The Gauss-Jordan Elimination Algorithm

Solving Systems of Real Linear Equations

A. Havens

Department of Mathematics
University of Massachusetts, Amherst

January 24, 2018

Outline

Row Echelon Form

Definition

A matrix A is said to be in *row echelon form* if the following conditions hold

1. all of the rows containing nonzero entries sit above any rows whose entries are all zero,
2. the first nonzero entry of any row, called the *leading entry* of that row, is positioned to the right of the leading entry of the row above it,

Observe: the above properties imply also that all entries of a column lying below the leading entry of some row are zero.

Row Echelon Form

- Such a matrix might look like this:

$$\begin{bmatrix} a & * & * & * & * & * \\ 0 & b & * & * & * & * \\ 0 & 0 & 0 & c & * & * \\ 0 & 0 & 0 & 0 & 0 & d \end{bmatrix},$$

  where $a, b, c, d \in \mathbb{R}^\times$ are *nonzero* reals giving the leading entries, and '$*$' means an entry can be an arbitrary real number.
- Note the staircase-like appearance hence the word *echelon* (from french, for ladder/grade/tier).
- Also note that not every column has a leading entry in this example.

Row Echelon Form

A square matrix in row echelon form is called an *upper triangular matrix*.

E.g.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{bmatrix}$$

is a 4×4 upper triangular matrix.

Reduced Row Echelon Form

Definition

A matrix A is said to be in *reduced row echelon form* if it is in row echelon form, and additionally it satisfies the following two properties:

1. In any given nonzero row, the leading entry is equal to 1,
2. The leading entries are the only nonzero entries in their columns.

We will often abbreviate row echelon form to REF and reduced row echelon form to RREF.
Recall, we encountered the idea of reduced row echelon form of a matrix when we considered solving a linear system of equations using an augmented matrix.

Connection to Systems and Row Operations

- An augmented matrix in reduced row echelon form corresponds to a solution to the corresponding linear system.
- Thus, we seek an algorithm to manipulate matrices to produce RREF matrices, in a manner that corresponds to the legal operations that solve a linear system.
- We already encountered row operations, and these will be the desired manipulations in building such an algorithm.
- Though our initial goal is to reduce augmented matrices of the form $\left[\mathrm{A} \mid \mathbf{b}\right]$ arising from a general real linear system, the algorithms we describe work for any matrix A with a nonzero entry.

Three Elementary Row Operations

Definition

Given any matrix A let R_i and R_j denote rows of A, and let $s \in \mathbb{R}$ be a nonzero real number. Then the elementary row operations are

1. We may swap two rows, just as we may write the equations in any order we please. We notate a swap of the ith and jth rows of an augmented matrix by $R_i \leftrightarrow R_j$.
2. We may replace a row R_i with the row obtained by scaling the original row by a *nonzero* real number. We notate this by $sR_i \mapsto R_i$.
3. We may replace a row R_i by the difference of that row and a multiple of another row. We notate this by $R_i - sR_j \mapsto R_i$.

Row Equivalence

Definition

Two matrices A and B are said to be *row equivalent* if and only if there is a sequence of row operations transforming A into B.

Observation

This notion is well defined and an *equivalence relation*. In particular, if A is row equivalent to B then B is row equivalent to A, since row operations are invertible. And if A is row equivalent to B, and B is row equivalent to C, then A is row equivalent to C, since we can concatenate sequences of row operations. (And of course, trivially, every matrix is row equivalent to itself.)

A Proposition

Proposition

For a given matrix A, there is a unique row equivalent matrix in reduced row echelon form.

For any matrix A, let's denote the associated reduced row echelon form by $\mathrm{RREF}(A)$.

Proof.

The Gauss-Jordan Elimination Algorithm! □

Wait, what's that?

Leading Entries and Pivot Positions

Definition

- A *pivot position* of a matrix A is a location that corresponds to a leading entry of the reduced row echelon form of A, i.e., a_{ij} is in a pivot position if an only if $\mathrm{RREF}(\mathrm{A})_{ij} = 1$.
- A column of a matrix A containing a pivot position is called a *pivot column*.
- A *pivot entry*, or simply, a *pivot* is a nonzero number in a pivot position, which may be used to eliminate entries in its pivot column during reduction.

The number of pivot positions in a matrix is a kind of *invariant* of the matrix, called *rank* (we'll define rank differently later in the course, and see that it equals the number of pivot positions)

Pivoting Down

We are ready to describe the procedure for *pivoting downward*:

Definition

Let a_{ij} denote the entry in the ith row and jth column of an $m \times n$ matrix A. Suppose $a_{ij} \neq 0$. To *pivot downward* on the (i, j)th entry a_{ij} of A is to perform the following operations:

(i.) $\dfrac{1}{a_{ij}} R_i \mapsto R_i$,

(ii.) For each integer $k > i$, $R_{i+k} - a_{i+k,j} R_i \mapsto R_{i+k}$.

Said more simply, make the nonzero entry a_{ij} into a 1, and use this 1 to eliminate (make 0) all other entries directly below the (i, j)th entry.

Pivoting Up

- In the algorithm, we'll first pivot down, working from the leftmost pivot column towards the right, until we can no longer pivot down.
- Once we've finished pivoting down, we'll need to *pivot up*.
- The procedure is analogous to pivoting down, and works from the rightmost pivot column towards the left. Simply apply row operations to use the pivot entries to eliminate entries in each pivot column above the pivots. This is an algorithmic way to accomplish back-substitution while working with matrices.

Overview of the algorithm - Initialization and Set-Up

We present an overview of the Gauss-Jordan elimination algorithm for a matrix A with at least one nonzero entry.

Initialize: Set B_0 and S_0 equal to A, and set $k = 0$. Input the pair $(\mathrm{B}_0, \mathrm{S}_0)$ to the forward phase, step (1).

Important: we will always regard S_k as a sub-matrix of B_k, and row manipulations are performed simultaneously on the sub-matrix S_k and on its parent matrix B_k.

Overview of the steps - Forward Phase

1. Given an input $(\mathrm{B}_k, \mathrm{S}_k)$, search for the leftmost nonzero column of S_k. If there is none or S_k is empty, proceed to the backwards phase, step (5), with input B_k.
2. After finding a nonzero column, exchange rows of B_k as necessary to bring the first nonzero entry up to the top row of S_k (Any exchanges in this step alter both B_k and S_k). Label the corresponding nonzero entry in B_k by p_k (for pivot).
3. Pivot downwards on p_k in B_k to form matrix B_{k+1}.
4. Narrow scope to the sub-matrix S_{k+1} of B_{k+1} consisting of entries strictly to the right and strictly below p_k. Repeat the procedures in steps (1)-(3) with input $(\mathrm{B}_{k+1}, \mathrm{S}_{k+1})$.

Completing the Forward Phase

So, one loops over the first four steps until all pivot columns have been located and pivoting down has occurred in each pivot column.

The matrix B_k is in row echelon form, with leading 1s in each pivot position.

This completes the *forward phase*. and so the *backwards phase* commences with, step (5).

Overview of the steps - Backwards Phase

5. Start at the rightmost pivot of B_k and pivot up. Call the result B_{k+1}.
6. Move left to the next pivot column of B_{k+1} and pivot up. Increment k, and repeat this step until there are no remaining pivots.
7. The matrix B_k returned by the previous step upon termination is the output $\mathrm{RREF}(\mathrm{A})$.

A familiar 3×4 Example

We'll work with the augmented matrix

$$\mathrm{A} = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 1 & -2 & 3 & 6 \\ 4 & -5 & 6 & 12 \end{array}\right]$$

from last time.

1. The entry $a_{11} = 1$, so we can pivot down, using the row operations $R_2 - R_1 \mapsto R_2$ and $R_3 - 4R_1 \mapsto R_3$. This transforms the matrix into the row equivalent matrix

$$\mathrm{B}_1 = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & -3 & 2 & 0 \\ 0 & -9 & 2 & -12 \end{array}\right].$$

A familiar 3×4 Example

2. Ignoring the first row and column, we look to the 2×3 sub-matrix

$$\mathrm{S}_1 = \left[\begin{array}{cc|c} -3 & 2 & 0 \\ -9 & 2 & -12 \end{array}\right].$$

The top entry is nonzero, and so we may pivot downwards. We first have to scale this entry to make it 1. In the matrix B_1 we would apply the row operation $-\frac{1}{3}R_2 \mapsto R_2$. Then we eliminate the -9 below our pivot using $R_3 + 9R_2 \mapsto R_3$. The result is the matrix

$$\mathrm{B}_2 = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & -2/3 & 0 \\ 0 & 0 & -4 & -12 \end{array}\right],$$

which is row equivalent to A.

A familiar 3×4 Example

3. We now consider the sub-matrix

$$\mathrm{S}_2 = \begin{bmatrix} -4 & | & -12 \end{bmatrix}.$$

the only thing to do in the pivoting down algorithm is to make the first entry into a leading 1 by scaling, so we apply $-\frac{1}{4}R_3 \mapsto R_3$ to B_2. We now have an REF matrix row equivalent to A, with leading 1s in each pivot position:

$$\mathrm{B}_3 = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & -2/3 & 0 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

A familiar 3×4 Example

4. We've completed the *forward phase*, so we now we begin the *backwards phase*, searching from the right for a pivot column to begin pivoting up.
   The right-most column is not a pivot column, since 3 is not the leading entry in the bottom row. Thus, the column to its immediate left is where we begin pivoting up, applying the row operations $R_2 + \frac{2}{3}R_3 \mapsto R_2$ and $R_1 - R_3 \mapsto R_1$ to B_3 get

$$\mathrm{B}_4 = \left[\begin{array}{ccc|c} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

A familiar 3×4 Example

5. Moving left once more, we use the pivot in the $(2,2)$ position to pivot up in the second column from the left. The only row operation we need is $R_1 - R_2 \mapsto R_1$, yielding

$$\mathrm{B}_5 = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

By construction, this is row equivalent to A.

6. There are no more pivots, and the matrix is clearly in reduced row echelon form. Thus $\mathrm{RREF}(\mathrm{A}) = \mathrm{B}_5$.

An Exercise

Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Show that

$$\mathrm{RREF}(A) = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Solved Systems with a Unique Solution

Proposition (Unique solutions for n linear equations in n variables)

Suppose A *is an* $n \times n$ *matrix, and* $\mathbf{b}$ *is a column vector with* n *entries. If every column of* A *is a pivot column, then the reduced row echelon form of* A *is*

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{bmatrix},$$

the matrix with entries $a_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$.

Solved Systems with a Unique Solution

Proposition (Unique solutions for n linear equations in n variables)

In this case, the corresponding linear system of equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \vdots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= b_n \end{cases}$$

has a unique solution.

Solved Systems with a Unique Solution

Proposition (Unique solutions for n linear equations in n variables)

In particular, if

$$\mathrm{RREF}\left(\left[\mathrm{A} \mid \mathbf{b}\right]\right) = \left[\begin{array}{cccc|c} 1 & 0 & \dots & 0 & v_1 \\ 0 & 1 & \dots & 0 & v_2 \\ \vdots & \dots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & v_n \end{array}\right],$$

then the corresponding system of linear equations has solution $x_1 = v_1, \dots, x_n = v_n$.

An Example with a Free Variable

Consider the system

$$\begin{cases} x + 2y + 3z &= -1 \\ 4x + 5y + 6z &= 0 \\ 7x + 8y + 9z &= 1\,. \end{cases}$$

The corresponding augmented matrix is

$$\mathrm{M} = \left[\begin{array}{ccc|c} 1 & 2 & 3 & -1 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 1 \end{array}\right].$$

An Example with a Free Variable

Using row operations from the exercise above, one easily obtains

$$\mathrm{RREF}(\mathrm{M}) = \left[\begin{array}{ccc|c} 1 & 0 & -1 & 5/3 \\ 0 & 1 & 2 & -4/3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

This tells us that the variable z is *free*, as any real value of z satisfies the equation $0x + 0y + 0z = 0$.

An Example with a Free Variable

The remaining equations are

$$\begin{cases} x - z = 5/3\,, \\ y + 2z = -4/3\,. \end{cases}$$

Thus, solving for x and y in terms of the free variable z, we can express the solution as $(5/3 + z, -4/3 - 2z, z)$, for any real number z. This system therefore has infinitely many solutions.

Consistency

- If no solution exists, the system is said to be *inconsistent*. Otherwise, it is said to be a consistent system.
- For an augmented matrix $\left[\mathrm{A} \mid \mathbf{b}\right]$, consistency requires that the pivots all occur in positions within the coefficient matrix A. Why?
- If there is a pivot in the column $\mathbf{b}$, then the corresponding row of $\mathrm{RREF}(\mathrm{A})$ is a row of zeros. This corresponds to an equation of the form $0 = a$ for nonzero a, which is inconsistent.
- An important fact we'll use later is that such an inconsistency arises when there is a way to some combine rows (and columns) of A nontrivially to obtain a zero row (or zero column). To understand the significance of this, we must study the geometry of *vectors*.

An Inconsistent System

Consider the system

$$\begin{cases} x + 2y + 3z &= 12 \\ 4x + 5y + 6z &= 11 \\ 7x + 8y + 9z &= -10\,. \end{cases}$$

The corresponding augmented matrix is

$$\mathrm{A} = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 12 \\ 4 & 5 & 6 & 11 \\ 7 & 8 & 9 & -10 \end{array}\right].$$

An Inconsistent System

After applying the appropriate row operations, one will find that

$$\mathrm{RREF}(\mathrm{A}) = \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Since $0 \neq 1$, this system is not consistent.

There is no set of values for x, y, and z that can satisfy all three equations at once.

Any pair of planes from the above system intersect in a set of lines, and one can show that the three lines of intersection never meet.

Any Two Lines. . .

- We'll apply the Gauss-Jordan elimination algorithm to abstractly give a complete general solution to systems of two equations in two variables:

$$\begin{cases} ax + by &= e \\ cx + dy &= f \end{cases} \longleftrightarrow \left[\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right] .$$

- We assume temporarily that $a \neq 0$. We will discuss this assumption in more depth later.
- With this assumption, we may pivot down from the top-left entry.
- Thus we apply the row operation $aR_2 - cR_1 \mapsto R_2$.

Any Two Lines. . .

- Applying $aR_2 - cR_1 \mapsto R_2$ yields:

$$\left[\begin{array}{cc|c} a & b & e \\ c & d & f \end{array}\right] \longmapsto \left[\begin{array}{cc|c} a & b & e \\ 0 & ad - bc & af - ce \end{array}\right].$$

- We see that if $ad - bc = 0$, then either there is no solution, or we must have $af - ce = 0$.
- Let's plug on assuming that $ad - bc \neq 0$. We may eliminate the upper right position held by b in the coefficient matrix by $(ad - bc)R_1 - bR_2 \mapsto R_1$.

Any Two Lines. . .

- Applying $(ad - bc)R_1 - bR_2 \mapsto R_1$, yields

$$\left[\begin{array}{cc|c} a(ad-bc) & 0 & (ad-bc)e - b(af-ce) \\ 0 & ad-bc & af-ce \end{array}\right]$$

$$= \left[\begin{array}{cc|c} a(ad-bc) & 0 & ade-abf \\ 0 & ad-bc & af-ce \end{array}\right].$$

Any Two Lines. . .

- Since we assumed a and $ad - bc$ nonzero, we may apply the final row operations $\frac{1}{a(ad-bc)}R_1 \mapsto R_1$ and $\frac{1}{ad-bc}R_2 \mapsto R_2$ to obtain
$$\left[\begin{array}{cc|c} 1 & 0 & (de - bf)/(ad - bc) \\ 0 & 1 & (af - ce)/(ad - bc) \end{array}\right],$$
so we obtain the solution as
$$x = \frac{de - bf}{ad - bc}, \quad y = \frac{af - ce}{ad - bc}.$$
- About that assumption, $a \neq 0 \ldots$

Any Two Lines. . .

- Note that if $a = 0$ but $bc \neq 0$, the solutions are still well defined.
- One can obtain the corresponding expressions with $a = 0$ substituted in by instead performing elimination on

$$\left[\begin{array}{cc|c} 0 & b & e \\ c & d & f \end{array}\right],$$

where the first step would be a simple row swap.

Any Two Lines...

- However, if $ad - bc = 0$, there is no hope for the unique solution expressions we obtained, though there may still be solutions, or there may be none at all.
- How do we characterize this failure geometrically?
- A solution is unique precisely when the two lines $ax + by = e$ and $cx + dy = f$ have distinct slopes, and thus intersect in a unique point. One can show that $ad - bc$ measures whether the slopes are distinct!
- If $ad - bc = 0$, there could be no solutions at all (two distinct parallel lines) or infinitely many solutions!

Existence and Uniqueness of Solutions for Two-Dimensional Systems

Proposition

For a given two variable linear system described by the equations

$$\begin{cases} ax + by &= e \\ cx + dy &= f \end{cases}$$

the quantity $ad - bc = 0$ *if and only if the lines described by the equations have the same slope.*

Corollary

There exists a unique solution to the system above if and only if $ad - bc$ *is nonzero.*

Proofs?

Proof.

The proof of this corollary follows immediately from our application and discussion of Gauss-Jordan applied to the system. □

In the past, I've assigned the proof of the above proposition as an exercise, as all it involves is a little algebra and an attention to the different cases.

This time, I'll leave the proof here in the slides, for your perusal.

Another way it can be proven will be uncovered later in the course, when we discuss determinants. This gives an alternate (and in some sense, dual) geometric interpretation, which involves vectors and areas.

Proof of the Proposition?

Proof.

We must show two directions, since this is an *if and only if* statement.

Namely, we must show that if the lines have the same slopes, then $ad - bc = 0$, and conversely, if we know only that $ad - bc = 0$, we must deduce the corresponding lines possess the same slopes.

Let's prove the former. We have several cases we need to consider.

Proof of the Proposition?

Proof.

First, let's suppose that none of the coefficients are zero, in which case we can write each equation in *slope-intercept form*:

$$ax + by = e \longleftrightarrow y = -\frac{a}{b}x + \frac{e}{b}\,,$$

$$cx + dy = f \longleftrightarrow y = -\frac{c}{d}x + \frac{f}{d}\,,$$

Applying the assumption that the lines have identical slopes, we obtain

$$-\frac{a}{b} = -\frac{c}{d} \implies ad = bc \implies ad - bd = 0\,. \qquad (\star)$$

Proof of the Proposition?

Proof.

On the other hand, if for example, $a = 0$, then the first equation is $by = e$, which describes a horizontal line (we must have $b \neq 0$ if this equation is meaningful).

This tells us that the other equation is also for a horizontal line, so $c = 0$ and consequently $ad - bc = 0 \cdot d - b \cdot 0 = 0$.

A nearly identical argument works when the lines are vertical, which happens if and only if $b = 0 = d$.

Proof of the Proposition?

Proof.

It now remains to show the converse, that if $ad - bc = 0$, we can deduce the equality of the lines' slopes.
Provided neither a nor d are zero, we can work backwards in the equation $(\star)$:

$$ad - bc = 0 \implies -\frac{a}{b} = -\frac{c}{d}\,.$$

Else, if $a = 0$ or $d = 0$ and $ad - bc = 0$, then since $0 = ad = bc$, either $b = 0$ or $c = 0$.

But a and b cannot both be zero if we have a meaningful system (or indeed, the equations of two lines).

Proof of the Proposition?

Proof.

Thus if $a = 0$ and $ad - bc = 0$, then $c = 0$ and the lines are both horizontal.

Similarly, if $d = 0$ and $ad - bc = 0$, then $b = 0$ the system consists of two vertical lines. □

Recall the general questions we want to ask about solutions to a linear system:

- Existence: does *some* solution exist, i.e. is there some point satisfying all the equations? Equivalently, is the system consistent?
- Uniqueness: If a solution exists is it the only one?

Our preceding discussion leaves us with a few valuable results answering these questions.

A General Existence and Uniqueness Proposition

Proposition

A system with augmented matrix $\left[\mathrm{A} \mid \mathbf{b}\right]$ *is consistent if and only if there is no row of the form*

$$\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & b \end{array}\right]$$

in $\mathrm{RREF}(\mathrm{A})$*, or equivalently, if all pivot positions of* $\left[\mathrm{A} \mid \mathbf{b}\right]$ *occur within the coefficient matrix* A *(and thus correspond to pivot positions of* A*).*

If the system is consistent, then either there is a unique solution and no free variables, or there are infinitely any solutions, which can be expressed in terms of free variables. The number of free variables is the number of non-pivot columns of the coefficient matrix.

Geometric Existence and Uniqueness for 2 Lines

For 2 linear equations in 2 variables, we have a nice geometric answer to the existence and uniqueness questions.

Proposition

For a system of 2 equations in 2 variables, there is a unique solution precisely when the lines have distinct slopes, in which case there are two pivot positions and the quantity $ad - bc$ is nonzero.

If $ad - bc = 0$, the lines have identical slopes, and the system is consistent if and only if $af - ce = 0 = de - bf$, in which case the equations describe the same line. The system is inconsistent precisely when the lines are parallel and distinct.

We'd like a geometric interpretation of the general existence and uniqueness theorem. At least, it would be nice to understand its implications for planes in three dimensions.

What About Planes?

- Will three nonparallel planes always intersect in a unique point? We saw above that the answer is no.
- For two nonparallel planes, the geometric intersection is a line. We don't get unique solutions in this case, but we can still use row reduction to describe the line. Our free variable example above shows algebraically how to describe such a line in terms of a parameter (in this case, arising from one of the variables).
- Try to picture an inconsistent system of planes.

Towards Geometry in $\mathbb{R}^3$ and $\mathbb{R}^n$

We're missing an important tool for understanding how these things fit together.

That tool is vectors. They will play a role for lines in $\mathbb{R}^3$ analogous to slopes (we need more information than a single real number to define direction of a line in three dimensions.)

With vectors we will be able to better visualize lines in space, and understand matrices more deeply.

In fact, we can rephrase linear systems as problems involving vector arithmetic. We can even use vectors to understand where the equation of a plane comes from.

A deep understanding requires that we discuss the geometry and algebra of vectors.

Homework for Week 1

- Go to `http://people.math.umass.edu/~havens/math235-4.html`, review the policy and expectations for this section and the overall course.
- Read the course overview on the section website.
- Use the course ID to log into MyMathLab.
- Please read sections 1.1-1.3 of the textbook for Friday.