Gaussian elimination

October 2, 2019

Contents

1 Introduction

The point of 18.700 is to understand vectors, vector spaces, and linear transformations. The text provides an enormous amount of powerful abstract information about these things. Sometimes it's helpful to be able to *compute* with these things, and matrices are a powerful tool for doing that. These notes concern the most fundamental and elementary matrix computation: solving systems of linear equations. The ideas should be familiar to you already; one reason to talk about them here is to connect those elementary computational ideas to the more theoretical ones introduced in the text. Another reason is that many people (well, many people at MIT) use some jargon about solving simultaneous equations (*pivots* and *row-echelon form*, for example) and you should know that language.

2 Some definitions and examples

Always F is a field; you can think of $F = \mathbb{Q}$, the field of rational numbers, which is where I'll put most of the examples, but any field will do. The vector spaces we will look at are F^n, the n-tuples of elements of F, for n a nonnegative integer. Almost always it will be most convenient to think of these as *columns* rather than rows:

$$F^n = \left\{ v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \mid v_j \in F \right\}. \tag{2.1}$$

An $m \times n$ *matrix* is a rectangular array of elements of F with m rows and n columns. The book uses the notation $F^{m,n}$ for the set of all $m \times n$ matrices, but in these notes I'll prefer $M_{m\times n}(F)$:

$$M_{m\times n}(F) = \left\{ A = \begin{pmatrix} a_{1\,1} & a_{1\,2} & \cdots & a_{1\,n} \\ a_{2\,1} & a_{2\,2} & \cdots & a_{2\,n} \\ \vdots & & \vdots & \vdots \\ a_{m\,1} & a_{m\,2} & \cdots & a_{m\,n} \end{pmatrix} \mid a_{i\,j} \in F \right\}. \tag{2.2}$$

(Getting used to several different notations for the same thing is good for you!) Recall from Chapter 3 of the text that such a matrix A defines a linear map (which I'll also call A) from F^n to F^m, by the formula

$$A(v) = w, \qquad w_i = \sum_{j=1}^{n} a_{i\,j} v_j. \tag{2.3}$$

A little more explicitly, this is

$$A \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{1\,1}v_1 + \cdots a_{1\,n}v_n \\ \vdots \\ a_{m\,1}v_1 + \cdots a_{m\,n}v_n \end{pmatrix} = v_1 \cdot \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \cdots + v_n \cdot \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \in F^m. \tag{2.4}$$

The matrix A is exactly what we need to talk about simultaneous linear equations. A *system of m linear equations in n unknowns* is

$$\begin{aligned} a_{1\,1}x_1 + a_{1\,2}x_2 + \cdots + a_{1\,n}x_n &= b_1 \\ a_{2\,1}x_1 + a_{2\,2}x_2 + \cdots + a_{2\,n}x_n &= b_2 \\ &\vdots \\ a_{m\,1}x_1 + a_{m\,2}x_2 + \cdots + a_{m\,n}x_n &= b_m \end{aligned} \tag{2.5a}$$

The idea is that we are given the matrix $A \in M_{m\times n}$ (for instance some kind of model of a business operation), and the vector $b \in F^m$ (some kind of desired set of outcomes), and we wish to solve for an unknown vector $x \in F^n$ (the input conditions we should put into the model to get the desired outcomes).

If we take advantage of matrix notation, the large and unwieldy equations in (2.5a) can be written

$$Ax = b \qquad (x \in F^n, b \in F^m). \tag{2.5b}$$

In terms of abstract linear algebra, (2.5b) can be phrased like this:

$$\text{Given } A \in \mathcal{L}(V, W) \text{ and } b \in W, \text{ find } x \in V \text{ so that } Ax = b. \tag{2.5c}$$

Here are some definitions for matrices related to the big ideas about linear transformations (null space and range).

Definition 2.6. The *null space* of an $m \times n$ matrix A is

$$\mathrm{Null}(A) = \{v \in F^n \mid Av = 0\} \subset F^n.$$

The *range* of A is

$$\mathrm{Range}(A) = \{Av \in F^m \mid v \in F^n\} \subset F^m.$$

We're going to show that the range is exactly equal to the *column space of* A, defined in a moment. Write

$$c_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \in F^m \tag{2.6a}$$

for the jth column of A. (The vector c_j is called $A_{\cdot,j}$ on page 76 of the text, but I don't like that notation much.) Then (2.4) says that

$$Av = v_1 \cdot c_1 + \cdots + v_n \cdot c_n : \tag{2.6b}$$

Av is a linear combination of the n columns of A, with coefficients the entries of v. It follows easily that

$$\mathrm{Col}(A) = \mathrm{span}(c_1, \ldots, c_n) = \mathrm{Range}(A) \subset F^m. \tag{2.6c}$$

The *column rank of A,* c-rank(A), is the dimension of the column space; equivalently, the dimension of the range of A:

$$\text{c-rank}(A) = \dim \mathrm{Col}(A) = \dim \mathrm{Range}(A). \tag{2.6d}$$

Although we don't have a linear algebra interpretation yet, it's natural to define a vector $r_i \in F^n$ to be the ith row of A (called $A_{i,\cdot}$ in the text), and then

$$\mathrm{Row}(A) = \mathrm{span}(r_1, \ldots, r_m) \subset F^n.$$

Notice that here F^n means *row* vectors of length n, rather than the column vectors we've usually been considering. The *row rank of A,* r-rank(A) is the dimension of the row space:

$$\text{r-rank}(A) = \dim \mathrm{Row}(A). \tag{2.6e}$$

Because this isn't a serial drama in which I have to keep you in suspense until next week, I'll spoil the story and tell you all the things we're going to find out about these ranks and nullities.

Proposition 2.7. *Suppose A is an $m \times n$ matrix.*

1. *The row rank and the column rank of A are equal, and equal to the dimension of the range of A:*

$$\text{r-rank}(A) = \text{c-rank}(A) = \dim \mathrm{Range}(A).$$

Their common value is called the rank of A, *and written* rank(A).

2. *The dimension of the null space of A plus the rank of A is equal to n.*

The proof will appear in Section 5.

You might think it strange that after the first result is a model of equal treatment for rows and columns, the second shows a sudden preference for n. Shouldn't our null space be called a "column null space" (because it is contained in F^n, and there are n columns); and shouldn't there be a "row null space" contained in F^m, with the property that

$$\dim(\text{row null space of } A) + \text{r-rank}(A) = m? \tag{2.8}$$

Sounds reasonable to me. Can you find such a definition?

I'll now pause to state some easy facts that are useful in their own right, and which can be taken as inspiration for the method of Gaussian elimination.

Proposition 2.9. *Suppose A is an $m \times n$ matrix, with rows $r_1, \ldots, r_m \in F^n$. Suppose B is a $p \times m$ matrix.*

1. *Each row of BA is a linear combination of the rows of A. More precisely, the ith row of BA is the linear combination with coefficients given by the ith row of B:*

$$\sum_{j=1}^{m} b_{ij} r_j.$$

2. *The row space of BA is a subspace of the row space of A:*

$$\text{Row}(BA) \subset \text{Row}(A) \subset F^n.$$

3. *Each $1 \times n$ row vector r_j may be regarded as a linear map*

$$r_j \colon F^n \to F, \qquad r_j(v) = r_j v$$

from column vectors to F, by matrix multiplication. With this notation,

$$Av = \begin{pmatrix} r_1(v) \\ r_2(v) \\ \vdots \\ r_m(v) \end{pmatrix} \in F^m.$$

4. *The null space of A is*

$$\begin{aligned} \text{Null}(A) &= \{v \in F^n \mid r_j(v) = 0 \quad (j = 1, \ldots, m)\} \\ &= \{v \in F^n \mid r(v) = 0 \quad (r \in \text{Row}(A))\}. \end{aligned}$$

There are parallel results about right multiplication of A by C and the column space of A.

Sketch of proof. The formula in (1) is just the definition of matrix multiplication. Then (2) follows. The formula in (3) is again the definition of matrix multiplication, and then (4) follows. $\square$

This proposition shows that the row space of A (consisting of row vectors of size n) can be interpreted as equations defining the null space of A (consisting of column vectors of size n).

We now return to our march toward solving systems of equations. The next definition singles out some special matrices corresponding to systems of equations that are easy to solve. The strategy of Gaussian elimination is to transform any system of equations into one of these special ones.

Definition 2.10. An $m \times n$ matrix A is said to be in *row-echelon form* if the nonzero entries are restricted to an inverted staircase shape. (The terminology comes from a French military description of troop arrangements; the word originally meant "rung of a ladder," and is descended from the Latin "scala," meaning ladder or stairs.) More precisely, we require

1. the first nonzero entry in each row is strictly to the *right* of the first nonzero entry in each earlier row; and

2. any rows consisting entirely of zeros must follow any nonzero rows.

The second requirement may be thought of as a special case of the first, if the "first nonzero entry" of a zero row is defined to be in position $+\infty$, and one says that $+\infty > +\infty > j$ for any finite position j. The *pivots* of a row-echelon matrix are the (finite) positions $(i, j(i))$ of the first nonzero entries of the nonzero rows $i = 1, \cdots, r$, with $r \leq m$ the number of nonzero rows. Here is a row-echelon matrix, with the three pivots at $(1, 2)$, $(2, 4)$, and $(3, 5)$ shown in bold:

$$\begin{pmatrix} 0 & \mathbf{-2} & 3 & 1 & 0 & 1 \\ 0 & 0 & 0 & \mathbf{3/2} & -4/3 & 17 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 11 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

The row-echelon matrix A is said to be in *reduced row-echelon form* if in addition

1. each pivot entry is equal to 1, and

2. all the other entries in the column of a pivot are equal to zero.

The example above is *not* in reduced row-echelon form, because the pivots -2 and $3/2$ are not equal to 1, and because of the two nonzero entries above the pivots $3/2$ and 1. A reduced example is

$$\begin{pmatrix} 0 & \mathbf{1} & -3/2 & 0 & 0 & 181/18 \\ 0 & 0 & 0 & \mathbf{1} & 0 & 190/3 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 11 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Suppose that the row-echelon matrix A has pivots in the first r rows, in columns

$$1 \leq j(1) < j(2) < \cdots < j(r) \leq n.$$

We call x_1, x_2,..., x_n the *variables*, having in mind a system of equations like (2.5a). The r variables $x_{j(i)}$ corresponding to the pivot columns are called *pivot variables*. The remaining $n-r$ variables are called *free variables*.

Proposition 2.11. *Suppose that A is in reduced row-echelon form, with r pivots in the entries $\{(i, j(i)) \mid 1 \le i \le r\}$.*

1. *The first r standard basis vectors $(f_1, \ldots, f_r)$ of F^m are a basis of* $\mathrm{Range}(A)$*. This is the column space of A, so* $\text{c-rank}(A) = r$
2. *The (first) r nonzero rows are a basis of the row space of A, so* $\text{r-rank}(A) = r$.
3. *For each free variable x_j, there is a vector in the null space*
$$n_j = e_j - \sum_{i=1}^{r} a_{i\,j} e_{j(i)};$$
the $n-r$ vectors n_j, with x_j a free variable, are a basis of $\mathrm{Null}(A)$*.*
4. *The equation $Ax = b$ (see* (2.5)*) has a solution if and only if $b_i = 0$ for all $i > r$. In that case, one solution is*
$$x_{j(i)} = b_i \quad (1 \le i \le r), \qquad x_j = 0 \quad (x_j \textit{ free variable}).$$
5. *Still assuming that $b_i = 0$ for all $i > r$, the most general solution of $Ax = b$ has arbitrary values x_j for the $n-r$ free variables, and*
$$x_{j(i)} = b_i - \sum_{j \textit{ free}} a_{i\,j} x_j \qquad (1 \le i \le r).$$
That is, we choose the $n-r$ free variables, and then define the r pivot variables by the equation above.

3 Elementary row operations

Proposition 2.11 provides very complete (and nearly obvious) information about how to solve $Ax = b$ when A is in reduced row-echelon form. The present section gives a theoretical description of what you probably already know how to do in practice: to transform an *arbitrary* system of simultaneous equations into another system

$$(A, b) \quad \rightsquigarrow \quad (C, d) \tag{3.1}$$

with three properties:

1. A and C are matrices of the same size $n \times m$, over the same field F, and b and d are vectors in F^m;

2. the two systems $Ax = b$ and $Cx = d$ have exactly the same solutions; that is, for $x \in F^n$, the equation $Ax = b$ is true if and only if $Cx = d$ is true; and

3. the matrix C is in reduced row-echelon form.

The procedure for doing this is called *Gaussian elimination*: *Gaussian* because it was systematized by Gauss (although the ideas are hundreds or thousands of years older), and *elimination* because the idea is to eliminate some of the variables x_j from some of the equations.

The procedure consists of a series of simple steps called *elementary row operations*, described in Definition 3.2 below. We will show that each elementary row operation changes (A, b) to a new system (A', b') satisfying the first two conditions of (3) above. Then we will explain how to perform a series of elementary row operations (the number depends on A, but the largest possibility is something like $m^2 - m$) at the end of which we get a system in reduced row echelon form. Here is the main definition.

Definition 3.2. Suppose $Ax = b$ is a system of m equations in n unknowns ((2.5)). An *elementary row operation* is one of the four procedures below.

1. *Multiply* the ith equation by a nonzero scalar λ. That is, multiply the ith row of A and the ith entry of b each by λ:

$$(a_{i\,1}, a_{i\,2}, \ldots, a_{i\,n}) \rightsquigarrow (\lambda a_{i\,1}, \lambda a_{i\,2}, \ldots, \lambda a_{i\,n}), \quad b_i \rightsquigarrow \lambda b_i.$$

2. *Add* a multiple μ of the jth equation to a *later* equation i, with $1 \le j < i \le m$. That is

$$\begin{gathered}(a_{i\,1}, a_{i\,2}, \ldots, a_{i\,n}) \rightsquigarrow (a_{i\,1} + \mu a_{j,\,1}, a_{i\,2} + \mu a_{j\,2}, \ldots, a_{i\,n} + \mu a_{j\,n}),\\ b_i \rightsquigarrow b_i + \mu b_j.\end{gathered}$$

3. *Add* a multiple μ of the jth equation to an *earlier* equation i, with $1 \le i < j \le m$. That is

$$\begin{gathered}(a_{i\,1}, a_{i\,2}, \ldots, a_{i\,n}) \rightsquigarrow (a_{i\,1} + \mu a_{j,\,1}, a_{i\,2} + \mu a_{j\,2}, \ldots, a_{i\,n} + \mu a_{j\,n}),\\ b_i \rightsquigarrow b_i + \mu b_j.\end{gathered}$$

4. *Exchange* equations i and j.

$$\begin{aligned}(a_{j\,1}, a_{j\,2}, \ldots, a_{j\,n}) &\rightsquigarrow (a_{i\,1}, a_{i\,2}, \ldots, a_{i\,n}), \quad & b_j &\rightsquigarrow b_i\\ (a_{i\,1}, a_{i\,2}, \ldots, a_{i\,n}) &\rightsquigarrow (a_{j\,1}, a_{j\,2}, \ldots, a_{j\,n}), \quad & b_i &\rightsquigarrow b_j\end{aligned}$$

In order to talk about these operations formally, it is helpful to give them names. We call them

$$\begin{aligned}&M(i; \lambda) && (1 \le i \le m,\ \lambda \in F - \{0\})\\ &L(i, j; \mu) && (1 \le j < i \le m,\ \mu \in F)\\ &U(i, j; \mu) && (1 \le i < j \le m,\ \mu \in F)\\ &E(i, j) && (1 \le j < i \le m).\end{aligned} \tag{3.3a}$$

The letters stand for *multiply*, *lower*, *upper*, and *exchange*. To each elementary row operation we associate an $m \times m$ *elementary row matrix*

$$M(i;\lambda) = \begin{pmatrix} 1 & 0 & & \cdots & & 0 \\ 0 & 1 & & \cdots & & 0 \\ & & \ddots & & & \\ 0 & 0 & \cdots & \lambda & \cdots & 0 \\ & & & & \ddots & \\ 0 & 0 & & & \cdots & 1 \end{pmatrix} \tag{3.3b}$$

with λ appearing in the (i,i) place;

$$L(i,j;\mu) = \begin{pmatrix} 1 & 0 & & \cdots & & & 0 \\ 0 & 1 & & \cdots & & & 0 \\ & & & \ddots & & & \\ 0 & \cdots & \mu & \cdots & 1 & \cdots & 0 \\ & & & & & \ddots & \\ 0 & 0 & & \cdots & & & 1 \end{pmatrix} \tag{3.3c}$$

with μ appearing in the (i,j) position $(i > j)$;

$$U(i,j;\mu) = \begin{pmatrix} 1 & 0 & & & \cdots & & & 0 \\ 0 & 1 & & & \cdots & & & 0 \\ & & \ddots & & & & & \\ 0 & 0 & \cdots & 1 & \cdots & \mu & \cdots & 0 \\ & & & & & \ddots & & \\ 0 & 0 & & & \cdots & & & 1 \end{pmatrix} \tag{3.3d}$$

with μ appearing in the (i,j) position $(i < j)$; and

$$E(i,j) = \begin{pmatrix} 1 & 0 & & & \cdots & & & 0 \\ 0 & 1 & & & \cdots & & & 0 \\ & & \ddots & & & & & \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ & & & & \ddots & & & \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ & & & & & & \ddots & \\ 0 & 0 & & & \cdots & & & 1 \end{pmatrix} \tag{3.3e}$$

with the off-diagonal ones appearing in positions (i,j) and (j,i) $(i < j)$.

Proposition 3.4. *Suppose that we are give a system of m simultaneous linear equations in n unknowns $Ax = b$ ((2.5)).*

1. *Performing an elementary row operation (Definition 3.2) is the same as multiplying A and b on the left by the corresponding elementary row matrix ((3.3)).*

2. *Multiplying A and b on the left by any $p \times m$ matrix C can only enlarge the set of solutions. That is, any solution x of $Ax = b$ is also a solution of $(CA)x = Cb$.*

3. *The elementary row matrices are all invertible. Explicitly,*

$$M(i;\lambda)^{-1} = M(i;\lambda^{-1}); \qquad L(i,j;\mu)^{-1} = L(i,j,-\mu);$$
$$U(i,j;\mu)^{-1} = U(i,j,-\mu); \qquad E(i,j)^{-1} = E(i,j).$$

4. *Elementary row operations do not change the solutions of $Ax = b$.*

Consequently any finite sequence of elementary row operations amounts to left multiplication of A and b by an invertible $m \times m$ matrix L, and does not change the set of solutions.

Sketch of proof. The assertion in (1) is best understood by looking at the definition of matrix multiplication, and trying some examples. One can make a formal proof by writing a formula for the entries of the elementary row matrices, like

$$M(i;\lambda)_{r\,s} = \begin{cases} \lambda & r = s = i \\ 1 & r = s \neq i \\ 0 & r \neq s, \end{cases}$$

and then plugging this formula into the definition of $M(i;\lambda)A$.

The second statement in (2) is obvious (by applying C to the equation $Ax = b$, and using the associative law for matrix multiplication). The first statement in (2) follows from the second.

The elementary row operations as described in Definition 3.2 are obviously reversible, and in each case the inverse is another elementary row operation of the same kind. For example, to reverse the operation of adding μ times the jth row to the ith row, we simply add $-\mu$ times the jth row to the ith row. Because of (1), it follows that

$$L(i,j,-\mu)L(i,j;\mu) = I_m.$$

This proves that $L(i,j;\mu)^{-1} = L(i,j,-\mu)$, and the other assertions are similar.

For (4), part (2) says that an elementary row operation L can only increase the set of solutions. So by (3),

$$\begin{aligned} (\text{solutions of } Ax = b) &\subset (\text{solutions of } LAx = Lb) \\ &\subset (\text{solutions of } L^{-1}LAx = L^{-1}Lb) \\ &= (\text{solutions of } Ax = b). \end{aligned}$$

So the containments must be equalities. □

4 Gaussian elimination

We now know some elementary things to do to a system of simultaneous equations that don't change the solutions; and we know everything about solving systems that are in reduced row-echelon form. All that remains is to see that we can do those elementary things and put any system in reduced row-echelon form. This is pretty easy; the system of rules for doing it is *Gaussian elimination*. Here are the details. I'm going to do it in three parts. (This arrangement is slightly different from most of the written versions that you'll see; it's chosen to get some nice theoretical facts as consequences.)

The first part of the algorithm finds (in succession) r special entries

$$\begin{aligned}&(i(1), j(1)),\ (i(2), j(2)),\ \ldots,\ (i(r), j(r)),\\ &1 \le j(1) < j(2) < \cdots < j(r) \le n,\\ &1 \le i(p) \le m \text{ all distinct}\end{aligned} \tag{4.1a}$$

These entries will become the pivots in the row-echelon form (Definition 2.10). After we perform the row operations in the first part of the algorithm, we will have a matrix with the following properties (which are in the direction of the requirements of row echelon form):

$$\text{the first entry of row } i(p) \text{ is a 1, in column } j(p); \tag{4.1b}$$

$$\begin{gathered}\text{entries in column } j(p) \text{ above row } i(p),\\ \text{except in rows } i(q) \text{ with } q < p, \text{ are zero; and}\end{gathered} \tag{4.1c}$$

$$\text{entries in column } j(p) \text{ below row } i(p) \text{ are zero.} \tag{4.1d}$$

We will find these entries and arrange for the these vanishing conditions one row at a time. We know we are finished when we finally have

$$\text{all entries of } A \text{ outside rows } i(1), \ldots, i(r) \text{ are zero.} \tag{4.1e}$$

A theoretically important fact about this part of Gaussian elimination is

$$\begin{gathered}\text{the row operations to achieve}\\ \text{(4.1b)–(4.1e) are of types } M \text{ and } L.\end{gathered} \tag{4.1f}$$

Here is how we accomplish this with a succession of elementary row operations. It is better to look at examples than to write down the formal description. First I'll do a "typical" 3×3 case:

$$\begin{aligned}&\begin{pmatrix}\mathbf{2} & 3 & 4\\ 2 & 2 & 2\\ 1 & 2 & 1\end{pmatrix} \xrightarrow{M(1;1/2)} \begin{pmatrix}\mathbf{1} & 3/2 & 2\\ 2 & 2 & 2\\ 1 & 2 & 1\end{pmatrix}\\ &\qquad \xrightarrow{L(2,1;-2)} \begin{pmatrix}\mathbf{1} & 3/2 & 2\\ 0 & -1 & -2\\ 1 & 2 & 1\end{pmatrix} \xrightarrow{L(3,1;-1)} \begin{pmatrix}\mathbf{1} & 3/2 & 2\\ 0 & -1 & -2\\ 0 & 1/2 & -1\end{pmatrix}\end{aligned} \tag{4.2a}$$

Here I pick the first nonzero entry in the first nonzero column, and marked it in **bold**. This marks the **first** row as the first of our special rows. I then multiply this special row by the inverse of the first entry, to make the first entry **1**. Then I subtract multiples of the first row from other rows to get rid of the other entries in the first column. Onward...

$$\begin{pmatrix} \mathbf{1} & 3/2 & 2 \\ 0 & -1 & -2 \\ 0 & 1/2 & -1 \end{pmatrix} \xrightarrow{M(2;-1)} \begin{pmatrix} \mathbf{1} & 3/2 & 2 \\ 0 & \mathbf{1} & 2 \\ 0 & 1/2 & -1 \end{pmatrix} \xrightarrow{L(3,2;-1/2)} \begin{pmatrix} \mathbf{1} & 3/2 & 2 \\ 0 & \mathbf{1} & 2 \\ 0 & 0 & -2 \end{pmatrix} \tag{4.2b}$$

Here I pick the first column that's nonzero *outside* the first row, then marked in **bold** its first nonzero entry *outside* the first row: now the **second** row is identified as the second of our special rows. Its leading entry is -1, so I multiply the row by its inverse -1. Then I subtract multiples of the second row from later rows to get rid of the later entries in this column. Onward again...

$$\begin{pmatrix} \mathbf{1} & 3/2 & 2 \\ 0 & \mathbf{1} & 2 \\ 0 & 0 & \mathbf{-2} \end{pmatrix} \xrightarrow{M(3;-1/2)} \begin{pmatrix} \mathbf{1} & 3/2 & 2 \\ 0 & \mathbf{1} & 2 \\ 0 & 0 & \mathbf{1} \end{pmatrix} \tag{4.2c}$$

For this last step in this first part, I notice that the **third** column is the first one that's nonzero outside the first two special rows. Its first entry outside the two special rows is the lower right corner entry -2, so that one becomes our third pivot, marked in bold. I multiply that third row by $-1/2$ to make the leading entry 1, and we end up with a matrix satisfying the conditions in (4.6)

Here's a more peculiar example.

$$\begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & \mathbf{3} & 6 & 9 \end{pmatrix} \xrightarrow{M(3;1/3)} \begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & \mathbf{1} & 2 & 3 \end{pmatrix} \tag{4.3a}$$

The first nonzero column is the second one, and its first nonzero entry is in the **third row**; so that row is our first special one, and the leading entry will be our first pivot. We multiply the third row by $-1/3$ to make the leading entry 1, and we're done with the first special row.

$$\begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & \mathbf{2} & 4 \\ 0 & \mathbf{1} & 2 & 3 \end{pmatrix} \xrightarrow{M(2;1/2)} \begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & \mathbf{1} & 2 \\ 0 & \mathbf{1} & 2 & 3 \end{pmatrix} \xrightarrow{L(3,2;-2)} \begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & \mathbf{1} & 2 \\ 0 & \mathbf{1} & 0 & -1 \end{pmatrix} \tag{4.3b}$$

Here the first column that's nonzero outside the special third row is the third column, and its first nonzero entry is in the **second row**; that leading entry will be our second pivot. We multiply the second row by $1/2$ to make the leading entry one, and then we clear the column entries below it.

$$
\begin{pmatrix} 0 & 0 & 0 & \mathbf{2} \\ 0 & 0 & \mathbf{1} & 2 \\ 0 & \mathbf{1} & 0 & -1 \end{pmatrix} \xrightarrow{M(1;1/2)} \begin{pmatrix} 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 2 \\ 0 & \mathbf{1} & 0 & -1 \end{pmatrix}
$$

$$
\xrightarrow{L(2,1;-2)} \begin{pmatrix} 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{1} & 0 & -1 \end{pmatrix} \xrightarrow{L(3,1;1)} \begin{pmatrix} 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{1} & 0 & 0 \end{pmatrix} \tag{4.3c}
$$

This time the new pivot is in fourth column, the entry in the **first** row. We multiply that row by $1/2$ to make the pivot entry equal to one, then clear the column entries below. We end up with a matrix satisfying the conditions in (4.6).

The second part of the algorithm starts with a matrix having r special entries as in (4.1b)–(4.1e), and rearranges the rows so that row $i(1)$ becomes row 1, row $i(2)$ becomes row 2, and so on. At the end of this part, our pivots will be in locations

$$
\begin{gathered} (1, j(1)),\ (2, j(2)),\ \ldots,\ (r, j(r)), \\ 1 \le j(1) < j(2) < \cdots < j(r) \le n. \end{gathered} \tag{4.4a}
$$

The matrix after this part of the algorithm will satisfy the following requirements, which mean in particular that it is in row echelon form (Definition 2.10).

$$
\text{the first entry of row } p \text{ is a 1, in column } j(p) \quad (1 \le p \le r); \tag{4.4b}
$$

$$
\text{entries in column } j(p) \text{ below row } p \text{ are zero.} \tag{4.4c}
$$

$$
\text{all entries of } A \text{ below rows } 1, \ldots, r \text{ are zero.} \tag{4.4d}
$$

The big theoretical fact about this part of Gaussian elimination is

$$
\begin{gathered} \text{the row operations to achieve} \\ \text{(4.4b)–(4.4d) are of type } E. \end{gathered} \tag{4.4e}
$$

The way to carry out this part is almost obvious: we exchange (if they are not already the same) the first special row $i(1)$ with row 1; then the second special row with row 2; and so on through the r special rows. The "typical" case (like (4.2)) has special rows 1 through r in that order, and this step of the algorithm does nothing. Here is what the second step looks like in the example (4.3).

$$
\begin{pmatrix} 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{1} & 0 & 0 \end{pmatrix} \xrightarrow{E(1,3)} \begin{pmatrix} 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 & \mathbf{1} \end{pmatrix} \tag{4.5a}
$$

After we exchange row $i(1) = 3$ with row 1, the first three rows are the special rows in order, and we are done.

The third and last part of the algorithm starts with a matrix having r special entries in row echelon form as in (4.4b)–(4.4d), with pivots in locations

$$
\begin{gathered} (1, j(1)),\ (2, j(2)),\ \ldots,\ (r, j(r)), \\ 1 \le j(1) < j(2) < \cdots < j(r) \le n. \end{gathered} \tag{4.6a}
$$

This part of the algorithm clears the column entries *above* the pivots. The matrix at the end of this last part will satisfy the following requirements, which mean that it is in reduced row echelon form (Definition 2.10).

$$\text{the first entry of row } p \text{ is a 1, in column } j(p) \quad (1 \le p \le r); \tag{4.6b}$$

$$\text{all other entries in column } j(p) \text{ are zero.} \tag{4.6c}$$

$$\text{all entries of } A \text{ below rows } 1, \ldots, r \text{ are zero.} \tag{4.6d}$$

The theoretical fact about this part of Gaussian elimination is

$$\begin{matrix}\text{the row operations to achieve} \\ \text{(4.6b)–(4.6d) are of type } U.\end{matrix} \tag{4.6e}$$

Here is how this looks in the example of (4.2). First we clear the column above the second pivot

$$\begin{pmatrix} \mathbf{1} & 3/2 & 2 \\ 0 & \mathbf{1} & 2 \\ 0 & 0 & \mathbf{1} \end{pmatrix} \xrightarrow{U(1,2;-3/2)} \begin{pmatrix} \mathbf{1} & 0 & -1 \\ 0 & \mathbf{1} & 2 \\ 0 & 0 & \mathbf{1} \end{pmatrix}, \tag{4.7a}$$

then the column above the third pivot

$$\begin{pmatrix} \mathbf{1} & 0 & -1 \\ 0 & \mathbf{1} & 2 \\ 0 & 0 & \mathbf{1} \end{pmatrix} \xrightarrow{U(1,3;1)} \begin{pmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 2 \\ 0 & 0 & \mathbf{1} \end{pmatrix} \xrightarrow{U(2,3;-2)} \begin{pmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} \end{pmatrix}. \tag{4.7b}$$

Here is a theorem summarizing the algorithm described above.

Theorem 4.8. *Suppose A' is an $m \times n$ matrix with entries in a field F. Then we can perform a finite sequence of elementary row operations on A' to obtain a new $m \times n$ matrix A' in reduced row-echelon form. More precisely, we perform*

1. *at most m row operations of type M (multiply a row by a nonzero scalar) interspersed with at most $m(m-1)/2$ operations of type L (add a multiple of a row to a later row); then*
2. *at most $m(m-1)/2$ operations of type E (exchange two rows); then*
3. *at most $m(m-1)/2$ operations of type U (add a multiple of a row to an earlier row).*

Consequently, we can write

$$A' = UELA, \qquad A = L^{-1}E^{-1}U^{-1}A'.$$

Here L and L^{-1} are $m \times m$ invertible lower-triangular matrices; E and E^{-1} are invertible $m \times m$ permutation matrices; and U and U^{-1} are invertible $m \times m$ upper-triangular matrices with ones on the diagonal. The reduced row echelon matrix A is unique (independent of how the row reduction is performed).

The detailed description of step (1) is in (4.1), illustrated in the examples (4.2) and (4.3). The detailed description of step (2) is in (4.4), illustrated in example (4.5). The detailed description of step (3) is in (4.6), illustrated in (4.7). These descriptions can easily be made into a proof of the theorem; all that requires some additional explanation is the uniqueness assertion: that if A_1 and A_2 are reduced row-echelon matrices, and it is possible to pass from A_1 to A_2 by a sequence of elementary row operations, then $A_1 = A_2$. That is not terribly difficult, but I won't explain it.

If all we care about is solving a system of equations, we might as well stop after step (1): the system is then in row-echelon form, except that the equations have been rearranged, so we can solve it by Proposition 2.11. After step (2), the system is in row-echelon form.

5 Rank and row reduction

In this section we'll prove Proposition 2.7, using Gaussian elimination. We begin with some general statements about how row operations affect row and column spaces, null spaces, and ranges.

Proposition 5.1. *Suppose A is an $m \times n$ matrix.*

1. *Elementary row operations do not change the null space* $\mathrm{Null}(A) \subset F^n$*. In particular, they do not change the nullity* $\dim \mathrm{Null}(A)$*.*

2. *Elementary row operations do not change the row space* $\mathrm{Row}(A) \subset F^n$*. In particular, they do not change the row rank* $\text{r-rank}(A)$*.*

3. *Applying a sequence of elementary row operations is equivalent to left multiplication of A by an invertible $m \times m$ matrix L. The effect of this is to apply L to* $\mathrm{Range}(A) \subset F^m$*:*
$$\mathrm{Range}(LA) = \mathrm{Col}(LA) = L(\mathrm{Col}(A)) = L(\mathrm{Range}(A)).$$

4. *Elementary row operations do not change the column rank* $\text{c-rank}(A)$*.*

Proof. Part (1) is a special case of Proposition 3.4(4). For part (2), write the rows of A as
$$(r_1, \ldots, r_m), \qquad r_i \in F^n.$$
The row space of A is equal to the span of these m vectors. The statement that (for example) $L(i,j;\mu)$ does not change the span can be written as
$$\mathrm{span}(r_1, \ldots, r_j, \ldots, r_i, r_m) = \mathrm{span}(r_1, \ldots, r_j, \ldots, r_i + \mu r_j, \ldots, r_m)$$
for any $1 \leq i, j \leq m$ and any $\mu \in F$. This is clear.

The first statement of (3) is Proposition 3.4(1) and (3); and then the second is clear. For (4), because L is invertible, it does not change the dimension of $\mathrm{Col}(A)$. $\square$

Proof of Proposition 2.7. Because of Proposition 5.1, it suffices to prove the proposition after applying a sequence of elementary row operations to A. By Theorem 4.8, we may therefore assume that A is in reduced row echelon form. In that case the equality of row and column ranks follows from Proposition 2.11. $\square$

Having come so far, here is an explicit description of subspaces of F^n.

Theorem 5.2. *Suppose n and r are nonnegative integers. There is a one-to-one correspondence between r-dimensional subspaces $U \subset F^n$, and $r \times n$ matrices A in reduced row-echelon form, with one pivot in each row; that is, with no rows equal to zero. The correspondence sends the matrix A to the span* $\mathrm{Row}(A)$ *of the rows of A. To go in the other direction, suppose U is an r-dimensional subspace of F_n. Choose a basis $(u_1, \ldots, u_r)$ of U, and let A' be the $r \times n$ matrix with rows u_i. Perform Gaussian elimination on A', getting an $r \times n$ matrix A in reduced row echelon form; this is the matrix corresponding to the subspace U.*

Sketch of proof. A matrix A of the desired form clearly has r pivots, and so has rank r (Proposition 2.7). Therefore the row space $\mathrm{Row}(A)$ is indeed an r-dimensional subspace of F^n. Conversely, given an r-dimensional U, the construction in the theorem produces an $r \times n$ matrix A' with $\mathrm{Row}(A') = U$. Now perform Gaussian elimination on A' (Theorem 4.8), obtaining a reduced row echelon matrix A with $\mathrm{Row}(A) = \mathrm{Row}(A') = U$, as desired. □

The theorem says that any subspace has a basis of a very specific form. For example, it says that any two-dimensional subspace of F^3 has as basis the rows of one of the matrices

$$\begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \end{pmatrix}, \quad \begin{pmatrix} 1 & c & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

That is, any two-dimensional subspace of F^3 is either

1. the graph of $z = ax + by$ (some a, b in F); or
2. the graph of $y = cx$ (some c in F); or
3. the y-z plane $x = 0$.

6 Some computational tricks

Although these notes were written to emphasize interesting theoretical consequences of Gaussian elimination, the method *was* designed for solving systems of equations, so I will include a few remarks about that. Suppose A is an $m \times n$ matrix, and $b \in F^m$. Recall from (2.5) the system of m simultaneous equations in n unknowns

$$A(x) = b \qquad (x \in F^n, b \in F^m). \tag{6.1}$$

The "augmented matrix" for this system is the $m \times (n+1)$ matrix

$$\widetilde{A} =_{\mathrm{def}} (A|b) \tag{6.2}$$

Performing Gaussian elimination on the augmented matrix leads to a row-echelon matrix $(A'|b')$, corresponding to an equivalent system of equations $A'(x) = b'$. Here's how this looks in the example of (4.2).

$$\begin{pmatrix} 2 & 3 & 4 \\ 2 & 2 & 2 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \tag{6.3a}$$

Now I'll perform on $\widetilde{A}$ the sequence of row operations that I explained in the previous section for A

$$
\begin{aligned}
\widetilde{A} = \left(\begin{array}{ccc|c} 2 & 3 & 4 & 1 \\ 2 & 2 & 2 & 2 \\ 1 & 2 & 1 & 3 \end{array}\right) &\xrightarrow{M(1;1/2)} \left(\begin{array}{ccc|c} 1 & 3/2 & 2 & 1/2 \\ 2 & 2 & 2 & 2 \\ 1 & 2 & 1 & 3 \end{array}\right) \\
\xrightarrow{L(2,1;-2)} \left(\begin{array}{ccc|c} 1 & 3/2 & 2 & 1/2 \\ 0 & -1 & -2 & 1 \\ 1 & 2 & 1 & 3 \end{array}\right) &\xrightarrow{L(3,1;-1)} \left(\begin{array}{ccc|c} 1 & 3/2 & 2 & 1/2 \\ 0 & -1 & -2 & 1 \\ 0 & 1/2 & -1 & 5/2 \end{array}\right) \\
\xrightarrow{M(2;-1)} \left(\begin{array}{ccc|c} 1 & 3/2 & 2 & 1/2 \\ 0 & 1 & 2 & -1 \\ 0 & 1/2 & -1 & 5/2 \end{array}\right) &\xrightarrow{L(3,2;-1/2)} \left(\begin{array}{ccc|c} 1 & 3/2 & 2 & 1/2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & -2 & 3 \end{array}\right) \\
\xrightarrow{M(3;-1/2)} \left(\begin{array}{ccc|c} \mathbf{1} & 3/2 & 2 & 1/2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -3/2 \end{array}\right) &\xrightarrow{U(1,2;-3/2)} \left(\begin{array}{ccc|c} 1 & 0 & -1 & 2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -3/2 \end{array}\right) \\
\xrightarrow{U(1,3;1)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & -3/2 \end{array}\right) &\xrightarrow{U(2,3;-2)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -3/2 \end{array}\right).
\end{aligned}
\tag{6.3b}
$$

The equivalent system of equations is

$$x_1 = 1/2, \quad x_2 = 2, \quad x_3 = -3/2, \tag{6.3c}$$

which solves itself. You should check that these values really do satisfy the system (6.3a).

There is a similar computational technique to compute a (left) inverse of A. (Such a left inverse exists if and only if the null space of A is zero (see notes on one-sided inverses), which is the same as requiring that $r = n$: there are no free variables, and there is a pivot in every row. The reduced row-echelon form of A must then be

$$A' = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} I_n \\ 0_{m\times n} \end{pmatrix} \tag{6.4}$$

Here is how to compute a left inverse.

Proposition 6.5. *Suppose that A is an $m \times n$ matrix of rank $r = n$ (so that $m \geq n$). Form an augmented matrix $\widetilde{A} = (A|I_m)$ of size $m \times m+n$. Perform Gaussian elimination:*

$$\widetilde{A} = (A|I_m) \xrightarrow{Gauss} (A'|L)$$

with A' the matrix in (6.4) and L the $m \times m$ matrix which is the product of all the elementary row matrices used to reduce A. Write B for the $n \times m$ matrix consisting of the first n rows of L. Then

$$LA = A', \qquad BA = I_n.$$

In particular, B is a left inverse of A.