Theory of Linear Systems

Wang Xiao

Shanghai University of Finance and Economics

March 23, 2026

We have seen in high school that we can view the unit circle $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$ either as the set of solutions of an equation or in terms of a parametric representation

$$\{\begin{bmatrix} \cos t \\ \sin t \end{bmatrix} : t \in [0, 2\pi)\}.$$

These are the implicit and explicit representations of this subset of $\mathbb{R}^2$. Similarly, any subspace $V \subset \mathbb{R}^n$ can be represented in two ways:

i. $V = \operatorname{Span}\{\mathbf{v}_1, ..., \mathbf{v}_k\}$ for appropriate vectors $\mathbf{v}_1, ..., \mathbf{v}_k \in \mathbb{R}^n$, this is the explicit or parametric representation;

ii. $V = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = 0\}$ for an appropriate $m \times n$ matrix A, this is the implicit representation, viewing V as the intersection of the hyperplanes defined by $A_i\mathbf{x} = 0$.

As we have seen the two expression of the unit circle, the implicit and explicit descriptions extends to even nonlinear settings.

Gaussian Elimination and the Theory of Linear Systems

In this section we give an explicit algorithm for solving a system of m linear equations in n vairables:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m.
\end{aligned}
$$

With all the notations and concepts of matrix multiplication, we can write this in the form $A\mathbf{x} = \mathbf{b}$.

Some linear systems are easy to solve, but general complicated systems of equations require some algebraic manipulations before we can easily read off the general solution in parametric form. There are three basic operations we can perform on systems of equations that will not affect the solution set. They are called the *elementary operations*.

i. interchange any pair of equations;
ii. multiply any equation by a nonzero real number;
iii. replace any equation by its sum with a multiple of any other equation.

Before using elementary operations to solve a general system, we should make sure that performing elementary operations on a system of equations does not change its solutions.

Proposition

If a system of equations $A\mathbf{x} = \mathbf{b}$ *is changed into the new system* $C\mathbf{x} = \mathbf{d}$ *by elementary operations, then the systems have the same set of solutions.*

Proof.

We need to show two results: 1. every solution of $A\mathbf{x} = \mathbf{b}$ is also a solution of $C\mathbf{x} = \mathbf{d}$, and 2. vice versa.

1. Start with a solution $\mathbf{u}$ of $A\mathbf{x} = \mathbf{b}$. Denoting the rows of A by $\mathbf{A}_1, \ldots, \mathbf{A}_m$, we have

$$\begin{gathered} \mathbf{A}_1 \cdot \mathbf{u} = b_1 \\ \vdots \\ \mathbf{A}_m \cdot \mathbf{u} = b_m \end{gathered}$$

If we apply an elementary operation of type (i), $\mathbf{u}$ still satisfies precisely the same list of equations. If we apply an elementary operation of type (ii), i.e., multiplying the kth equation by $r \neq 0$. Note that if $\mathbf{u}$ satisfies $\mathbf{A}_k \cdot \mathbf{u} = b_k$, then it must satisfy

$$(r\mathbf{A}_k) \cdot \mathbf{u} = rb_k.$$

As for an elementary operation of type (iii), suppose we add r times the kth equation to the lth; since $\mathbf{A}_k \cdot \mathbf{u} = b_k$ and $\mathbf{A}_l \cdot \mathbf{u} = b_l$, it follows that

$$(r\mathbf{A}_k + \mathbf{A}_l) \cdot \mathbf{u} = (r\mathbf{A}_k \cdot \mathbf{u}) + (\mathbf{A}_l \cdot \mathbf{u}) = rb_k + b_l,$$

and so $\mathbf{u}$ satisfies the new lth equation.

2. To prove conversely that if $\mathbf{u}$ satisfies $C\mathbf{x} = \mathbf{d}$, the it satisfies $A\mathbf{x} = \mathbf{b}$, note that the inverse of elementary operation is an elementary operation.

Definition

We call the first *nonzero* entry of a row its *leading entry*. A matrix is in *echelon form* if

1. The leading entries move to the right in successive rows.
2. The entries of the column below each leading entry are all 0.
3. All rows of 0's are at the bottom of the matrix.

A matrix is in *reduced echelon form* if it is in echelon form and, in addition,

4. Every leading entry is 1.
5. All the entries of the column above each leading entry are 0 as well.

If a matrix is in echelon form, we call the leading entry of any nonzero row a *pivot*. We refer to the column in which a pivot appears as pivot columns and to the correpsonding variables as pivot variables. The remaining variables are called *free variables*.

Existence, constraint equations, and rank

Definition

If the system of equation $A\mathbf{x} = \mathbf{b}$ has no solutions, the system is said to be *inconsistent*; if it has at least one solution, then it is said to be *consistent*.

Definition

The *rank* of a matrix A is the number of nonzero rows (the number of pivots) in any echelon form of A. It is usually denoted by r.

Then the number of rows of 0's in the echelon form is $m - r$, and $\mathbf{b}$ must satisfy $m - r$ constraint equations. Note that the rank of a matrix is well-defined, i.e., independent of the choice of echelon form.

Proposition

The linear system $A\mathbf{x} = \mathbf{b}$ *is consistent if and only if the rank of the augmented matrix* $[A|\mathbf{b}]$ *equals the rank of* A*. In particular, when the rank of* A *equals* m*, the system* $A\mathbf{x} = \mathbf{b}$ *will be consistent for all vectors* $\mathbf{b} \in \mathbb{R}^m$*.*

Proof.

Let $[U|\mathbf{c}]$ denote the echelon form of the augmented matrix $[A|\mathbf{b}]$. We know that $A\mathbf{x} = \mathbf{b}$ is consistent if and only if any zero row in U corresponds to a zero entry in the vector $\mathbf{c}$, which occurs if and only if the number of nonzero rows in the augmented matrix $[U|\mathbf{c}]$ equals the number of nonzero rows in U, i.e., the rank of A. When $r = m$, there is no row of 0's in U and hence no possibility of inconsistency. □

Uniqueness and nonuniqueness of solutions

Definition

A system $A\mathbf{x} = \mathbf{b}$ is called *inhomogeneous* when $\mathbf{b} \neq \mathbf{0}$; the corresponding equation $A\mathbf{x} = \mathbf{0}$ is called the associated *homogeneous system.*

Important! To relate the solutions of the inhomogeneous system $A\mathbf{x} = \mathbf{b}$ and those of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$, we need the following fundamental algebraic observation.

Proposition

Let A be an $m \times n$ matrix and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}.$$

Proof.

Ex.

Theorem

Assume the system $A\mathbf{x} = \mathbf{b}$ *is consistent, and let* $\mathbf{u}_1$ *be a particular solution. Then all the solution are of the form*

$$\mathbf{u} = \mathbf{u}_1 + \mathbf{v}$$

for some solution $\mathbf{v}$ *of the associated homogeneous system* $A\mathbf{x} = \mathbf{0}$.

Proof.

First, if $\mathbf{u} = \mathbf{u}_1 + \mathbf{v}$, then $A\mathbf{u} = A\mathbf{u}_1 + A\mathbf{v} = \mathbf{b} + \mathbf{0} = \mathbf{b}$. Conversely, every solution of $A\mathbf{x} = \mathbf{b}$ can be written in this form, for if $\mathbf{u}$ is arbitrary solution of $A\mathbf{x} = \mathbf{b}$, then, by distributivity:

$$A(\mathbf{u} - \mathbf{u}_1) = A\mathbf{u} - A\mathbf{u}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

So $\mathbf{v} = \mathbf{u} - \mathbf{u}_1$ is a solution of the associated homogeneous system. $\square$

Proposition

Suppose the system $A\mathbf{x} = \mathbf{b}$ *is consistent. Then it has a unique solution if and only if the associated homogeneous system* $A\mathbf{x} = \mathbf{0}$ *has only the trivial solution. This happens exactly when* $r = n$*.*

Definition

An $n \times n$ matrix of rank $r = n$ is called *nonsingular*. An $n \times n$ matrix of rank $r < n$ is called *singular*.

Let A be an $n \times n$ matrix. The following are equivalent:

1. A is nonsingular.
2. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
3. For every $\mathbf{b} \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique solution.

Elementary matrices and inverses

So far we have focused on the interpretation of matrix multiplication in terms of columns, namely, the fact that the jth column of AB is the product of A with the jth column vector of B. Equally, the following is true

the jth row of AB is the product of the ith row vector of A with B.

Just as multiplying the matrix A by a column vector $\mathbf{x}$ on the right, which gives us the linear combination $x_1\mathbf{a}_1 + ... + x_n\mathbf{a}_n$ of the columns of A, we can easily check that multiplying A on the left by the row vector $[x_1 \dots x_m]$,

$$\begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix} \begin{bmatrix} --- & \mathbf{A}_1 & --- \\ --- & \mathbf{A}_2 & --- \\ & \vdots & \\ --- & \mathbf{A}_m & --- \end{bmatrix}$$

yields the linear combination $x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + ... + x_m\mathbf{A}_m$ of the rows of A.

We can perform row operations on a matrix A by multiplying on the left by appropriately chosen matrices. For example, if

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

$$E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

then

$$E_1 A = \begin{bmatrix} 3 & 4 \\ 1 & 2 \\ 5 & 6 \end{bmatrix}, \quad E_2 A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 20 & 24 \end{bmatrix} \quad E_3 A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 5 & 6 \end{bmatrix}$$

Such matrices that give corresponding elementary row operations are called *elementary matrices*.

i. To interchange rows i and j, we should multiply by an elementary matrix of the form

$$\begin{array}{c} \\ \\ i \;\rightarrow \\ \\ j \;\rightarrow \\ \\ \\ \end{array}
\begin{bmatrix}
1 & & & & & & \\
 & \ddots & & & & & \\
 & \dots & 0 & \dots & 1 & \dots & \\
 & & & \ddots & & & \\
 & \dots & 1 & \dots & 0 & \dots & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{bmatrix}$$

- To multiply row i by a scalar c, we should multiply by an elementary matrix of the form

$$\begin{array}{c c} & \\ & \\ & \\ i \rightarrow & \\ & \\ & \\ & \end{array}\left[\begin{array}{ccccccc} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{array}\right]$$

iii. To add c times row i to row j, we should multiply by an elementary matrix of the form

$$\begin{array}{rl} \begin{array}{c} \\ \\ i \rightarrow \\ \\ j \rightarrow \\ \\ \\ \end{array} & \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & \ldots & c & \ldots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \end{array}$$

Recall that if we want to find the constraint equations that a vector $\mathbf{b}$ must satisfy in order for $A\mathbf{x} = \mathbf{b}$ to be consistent, we reduce the augmented matrix $[A|\mathbf{b}]$ to echelon form $[U|\mathbf{c}]$ and set equal to 0 those entries of $\mathbf{c}$ corresponding to the rows of zeros in U. That is, when A is an $m \times n$ matrix of rank r, the constraint equations are merely the equations $c_{r+1} = ... = c_m = 0$. Letting E be the product of the elementary matrices corresponding to the elementary row operations required to put A in echelon form, we have $U = EA$ and so

$$[U|\mathbf{c}] = [EA|E\mathbf{b}].$$

That is, the constraint equation are the equations

$$\mathbf{E}_{r+1} \cdot \mathbf{b} = 0, \quad \ldots, \quad \mathbf{E}_m \cdot \mathbf{b} = 0.$$

Here we can use equation $[U|\mathbf{c}] = [EA|E\mathbf{b}]$ to find a simple way to compute E: when we reduce the augmented matrix $[A|\mathbf{b}]$ to echelon form $[U|\mathbf{c}]$, E is the matrix so that $E\mathbf{b} = \mathbf{c}$.

Recall that the inverse of the $n \times n$ matrix A is the matrix A^{-1} satisfying $AA^{-1} = A^{-1}A = I_n$. It is convenient to have an inverse matrix if we wish to solve the system $A\mathbf{x} = \mathbf{b}$ for numerous vectors $\mathbf{b}$. If A is invertible, we can solve as follows:

$$A\mathbf{x} = \mathbf{b} \Rightarrow A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b} \Rightarrow (A^{-1}A)\mathbf{x} = A^{-1}\mathbf{b} \Rightarrow \mathbf{x} = I_n\mathbf{x} = A^{-1}\mathbf{b}$$

Parametric form of solutions to a linear system

We have seen the structure of solutions to a linear system, and investigated how to determine whether a linear system has a solution. To *solve* a system, we want to give a complete *parametric description* of the solutions, i.e., we want to write the *general solution* of the system in terms of a set of parameters. For example, the system

$$\begin{array}{rcrcrcr} x_1 & & & & & = & 1 \\ & & x_2 & & & = & 2 \\ & & & & x_3 & = & -1 \end{array}$$

has exactly one solution $\mathbf{x} = (1, 2, -1)^\top$. The following example

$$\begin{array}{rcrcrcr} x_1 & & & - & x_3 & = & 1 \\ & & x_2 & + & 2x_3 & = & 2 \end{array}$$

has x_3 as *free* variables, which is used to represent pivot variables x_1 and x_2. Thus, any solution of this system is of the form

$$\mathbf{x} = (1 + t, 2 - 2t, t)^\top = (1, 2, 0) + t(1, -2, 1) \text{ for some } t \in \mathbb{R}.$$

Example

$$\begin{array}{rcrcrcrcr} 3x_1 & - & 2x_2 & + & 2x_3 & + & 9x_4 & = & 4 \\ 2x_1 & + & 2x_2 & - & 2x_3 & - & 4x_4 & = & 6 \end{array}$$

After a series of elementary operation on the augmented matrix $[A|b]$, or equivalently, applying Gaussian Elimination on the linear system, we end up with the following system corresponding to the echelon form of $[A|b]$:

$$\begin{array}{rcrcrcrcr} x_1 & & & & & + & x_4 & = & 2 \\ & & x_2 & - & x_3 & - & 3x_4 & = & 1 \end{array}$$

We can read off the general solution of the system:

$$\begin{array}{rcrcrcr} x_1 & = & 2 & & & - & x_4 \\ x_2 & = & 1 & + & x_3 & + & 3x_4 \\ x_3 & = & & & x_3 & & \\ x_4 & = & & & & & x_4 \end{array} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

which is a parametric representation of a plane in $\mathbb{R}^4$.