T = T_A$ is linear (Proposition 3.2.2). Its columns are $T(\mathbf{e}_1) = (0, 1)$ and $T(\mathbf{e}_2) = (1, -3)$, so A is the standard matrix. (b) $R_{180^\circ} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$, so $\mathbf{x} \mapsto -\mathbf{x}$: a half turn, the point reflection through $\mathbf{0}$. (c) $T(2(1, 1)) = T(2, 2) = (4, 0)$, but $2T(1, 1) = 2(1, 0) = (2, 0)$. The scaling rule fails, so T is not linear. (d) Projection onto the second axis: $P = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Reflection across it: $Q = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$.

Section 3.3

(a) The columns of AB are $A(1, 0) = (2, 1)$ and $A(3, 1) = (6, 2)$, so $AB = \begin{bmatrix} 2 & 6 \\ 1 & 2 \end{bmatrix}$. (b) $BA = \begin{bmatrix} 1 \cdot 2 + 3 \cdot 1 & 1 \cdot 0 + 3 \cdot (-1) \\ 0 \cdot 2 + 1 \cdot 1 & 0 \cdot 0 + 1 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 5 & -3 \\ 1 & -1 \end{bmatrix} \neq AB$, so A and B do not commute. (c) For example $M = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ or $M = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}$; in both cases $M^2 = O$.

Section 3.4

(a) $A^T = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 0 & 3 \end{bmatrix}$. With columns $\mathbf{a}_1 = (1, 2, 0)$, $\mathbf{a}_2 = (-1, 0, 3)$: $\mathbf{a}_1 \cdot \mathbf{a}_1 = 5$, $\mathbf{a}_1 \cdot \mathbf{a}_2 = -1$, $\mathbf{a}_2 \cdot \mathbf{a}_2 = 10$, so $A^T A = \begin{bmatrix} 5 & -1 \\ -1 & 10 \end{bmatrix}$, symmetric (Proposition 3.4.8). (b) $\text{tr}(A^T A) = 5 + 10 = 15$, and $\text{tr}(AA^T) = \text{tr}(A^T A) = 15$ by Theorem 3.4.16(c). (c) False. $(AB)^T = B^T A^T = BA$, so AB is symmetric exactly when $AB = BA$. For $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$: $AB = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, not symmetric.

Section 3.5

(a) Here $ad - bc = 3 \cdot 2 - 1 \cdot 5 = 1$, so the inverse is $\begin{bmatrix} 2 & -1 \\ -5 & 3 \end{bmatrix}$. Check: $\begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -5 & 3 \end{bmatrix} = \begin{bmatrix} 6 - 5 & -3 + 3 \\ 10 - 10 & -5 + 6 \end{bmatrix} = I_2$. (b) $R_2 \rightarrow R_2 - 2R_1$ gives row 2 = $[0 \ 1 \ 0 \mid -2 \ 1 \ 0]$; then $R_3 \rightarrow R_3 - R_2$ gives row 3 = $[0 \ 0 \ 1 \mid 2 \ -1 \ 1]$. The left block is I_3 , so the inverse is $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & -1 & 1 \end{bmatrix}$ (Theorem 3.5.10(a)). (c) No. $R_3 \rightarrow R_3 - R_1$ gives row 3 = $(0, 1, 1)$, and $R_3 \rightarrow R_3 - R_2$ makes it a zero row, so the reduced row echelon form is not I_3 and the matrix is not invertible (Theorem 3.5.10(b)). With $x_3 = 1$ free, $x_2 = -1$ and $x_1 = -2x_2 - 3x_3 = -1$: indeed $A(-1, -1, 1) = (-1 - 2 + 3, -1 + 1, -1 - 3 + 4) = (0, 0, 0)$.

Section 3.6

(a) $E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$, and $E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{bmatrix}$ is the matrix of the reverse operation $R_2 \rightarrow R_2 - 4R_3$ (Theorem 3.6.6). By Theorem 3.6.3, EE^{-1} is E^{-1} after $R_2 \rightarrow R_2 + 4R_3$: row 2 becomes $(0, 1, -4) + 4(0, 0, 1) = (0, 1, 0)$, so $EE^{-1} = I_3$. (b) $R_2 \rightarrow R_2 - 3R_1$ gives rows $(2, 1)$, $(0, 1)$; $R_1 \rightarrow R_1 - R_2$ gives $(2, 0)$, $(0, 1)$; $R_1 \rightarrow \frac{1}{2}R_1$ gives I_2 . So $E_1 = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix}$, and by Theorem 3.6.8 $A^{-1} = E_3 E_2 E_1 = E_3 \begin{bmatrix} 4 & -1 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -\frac{1}{2} \\ -3 & 1 \end{bmatrix}$ (the 2×2 formula agrees: $ad - bc = 2$), while

$A = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 6 & 4 \end{bmatrix}$. (c) (i) True. If $A\mathbf{u} = A\mathbf{v} = \mathbf{e}_1$ with $\mathbf{u} \neq \mathbf{v}$, then $A(\mathbf{u} - \mathbf{v}) = \mathbf{0}$ with $\mathbf{u} - \mathbf{v} \neq \mathbf{0}$, so statement (3) of Theorem 3.6.10 fails, and so does (8). (ii) False. $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ has $A\mathbf{e}_1 = \mathbf{e}_1$, but its second column is zero, so its columns are dependent and A is not invertible. Statement (7) asks for a solution for *every* $\mathbf{b}$, not just one.

Section 3.7

(a) The multiplier is $l_{21} = 3/1 = 3$, and $R_2 \rightarrow R_2 - 3R_1$ gives row 2 = (0, 2). So $L = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$ (Theorem 3.7.4). Check: $LU = \begin{bmatrix} 1 & 2 \\ 3 & 6+2 \end{bmatrix} = A$. (b) Forward substitution in $L\mathbf{y} = (1, 7)$: $y_1 = 1$, then $3y_1 + y_2 = 7$ gives $y_2 = 4$. Back substitution in $U\mathbf{x} = \mathbf{y}$: $2x_2 = 4$ gives $x_2 = 2$, then $x_1 + 2x_2 = 1$ gives $x_1 = -3$ (Theorem 3.7.5). So $\mathbf{x} = (-3, 2)$. Check: $A\mathbf{x} = (-3 + 4, -9 + 16) = (1, 7)$. (c) $R_2 \rightarrow R_2 - 2R_1$ gives row 2 = (0, 1, 1), and row 3 already has a 0 in column 1, so $l_{31} = 0$. Then $l_{32} = 4/1 = 4$, and $R_3 \rightarrow R_3 - 4R_2$ gives row 3 = (0, 0, 1). Writing each multiplier where it cleared an entry, $L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.