211 LECTURE 14

Solution sets to systems of linear equations

Today we shall consider the problem of finding the solutions to a system of linear equations. This contains two subproblems. One is the existence problem: When can we be sure to find a solution? Second is the uniqueness problem: How many solutions do we get?

1. Some low dimensional cases–including geometry

The important thing here is to realize our set of solutions to a system of linear equations as the intersection of a bunch of things we can understand more easily–namely "hyperplanes" in the appropriate coordinate space. Here, a "hyperplane" is a linear subspace of one dimension less that the ambient space. We illustrate this principle with examples.

1.1. **Two unknowns.** We consider the systems which involve the equations $7x+2y = 5$, $x - y = 1$ and $2x + y = 0$. The idea is to use the geometry of lines in the plane to see the solution set. Also discuss the homogeneous versions.

1.2. **Three unknowns.** Here we discuss systems involving the equations $x + y = 0$, $y + z = 0$, $x + y + z = 1$ and $2x - y = 0$. Again, we use the geometry of planes in three-space to see the solution set. Also discuss the homogeneous versions.

2. The general picture

Abstracting these situations in fact works out pretty well. A system of m equations in n unknowns can be interpreted as a matrix system where the matrix is $m \times n$ and the vectors are column vectors of size n and m, respectively. The geometry is that we are intersecting m "hyperplanes" in the n-dimensional coordinate space in which the vectors are supposed to live.

2.1. **Our basic scheme.** The basic principle is that a system of m equations in n unknowns (i.e. one for which the associated matrix is of size $m \times n$) will fall into the following scheme.

- $m < n$**:** There are solutions. In fact, we should have a translate an $(n - m)$-dimensional subspace which consists of solutions.
- $m = n$**:** We expect exactly one solution.
- $m > n$**:** We expect no solutions.

Of course, stupid things can always happen. One can design systems which break the "rules" above in all sorts of ways.

3. A problem, and a hint at how to proceed

A big problem is that our system can contain lots of redundancies. This keeps us from using our "basic principle" effectively. By way of example, consider the following systems.

$$\begin{aligned} 2x - 3y &= 1 \\ -4x + 6y &= -2 \end{aligned}$$

This system is redundant because the second equation is just -2 times the first equation. It really should be just one equation.

$$\begin{aligned} 3x + 2y - z &= 4 \\ 3x - 2y \quad &= 1 \\ 4y - z &= 3 \end{aligned}$$

This system also has redundancy, but it is not as obvious. The third equation is unnecessary because it is the difference of the first two.

We need to find a method of solving the equations which helps us sort out these types of redundancies. If we can do this, it will hopefully show that our intuition in the scheme above holds true as well as it can. It turns out that this is the case! Our first hint is in a method we already know from high school.

3.1. **Gaussian Elimination in two unknowns.**

We already have a good method of solving systems in two unknowns. It is called *elimination.* The idea is to take linear combinations of the equations involved to simplify their form. For two equations a simple example is

$$\begin{aligned} 2x - 3y &= 1 \\ x + y &= 2 \end{aligned}$$

First, let's switch the order of these equations. (This seems silly, but the matrix analog is something we will do from time to time.)

$$\begin{aligned} x + y &= 2 \\ 2x - 3y &= 1 \end{aligned}$$

Now add (-2) times equation one to equation two to obtain the following equivalent system.

$$\begin{aligned} x \ +y &= 2 \\ -5y &= -3 \end{aligned}$$

Then we multiply equation two by $(-1/5)$ to get the system below.

$$\begin{aligned} x+y &= 2 \\ y &= 3/5 \end{aligned}$$

Finally, we add (-1) times equation two to equation one to see the solution.

$$\begin{aligned} x \quad &= 7/5 \\ y &= 3/5 \end{aligned}$$

It is not difficult to check that this works.

This seems very simple (it is), but if you are careful you already see exactly how to proceed in the general case. We will discuss the general case more precisely after break.

3.2. **the matrix formulation.** Note that in our solution, we didn't really need to know the names of the variables involved to do the process. The different variables just helped keep everything lined up in columns properly. Well, we can also use the matrix formulation of things to this!

Our problem is given in matrix form by the following equation.

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

What we do is form the *augmented matrix* $[A, \mathbf{b}]$, in our case given by

$$[A, \mathbf{b}] = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

And now perform operations as above *on the rows of the augmented matrix*! That is, we should use the operations of switching rows, multiplying a row by a number and adding a multiple of one row to another to try and make the left hand portion of the augmented matrix (The A part) look like the identity matrix $\mathbf{I}$. If we can do this, then the vector on the right side (the $\mathbf{b}$ part) will be our solution.

This is what we shall need to do in general, but there is more to discuss.

- We need an efficient way to do the elimination process. In fact, *Gaussian* or *Gauss-Jordan* elimination is the effective way.

- The process doesn't always lead to the result above of $\begin{bmatrix}\mathbf{I} & \mathbf{d}\end{bmatrix}$. One can often get a whole row of zeros. Sometimes, you get a row which is all zeros in the A part, but nonzero in the **b** part. Each of these situations needs to be analyzed.
- Finally, it can be useful to know in advance how big the solution set is supposed to be. This leads us to our next topic.

3.3. **homogeneous and inhomogeneous equations.** A system of linear equations is called *homogeneous* when the constants on the right hand side are all zero. It is the same thing to say that in matrix form it looks like $A \cdot \mathbf{x} = \mathbf{0}$.

The important thing about a homogeneous system is that it *always* has at least one solution, $\mathbf{x} = 0$. This corresponds to the geometric fact that all of the "hyperplanes" pass through the origin.

A system of equations is called *inhomogeneous* if it is not homogeneous. This case is more difficult. Of course, given an inhomogeneous system $A \cdot \mathbf{x} = \mathbf{b}$, we can always form the associated homogeneous system $A \cdot \mathbf{x} = \mathbf{0}$.

From our examples above, it seems that the dimension of the space of solutions to the inhomogeneous equation is the same as the dimension of the space of solutions to the homogeneous equation. This can be formalized by using a variant of a result we had about first order linear differential equations.

Proposition Suppose that v is a solution to the system of linear equations given by $A \cdot \mathbf{x} = \mathbf{b}$, and that w is a solution to the associated homogeneous equation $A \cdot \mathbf{x} = \mathbf{0}$. Then $v + w$ is a solution to the general equation $A \cdot \mathbf{x} = \mathbf{b}$.

Conversely, if v and w are two solutions to the system $A \cdot \mathbf{x} = \mathbf{b}$, then their difference $v - w$ is a solution to the corresponding homogeneous system $A \cdot \mathbf{x} = \mathbf{0}$.

This proposition tells us how to find the dimension of the solution set. Since solutions always look like

$$\text{particular solution} + \text{solution to the homogeneous version}$$

We see that the solution set is generally a "translate" of the solution set to the homogeneous problem.

We shall reinterpret our matrix equations as statements about linear mappings from one coordinate space to another, and then this problem will become one of finding the *null space* of the map. Then we shall have to find an effective way of describing this set, and we shall need the concept of *vector subspace* and *basis*.

All of this and more after break.