MATH 115, SUMMER 2012
LECTURE 23

JAMES MCIVOR

1. Intro to Diophantine Equations - One Linear Equation

We spent a lot of time studying solutions to congruences.

- A **Diophantine Equation** is an equation, usually polynomial, with integer coefficients, for which we seek solutions in integers.

- We'll spend the next week or so studying these, with an aim to introduce some geometric ideas to their solutions. The interaction of number theory and geometry is the source of some very exciting ideas in modern math.

- we begin with a linear diophantine equation: what are the solutions of

$$ax + by = c,$$

where a, b, c are given and x and y are to be determined?

- let $g = (a, b)$. we saw a long time ago that there is a solution if and only if $g|c$.

- if there is one solution, there are infinitely many. Why?

- draw picture.

- Write $y = f(x) = -\frac{a}{b}x + \frac{c}{b}$

- for which integers δ is $f(x + \delta)$ also an integer? Smallest such δ is b/g.

- when x increases by b/g, y decreases by a/g,

- conclusion: given one solution (x_0, y_0), the other solutions are given by

$$(x_0 + kb/g, y_0 - ka/g),$$

where k ranges over all integers.

2. Many Linear Equations

Now we consider the case of a sytem of m linear equations with integer coefficients in n unknowns $x_1, \ldots, x_n$, such as

$$\begin{aligned} a_{1,1}x_1 + \ldots + a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + \ldots + a_{2,n}x_n &= b_2 \\ \vdots \qquad & \quad \vdots \\ a_{m,1}x_1 + \ldots + a_{m,n}x_n &= b_m \end{aligned}$$

Our method of solution is similar to that of a linear algebra class - we use rwo and column operations. But since we work only with integers, we have to restrict slightly the row operations that are allowed. Specifically, we can do the following:

(1) (R1) Add m times one row to another row (here m must be an integer!)
(2) (R2) Interchange two rows
(3) (R3) Multiply one row by negative one (not just any integer!)

Similarly, we have the following column operations:

1. (C1) Add m times one column to another column (m an integer!)
2. (C2) Interchange two columns
3. (C3) Multiply one column by negative one (not just any integer!)

These row and column operations are more limited, since our scalars must be integers. Nevertheless, we have the following powerful theorem:

Theorem 1 (Smith Normal Form)**.** *If A is any $m \times n$ integer matrix, then there is an invertible $m \times m$ matrix L and an invertible $n \times n$ matrix R such that*

$$D = LAR,$$

where D is an $m \times n$ matrix that is "diagonal" (although not necessarily square, meaning that $d_{i,j} = 0$ if $i \neq j$.

Proof. (Sketch) First we note that row and column operations can be expressed as multiplication by some invertible matrix. So it suffices to show that we can reduce A to this "diagonal" form by row and column operations.

- switching the columns around if necessary, we may assume the first column is nonzero.

- switching rows around if necessary, we can assume the first entry of this column is nonzero, and has smallest absolute value out of all the nonzero entries in that column.

- multiply the first row by -1 if necessary to make the 1,1 entry positive

- the division algorithm shows that by adding appropriate multiples of the first row to the other rows, we can make all the entries beneath the first one in this first column zero.

- now move on to the second column. Repeat the above procedure without touching the first row.

- keep on doing this until you have an upper triangular matrix.

- now do column operations in the same spirit to make it diagonal.

- note that unlike when working over a field ($\mathbb{R}$, for example), we may not be able to ensure that the diagonal entries are all 1.

□

In fact, with a little more care, we can arrange it so that the diagonal entries $d_1, d_2, \ldots$ of D satisfy the divisibility relations $d_1 | d_2$, $d_2 | d_3$, etc., and having done this, the matrix D is unique. But we won't need this uniqueness, so I omitted it from the statement of the theorem.

- Why do we care? This allows us to solve the system of linear equations above!

- Here's how: Write the system of linear integral equations in matrix form:

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

where $\mathbf{x}$ is the $n \times 1$ column vector of the variables $x_1, \ldots, x_n$, and $\mathbf{b}$ is the $m \times 1$ column vector of the numbers $b_1, \ldots, b_m$.

- Write the smith normal form (SNF) of A as $D = LAR$. Then set $\mathbf{y} = R^{-1}\mathbf{x}$ (possible since R is invertible), so $\mathbf{x} = R\mathbf{y}$, and set $\mathbf{c} = L\mathbf{b}$.

- Then the equation $A\mathbf{x} = \mathbf{b}$ is equivalent to $D\mathbf{y} = \mathbf{c}$.

- then we have the following useful result, which is immediate by the invertibility of L and R:

Theorem 2. *With notation as above, the equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution in integers if and only if the equation $D\boldsymbol{y} = \boldsymbol{c}$ has a solution in integers. Letting the diagonal entries of D be $d_1, \ldots, d_r$, and the entries of $\boldsymbol{c}$ be $c_1, \ldots, c_m$, this happens if and only if $d_i | c_i$ for $i = 1, \ldots, r$, and $c_i = 0$ for $i > r$.*

- There are two ways the system may fail to have a solution:

- first, it may be inconsistent, even over $\mathbb{R}$ - this happens when some $c_i \neq 0$ for $i > r$.

- second, when working over the integers we have an additional problem, of divisibility. For instance, the first row of our "diagonal" system could be $3y_1 = 2$, which is OK over $\mathbb{R}$, but has no solution over $\mathbb{Z}$.

- note also that if there is a solution, then two things can happen:

1) $r = n$, and the solution is unique, or

2) $r < n$, and then there are infinitely many integer solutions, coming from the free variables y_i for $i > r$.

3. How to Actually Compute the SNF

I find the book's treatment of this procedure a little confusing: here's what I recommend you do:

(1) Write down $I_m A I_n$, where I_m is the $m \times m$ identity matrix
(2) Perform your row and column operations on A
(3) Every time you do a row operation, do it also to the left matrix I_m
(4) Every time you do a column operation, do it also to the right identity matrix I_n
(5) Eventually the middle matrix will be your diagonal matrix D
(6) The matrix on the left is L, the matrix on the right is R
(7) The solution depends on D and $\mathbf{c}$, so compute $\mathbf{c} = L\mathbf{b}$
(8) Now check whether D and $\mathbf{c}$ satisfy the conditions in the theorem.
(9) If no, you're done, if yes, get your solution $\mathbf{y}$, and then compute $\mathbf{x} = R\mathbf{y}$

4. An Example

Find all integer solutions to the system of equations

$$\begin{aligned} 2x_1 + x_2 - 3x_3 - x_4 &= 10 \\ x_1 - x_2 - 3x_3 + x_4 &= 2 \\ 4x_1 - 4x_2 + 16x_4 &= 20 \end{aligned}$$

Solution

We set

$$A = \begin{pmatrix} 2 & 1 & -3 & -1 \\ 1 & -1 & -3 & 1 \\ 4 & -4 & 0 & 16 \end{pmatrix}$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

$$\mathbf{b} = \begin{pmatrix} 10 \\ 2 \\ 20 \end{pmatrix}$$

Do the prescribed operations (computations omitted) - you get

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 12 & 0 \end{pmatrix}$$

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

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

so

$$\mathbf{c} = L\mathbf{b} = \begin{pmatrix} 2 \\ 6 \\ 12 \end{pmatrix}$$

Since $1|2$, $3|6$, and $12|12$, we have a solution. It's not unique, since here $r = 3 < 4 = n$. The solution for $\mathbf{y}$ is

$$\mathbf{y} = \begin{pmatrix} 2 \\ 2 \\ 1 \\ k \end{pmatrix},$$

where $y_4 = k$ can be any integer, which gives

$$\mathbf{x} = R\mathbf{y} = \begin{pmatrix} 6 - 24 \\ 1 + 2k \\ 1 - k \\ k \end{pmatrix}$$