

GAUSSIAN ELIMINATION AND LU DECOMPOSITION (SUPPLEMENT FOR MA511)

D. ARAPURA

Gaussian elimination is the go to method for all *basic* linear classes including this one. We go summarize the main ideas.

1. MATRIX MULTIPLICATION

The rule for multiplying matrices is, at first glance, a little complicated. If A is $m \times n$ and B is $n \times p$ then $C = AB$ is defined and of size $m \times p$ with entries

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

The main reason for defining it this way will be explained when we talk about linear transformations later on.

THEOREM 1.1.

- (1) *Matrix multiplication is associative, i.e. $A(BC) = (AB)C$. (Henceforth, we will drop parentheses.)*
- (2) *If A is $m \times n$ and $I_{m \times m}$ and $I_{n \times n}$ denote the identity matrices of the indicated sizes, $AI_{n \times n} = I_{m \times m}A = A$. (We usually just write I and the size is understood from context.)*

On the other hand, matrix multiplication is almost never commutative; in other words generally $AB \neq BA$ when both are defined. So we have to be careful not to inadvertently use it in a calculation. Given an $n \times n$ matrix A , the inverse, if it exists, is another $n \times n$ matrix A^{-1} such that

$$AA^{-1} = I, A^{-1}A = I$$

A matrix is called invertible or nonsingular if A^{-1} exists. In practice, it is only necessary to check one of these.

THEOREM 1.2. *If B is $n \times n$ and $AB = I$ or $BA = I$, then A invertible and $B = A^{-1}$.*

Proof. The proof of invertibility of A will need to wait until we talk about determinants. Assuming this and $AB = I$. Multiply both sides on the left by A^{-1} to get

$$A^{-1}AB = A^{-1}I$$

which simplifies to

$$IB = A^{-1}$$

or $B = A^{-1}$. If $BA = I$ the argument is similar. □

2. ELEMENTARY ROW OPERATIONS

Given a matrix A , we can try to simplify it by performing 3 types of *elementary row operations*:

- I Add a multiple of one row to another.
- II Interchange two rows
- III Multiply a row by a *nonzero* number.

A matrix is elementary if can be obtained from the identity by a single elementary row operation.

THEOREM 2.1. *If A is an $n \times m$ matrix and E an elementary $n \times n$ matrix, then EA is the same thing as doing the corresponding elementary row operation on A .*

Rather than giving a detailed proof, let us consider a few examples

$$E_1 = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}, E_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, E_3 = \begin{pmatrix} x & 0 \\ 0 & 1 \end{pmatrix}, A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then we can see by multiplying

$$E_1 A = \begin{pmatrix} a & b \\ c + xa & d + xb \end{pmatrix}, E_2 A = \begin{pmatrix} c & d \\ a & b \end{pmatrix}, E_3 A = \begin{pmatrix} xa & xb \\ c & d \end{pmatrix},$$

that this is exactly what the theorem says.

QUESTION 2.2. *What happens if we multiply the other way AE ?*

Another important fact is:

THEOREM 2.3. *An elementary matrix is invertible (or nonsingular).*

Proof. We have to check this for each of the three types. We just treat type I. given an elementary matrix E , corresponding to adding x times row i to row j , E^{-1} is the elementary matrix corresponding to adding $-x$ times row i to row j . $\square$

3. GAUSSIAN ELIMINATION

A matrix is in *reduced row echelon form* (RREF for short) if it looks like

$$(3.1) \quad \begin{pmatrix} 1 & * & 0 & \dots \\ 0 & 0 & 1 & \dots \\ 0 & 0 & 0 & \dots \end{pmatrix}$$

More precisely,

- (1) The first nonzero entry of a row (called a pivot) is 1 .
- (2) A pivot is the only nonzero entry in its column.
- (3) Rows of 0's have to be at the bottom of the matrix.

THEOREM 3.1 (Gaussian Elimination). *Any matrix A can be taken to a reduced row echelon matrix B by a finite sequence of elementary row operations. Furthermore, B depends only on A and not on the specific sequence of operations.*

Note that the statement is not true if B only in echelon form.

Proof. To prove that B exists, we just have to describe the algorithm and prove that it works. We do this in informal terms:

- (1) We'll use a "pointer", which is really a pair of integers to tell us what entry of A we're looking at as the algorithm progresses. We'll refer to this location (entry, row or column) as the *current* location. Initially, set the pointer to $(1, 1)$.
- (2) If the current entry of A (e.g. a_{11}) is zero switch the current row with the first row below it with a nonzero entry in the current column. If no such row exists, advance the pointer by $(1, 1)$ and repeat the algorithm from step 2.
- (3) Multiply current row by a constant so that it leads with 1.
- (4) Subtract multiples of the current row from rows above and below it to clear the current column.
- (5) Advance the pointer by $(1, 1)$, if possible, then repeat from step 2, otherwise stop.

The proof that the algorithm is correct can be done by mathematical induction, but we won't go into details.

Next, we have to prove that if we have two sequences of elementary matrices $E_1, \dots, E_N$ and $E'_1, \dots, E'_M$ such that $B = E_N \dots E_1 A$ and $B' = E'_M \dots E'_1 A$ are both in RREF, then $B = B'$. Note that $B' = CB$ where

$$C = E'_M \dots E'_1 E_1^{-1} \dots E_N^{-1}$$

Suppose that B is as in (3.1), then

$$CB = \begin{pmatrix} c_{11} & *c_{11} & 0 & \dots \\ 0 & 0 & c_{22} & \dots \\ 0 & 0 & 0 & \dots \end{pmatrix}$$

The only way this is in RREF is if $c_{11} = 1$, $c_{22} = 1$ etc. This forces $B = B'$. $\square$

The last part of the theorem says that B is really a function of A . A lot of software, such as MATLAB, will compute this.

4. SOLVING EQUATIONS

The original *raison d'être* for linear algebra is to understand solutions of systems of linear equations. In matrix form, we are given a matrix A and a column vector B , then we want to find all column vectors X such that

$$(4.1) \quad AX = B$$

The easiest strategy is to form the augmented matrix $[A|B]$. The notation means append B as the last column of A . Take it to RREF, or just to some intermediate stage that's simpler than the original. Then solve the new simpler system. The obvious question is why is this the same as solving the original system? To answer this, list the elementary matrices $E_1, \dots, E_N$ corresponding the operations used in the order they appear. Set $C = E_N \dots E_1$. Then the new system is

$$(4.2) \quad CAX = CB$$

THEOREM 4.1. *The systems (4.1) and (4.2) have the same solutions.*

Proof. Suppose that X is solution of (4.2). Then multiplying both sides by C^{-1} on the left and simplifying shows that X also satisfies (4.1). Multiplying (4.1) on the left by C shows the converse. $\square$

Suppose A is $n \times n$, is invertible and we happen to know what A^{-1} is, then the solution to (4.1) is simply

$$X = A^{-1}B$$

So how to we find A^{-1} ? The algorithm, called Gauss-Jordan elimination goes as follows:

- (1) Form $[A|I]$
- (2) Do elementary row operations to get it to $[L|R]$, such that L is in RREF.
- (3) There are two possibilities either L is I or it has a row of zeros. In the first case, $R = A^{-1}$. In the second case, A^{-1} does not exist.

To show this works, we need

THEOREM 4.2.

- (1) Suppose that $E_1, \dots, E_N$ are elementary matrices such that

$$E_N \dots E_1 [A|I] = [I|R]$$

Then $R = A^{-1}$.

- (2) Suppose that $E_1, \dots, E_N$ are elementary matrices such that

$$E_N \dots E_1 [A|I] = [L|R]$$

where C has a row of zeros, then A is not invertible.

Proof. In case (1), we can see that

$$E_N \dots E_1 [A|I] = [E_N \dots E_1 A | E_N \dots E_1 I]$$

Comparing this to what we are given, we obtain

$$E_N \dots E_1 A = I$$

$$E_N \dots E_1 I = R$$

From theorem 1.2

$$A^{-1} = E_N \dots E_1$$

This is R by the second equation.

The proof of (2) will be given later on □

5. LU DECOMPOSITION

A square matrix is called upper (respectively lower) triangular if all entries below (resp. above) the diagonal are zero. A triangular matrix will be called a unit triangular matrix if it has 1's on the diagonal. Instead of taking a square matrix all the way to RREF, it is simpler, and often just as useful, to take it to upper triangular form. For example, solving the system

$$AX = B$$

when A is upper triangular is easy. Start with the last equation,

$$a_{nn}x_n = b_n$$

this tells the value for x_n assuming the coefficient is nonzero. Then back substitute into the second to last

$$a_{n-1,n-1}x_{n-1} + a_{n-1,n}x_n = b_{n-1}$$

to find x_{n-1} etc. We can summarize this by the following

THEOREM 5.1. *If A is upper triangular with no zeros on the diagonal, the solution to the previous equation is given by*

$$(5.1) \quad \begin{aligned} x_n &= \frac{b_n}{a_{nn}} \\ x_{n-1} &= \frac{b_{n-1} - a_{n-1,n}x_n}{a_{n-1,n-1}} \\ &\dots \end{aligned}$$

Furthermore A^{-1} exists and is again upper triangular.

Proof. The first part follows from we said above. To prove the statement about the inverse, we apply the first part to where B are the columns of the identity. If $B = (0 \dots 0, 1, 0 \dots)^T$ is the i th column of I , then solution of

$$AX = B$$

gives the i th column of A^{-1} . The formulas (5.1) show that

$$x_n = \dots x_{i+1} = 0$$

This says that entries below the diagonal are zero. □

COROLLARY 5.2. *A lower triangular matrix A with nonzero diagonals is invertible, and A^{-1} is lower triangular.*

Proof. If A^T is upper triangular so $(A^T)^{-1}$ is upper triangular. One can check that this is the same thing as $(A^{-1})^T$. Therefore A^{-1} is lower triangular. □

It is clear we can take A to an upper triangular matrix by elementary row operations. In good cases, we use only operations of type I, where we add a multiple of one row to a row below it. This means that we can find a sequence of elementary matrices E_i which are unit lower triangular such that

$$E_N \dots E_1 A = U$$

where U is upper triangular. Setting $L = E_1^{-1} \dots E_N^{-1}$, we see that it is unit lower triangular. Therefore

$$A = LU$$

is a product of a unit lower triangular matrix with an upper triangular matrix. If such a factorization is possible, then we say that A admits an LU decomposition. Unfortunately, it is not always possible. For instance,

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

does not admit an LU factorization. In this case, we can fix the problem by switching rows. A matrix is called a permutation matrix if is obtained from the identity by permuting rows, or equivalently by a sequence of elementary operations of type II.

THEOREM 5.3. *Most invertible matrices admit a unique LU decomposition. If A admits an LU decomposition, then L and U are unique. Given an invertible A , there exists a permutation matrix P such that PA admits an LU decomposition.*

“Most” is not a precise mathematical term. The actual statement is that the set of invertible matrices admitting LU decompositions is *dense* in the set of all invertible matrices. Intuitively this means that given any matrix a small perturbation of coefficients will yield one admitting such a decomposition.

Proof. We give a proof in the 2×2 case. We do this directly rather than using Gaussian elimination. We have to solve

$$\begin{pmatrix} 1 & 0 \\ l_{21} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

or

$$u_{11} = a_{11}$$

$$u_{12} = a_{12}$$

$$l_{21}u_{11} = a_{21}$$

$$l_{21}u_{12} + u_{22} = a_{22}$$

This is very easy to solve. The first two equations give us the values for u_{11}, u_{12} , and

$$l_{12} = \frac{a_{21}}{u_{11}}$$

$$u_{22} = a_{22} - l_{12}u_{12}$$

gives the remaining variables. If $a_{11} = 0$ then these solutions won't be valid. In this case, we choose P in advance to switch the rows.

□

Given

$$PA = LU$$

as above, we can find the inverse

$$A^{-1} = U^{-1}L^{-1}P$$

easily. The obvious question is why bother with this since we already have a method for computing the inverse? The answer is that Gauss-Jordan is good for hand computation, but not as good for implementing on a computer. In this case, we need an algorithm which is both efficient and numerically stable in the sense that we can control round off errors. In classic Gaussian elimination, you might divide by a very small number, which can magnify the round-off error. This is a big topic, and we won't have much more to say about it in the future.