What you think you want to know:

The algorithm for solving a system

$$
\begin{array}{rcrcrcrcrcl}
x & + & 3y & + & 2z & + & 10u & - & 7v & = & a \\
2x & + & 6y & + & 6z & + & 28u & - & 20v & = & b \\
x & + & 3y & + & 5z & + & 22u & - & 16v & = & c
\end{array}
$$

$$
\rightarrow \quad \left[\begin{array}{ccccr|c}
1 & 3 & 2 & 10 & -7 & a \\
2 & 6 & 6 & 28 & -20 & b \\
1 & 3 & 5 & 22 & -16 & c
\end{array}\right]
$$

$$\left[\begin{array}{ccccc|c} \mathbf{1} & 3 & 2 & 10 & -7 & a \\ {\color{red}2} & 6 & 6 & 28 & -20 & b \\ {\color{red}1} & 3 & 5 & 22 & -16 & c \end{array}\right]$$

Reverse rows, if necessary (it isn't here), so that the first row has its first (from the left) nonzero entry at least as far left as any other. That is the first pivot. Use it to make the entries below it by row operations. (This is "elimination".)

(Strang doesn't want you to make the pivot a 1. I disagree, unless it introduces fractions. This is already a 1.)

$$\xrightarrow[\boxed{3} - \boxed{1}]{\boxed{2} - 2\boxed{1}}$$

Interlude

Numerical analysts would say pick pivots to be, not just nonzero, but as large (in absolute value) as possible, to minimize roundoff error.

But pivots of 1 make the arithmetic easier to do by hand.

$$\left[\begin{array}{ccccc|c} 1 & 3 & 2 & 10 & -7 & a \\ 0 & 0 & \mathbf{2} & 8 & -6 & b-2a \\ 0 & 0 & \color{red}{3} & 12 & -9 & c-a \end{array}\right]$$

Leave the first row alone. From the remaining rows, pick one that has its first nonzero entry at least as far left as any other, and if necessary switch it into the second row. That is the second pivot. Use it to make the entries below it 0.

(This pivot is not 1; I'll follow Strang's lead, reluctantly.)

$$\xrightarrow{\boxed{3} - (3/2)\boxed{2}}$$

$$\left[\begin{array}{ccccc|r} 1 & 3 & 2 & 10 & -7 & a \\ 0 & 0 & 2 & 8 & -6 & b-2a \\ 0 & 0 & 0 & 0 & 0 & c-a-(3/2)(b-2a)=c-(3/2)b+2a \end{array}\right]$$

If there were more nonzero rows, we would go on to find more pivots, each further to the right as we go down. But there are no more nonzero rows (in the coefficient part of the matrix), so there are no more pivots. The coefficient matrix is now in *(regular) echelon form:*

- pivots are in the first few rows,
- further to the right as we go down, and
- any rows of all zeros come at the end.

We can now read off: The system has a solution exactly when the constants satisfy the equation

$$c-(3/2)b+2a=0\ .$$

$$\left[\begin{array}{ccccc|r} 1 & 3 & 2 & 10 & -7 & a \\ 0 & 0 & 2 & 8 & -6 & b-2a \\ 0 & 0 & 0 & 0 & 0 & c-(3/2)b+2a \end{array}\right]$$

Suppose we set $a = 5$, $b = 4$, $c = -4$, satisfying $c - (3/2)b + 2a = 0$:

$$\left[\begin{array}{ccccc|r} 1 & 3 & 2 & 10 & -7 & 5 \\ 0 & 0 & \mathbf{2} & 8 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Put the coefficient matrix into *reduced row echelon form*:

- pivots turned into 1's ("leading 1's")
- with 0's above them

(working right to left and eliminating up).

$$\xrightarrow[\boxed{1} - 2\boxed{2}]{\boxed{2} \times (1/2)}$$

$$\left[\begin{array}{ccccc|c} 1 & 3 & 0 & 2 & -1 & 11 \\ 0 & 0 & 1 & 4 & -3 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

5 variables − 2 (useful) equations = 3 "degrees of freedom"

For easier algebra:

- *Free variables* (with arbitrary values) are the ones without pivot coefficients;
- write the others in terms of them.

So:

- Free variables: y, u, v
- From $z + 4u - 3v = -3$, get $z = -3 - 4u + 3v$.
- From $x + 3y + 2u - v = 11$, get $x = 11 - 3y - 2u + v$.

Before we look again at the last example, let's do a different example, with more than 2 pivots, of the process of going from echelon form to reduced row echelon form:

$$\begin{bmatrix} 2 & 1 & -3 & 4 \\ 0 & 3 & 9 & 6 \\ 0 & 0 & 0 & \mathbf{5} \end{bmatrix} \xrightarrow{\substack{\boxed{3} \times (1/5) \\ \boxed{1} - 4\boxed{3} \\ \boxed{2} - 6\boxed{3}}} \begin{bmatrix} 2 & 1 & -3 & 0 \\ 0 & \mathbf{3} & 9 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{\substack{\boxed{2} \times (1/3) \\ \boxed{1} - \boxed{2}}} \begin{bmatrix} \mathbf{2} & 0 & -6 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\boxed{1} \times (1/2)} \begin{bmatrix} 1 & 0 & -3 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

What you <u>really</u> want to know:

The relevant concepts
(and terms and notations)

Now let's go back to the first example. Strang consistently uses A for the coefficient matrix and $\mathbf{b}$ for the vector of constant terms:

$$A = \begin{bmatrix} 1 & 3 & 2 & 10 & -7 \\ 2 & 6 & 6 & 28 & -20 \\ 1 & 3 & 5 & 22 & -16 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Note two b's are in different fonts; one is a vector, the other a scalar (one of the components of the vector). The text uses subscripts.

He uses U for the regular echelon form of the coefficient matrix, and occasionally $\mathbf{c}$ for the new vector of constant terms:

$$U = \begin{bmatrix} 1 & 3 & 2 & 10 & -7 \\ 0 & 0 & 2 & 8 & -6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} a \\ b - 2a \\ c - (3/2)b + 2a \end{bmatrix}$$

At this point Strang would expect you to substitute values of $a = 5$, $b = 4$, $c = -4$, which gives a specific $\mathbf{c}$:

$$U = \begin{bmatrix} 1 & 3 & 2 & 10 & -7 \\ 0 & 0 & 2 & 8 & -6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} 5 \\ -6 \\ 0 \end{bmatrix}$$

Then he might encourage you to solve the system by back-substitution. This might give you a solution that looks different from the one we found, but it would be the same, for different values of the free variables. Then he would encourage you to put the coefficient matrix into rref form:

$$R = \begin{bmatrix} 1 & 3 & 0 & 2 & -1 \\ 0 & 0 & 1 & 4 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{d} = \begin{bmatrix} 11 \\ -3 \\ 0 \end{bmatrix}$$

Back up: Matrix A is 3×5, so multiplying by it (on the left) takes vectors in $\mathbb{R}^5$ to vectors in $\mathbb{R}^3$; and $A\mathbf{x} = \mathbf{b}$ has solution when $\mathbf{b}$ satisfies $c - (3/2)b + 2a = 0$.

That's when $\mathbf{b}$ is in the column space $\mathbf{C}(A)$ (in $\mathbb{R}^3$).

Only one equation, so $\mathbf{C}(A)$ is a plane (through the origin, like any subspace).

Dimension: geometry → algebra. Step 1

Definition

Let V be a vector space. A *basis* for V is a set of vectors in V that is (linearly) independent and spans V. (In other words, every element of V can be written as a linear combination of the basis vectors in exactly one way.)

Example

The *standard basis* for $\mathbb{R}^3$ is

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

because $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$

Dimension: geometry $\rightarrow$ algebra. Step 2

"Lin ind" means that each new vector goes off in a different direction from the earlier ones, giving a new dimension to the span. So a basis for V is a lin ind set whose span has as many dimensions as possible within V — its span is all of V.

Of course a given vector space has lots of bases. But they all have the same number of vectors in them. (The proof is simple but ugly and uninformative.)

Definition

The number of vectors in any basis for a vector space V is the *dimension* of V.

Spaces (and their dimensions) from $A\mathbf{x} = \mathbf{b}$

First space: $\mathbf{C}(A)$, i.e., $\mathbf{b}$'s for which $A\mathbf{x} = \mathbf{b}$ had a solution.
In our example, $\mathbf{C}(A)$ was the plane $c - (3/2)b + 2a = 0$ — dimension 2.
We can pick a basis, inspired by the standard basis for the plane $\mathbb{R}^2$:

- Let $a = 1$ and $b = 0$, get $c = -2$;
- Let $a = 0$ and $b = 1$, get $c = 3/2$.

Then a basis for $\mathbf{C}(A)$ is

$$\begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 3/2 \end{bmatrix}.$$

Definition

The dimension of the column space $\mathbf{C}(A)$ of a matrix A is called the *rank* of A.

We can find a basis for the column space by picking out the ones that are linearly independent of the ones that came before.

WARNING: Doing row operations will change the columns and the column space of a matrix.

But if one column is a linear combination of earlier ones, row operations won't change that (because they don't change solutions of a system). So independent columns in U or R come from independent columns in A. But in U or R, the pivot columns are clearly the ones that are not combinations of earlier ones. So:

THE RANK OF A MATRIX IS THE NUMBER OF PIVOTS IN ITS ECHELON OR RREF FORM.

rank(A) = rank(A^T)

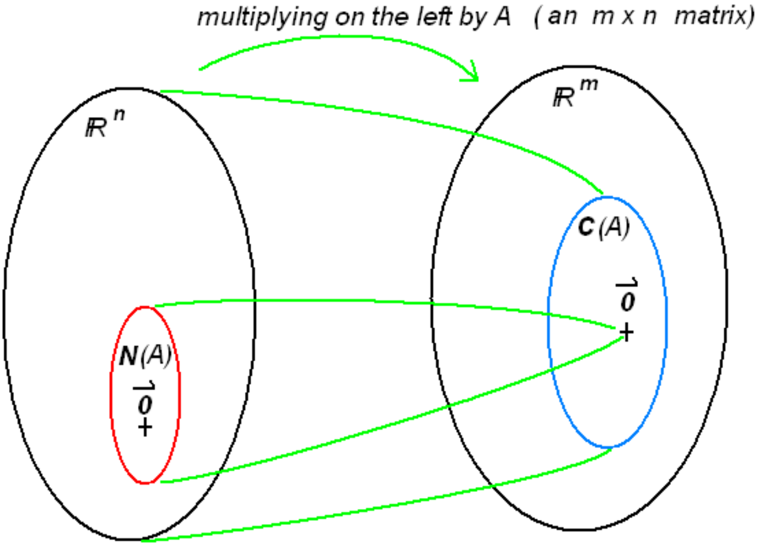
$$\begin{aligned}
\text{rank}(A) &= \text{number of ind columns of } A \\
&= \text{number of pivots in rref form } R \text{ of } A \\
&= \text{number of ind } \underline{\text{rows}} \text{ in } R \\
&= \text{dim of span of rows of } R \\
&= \text{dim of span of rows of } A \\
&\qquad \text{(because row ops don't change row span)} \\
&= \text{number of ind rows of } A \\
&= \text{number of ind } \underline{\text{columns}} \text{ of } A^T \\
&= \text{rank}(A^T)
\end{aligned}$$

Spaces (and their dimensions) from $A\mathbf{x} = \mathbf{b}$ (ctnd)

(As usual, A is $m \times n$.)

If $\mathbf{b}$ isn't the zero vector in $\mathbb{R}^m$, then the set of solutions to $A\mathbf{x} = \mathbf{b}$ is <u>not</u> a vector subspace. (For example, the zero vector in $\mathbb{R}^n$ isn't in it.) BUT ...

Second space: The set of solutions in $\mathbb{R}^n$ to $A\mathbf{x} = \mathbf{0}$ (the zero vector in $\mathbb{R}^m$) is the *nullspace* of A, denoted $\mathbf{N}(A)$.


multiplying on the left by A (an m x n matrix)
ℝ n
ℝ m
N(A)
0
C (A)
0

Example

What is $\mathbf{N}(A)$ for the 3×5 matrix A from earlier?

$$\left[\begin{array}{ccccc|c} 1 & 3 & 2 & 10 & -7 & 0 \\ 2 & 6 & 6 & 28 & -20 & 0 \\ 1 & 3 & 5 & 22 & -16 & 0 \end{array}\right] \longrightarrow \left[\begin{array}{ccccc|c} 1 & 3 & 0 & 2 & -1 & 0 \\ 0 & 0 & 1 & 4 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

so y, u, v are free variables, and

$$z = -4u + 3v, \qquad x = -3y - 2u + v.$$

A general element of that nullspace is $\begin{bmatrix} -3y-2u+v \\ y \\ -4u+3v \\ u \\ v \end{bmatrix}$.

Strang calls this the "complete solution".

3 free variables, so a basis inspired by the standard basis for $\mathbb{R}^3$:

$$\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Strange calls these the "special solutions".

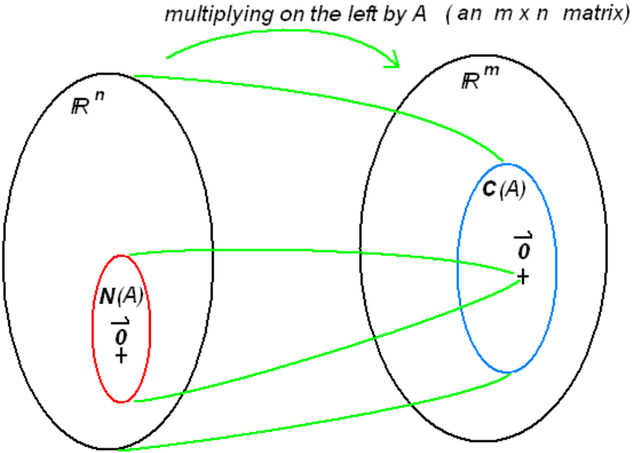
Notice:

$$\begin{array}{cccccl} 1 & 3 & 0 & 2 & -1 & \\ 0 & 0 & 1 & 4 & -3 & \\ 0 & 0 & 0 & 0 & 0 & \\ +1 & & +1 & & & \dim(\mathbf{C}(A)) \\ & +1 & & +1 & +1 & \dim(\mathbf{N}(A)) \end{array}$$

In general:

THE DIMENSION OF THE NULLSPACE OF A MATRIX PLUS THE RANK EQUALS THE NUMBER OF COLUMNS.

In algebra: If an $m \times n$ matrix has rank r, then its nullspace has dimension $n - r$.



I like to think of matrix multiplication as taking the n dimensions of its domain $\mathbb{R}^n$ and collapsing $n - r$ of them, where r is its rank, leaving r of them in its column space inside its range $\mathbb{R}^m$.

Other nonstandard terminology

Strang calls a matrix whose columns form a basis for the nullspace a "nullspace matrix".

Example

For our favorite A, a nullspace matrix is

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

He has Matlab code that computes such a matrix; the code is called "nullbasis". As far as I know, no one else uses these terms.

On the other hand, some books use "nullity" for the dimension of the null space.

Simple facts about rank

A is $m \times n$, $\text{rank}(A) = r$:

- $r \leq m$. $r \leq n$.
- If $r = m$, then $A\mathbf{x} = \mathbf{b}$ has at least one solution for every $\mathbf{b}$.
- If $r = n$, then $A\mathbf{x} = \mathbf{b}$ has at most one solution of a given $\mathbf{b}$. (More about this one later.)
- If $m = n = r$, then A is invertible.