5. And over rings?

For a general ring, not much can be said in general about using Gaussian elimination to solve systems of linear equations. That we can't divide by some non-zero elements generally ensures that the method doesn't work. However, the integers $\mathbb{Z}$ have some special properties which allow something similar to work.

For example, even though we can't divide by 3, we can still solve $3x = 6$: $3x = 6$ implies $3x - 6 = 3(x - 2) = 0$ and thus either $3 = 0$ or $x - 2 = 0$. Since $3 \neq 0$ it must be the case that $x = 2$. Indeed, if two integers r, s have product $rs = 0$ then one or other of r and s must be zero.

Definition 15. *A ring R is an* integral domain *if the product of two non-zero elements is non-zero. More formally, if $r, s \neq 0$ in R then $rs \neq 0$.*

Examples:

(1) The integers $\mathbb{Z}$ is an integral domain. This lets us solve equations such as . We can't divide by 3 inside $\mathbb{Z}$, but
(2) $\mathbb{Z}/6\mathbb{Z}$ is not an integral domain since $2 \cdot 3 = 0$.

We also have a notion of divisor:

Definition 16. *Let $a, r \in R$ when R is an integral domain. Then r divides a, denoted by $r|a$, if there is an element $s \in R$ such that $a = rs$.*

(Question 5.1) Show that any unit in $R^\times$ divides any element in R

(Question 5.2) Show that $a|a'$ and $a'|a$ if and only if $a = ua'$ for some unit $u \in R^\times$

(Question 5.3) What integers divide 7 in $\mathbb{Z}$? Be careful about negative numbers!

However, the properties of the integers that really helps us is

If $a \neq 0$ and $a|b$ then $|a| \leq |b|$
If $a, b \in \mathbb{Z}$ and $b \neq 0$ then there exists unique integers q, r such that $a = qb + r$ with $0 \leq r < |b|$.

For example, suppose we had the 1×2 matrix of integes $[12 \ -54]$. Then we can apply (column) operations using property 2) to reduce the overall size of the entries. Since $-54 = -5 \cdot 12 + 6$ if we subtract add 5 times column I to column II we get $[12\ 6]$. We can then subtract 2 times column II from column I to get $[0\ 6]$ and then switch to get $[6\ 0]$. Notice that 6 is the greatest common divisor of -54 and 12

Given a row (or column) of integers $\mathbf{r} = [j_1\ j_2\ \cdots\ j_k]$ with at least one $j_i \neq 0$, let $c(\mathbf{r}) = \sum |j_i| \geq 0$. If two entries j_l and j_m are non-zero, then we can repeat the preceding reasoning to find a column operation which replaces $\mathbf{r}$ with $\mathbf{r}'$ so that $0 \leq c(\mathbf{r}') < c(\mathbf{r})$. Repeating this constantly reduces the value of $c(\mathbf{r})$, but this is bounded below by $0 \in \mathbb{Z}$. Thus, we must eventually be unable to continue: we can arrange to get a row $[g\ 0\ \cdots\ 0]$ with at most one non-zero entry, which is positive and occurs in the first position. This non-zero entry g must divide all the original j_i:

(Question 5.4) Suppose we have a sequence of numbers $[s_1\ s_2\ \cdots\ s_k]$ and $d|s_i$ for all i. Show that if we replace our sequence with one obtained by either

(1) adding a multiple of the l^{th} entry to the l^{th} entry ($l \neq m$), or

(2) switch two entries, or
(3) multiply all the entries by ±1, a unit in $\mathbb{Z}$

we obtain a new sequence all of whose entries are still divisible by d.

Since all the entries of $[g\ 0\ \cdots\ 0]$ are divisible by g, and we can obtain $[j_1\ j_2\ \cdots\ j_k]$ by reversing the column operations) we must have $g|j_i$. In fact, this g is the greatest common divisor of the j_i: for if $d|j_i$ for all i then, by the preceding question $d|g$ as well.

We now prove two results about matrices of integers

Proposition 17. *A matrix* $\mathbf{M} \in M_{n\times n}(\mathbb{Z})$ *is invertible if and only if it is the product of elementary matrices.*

We will prove this by showing that there are elementary matrices $\mathbf{E}_1, \ldots, \mathbf{E}_k$ such that

$$\mathbf{E}_1\mathbf{E}_2\cdots\mathbf{E}_k\mathbf{M} = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix whose diagonal entries are all equal to one of the units ±1. $\mathbf{D}$ is then invertible. So if such elementary matrices exist, $\mathbf{M}$ will be invertible.

We prove the other direction of the proposition by induction on n. First, if $n = 1$, then $\mathbf{M} = [a]$. If $\mathbf{M}$ is invertible, there exists a $b \in \mathbb{Z}$ such that $[a][b] = [ab] = [1]$. Thus, $ab = 1$ and a is a unit. Consequently, $\mathbf{M}$ is a diagonal matrix with units on the diagonal, as required. Now assume the result is true for all invertible matrices in $M_{n\times n}(\mathbb{Z})$. We will show that is it also true for all matrices in $M_{(n+1)\times(n+1)}(\mathbb{Z})$. Start with the first column, as above, there are row operations which will convert our matrix into the form $\left[\begin{array}{c|c} u & * \\ \hline 0 & \mathbf{M}' \end{array}\right]$

(Question 5.5) By explicitly writing done $\mathbf{M}\mathbf{M}^{-1} = \mathbf{I}$, show that if

$$\mathbf{M} = \left[\begin{array}{c|c} u & * \\ \hline 0 & \mathbf{A} \end{array}\right] \qquad \text{then} \qquad \mathbf{M}^{-1} = \left[\begin{array}{c|c} v & * \\ \hline 0 & \mathbf{B} \end{array}\right]$$

with $uv = 1$ and $\mathbf{AB} = \mathbf{I}$. That is, u is a unit in R and $\mathbf{A}$ is invertible in $M_{n\times n}(R)$.

(Question 5.6) Use the induction assumption on $\mathbf{A}$ to replace the lower right matrix with a diagonal matrix whose diagonal entries are units. Explain how knowing that these are units allows us to find additional row operations that eliminate the $*$'s, so that we end with a diagonal matrix whose diagonal entries are all units.

(Question 5.7) Find the inverse of a the following matrix in $GL_3(\mathbb{Z})$. Do not ever divide!

$$\begin{bmatrix} 5 & 1 & -1 \\ -2 & -3 & -2 \\ 2 & 2 & 1 \end{bmatrix}$$

While there is no analog of RREF for integer matrices, there is a similar form we can hope to achieve, called *Smith Normal Form*:

Theorem 18. *Let $M \in M_{n\times m}(\mathbb{Z})$. Then there exist matrices $\mathbf{P} \in GL_n(\mathbb{Z})$ and $\mathbf{Q} \in GL_m(\mathbb{Z})$ such that*

$$\mathbf{PMQ}^{-1} = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & \end{bmatrix}$$

whose only non-zero entries occur on the main diagonal, and $d_1|d_2|d_3|\cdots|d_k$.

Notes:

(1) It is possible for $d_i = 0$ provided $d_{i+1} = d_{i+2} = \cdots = d_k = 0$ as well.
(2) This form is necessarily unique: to make it such we would need to multiply the entries by units to make them positive.
(3) The "diagonal" here occurs in a potentially non-square matrix and thus can terminate on one of the sides.

We start as above by using row operations to obtain a matrix $\left[\begin{array}{c|c} g_1 & * \\ \hline \mathbf{0} & \mathbf{M}_1 \end{array}\right]$ where e need *not* be a unit. Since e is not a unit, we should not hope to arrange for units on the diagonal. However, what can proceed by doing column operations, at the expense of multiplying on the right by an invertible matrix such as $\mathbf{Q}^{-1}$.

That is, suppose we have used row operations to obtain a matrix, as above, where $m_{i1} = 0$ for $i > 1$ and let $g_1 = m_{11}$. The top row then looks like $[g_1\ m_{12}\ \cdots\ m_{1m}]$. Now we perform column operations on the matrix until we arrive at a top row like $[g_2\ 0\ \cdots\ 0]$. By the arguments above g_2 is the greatest common divisor of the elements in the original row, so $g_2|g_1$. Thus, $|g_2| \le |g_1|$. At this point we have matrix of the form $\left[\begin{array}{c|c} g_2 & \mathbf{0} \\ \hline * & \mathbf{M}_2 \end{array}\right]$ where, due to the column operations, the first column need not consist only of 0's (for instance we may have to switch columns to get g_2 in the top left.

However, we can perform row operations on the first column again to get g_3 in the top left, with $g_3|g_2$, and the other entries in the first column all 0. This can mess up the first row, so we repeat with the first row. Alternating between row and column operations, using the first column and row to guide us, gives $g_1, g_2, g_3, \ldots$ with $g_{i+1}|g_i$ and $|g_i| \ge |g_{i+1}|$. Indeed, if one of the entries in the first row (or column) is not divisible by g_i then we can ensure that $|g_{i+1}| < |g_i|$. Thus, we will get a decreasing sequence of natural numbers $|g_i|$, so this sequence must become constant for some $i = n$. This can only happen if we all the other entries in the first row and column are divisible by g_n, and we can then use row and column operations to make them all 0 by adding multiples of the first row and column. Thus, we have a matrix of the form

$$\begin{bmatrix} g_n & 0 & 0 & \cdots & 0 \\ 0 & * & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & * & * & \cdots & * \end{bmatrix}$$

Now we can apply induction to the lower $(n-1)\times(m-1)$ matrix of $*$'s to get a matrix of the form

$$\mathbf{PMQ}^{-1} = \begin{bmatrix} d_1' & 0 & 0 & \cdots & 0 \\ 0 & d_2' & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & \end{bmatrix}$$

This has the right diagonal form, but we don't yet know that $d'_1|d'_2|\cdots|d'_k$ as required. To do that requires an additional observation: since we can turn $[a\ b]$ into $[g\ 0]$ by column operations, we must have that $g = ax + by$ for some elements $x, y \in \mathbb{Z}$. If we let $a = g\alpha$ and $b = g\beta$ for some $\alpha, \beta \in R$, then

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} a & 0 \\ ax & b \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} a & 0 \\ ax+by & b \end{bmatrix} = \begin{bmatrix} a & 0 \\ g & b \end{bmatrix}$$
$$\xrightarrow{(3)} \begin{bmatrix} 0 & -\alpha b \\ g & b \end{bmatrix} \xrightarrow{(4)} \begin{bmatrix} 0 & -\alpha b \\ g & 0 \end{bmatrix} \xrightarrow{(5)} \begin{bmatrix} g & 0 \\ 0 & -\alpha b \end{bmatrix}$$

Where we have performed the following row/column operations at each step:

(1) $RII + xRI \to RII$
(2) $CI + yCII \to CI$
(3) $RI - \alpha RII \to RI$
(4) $CII - \beta CI \to CII$
(5) $RI \leftrightarrow RII$

Note that we can also reverse the sequence above. For example, if a and b are relatively prime, then we can find row/column operations implementing

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \leftrightarrow \begin{bmatrix} 1 & 0 \\ 0 & ab \end{bmatrix}$$

Since $g|(-\alpha b)$, the last matrix has the required division property along the diagonal. By choosing different pairs of the d'_is along the diagonal, we can obtain the Smith Normal Form. We illustrate how to choose pairs with the following exercise:

(Question 5.8) Explain why

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$$

is not in Smith Normal Form in $M_{4\times 4}(\mathbb{Z})$ and then find the Smith Normal Form.

(Question 5.9) Show that computing the Smith normal form gives

$$\begin{bmatrix} 8 & 4 & 8 \\ 4 & 8 & 4 \end{bmatrix} \qquad \longrightarrow \qquad \begin{bmatrix} 4 & 0 & 0 \\ 0 & 12 & 0 \end{bmatrix}$$

Then use the Smith Normal form calculation to find all solutions to the system

$$8x + 4y + 8z = a$$
$$4x + 8y + 4z = b$$

when $a = -4, b = 4$ and when $a = 3, b = 12$. (*Hint:* Be careful about what happens when you do column operations: use $\big(\mathbf{PMQ}^{-1}\big)\big(\mathbf{Qx}\big) = \mathbf{Pb}$ and find $\mathbf{Q}$ by keeping track of the elementary matrices used for the column operations).

If you look closely at the arguments above, we only use a few properties of the integers. We can try to find other rings with the same properties so that the results above apply immediately to them as well. So,

Definition 19. *A* Euclidean domain *is an integral domain R which admits a function $\phi\colon R \longrightarrow \mathbb{Z}$ such that*

1. *$\phi(0) = 0$ and $\phi(r) > 0$ when $r \neq 0$*
2. *If a divides b and $b \neq 0$ the $\phi(a) \leq \phi(b)$*
3. *For any $a, b \in R$ with $b \neq 0$, then there exist integers q, r such that $a = qb + r$ with $0 \leq \phi(r) < \phi(b)$.*

The integers $\mathbb{Z}$ are Euclidean with $\phi(n) = |n|$ by the quotient-remainder theorem. The name comes from the that of the process for reducing $[ab] \longrightarrow [g0]$, which is a version of what's better known as the Euclidean Algorithm.

By following the arguments above, and replacing $|n|$ with $\phi(r)$ we obtain proofs of

Proposition 20. *If R is a Euclidean domain, then a matrix $\mathbf{M} \in M_{n\times n}(R)$ is invertible if and only if it is the product of elementary matrices.*

Theorem 21. *Let R be a Euclidean domain, and let $M \in M_{n\times m}(R)$. Then there exist matrices $\mathbf{P} \in GL_n(R)$ and $\mathbf{Q} \in GL_m(R)$ such that*

$$\mathbf{PMQ}^{-1} = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & \end{bmatrix}$$

whose only non-zero entries occur on the diagonal, and $d_1|d_2|d_3|\cdots|d_k$.

Incidentally, this implies that we can find a "greatest" common divisor of any two elements $a, b \in R$: simply form $[a\ b]$ and use row operations to reduce it to $[g\ 0]$. The resulting g is a common divisor: $g|a$ and $g|b$ and is "greatest" in the following sense: if $h|a$ and $h|b$ then $h|g$. However, we can multiply g by a unit to get another element with the same properties. In $\mathbb{Z}$ we could make the greatest common divisor unique by requiring it to be positive, but there may be no such normalization in other Euclidean rings.

However, Euclidean rings are rare. Smith Normal Forms exist for matrices over a more general type of ring: a principal ideal domain, but these are also rare. There are many types of rings for which the methods here just do not work. Nevertheless, a ring we will care about *is* Euclidean:

Theorem 22. *Let $\mathbb{F}$ be a field, and suppose $a(x), b(x) \in \mathbb{F}[x]$ and $b(x) \neq 0$. Then there exist unique polynomials $q(x)$ and $r(x)$ with $0 \leq \deg\big(r(x)\big) < \deg\big(b(x)\big)$ such that*

$$a(x) = q(x)b(x) + r(x)$$

This follows from the algorithms for dividing polynomials:

Example: Suppose $b(x) = 3x^2 + 2x$ and $a(x) = 2x^4 - 3x^2 + x + 1$, then following the recipe above we see that $a(x) - \frac{2}{3}x^2\, b(x) = -4/3\, x^3 - 3x^2 + x + 1$. The latter polynomial minus $-4/9\, x\, b(x)$ is $-19/9\, x^2 + x + 1$, which in turn is $-19/27\, b(x) + 65/27\, x + 1$. Putting these together gives us

$$\begin{aligned} a(x) &= \frac{2}{3}x^2 b(x) - \frac{4}{9}\, x\, b(x) - \frac{19}{27}\, b(x) + \frac{65}{27}x + 1 \\ &= \big(\frac{2}{3}x^2 - \frac{4}{9}\, x - \frac{19}{27}\big)\, b(x) + \frac{65}{27}x + 1 \end{aligned}$$

(Question 5.10) Show that for polynomials in $\mathbb{F}[x]$ the values for $q(x)$ and $r(x)$ are unique: that is, if $q(x)b(x) + r(x) = q'(x)b(x) + r'(x)$ with the degrees of $r(x)$ and $r'(x)$ being smaller than that for b then $q = q'$ and $r = r'$.

(Question 5.11) We prove the theorem by induction on the degree of $a(x)$:

(Base Case) Suppose $b(x) \in \mathbb{F}[x]\backslash\{0\}$ and $\deg\big(a(x)\big) < \deg\big(b(x)\big)$. Then we can take $q(x) = 0$ and $r(x) = a(x)$.

(1) Suppose $\deg\big(b(x)\big) = m$, then $b = b_m x^m +$ (lower order terms) for $b_m \neq 0$. Let $n \in \mathbb{N}$ with $n \geq m$. Assume the result is true for all polynomials $a'(x) \in \mathbb{F}[x]$ with $\deg\big(a'(x)\big) < n$. Let $a(x) \in \mathbb{F}[x]$ be some polynomial of degree n. Then $a = a_n x^n +$ (lower order terms) with $a_n \neq 0$. Use induction on $a(x) - \dfrac{a_n}{b_m} \cdot x^{n-m}\, b(x)$ to prove the result.

The degree of the polynomial plays the role that $|n|$ did before. However, to get the definition of ϕ exactly the same we use $\phi\big(p(x)\big) = 2^{\deg(p)}$ and $\phi(0) = 0$.

Moreover, these polynomial rings are integral:

(Question 5.12) Prove that $p(x), q(x) \neq 0$ implies $p(x)q(x) \neq 0$ in $R[x]$ when R is an integral domain.

(Question 5.13) Show that $R = \mathbb{Z}/6\mathbb{Z}[x]$ is not an integral domain. In fact, when R is not an integral domain then $R[x]$ is not either.

and we can normalize the greatest common divisors. In $\mathbb{F}[x]$ a polynomial $a_k x^k +$(lower order terms) is called *monic* if $a_k = 1$. When $a_k \neq 0$ in $\mathbb{F}$ we can divide the polynomial by a_k, a unit, to get a *monic* polynomial. So we will require our greatest common divisors to be monic polynomials. Notice that the preceding arguments give us an algorithm for finding this, similar to the Euclidean algorithm.

Thus, since $\mathbb{F}[x]$ is Euclidean, any matrix with entries in $\mathbb{F}[x]$ has a Smith Normal Form. This fact will prove important to us later.

(Question 5.14) Find the Smith Normal forms for

$$\begin{bmatrix} x-1 & 0 & 0 \\ 0 & x-1 & 0 \\ 0 & 0 & x-2 \end{bmatrix} \qquad \begin{bmatrix} x-1 & 3 & 0 \\ 0 & x-1 & 0 \\ 0 & 0 & x-2 \end{bmatrix}$$

in $M_{3\times 3}(\mathbb{R}[x])$

Note: Euclidean rings have another nice property: any element is the product, up to multiplication by units, of a unique collection of prime elements, where an element $p \in R$ is prime if $p = rs$ implies that either r or s is a unit[7]. That is, the fundamental theorem of arithmetic holds in these rings (which, as it is proved using the greatest common divisor methods above, shouldn't be too surprising.)

[7] This is true in Euclidean rings. In more general rings we need to distinguish being prime and being irreducible