Solving linear systems of equations

Renzo: MATH 369

1 Step 1: translation.

We observe that finding the solutions of a linear systems of equations is equivalent to determining the inverse image of one vector via a linear function.

Suppose you have a linear system of m equations in n unknowns. For example:

$$\begin{cases} 3x + y - 2z &= 7 \\ 2x - y + 2z &= 3 \end{cases}$$

Look at what you see on the left hand side of the expression. That is the definition of a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$. Observe that the number of unknowns is the dimension of the input space, the number of equations is the dimension of the output space. In the above example, the linear function is:

$$\begin{array}{rccl} L: & \mathbb{R}^3 & \to & \mathbb{R}^2 \\ & (x, y, z) & \mapsto & (3x + y - 2z, 2x - y + 2z) \end{array}$$

Using the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^2$, we can represent L by the matrix:

$$M_L = \begin{bmatrix} 3 & 1 & -2 \\ 2 & -1 & 2 \end{bmatrix}$$

Finding the output for a generic input vector is obtained by matrix multiplication:

$$L(x, y, z) = \begin{bmatrix} 3 & 1 & -2 \\ 2 & -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Now notice that solving the linear system above corresponds to finding ALL values of x, y and z such that $L(x, y, z)$ is the vector $(7, 3)$. This is precisely the definition of $L^{-1}(7, 3)$.

In summary, we have the following dictionary:

Linear system of equations	Linear function $L : \mathbb{R}^n \to \mathbb{R}^m$
Number of equations	dimension output space
Number of variables	dimension input space
Solutions	inverse image of a vector.

2 Moment of reflection: what does this buy us?

We have spent a fair amount of time thinking about the geometry of linear functions; this should give us some insight as to what to expect when solving a linear system of equations. Here are some nice consequences of our translation. Suppose you are given a system of linear equations corresponding to finding $L^{-1}(w)$ for some vector w.

1. Solutions exist if and only if $w \in Range(L)$.
2. So, in particular, if L is onto, then solutions exist, whatever the value of w.
3. Solutions are a translate of $Ker(L)$, so in particular, they must depend on as many free parameters as the dimension of $Ker(L)$.
4. Solutions (when they exist) are unique if and only if L is $1:1$.
5. Solutions always exist and are unique when L is $1:1$ and onto, i.e. if L is an isomorphism.
6. The Rank-Nullity theorem gives us some a priori information on possibilities for the set of solutions.

For example, let us look at the system of linear equations in Section 1. Since we have a non-zero map from $\mathbb{R}^3 \to \mathbb{R}^2$, we know that it is not possible to have $Ker(L) = 0$. So there are two possibilities:

1. The dimension of $Range(L)$ is equal to one, and the dimension of $Ker(L)$ is equal to 2. In this case the function is not onto, so we have to worry whether the vector $(7, 3)$ belongs to the range or not.
2. The dimension of $Range(L)$ is equal to two, and the dimension of $Ker(L)$ is equal to 1. In this case the function is onto, so we know that our system would have a line worth of solutions.

How do we decide between the two cases? Well, we see for example that $L(e_1)$ and $L(e_2)$ are non-proportional, hence linearly independent. This tells us that the dimension of the $Range(L)$ is at least two. But then it must be 2. L is onto and we are in the second situation.

So without doing any computations, we know that the linear system from Section 1 must have a line worth of solutions!

3 That's fine and dandy, but how do we find the solutions?

Our goal now is to understand what the meaning of the Gaussian Elimination Algorithm in terms of our translation from Section 1. We start by making a couple wishes:

Wish 1: Since the solutions should correspond to values of $x, y, z, \ldots$ we would like to truly express the solutions by saying what $x, y, z, \ldots$ should be. But the values of $x, y, z\ldots.$ are the coordinates of vectors in $\mathbb{R}^n$ in the standard basis. Hence in our translation this means that we want to describe the subset $L^{-1}(w)$ in terms of the standard basis of $\mathbb{R}^n$. To be fully honest though, we will allow this small modification to the standard basis: it will be convenient to be able to reorder the vectors of the standard basis. This has the effect of correspondingly reordering the coordinates on the input space. (For example if I switch e_1 and e_3, z will be the first coordinate and x the third).

Wish 2: It would be nice if we could choose a new basis for the output space, that makes it so that it becomes immediate to read off the solutions.

Let us go back to our example from Section 1, and see how we can make both wishes come true. In our example, we have:

$$L(e_1) = (3, 2) \qquad L(e_2) = (1, -1) \qquad L(e_3) = (-2, 2)$$

So what happens if we take $L(e_1)$ and $L(e_2)$ to be a basis for $\mathbb{R}^2$? We note that

$$L(e_3) = 0L(e_1) - 2L(e_2)$$

and therefore the matrix for L (using the standard basis for $\mathbb{R}^3$, but our new basis in $\mathbb{R}^2$) is

$$M'_L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \end{bmatrix}$$

That is a pretty nice matrix... notice for example that figuring out the Kernel of L using this matrix is pretty simple:

$$Ker(L) = \left\{ \begin{array}{ccc} x & = & 0 \\ y - 2z & = & 0 \end{array} \right. = \{(0, 2z, z) | z \in \mathbb{R}\} = Span([0, 2, 1]).$$

Now, if we want to solve our linear system, we must find the coordinates of w in the new basis. Observe that

$$\begin{bmatrix} 7 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

This tells us that the coordinates of w in the new base are 2 and -1. So we have the weird equation:

$$\begin{bmatrix} 7 \\ 3 \end{bmatrix}_{st} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}_{new}$$

The superscripts there are written to remember that the two column vectors describe the same vector as a linear combination of two different bases.

So now we are looking to solve the linear system:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

which is equivalent to:

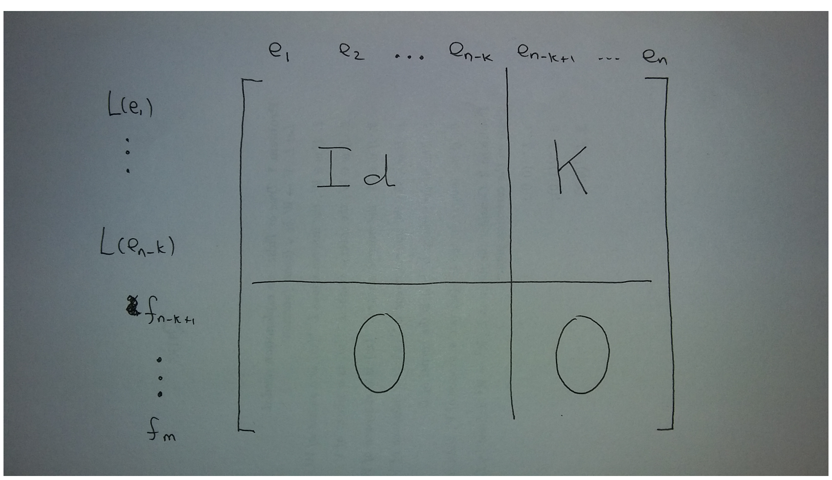
$$\begin{cases} x & = & 2 \\ y - 2z & = & 1 \end{cases}$$

And now it is immediate to see that $x = 2, y = 1, z = 0$ is a solution of this system.

But we know that $L^{-1}(w)$ is a translate of the Kernel of L by a particular solution, therefore:

$$L^{-1}(w) = \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 2z \\ z \end{bmatrix} | z \in \mathbb{R} \right\}$$

OK, let us now generalize what we just learned: suppose we have a linear function $L : \mathbb{R}^n \to \mathbb{R}^m$. Suppose that the dimension of Ker(L) is k, and the dimension of the *Range*(L) is $n - k$. Up to reordering, we can assume that the first $n - k$ vectors of $\mathbb{R}^n$ map to a basis for *Range*(L). Then we choose $L(e_1), \ldots, L(e_{n-k})$ to be the first $n - k$ vectors of a basis for $\mathbb{R}^m$, and we complete such list arbitrarily to a basis B' for $\mathbb{R}^m$. With this choice of basis, the matrix representing L looks like this:



Spelled out in words, we have:

1. The North-West block is a $(n-k) \times (n-k)$ identity matrix.
2. All rows past the $(n-k)$-th are all 0's.
3. The North-East block we have no control over.

We now observe that when the linear function has this form, these very nice things happen.

1. It is very easy to describe the Kernel of L:

   The $n-k+1$-th to n-th coordinates of a vector can be chosen freely, the first k coordinates are determined by the submatrix K.

2. It is easy to decide if a vector w belongs to the Range of L:

   A vector w belongs to Range(L) if and only if its coordinates starting from the $n-k+1$-th are all zero.

3. If $w \in Range(L)$, it is easy to find a particular solution to the linear system of equations $L(\vec{x}) = w$.

   Set the first $n-k$ coordinates equal to the first $n-k$ coordinates of w, and all the coordinates after equal 0.

4 OK, fine, but how do we do this in practice?

Indeed, even after we have decided what we want our basis for W to be, implementing the change of basis might be tricky. So, we study the following elementary changes of basis. Assume W has a basis $B = \{f_1, \ldots, f_m\}$. This is very important, so make sure you understand it well! We are now going to do some simple modification on the basis of W and describe how the coordinates of a vector w change!

scaling a vector Suppose we now replace the vector f_1 with the vector λf_1, for some $\lambda \neq 0$. So $B' = \{\lambda f_1, \ldots, f_m\}$. Then the coordinates of a vector w in the two bases are as follows:

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_n \end{bmatrix}_B = \begin{bmatrix} \frac{w_1}{\lambda} \\ w_2 \\ \ldots \\ w_n \end{bmatrix}_{B'}$$

swapping two vectors Suppose we now swap the vectors f_1 and f_2. So $B' = \{f_2, f_1, f_3 \ldots, f_m\}$. Then the coordinates of a vector w in the two bases are as follows:

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_n \end{bmatrix}_B = \begin{bmatrix} w_2 \\ w_1 \\ \ldots \\ w_n \end{bmatrix}_{B'}$$

subtracting one vector from another Suppose we replace f_2 with $f_2 - f_1$. So $B' = \{f_1, f_2 - f_1, f_3 \ldots, f_m\}$. Then the coordinates of a vector w in the two bases are as follows:

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_n \end{bmatrix}_B = \begin{bmatrix} w_1 + w_2 \\ w_2 \\ \cdots \\ w_n \end{bmatrix}_{B'}$$

Now suppose you have a linear function $L : V \to W$, that in some bases $\{e_1, \ldots, e_n\}$ for V and $\{f_1, \ldots, f_m\}$ for W is represented by the matrix M. Then, remembering that the columns of M are the coordinates of the vectors $L(e_1) \ldots L(e_n)$, in the basis $\{f_1, \ldots, f_m\}$, we have that if we change the basis of W in one of the three above ways, the matrix of L changes accordingly:

1. If the first basis vector f_1 is replaced by λf_1, then the entire first row or M is scaled by $\frac{1}{\lambda}$.
2. If the two basis vectors f_i and f_j are swapped, so are the i-th and j-th rows of M.
3. If the second basis vector is replaced by $f_2 - f_1$, then the first row of M is replaced by the sum of the first two rows.

But hey! These are precisely ROW operations on the matrix representing the linear function L. Note, that if you want to solve a linear system $L(x) = w$, then you want to make sure that if you change basis for the vector space W, you also change the coordinates of the output vector w. That is why you should perform row operations not ONLY on the matrix M but also on the column vector w. Doing these operations simultaneously are the same things as doing row operation on the augmented matrix $[M|w]$.

5 Gaussian Elimination, with no secrets!

Now we understand what happens when we do Gaussian eliminiation. By performing row operations on the augmented matrix of a linear system of equations, we change (in several small steps) basis for W until we obtain a basis as the one described in Section 3. At that point one can read off the solutions of the system as described in the end of Section 3.