2. Echelon form

It is time to make some more definitions and actually prove that Gaussian elimination works:

Definition 2.1. *Let m and $n \in \mathbb{Z}$ be two positive integers. Let*

$$I = \{\, i \in \mathbb{Z} \,|\, 0 < i \leq m \,\} \qquad \text{and} \qquad J = \{\, j \in \mathbb{Z} \,|\, 0 < i \leq n \,\}.$$

*A **real** $m \times n$ **matrix** is a function $A \colon I \times J \longrightarrow \mathbb{R}$.*

Let A be an $m \times n$ matrix. Let $i \in I$ and $j \in J$ be two indices, so that $1 \leq i \leq m$ and $1 \leq j \leq n$. The (i, j) entry of the matrix A is the value of the function A at (i, j), that is $A(i, j)$. It is customary to write $A = (a_{ij})$ which means that $A(i, j) = a_{ij}$ and we refer to this as the entry in the ith row and j column (in the usual fashion).

Note however that the definition given in (2.1) is convenient for two reasons. First it is immediate that two matrices are equal if and only if they have the same entries (since this is what it means for two functions to be the same). It is also straightforward to generalise this definition to complex matrices, integer matrices and so on (the possibilities are endless); just change the range of the function to the appropriate set.

Now let us define the end product of Gaussian elimination. This is a little harder to define than one might first imagine:

Definition 2.2. *Let A be a matrix. We say that A is in **echelon form** if A satisfies the following properties:*

- *The first (reading left to right) non-zero entry in every row is a one. Such entries are called **pivots**.*
- *Every row which contains a pivot occurs before every row which does not contain a pivot.*
- *Pivots in earlier rows (again, reading left to right) come in earlier columns (reading top to bottom).*

One can express the first and third rules using indices. The first rule says that if we fix i, and j has the property that $a_{ij} \neq 0$ whilst $a_{ij'} = 0$ for all $j' < j$ then $a_{ij} = 1$. The third rule says that if there is a pivot at position (i, j) and there is a pivot at position (i', j'), where $i < i'$ then $j < j'$.

Okay, so what are the basic steps of Gaussian elimination?

Definition 2.3. *Let A be a matrix. An **elementary row operation** is one of three operations:*

1. *Multiply a row by a non-zero scalar.*
2. *Multiply a row by a non-zero scalar and add the result to another row.*

(3) *Swap two rows.*

We say that two matrices A_1 and A_2 are ***row equivalent*** *if we can get from A_1 to A_2 by a (finite) sequence of elementary row operations.*

So Gaussian elimination takes any matrix A, performs a sequence of elementary row operations until A is in echelon form. By definition this means that A is row equivalent to a matrix in echelon form. Let us first see that performing row operations does not change the set of solutions:

Theorem 2.4. *Let $C_1 = (A_1|b_1)$ and $C_2 = (A_2|b_2)$ be two row equivalent augmented matrices.*

Then the system of equations $A_1v = b_1$ and $A_2v = b_2$ have the same solution set.

Proof. By an obvious induction it suffices to prove this result if C_2 is obtained from C_1 by an elementary row operation. Note first that all of the elementary row operations are reversible. If we multiply the ith row of C_1 by λ to get C_2 then if we multiply the ith row of C_2 by $1/\lambda$ we get C_1. If we swap the ith row and jth row of C_1 to get C_2 then we simply swap the same rows to get from C_2 to C_1. Finally if we take the ith row of C_1, multiply by λ and add it to the jth row then, to get back, take the ith row of C_2, multiply by $-\lambda$ and add it to the jth row. In other words if there is an elementary row operation to get from C_1 to C_2 then there is an elementary row operation to get from C_2 to C_1 (of the same type even).

Let S_1 be the solution set of $A_1v = b_1$ and let S_2 be the solution set of $A_2v = b_2$. Then it suffices to show that $S_1 \subset S_2$. Indeed by symmetry, $S_2 \subset S_1$ and in this case $S_1 = S_2$.

Note that swapping rows has no effect on the solutions. Swapping rows of both C_1 and C_2, we may therefore assume that the elementary row operation only acts on the first two rows. Let F_i be the solution set of the equations corresponding to those rows of C_i and let L_i be the solution set of the remaining rows. Then the solutions sets are precisely $S_1 = F_1 \cap L_1$ and $S_2 = F_2 \cap L_2$. But $L_1 = L_2$ since they are the solutions sets corresponding to the same rows and so to the same equations. So we might as well assume that $m \leq 2$ (the two cases, $m = 1$ and $m = 2$ depend on which elementary row operation we use).

Consider multiplying the first row by a non-zero scalar λ. In this case $m = 1$ and the two equations read,

$$a_1x_1+a_2x_2+\cdots+a_nx_n = b \qquad (\lambda a_1)x_1+(\lambda a_2)x_2+\cdots+(\lambda a_n)x_n = \lambda b.$$

Suppose that $v = (r_1, r_2, \dots, r_n) \in S_1$. By definition this means

$$a_1 r_1 + a_2 r_2 + \cdots + a_n r_n = b.$$

Multiply both sides by λ to get

$$(\lambda a_1)\lambda r_1 + (\lambda a_2) r_2 + \cdots + (\lambda a_n) r_n = \lambda b.$$

But then v is a solution of the second equation, that is $v \in S_2$. Hence $S_1 \subset S_2$ and so by symmetry $S_1 = S_2$.

Now consider the only remaining case, multiplying the first row by a non-zero scalar λ and adding it to the second row. In this case $m = 2$. If the equations corresponding to C_1 are

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \end{aligned}$$

then the equations corresponding to C_2 are

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ (\lambda a_{11} + a_{21})x_1 + (\lambda a_{11} + a_{22})x_2 + \cdots + (\lambda a_{1n} + a_{2n})x_n &= (\lambda b_1 + b_2). \end{aligned}$$

Suppose that $v = (r_1, r_2, \dots, r_n) \in S_1$. By definition this means

$$\begin{aligned} a_{11}r_1 + a_{12}r_2 + \cdots + a_{1n}r_n &= b_1 \\ a_{21}r_1 + a_{22}r_2 + \cdots + a_{2n}r_n &= b_2. \end{aligned}$$

But then

$$\begin{gathered} (\lambda a_{11} + a_{21})r_1 + (\lambda a_{11} + a_{22})r_2 + \cdots + (\lambda a_{1n} + a_{2n})r_n = \\ \lambda a_{11}r_1 + a_{21}r_1 + \lambda a_{11}r_2 + a_{22}r_2 + \cdots + \lambda a_{1n}r_n + a_{2n}r_n = \\ \lambda(a_{11}r_1 + a_{11}r_2 + \cdots + a_{1n}r_n) + (a_{21}r_1 + a_{22}r_2 + \cdots + a_{2n}r_n) = \lambda b_1 + b_2, \end{gathered}$$

so that v satisfies the second equation of the second system. Since v clearly satisfies the first equation of the second system (which is nothing but the first equation of the first system), $v \in S_2$. Hence $S_1 \subset S_2$ and so $S_1 = S_2$ by symmetry. $\square$

Now we describe Gaussian elimination as an algorithm.

Algorithm 2.5 (Gaussian elimination)**.** *Let A be an $m \times n$ matrix.*

(1) *Let $l = 0$.*
(2) *Set $i = l+1$. If there is no row below the lth row which contains a non-zero entry then* ***STOP****.*
(3) *Otherwise pick the smallest j such that there is an index $i' > i$ such that $a_{i'j} \neq 0$.*
(4) *Swap rows i and i'.*
    (a) *Let $\lambda = 1/a_{ij}$ and multiply the ith row by $\lambda = 1/\mu$.*

(b) *If there are no non-zero elements below the ith row in the jth column then increase l by 1 and then* ***return*** *to (2).*
(c) *Pick the smallest j such that there is an index $i' > i$ with $\lambda = a_{i'j} \neq 0$. Multiply the ith row by $-\lambda$ and add it to the row i' and* ***return*** *to (b).*

The basic idea is that the index l measures how many rows are in echelon form. At the beginning we don't know any rows are in echelon form, so we put $l = 0$. If the only rows left are rows of zeroes then we are done. Otherwise we pick the first non-zero entry to the left below row l and we put this in the next row (row $l+1$). We make this number a pivot and eliminate all the entries in the same column below this entry. At this stage we know that the first $l+1$ rows are in echelon form and we repeat the whole process.

We need a little notation. Let $A = (a_{ij})$ a matrix. A submatrix B of A is any matrix which is obtained from A by deleting rows and columns.

Theorem 2.6. *Gaussian elimination always terminates with a matrix in echelon form.*

In particular every matrix is row equivalent to a matrix in row echelon form.

Proof. Let B_l be the $l \times n$ submatrix of A obtained by deleting the last $m = l$ rows of A. I claim that at every stage of the algorithm the matrix B_l is in echelon form.

This is certainly true at the beginning (since then $l = 0$ and this is a vacuous statement). On the other hand, the steps of the algorithm never alter the first l rows, so if don't change l we never lose the fact that B_l is in echelon form. Finally every time we increase l by one then it easy to see that the matrix B_{l+1} is in echelon form.

Suppose that Gaussian elimination stops. The only place we can stop is at step (2) in which case all the entries below the lth row are zeroes. But then the matrix A is in echelon form.

Therefore it suffices to prove that Gaussian elimination always terminates. Note that there are two places where the algorithm loops (and this is of course the only reason the algorithm could continue indefinitely).

Suppose that we are at step (c). Every time we go to step (b), we create one more zero in the jth column in a row below the ith row. Since there are at most $m - i$ rows below the ith row, this can only happen at most $m - i$ times.

Now suppose that we are at step (b) and we return to step (2). Since l increases by one every time and $l \leq m$ this can only happen at most m times. $\square$