HOW TO SOLVE LINEAR SYSTEMS BY HAND AND WITH MATLAB

DIEGO MALDONADO

Abstract. Notes on linear systems, reduced row echelon form and rank of a matrix, and criterion of solvability of systems.

Math551:Applied Matrix Theory
Department of Mathematics
Kansas State University

1. Introduction

The purpose of these notes is to introduce the matrix viewpoint towards solving systems of linear equations. Here you will learn how to perform elementary row operations, how to obtain the reduced row echelon form and rank of a given matrix, and how to solve systems of any size.

2. Solving linear systems

2.1. $m \times n$ **systems of linear equations.** An $m \times n$ system of linear equations has the form

$$(S) \left\{ \begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & = & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & \cdots & + & a_{mn}x_n & = & b_m \end{array} \right.$$

where m is the number of equations and n is the number of unknowns. We collet the coefficients a_{ij} (where $i = 1, \dots, m$ and $j = 1, \dots, n$) as the entries of the following $m \times n$ matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

This matrix A is called the *matrix of coefficients* of the system (S). The right-hand side constants in (S) form the coefficients of the $m \times 1$ *right-hand side vector*

Date: January 30, 2013.

associated to (S), usually denoted as

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

When the right-hand side vector b equals the zero vector, the system (S) is called a *homogeneous system*; otherwise, it is called an *inhomogeneous system.*

The *unknowns* $x_1, x_2, \cdots, x_n$ make up the entries of the $n \times 1$ *vector of unknowns*, that is,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

With this notation, and using the definition of matrix product, the system (S) is concisely expressed in *matrix form* as $Ax = b$.

If we further denote by u_j the j-th column of A, that is,

$$u_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad u_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}, \quad \cdots, \quad u_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix},$$

then the system (S) can be expressed in *vector form* as follows

$$x_1 u_1 + x_2 u_2 + \cdots + x_n u_n = b. \tag{2.1}$$

The matrix and vector forms of a system will play a crucial role in the process of solving that system.

For example, the system

$$(S) \left\{ \begin{array}{rcrcrcr} & & 2x_2 & + & 3x_3 & = & 8 \\ 2x_1 & + & 3x_2 & + & x_3 & = & 5 \\ x_1 & - & x_2 & - & 2x_3 & = & -5 \end{array} \right.$$

is a 3×3 system (that is, with 3 equations and 3 unknowns) with matrix of coefficients

$$A = \begin{bmatrix} 0 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & -1 & -2 \end{bmatrix},$$

right-hand side vector

$$b = \begin{bmatrix} 8 \\ 5 \\ -5 \end{bmatrix}$$

and vector of unknowns

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Thus, the system (S) can be written in matrix form $Ax = b$, that is,

$$\begin{bmatrix} 0 & 2 & 3 \\ 2 & 3 & 1 \\ 1 & -1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ -5 \end{bmatrix}$$

and it can be written in vector form as

$$x_1 \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ -5 \end{bmatrix}.$$

2.2. **Elementary operations on the equations of a system.** Our main method for solving a linear system will be based on the execution of three *elementary operations* on its equations. Namely, interchanging two equations, updating an equation by multiplying it by a non-zero scalar, and updating an equation by adding to it a multiple of another equation. Here is how they work.

2.2.1. *Interchanging two equations.* The most basic operation on the equations of a system is the interchange of two equations. For instance, consider the system

$$(S) \left\{ \begin{array}{rcrcl} 3x_1 & + & x_2 & = & 2 \\ -x_1 & + & 2x_2 & = & 5 \\ x_1 & + & x_2 & = & 1 \end{array} \right.$$

The action of interchanging the first and second equations of (S) is denoted as $(E1) \overset{(S)}{\leftrightarrow} (E2)$; clearly, the '$E$' stands for 'Equation'. Such an interchange yields the system

$$(S_1) \left\{ \begin{array}{rcrcl} -x_1 & + & 2x_2 & = & 5 \\ 3x_1 & + & x_2 & = & 2 \\ x_1 & + & x_2 & = & 1 \end{array} \right.$$

Notice that from (S_1) we can recover (S) by doing $(E1) \overset{(S_1)}{\leftrightarrow} (E2)$. Technically, the systems (S) and (S_1) are not the same and we cannot say that (S) equals (S_1). Instead we say that (S) and (S_1) are *equivalent*, in the sense that a pair (x_1, x_2) is a solution to (S) if and only if it is a solution to (S_1). Interchanging equations in a system is then a reversible operation that yields equivalent systems.

2.2.2. *Updating an equation by multiplying it by a non-zero scalar.* Consider the system

$$(S) \left\{ \begin{array}{rcrcl} x_1 & + & x_2 & = & 7 \\ 4x_1 & - & 2x_2 & = & 8 \end{array} \right.$$

Instead of interchanging equations in (S) we now update one, say, the second equation, by multiplying it by a non-zero real number, for instance, $\frac{1}{4}$. This operation is denoted by $\frac{1}{4}(E2) \overset{(S)}{\mapsto} (E2)$ and yields the system

$$(S_1)\left\{\begin{array}{ccccc} x_1 & + & x_2 & = & 7 \\ x_1 & - & \frac{1}{2}x_2 & = & 2 \end{array}\right.$$

Notice that the original system (S) can be recovered from (S_1) by doing $4(E2) \overset{(S_1)}{\mapsto} (E2)$. This tells us that the operation of updating an equation by replacing it with a non-zero multiple of itself produces equivalent systems. It is important that the multiple is a non-zero multiple. If we updated an equation by multiplying it by zero, we would get $0 = 0$ and completely lose the information from that equation! Then, that operation would not yield an equivalent system since part of the original information would be lost. Back to the main point: if we update an equation by multiplying it by a non-zero number, say $\lambda \in \mathbb{R} \setminus \{0\}$, we can always undo that operation (and come back to the original system) through multiplication by $\frac{1}{\lambda}$.

2.2.3. *Updating an equation by adding to it a multiple of another equation.* Let us exemplify this operation with the following system

$$(S)\left\{\begin{array}{ccccccc} 3x_1 & + & 2x_2 & - & x_3 & = & 9 \\ 5x_1 & - & x_2 & + & x_3 & = & 8 \\ -x_1 & + & 3x_2 & + & 4x_3 & = & 5 \end{array}\right.$$

We now update an equation in (S) by adding to it a multiple of another equation in (S). For instance, let us update the second equation by adding to it twice the first one. In symbols, this operation is denoted by $(E2) + 2(E1) \overset{(S)}{\mapsto} (E2)$. Of course, by $(E2) + 2(E1)$ we mean

$$\begin{array}{ccccccc} 5x_1 & - & x_2 & + & x_3 & = & 8 \\ 6x_1 & + & 4x_2 & - & 2x_3 & = & 18 \\ \hline \\ 11x_1 & + & 3x_2 & - & x_3 & = & 26, \end{array}$$

so that the update $(E2) + 2(E1) \overset{(S)}{\mapsto} (E2)$ produces

$$(S_1)\left\{\begin{array}{ccccccc} 3x_1 & + & 2x_2 & - & x_3 & = & 9 \\ 11x_1 & + & 3x_2 & - & x_3 & = & 26 \\ -x_1 & + & 3x_2 & + & 4x_3 & = & 5 \end{array}\right.$$

As you might have guessed, in this case we can also retrieve the original system (S). Indeed, that is achieved by doing $(E2) - 2(E1) \overset{(S_1)}{\mapsto} (E2)$. The operation of updating an equation by adding to it a multiple of another equation is then a reversible operation that produces equivalent systems.

2.2.4. *Why do we care about the elementary operations on the equations?* Suppose that you run into the German word “Stuhl”. Suppose that you don't recognize that word and that you are curious about its meaning. Luckily, you have with you translation dictionaries: German-French, French-Italian, and Italian-English translation dictionaries. It's clear what to do. The German-French dictionary shows “Stuhl= chaise”, if you can interpret the French word “chaise”, you're done. Otherwise, you just translated your problem into an equivalent problem and keep going. The French-Italian dictionary shows “chaise=sedia”. If you are not familiar with that Italian word, at least you know that “Stuhl” means “sedia” and go for the last dictionary. Finally, the Italian-English dictionary shows “sedia=chair”, so that “Stuhl = chair”, and now you interpret the word “chair” and understand the meaning of “Stuhl”.

When it comes to linear systems, the elementary operations function as translation dictionaries. They help us translate a given system into equivalent systems, until we get to an equivalent system that we can easily interpret. Next, we describe the method of translating and interpreting systems in terms of elementary operations.

2.3. **The elementary operations on the equations in action: Gaussian elimination.** The simple form of the following system makes it immediate to interpret

$$(S_0)\left\{\begin{array}{rcrcl} x_1 & + & 0x_2 & = & 3 \\ 0x_1 & + & x_2 & = & 2 \end{array}\right.$$

Indeed, it's obvious that$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ solves the system. In what follows, we examine the possibility of taking an arbitrary system (S) and translating it (via elementary operations on the equations) into its simplest (but equivalent) expression. This procedure is called *Gaussian elimination* and we first exemplify it by considering the system

$$(S)\left\{\begin{array}{rcrcl} 2x_1 & + & x_2 & = & 11 \\ -x_1 & + & 3x_2 & = & 12 \end{array}\right.$$

In the notation of the elementary operations we obtain

$$(S)\left\{\begin{array}{rcrcl} 2x_1 & + & x_2 & = & 11 \\ -x_1 & + & 3x_2 & = & 12 \end{array}\right. \qquad (E1) \overset{(S)}{\leftrightarrow} (E2)$$

$$(S_1)\left\{\begin{array}{rcrcl} -x_1 & + & 3x_2 & = & 12 \\ 2x_1 & + & x_2 & = & 11 \end{array}\right. \qquad (-1)(E1) \overset{(S_1)}{\mapsto} (E1)$$

$$(S_2)\left\{\begin{array}{rcrcr} x_1 & - & 3x_2 & = & -12 \\ 2x_1 & + & x_2 & = & 11 \end{array}\right. \qquad (E2) + (-2)(E1) \overset{(S_2)}{\mapsto} (E2)$$

$$(S_3)\left\{\begin{array}{rcrcr} x_1 & - & x_2 & = & -12 \\ 0x_1 & + & 7x_2 & = & 35 \end{array}\right.$$

At this point we have eliminated x_1 from the second equation in (S_3) which renders (S_3) as an easier to interpret system (because in this system we can readily see that x_2 must be $35/7 = 5$, which is not apparent from the previous systems). This elimination process is called *forward elimination.* If we solved for x_2 in $(E2)$ of (S_3) and substituted its value in $(E1)$ (of (S_3)), we would be doing *backward substitution.* Instead of doing that, we will eliminate x_2 from $(E1)$ as follows

$$(S_3)\left\{\begin{array}{rcrcr} x_1 & - & x_2 & = & -12 \\ 0x_1 & + & 7x_2 & = & 35 \end{array}\right. \qquad\qquad \tfrac{1}{7}(E2) \overset{(S_3)}{\mapsto} (E2)$$

$$(S_4)\left\{\begin{array}{rcrcr} x_1 & - & 3x_2 & = & -12 \\ 0x_1 & + & x_2 & = & 5 \end{array}\right. \qquad\qquad (E1)+3(E2) \overset{(S_4)}{\mapsto} (E1)$$

$$(S_5)\left\{\begin{array}{rcrcl} x_1 & + & 0x_2 & = & 3 \\ 0x_1 & + & x_2 & = & 5 \end{array}\right.$$

and immediately interpret that $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ is the only solution to (S_5) and, by equivalence, the only solution to (S). This process of coming back to eliminate x_2 is an instance of *backward elimination.* Performing forward elimination followed by backward elimination is algorithmically more convenient than doing forward elimination followed by backward substitution. The process of doing forward elimination followed by backward elimination is called Gaussian elimination.

2.3.1. *Translation is not interpretation.* For instance, if you don't know the meaning of the English word *anthropogenic*, most likely you don't know its meaning in any other language. Then, translation dictionaries will only provide you with equivalent and equally unfamiliar words. On the other hand, if instead of translation dictionaries you use an English dictionary, you will be able to interpret the word anthropogenic and find out that it means 'caused or produced by humans', as in 'anthropogenic global warming'.

By all that we mean that it should be clear that the process of Gaussian elimination translates a system into equivalent systems that are easier to interpret. However, there is always the task of interpreting at least one of those simpler-looking systems. In the previous example we translated (S) into (S_5) and immediately interpreted (that is, solved) (S_5). There will be cases when even after translating a system into its simplest expression, the task of interpreting it (that is, solving it) will be a non-trivial one (or, at least, not as easy as (S_5) was). Before addressing that issue, let's simplify the notation in the Gaussian elimination.

2.4. **Wake up, System, the (augmented) matrix has you.** By now, you must have realized that, when performing the elementary operations on the equations of a system, we kept writing the variables (e.g., x_1 and x_2) when such operations mainly involve only the coefficients of the variables. With a little bit of bookkeeping and the use of matrix notation we can avoid writing the variables on and on.

All the information of a system (S) is encoded in the *augmented*[1] *matrix of (coefficients of) the system* (S), which is defined as the $m \times (n+1)$ matrix $[A, b]$, where

[1]To augment=to make larger; enlarge in size, number, strength, or extent; increase.

A is the matrix of coefficients and b is the right-hand side vector of (S). That is, the augmented matrix of a system is obtained by attaching the right-hand side vector to the matrix of coefficients as an $(n+1)$-th column.

For example, the system

$$(S)\left\{\begin{array}{rcrcr} 2x_1 & + & x_2 & = & 11 \\ -x_1 & + & 3x_2 & = & 12 \end{array}\right.$$

from Subsection 2.3 has the following augmented matrix

$$M = \begin{bmatrix} 2 & 1 & 11 \\ -1 & 3 & 12 \end{bmatrix}.$$

As we can see, each row of M encodes the information of a corresponding equation in (S).

Let us introduce the concept of *elementary row operations* (analogous to the elementary operations on equations) and use, for instance, the notation $(R1) \overset{(M)}{\leftrightarrow} (R2)$ (which is analogous to the notation $(E1) \overset{(S)}{\leftrightarrow} (E2)$) to indicate the interchange of the first and second row of the matrix M, and similarly for the other elementary operations. Of course, now 'R' stands for 'Row'. The chain of translations in Subsection 2.3 can be recast as

$$\begin{aligned}
M &= \begin{bmatrix} 2 & 1 & 11 \\ -1 & 3 & 12 \end{bmatrix} & (R1) &\overset{(M)}{\leftrightarrow} (R2) \\
M_1 &= \begin{bmatrix} -1 & 3 & 12 \\ 2 & 1 & 11 \end{bmatrix} & (-1)(R1) &\overset{(M_1)}{\mapsto} (R1) \\
M_2 &= \begin{bmatrix} 1 & -3 & -12 \\ 2 & 1 & 11 \end{bmatrix} & (R2) + (-2)(R1) &\overset{(M_2)}{\mapsto} (R2) \\
M_3 &= \begin{bmatrix} 1 & -3 & -12 \\ 0 & 7 & 35 \end{bmatrix} & \tfrac{1}{7}(R2) &\overset{(M_3)}{\mapsto} (R2) \\
M_4 &= \begin{bmatrix} 1 & -3 & -12 \\ 0 & 1 & 5 \end{bmatrix} & (R1) + 3(R2) &\overset{(M_4)}{\mapsto} (R1) \\
M_5 &= \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 5 \end{bmatrix},
\end{aligned}$$

where the matrix M_j is the augmented matrix of the system (S_j), for $j = 1, \ldots, 5$. Now, the matrix M_5 is the simplest possible expression of the matrix M and we go back to the system notation and recover (S_5), which is the simplest expression for (S).

We now turn to the fact that every matrix can be reduced to its simplest expression.

2.5. **The elementary row operations and Gaussian elimination.** What do we mean by the simplest possible (and equivalent) form of a matrix? If you look carefully, what makes the systems (S_0) and (S_5) in Subsection 2.3 extremely approachable is that they have the 2×2 identity matrix I_2 as their matrix of coefficients. You can see the matrix I_2 in M_5 (disregard its last column).

This is very insightful because it's telling us that, in order to obtain an easier-to-interpret system, we should perform our elementary row operations in such a way as to produce an identity matrix (or something close to it). Let's practice.

Example 2.1. A system with exactly one solution. Consider the 3×3 system

$$(S_1)\begin{cases} & & 2x_2 & + & 3x_3 & = & 8 \\ 2x_1 & + & 3x_2 & + & x_3 & = & 5 \\ x_1 & - & x_2 & - & 2x_3 & = & -5 \end{cases}$$

whose associated augmented matrix is given by

$$M_1 = \begin{bmatrix} 0 & 2 & 3 & 8 \\ 2 & 3 & 1 & 5 \\ 1 & -1 & -2 & -5 \end{bmatrix}.$$

The idea is to perform elementary row operations on M_1 so as to produce the 3×3 identity matrix I_3 (or perhaps something close to it), inside M_1. Here is one way (you can find other ways) to proceed in that direction. In what follows we will make the notation lighter, but still self-explanatory. The first column of I_3 is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, let's try to make the first column of M_1 look like $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ through elementary row operations. We can do

$$M_1 = \begin{bmatrix} 0 & 2 & 3 & 8 \\ 2 & 3 & 1 & 5 \\ 1 & -1 & -2 & -5 \end{bmatrix} \qquad (R3) \leftrightarrow (R1)$$

$$M_2 = \begin{bmatrix} 1 & -1 & -2 & -5 \\ 2 & 3 & 1 & 5 \\ 0 & 2 & 3 & 8 \end{bmatrix} \qquad (R2) + (-2)(R1) \mapsto (R2)$$

$$M_3 = \begin{bmatrix} 1 & -1 & -2 & -5 \\ 0 & 5 & 5 & 15 \\ 0 & 2 & 3 & 8 \end{bmatrix}$$

Good. The first row operation created a "1" in the right place of M_2 and the second one eliminated the "2" in $(M_2)_{21}$. The next step is to make the second column of M_3

look like $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, which is the second column of I_3. Again, there are several ways to accomplish that. Let's first turn the "5" at $(M_3)_{22}$ into a "1", and then use that "1" to turn all other entries along the second column into zeros. Like this

$$M_3 = \begin{bmatrix} 1 & -1 & -2 & -5 \\ 0 & 5 & 5 & 15 \\ 0 & 2 & 3 & 8 \end{bmatrix} \qquad \tfrac{1}{5}(R2) \mapsto (R2)$$

$$M_4 = \begin{bmatrix} 1 & -1 & -2 & -5 \\ 0 & 1 & 1 & 3 \\ 0 & 2 & 3 & 8 \end{bmatrix} \qquad (R1) + (R2) \mapsto (R1)$$

$$M_5 = \begin{bmatrix} 1 & 0 & 1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 2 & 3 & 8 \end{bmatrix}$$

and now we take care of the "2" at $(M_5)_{32}$ as follows

$$M_5 = \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 2 & 3 & 8 \end{bmatrix} \qquad (R3) + (-2)(R2) \mapsto (R3)$$

$$M_6 = \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

Almost there. The final step is to make the third column of M_6 look like $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, which is the third column of I_3. We got the "1" at $(M_6)_{33}$ for free. Let's use it to turn all the other coefficients in that column into zero. Here is one way

$$M_6 = \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 2 \end{bmatrix} \qquad (R1) + (R3) \mapsto (R1)$$

$$M_7 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 2 \end{bmatrix} \qquad (R2) + (-1)(R3) \mapsto (R2)$$

$$M_8 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

And there is I_3 inside M_8! Success!

Interpretation. Next, we must interpret M_8, that is, we think of the linear system that has M_8 as its associated augmented matrix, that is,

$$(S_8)\left\{\begin{array}{rcrcrcr} x_1 & + & 0x_2 & + & 0x_3 & = & 0 \\ 0x_1 & + & x_2 & + & 0x_3 & = & 1 \\ 0x_1 & + & 0x_2 & + & x_3 & = & 2 \end{array}\right.$$

from which we immediately see that the system (S_8) has the only solution $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$. Since (S_8) is equivalent to the original system (S), the system (S) also has $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$ as its only solution. Interpreting the matrix M_8 was very easy. Will it always be that simple? No. Sometimes it will be simpler and sometimes more complicated. Let's see how.

Example 2.2. A system with no solutions. Consider the system

$$(S)\left\{\begin{array}{rcrcrcr} x_1 & + & x_2 & - & x_3 & = & 2 \\ 2x_1 & - & 3x_2 & + & 8x_3 & = & 9 \\ 3x_1 & - & 2x_2 & + & 7x_3 & = & 5 \end{array}\right.$$

whose augmented matrix is given by

$$M = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 2 & -3 & 8 & 9 \\ 3 & -2 & 7 & 5 \end{bmatrix}.$$

Our goal is to simplify M through the execution of elementary row operations. The idea is always the same, forcing the identity matrix I_3 (or something close to it) to appear inside an equivalent form of M. We deal with the first column as follows

$$M = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 2 & -3 & 8 & 9 \\ 3 & -2 & 7 & 5 \end{bmatrix} \qquad (R2) + (-2)(R1) \mapsto (R2)$$

$$M_1 = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 0 & -5 & 10 & 5 \\ 3 & -2 & 7 & 5 \end{bmatrix} \qquad (R3) + (-3)(R1) \mapsto (R3)$$

$$M_2 = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 0 & -5 & 10 & 5 \\ 0 & -5 & 10 & -1 \end{bmatrix}$$

Next we address the second column

$$M_2 = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 0 & -5 & 10 & 5 \\ 0 & -5 & 10 & -1 \end{bmatrix} \qquad (-\tfrac{1}{5})(R2) \mapsto (R2)$$

$$M_3 = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & -2 & -1 \\ 0 & -5 & 10 & -1 \end{bmatrix} \qquad (R1) + (-1)(R2) \mapsto (R1)$$

$$M_4 = \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -2 & -1 \\ 0 & -5 & 10 & -1 \end{bmatrix} \qquad (R3) + 5(R2) \mapsto (R3)$$

$$M_5 = \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & -6 \end{bmatrix}$$

and we are done with the second column. At this point we realize that $(M_5)_{33} = 0$. If $(M_5)_{33}$ were different from zero, say, if it were 8, then we could turn it into a 1 by multiplying by $\frac{1}{8}$, and then use that 1 to turn the other coefficients in that column into zero (as we did with the first and second columns), but $(M_5)_{33}$ being zero precludes us from doing that. We could also be tempted to interchange the first and third rows, so that we get a 1 in the right place. However, doing that would spoil our first column, which would end up looking like $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ instead of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ as desired. What to do? We move on to the next column and use $(M_5)_{34} = -6$ which is different from zero. Here we go

$$M_5 = \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & -6 \end{bmatrix} \qquad (-\tfrac{1}{6})(R3) \mapsto (R3)$$

$$M_6 = \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (R2) + (R3) \mapsto (R2)$$

$$M_7 = \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (R1) + (-3)(R3) \mapsto (R1)$$

$$M_7 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Next, we come back to systems. The matrix M_7, which is equivalent to the original matrix M, is the augmented matrix of the system

$$(S_7)\left\{\begin{array}{rcrcrcl} x_1 & + & 0x_2 & + & x_3 & = & 0\\ 0x_1 & + & x_2 & - & x_3 & = & 0\\ 0x_1 & + & 0x_2 & + & 0x_3 & = & 1\end{array}\right.$$

Interpretation: We immediately see that the system (S_7) admits no solution. Indeed, there is no way to find real numbers x_1, x_2 and x_3 that satisfy the third equation in (S_7), since this equation says that $0 = 1$. Absurd! Hence, (S_7) has no solutions. Consequently, the original system (S), being equivalent to (S_7), also has no solutions.

The lack of solutions of the system (S_7) can already be seen on the third row of M_7.

Example 2.3. A system with infinitely many solutions. Now let's study the system

$$(S)\left\{\begin{array}{rcrcrcl} x_1 & + & x_2 & + & x_3 & = & 4\\ -2x_1 & + & x_2 & + & x_3 & = & 1\\ 0x_1 & + & 3x_2 & + & 3x_3 & = & 9\end{array}\right.$$

whose augmented matrix is given by

$$M = \begin{bmatrix} 1 & 1 & 1 & 4\\ -2 & 1 & 1 & 1\\ 0 & 3 & 3 & 9\end{bmatrix}.$$

Elementary row operations yield

$$M = \begin{bmatrix} 1 & 1 & 1 & 4\\ -2 & 1 & 1 & 1\\ 0 & 3 & 3 & 9\end{bmatrix} \qquad (R2) + 2(R1) \mapsto (R2)$$

$$M_1 = \begin{bmatrix} 1 & 1 & 1 & 4\\ 0 & 3 & 3 & 9\\ 0 & 3 & 3 & 9\end{bmatrix} \qquad \tfrac{1}{3}(R2) \mapsto (R2)$$

$$M_2 = \begin{bmatrix} 1 & 1 & 1 & 4\\ 0 & 1 & 1 & 3\\ 0 & 3 & 3 & 9\end{bmatrix} \qquad (R1) + (-1)(R2) \mapsto (R1)$$

$$M_3 = \begin{bmatrix} 1 & 0 & 0 & 1\\ 0 & 1 & 1 & 3\\ 0 & 3 & 3 & 9\end{bmatrix} \qquad (R3) + (-3)(R2) \mapsto (R3)$$

$$M_4 = \begin{bmatrix} 1 & 0 & 0 & 1\\ 0 & 1 & 1 & 3\\ 0 & 0 & 0 & 0\end{bmatrix}.$$

As opposed to the case in Example 2.2, now there is no non-zero coefficient on the third row of the resulting matrix. Thus, there is nothing else we can do. The system cannot be simplified any further.

Interpretation: The matrix M_4 is the augmented matrix of the system

$$(S_4)\begin{cases} x_1 + 0x_2 + 0x_3 = 1 \\ 0x_1 + x_2 + x_3 = 3 \\ 0x_1 + 0x_2 + 0x_3 = 0 \end{cases}$$

In this case, the last equation in (S_4) yields $0 = 0$, which is no contradiction, but also carries no information. Then we can completely disregard that equation.

The first two equations in (S_4) yield that $x_1 = 1$ and that, say, $x_2 = 3 - x_3$, and there is no restriction or constraint on x_3. We say that x_3 is a *free variable*. That is, the solutions to the system (S_4), and, consequently, to the original system (S), can be written in the form

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 - x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \tag{2.2}$$

where x_3 is a free variable. This means that any choice of x_3 will produce a solution x according to the formula in (2.2). For instance, the choice $x_3 = 0$ produces the solution $x = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$, the choice $x_3 = 1$ produces $x = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, and so on. Since there are infinitely many choices for x_3 (one for each real number), the system (S) admits infinitely many solutions.

Notice that we could have done $x_1 = 1$ and $x_3 = 3 - x_2$ and no restriction on x_2. That is, we could have chosen x_2 as the free variable to conclude that the solutions to (S) can be written as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ x_2 \\ 3 - x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \tag{2.3}$$

with x_2 free. The point is that both formulas (2.3) and (2.2) completely describe the solutions to (S). We can use either one.

There is no way, however, of choosing x_1 to be the free variable, since x_1 is stuck with the value 1.

2.6. **The reduced row echelon form (rref) of a matrix. Rank of a matrix.** Look at the matrices M_8 in Example 2.1, M_7 in Example 2.2, and M_4 in Example 2.3. They all represent the simplest form of their original counterparts. Here is what we mean by "the simplest form"

We say that a matrix R is in *reduced row echelon form (rref)* if the following conditions hold true

(i) All zero rows of R lie below any non-zero row.
(ii) The first nonzero entry in any row equals 1.

(iii) The first nonzero entry in any given row (which is a 1) lies to the right of the first nonzero row (another 1) in the row immediately above it.
(iv) The first nonzero entry in any given row (which is a 1) is the only nonzero coefficient in that column.

Practice the definition of *rref* by checking that the matrices M_8 in Example 2.1, M_7 in Example 2.2, and M_4 in Example 2.3 are in reduced row echelon form.

It is a theorem that every $m \times n$ matrix M can be brought (through elementary row operations) to exactly one matrix in reduced row echelon form, we denote that matrix by $rref(M)$.

The *rank* of an $m \times n$ matrix M is defined as the number of nonzero rows in $rref(M)$, and it is denoted as $rank(M)$. For instance, the ranks of the matrices M_8 in Example 2.1, M_7 in Example 2.2, and M_4 in Example 2.3, are 3, 3, and 2, respectively.

It is insightful to see how the matrix $rref(M)$ can be created from M through elementary row operations. However, after doing that several times, we realize that that task is best left to a machine. After all, it is an algorithmic procedure easily implemented in a computer. Here are some examples on how we compute $rref(M)$ and $rank(M)$ using Matlab

Example 2.4. Compare with M in Example 2.3

```
>> B=[1 1 1 4; -2 1 1 1; 0 3 3 9]
B =

     1     1     1     4
    -2     1     1     1
     0     3     3     9

>> rref(B)

ans =

     1     0     0     1
     0     1     1     3
     0     0     0     0

>> rank(B)

ans =

     2
```

Example 2.5. Compare with M in Example 2.2.

```
>> Q=[1 1 -1 2; 2 -3 8 9; 3 -2 7 5]
```

```
Q =

     1     1    -1     2
     2    -3     8     9
     3    -2     7     5

>> rref(Q)

ans =

     1     0     1     0
     0     1    -2     0
     0     0     0     1

>> rank(Q)

ans =

     3
```

Example 2.6. `>> D=[1 3 4 5 6; 8 9 0 3 2; 2 2 2 2 2]`

```
D =

     1     3     4     5     6
     8     9     0     3     2
     2     2     2     2     2

>> rref(D)

ans =

    1.0000         0         0   -0.6316   -1.0526
         0    1.0000         0    0.8947    1.1579
         0         0    1.0000    0.7368    0.8947

>> rank(D)

ans =

     3
```

2.7. **Criterion for solvability of systems.** The results in Examples 2.1, 2.2, and 2.3 are typical in the sense that any given system (S) will have either exactly one solution, no solutions, or infinitely many solutions. This rules out the hypothetical situation of, say, a system having exactly 4 solutions. Impossible. If it has (at least) 4 solutions, then it must have infinitely many.

Theorem 2.7. *Given an $m \times n$ system (S) with matrix of coefficients A, right-hand side vector b, and augmented matrix $M = [A, b]$ (that is, (S) is represented as $Ax = b$), then one and only one of the following holds true*

(i) $rank(A) = rank(M) = n$, in which case the system (S) has exactly one solution.
(ii) $rank(A) < rank(M)$, in which case the system (S) has no solutions.
(iii) $rank(A) = rank(M) < n$, in which case the system has infinitely many solutions and the number of free variables equals $n - rank(A)$.

When a system has solutions (either just one or infinitely many) we say that *the system is consistent*; otherwise, we say that *the system is inconsistent*.

2.8. **Solving linear systems with Matlab.** Matlab will help us to obtain the $rref$ of a matrix. That is, Matlab will translate the original augmented matrix into its simplest form for us. However, the interpretation of the associated system will be still up to us. Here is how it works

Example 2.8. Let's solve the system

$$(S)\left\{\begin{array}{rcrcrcrcl} x_1 & + & 4x_2 & + & 2x_3 & + & 6x_4 & = & 1 \\ 4x_1 & + & 5x_2 & + & 0x_3 & + & x_4 & = & 1 \\ -x_1 & - & 2x_2 & + & 3x_3 & - & x_4 & = & 5 \end{array}\right.$$

This system has the matrix of coefficients

```
A =

     1     4     2     6
     4     5     0     1
    -1    -2     3    -1
```

and augmented matrix

```
M =

     1     4     2     6     1
     4     5     0     1     1
    -1    -2     3    -1     5
```

Let's compute $rank(A)$ and $rank(M)$ to see what to expect.

```
>> rank(A)

ans =

     3

>> rank(M)

ans =

     3
```

Since the system is a 3×4 system, we have $m = 3$ and $n = 4$. Then, we are in the situation $rank(A) = rank(M) = 3 < n = 4$, and, by Theorem 2.7, the system (S) is a consistent system with infinitely many solutions and $4 - 3 = 1$ free variable. Of course, now we have to find those solutions.

We begin by translating the matrix M into its simples form, that is,

```
>> rref(M)

ans =

    1.0000         0         0   -2.1538    1.3077
         0    1.0000         0    1.9231   -0.8462
         0         0    1.0000    0.2308    1.5385
```

Done. Matlab did the translation fast (much faster than us doing it by hand). Now comes the interpretation. The matrix $rref(M)$ is the augmented matrix of the system

$$(S') \begin{cases} x_1 + 0x_2 + 0x_3 - 2.1538x_4 = 1.3077 \\ 0x_1 + x_2 + 0x_3 + 1.9231x_4 = -0.8462 \\ 0x_1 + 0x_2 + x_3 + 0.2308x_4 = 1.5385 \end{cases}$$

so that $x_1 = 1.3077 + 2.1538x_4$, $x_2 = -0.8462 - 1.9231x_4$, $x_3 = 1.5385 - 0.2308x_4$, and x_4 is the free variable (we already knew there would be one free variable). The way coefficients are arranged in the system (S') makes it more convenient to choose x_4 as the free variable, as we just did.

Finally, we completely describe the solutions to the system (S) by writing that any solution x can be expressed as

$$(2.4) \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1.3077 + 2.1538x_4 \\ -0.8462 - 1.9231x_4 \\ 1.5385 - 0.2308x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1.3077 \\ -0.8462 \\ 1.5385 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2.1538 \\ -1.9231 \\ -0.2308 \\ 1 \end{bmatrix},$$

with x_4 is free. That is, each choice of x_4, when plugged into the equation (2.4), will produce a vector x whose entries will satisfy the equations in (S) (and in (S'), of course).

Example 2.9. Consider a system (S) with augmented matrix given by

```
>> M=[2 6 6 4; 3 1 1 6; -3 -2 1 7]

M =

     2     6     6     4
     3     1     1     6
    -3    -2     1     7
```

From which we deduce that (S) is a 3×3 system whose matrix of coefficients is

```
A =
```

```
     2     6     6
     3     1     1
    -3    -2     1
```

Let's compute ranks to see what's going on

```
>> rank(A)

ans =

     3

>> rank(M)

ans =

     3
```

Now, we have $m = n = 3$ and $rank(A) = rank(M) = 3$. By Theorem 2.7 it follows that (S) is a consistent system with exactly one solution. Let's find it by taking M to its simplest form and then interpreting. First the simplest form of M

```
>> rref(M)

ans =

    1.0000         0         0    2.0000
         0    1.0000         0   -4.3333
         0         0    1.0000    4.3333
```

Now we quickly interpret that the only solution to (S) is given by

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -4.3333 \\ 4.3333 \end{bmatrix} \tag{2.5}$$

Example 2.10. We are given the matrix

```
M =

     1     4     4     6
     0     2     1     2
     1     6     5    -1
```

and we are told that M is the augmented matrix of a system (S). Does (S) have solutions? Clearly, (S) is a 3×3 system whose matrix of coefficients is

```
A =

     1     4     4
     0     2     1
     1     6     5
```

Checking out the ranks

```
>> rank(A)

ans =

     2

>> rank(M)

ans =

     3
```

Then, we are dealing with $m = n = 3$, and $rank(A) = 2 < 3 = rank(M)$. By Theorem 2.7, it follows that (S) is an inconsistent system, since the rank of the matrix of coefficients is strictly smaller than the rank of the corresponding augmented matrix.

We could also conclude that such a system (S) doesn't have solutions by looking at $rref(M)$. Indeed, we have

```
>> rref(M)

ans =

    1.0000         0    2.0000         0
         0    1.0000    0.5000         0
         0         0         0    1.0000
```

and realize that the last row of $rref(M)$ leads to a contradiction $(0 = 1)$. Inconsistency!

KANSAS STATE UNIVERSITY. DEPARTMENT OF MATHEMATICS. 138 CARDWELL HALL. MANHATTAN, KS 66502

E-mail address: dmaldona@math.ksu.edu