Fall 2019 Math 201-003 at Drexel: blackboard notes of 2019-10-09

Darij Grinberg

December 5, 2019

1. Gaussian elimination (continued)

1.1. Spanning (continued)

1.1.1. Connecting independence and spanning

Recall our method for checking for linear dependence ([Strickland, Method 8.8]):

Theorem 1.1.1. Let $v_1, v_2, \ldots, v_m$ be m vectors in $\mathbb{R}^n$. To see whether they are dependent, the following algorithm works:

(a) Form the $n \times m$-matrix A whose columns are $v_1, v_2, \ldots, v_m$. This is usually written as follows:

$$A = [v_1 \mid v_2 \mid \cdots \mid v_m].$$

(b) Row-reduce A to get an $n \times m$-matrix B in RREF.

(c) If every column of B has a pivot, then $v_1, v_2, \ldots, v_m$ are independent.

(d) If some column of B has no pivot, then $v_1, v_2, \ldots, v_m$ are dependent. Moreover, solutions to the system $B\lambda = 0$ (where $\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix}$) correspond precisely to the relations between $v_1, v_2, \ldots, v_m$.

Recall our method for checking whether a list of vectors spans $\mathbb{R}^n$ ([Strickland, Method 9.7]):

Theorem 1.1.2. Let $\mathcal{V} = (v_1, v_2, \ldots, v_m)$ be a list of vectors in $\mathbb{R}^n$. We can check whether this list spans $\mathbb{R}^n$ as follows:

(a) Form the $m \times n$-matrix

$$C = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_m^T \end{bmatrix}.$$

This is the matrix whose rows are $v_1^T, v_2^T, \ldots, v_m^T$ (the transposes of the vectors in $\mathcal{V}$).

(b) Row-reduce C to get a matrix D in RREF.

(c) If D has a pivot in each column, then $\mathcal{V}$ spans $\mathbb{R}^n$.

(d) If D has no pivot in some column, then $\mathcal{V}$ does not span $\mathbb{R}^n$.

Note that the matrix C in the latter theorem is the transpose of the matrix A in the former theorem. This does not mean that the answers to the "are $v_1, v_2, \ldots, v_m$ dependent?" and "do $v_1, v_2, \ldots, v_m$ span $\mathbb{R}^n$?" questions are directly related. However, it means the following:

Proposition 1.1.3. Let P be any $m \times n$-matrix.

(a) The columns of P are linearly independent (in $\mathbb{R}^m$) if and only if the columns of P^T span $\mathbb{R}^n$.

(b) The columns of P span $\mathbb{R}^m$ if and only if the columns of P^T are linearly independent (in $\mathbb{R}^n$).

Proof. **(a)** Compare the two previous theorems. If $v_1, v_2, \ldots, v_m$ are the columns of P, then $P = [v_1 \mid v_2 \mid \cdots \mid v_m]$ and $P^T = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_m^T \end{bmatrix}$. Therefore, the method for checking whether $v_1, v_2, \ldots, v_m$ are linearly independent does the exact same thing as the method for checking whether the columns of P^T span $\mathbb{R}^n$. This proves **(a)**.

(b) follows from **(a)** by applying **(a)** to P^T instead of P, since $\left(P^T\right)^T = P$. $\square$

1.2. Bases

Definition 1.2.1. A **basis** of $\mathbb{R}^n$ means a list $(v_1, v_2, \ldots, v_m)$ of vectors in $\mathbb{R}^n$ that is both independent and spans $\mathbb{R}^n$.

Example 1.2.2. Let

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \text{and} \qquad v_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

I claim that (v_1, v_2) is a basis of $\mathbb{R}^2$.

Indeed, to see that it is independent, we row-reduce

$$A = [v_1 \mid v_2] = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix};$$

this has two pivots, so (v_1, v_2) is independent.

To see that it spans $\mathbb{R}^2$, we row-reduce

$$C = \begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix};$$

this has two pivots, so (v_1, v_2) spans $\mathbb{R}^2$.

Theorem 1.2.3. Let $v_1, v_2, \ldots, v_m$ be column vectors in $\mathbb{R}^n$.

(a) The list $(v_1, v_2, \ldots, v_m)$ spans $\mathbb{R}^n$ if and only if each $w \in \mathbb{R}^n$ can be written as $w = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m$ for **at least one** m-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbb{R}^m$.

(b) The list $(v_1, v_2, \ldots, v_m)$ is linearly independent if and only if each $w \in \mathbb{R}^n$ can be written as $w = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m$ for **at most one** m-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbb{R}^m$.

(c) The list $(v_1, v_2, \ldots, v_m)$ is a basis of $\mathbb{R}^n$ if and only if each $w \in \mathbb{R}^n$ can be written as $w = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m$ for **exactly one** m-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbb{R}^m$.

In the latter case, these $\lambda_1, \lambda_2, \ldots, \lambda_m$ are called the **coordinates** of w with respect to the basis $(v_1, v_2, \ldots, v_m)$.

Proof of Theorem 1.2.3. **(a)** We have the following equivalence:

$$\begin{aligned}
&(\text{the list } (v_1, v_2, \ldots, v_m) \text{ spans } \mathbb{R}^n) \\
\iff &(\mathbb{R}^n = \operatorname{span}(v_1, v_2, \ldots, v_m)) \\
\iff &(\mathbb{R}^n \subseteq \operatorname{span}(v_1, v_2, \ldots, v_m)) \\
&\qquad (\text{since } \operatorname{span}(v_1, v_2, \ldots, v_m) \subseteq \mathbb{R}^n \text{ is true either way}) \\
\iff &(\text{each } w \in \mathbb{R}^n \text{ lies in } \operatorname{span}(v_1, v_2, \ldots, v_m)) \\
\iff &(\text{each } w \in \mathbb{R}^n \text{ is a combination of } v_1, v_2, \ldots, v_m) \\
\iff &\begin{pmatrix} \text{each } w \in \mathbb{R}^n \text{ can be written as } \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m \\ \text{for at least one choice of } \lambda_1, \lambda_2, \ldots, \lambda_m \end{pmatrix}.
\end{aligned}$$

(b) $\Longrightarrow$: Assume that $(v_1, v_2, \ldots, v_m)$ is linearly independent.

We must prove that each $w \in \mathbb{R}^n$ can be written as $w = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m$ for **at most one** m-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbb{R}^m$.

So let $w \in \mathbb{R}^n$. Let $(\alpha_1, \alpha_2, \ldots, \alpha_m)$ and $(\beta_1, \beta_2, \ldots, \beta_m)$ be two m-tuples of real numbers such that

$$w = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m \qquad \text{and} \qquad w = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_m v_m.$$

We must prove that $(\alpha_1, \alpha_2, \ldots, \alpha_m) = (\beta_1, \beta_2, \ldots, \beta_m)$; in other words, we must prove that $\alpha_i = \beta_i$ for each i.

Subtracting the two equalities

$$w = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m \qquad \text{and} \qquad w = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_m v_m,$$

we obtain

$$\begin{aligned} 0 &= (\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_m v_m) - (\beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_m v_m) \\ &= (\alpha_1 v_1 - \beta_1 v_1) + (\alpha_2 v_2 - \beta_2 v_2) + \cdots + (\alpha_m v_m - \beta_m v_m) \\ &= (\alpha_1 - \beta_1)\, v_1 + (\alpha_2 - \beta_2)\, v_2 + \cdots + (\alpha_m - \beta_m)\, v_m. \end{aligned}$$

This is a relation between the $v_1, v_2, \ldots, v_m$, and thus must be the trivial relation, because we assumed that the $v_1, v_2, \ldots, v_m$ are independent (so each relation between them is the trivial relation). This means that

$$\alpha_1 - \beta_1 = 0, \qquad \alpha_2 - \beta_2 = 0, \qquad \ldots, \qquad \alpha_m - \beta_m = 0.$$

In other words,

$$\alpha_1 = \beta_1, \qquad \alpha_2 = \beta_2, \qquad \ldots, \qquad \alpha_m = \beta_m.$$

In other words, $(\alpha_1, \alpha_2, \ldots, \alpha_m) = (\beta_1, \beta_2, \ldots, \beta_m)$. This proves that there is **at most one** m-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbb{R}^m$ such that $w = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m$.

$\Longleftarrow$: Assume that each $w \in \mathbb{R}^n$ can be written as $w = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m$ for **at most one** m-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbb{R}^m$. We must prove that $(v_1, v_2, \ldots, v_m)$ is linearly independent.

Indeed, let $\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_m v_m = 0$ be a relation between $v_1, v_2, \ldots, v_m$. Then, the zero vector 0 can be written as $0 = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m$ for **two** m-tuples $(\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbb{R}^m$: namely, for

$$(\lambda_1, \lambda_2, \ldots, \lambda_m) = (\mu_1, \mu_2, \ldots, \mu_m) \qquad (\text{since } 0 = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_m v_m),$$

and also for

$$(\lambda_1, \lambda_2, \ldots, \lambda_m) = (0, 0, \ldots, 0) \qquad (\text{because } 0 = 0 v_1 + 0 v_2 + \cdots + 0 v_m).$$

But by our assumption, there is **at most one** such m-tuple. Thus, the two m-tuples

$$(\mu_1, \mu_2, \ldots, \mu_m) \qquad \text{and} \qquad (0, 0, \ldots, 0)$$

must be identical. In other words, $(\mu_1, \mu_2, \ldots, \mu_m) = (0, 0, \ldots, 0)$. In other words, $\mu_i = 0$ for all i. This means that our relation $\mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_m v_m = 0$ must be trivial. Hence, we have shown that any relation between $v_1, v_2, \ldots, v_m$ is trivial. In other words, $(v_1, v_2, \ldots, v_m)$ is linearly independent.

(c) The claim of **(c)** is just obtained by combining **(a)** with **(b)**, because

$$\text{“basis”} = \text{“span”} \wedge \text{“linearly independent”}$$

and

$$\text{“exactly one”} = \text{“at least one”} \wedge \text{“at most one”}.$$

□

Example 1.2.4. Let

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Then, (v_1, v_2, v_3) is a basis of $\mathbb{R}^3$.

Indeed, to see that it is independent, we have to argue that

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0 \qquad \Longrightarrow \qquad \lambda_1 = \lambda_2 = \lambda_3 = 0.$$

But this is easy:

$$\begin{aligned} & (\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0) \\ \Longrightarrow\ & \left(\lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right) \\ \Longrightarrow\ & \left(\begin{pmatrix} \lambda_1 + \lambda_2 + \lambda_3 \\ \lambda_2 + \lambda_3 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right) \\ \Longrightarrow\ & (\lambda_1 + \lambda_2 + \lambda_3 = 0 \text{ and } \lambda_2 + \lambda_3 = 0 \text{ and } \lambda_3 = 0) \\ \Longrightarrow\ & (\lambda_1 = \lambda_2 = \lambda_3 = 0). \end{aligned}$$

As to why the vectors v_1, v_2, v_3 span $\mathbb{R}^3$: We can write each vector $w = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \in \mathbb{R}^3$ as a combination of v_1, v_2, v_3, namely as

$$w = (a_1 - a_2)\, v_1 + (a_2 - a_3)\, v_2 + a_3 v_3,$$

since

$$\begin{aligned} & (a_1 - a_2)\, v_1 + (a_2 - a_3)\, v_2 + a_3 v_3 \\ &= (a_1 - a_2) \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (a_2 - a_3) \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + a_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} (a_1 - a_2) + (a_2 - a_3) + a_3 \\ (a_2 - a_3) + a_3 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = w. \end{aligned}$$

Let us restate Theorem 1.2.3 as follows:

Theorem 1.2.5. Let $v_1, v_2, \ldots, v_m$ be column vectors in $\mathbb{R}^n$. Let

$$A = [v_1 \mid v_2 \mid \cdots \mid v_m] \qquad \text{(an } n \times m\text{-matrix)}.$$

(a) The list $(v_1, v_2, \ldots, v_m)$ spans $\mathbb{R}^n$ if and only if for each $w \in \mathbb{R}^n$, the system $A\lambda = w$ has **at least one** solution $\lambda \in \mathbb{R}^m$.

(b) The list $(v_1, v_2, \ldots, v_m)$ is linearly independent if and only if for each $w \in \mathbb{R}^n$, the system $A\lambda = w$ has **at most one** solution $\lambda \in \mathbb{R}^m$.

(c) The list $(v_1, v_2, \ldots, v_m)$ is a basis of $\mathbb{R}^n$ if and only if for each $w \in \mathbb{R}^n$, the system $A\lambda = w$ has **exactly one** solution $\lambda \in \mathbb{R}^m$.

Proof of Theorem 1.2.5. If we write the vector $\lambda \in \mathbb{R}^m$ as $\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix}$, then the first lemma from last time says

$$\underbrace{[v_1 \mid v_2 \mid \cdots \mid v_m]}_{=A} \underbrace{\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix}}_{=\lambda} = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m,$$

that is,

$$A\lambda = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m.$$

Thus, a solution to $A\lambda = w$ is the same as a column vector $\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} \in \mathbb{R}^m$ satisfying

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_m v_m = w.$$

Thus, Theorem 1.2.5 follows from Theorem 1.2.3. □

This theorem shows that spanning and independence of vectors are directly related to the kind of solutions that a linear system can have.

The next proposition is [Strickland, Proposition 10.11]:

Proposition 1.2.6. Let A be an $n \times n$-matrix. Then, the columns of A form a basis of $\mathbb{R}^n$ if and only if the columns of A^T form a basis of $\mathbb{R}^n$.

Proof. This follows from Proposition 1.1.3. □

The next proposition is [Strickland, Proposition 10.12]:

Proposition 1.2.7. Let $v_1, v_2, \ldots, v_n$ be n vectors in $\mathbb{R}^n$. (Yes, we need the same n here.)

(a) If $(v_1, v_2, \ldots, v_n)$ is independent, then $(v_1, v_2, \ldots, v_n)$ is a basis of $\mathbb{R}^n$ (and thus, in particular, spans $\mathbb{R}^n$).

(b) If $(v_1, v_2, \ldots, v_n)$ spans $\mathbb{R}^n$, then $(v_1, v_2, \ldots, v_n)$ is a basis of $\mathbb{R}^n$ (and thus, in particular, is independent).

We will prove this in a moment. First, here is an imprecise but memorable way to formulate this proposition:

- **(a)** Sufficiently many independent vectors always form a basis.
- **(b)** Sufficiently few spanning vectors always form a basis.

Here, "sufficiently many" means "n" both times, where the vectors are in $\mathbb{R}^n$.

Example 1.2.8. **(a)** The three vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ are independent and thus (by part **(a)** of this proposition) span $\mathbb{R}^3$.

(b) The three vectors $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ span $\mathbb{R}^3$ and thus (by part **(b)** of this proposition) are independent.

(c) The two vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ in $\mathbb{R}^3$ are independent, but don't form a basis. This does not contradict the proposition, because they are just 2 vectors, not 3 vectors.

Proof of Proposition 1.2.7. Let $A = [v_1 \mid v_2 \mid \cdots \mid v_n]$.

(a) Suppose that $(v_1, v_2, \ldots, v_n)$ is independent. Let B be the RREF of A. By our method for testing independence, we thus see that B has a pivot in each column. Hence, $B = I_n$ (since B is an $n \times n$-matrix). Thus, there is a way to transform the matrix A into I_n by row operations.

We must show that $(v_1, v_2, \ldots, v_n)$ spans $\mathbb{R}^n$. In other words, we must show that the system $A\lambda = w$ has at least one solution for each $w \in \mathbb{R}^n$.

Fix $w \in \mathbb{R}^n$. The augmented matrix of the system $A\lambda = w$ is

$$[v_1 \mid v_2 \mid \cdots \mid v_n \mid w] .$$

Let us now recall the row operations that were used to transform the matrix A into I_n. Applying the same row operations to the augmented matrix $[v_1 \mid v_2 \mid \cdots \mid v_n \mid w]$, we obtain

$$[I_n \mid *]$$

(where the asterisk "$*$" stands for a last column that we don't know), because the first n columns of the augmented matrix $[v_1 \mid v_2 \mid \cdots \mid v_n \mid w]$ are precisely the n columns of A. But this matrix $[I_n \mid *]$ is in RREF (no matter what its last column is), and has a pivot in each of its first n columns, and thus has **no** pivot in its last column. Thus, we conclude that the system $A\lambda = w$ has a solution (according to the Gaussian elimination method). So we have shown that $(v_1, v_2, \ldots, v_n)$ spans $\mathbb{R}^n$. Hence, the list $(v_1, v_2, \ldots, v_n)$ is a basis of $\mathbb{R}^n$ (since this list is independent).

(b) Suppose that $(v_1, v_2, \ldots, v_n)$ spans $\mathbb{R}^n$. In other words, the columns of the matrix A span $\mathbb{R}^n$. By part **(b)** of the first proposition today, this yields that the columns of A^T are linearly independent. Thus, Proposition 1.2.7 **(a)** (applied to these columns) shows that the columns of A^T form a basis of $\mathbb{R}^n$. Thus, Proposition 1.2.6 shows that the columns of A form a basis of $\mathbb{R}^n$. In other words, $v_1, v_2, \ldots, v_n$ form a basis of $\mathbb{R}^n$. □

Corollary 1.2.9. Let $v_1, v_2, \ldots, v_m$ be m vectors in $\mathbb{R}^n$.

(a) If $m < n$, then $(v_1, v_2, \ldots, v_m)$ cannot span $\mathbb{R}^n$.

(b) If $m > n$, then $(v_1, v_2, \ldots, v_m)$ cannot be independent.

(c) If $m \neq n$, then $(v_1, v_2, \ldots, v_m)$ cannot be a basis of $\mathbb{R}^n$.

(d) If $m = n$ and $(v_1, v_2, \ldots, v_m)$ is either independent or spans $\mathbb{R}^n$, then $(v_1, v_2, \ldots, v_m)$ is a basis of $\mathbb{R}^n$.

Proof. **(a)** was proven last time.

(b) was proven last time.

(c) follows by combining **(a)** and **(b)**.

(d) This is Proposition 1.2.7. □

Here is an imprecise "fact": If you pick n vectors in $\mathbb{R}^n$ "at random", it is almost sure that you get a basis of $\mathbb{R}^n$. To **not** get a basis, you need a "coincidence". (This is imprecise because there is no such thing as a "random" real number, thus no "random" vectors either. Formalizing this rigorously would take us too far afield.)

1.3. Elementary matrices

We have defined row operations as operations in which rows get added, scaled or swapped. In a way, this means we have been breaking our matrices apart (into rows, at least) and assembling them back again. Can we achieve the same result through matrix addition, multiplication etc., without breaking them apart? In other words, can we rewrite our row operations in terms of matrices, not in terms of rows?

Yes, but we need some special matrices for this. We follow [Strickland, Definition 11.1] in the following:

Definition 1.3.1. Let $n \in \mathbb{N}$. We define the following $n \times n$-matrices:

(a) If $p \in \{1, 2, \ldots, n\}$ and if $\lambda \in \mathbb{R}$ is nonzero, then $D_p(\lambda)$ shall mean the $n \times n$-matrix $\operatorname{diag}(1, 1, \ldots, 1, \lambda, 1, 1, \ldots, 1)$; this is the matrix whose diagonal entries are $1, 1, \ldots, 1, \lambda, 1, 1, \ldots, 1$ with the λ in position p, and whose off-diagonal entries are 0. Equivalently, $D_p(\lambda)$ is the $n \times n$-matrix which is I_n but with the (p, p)-th entry replaced by λ.

(b) If $p, q \in \{1, 2, \ldots, n\}$ are distinct, and if $\mu \in \mathbb{R}$, then $E_{p,q}(\mu)$ shall mean the $n \times n$-matrix which is I_n but with the (p, q)-th entry replaced by μ.

(c) If $p, q \in \{1, 2, \ldots, n\}$ are distinct, then $F_{p,q}$ shall mean the $n \times n$-matrix which is I_n but with rows p and q swapped.

Example 1.3.2. Let $n = 4$.

(a) We have
$$D_2(\lambda) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(b) We have
$$E_{2,4}(\mu) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \mu \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

(c) We have
$$F_{2,4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

The three kinds of matrices we just defined ($D_p(\lambda)$, $E_{p,q}(\mu)$ and $F_{p,q}$) are called the **elementary matrices**. They can be used to re-encode row operations as follows ([Strickland, Proposition 11.3]):

Proposition 1.3.3. Let A be an $n \times m$-matrix, and let A' be obtained from A by a single row operation. Then, $A' = UA$ for some elementary matrix U. In more detail:

(a) If A' is obtained from A by scaling the p-th row by λ, then $A' = D_p(\lambda) A$.

(b) If A' is obtained from A by adding μ times the q-th row to the p-th row, then $A' = E_{p,q}(\mu) A$.

(c) If A' is obtained from A by swapping the p-th and q-th rows, then $A' = F_{p,q} A$.

Proof. Let $A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix}$.

(a) We have

$$D_2(\lambda) A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ \lambda a_{2,1} & \lambda a_{2,2} & \lambda a_{2,3} & \lambda a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix}.$$

(b) We have

$$\begin{aligned} E_{2,4}(\mu) A &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \mu \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} \\ &= \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} + \mu a_{4,1} & a_{2,2} + \mu a_{4,2} & a_{2,3} + \mu a_{4,3} & a_{2,4} + \mu a_{4,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix}. \end{aligned}$$

(c) We have

$$F_{2,4} A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{pmatrix} = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \end{pmatrix}.$$

□

Remark: My notations $D_p(\lambda)$, $E_{p,q}(\mu)$ and $F_{p,q}$ are from Strickland's [Strickland], except that he does not place commas between the two subscripts (so he calls them $D_p(\lambda)$, $E_{pq}(\mu)$ and F_{pq}). They correspond to S_p^λ, $A_{p,q}^\mu$ and $T_{p,q}$ in [lina].

Remark: So multiplying elementary matrices to a matrix A on the left corresponds to row operations on A. Likewise, multiplying them to A on the right corresponds to column operations on A. (Beware, however: $E_{p,q}(\mu)$ will now add μ times the p-th column to the q-th column.)

Corollary 1.3.4. Let A and B be $n \times m$-matrices such that A can be transformed into B by row operations. Then, $B = UA$ for some matrix U that can be written as a product of elementary matrices.

Proof. Let

$$A \longrightarrow C_1 \longrightarrow C_2 \longrightarrow \cdots \longrightarrow C_s \longrightarrow B$$

be a way to get from A to B by row operations (where each arrow stands for a single row operation). Then, by Proposition 1.3.3, we have

$$C_1 = U_1 A,$$

$$\begin{aligned} C_2 &= U_2 C_1, \\ C_3 &= U_3 C_2, \\ &\vdots \\ C_s &= U_s C_{s-1}, \\ B &= U_{s+1} C_s \end{aligned}$$

for some elementary matrices $U_1, U_2, \ldots, U_{s+1}$. Combining these equalities, we find

$$B = \underbrace{U_{s+1} U_s U_{s-1} \cdots U_1}_{\text{a product of elementary matrices}} A.$$

□

1.4. Invertibility

We follow partly [Strickland, §11], partly [lina, §3.2].

Definition 1.4.1. Let A be an $m \times n$-matrix.

(a) A **left inverse** of A means an $n \times m$-matrix L such that $LA = I_n$.

(b) A **right inverse** of A means an $n \times m$-matrix R such that $AR = I_m$.

(c) An **inverse** (or **two-sided inverse**) of A means an $n \times m$-matrix B such that $BA = I_n$ and $AB = I_m$.

Thus, an inverse of A is the same as a matrix that is simultaneously a left inverse and a right inverse of A.

Example 1.4.2. If $A = \begin{pmatrix} 1 & 3 \end{pmatrix}$, then $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ is a right inverse of A, since

$$\begin{pmatrix} 1 & 3 \end{pmatrix} \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \cdot (-2) + 3 \cdot 1 \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix} = I_1.$$

But $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is also a right inverse of A in this case. So right inverses are not unique.

Nor do they always exist; for example, $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ has none.

Similarly for left inverses.

Note that if L is a left inverse of A, then L^T is a right inverse of A^T, and vice versa; therefore, the properties of right inverses mirror the properties of left inverses.

Theorem 1.4.3. Let A be a matrix. Let L be a left inverse of A. Let R be a right inverse of A. Then, $L = R$, and furthermore this matrix $L = R$ is an inverse of A.

Proof. Comparing

$$\underbrace{LA}_{=I_n} R = I_n R = R \qquad \text{and} \qquad L \underbrace{AR}_{=I_m} = L I_m = L,$$

we get $L = R$. Now, $LA = I_n$ and $A \underbrace{L}_{=R} = AR = I_m$, so we conclude that L is an inverse of A. □

Corollary 1.4.4. Assume that a matrix A has both a left inverse and a right inverse. Then:

(a) The matrix A has a **unique** inverse.

(b) This inverse is the **only** left inverse of A and the **only** right inverse of A.

Proof. We have assumed that A has a left inverse and a right inverse. Call them L and R. Then, Theorem 1.4.3 shows that $L = R$, and that furthermore this matrix $L = R$ is an inverse of A. Now, if L' is any other left inverse of A, then Theorem 1.4.3 (applied to L' instead of L) also shows that $L' = R$, whence $L' = R = L$. This shows that the inverse $L = R$ is the **only** left inverse of A. Similarly, it is the **only** right inverse of A. This proves part **(b)**. From this, the uniqueness part of part **(a)** follows (since any other inverse would also be a left inverse). □

Corollary 1.4.5. If A has an inverse, then this inverse is unique.

Proof. An inverse is both a left inverse and a right inverse. Thus, this corollary follows from Corollary 1.4.4 **(a)**. □

Definition 1.4.6. If a matrix A has an inverse, then this inverse is denoted by A^{-1}. (This notation is harmless, because Corollary 1.4.5 says that this inverse is unique.)

A matrix is said to be **invertible** if it has an inverse.

Proposition 1.4.7. Let A and B be two invertible matrices. Then, their product AB is also invertible, and its inverse is $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. We have to check that $B^{-1}A^{-1}$ is an inverse of AB. In other words, we have to check that $\left(B^{-1}A^{-1}\right)(AB) = I$ and $(AB)\left(B^{-1}A^{-1}\right) = I$ (where I stands for the identity matrix of the appropriate size). But this follows from

$$\left(B^{-1}A^{-1}\right)(AB) = B^{-1} \underbrace{A^{-1}A}_{=I} B = B^{-1} \underbrace{IB}_{=B} = B^{-1}B = I \qquad \text{and}$$

$$(AB)\left(B^{-1}A^{-1}\right) = A \underbrace{BB^{-1}}_{=I} A^{-1} = \underbrace{AI}_{=A} A^{-1} = AA^{-1} = I.$$

□

References

[lina] Darij Grinberg, *Notes on linear algebra*, version of 13 December 2016. https://github.com/darijgr/lina

[Strickland] Neil Strickland, *MAS201 Linear Mathematics for Applications*, lecture notes, 28 September 2013. http://neil-strickland.staff.shef.ac.uk/courses/MAS201/