Solving linear equations

1 Solving one equation: review

Given a system of linear equations $Ax = b$, we can solve this equation by Gauß-Jordan elimination. By solve we mean:

- determine if a solution exists
- if so, find all possible solutions, which effectively is:
  - find the kernel of A (solve the homogeneous equation $Ax = 0$)
  - find a particular solution

We do this by reducing the augmented matrix $[A|b]$ to a row-reduced echelon matrix $[A'|b']$ (note that b changes to b' in the process), from which we can read off the answer.

Suppose A' has k rows (that is, A has rank k), and the pivots are in columns $c_1, \ldots, c_k$. Then:

- a solution exists iff b' vanishes after the kth entry:
  i.e., it's $b' = (b'_1, b'_2, \ldots, b'_k, 0, \ldots, 0)$ – this is because otherwise we have a $0 = 1$ row.
- to find all possible solutions
  - the kernel of A is exactly "the pivot equals minus the later terms in that row"
  - a particular solution is "put the b'_i coordinate in the c_ith entry of x (formally, $x_{c_i} = b'_i$).

The above notation is confusing, but the idea is simple – let's illustrate.

For instance, if $[A'|b']$ is

$$\left[\begin{array}{ccc|c} 1 & 2 & & 4 \\ & & 1 & 7 \\ & & & 0 \end{array}\right]$$

(I've left spaces for zeros, for readability), then:

The *pivots*[1] (the leading 1s in some rows) are x_1 and x_3, while x_2 is the *free variable*: we can set it to whatever we want (b/c the pivots can be adjusted to compensate).

The solution is:

- the system is *consistent* (meaning it has a solution), as we have zeros opposite the zero rows
- the solution is given by:
  - the kernel is the subspace of $\mathbf{R}^3$ such that $x_1 + 2x_2 = 0$ (so $x_1 = -2x_2$) and $x_3 = 0$; in general we'd have more equations, but they'd all be of this form: some pivot equals negative the later coordinates in that row (or zero if there aren't any). So $\ker A$ is

$$\begin{bmatrix} -2x_2 \\ x_2 \\ 0 \end{bmatrix}$$

    For clarity, one often writes r, s, t for the free variables, so

$$\ker A = \begin{bmatrix} -2r \\ r \\ 0 \end{bmatrix}$$

  - A specific solution in $x_0 = (4, 0, 7)$ (i.e., $x_1 = 4, x_2 = 0, x_3 = 7$): the pivots tell us where to put the numbers so they'll end up in the right place.

Thus the general solution is:

$$x_0 + \ker A = \begin{bmatrix} 4 \\ 0 \\ 7 \end{bmatrix} + \begin{bmatrix} -2r \\ r \\ 0 \end{bmatrix} = \begin{bmatrix} 4 - 2r \\ r \\ 7 \end{bmatrix}$$

[1] I don't know if this usage is correct: the point is that these are the "unfree" variables.

2 Solving many equations

The above shows how you can solve one particular equation, but what if you want to solve many, like $Ax = b_i$. Maybe every day your boss comes by and gives you a b_i to solve, and you really don't want to have to do everything *again*.

What's the solution? Well, every time you row reduce $[A|b]$, you'll get the same A', but maybe a different b' – so all you need to do is record the sequence of row moves and apply this to each b_i.

Recall that a row operation is just left-multiplication by an elementary matrix, so multiplying these together yields a matrix such that left-multiplication by it does all of the row operations. Call this matrix R, so $RA = A'$, and thus $R[A|b] = [A'|b']$.

So given A, compute R (and in the process A') and you're set! Now given b, just compute $b' = Rb$ and you can read the solutions to $A'x = b'$ straight off the matrix as above.

How can you compute R? Concretely, you can do this by looking at the augmented matrix $[A|I]$ – then Gauß-Jordan yields $[A'|R]$: each row operation you do on the A part gets recorded on the I part.

3 Special cases

3.1 Span

Given a collection of vectors $\langle v_1, \ldots, v_l \rangle$ where $v_i \in \mathbf{R}^n$, asking "is w in the span of $\langle v_i \rangle$, and if so, express it" is exactly the same as solving the equation $Ax = b$ where A is the column matrix of the v_i and $b = w$.

If we want to solve this for any w that's handed to us, just find A' and R (as above), and run the machine.

3.2 Invertible Matrices

A particularly interesting case is when the rref matrix A' is the identity matrix I_n (just 1s on the diagonal); note that this can only possible occur if A is square, i.e., a map $A\colon \mathbf{R}^n \to \mathbf{R}^n$ between spaces of the same dimension.

In that case there's no kernel (the kernel is $x_i = 0$ for all i), and the image is the whole space – the solution is exactly $x = b'$.

In that case $RA = I_n$, and $R = A^{-1}$.

For instance, for $A = \begin{pmatrix} 5 & 6 \\ 4 & 5 \end{pmatrix}$, we get $A^{-1} = \begin{pmatrix} 5 & -6 \\ -4 & 5 \end{pmatrix}$; in equations, $Ax = b$ is solved by applying A^{-1} to both sides and getting $A^{-1}Ax = A^{-1}b$ so $x = A^{-1}b$ (which we've also called b').