Linear systems

A. Eremenko

August 22, 2024

A linear system is a system of equations of the form

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,n}x_n &= b_1, \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,n}x_n &= b_2, \\ \ldots &= \ldots \\ a_{m,1}x_1 + a_{m,2}x_2 + \ldots + a_{m,n}x_n &= b_m. \end{aligned}$$

Here m, n are positive integers, $a_{i,j}$ are given numbers; they are called *coefficients* of the system, b_i are given numbers; they are called right hand sides, and x_j are unknown numbers to find. So we have m equations with n unknowns. A *solution* of the system is an ordered set of numbers $(x_1, \ldots, x_n)$ which satisfy all equations of the system.

When writing the system we align it so that each entry x_j occupy certain column. If some $a_{i,j} = 0$ we can leave the empty space in this column. When the number n of unknowns is ≤ 3 we usually denote them by x, y, z instead of x_1, x_2, x_3.

When $m = n = 1$ we have

$$ax = b.$$

Three cases may occur:

1. $a \neq 0$. Then the equation has a unique solution $x = b/a$.

2. $a = 0$ but $b \neq 0$. Then the equation has no solutions.

3. $a = b = 0$. Then the equation has infinitely many solutions, namely every number x is a solution.

We will see that this pattern persists in the general case: there are only three possibilities: a unique solution, no solutions or infinitely many.

To solve an arbitrary system of linear equations we will use an algorithm the main part of which will consist of applying three kinds of operations:

Operation 1. Replace an equation by the sum of this equation and a multiple of another one.

Operation 2. Interchange two equations.

Operation 3. Multiply an equation by a non-zero number.

A new system obtained as a result of applying these operations is *equivalent* to the original one, in the sense that these two systems have the same solutions.

Example 1.

$$\begin{aligned} x + y &= 3, \\ 2x - y &= 1 \end{aligned}$$

Apply Operation 1: replace the second equation by its sum with the first one multiplied by -2. This will eliminate x from the second equation:

$$\begin{aligned} x + y &= 3, \\ -3y &= -5. \end{aligned}$$

Now apply Operation 3: multiply the second equation by $-1/3$. We obtain $y = 5/3$.

Once we know y, x is obtained from the first equation: $x = 3 - 5/3 = 4/3$. Thus we obtained the unique solution $(x, y) = (5/3, 4/9)$.

The same method works in general. Consider the generic case first.

Suppose first that $a_{1,1} \neq 0$. Then we leave the first equation unchanged, and add to each equation beginning from the second one a multiple of the first equation, to eliminate x_1 from all equations except the first one. In other words, for $i \geq 2$ replace the i-th equation by its sum with the first equation multiplied by $-a_{i,1}/a_{1,1}$. This $a_{1,1}$ is called the first *pivot*.

In the new system, suppose that $a_{2,2} \neq 0$. Then we leave the first two equations unchanged, and eliminate x_2 from the rest, by replacing each equation beginning from the third one by its sum with a multiple of the second equation. This $a_{2,2}$ is the second pivot.

Then continue this procedure until the system acquires the *row echelon form* (REF): the left-most non-zero coefficient moves to the right as we move down. These left-most non-zero coefficients in each row are called *pivots*.

When the system is brought to this form, it is easy to solve: if there are unknowns in the columns which do not contain pivots, assign arbitrary values to them. These unknowns are called *free variables*. Then determine the rest of the unknowns, starting from the last one and moving upward. When doing this, one has to divide by pivots, which is permissible because they are non-zero by definition.

It may happen that one of the equations of the row echelon form is $0x_1 + 0x_2 + \ldots + 0x_n = b_k$, where $b_k \neq 0$. Then the system has no solutions. In all other cases the system has a solution. The solution is unique if each column of x_j contains a pivot, so there are no free variables. Otherwise we have some free variables, and there are infinitely many solutions.

Example 2.

$$\begin{aligned} x + y + z &= 4 \\ x + 2y - z &= 2 \\ 2x + 3y &= 6 \end{aligned}$$

The row echelon form is

$$\begin{aligned} x + y + z &= 4 \\ y - 2z &= -2 \\ 0 &= 0 \end{aligned}$$

Here z is a free variable (columns of x and y contain pivots, each equal to 1). Assign $z = t$, where t is arbitrary. Then from the second equation $y = -2 + 2t$, and from the first equation

$$x = 4 - y - z = 4 - (-2 + 2t) - t = 6 - 3t.$$

So the general solution is

$$(x, y, z) = (6 - 3t, -2 + 2t, t), \tag{1}$$

where t is an arbitrary number.

If in the original system the right hand side of the third equation is replaced by 5, then the last equation of the REF will be $0 = -1$, and the system will have no solutions.

Matrix notation. In our calculations we only process certain rectangular arrays of numbers, the unknowns x_j are only needed to align the columns. So we can dispose of them by keeping the alignment.

A *matrix* is a rectangular array of numbers. The set of all matrices with m rows and n columns is denoted by $\mathrm{Mat}(m \times n)$. Sometimes an indication is necessary what number system is used, so, for example $\mathrm{Mat}_{\mathbf{R}}(m \times n)$ will stand for the set of all matrices with real entries. To a system of linear equations as above we associate two matrices: the *matrix of coefficients*

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \dots & \dots & \dots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}$$

and the *augmented matrix* of size $m \times (n+1)$:

$$[A, b] = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} & b_1 \\ \dots & \dots & \dots & \dots \\ a_{m,1} & \dots & a_{m,n} & b_m \end{pmatrix}.$$

To solve the system we perform row operations on the augmented matrix to reduce it to the a row echelon form. A solution exists if and only if the last column does not contain a pivot. If this is so, we assign free variables, if necessary, and solve for the rest of the variables.

Our algorithm as described uses only Operation 1, but it assumes that the matrix is generic, that is certain elements are not zero. So next we describe the algorithm in the general form.

a) Consider the left-most column which contains a non-zero element. If this non-zero element is in the first row, then this is the first pivot. If this element stands in some other row, make this row first by a row exchange (Operation 2). The first pivot is the left-most non-zero element of this new first row.

b) Add appropriate multiples of the first row to other rows to make all elements directly under the pivot zero.

c) Leave the first row unchanged, and repeat this procedure to the matrix consisting of rows 2 to m, then to the matrix consisting of rows 3 to m and so on.

This is the first step of the algorithm. The result is a matrix in the REF. In this matrix, if there is a row which has zeros at all places except the last one, and the last element is not zero, then the system has no solutions.

Otherwise, consider the free variables: x_j is free if its column contains no pivot. Assign to the free variables arbitrary values, and solve for the rest of the variables, beginning from the last non-zero equation, and moving upwards step by step. This second step of the algorithm is called the the *back substitution*

Column vectors. A column *vector* of size n is a matrix $n \times 1$:

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

By default, we call them simply "vectors". In the future we will introduce a more general notion of a vector. The set of all column vectors of size n with real entries is denoted by $\mathbf{R}^n$

Similarly, a row vector is a matrix $1 \times n$.

Vectors of the same size can be multiplied on numbers and added (entry-wise). If $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k$ are column vectors of the same size, then the expression

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \ldots + c_k\mathbf{a}_k$$

with numbers c_j is called a *linear combination* of $\mathbf{a}_1, \ldots \mathbf{a}_k$. These numbers c_j are called *coefficients* of the linear combination.

The general solution of a system of linear equations can be always written as a linear combination of some vectors where some coefficients may be arbitrary. For example, the solution (1) of the system in Example 2 can be written as a column vector

$$\mathbf{x} = \begin{pmatrix} 6 \\ -2 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 2 \\ 1 \end{pmatrix}.$$

We will always write the unknowns and solutions of linear systems as column vectors (not row vectors as we did in (1).

Geometric interpretation. Consider a system with two equations and two unknowns, like in Example 1. Pairs (x, y) can be interpreted as coordinates of a point in the plane. Then two equations in Example 1 represent two lines which can be plotted in the plane, for example by using their intercepts with the coordinate axes. Solutions are those pairs (x, y) which satisfy both equations, so they are represented by the points which belong to both lines. Two lines can intersect at one point. In Example 1, they intersect at the single point $(5/3, 4/9)$.

In general, two lines can be parallel. This corresponds to the case when the system has no solutions.

Two lines can also coincide. Then all points on this line represent infinitely many solutions. In this case we say that the set of solutions is a line.

Finally, one more case is possible: both equations can be of the form $0x + 0y = 0$. Then each point (x, y) in the plane is a solution.

So we see geometrically that a system of two equations can have a unique solution, no solutions or infinitely many. But in the last case we have two possibilities: the solutions can form a line or the whole plane.

There is a second way to view a system of linear equations geometrically. Write Example 1 in the form

$$x \begin{pmatrix} 1 \\ 2 \end{pmatrix} + y \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

This suggests that we are looking for a linear combination of two given vectors which is equal to some third given vector. We can plot these two given vectors, their multiples will have the same directions and the lengths multiplied by x and y (opposite directions if x and/or y is negative) and the sum is determined by the parallelogram rule.

So we have two geometric interpretations of a system of two equations with two unknowns:

a) finding intersection of two lines, and

b) finding a linear combination of two given vectors which is equal to the third given vector.

As in the first interpretation, we can see geometrically when the problem has a unique solution, or no solutions or infinitely many. The solution is unique if the two given vectors are not collinear. When they are collinear, the system has infinitely many solutions when the third vector is on the same line, and no solutions if it is not on the same line as the first two.

Similar interpretations can be given to systems of three equations with three unknowns. A single linear equation in three variables represents a plane in the space. A sketch of this plane can be made by plotting intercepts with the co-ordinate axes. First interpretation is finding the intersections of three planes. Several cases may occur:

a) Three planes intersect at one point. This point represents a unique solution.

b) Two of the three planes are parallel (and the third is arbitrary). There are no solutions in this case.

c) Three planes may have a common line. This line represents infinitely many solutions.

d) Two planes can coincide. The third plane can be parallel to these two (no solutions). Or the third plane can intersect the coinciding two by a line (infinitely many solutions forming a line). Or all three planes can coincide (infinitely many solutions forming a plane). Finally, all three equations can be of the form $0 = 0$, in which case the set of solutions is represented by the whole space.

The second geometric interpretation is obtained if we write the system as a linear combination of columns of the coefficient matrix. For example, the system from Example 2 can be written as

$$x \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} + y \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + z \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 6 \end{pmatrix}.$$

So the problem is to find a representation of a given vector as a linear combination of three given vectors.

With larger number of variables, the geometric intuition does not work so well, and we have to resort to algebra. Still intuition in dimensions 2 and 3 helps, and many statements become clearer when formulated in geometric terms.

Some general statements about linear systems.

Consider the REF of come matrix A. The number of non-zero rows (which is equal to the number of pivots, since every non-zero row contains exactly one pivot) is called the *rank* of the matrix; it is denoted by $r(A)$. We will later prove that the rank does not depend on how we obtain the REF, it is defined by the original matrix.

Now consider the system with coefficient matrix A and the right hand side $\mathbf{b}$, so that the augmented matrix of this system is $[A, \mathbf{b}]$. and state some general results on the existence and uniqueness of solutions.

Theorem 1. *Let A be an $m \times n$ matrix, and $\mathbf{b} \in \mathbf{R}^m$. If $r(A) < r([A, b])$ then the system has no solutions.*

If $r(A) = r([A, b])$ then the system has at least one solution.

The solution is unique if the REF of A has n pivots, in other words, there is a pivot in each column of the REF of $[A, \mathbf{b}]$ except the last one.

An important special case is $\mathbf{b} = \mathbf{0}$, where $\mathbf{0}$ is the zero-vector (a column consisting of zeros). Such systems are called *homogeneous*. A homogeneous system always has at least one solution, namely $\mathbf{x} = \mathbf{0}$. This solution is called *trivial* and all other solutions are called *non-trivial*.

Theorem 2. *If A is of size $m \times n$ with $n > m$ (the width is greater than the height), then the homogeneous system with this matrix has a non-trivial solution.*

Proof. Since $n > m$, there is at least one column in the REF corresponding to a free variable. So we have infinitely many solutions.