Chapter 14

Linear Algebra I: Systems of Equations and Matrices

14.1 Motivation and Introduction

Example 14.1. *The* 2×2 *system*

$$\begin{aligned} 2x - y &= 0, \\ -x + 2y &= 3 \end{aligned}$$

has the unique solution $x = 1,\ y = 2$. *Geometrically: two lines in the plane which intersect at the point P*$(1, 2)$.

Example 14.2. *The* 2×3 *system of equations*

$$\begin{aligned} x + 3y + 2z &= 3, \\ 2x + 6y + 4z &= 4 \end{aligned}$$

has no solutions. Geometrically: 2 *parallel but non-intersecting planes in* $\mathbb{R}^3$*: note they have parallel normal vectors* $(2, 6, 4) = 2(1, 3, 2)$.

Example 14.3. *The* 2×3 *system of equations*

$$\begin{aligned} x + y + z &= 3, \\ 2x + 3y + z &= 10 \end{aligned}$$

admits a one-parameter family of solutions $(x, y, z) = (-1, 4, 0) + k(-2, 1, 1),\ k \in \mathbb{R}$. *Geometrically:* 2 *planes in* $\mathbb{R}^3$ *intersect in a line, the system has infinitely many solutions.*

14.1.1 Basic Notation and Definitions; row and column picture

Consider again the linear system of equations

$$\begin{aligned} x + 3y + 2z &= 3, \\ 2x + 6y + 4z &= 4. \end{aligned}$$

We can write it in the form

$$\begin{pmatrix} 1 & 3 & 2 \\ 2 & 6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix},$$

for which we have the shorthand

$$A\underline{\mathbf{x}} = \underline{\mathbf{b}}$$

where

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 6 & 4 \end{pmatrix}, \quad \underline{\mathbf{x}} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad \underline{\mathbf{b}} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

Here, A is a *matrix* and $\underline{\mathbf{x}}, \underline{\mathbf{b}}$ are column vectors. The general system of m linear equations in n variables, as above can be written in the form

$$A\underline{\mathbf{x}} = \underline{\mathbf{b}}, \tag{14.1}$$

where $A = (a_{ij})\quad i = 1, \cdots, m,\, j = 1, \ldots, n$ and $\underline{\mathbf{x}}, \in \mathbb{R}^n$ $\underline{\mathbf{b}} \in \mathbb{R}^m$ are **column vectors**

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ : & : & : & & : \\ : & : & : & & : \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix}, \quad \underline{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}, \quad \underline{\mathbf{b}} = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}.$$

We say A is an $m \times n$ matrix, with m rows and n columns. In the *general* element a_{ij}; i is called the row index, and j the column index. In the previous example, we had 2 equations in 3 unknowns giving a 2×3 matrix. Matrix entries can be any mathematical object, not just numbers: variables, functions and other matrices are common.

Solving the matrix equation $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$, for large systems of equations is one of the central points of this and the next chapter. It has applications in Signal Processing, Communications, Sound/Image/Video processing and transmission, animation, and, yes, also in plain old pure Mathematics.

There are two interesting ways to look at $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$: the row picture and the column picture.

In the **row picture**, we take

$$\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix},$$

and each row represents the equation of a line in $\mathbb{R}^2$.

For picture, see lectures.

The solution of the system is $(1, 2)$, the intersection of the lines.

In the **column picture** we write the system $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ in the form

$$x \begin{pmatrix} 2 \\ -1 \end{pmatrix} + y \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix},$$

so now finding the solution of the system is asking the question:

Which linear combination of the vectors $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ *and* $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ *produces the vector* $\begin{pmatrix} 0 \\ 3 \end{pmatrix}$ *?*

To solve the system, we want that particular linear combination $x\underline{\mathbf{u}} + y\underline{\mathbf{v}} = \underline{\mathbf{b}}$ where $\underline{\mathbf{u}}, \underline{\mathbf{v}}$ are the columns of the matrix A.

For picture, see lectures.

One of the first questions we always ask is: if we can find the right linear combination to produce the vector $(0, 3)$ from the columns of the matrix, can we produce any vector in the plane that way? Do all possible linear combinations of the columns yield the entire 2D space?

Both row and column pictures are equivalent ways of seeing the matrix equation $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$, but the colum picture will turn out to be the more interesting one for us.

Let's look at a 3D example. We saw before that

$$\begin{aligned} 2x \quad &+y + 2z &&= 2 \\ 2x \quad &-y - z &&= 1 \\ x \quad &+y + 2z &&= 1 \end{aligned}$$

has solution $(1, 2, -1)$. In Matrix form, this is

$$\begin{pmatrix} 2 & 1 & 2 \\ 2 & -1 & -1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

In the row picture, each row gives the equation of a plane in $\mathbb{R}^3$, and the solution is the intersection of the three planes.

For picture, see lectures.

In the column picture we write

$$x \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + z \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

and so solving the system is now finding the right linear combination of the three columns of A to produce $x\underline{\mathbf{u}}+y\underline{\mathbf{v}}+z\underline{\mathbf{w}} = \underline{\mathbf{b}}$. We already know which combination, it is the unique solution $(x, y, z) = (1, 2, -1)$

For picture, see lectures.

It's always a good idea to think about how this might fail. Consider the first two columns: we know that the linear combinations of two vectors form a plane $x\underline{\mathbf{u}} + y\underline{\mathbf{v}}$ so if we want to produce a vector $\underline{\mathbf{b}}$ outside that plane, the vector given by the third column $\underline{\mathbf{w}}$ had better not be in the same plane as $\underline{\mathbf{u}}$ and $\underline{\mathbf{w}}$.

In other words, if it happened that the third and first column were equal, $\underline{\mathbf{w}} = \underline{\mathbf{u}}$, or if the third column is some linear combination of the first two, such as $\underline{\mathbf{w}} = 2\underline{\mathbf{u}} - 3\underline{\mathbf{v}}$, then the linear combinations $x\underline{\mathbf{u}} + y\underline{\mathbf{v}} + z\underline{\mathbf{w}}$ would only ever produce vectors in the plane $x\underline{\mathbf{u}} + y\underline{\mathbf{v}}$, so if $\underline{\mathbf{b}}$ is outside that plane, we would have no solutions. And if $\underline{\mathbf{b}}$ is in that plane, it turns out we would have infinitely many solutions.

We will see later that the two cases, unique solution or otherwise, are easy to distinguish. When we have a unique solution, the matrix is **invertible**; when there is no solution or infinitely many solutions, the matrix is not invertible, or **singular.**

It also turns out that in the case of a unique solution, we can use the linear combinations of the columns of A to produce **any** vector $\underline{\mathbf{b}}$. In other words, the linear combinations $x\underline{\mathbf{u}} + y\underline{\mathbf{v}} + z\underline{\mathbf{w}}$ yield the entire 3D space.

14.2 Matrix Algebra

Let $A,\ B$ be two $m \times n$ matrices. The sum/difference of the two matrices $C = A \pm B$ is the $m \times n$ matrix whose components are

$$c_{ij} = a_{ij} \pm b_{ij}, \quad i = 1, \dots m,\ j = 1, \dots n.$$

Similarly, we can define the product of a matrix $A \in \mathbb{R}^{m\times n}$ with a real number $\lambda \in \mathbb{R}$: $C = \lambda A, \quad c_{ij} = \lambda a_{ij}, \quad i = 1, \dots m,\ j = 1, \dots n.$ These definitions are the same as for vector addition and scalar multiplication. We will see that a vector is just a matrix with a single row/column.

Example 14.4.

$$\begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 9 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 1 \\ 2 & 1 & 6 \end{pmatrix} = \begin{pmatrix} 4 & 1 & 1 \\ 4 & 2 & 15 \end{pmatrix}.$$

$$3 \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 9 \end{pmatrix} = \begin{pmatrix} 3 & 3 & 0 \\ 6 & 3 & 27 \end{pmatrix}$$

Let the **transpose** A^T of A be the matrix obtained from A by interchanging columns and rows in A, with general elements $a^T_{ij} = a_{ji}$ .. Then

$$\begin{pmatrix} 3 & 3 & 0 \\ 6 & 3 & 27 \end{pmatrix}^T = \begin{pmatrix} 3 & 6 \\ 3 & 3 \\ 0 & 27 \end{pmatrix}$$

14.2.1 Matrix multiplication

The product of an $m \times n$ matrix A with an $n \times p$ matrix B is an $m \times p$ matrix $C = AB$ whose ijth component is

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}, \quad i = 1, \ldots, m, \quad j = 1, \ldots p. \tag{14.2}$$

To see more, expand the sum in the formula. The ij entry of C is

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \ldots a_{in}b_{nj}$$

The elements from A are $a_{i1}, a_{i2}, a_{i3}, \ldots, a_{in}$, this is the i^{th} row of A. Similarly, $b_{1j}, b_{2j}, b_{3j}, \ldots, b_{nj}$ is the j^{th} column of B. Seen that way,

the ij entry of C is the scalar product of row i from A and column j from B.

Hence matrices can be multiplied together only if the *length* of a row in A is the same as the length of a column in B: $(m \times n)$X$(n \times p)$ must agree in n.

Example 14.5. *Calculate the product of the matrices*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ -4 & 5 & -6 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -2 \\ -3 & 4 \\ -5 & 6 \end{pmatrix}.$$

A is (2×3) and B is (3×2), so we can proceed.

$$\begin{aligned} AB &= \begin{pmatrix} 1 & 2 & 3 \\ -4 & 5 & -6 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ -3 & 4 \\ -5 & 6 \end{pmatrix} \\ &= \begin{pmatrix} 1(1) + 2(-3) + 3(-5) & 1(-2) + 2(4) + 3(6) \\ -4(1) + 5(-3) + (-6)(-5) & -4(-2) + 5(4) + (-6)6 \end{pmatrix} \\ &= \begin{pmatrix} -20 & 24 \\ 11 & -8 \end{pmatrix} \end{aligned}$$

Notice that $C = AB$ is a 2×2, whereas $D = BA$ is a 3×3 matrix. It is not only that $AB \neq BA$, these two matrices do not even have the same dimensions.

Exercise: Show that D is a 3×3 matrix with entries to be completed:

$$D = \begin{pmatrix} 9 & -8 & 15 \\ * & * & * \\ * & * & * \end{pmatrix}$$

That both AB and BA exist is not a given, this only works if A is $m \times n$ and B is $n \times m$, as above. If $m = n$, we say the matrices are *square*, but even then we normally have $AB \neq BA$, even though both products exist.

14.2.2 Properties of matrix multiplication

There is another way to look at matrix multiplication. We have seen how $A\underline{\mathbf{x}}$ can be thought of as a linear combination of the columns of A:

$$\begin{pmatrix} 2 & 1 & 2 \\ 2 & -1 & -1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} + z \begin{pmatrix} 2 \\ -11 \\ 2 \end{pmatrix}$$

and the result is a column vector.

If we are multiplying two matrices $AB = C$, then we can think of A multiplying a single column of B, giving a column of C in the same way: each column of C is a linear combination of the columns of A, with the coefficients for the linear combination coming from the given column of B.

We can turn this around and get yet another way of multiplying matrices. Take a row of A and multiply B, thinking of B in terms of rows.

$$(x, y, z) \begin{pmatrix} 2 & 1 & 2 \\ 2 & -1 & -1 \\ 1 & 1 & 2 \end{pmatrix} = x(2, 1, 2) + y(2, -1, -1) + z(1, 1, 2)$$

and the result is a row vector.

If (x, y, z) is row R_i of A, then the linear combination of rows of B obtained by multiplying $(x, y, z)B$ is row R_i of C. As before, every row of C is a linear combination of rows of B, and the coefficients for that linear combination come from a row of A. At this point, notation becomes difficult, because a row acts differently to a column. To disambiguate, we agree to take a column as default for a vector, and a row will the transpose of a vector. So now

$$\underline{\mathbf{y}} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} \quad \Rightarrow \quad \underline{\mathbf{y}}^T = (2, 2, 1)\,.$$

Theorem 14.1. *For matrices A, B, C and $\lambda \in \mathbb{R}$,*

1. *$A\,(B\,C) = (A\,B)\,C$.*
2. *$(\lambda A)B = A(\lambda B) = \lambda(AB)$.*
3. *If A, B are square matrices, then $(AB)^T = B^T A^T$.*

(Properties (i) and (ii) apply to non-square matrices, as long as the products exist.)

Proof: For (i), consult any textbook on Linear Algebra; for (ii), easy, using definition; for (iii), here's a sketch proof.

Begin with $B = \underline{\mathbf{x}}$, a single column vector. Then $A\underline{\mathbf{x}}$ is a linear combination of the columns of A. On the other hand, if we multiply A on the left with a row, $\underline{\mathbf{x}}^T A^T$ is a linear combinatin of the rows of A^T. In both cases, the coefficients of the linear combinations are the same: $x_1, x_2, \ldots$ and the rows of A^T are the columns of A: it's the **same** linear combination, but $A\underline{\mathbf{x}}$ is a column, while $\underline{\mathbf{x}}A$ is a row, the transpose of the column $A\underline{\mathbf{x}}$. Hence

$$\underline{\mathbf{x}}^T A^T = (A\underline{\mathbf{x}})^T\,,$$

so the claim is true if B has a single column. Now let B have two columns: $B = (\underline{\mathbf{x}}_1\ \underline{\mathbf{x}}_2)$. We apply the same idea to each column separately: the columns of AB are $A\underline{\mathbf{x}}_1$ and $A\underline{\mathbf{x}}_2$:

$$AB = (A\underline{\mathbf{x}}_1\ A\underline{\mathbf{x}}_2)$$

and similarly these columns transposed give rows:

$$(AB)^T = (A\underline{\mathbf{x}}_1\ A\underline{\mathbf{x}}_2)^T = \begin{pmatrix} (A\underline{\mathbf{x}}_1)^T \\ (A\underline{\mathbf{x}}_2)^T \end{pmatrix}$$

where the last is a matrix with two rows, $(A\underline{\mathbf{x}}_1)^T = \underline{\mathbf{x}}_1^T A^T$ and $(A\underline{\mathbf{x}}_2)^T = \underline{\mathbf{x}}_2^T A^T$, so that

$$(AB)^T = \begin{pmatrix} \underline{\mathbf{x}}_1^T A^T \\ \underline{\mathbf{x}}_2^T A^T \end{pmatrix}\,.$$

The last is a matrix with two rows, and each one of these rows is exactly what you get if you consider a row of B^T multiplying A^T, i.e. the matrix is equal to $B^T A^T$, which generalizes nicely for more than two rows/columns.

Example 14.6. *If $A = \begin{pmatrix} 2 & 1 \\ 2 & -1 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 3 \\ -2 & 4 \end{pmatrix}$ multiply to confirm that $(AB)^T = B^T A^T$.*

Exercise: Given an $n \times n$ matrix A, we can define its powers

$$A^1 = A, A^2 = A \times A,\ A^3 = A^2 \times A, \ldots, A^k = A^{k-1} \times A$$

Find the n-th power of the following matrices

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} a & a-b \\ 0 & b \end{pmatrix}$$

Let $\underline{\mathbf{x}} = (x_1, x_2, \ldots x_n)$ be a row vector and let $\underline{\mathbf{x}}^T$ be its transpose:

$$\underline{\mathbf{x}}^T = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}.$$

We think of a row vector as a $1 \times n$ matrix: a matrix with one row. Similarly, a column vector is an $n \times 1$ matrix, having a single column. We can multiply them as matrices. Then we recognise $\underline{\mathbf{x}}\underline{\mathbf{x}}^T$ as the scalar product, a real number (and 1×1 matrix), the square of the **length/norm** of the vector $\underline{\mathbf{x}}$:

$$\underline{\mathbf{x}}\,\underline{\mathbf{x}}^T = \sum_{i=1}^{n} x_i^2 = |\underline{\mathbf{x}}|^2$$

The **unit/identity matrix** of order n in the $n \times n$ matrix with $a_{ij} = 1,\ i = j$ and $a_{ij} = 0,\ i \neq j$. We can write this in the form

$$a_{ij} = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

δ_{ij} is called **Kronecker's delta**.

For $n = 3$ we have

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

We clearly have $I^2 = I,\ IA = AI = A$. We can now extend the notion of powers of matrices, defining $A^0 = I_n$ for any $n \times n$ matrix A.

Example 14.7. *If A and B* **commute***, i.e. $AB = BA$ then*

$$(A+B)^k = \sum_{i=0}^{k} \binom{k}{i} A^i B^{k-i}$$

14.2.3 Special classes of matrices

An $n \times n$ matrix is called **symmetric** provided that $A = A^T$, i.e. $a_{ij} = a_{ji},\ i, j = 1, \ldots n$.

Example 14.8. *The matrix*

$$A = \begin{pmatrix} 10 & 3 & 10 \\ 3 & 9 & -6 \\ 10 & -6 & 20 \end{pmatrix}$$

is symmetric.

More interesting, when a matrix isn't symmetric, is a very simple answer to the question "how do i get a symmetric matrix from this non-symmetric one?" For example

Example 14.9.

$$R = \begin{pmatrix} 1 & 3 \\ 3 & 0 \\ -2 & 4 \end{pmatrix}, \quad \Rightarrow \quad R^T = \begin{pmatrix} 1 & 3 & -2 \\ 3 & 0 & 4 \end{pmatrix}$$

where in this case, R and R^T aren't even the same size!

The answer is: for any matrix R, the matris RR^T is symmetric, because

$$(RR^T)^T = (R^T)^T R^T = RR^T$$

because $(AB)^T = B^T A^T$.

We note that RR^T always exists because to be able to multiply the two, we need the length of a row of the first to be equal to the length of a column of the second. As the rows of R *are* the columns of R^T, this is trivially satisfied.

So we find

$$RR^T = \begin{pmatrix} 1 & 3 \\ 3 & 0 \\ -2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 3 & -2 \\ 3 & 0 & 4 \end{pmatrix} = \begin{pmatrix} 10 & 3 & 10 \\ 3 & 9 & -6 \\ 10 & -6 & 20 \end{pmatrix}.$$

Note that $R^T R$ also exists and is symmetric for the same reason, but in this case would be a 2×2 matrix.

Exercise. Calculate $R^T R$ for the above example.

A square matrix is called **orthogonal** provided that $A^T = A^{-1}$.

Example 14.10. *The matrix*

$$A = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is orthogonal: $AA^T = A^T A = I$.

Example 14.11. *A matrix can have complex entries. For example consider the matrix*

$$A = \begin{pmatrix} 1+i & 7+2i \\ 3-i & 2+i \end{pmatrix}$$

A complex-valued matrix is called **hermitian** *if $A^T = \overline{A}$. It will be called* **unitary** *if $A^{-1} = A^T$.*

14.2.4 The inverse of a matrix

Let A be an $n \times n$ matrix and assume that there exists an $n \times n$ matrix B such that

$$AB = BA = I. \tag{14.3}$$

We will say that A is **invertible** and we will call B the **inverse** of the matrix A. We will use the notation A^{-1} for the inverse of an invertible matrix:

$$AA^{-1} = A^{-1}A = I.$$

Example 14.12. *Let*

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

Check that the inverse of this matrix is

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

Consider the linear system of equations

$$A\underline{\mathbf{x}} = \underline{\mathbf{b}} \tag{14.4}$$

and assume that the matrix A is invertible. Then the solution of (14.4) is

$$\underline{\mathbf{x}} = A^{-1}\underline{\mathbf{b}}.$$

Why this is the case is easy to see: multiply the matrix equation on the left by the inverse matrix:

$$A^{-1}(A\underline{\mathbf{x}}) = A^{-1}\underline{\mathbf{b}} \Rightarrow (A^{-1}A)\underline{\mathbf{x}} = A^{-1}\underline{\mathbf{b}} \Rightarrow (I_n)\underline{\mathbf{x}} = A^{-1}\underline{\mathbf{b}} \Rightarrow \underline{\mathbf{x}} = A^{-1}\underline{\mathbf{b}}$$

where we have used the associative property of matrix multiplication in the first step, the definition of the inverse matrix in the second step, and a property of the identity matrix in the last step.

Example 14.13. *The solution of the linear system of equations*

$$\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

An interesting property of the inverse is seen when we combine inverse with transpose:

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

To begin with, we apply the property of the transpose of a product of matrices:

$$AA^{-1} = I \quad \Rightarrow \quad (AA^{-1})^T = I^T = I$$

as the identity matrix is symmetric. Hence

$$(A^{-1})^T A^T = I$$

so the first of these two, $(A^{-1})^T$ is the inverse of the second, A^T, as claimed.

A well-known formula exists for the 2×2 inverse. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{then} \quad A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

14.3 Determinants

The formula for the inverse of the 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ depended on division by the scalar $ad - bc$. Clearly, this determines existence or not of the inverse. Should we have $ad - bc = 0$ then there is no inverse, the matrix A is singular. If we have $ad - bc \neq 0$, then A is invertible, and the inverse is as given.

This is the **determinant** of the 2×2 matrix, denoted det(A) or $|A|$:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Rightarrow \quad \det\,(A) = |A| = ad - bc\,.$$

14.3.1 Determinant of a 3×3 matrix

The determinant of a 3×3 matrix $A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ can be written as

$$\det(A) = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13} \tag{14.5}$$

where M_{ij} is called the *Minor*: a determinant obtained by deleting rom i and column j from A:

$$M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \quad M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{23} \end{vmatrix}, \quad M_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \tag{14.6}$$

The $+/-/+$ pattern comes from multiplying the Minor by ± 1 given by

$$C_{ij} = (-1)^{i+j} M_{ij}$$

and we call the term C_{ij} the ij-*Cofactor*

The expression for the determinant above is expansion in first row. We can also expand any other row or column. For example, by the second column

$$\det(A) = -a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{22} \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} - a_{13} \begin{vmatrix} a_{12} & a_{12} \\ a_{22} & a_{22} \end{vmatrix}$$

Note the operation of $(-1)^{i+j}$ giving a $-/+/-$ pattern here.

Exercise: Calculate the determinant

$$\begin{vmatrix} 1 & 4 & 1 \\ 2 & 1 & 3 \\ 3 & 2 & 2 \end{vmatrix}$$

The strategy is to exploit the structure of the determinant: we expand in the column or row that simplifies the calculation.

Exercise: Calculate the determinant

$$\begin{vmatrix} 63 & 124 & 10 \\ 0 & 0 & 1 \\ 2 & 1 & 6 \end{vmatrix}$$

We can also define higher order determinants. For a 4×4 matrix, expanding by any row/column. For example, by the third column:

det(A)=

$= (-1)^{1+3}a_{13}M_{13} + (-1)^{2+3}a_{23}M_{23} + (-1)^{3+3}a_{33}M_{33} + (-1)^{4+3}a_{43}M_{43} \quad = a_{13}M_{13} - a_{23}M_{23} + a_{33}M_{33} - a_{43}M_{43}$

14.3.2 Properties of determinants

1. The value of a determinant remains unchanged under transposition:

   $$\det(A) = \det(A^T).$$

   Example 14.14. *Calculate the determinant of*

   $$\begin{pmatrix} 1 & 4 & 1 \\ 2 & 1 & 3 \\ 3 & 2 & 2 \end{pmatrix}^T$$

2. If B is obtained from A by exchanging *exactly two* rows (or two columns) then det(A) $= -$det(B).

   Example 14.15. *Calculate the determinant of*

   $$\begin{pmatrix} 1 & 4 & 1 \\ 3 & 2 & 2 \\ 2 & 1 & 3 \end{pmatrix}$$

3. If 2 rows (or columns) of a determinant are the same or multiples of each other, then $\det(A) = 0$. A matrix with zero determinant is called *singular*. A matrix with non-zero determinant is called $invertible$. We will see later how the determinant and the inverse are related, but note that the determinant in the 2×2 case where an inverse matrix was given, was not zero!

Example 14.16. *Calculate the determinant of*

$$\begin{pmatrix} 1 & 4 & 2 \\ 3 & 2 & 6 \\ 2 & 1 & 4 \end{pmatrix}$$

4. If the elements of any column (or row) are multiplied by a factor λ then the determinant is multiplied by λ.

   Example 14.17. *Calculate the determinant of*

$$\begin{pmatrix} 1 & 3(4) & 1 \\ 2 & 3(1) & 3 \\ 3 & 3(2) & 2 \end{pmatrix}$$

5. If a multiple of one row (or column) is added to another row (or column), the determinant is unchanged.

   Example 14.18. *Calculate the determinant of*

$$\begin{pmatrix} 1 & 4 & 1 \\ 2 & 1 & 3 \\ 3+2(2) & 2+2(1) & 2+2(3) \end{pmatrix}$$

   (adding twice row 2 to row 3)

6. A square matrix is **upper (resp. lower) triangular** if it has all entries below (resp. above) the diagonal equal to zero. The determinant of such a matrix is the product of the diagonal elements.

   Example 14.19. *Calculate the determinant of the upper-triangular matrix*

$$det \begin{pmatrix} -1 & 2 & 3 \\ 0 & 3 & 2 \\ 0 & 0 & -4 \end{pmatrix}$$

   Left as exercise.

7. det(AB) =det(A)det(B).

Exercise: Calculate the determinant, expanding by a row and by a column.

$$\begin{vmatrix} 1 & 4 & 1 \\ 2 & 1 & 3 \\ 3 & 2 & 2 \end{vmatrix}$$

Exercise: Show that

$$\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = (b-a)(c-a)(c-b).$$

Chapter 15

LA II: Gaussian Elimination and applications

The fundamental task of Linear Algebra is the solution of the humble equation $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$. We've seen a little bit how this can turn out, at the beginning of the last chapter. There is either

- a unique solution;
- infinitely many solutions;
- no solution.

If solutions exist, the system is called **consistent**. If no solutions exist the system is called **inconsistent**

Using the technique known as Gaussian elimination, we can investigate the question of existence/uniqueness of solutions efficiently. The notation used in the technique is due to Carl Friedrich Gauss, one of the great Mathematicians of the 19th century, who also contributed to Electromagnetic theory. But though he systematised the notation, Gauss did not invent the technique: it first appears in the "The Nine Chapters on the Mathematical Art", a Chinese text written between the tenth and second centuries BCE.

15.1 Gaussian elimination

Consider a system of m linear equations in n unknowns:

$$\begin{aligned}
a_{11}x_1 \quad &+a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\
a_{21}x_1 \quad &+a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\
\vdots \quad & &\vdots \\
\vdots \quad & &\vdots \\
a_{m1}x_1 \quad &+a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m
\end{aligned}$$

which is more efficiently rewritten as $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ where A is an $m \times n$ matrix, $\underline{\mathbf{x}} \in \mathbb{R}^n$ and $\underline{\mathbf{b}} \in \mathbb{R}^m$.

We can also think of this as a matrix multiplication: $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ is $(m \times n) \times (n \times 1) = (m \times 1)$ with the usual formula $\displaystyle\sum_{j=1}^{n} a_{ij}x_j = b_i, \;\; i = 1, \ldots, m.$ where each i gives one equation.

To avoid the endless elimination of variables, we'd like to solve this without reference to x_i.

Example 15.1. *We use Gaussian elimination to solve the linear system of 3 equations in three unknowns.*

$$\begin{aligned}
x \quad &+y + 2z &= 1 \\
2x \quad &+y + 2z &= 2 \\
2x \quad &-y - z &= 1
\end{aligned}$$

We begin with the traditional approach, first eliminating x, then y:

$$\begin{array}{rrrl} x & +y & +2z & =1 \\ 2x & +y & +2z & =2 \\ 2x & -y & -z & =1 \end{array} \quad \underset{\substack{Eq2-2\times Eq1 \\ Eq3-2\times Eq1}}{\Rightarrow} \quad \begin{array}{rrrl} x & +y & +2z & =1 \\ & -y & -2z & =0 \\ & -3y & -5z & =-1 \end{array} \quad \underset{Eq3-3\times Eq2}{\Rightarrow} \quad \begin{array}{rrrl} x & +y & +2z & =1 \\ & -y & -2z & =0 \\ & & z & =-1 \end{array}$$

The last equation has the solution $z=-1$, which we substitute into $-y-2z=0 \Rightarrow y=2$ and then substitute into the first, to get $x=1$. Doing this with a larger system would be cumbersome, so we need the efficiency given by $A\underline{\mathbf{x}}=\underline{\mathbf{b}}$:

$$\begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 2 \\ 2 & -1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix},$$

But even that isn't enough. We form the **augmented coefficient matrix** $(A:\underline{\mathbf{b}})$:

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 2 & 1 & 2 & 2 \\ 2 & -1 & -1 & 1 \end{array}\right)$$

and operate on its rows exactly as the operations used to eliminate first x, then y from the equations. The $x-$coefficient we keep is identified by a box, we call this a **pivot** and use it to eliminate the other $x-$coefficients, i.e below it in the first column. The operations Eq2-2$\times$ Eq1 and Eq3-2$\times$ Eq1 are now operations on rows: R_2-2R_1 and R_3-2R_1. We read the first of these as "subtract twice the first row from the second row". Then:

$$(A:\underline{\mathbf{b}}) = \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 2 & 1 & 2 & 2 \\ 2 & -1 & -1 & 1 \end{array}\right) \underset{\substack{R_2-2R_1 \\ R_3-2R_1}}{\sim} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 0 & \boxed{-1} & -2 & 0 \\ 0 & -3 & -5 & -1 \end{array}\right) \underset{R_3-3R_2}{\sim} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 0 & \boxed{-1} & -2 & 0 \\ 0 & 0 & \boxed{1} & -1 \end{array}\right)$$

In the first two steps, the pivot is $a_{11}=1$, used to eliminate the entries in a_{21} and a_{31}. For the next step, the pivot is $a_{22}=-1$ used to eliminate the entry in a_{32}. The last pivot is $a_{33}=1$: there are three pivots, one corresponding to each unknown. The symbol " " reads as "is row equivalent to", and the statement C D means that matrices C and D are row equivalent, i.e. that one can be obtained from the other by row operations. and the equations they embody have the same solutions.

The next part of the process is **back-substitution**, corresponding exactly to the steps when we had equations: the last row gives the equation $z=-1$. The middle row gives the equation $-y-2z=0$ and substituting the value $z=-1$ gives $y=2$. The top row then gives x, as before.

That went smoothly, but it's also a good idea to look at what else can happen. For example, if the inital set-up were $a_{22}=2$, then after the first step we'd have a zero in the desired pivot position. Fortunately, that's not a problem as there isn't a zero below, at a_{32}. We exchange rows, using the notation $R_2 \leftrightarrow \mathbb{R}_3$:

$$(A:\underline{\mathbf{b}}) = \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 2 & 2 & 2 & 2 \\ 2 & -1 & -1 & 1 \end{array}\right) \underset{\substack{R_2-2R_1 \\ R_3-2R_1}}{\sim} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 0 & 0 & -2 & 0 \\ 0 & -3 & -5 & -1 \end{array}\right) \underset{R_2 \leftrightarrow R_3}{\sim} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 0 & \boxed{-3} & -5 & -1 \\ 0 & 0 & \boxed{-2} & 0 \end{array}\right)$$

and in this case the procedure is done, and we have three pivots and proceed to back-substitution.

Quite another problem might arise if the initial set-up had $a_{33}=-2$:

$$(A:\underline{\mathbf{b}}) = \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 2 & 1 & 2 & 2 \\ 2 & -1 & -2 & 1 \end{array}\right) \underset{\substack{R_2-2R_1 \\ R_3-2R_1}}{\sim} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 0 & \boxed{-1} & -2 & 0 \\ 0 & -3 & -6 & -1 \end{array}\right) \underset{R_3-3R_2}{\sim} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 0 & \boxed{-1} & -2 & 0 \\ 0 & 0 & 0 & -1 \end{array}\right)$$

Now the last row gives the equation $0x+0y+0z=-1$, and the system has no solutions, is inconsistent. If the intial third equation were $2x-y-2z=2$, then after the same elimination steps, the entire last row would consist of zeroes, and hence no contradiction in the equation $0x+0y+0z=0$. Thus, the second row would have a free variable and we end up with a line in $\mathbb{R}^3$, infinitely many solutions. Recall that these are the prism and star case, if we think of equations of planes. Crucially, in both cases we have two interesting observations:

- after elimination, there are two pivots in A;
- looking at the original columns of A, we see that C_3 is twice C_2.

When we were looking at the multiplication $A\underline{\mathbf{x}}$ as a linear combination of columns of A, we saw that $xC_1 + yC_2$ produces a plane in $\mathbb{R}^3$. If C_3 is in that plane, the linear combinations $xC_1 + yC_2 + zC_3$ add nothing further and the linear combinations do not produce the entire 3D space, only a 2D plane. In that case, if $\underline{\mathbf{b}}$ is outside that plane, there is no way to yield it as a linear combination of the columns and there is no solution, and if $\underline{\mathbf{b}}$ is inside the plane, there are infinitely many ways to produce it as a linear combination of columns of A.

When there were three pivots in A we had a unique solution, when there were only two, we had no solution or infinitely many solutions.

15.1.1 Elimination matrices

After Gaussian elimination, the part of the augmented matrix corresponding to A is now upper triangular:

$$U = \begin{pmatrix} \boxed{1} & 1 & 2 \\ 0 & \boxed{-1} & -2 \\ 0 & 0 & \boxed{1} \end{pmatrix}$$

Our next goal is to represent each of the elimination steps which change A to U as multiplication by a matrix. Recall that multiplying a matrix A by a row can be seen as giving a linear combination of the rows of A, which is another row:

$$(1, -2, 7) \begin{pmatrix} 3 & 1 & -4 \\ -2 & 0 & 5 \\ -1 & 1 & -3 \end{pmatrix} = 1(3, 1, -4) - 2(-2, 0, 5) + 7(-1, 1, 3) = (-8, 8, 7)\ .$$

Similarly, if we multiply two matrices EA, the product matrix has row r_i given by linear combinations of the rows of A, and the coefficients of the linear combination are in row r_i in E.

Hence, looking at the elimination steps, we begin with the second row. In the first step, only row two changes. If we want only row two of A to change, we need to take E with rows that on multiplication do nothing to the first and third rows of A and give the right elimination on row two.

Say A has rows R_1, R_2, R_3. Leaving the first row of A untouched means E needs to have as first row the entries $(1, 0, 0)$ because those values give the first row of the product matrix EA as $1R_1 + 0R_2 + 0R_3 = R_1$, unchanged. Likewise, if we want the third row unchanged, then the third row of E must have the entries $(0, 0, 1)$, so that the third row of the product EA becomes $0R_1 + 0R_2 + 1R_3 = R_3$, unchanged.

The second row of E, in our example needs to yield $R_2 - 2R_1$, and the entries would be $(-2, 1, 0)$ giving $-2R_1 + R_2 + 0R_3$ as the second row of EA as required. Putting everything together, we have our first **Elimination matrix** E_{21}

$$E_{21}A = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 2 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -2 \\ 2 & -1 & -1 \end{pmatrix}$$

We label the elimination matrix E_{21} because its action is to produce a zero in the 21-position of A. Note that this can be obtained by taking the identity matrix and putting the correct multiplier -2 in the $21-$ position. The next step was $R_3 - 2R_1$ to obtain a zero in the $31-$ position, acting only on the bottom row, leaving the top two rows unchanged. The multiplier is again -2 and so E_{31} will have -2 in the $31-$ position and otherwise look like the identity matrix:

$$E_{31}(E_{21}A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -2 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & -3 & -5 \end{pmatrix}$$

For the last step, getting a zero in the $32-$position, with multiplier -3 we proceed similarly:

$$E_{32}(E_{31}E_{21}A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & -3 & -5 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{pmatrix} = U$$

If we multiply the three elimination matrices together, we get one matrix multiolying A, call it $E = E_{32}E_{31}E_{21}$ so that

$$EA = U \Rightarrow A = LU$$

where we would like to know what is L, because this is one of the key steps here: factorizing A into the product LU, called $LU-$decomposition. One obvious point: if E is invertible, and we will see that it is, then L is that inverse. So first we need to talk about inverses.

15.2 Finding the inverse matrix

The first thing to think about is, what is the inverse of a product AB? We need a matrix $(AB)^{-1}$ so that multiplied together, the two give the identity:

$$AB(AB)^{-1} = I$$

The obvious answer is that it must be $B^{-1}A^{-1}$ which we check by multiplying:

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

Crucially, we observe that if the inverse of a product matrix AB is to exist, both A and B need to be invertible. So if we want the inverse of $E = E_{32}E_{31}E_{21}$ this would be

$$E^{-1} = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}\ .$$

It turns out that these inverses of elimination matrices are very easy. If we look at the first one we met, then

$$E_{21}^{-1}E_{21} = I \quad \text{and} \quad E_{21}^{-1}E_{21}A = IA = A$$

so what is the action of the two matrices multiplying A? The first one, E_{21} took twice row one and subtracted it from row two, leaving the first and third row untouched. The second one, E_{21}^{-1} must undo this action, so that we end up back with A after two multiplications. It should be obvious that to reverse the action we need to take the same multiple, twice row one, and *add* it to row two. Hence

$$E_{21} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \Rightarrow \quad E_{21}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\ .$$

Similarly

$$E_{31}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \quad \text{and} \quad E_{32}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}\ .$$

We can multiply any pair together to confirm they are inverses:

$$E_{21}E_{21}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I_3\ .$$

What about general inverses? The first thing to establish, is when do inverses exist and when not? We've already talked about it in terms of determinant for a 2×2 matrix. Let's look at one of these which doesn't have an inverse, and see why. This will also give us an insight into how to find the inverse, when one exists.

Example 15.2. *Why is $A = \begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix}$ singular, not invertible?*

We can observe easily that the determinant is $2(10) - 4(5) = 0$. But there's another good way to look at why this matrix has no inverse. Let's say an inverse exists, say

$$A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then we must have

$$AA^{-1} = \begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Remembering that (i) we can think of multiplication of A times the first column of A^{-1} giving the first column of the identity, and (ii) multiplication of A times a column is a linear combination of the columns of A, we observe

$$\begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix}\begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \Rightarrow \quad a\begin{pmatrix} 2 \\ 4 \end{pmatrix} + b\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and we observe immediately that the linear combination of the columns of A on the right of the last equation lies on the line with direction $\begin{pmatrix} 2 \\ 5 \end{pmatrix}$ because both columns lie on that line, so any linear combination must lie on that line and there is no way to obtain the vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ from such a linear combination. Looking at the second column of the identity, we'd have the same problem. So there is no inverse.

Another way, even more important, of observing when a matrix is singular is the following argument. Can we find a non-zero vector $\underline{\mathbf{x}}$ so that $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$? For our matrix we can, there's an obvious one:

$$\begin{pmatrix} 2 & 5 \\ 4 & 10 \end{pmatrix}\begin{pmatrix} -5 \\ 2 \end{pmatrix} = -5\begin{pmatrix} 2 \\ 4 \end{pmatrix} + 2\begin{pmatrix} 5 \\ 10 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and we observe that $(5, -2)^T$ would have worked just as well, in fact any multiple of that vector would work. The fact that we can find a non-zero $\underline{\mathbf{x}}$ satisfying $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ immediately implies that A is singular. Suppose A^{-1} exists, then we can multiply both sides to get

$$A^{-1}A\underline{\mathbf{x}} = A^{-1}\underline{\mathbf{0}} \Rightarrow \underline{\mathbf{x}} = \underline{\mathbf{0}}$$

which says that if an inverse exists, then the only possible vector $\underline{\mathbf{x}}$ is the zero-vector. But we just found one that isn't the zero vector. The key observation is there exists a linear combinations of the columns of A that yields the zero vector and therefore the matrix is singular.

So now let's look at a matrix that does have an inverse. For a 2×2 example, it's easy, just pick one that doesn't have both columns pointing in the same direction.

Example 15.3. *Find the inverse of* $A = \begin{pmatrix} 2 & 5 \\ 4 & 9 \end{pmatrix}$.

This time we know the inverse exists, so begin again with

$$AA^{-1} = \begin{pmatrix} 2 & 5 \\ 4 & 9 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and think of multiplication of A with a column of A^{-1} giving a column of the identity. So

$$\begin{pmatrix} 2 & 5 \\ 4 & 9 \end{pmatrix}\begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 5 \\ 4 & 9 \end{pmatrix}\begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Gaussian elimination would give both unknown columns $(a, c)^T$ and $(b, d)^T$, doing row operations on

$$\left(\begin{array}{cc|c} 2 & 5 & 1 \\ 4 & 9 & 0 \end{array}\right) \quad \text{and} \quad \left(\begin{array}{cc|c} 2 & 5 & 0 \\ 4 & 9 & 1 \end{array}\right)$$

The improvement on Gaussian elimination, called Gauss-Jordan elimination, to obtain the inverse comes from observing that the row operations are the same in both cases, because we're working on A both times, so why not do both at once: augment A with both columns of the identity and work on that:

$$(A : I) \sim \left(\begin{array}{cc|cc} 2 & 5 & 1 & 0 \\ 4 & 9 & 0 & 1 \end{array}\right) \underset{R_2 - 2R_1}{\sim} \left(\begin{array}{cc|cc} 2 & 5 & 1 & 0 \\ 0 & -1 & -2 & 1 \end{array}\right)$$

The aim of Gauss-Jordan is to use row operations to make the right half of this augmented matrix into the identity:

$$(A : I) \sim \left(\begin{array}{cc|cc} 2 & 5 & 1 & 0 \\ 0 & -1 & -2 & 1 \end{array}\right) \underset{R_1 + 5R_2}{\sim} \left(\begin{array}{cc|cc} 2 & 0 & -9 & 5 \\ 0 & -1 & -2 & 1 \end{array}\right)$$

The last step consists in multiplying row one by $1/2$ and row two by -1. This is a legitimate row operation, albeit one we haven't considered yet. Each row represents two equations - and the right-hand side is the same in both - and multiplying an equation by a non-zero constant does not alter the equation. Hence

$$(A:I) \sim \left(\begin{array}{cc|cc} 2 & 0 & -9 & 5 \\ 0 & -1 & -2 & 1 \end{array} \right) \underset{\substack{\frac{1}{2}R_1 \\ -R_2}}{\sim} \left(\begin{array}{cc|cc} 1 & 0 & -\frac{9}{2} & \frac{5}{2} \\ 0 & 1 & 2 & -1 \end{array} \right) .$$

We now claim this last form is $(I : A^{-1})$. The left half, corresponding to A has been transformed into I and the right half, corresponding to I has been transformed into A^{-1}. How can this be? Recall that the first two row operation correponds to multiplication by an elimination matrix. The last two row operations can be represented by a scaling matrix:

$$S = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -1 \end{pmatrix}$$

Multiplying SA has the same effect as the two row ops $\frac{1}{2}R_1$ and $-R_2$, check! Crucially, the action of the three matrices, one scaling and two elimination matrices, can be seen as multiplying on the left by a single matrix E, the product of the three, as we saw before. The left-hand half is $EA = I$, and the right-hand half, having the same row operations perfomed, is therefore $EI = E$. But if $EA = I$, then E is the inverse, and the right-hand half is E, the inverse. Hence

$$A^{-1} = \begin{pmatrix} -\frac{9}{2} & \frac{5}{2} \\ 2 & -1 \end{pmatrix} .$$

15.3 LU-decomposition

We had arrived two sections back, using elimination matrices to obtain

$$EA = U ,$$

where U is upper triangular, and we'd now like to write

$$A = LU .$$

After the previous section, we know what L is: it must be E^{-1} because

$$E^{-1}(EA) = E^{-1}U \quad \Rightarrow \quad IA = A = E^{-1}U = LU .$$

We continue with the example we already had:

$$EA = (E_{32}E_{31}E_{21})A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 2 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{pmatrix} = U .$$

But we saw how to find the inverse of E in the last section, though we didn't calculate it:

$$E^{-1} = (E_{32}E_{31}E_{21})^{-1} = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & 3 & 1 \end{pmatrix} = L .$$

What's nice about using the inverses E_{21}^{-1}, E_{31}^{-1}, and E_{32}^{-1} to obtain L is that the multipliers $2, 2$ and 3 from each elimination matrix turn up in their respective positions in L. We could have found $E = E_{32}E_{31}E_{21}$ first, and then inverted, why not? Check for yourself: if you obtain E first, you don't have the nice short-cut with the multipliers $-2, -2$ and -3; you would find E has -8 in the $31-$position. Finally we have our first matrix factorization, the LU-decomposition:

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 2 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{pmatrix} = LU .$$

And now it's also obvious why we call it L: it will always be a lower-triangular matrix, as it is obtained from the identity by a series of row operations where multiples of rows are added to other rows, further down, hence the multipliers appearing below the diagonal only.

Sometimes it's also written as $A = LDU$ where D is a diagonal matrix containing the scaling factors which give the pivots in U, so that in this form, all pivots in U are equal to one:

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 2 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} = LDU\,,$$

which is not very impressive in our example, as there is a single -1 taken out of U. If for example we had

$$U = \begin{pmatrix} 2 & 3 & -2 \\ 0 & -3 & 6 \\ 0 & 0 & 4 \end{pmatrix} \quad \text{then} \quad DU = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & \frac{3}{2} & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

Symmetric matrices are 'nice' in that they make a lot of problems easier to solve. This has an interesting consequence here. Let $A = A^T$ be a symmetric matrix, then if we have the $LU-$ decomposition:

$$A = LDU \quad \Rightarrow \quad A^T = (LDU)^T = U^T D^T L^T = U^T D L^T = A \quad \Rightarrow \quad U = L^T \quad \text{and} \quad L = U^T\,.$$

This means a problem involving a symmetric matrix will be computationally less expensive!

There's one situation we haven't discussed in the LU-context yet. When we first used Gaussian elimination we looked at a situation where we had to exchange two rows because after the first row operation, the next pivot position had a zero:

$$A : \underline{\mathbf{b}} = \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 2 & 2 & 2 & 2 \\ 2 & -1 & -1 & 1 \end{array}\right) \underset{\substack{R_2 - 2R_1 \\ R_3 - 2R_1}}{\sim} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 0 & 0 & -2 & 0 \\ 0 & -3 & -5 & -1 \end{array}\right) \underset{R_2 \leftrightarrow R_3}{\sim} \left(\begin{array}{ccc|c} \boxed{1} & 1 & 2 & 1 \\ 0 & \boxed{-3} & -5 & -1 \\ 0 & 0 & \boxed{-2} & 0 \end{array}\right)$$

We need to think about how to effect exchanging two rows through left multiplication by a matrix. These are permuation matrices.

15.3.1 Permutation matrices

First, we give a thought to what a permutation matrix must look like if we use it to multiply a 3×3 matrix A on the left, to exchange two rows in A, say the second and third, as the row operation above did. First off, row one stays the same, so P must have top row $(1, 0, 0)$. Then, row two becomes row three, so the middle row of P should be $(0, 0, 1)$ and similarly, with row three becoming row two, the bottom row of P must be $(0, 1, 0)$ giving

$$P_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Note how P is obtained from the identity with the same row operation we want to effect on A, the same as the elimination and scaling matrices we saw earlier. There's a general principle here:

If B is obtained from the identity matrix by
(i) performing a row exchange,
(ii) adding a multiple of one row to another, or
(iii) multiplying rows by non-zero scalars,
then multiplying BA will reproduce the same row operation on A.

In fact we can exchange more than one row at a time in a permutation matrix. For example if we multiply A by

$$P_{132} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

the new version of A would have the old rows in the New Order, from top to bottom: R_2, R_3, R_1. The notation P_{132} indicates that $R_1 \to R_3$ and $R_3 \to R_2$, so implicitly $R_2 \to R_1$.

Exercise. Obtain all 3×3 permutation matrices. Above are two, one more is the identity: if we're going to exchange however many rows, we have to include the possibility of exchanging zero rows. How many 4×4 permutation matrices do you expect? And 5×5?

Digression, not examinable: permutation matrices of a given size form what is known as a **group**. A group consists of several elements, here matrices, and a mutliplication operation, here matrix multiplication. Among the groups there must be a multilicative identity element, here the identity matrix. Furthermere, (i) for every element A in the group, there must be a multiplicative inverse A^{-1} in the group, (ii) the multiplication must be associative, and (iii) the group is **closed** under multiplication, i.e. any product of two elements must be in the group.

We already have the identity and associative property. It is easy to see that

$$P_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

is its own inverse: it you exchange two rows, do it again and return to the original state; likewise for any permutation matrix that exchanges only two rows. For

$$P_{132} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

the inverse must move R_1 to R_2, as well as R_2 to R_3 and R_3 to R_1, hence

$$P_{132}^{-1} = P_{123} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} .$$

Exercise: Confirm that the permutation matrices of order 3 form a group. You need to (i) show that every element has an inverse in the group, and (ii) multiplying any two elements together yields an element of the group. There is a rich body of pure maths called Group Theory: great fun if you like abstraction. End of digression, back to LU! (Awww...)

An example will help see the way forward, we use the 3×3 example above.

Example 15.4. *Using row operations we had*

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 2 \\ 2 & -1 & -1 \end{pmatrix} \underset{\substack{R_2-2R_1 \\ R_3-2R_1}}{\sim} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & -2 \\ 0 & -3 & -5 \end{pmatrix} \underset{R_2 \leftrightarrow R_3}{\sim} \begin{pmatrix} 1 & 1 & 2 \\ 0 & -3 & -5 \\ 0 & 0 & -2 \end{pmatrix} = U$$

This is equivalent to

$$E_{31}E_{21}A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 2 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & -2 \\ 0 & -3 & -5 \end{pmatrix} .$$

Next, we multiply by P_{23} and finish:

$$P_{23}(E_{31}E_{21}A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 0 & -2 \\ 0 & -3 & -5 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & -3 & -5 \\ 0 & 0 & -2 \end{pmatrix} = U$$

What's nice to observe is that if we can identify the row exchange or exchanges (there may be more than one!) involved in obtaining U this way, we can first multply A by the appropriate permutation matrix that carries out these exchanges and then do a straightforward $LU-$decompostion on PA, obtaining

$$PA = LU .$$

We observe that for the above example,

$$P_{23}A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 2 & 2 & 2 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 2 & -1 & -1 \\ 2 & 2 & 2 \end{pmatrix}$$

and it is left as an exercise to confirm that we obtain the same U with two elimination matrices, whose inverses then form L, so that $PA = LU$. Warning: these elimination matrices are similar, but *not* the same as in the first version $P_{23}(E_{31}E_{21}A)$.

Exercise: For the 4×4 matrix

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

we note that there is no pivot in row one: the very first operation needs to be a row exchange. After this, elimination will put another zero in a pivot position, requiring a second row exchange. Combine the two row exchanges into a single permutation matrix P and then finish by finding the required elimination matrices, and combine their inverses into L, giving the decomposition $PA = LU$.

Chapter 18

LA III: Vector Spaces and Subspaces

In the first part of Linear Algebra, we met some interesting ideas involving systems of linear equations and how to solve them using matrices and Gaussian elimination, but these have been rather specialized, involving mostly square matrices and systems that worked nicely.

Something we haven't mentioned so far: what happens with $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ when A is not a square matrix, or when there are zeroes in pivot positions we can't get rid of? The case we've examined in any detail for $LU-$decomposition is of a square, invertible matrix A. We now look at the general case and to do that, we first need to develop some of the ideas involved in **vector spaces** containing **subspaces.**

First off, what is a vector space? The simplest idea is that $\mathbb{R}^2$ and $\mathbb{R}^3$ are very familiar vector spaces, as is any $\mathbb{R}^n$, but that tells us nothing. The key idea is that we have a set, call it V: a collection of vectors with real number entries and satisfying two crucial properties. If $\underline{\mathbf{u}}$ and $\underline{\mathbf{v}}$ are in our vector space V, then

- $a\underline{\mathbf{v}} \in V$ for any real a;
- $\underline{\mathbf{u}} + \underline{\mathbf{v}} \in V$.

This property is referred to by saying the set V is **closed under** scalar multiplication and vector addition. In other words, we can add any pair of vectors without leaving the set, and we can multiply a vector by any scalar and also not leave the set. The first obvious consequence of this is that any vector space must contain the zero vector, as $0\underline{\mathbf{u}} = \underline{\mathbf{0}}$ for any vector $\underline{\mathbf{u}}$, so this must be in V. And if you look at $\mathbb{R}^2$, say, it contains vectors like

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \begin{pmatrix} e \\ \pi \end{pmatrix},$$

and so on. And it's obvious that this set is closed under vector addition and scalar multiplication. One nice way to sum this up is that if you have any two vectors $\underline{\mathbf{u}}, \underline{\mathbf{v}} \in V$, then **all** linear combinations $a\underline{\mathbf{u}} + b\underline{\mathbf{v}}$ are also in V.

In $\mathbb{R}^3$ we look at 3D vectors:

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} e \\ \pi \\ 1 \end{pmatrix},$$

etc.

[Note: a more formal way of defining a vector space is to list a few more properties which we take for granted in $\mathbb{R}^n$, including vector addition being commutative and associative, some distributive laws for scalar multiplication, and a few more. This is superfluous for $\mathbb{R}^n$, but not if we look at other sets that can be seen as vector spaces. For example the set of polynomials of degree $\leq n$ can be thought of as a vector space: the vectors are polynomials; vector addition and scalar multiplication the obvious ones.]

18.1 Subspaces

A **subspace** is a subset of a vector space, which is also a vector space in its own right: a vector space inside a vector space. For example, in $\mathbb{R}^2$, we can think of a line through the origin with equation

$$\underline{\mathbf{r}} = \lambda \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

This is indeed a vector space: add any two vectors $a\begin{pmatrix}1\\2\end{pmatrix}$ and $b\begin{pmatrix}1\\2\end{pmatrix}$ on the line and you get $(a+b)\begin{pmatrix}1\\2\end{pmatrix}$, staying on the line; multiply any vector $b\begin{pmatrix}1\\2\end{pmatrix}$ on the line by a scalar a and you get $ab\begin{pmatrix}1\\2\end{pmatrix}$, on the line.

What doesn't work? Try the first quadrant in the $xy-$plane, including the axes, so $x, y \geq 0$. This is a subset of $\mathbb{R}^2$ and it is closed under vector addition: add any two vectors (a, b) and (c, d) satisfying $a, b, c, d \geq 0$ you obviously get $(a+c, b+d)$ still in the first quadrant. But not multiplication by a scalar, as $-1(a, b)$ is in the third quadrant. Or try a line that's not through the origin.

$\mathbb{R}^2$ itself, being a subset of $\mathbb{R}^2$ is also a subspace of $\mathbb{R}^2$, the subspace containing all of $\mathbb{R}^2$, the biggest subspace. Another, not so obvious subspace of $\mathbb{R}^2$ is the zero vector alone, the set $V = \{(0,0)\}$, the smallest subspace. It obeys both conditions: $a(0,0) = (0,0)$ for any real a and $(0,0)+(0,0) = (0,0)$, so the set is closed as required. It's not the most interesting subspace, but it is one, nevertheless and we will see how it is important.

So $\mathbb{R}^2$ has three kinds of subspace:

- the whole of $\mathbb{R}^2$;
- lines through the origin;
- the zero vector.

Similarly, $\mathbb{R}^3$ has four kinds of subspaces:

- the whole of $\mathbb{R}^3$;
- planes through the origin;
- lines through the origin;
- the zero vector.

If you're not convinced about the second one of these, think about a plane through the origin as generated by a point, $(0,0,0)$ and adding to this all possible linear combinations of two vectors which don't lie on a line, such as

$$a\begin{pmatrix}-1\\2\\3\end{pmatrix} + b\begin{pmatrix}2\\3\\0\end{pmatrix}.$$

It shouldn't be hard to see that if you take any vector of this form and multiply it by a scalar, you get another vector of this form. Or add two vectors.

18.1.1 Unions and Intersections of subspaces

If we take two subspaces of, say $\mathbb{R}^3$, a plane P and a line, L these are two subsets of $\mathbb{R}^3$. What about their union? We assume the line does not lie in the plane, but both contain the origin.

For picture, see lectures.

If we take $V = P \cup L$, there's no problem with closure if we take $\underline{\mathbf{u}}, \underline{\mathbf{v}}$ from one either L or P. Both L and P are subspaces of $\mathbb{R}^3$, so $a\underline{\mathbf{u}} + b\underline{\mathbf{v}}$ will be in either L or P and so in the union. But if we take one from each, their sum will be outside the union.

For picture, see lectures.

For the intersection, on the other hand, we have $L \cap P$ with only a single element, the zero vector, which we know is a subspace. It turns out that *every* intersection of two subspaces A, B of a vector space is also a subspace.

Take two elements, $\underline{\mathbf{u}}, \underline{\mathbf{v}} \in L \cap P$. Then $\underline{\mathbf{u}}, \underline{\mathbf{v}}$ is in L, so $a\underline{\mathbf{u}} + b\underline{\mathbf{v}}$ is in L because L is a subspace. Likewise, $\underline{\mathbf{u}}, \underline{\mathbf{v}}$ is in P, so $a\underline{\mathbf{u}} + b\underline{\mathbf{v}}$ is in P because P is a subspace. As $\underline{\mathbf{u}}, \underline{\mathbf{v}}$ is in both L, P, it is in the intersection $L \cap P$, so this set is closed under scalar multiplication and vector addition, and is therefore a subspace. The intersection can be just the zero vector, or something bigger. Which one will turn out to be important!

18.2 Column space

A special kind of subspace is related to a matrix: remember, this is about $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$. Take the matrix

$$A = \begin{pmatrix} -1 & 2 \\ 2 & 3 \\ 3 & 4 \end{pmatrix},$$

and look at its columns. They will be in a subspace of $\mathbb{R}^3$ containing, among others, *those* columns:

$$\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

This subspace will also contain the zero vector and all linear combinations of the two columns:

$$a\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} + b\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}.$$

In fact, those linear combinations are the whole subspace, a plane through the origin in $\mathbb{R}^3$. This is the **column space** of the matrix A, denoted $C(A)$ which contains all possible linear combinations of the columns of A. The column space is always a subspace of $\mathbb{R}^n$, where n is the length of the column of the $m \times n$ matrix A.

For picture, see lectures.

By the way we've constructed it, it should be obvious that $C(A)$ will always be a subspace of some $\mathbb{R}^n$, being clearly closed under vector addition and scalar multiplication. Here we have a plane, because the columns don't lie on a line, but for

$$A = \begin{pmatrix} -1 & -2 \\ 2 & 4 \\ 3 & 6 \end{pmatrix}, \Rightarrow C(A) = \left\{ a\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} + b\begin{pmatrix} -2 \\ 4 \\ 6 \end{pmatrix} \right\} = \left\{ k\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} \right\}$$

which is a line. So we don't always know what sort of subspace we'll get: it depends on the columns.

18.2.1 Column space and Linear equations

How does column space of a matrix A relate to the system of linear equations $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$? Consider the column space of a matrix

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix}$$

so the column space $C(A)$ is all linear combinations of the columns of A which are 4-D vectors:

$$C(A) = \left\{ a\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + b\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + c\begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \right\}$$

and the first question we consider is

"Can this be the whole of $\mathbb{R}^4$?"

In other, words, is every vector in $\mathbb{R}^4$ found in $C(A)$? Instinct should tell us "no" : if we begin with the linear combinations of three vectors, we can't get all of 4D-space. So $C(A)$ is a smaller subspace than the whole vector space. Question: how much smaller?

Let's go back to $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$, and the question above can be rephrased as

"Does $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ have a solution for every $\underline{\mathbf{b}}$?"

Can we solve

$$A\underline{\mathbf{x}} = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + y \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + z \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \underline{\mathbf{b}}$$

for any $\underline{\mathbf{b}}$? If we can't reach all of $\mathbb{R}^4$ with the linear combinations in $C(A)$, then there will be some $\underline{\mathbf{b}}$ which is not such a linear combination, and the equation will have no solution.

The next question then arises: if not all possible $\underline{\mathbf{b}}$ provide solutions, which $\underline{\mathbf{b}}$ do? There's an obvious one: let $\underline{\mathbf{b}} = \underline{\mathbf{0}}$. The equation $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ always has a solution for any A, because $A\underline{\mathbf{0}} = \underline{\mathbf{0}}$. Or, in the terms of column space, $\underline{\mathbf{b}} = \underline{\mathbf{0}}$ is always in any subspace, so it's in $C(A)$.

Or just let $\underline{\mathbf{b}} = (1, 2, 3, 4)^T$ or any of the other columns. Then the solution is $x = 1$ and $y = z = 0$. Or we just think of some solution like $x = 23, y = 42$ and $z = \pi$. Then set

$$\underline{\mathbf{b}} = 23 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + 42 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \pi \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$$

and we know that $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has a solution, it's $\underline{\mathbf{x}} = (23, 42, \pi)$! It should by now be obvious: solutions to $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ exist *exactly* when $\underline{\mathbf{b}} \in C(A)$.

So it's worth looking at $C(A)$ a bit more, and ask the question:

"Are the columns of A independent?"

We want to build $C(A)$, taking linear combinations of columns, and then ask if every column contributes something here, does it reach where none of the other columns reach? Or, does one of the columns contribute nothing extra, could we throw it out and still have the same column space?

The answer here is yes to the second question: we *can* throw out one of the columns, for example the third one, because clearly $C_1 + C_2 = C_3$. So the third column contributes nothing further to the linear combinations $xC_1 + yC_2$, because it is one of them, C_3 is in the same plane as C_1 and C_2. In other words, the whole of $C(A)$ can be generated by those linear combinations $xC_1 + yC_2$ and C_3 contributes nothing. We call C_1 and C_2 **pivot columns**.

It should be clear that there is a choice here: $C_3 - C_2 = C_1$, so we could just as easily throw out C_1 and keep C_2 and C_3 as the columns that generate $C(A)$ through their linear combinations. If C_3 is in the same plane as C_1 and C_2, then we could just as easily say that C_1 is in the same plane as C_2 and C_3, and disregard C_1. The point is independence. There are two independent columns and $C(A)$ is a 2-D subspace of $\mathbb{R}^4$. We will return to this idea later!

18.3 Nullspace

We observed just now that $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ always has a solution for any A: it is $\underline{\mathbf{x}} = \underline{\mathbf{0}}$. An interesting question is: what other solutions are there when $\underline{\mathbf{b}} = \underline{\mathbf{0}}$? This is the **nullspace**, or $N(A)$: all solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$. Unlike the column space, the nullspace depends on the other dimension of A: we saw for a 4×3 matrix like

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix}$$

the column space is a subspace of $\mathbb{R}^4$, but the nullspace is all the solutions $\underline{\mathbf{x}}$ in $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$, and these are from $\mathbb{R}^3$. And we call it a subspace, so let's show it is one.

We already know $\underline{\mathbf{0}}$ is always in $N(A)$ because $A\underline{\mathbf{0}} = \underline{\mathbf{0}}$, for any A. Let $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2 \in N(A)$. Then

$$A(a\underline{\mathbf{x}}_1 + b\underline{\mathbf{x}}_2) = a(A\underline{\mathbf{x}}_1) + b(A\underline{\mathbf{x}}_2) = a\underline{\mathbf{0}} + \mathbf{b}\underline{\mathbf{0}} = \underline{\mathbf{0}}$$

so any linear combination of vectors in $N(A)$ is also in $N(A)$, so it's a subspace.

What other vectors are in $N(A)$? For our matrix, that means finding (x, y, z) to satisfy

$$x \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + y \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + z \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Well, we already know how the columns are related: $C_1 + C_2 = C_3$ so $C_1 + C_2 - C_3 = \underline{\mathbf{0}}$ and a first solution is $(1, 1, -1)$. In fact, any multiple of this works: $2C_1 + 2C_2 - 2C_3 = \underline{\mathbf{0}}$ and so on, so

$$\lambda \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \in N(A) \qquad\qquad \text{for any } \lambda \in \mathbb{R} \; .$$

This is a line through the origin and is, in fact, the whole of the nullspace.

Now go back from $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ to $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ and look at the solutions of

$$A\underline{\mathbf{x}} = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 4 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

Do these solutions form a subspace? Without solving the equation, the answer is still clear. Because $A\underline{\mathbf{0}}$ is always equal to $\underline{\mathbf{0}}$, it can never be equal to $\underline{\mathbf{b}} \neq \underline{\mathbf{0}}$. Therefore $\underline{\mathbf{x}} = \underline{\mathbf{0}}$ is not a solution of this equation. Hence the solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{b}} \neq \underline{\mathbf{0}}$ do not form a subspace, because they don't $-$ can't! $-$ include $\underline{\mathbf{x}} = \underline{\mathbf{0}}$. In this case the solutions are a line in $\mathbb{R}^3$ which does not go through the origin.

For picture, see lectures

As an exercise, confirm that this line and the line forming the nullspace are parallel. This simple observation gives the basis for a key step in solving the general case of $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$. But first, how do we find the nullspace, every time, for any A?

18.3.1 Algorithm for finding the Nullspace

We use Gaussian Elimination to do this, but need to take care about how we handle what previously looked problematic: zeroes in pivot positions. The algorithm needs to be refined a little in terms of how we handle the solutions.

Example 18.1. *Use row operations on*

$$A = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix}$$

to find the nullspace $N(A)$.

Note that the columns and rows are not independent, this plays a role as we will see later:$C_2 = 2C_1$ and $R_1 + R_2 = R_3$. We are looking for solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$, so no need to augment the matrix A with the extra column for elimination: the row operations don't do anything to $\underline{\mathbf{0}}$. Hence we use the first pivot in the $11-$position:

$$A = \begin{pmatrix} \boxed{1} & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix} \underset{\substack{R_2 - 2R_1 \\ R_3 - 3R_1}}{\sim} \begin{pmatrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{pmatrix}$$

and immediately note there is no pivot in the $22-$position. To proceed, we need a pivot in the second row, this is the 2 in the $23-$position, so

$$A \sim \begin{pmatrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & \boxed{2} & 4 \\ 0 & 0 & 2 & 4 \end{pmatrix} \underset{R_3 - R_2}{\sim} \begin{pmatrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & \boxed{2} & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} = U$$

This is the so-called **echelon** form, referring to the staircase shape that results. This is still an upper-triangular matrix, as all entries below the main diagonal are zero. In echelon form, the columns containing the pivots, C_1 and C_3, are still called **pivot columns** and the other columns, C_2 and C_4, are called **free columns**. The number of pivots or pivot columns is called the **rank** of the matrix. In this case, rank$(A) = r = 2$.

Recall that

$$A\underline{\mathbf{x}} = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} + x_3 \begin{pmatrix} 2 \\ 6 \\ 8 \end{pmatrix} + x_4 \begin{pmatrix} 2 \\ 8 \\ 10 \end{pmatrix}$$

and this relationship is not changed by the row operations. Hence each column is associated with a variable in $\underline{\mathbf{x}}$ and so x_1 and x_3 are **pivot variables** while x_2 and x_4 are **free variables.**

The algorithm for finding the nullspace hinges on how we deal with the free variables. We have seen before that a free variable can be any real number. We choose $x_2 = 1$ and $x_4 = 0$. These are substituted into the two equations left in the echelon form:

$$x_1 + 2x_2 + 2x_3 + 2x_4 = 0 \quad \text{and} \quad 2x_3 + 4x_4 = 0$$

giving the values of the pivot variables, $x_3 = 0$ and $x_1 = -2$. We now have one solution, one vector in the nullspace, a **special solution**

$$\underline{\mathbf{x}}_1^s = \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} \in N(A)\,.$$

Note that in terms of columns of A, this says that $-2C_1 + C_2 = \underline{\mathbf{0}}$, as we've already noticed. Also, given that any multiple of a vector in a subspace is also in the subspace, we have $p(-2, 1, 0, 0)^T \in N(A)$ for any real p.

Next we let $x_4 = 1$ and $x_2 = 0$. This gives $x_3 = -2$ and $x_1 = 2$, and we have a second special solution:

$$\underline{\mathbf{x}}_2^s = \begin{pmatrix} 2 \\ 0 \\ -2 \\ 1 \end{pmatrix} \in N(A)$$

and in terms of the columns, this vector says that $2C_1 - 2C_3 + C_4 = \underline{\mathbf{0}}$, which isn't so obvious when looking at A. Any multiple of this vector is also in the nullspace. So all linear combinations of the two vectors are in the nullspace, in fact, all linear combinations of these vectors *are* the nullspace.

This is the algorithm:

- Set the first free variable equal to one, all other free variables equal to zero;
- Find the corresponding values of the pivot variables;
- Any multiple of the resulting vector is in $N(A)$;
- Repeat for each free variable;
- The resulting linear combinations are the nullspace.

The nullspace in our example, is therefore given by

$$\underline{\mathbf{x}}_n = p\underline{\mathbf{x}}_1^s + q\underline{\mathbf{x}}_2^s = p \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + q \begin{pmatrix} 2 \\ 0 \\ -2 \\ 1 \end{pmatrix}$$

where $p, q \in \mathbb{R}$.

Some observations:

1) The number of free variables and pivot variables is the number of variables, n for an $m \times n$ matrix A. Here $n = 4$. So the number of free variables is always $n - r$ where r =rank(A).

2) The number $r = 2$ tells us that of the three equations involved in the original $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$, only two contributed. The system had three equations, but as we observed at the beginning, the third row of A is the sum of the first two rows, which says that the third equation contributes nothing, confirmed by $r = 2$ and by the echelon form of A which gave us only two equations.

3) The nullspace of solutions has **dimension** $n - r$, by which we mean that we have r pivot variables giving $n - r$ free variables, each one giving an independent solution vector, a special solution, and the linear combinations of these form the nullspace. The nullspace here is 2D: algebraically like a plane, but in $\mathbb{R}^4$. A line through $\underline{\mathbf{0}}$, in whatever space, is a 1D subspace. There are also 3D and 4D subspaces of higher-dimensional spaces, but they're harder to visualize. Algebraically they behave the same: a plane has 2 independent directions, in any space. So a 3D subspace will have 3 independent directions, and a general $r-$dimensional subspace can be thought of as the linear combinations of r independent vectors.

4) The column space $C(A)$ is also made up of linear combinations, in this case of the columns of A. There are r independent columns, so $C(A)$ has dimension r; in our example also 2D. But we note that $C(A)$ is a subspace of $\mathbb{R}^3$, while $N(A)$ is a subspace of $\mathbb{R}^4$.

5) In general, for an $m \times n$ matrix, $C(A)$ is a subspace of $\mathbb{R}^m$, while $N(A)$ is a subspace of $\mathbb{R}^n$.

We are now ready to combine what we have done with nullspace and column space and tackle the complete solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$.

18.4 Solving linear equations: general case

We finally turn to finding the solutions in the general case $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$, but as usual, there's something else to be established first: do solutions even exist? We've seen that $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ always has a solution, which we call the nullspace. But we've also seen $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ where there is no solution, and established that we need $\underline{\mathbf{b}}$ to be in the column space $C(A)$ for solutions to exist. But we can't just stare at A and $\underline{\mathbf{b}}$ and say yes or no, unless there's some obvious case, like when $\underline{\mathbf{b}}$ *is* one of the columns, or maybe the sum of two columns. But in general, and for larger matrices, we need something better.

18.4.1 Solvability

It turns out that this is not hard to establish with – once again – Gaussian elimination, and we return to the example we had before:

Example 18.2. *Establish a condition on* $\underline{\mathbf{b}}$ *so that* $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ *has solutions, for*

$$A = \begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix}$$

Begin with the augmented matrix $(A : \underline{\mathbf{b}})$ and use the same row operations:

$$A = \begin{pmatrix} \boxed{1} & 2 & 2 & 2 & | & b_1 \\ 2 & 4 & 6 & 8 & | & b_2 \\ 3 & 6 & 8 & 10 & | & b_3 \end{pmatrix} \underset{\substack{R_2-2R_1 \\ R_3-3R_1}}{\sim} \begin{pmatrix} \boxed{1} & 2 & 2 & 2 & | & b_1 \\ 0 & 0 & 2 & 4 & | & b_2 - 2b_1 \\ 0 & 0 & 2 & 4 & | & b_3 - 3b_1 \end{pmatrix} \underset{R_3 - R_2}{\sim} \begin{pmatrix} \boxed{1} & 2 & 2 & 2 & | & b_1 \\ 0 & 0 & \boxed{2} & 4 & | & b_2 - 2b_1 \\ 0 & 0 & 0 & 0 & | & b_3 - b_2 - b_1 \end{pmatrix}$$

It's easy to spot when the system has a solution: the last row represents the equation

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = b_3 - b_2 - b_1$$

and solutions exist **only if** $b_3 - b_2 - b_1 = 0$.

How did we get here? Row operations on the $A-$side produced a zero row and the same row operations on $\underline{\mathbf{b}}$ produced $b_3 - b_2 - b_1$. In words, if some combination of the rows of A produces the zero row and the same combination of the entries of $\underline{\mathbf{b}}$ gives zero, then the system $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ is called **consistent** and has a solution: this is the **solvability condition**.

Note that we have the same row operations here and for the nullspace, so it makes sense to find both in one go. Find the echelon form of $(A : \underline{\mathbf{b}})$ and proceed.

18.4.2 Particular solution and nullspace

Once we've established that a system is indeed solvable, we need a **particular solution**, let's call it x_p which satisfies $A\underline{\mathbf{x}}_p = \underline{\mathbf{b}}$. Then the solution is given by the sum of the particular solution and all vectors in the nullspace.[1] In other words if our nullspace is the linear combinations of the special solutions, then the full solution is

$$\underline{\mathbf{x}} = \underline{\mathbf{x}}_p + \underline{\mathbf{x}}_n = \underline{\mathbf{x}}_p + c_1\underline{\mathbf{x}}_1^s + c_2\underline{\mathbf{x}}_2^s + c_3\underline{\mathbf{x}}_3^s + \cdots$$

Here's the complete algorithm to solve the general case of

$$A\underline{\mathbf{x}} = \underline{\mathbf{b}}\,,$$

- Find the echelon form of the augmented coefficient matrix;
- Establish the system is solvable;
- Solve the homogenous equation $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ for the nullspace ; $\underline{\mathbf{x}}_n = c_1\underline{\mathbf{x}}_1^s + c_2\underline{\mathbf{x}}_2^s + c_3\underline{\mathbf{x}}_3^s + \cdots$;
- Find a particular solution of $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$, call this $\underline{\mathbf{x}}_p$;
- The solution is the sum of particular solution and nullspace:

$$\underline{\mathbf{x}} = \underline{\mathbf{x}}_p + \underline{\mathbf{x}}_n = \underline{\mathbf{x}}_p + c_1\underline{\mathbf{x}}_1^s + c_2\underline{\mathbf{x}}_2^s + c_3\underline{\mathbf{x}}_3^s + \cdots$$

To finish our example we need to choose a vector $\underline{\mathbf{b}}$ which meets the solvability condition, say $\underline{\mathbf{b}} = (1, 5, 6)^T$ given that $6 - 5 - 1 = 0$, so we're ok. Finally,

Example 18.3. *Solve*

$$\begin{pmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix} \underline{\mathbf{x}} = \begin{pmatrix} 1 \\ 5 \\ 6 \end{pmatrix}.$$

In echelon form, the augmented matrix now looks like this:

$$(A : \underline{\mathbf{b}}) \sim \begin{pmatrix} \boxed{1} & 2 & 2 & 2 & | & 1 \\ 0 & 0 & \boxed{2} & 4 & | & 3 \\ 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}$$

All we need is *one* solution: remember the free variables, here x_2 and x_4. If we can choose any value we want for these, we can always choose them to be zero, so let $x_2 = x_4 = 0$, and find the value of the pivot variables. The equations represented by the two non-zero rows are now

$$x_1 + 2x_2 + 2x_3 + 2x_4 = 1 \quad \text{and} \quad 2x_3 + 4x_4 = 3 \quad \Rightarrow \quad x_1 + 2x_3 = 1 \quad \text{and} \quad 2x_3 = 3\,.$$

So $x_3 = 3/2$ and $x_1 = -2$ giving a the particular solution

$$x_p = \begin{pmatrix} -2 \\ 0 \\ 3/2 \\ 0 \end{pmatrix}.$$

In our example, we've already calculated the nullspace, so the solution is

$$\underline{\mathbf{x}} = \begin{pmatrix} -2 \\ 0 \\ 3/2 \\ 0 \end{pmatrix} + p\begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + q\begin{pmatrix} 2 \\ 0 \\ -2 \\ 1 \end{pmatrix},$$

[1] Recall the solution of linear ODEs: particular integral x_p and complementary function x_c, where x_c is a linear combination of independent functions that solve the homogenous ODE. This is no coincidence: there is an interesting way of thinking of functions as vectors with infinitely many entries. The ideas in vector spaces become even more useful. If interested, see for example the theory of Hilbert spaces.

for any $p, q \in \mathbb{R}$.

What is this object? We already know the nullspace is a 2-D subspace of $\mathbb{R}^4$, a plane through the origin. We've taken that plane and added the vector $\underline{\mathbf{x}}_p$, a translation. The result is a plane parallel to the nullspace plane, but through the point $(-2, 0, 3/2, 0)$. So it's not a subspace, though geometrically they are similar. Remember the earlier example, where the solution and the nullspace were parallel lines: one through the origin, one not.

For picture, see lectures

We've asserted the method as a recipe, but proving it is not hard. We'll assume the nullspace consists of two special solutions, but it's easy to generalize for more. First, we show that this is a solution. It's clear that

$$A\underline{\mathbf{x}} = A(\underline{\mathbf{x}}_p + \underline{\mathbf{x}}_n) = A\underline{\mathbf{x}}_p + A\underline{\mathbf{x}}_n = \underline{\mathbf{b}} + \underline{\mathbf{0}} = \underline{\mathbf{b}}$$

so the sum of particular solution and nullspace does indeed solve the equation. But how do we know there isn't another one? How do we know these are *all* the solutions?

Assume the solutions are $\underline{\mathbf{x}} = \underline{\mathbf{x}}_p + a\underline{\mathbf{x}}_1^s + b\underline{\mathbf{x}}_2^s$ satisfying $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$, so we have

$$A\underline{\mathbf{x}} = A(\underline{\mathbf{x}}_p + a\underline{\mathbf{x}}_1^s + b\underline{\mathbf{x}}_2^s) = \underline{\mathbf{b}}$$

Now assume there is another solution, not included in $\underline{\mathbf{x}}$, say $\underline{\mathbf{y}}$. Then we must have

$$A\underline{\mathbf{y}} = \underline{\mathbf{b}} = A(\underline{\mathbf{x}}_p + a\underline{\mathbf{x}}_1^s + b\underline{\mathbf{x}}_2^s) \quad \Rightarrow \quad A\underline{\mathbf{y}} - A(\underline{\mathbf{x}}_p + a\underline{\mathbf{x}}_1^s + b\underline{\mathbf{x}}_2^s) = \underline{\mathbf{0}} \quad \Rightarrow \quad \mathbf{A}(\underline{\mathbf{y}} - \underline{\mathbf{x}}_\mathbf{p} - \mathbf{a}\underline{\mathbf{x}}_\mathbf{1}^\mathbf{s} - \mathbf{b}\underline{\mathbf{x}}_\mathbf{2}^\mathbf{s}) = \underline{\mathbf{0}}$$

so $\underline{\mathbf{y}} - \underline{\mathbf{x}}_p - a\underline{\mathbf{x}}_1^s - b\underline{\mathbf{x}}_2^s$ is in the nullspace. So there exist constants c, d such that

$$\underline{\mathbf{y}} - \underline{\mathbf{x}}_p - a\underline{\mathbf{x}}_1^s - b\underline{\mathbf{x}}_2^s = c\underline{\mathbf{x}}_1^s + d\underline{\mathbf{x}}_2^s \quad \Rightarrow \quad \underline{\mathbf{y}} = \underline{\mathbf{x}}_p + f\underline{\mathbf{x}}_1^s + g\underline{\mathbf{x}}_2^s$$

for some $f, g \in \mathbb{R}$. In other words, $\underline{\mathbf{y}}$ is a sum of the particular solution and some combination of the special solutions! But the solution $\underline{\mathbf{x}}_p + c_1\underline{\mathbf{x}}_1^s + c_2\underline{\mathbf{x}}_2^s$ includes **all** linear combinations of the special solutions, including $f\underline{\mathbf{x}}_1^s + g\underline{\mathbf{x}}_2^s$. So $\underline{\mathbf{y}}$ is one of the solutions we already had, and these are therefore *all* the solutions.

That's the recipe in general, but there's a lot of mileage in the interpretation of some of these results and in the next chapter we look more closely at rank.

Chapter 19

LA IV: Independence, Basis and the Four Fundamental Subspaces

We noticed in the last chapter how so many things involve the rank of a matrix, and begin with a closer examination of this idea.

19.1 Column and Row rank

We begin with an $m \times n$ matrix A, with m rows and n columns. Let the rank be r. Obviously, there can be at most n pivots, because a column can't have more than one pivot. Similarly, there can be at most m pivots, because a row also can have at most one pivot. The number of pivots is the rank, so we have $r \leq n$ and $r \leq m$. We're especially interested in what happens when r is maximal, when it is as big as possible, given A.

19.1.1 More rows than columns

Suppose we have a "tall" matrix: more rows than colums, such as

$$A = \begin{pmatrix} 1 & 3 \\ 2 & 1 \\ 6 & 1 \\ 5 & 1 \end{pmatrix}$$

which is 4×2 and elimination gives

$$A \sim \begin{pmatrix} \boxed{1} & 3 \\ 0 & \boxed{1} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

We introduce the so-called **reduced echelon form** in which all pivots are equal to 1 and entries above pivots are equal to zero. Matlab has the command rref(A) for this. For A, the reduced echelon form R is

$$A \sim \begin{pmatrix} \boxed{1} & 0 \\ 0 & \boxed{1} \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = R$$

In this case, we have $r = n = 2$, its maximal value, where every column has a pivot, also called **full column rank**. A few observations:

- There are no free columns, so no free variables;
- The special solutions that make up the nullspace arise from free variables, and in this case there aren't any: the only solution to $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ is $\underline{\mathbf{0}}$, and the nullspace $N(A)$ contains only the zero vector;

- Therefore, if $\underline{\mathbf{b}}$ is not in the column space of A, the system has no solutions or
- If $\underline{\mathbf{b}}$ is in $C(A)$, then the system $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has a solution: it is the particular solution $\underline{\mathbf{x}}_p$, a unique solution;
- R has the identity matrix at the top, with rows of zeroes below;
- The columns in this case are independent, a concept we've mentioned a few times, and to which we will return.

19.1.2 More columns than rows

Suppose we have a "flat" matrix, with more columns than rows, such as

$$A = \begin{pmatrix} \boxed{1} & 2 & 6 & 5 \\ 3 & 1 & 1 & 1 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & 2 & 6 & 5 \\ 0 & -5 & -17 & -14 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & 2 & 6 & 5 \\ 0 & \boxed{1} & 17/5 & 14/5 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & 0 & -4/5 & -3/5 \\ 0 & \boxed{1} & 17/5 & 14/5 \end{pmatrix} = R$$

This time we have $r = m = 2$, again its maximal value, called **full row rank**. Again, some observations:

- There are $n - m = 2$ free variables and free columns;
- The Nullspace is given by the linear combinations of $n - m$ special solution vectors, two of them;
- R has no rows of zeroes, so $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has solutions for every $\underline{\mathbf{b}}$;
- The system always has solutions, infinitely many: $\underline{\mathbf{x}}_p + \underline{\mathbf{x}}_n$;
- R has the identity matrix on the left, with non-zero columns on the right, the free columns, call therse F;
- There are two independent rows.

In a previous example, we had a "flat" matrix with 2 pivot columns and 2 free columns, but they were not ordered nicely as above, recall

$$A = \begin{pmatrix} \boxed{1} & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & 2 & 2 & 2 \\ 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & 2 & 0 & -2 \\ 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix} = R$$

This example doesn't have full row rank, but the point here is that the pivot columns will always be columns of the identity matrix and if we want to, then the order of the unknowns can be rearranged in the original system, so that the pivot columns are together and contain the identity matrix, on the left of R, with the free columns F on the right.

19.1.3 Square matrix A: same number of rows and columns

We now have $m = n$ and again consider the case where rank r is maximal: $r = m = n$, this is full row rank and full column rank, so we just call it **full rank**. The square matrix might be

$$A = \begin{pmatrix} \boxed{1} & 2 \\ 3 & 1 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & 0 \\ 0 & \boxed{1} \end{pmatrix} = R = I$$

We observe:

- A is invertible;
- There are no free variables and the nullspace is only the zero vector;
- The system $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ always has a solution, the unique solution $\underline{\mathbf{x}} = A^{-1}\underline{\mathbf{b}}$;
- R is the identity matrix.

The last observation is what we used implicitly, when we inverted matrices with elimination earlier, using row operations to get $(A : I) \sim (I : A^{-1})$ where after row operations we had the identity on the left, this was the reduced echelon form, though we didn't have that name then.

19.1.4 Summary

What we've seen is that rank tells you everything we want to know about the nature of the solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$:

$r = m = n$	$r = n < m$	$r = m < n$	$r < m$ and $r < n$
full rank	full column rank	full row rank	
square	tall	flat	any of these
$R = I$	$R = \begin{pmatrix} I \\ 0 \end{pmatrix}$	$R = (IF)$	$R = \begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix}$
1 sol	0 or 1 sol	∞ many sols	0 or ∞ many sols

19.2 Independence, span, basis, and dimension

We don't begin with independence, but rather its opposite: dependence. We have some vectors $\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \underline{\mathbf{x}}_3, \ldots, \underline{\mathbf{x}}_n$ and look at linear combinations of them:

$$c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + c_3\underline{\mathbf{x}}_3 + \ldots + c_n\underline{\mathbf{x}}_n$$

and we ask: is there a linear combination that gives me the zero vector? Can I find coefficients $c_1 \ldots c_n$ so that

$$c_1\underline{\mathbf{x}}_1 + c_2\underline{\mathbf{x}}_2 + c_3\underline{\mathbf{x}}_3 + \ldots + c_n\underline{\mathbf{x}}_n = \underline{\mathbf{0}}\ ?$$

Of course there is: let all the coefficients be zero: $c_1 = c_2 = \ldots c_n = 0$, that's always the case. So what we really want to know here: is there a non-zero combination of the vectors that gives the zero vector? For example in the matrix

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{pmatrix}$$

the columns, which we know are vectors, add up: $C_1 + C_2 = C_3$ or:

$$1\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + 1\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - 1\begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} = \underline{\mathbf{0}}$$

In that case we say the column vectors are **linearly dependent** because we found a non-zero combination of the vectors that gave us $\underline{\mathbf{0}}$. If that's not possible, if the *only* combination that gives the zero vector, is when we multiply every vector by zero:

$$0\underline{\mathbf{x}}_1 + 0\underline{\mathbf{x}}_2 + \ldots 0\underline{\mathbf{x}}_n = \underline{\mathbf{0}}$$

then we say the vectors are **linearly independent**.

Let's try this with some pictures, begin in $\mathbb{R}^2$.

for pictures, see lectures.

Here are two parallel vectors $\underline{\mathbf{v}}_1$ and $\underline{\mathbf{v}}_2$. They are linearly dependent because $\underline{\mathbf{v}}_2 = 2\underline{\mathbf{v}}_1$ so $2\underline{\mathbf{v}}_1 - \underline{\mathbf{v}}_2 = \underline{\mathbf{0}}$.

Or take $\underline{\mathbf{v}}_1$ and the zero-vector. We can always write $0\underline{\mathbf{v}}_1 + 23\underline{\mathbf{0}} = \underline{\mathbf{0}}$ and see linear dependence. So if the zero vector is in the set, it's always linearly dependent.

So taking two non-parallel vectors, $\underline{\mathbf{v}}_1$ and $\underline{\mathbf{v}}_2$ we can see that they are linearly independent, because no linear combination will give us $\underline{\mathbf{0}}$. Now add a third vector, not parallel to either $\underline{\mathbf{v}}_1$ or $\underline{\mathbf{v}}_2$, we get linear dependence, clear in the picture, because we can always multiply the vectors by the right amount to construct a triangle.

How does this relate to matrices? We construct a matrix using the vectors as columns. Guessing at the vectors, we might have

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & -1 \end{pmatrix}$$

and the linear combination is given by

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 2 & -1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \underline{\mathbf{0}} = \mathbf{c_1} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \mathbf{c_2} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \mathbf{c_3} \begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

But that's saying that we are looking for the coefficient vector in the nullspace. And we already know that for a flat matrix, with more columns than rows, the nullspace always has infinitely many vectors, including a non-zero vector. In other words, saying that some vectors are dependent is the same thing as saying the matrix with those vectors as columns has more than $\underline{\mathbf{0}}$ in its nullspace!

In general, take vectors $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$ and make them into the columns of a matrix:

$$A = (\underline{\mathbf{v}}_1\, \underline{\mathbf{v}}_2\, \ldots \underline{\mathbf{v}}_n)$$

If $N(A) = \underline{\mathbf{0}}$, the vectors are linearly independent. If $N(A) \neq \underline{\mathbf{0}}$, which means that $N(A)$ has non-zero vectors, then the vectors are linearly dependent.

We can rephrase this in terms of rank. Let the vectors have length m, and we take n vectors, then$A$ is $m \times n$. If $m < n$, as in the above example of a flat matrix, we immediately have linear dependence.

So let $m > n$, a tall matrix, more rows than columns, or $m = n$ a square matrix. We already know that when we have full column rank, $n = r$, then every column is a pivot column, there are no free variables, and the nullspace is only $\underline{\mathbf{0}}$: the only solution to $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ is the zero vector. So the column vectors are linearly independent.

Still with $m \geq n$, what if we don't have full colum rank? If $r < n$, not every column has a pivot, so there are free variables, and the nullspace has infinitely many vectors including a non-zero vector, the equation $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ has a non-zero solution, giving a non-zero combination of the columns to equal $\underline{\mathbf{0}}$, so the column vectors are linearly dependent.

The idea of vectors that span a space ties in with independene and other ideas, but first we need to define it. We say that the set of vectors $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$ **span** a vector space if the vector space consists of all linear combinations of those vectors. We've already seen this: the columns of a matrix span the subspace called column space.

If we have a set of vectors $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$ and take all their linear combinations and call that a space S, then that's the *smallest* subspace which contains those vectors. We could take a bigger subspace which also contains those combinations of vectors, but some others that aren't combinations of $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$. Then we can no longer say that $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$ span that space. In that sense, the space spanned by $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$ is the smallest space containing $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$.

We are interested in span and dependence. So take a matrix A with n columns and m rows and form its column space $C(A)$. Then we know that the columns of A span the subspace $C(A)$. But are the columns dependent or independent? That depends on the columns, we have seen both are possible. We would very much like a set that is independent and which spans some space. The number of vectors needs to be just right: if there aren't enough vectors, we might not get to the whole of the space. If there are too many, they might be dependent, as we saw in the $\mathbb{R}^2$ example. This leads to the idea of **basis**.

Definition: A basis of a vector space V is a set of vectors $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$ with two properties:
(i) $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$ are linearly independent;
(ii) $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n$ span the vector space V.

Example 19.1. *A basis for* $\mathbb{R}^3$ *is given by the vectors* $\underline{\mathbf{i}}, \underline{\mathbf{j}}, \underline{\mathbf{k}}$*:*

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Are they independent? We could form

$$c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \underline{\mathbf{0}}$$

and solve the three equations to give $c_1 = c_2 = c_3 = 0$. Or we could form them into a matrix, taking these vectors as columns, we get identity matrix and ask, what is the nullspace? What are the solutions of $I\underline{\mathbf{x}} = \underline{\mathbf{0}}$? Clearly, that is just $\underline{\mathbf{x}} = \underline{\mathbf{0}}$, no calculations needed. And if the nullspace is only the zero vector, we have the same thing again: independence.

To see that they span the space is almost trivial. If these vectors $\underline{\mathbf{i}}, \underline{\mathbf{j}}, \underline{\mathbf{k}}$ span $\mathbb{R}^3$, then every vector in $\mathbb{R}^3$ must be a linear combination of $\underline{\mathbf{i}}, \underline{\mathbf{j}}, \underline{\mathbf{k}}$. So take any vector $\underline{\mathbf{v}} = (a, b, c) \in \mathbb{R}^3$ and we can write $\underline{\mathbf{v}} = a\underline{\mathbf{i}} + b\underline{\mathbf{j}} + c\underline{\mathbf{k}}$ and we have the desired linear combination.

The vectors $\underline{\mathbf{i}}, \underline{\mathbf{j}}, \underline{\mathbf{k}}$ are linearly independent and span $\mathbb{R}^3$, so they form a basis.

Example 19.2. *Take the vectors*

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

Do these form a basis for $\mathbb{R}^3$*? If not, find a basis including these.*

Clearly they are independent, two vectors are dependent only if they are parallel, and these aren't parallel. But do they span $\mathbb{R}^3$? Rather than trying to form an arbitrary vector (a, b, c) as a linear combination, let's look at all the linear combinations:

$$p \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + q \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

and observe that this is a plane, as we've seen before. It's not the whole of $\mathbb{R}^3$, so these two vectors don't span $\mathbb{R}^3$, we need another one, a third independent direction. If we were to add the vector $(3, 4, 5)^T$ this wouldn't work, because this one is clearly dependent on the first two, so we'd lose independence...and we'd still be in the same plane! So try

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 \\ 4 \\ 6 \end{pmatrix}$$

It feels as if this should be a basis, but how do we test it? Form a matrix from the three vectors, we get a square matrix A. If it's invertible, then $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ has only one solution, namely $\underline{\mathbf{0}}$, so the vectors are independent. Do they also span $\mathbb{R}^3$? Can we say, for any vector $\underline{\mathbf{b}} \in \mathbb{R}^3$, that $\underline{\mathbf{b}}$ is a linear combination of those vectors, can we find c_1, c_2, c_3 so that

$$c_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} + c_3 \begin{pmatrix} 3 \\ 4 \\ 6 \end{pmatrix} = \underline{\mathbf{b}}\,?$$

That's the same thing as finding $\underline{\mathbf{c}} = (c_1, c_2, c_3)^T$ so that $A\underline{\mathbf{c}} = \underline{\mathbf{b}}$. But if the matrix is invertible, then $\underline{\mathbf{c}} = A^{-1}\underline{\mathbf{b}}$ and we have the required linear combination, and the vectors span $\mathbb{R}^3$. In general, we can say that a set of vectors $\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n \in \mathbb{R}^n$ forms a basis for $\mathbb{R}^n$ if and only if the matrix which has those vectors as columns is invertible.

In both cases, the basis we found for $\mathbb{R}^3$ had three vectors. Could a basis have four? No, because if we form the 3×4 matrix A with those four column vectors in $\mathbb{R}^3$, then we have a flat matrix and $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$ has non-zero solutions, so they're not independent anymore. In any $\mathbb{R}^n$ you can have *at most* n independent vectors.

On the other hand, as we have seen, we need n vectors. In $\mathbb{R}^3$, two vectors spanned only a plane, they don't span the whole of $\mathbb{R}^3$, a third, independent direction is needed. Another way of expressing this is

(i) a basis is a **minimal** spanning set for V: take away any vector and you no longer have a spanning set.

(ii) a basis is a **maximal** linearly independent set: add any other vector and you no longer have linear independence.

This is the idea of **dimension**: the number of linearly independent vectors that span a space. The dimension of $\mathbb{R}^3$ is 3. The dimension of a plane is 2, because we need two vectors to span the plane, and any more would be linearly dependent. Each basis has infinitely many bases, but every basis has the same number of vectors: this is the dimension.

Note, that the mathematical definition of dimension is consistent with our informal understanding of a plane as a 2-dimensional thing, and the space we inhabit, mathematically equivalent to $\mathbb{R}^3$, as 3-dimensional.

Example 19.3. *Find the dimension and a basis for the column space $C(A)$, where*

$$A = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{pmatrix}$$

So we look at the columns of A, we know already that they span the column space, because that's how $C(A)$ was constructed, by taking all linear combinations of the columns. But are they independent? No: the fourth column is equal to the first column. But if it wasn't that obvious, how would we work it out? We could find the nullspace: does it have a non-zero element? So if we can't spot the dependence, we know how to find the nullspace and answer that yes, these vectors are dependent. Or recall that you can have at most 3 independent vectors in $\mathbb{R}^3$ and here we have 4: again, they must be dependent.

To find the basis for the column space, we need to find independent columns. We look at the first two columns

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

They are clearly independent. Do they span the column space? If we add the third column we get linear dependence:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 3 \end{pmatrix}$$

so that's no good: the third column is in the space spanned by the first two, a plane . If we add the fourth column, again no good, as it's equal to the first column. So the column space is spanned by the first two columns, which are independent, so form a basis for $C(A)$. But we could just as easily use the first and the third column, thinking that the second column is in the plane spanned by those two. Or the second and third. In fact, the only pair of columns we can't use is the first and fourth.

But if it's not that obvious, how can we guarantee to identify a full set of independent columns? For a larger matrix, we can't rely on just spotting the obvious dependence. Imagine a 23×42 system where it just so happens that

$$C_1 - 13.5C_6 + 17.3C_{13} + 9.2C_{17} = C_{21} \, ,$$

however would you notice?

There's an easy answer to the key question: Which columns can we choose that are guaranteed to be linearly independent and span the whole of the column space? One set is the pivot columns, in this case the first two columns, as will be clear after one-step elimination:

$$A = \begin{pmatrix} \boxed{1} & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{pmatrix} \sim \begin{pmatrix} \boxed{1} & 2 & 3 & 1 \\ 0 & \boxed{-1} & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

So the first two columns are the pivot columns and a basis for the column space is

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

and the dimension of $C(A)$ is 2.

Warning: we take the pivot columns of A before elimination, not afterwards. If we took the two after elimination, all linear combinations of those columns

$$p\begin{pmatrix}1\\0\\0\end{pmatrix} + q\begin{pmatrix}2\\-1\\0\end{pmatrix}$$

would have zero in the third component, which is clearly not going to be the case in a lot of vectors in $C(A)$. For a start, $C(A)$ contains the columns themselves, all of which have non-zero third entries. Elimination merely identifies **which** are the pivot columns.

So now another thing becomes clear: if the pivot columns give a basis, we've already counted them once, when we called the number of pivots the rank of the matrix.

Theorem: The dimension of the column space of a matrix A is equal to the rank:

$$\mathrm{rank}(A) \;=\; \mathrm{dim}C(A)\ .$$

And while we're at it, we have another subspace with a dimension. We saw how the nullspace was obtained as a linear combination of the special solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$, and these corresponded to the free variables, of which there were $n-$rank(A). So these special solutions span the nullspace, and the way they were constructed guarantees linear indenpendence. So we have another

Theorem: The dimension of the nullspace of an $m \times n$ matrix A is equal to the number of columns minus the rank:

$$n - \mathrm{rank}(A) \;=\; \mathrm{dim}N(A)\ .$$

At the end of the previous section, we had a table summarizing the facts we had established about the nature of the solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$. With the notions of independence and basis, we can now extend those facts, in terms of existence and uniqueness of solutions, with the following

Theorem: Existence and Uniqueness of Solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$

Let A be an $m \times n$ matrix. Then

(i) [*Existence*] The following statements are equivalent:

- $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has a solution for all $\underline{\mathbf{b}} \in \mathbb{R}^m$;
- The column space of A is the entire $\mathbb{R}^m$;
- rank$(A) = m$;
- The rows of A are linearly independent: A has full row rank.
- If $m = n$, the rows of A form a basis for $\mathbb{R}^n$.

Otherwise, the following statements are equivalent:

- $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has a solution for some, but not all $\underline{\mathbf{b}} \in \mathbb{R}^m$;
- The column space of A is subspace of $\mathbb{R}^m$, not equal to $\mathbb{R}^m$;
- rank$(A) < m$;
- The rows of A are linearly dependent: A doen not have full row rank.
- If $m = n$, the rows of A do not form a basis for $\mathbb{R}^n$.

The above is only possible for $m \leq n$: flat/square matrix.

(ii)[*Uniqueness*] The following statements are equivalent:

- If $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ has a solution, it is unique;
- The null space of A is the zero vector;

- rank$(A) = n$;
- The columns of A are linearly independent: A has full column rank.
- If $m = n$, the columns of A form a basis for $\mathbb{R}^n$.

Otherwise, the following statements are equivalent:

- $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ no solution or infinitely many solutions;
- The null space of A contains more than the zero vector;
- rank$(A) < n$;
- The columns of A are linearly dependent: A does not have full column rank.
- If $m = n$, the columns of A do not form a basis for $\mathbb{R}^n$.

The above is only possible for $m \geq n$: tall/square matrix.

19.3 Four fundamental subspaces

We've seen how the column space and nullspace are important to determining the nature of the solutions of $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$ as well as finding them. We now look at two more subspaces related to a matrix, which together form the **four fundamental subspaces.** The missing two spaces are found by transposing the matrix A and taking the column and nullspace of this, giving the **row space** and the **left nullspace**.

The third subspace, the row space of A is given by the linear combinations of the rows of A. If we look at the transpose A^T, its columns are the rows of A, so $C(A^T)$, the column space of A^T gives the row space of A. As with columns and $C(A)$, if the rows of A are independent, they form a basis for the row space.

The fourth subspace, where we find the nullspace of A^T, denoted $N(A^T)$, is called the left nullspace of A. We'll see where the name comes from.

If A is $m \times n$, the nullspace consists of solutions to $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$, so $\underline{\mathbf{x}}$ has n components, and so $N(A)$ is a subspace of $\mathbb{R}^n$. Similarly, the column space is made up of columns of A, which have m components, so $C(A)$ is a subspace of $\mathbb{R}^m$. Hence A^T is $n \times m$ and the row space will be a subspace of $\mathbb{R}^m$ while the left nullspace will be a subspace of $\mathbb{R}^n$.

At this point it's instructive to draw a picture of the four subspaces.

for picture, see lectures

For each subspace, we want a basis and the dimension, and we already have these for the first two. For the column space $C(A)$ the basis is the linearly independent colums of A, the pivot columns, and their number, r, the rank of A, gives the number of vectors in the basis, so dim $C(A)$ =rank$(A) = r$. For the nullspace $N(A)$, we know that the basis is given by the special solutions to $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$, and their number is the number of free variables, $n - r$, so dim$N(A) = n - r$.

For the row space, $C(A^T)$, the basis is given by the linearly independent rows of A, corresponding to the pivot columns of A^T. We will find that these are the non-zero rows in R, the reduced echelon form of A. Each of these rows has a pivot, so their number is, again, the rank of A, giving the dimension of the row space, dim$C(A^T) = r$.

For the left nullspace, $N(A^T)$ we reason as follows. In $\mathbb{R}^n$ we have $C(A^T)$ and $N(A)$ with dimension r and $n-r$, respectively. In $\mathbb{R}^m$ we have $C(A)$ and $N(A^T)$ with dimension r and, we expect the symmetry: $m - r$. But not just the symmetry: A^T has m columns, and therefore when solving the equations $A^T\underline{\mathbf{x}} = \underline{\mathbf{0}}$ we will have $m - r$ free variables, giving the number of special solutions, which forms the basis for the nullspace of A^T, as expected.

Let's see how all of these can be found with a single row-reduction operation: taking A to rref$(A) = R$.

Example 19.4. *Obtain the reduced echelon form of*

$$A = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{pmatrix}$$

and use this to obtain bases for the four fundamental subspaces of A.

Gaussian elimination:

$$A = \begin{pmatrix} \boxed{1} & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{pmatrix} \underset{\substack{R_2-R_1 \\ R_3-R_1}}{\sim} \begin{pmatrix} \boxed{1} & 2 & 3 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \underset{-R_2}{\sim} \begin{pmatrix} \boxed{1} & 2 & 3 & 1 \\ 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \underset{R_1-2R_2}{\sim} \begin{pmatrix} \boxed{1} & 0 & 1 & 1 \\ 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = R = \begin{pmatrix} I & F \\ 0 & 0 \end{pmatrix}$$

1. Column Space We begin with the column space: the first two columns are the pivot columns, so we already had the basis for $C(A)$ as

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

and dim$C(A) = 2 =$rank(A). We saw how $C(A) \neq C(R)$.

2. Row Space For the row-space, however, we argue that the non-zero rows of R are linear combinations of the rows of A. Any linearly dependent rows have been eliminated, in this case the last one. So how many non-zero rows in R? Each one has a pivot, their number is r the rank of A. So the first r rows of R, the reduced echelon form of A will be a basis for the row space, $C(A^T)$. We know these rows are *in* the row space, but do they span it? Is every vector in the row space a linear combination of those first r rows of R? Yes, because every row in A is a linear combination of the rows of R, obtainable by inverting the row operations, and every vector in the row space is a linear combination of the rows of A, therefore every element of $C(A^T)$ can be obtained as a linear combinations of non-zero rows in R. As columns, they are

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

and form the basis of $C(A^T)$, and so dim$C(A^T) = 2 =$ rank(A).

3. Nullspace For the nullspace, recall that we need the special solutions. The free variables in this case are x_3 and x_4, pivot variables are x_1 and x_2. From $A\underline{\mathbf{x}} = \underline{\mathbf{0}}$, the two non-zero rows of $R\underline{\mathbf{x}} = \underline{\mathbf{0}}$ give

$$x_1 + x_3 + x_4 = 0 \quad \text{and} \quad x_2 + x_3 = 0$$

Setting $x_3 = 1\,, x_4 = 0$ and vice-versa, we get the start of the two linearly independent special solutions

$$\begin{pmatrix} \\ \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \\ \\ 0 \\ 1 \end{pmatrix}$$

and solving the equations gives x_1 and x_2 so that

$$\underline{\mathbf{x}}_1^s = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \underline{\mathbf{x}}_2^s = \begin{pmatrix} -1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

as the basis of $N(A)$ with dimension $n - r = 2$

4. Left Nullspace Finally, we look at the left nullspace, the nullspace of the transpose, or $N(A^T)$. Let $\underline{\mathbf{y}}$ be a vector in this space. Then

$$A^T\underline{\mathbf{y}} = \underline{\mathbf{0}}\,.$$

We can transpose the equation:

$$(A^T\underline{\mathbf{y}})^T = \underline{\mathbf{0}}^{\mathbf{T}} \quad \Rightarrow \quad \underline{\mathbf{y}}^{\mathbf{T}}\mathbf{A} = \underline{\mathbf{0}}^{\mathbf{T}}$$

On the right, $\underline{\mathbf{0}}^{\mathbf{T}}$ is a row of zeroes. On the left, we also have a row: the multiplication of A on the left by the row vector $\underline{\mathbf{y}}^T$, hence the name left nullspace. The basis for this is also identifiable in the above sequence of row reduction taking A to R, just not as obvious as before.

Begin with the row operations that transform A, step-by-step, to its reduced echelon form R. We have seen how each row operation can be represented by an elimination, permutation or scaling matrix, multiplying on the left. Multiplied together, they give E so that

$$EA = R\,.$$

As A and R are both $m \times n$, we need the product E of elimination and other matrices to be square $m \times m$. Crucially, we recall how each matrix is obtained by performing the same row operation on the identity. If we operate on the identity in the right order we have

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \underset{\substack{R_2-R_1 \\ R_3-R_1}}{\sim} \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \underset{-R_2}{\sim} \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \underset{R_1 - 2R_2}{\sim} \begin{pmatrix} -1 & 2 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix} = E$$

So

$$EA = \begin{pmatrix} -1 & 2 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} \boxed{1} & 0 & 1 & 1 \\ 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = R$$

We're looking for rows $\underline{\mathbf{y}}^T$ which, when multiplying A on the left give a row of zeroes:

$$\underline{\mathbf{y}}^T A = \underline{\mathbf{0}}^{\mathbf{T}}.$$

Recall that we've seen earlier that when we multiply matrices like EA together to give R, then we can think of the first row of E multiplying A on the left, giving a combination of the rows of A, which then yields the first row of R. We note R has a row of zeroes, the third row. So in terms of multiplication, we can look at

$$(-1, 0, 1) \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{pmatrix} = (0, 0, 0, 0)$$

This is saying that $-1R_1(A) + 0R_2(A) + 1R_3(A) =$ a zero row in R. Check:

$$-1(1, 2, 3, 1) + 0(1, 1, 2, 1) + 1(1, 2, 3, 1) = (0, 0, 0, 0)\,,$$

as expected. We have a combination of rows of A giving the zero row, which is a combination of columns of A^T giving the zero column: this is the combination required for the nullspace of A^T, which we seek.

Hence $\underline{\mathbf{y}}^T$ is the third row of E and if we note that any multiple of this will still give zero, we have our basis for the left nullspace, $(-1, 0, 1)^T$, which now has dim$N(A^T) = 1 = m - r$.

If R had more than one zero row, we would have more than one row of E, giving more than one basis vector.

We've seen a lot of abstract detail regarding vector spaces and subspaces. In the next chapter, and in the next year, we will apply some of these ideas.