

Section 2: Linear Systems, Matrices and Vector Spaces

MATH241 Lecture Notes, notes pages 7–21

Notation. Scalars are plain italic (a, α, z), vectors are bold ($\mathbf{u}, \mathbf{x}$), and matrices are italic capitals (A, B ; zero matrix O , identity I). The handwritten notes underline vectors once and matrices twice.

These notes are built on top of Orhan Arikan's linear algebra course notes.

— notes page 7 —

Systems of Linear Algebraic Equations

★ (Chapter 8 in the textbook)

Many physical systems are modeled by linear equations, so being able to set up and solve them is of fundamental importance in engineering. This is where linear algebra begins.

The general form of a system of m linear algebraic equations in n unknowns is

$$\left. \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = c_m \end{array} \right\} m \text{ equations.}$$

Here the a_{ij} are the *coefficients* and the c_i are the *right-hand sides*; both can be real or complex. The x_j are the *unknowns*, which can also be real or complex.

A few words describe the possible outcomes:

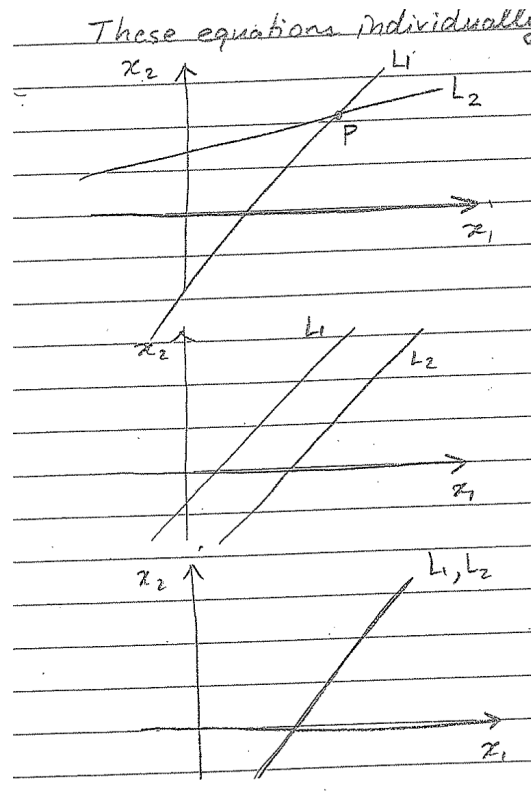
- > If replacing $x_1, \dots, x_n$ by numbers $s_1, \dots, s_n$ makes all the equations true, then $s_1, \dots, s_n$ is a *solution* of the system.
- > A system with at least one solution is *consistent*.
- > If there is exactly one solution, the solution is *unique*.
- > If there is more than one solution, the solutions are *non-unique*.
- > A system with no solution is *inconsistent*.
- > The collection of all solutions is the *solution set*.

— notes page 8 —

examples Take the case $m = n = 2$ with real coefficients:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= c_1 \\ a_{21}x_1 + a_{22}x_2 &= c_2. \end{aligned}$$

Each of these equations describes a line in the plane, and a solution of the system is a point that lies on both lines. There are three possibilities, shown in the figure: the lines cross at a single point, they are parallel and never meet, or they are the same line.



- > *Consistent with a unique solution*, since the lines cross once: $2x_1 - x_2 = 5$, $x_1 + 3x_2 = -1$.
- > *Inconsistent*, with no solution, since the lines are parallel: $x_1 + 3x_2 = 1$, $x_1 + 3x_2 = 2$.
- > *Consistent with non-unique solutions*, since the second equation is three times the first and so describes the same line: $x_1 + 3x_2 = 1$, $3x_1 + 9x_2 = 3$.

In general, each equation defines a hyperplane in n -dimensional space. The system is consistent if these hyperplanes have a common point, and the solution is unique if they meet in exactly one point.

Obtaining solutions: Gauss or Gaussian Elimination

Gaussian elimination solves a system by simplifying it step by step, using only operations that do not change the solutions, until the answer can be read off. Take $m = n = 3$ as an example. To start, add -3 times equation 1 to equation 2 and -1 times equation 1 to equation 3. This eliminates x_1 from the second and third equations:

$$\begin{cases} x_1 + x_2 - x_3 = 1 \\ 3x_1 + x_2 + x_3 = 9 \\ x_1 - x_2 + 4x_3 = 8 \end{cases} \Rightarrow \begin{cases} x_1 + x_2 - x_3 = 1 \\ -2x_2 + 4x_3 = 6 \\ -2x_2 + 5x_3 = 7. \end{cases}$$

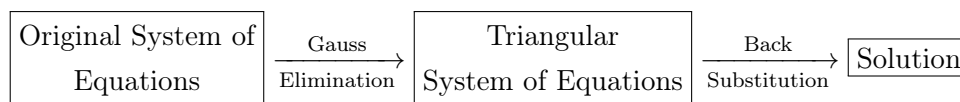
— notes page 9 —

Continuing the example, we now eliminate x_2 from the third equation by subtracting the second equation from it. This leaves a *triangular* system, in which each equation involves fewer unknowns than the one above it (left). Dividing the second equation by -2 makes its leading coefficient 1 (right):

$$\begin{cases} x_1 + x_2 - x_3 = 1 \\ -2x_2 + 4x_3 = 6 \\ x_3 = 1 \end{cases} \Rightarrow \begin{cases} x_1 + x_2 - x_3 = 1 \\ x_2 - 2x_3 = -3 \\ x_3 = 1. \end{cases}$$

The last equation already gives $x_3 = 1$. Substituting this into the second equation gives $x_2 = 2x_3 - 3 = -1$. Substituting x_2 and x_3 into the first equation gives $x_1 = -x_2 + x_3 + 1 = 3$. So the system has the unique solution $x_1 = 3, x_2 = -1, x_3 = 1$.

Important Observation: the method has two stages. First, Gauss elimination turns the original system into a triangular one. Second, back substitution solves the triangular system from the bottom up.



Question Does the triangular system have the same solutions as the original system? We need this, otherwise the method would solve the wrong problem.

Answer Yes. Gauss elimination uses only the three *elementary row operations* below, and none of them changes the solution set.

Definition The elementary row operations are:

1. Adding a multiple of one equation to another, $(\text{eq. } j) \longrightarrow (\text{eq. } j) + \alpha(\text{eq. } k)$ with $j \neq k$.
2. Multiplying an equation by a nonzero constant, $(\text{eq. } j) \longrightarrow \alpha(\text{eq. } j)$.
3. Interchanging two equations, $(\text{eq. } j) \longleftrightarrow (\text{eq. } k)$.

Thm 8.3.1: row operations do not change the solutions

If one linear system is obtained from another by a finite number of elementary row operations, then the two systems are equivalent, that is, they have the same solution set.

Proof Call the original system LS_1 and the system obtained by the row operations LS_2 , so $LS_1 \xrightarrow{E_1} \dots \xrightarrow{E_q} LS_2$, where $E_1, \dots, E_q$ are elementary row operations. We prove the claim by induction on q , the number of operations. The main work is the case $q = 1$.

— notes page 10 —

For $q = 1$ we have $LS_1 \xrightarrow{E_1} LS_2$, and E_1 is one of the three elementary operations. We write S_{LS_1} and S_{LS_2} for the two solution sets, and show for each operation that $S_{LS_1} \subset S_{LS_2}$ and $S_{LS_2} \subset S_{LS_1}$, which together mean $S_{LS_1} = S_{LS_2}$.

(i) *Adding a multiple of one equation to another,*

$$(\text{eq. } j) \longrightarrow (\text{eq. } j) + \alpha \cdot (\text{eq. } k).$$

Take a solution $(s_1, \dots, s_n) \in S_{LS_1}$. It satisfies every equation of LS_1 , in particular $(\text{eq. } j)$ and $(\text{eq. } k)$, so it also satisfies their combination $(\text{eq. } j) + \alpha(\text{eq. } k)$, which is the j^{th} equation of LS_2 . All the other equations are unchanged, so $(s_1, \dots, s_n) \in S_{LS_2}$. Hence $S_{LS_1} \subset S_{LS_2}$.

Conversely, take $(s_1, \dots, s_n) \in S_{LS_2}$. It satisfies $(\text{eq. } k)$ and the new j^{th} equation $(\text{eq. } j) + \alpha(\text{eq. } k)$ of LS_2 , so it also satisfies their difference $(\text{eq. } j) + \alpha(\text{eq. } k) - \alpha(\text{eq. } k) = (\text{eq. } j)$, the original j^{th} equation of LS_1 . So $(s_1, \dots, s_n) \in S_{LS_1}$, and $S_{LS_2} \subset S_{LS_1}$. Therefore $S_{LS_1} = S_{LS_2}$.

(ii) *Multiplying an equation by a non-zero constant,*

$$(\text{eq. } j) \longrightarrow \alpha(\text{eq. } j), \quad \alpha \neq 0.$$

Here LS_2 differs from LS_1 only in its j^{th} equation. If $(s_1, \dots, s_n) \in S_{LS_1}$, it satisfies $(\text{eq. } j)$ of LS_1 , and multiplying both sides of that equation by α keeps it true, so $(s_1, \dots, s_n) \in S_{LS_2}$. Hence $S_{LS_1} \subset S_{LS_2}$.

— notes page 11 —

Conversely, if $(s_1, \dots, s_n) \in S_{LS_2}$, it satisfies $\alpha(\text{eq. } j)$. Because $\alpha \neq 0$ we can divide by α , so it satisfies $\frac{1}{\alpha} \cdot \alpha(\text{eq. } j) = (\text{eq. } j)$ of LS_1 . Hence $S_{LS_2} \subset S_{LS_1}$, and $S_{LS_1} = S_{LS_2}$.

(iii) *Interchanging two equations*, $(\text{eq. } j) \longleftrightarrow (\text{eq. } k)$. This case is the simplest: LS_2 contains exactly the same equations as LS_1 , only in a different order. A point satisfies all of them or it does not, whatever the order, so $S_{LS_1} = S_{LS_2}$.

This completes the case $q = 1$.

Induction step. Assume the claim holds after q operations, that is, $S_{LS_1} = S_{LS_q}$, where LS_q is the system obtained after q operations. The next system $LS_{(q+1)}$ comes from LS_q by one more elementary operation, so by the case $q = 1$ we have $S_{LS_{(q+1)}} = S_{LS_q}$. Combining the two equalities,

$$S_{LS_{(q+1)}} = S_{LS_q} = S_{LS_1}.$$

This concludes the proof. ■

— notes page 12 —

examples

ex/ *An inconsistent system.* We apply the row operations step by step. After each system, the operations used to reach the next one are written on its right.

$$\begin{aligned} & \begin{cases} 2x_1 + 3x_2 - 2x_3 = 4 \\ x_1 - 2x_2 + x_3 = 3 \\ 7x_1 - x_3 = 2 \end{cases} & (\text{eq. } 1) \longrightarrow \frac{1}{2}(\text{eq. } 1) \\ \\ & \begin{cases} x_1 + \frac{3}{2}x_2 - x_3 = 2 \\ x_1 - 2x_2 + x_3 = 3 \\ 7x_1 - x_3 = 2 \end{cases} & \begin{array}{l} (\text{eq. } 2) \longrightarrow (\text{eq. } 2) - (\text{eq. } 1) \\ (\text{eq. } 3) \longrightarrow (\text{eq. } 3) - 7(\text{eq. } 1) \end{array} \\ \\ & \begin{cases} x_1 + \frac{3}{2}x_2 - x_3 = 2 \\ -\frac{7}{2}x_2 + 2x_3 = 1 \\ -\frac{21}{2}x_2 + 6x_3 = -12 \end{cases} & \begin{array}{l} (\text{eq. } 2) \longrightarrow -\frac{2}{7}(\text{eq. } 2) \\ (\text{eq. } 3) \longrightarrow (\text{eq. } 3) + \frac{21}{2}(\text{eq. } 2) \end{array} \\ \\ & \begin{aligned} & x_1 + \frac{3}{2}x_2 - x_3 = 2 \\ & x_2 - \frac{4}{7}x_3 = -\frac{2}{7} \\ & 0 = -15 \end{aligned} & \begin{array}{l} \text{cannot be satisfied by} \\ \text{any choice of } x_1, x_2 \text{ and } x_3. \end{array} \end{aligned}$$

The last equation, $0 = -15$, can never hold, so the system is *inconsistent*.

The right-hand side of the third equation decided this. If it had been 17 instead of 2, the same elimination would have ended with $0 = 0$ instead:

$$\begin{aligned} & \begin{array}{l} \text{(p)} \quad x_1 + \frac{3}{2}x_2 - x_3 = 2 \\ \text{(f)} \quad x_2 - \frac{4}{7}x_3 = -\frac{2}{7} \\ 0 = 0. \end{array} \end{aligned}$$

Now the system is consistent. The equation $0 = 0$ says nothing, so only two equations remain for three unknowns. We call x_1 and x_2 the *pivot variables* (p), because they sit at the start of an equation, and x_3 a *free variable* (f), because it can be chosen freely. Setting $x_3 = \alpha$ and solving upward gives

$$x_3 = \alpha \quad (\text{a free parameter})$$

$$\begin{aligned}x_2 &= -\frac{2}{7} + \frac{4}{7}\alpha \\x_1 &= \frac{17}{7} + \frac{1}{7}\alpha.\end{aligned}$$

Every value of α gives a solution, so the solution set is a whole line of points:

Solution set

$$S = \{(x_1, x_2, x_3) : x_1 = \frac{17}{7} + \frac{1}{7}\alpha, \quad x_2 = -\frac{2}{7} + \frac{4}{7}\alpha, \quad x_3 = \alpha, \quad \alpha \in \mathbb{R}\}.$$

— notes page 13 —

Suppose a consistent system has r pivot variables and n unknowns, with $r < n$. Then only r of the unknowns are determined by the equations, and the remaining $n - r$ are free. The solution set can be written with $n - r$ free parameters, and it is called an “ $(n - r)$ -parameter family of solutions”. Every choice of the parameters gives one “particular” solution.

Theorem 8.3.2 Existence/Uniqueness of Solutions for Linear Systems

The following theorem answers, for any linear system, whether a solution exists and whether it is unique. Consider a linear system (LS) of m equations in n unknowns:

$$\left. \begin{array}{l} a_{11}x_1 + \cdots + a_{1n}x_n = c_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = c_m \end{array} \right\} m \text{ equations with } n \text{ unknowns.}$$

Apply Gauss elimination to obtain an upper triangular (row echelon) system, and let r be the number of pivot variables. Then:

Existence and uniqueness

1. If $r = m$, the system is always consistent.
2. If $r < m$, the system is consistent exactly when the last $(m - r)$ equations have zero right-hand sides.
3. If $r < m$ and any of the last $(m - r)$ equations has a non-zero right-hand side, the system is inconsistent.
4. If the system is consistent and $r = n$, the solution is unique.
5. If the system is consistent and $r < n$, the solution set is an $(n - r)$ -parameter family of solutions.

In words: after elimination, an equation that reads $0 = (\text{non-zero number})$ makes the system impossible, and unknowns without a pivot are free.

Theorem 8.3.4 Existence and Uniqueness of Solutions for $c_1 = \cdots = c_m = 0$

A system whose right-hand sides are all zero is called *homogeneous*. It is a special case of Theorem 8.3.2 and has a simpler answer.

Homogeneous systems

1. A homogeneous system is always consistent, because $x_1 = \cdots = x_n = 0$ is a solution.
2. If $r < n$, the solution is non-unique: there are non-trivial solutions as well.

Matrix Notation

Writing out all the unknowns at every step is tedious. Only the numbers matter, so we collect them into a *coefficient matrix* and a vector of right-hand sides:

$$\left. \begin{array}{l} a_{11}x_1 + \cdots + a_{1n}x_n = c_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = c_m \end{array} \right\} \longrightarrow \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{\text{coefficient matrix}}, \quad \underbrace{\begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}}_{\text{vector of right-hand sides}}.$$

Gauss elimination is easiest to carry out on the *augmented coefficient matrix*, which puts the right-hand sides next to the coefficients:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & c_1 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & c_m \end{bmatrix}$$

The elementary row operations act on the rows of this matrix exactly as they acted on the equations.

Def'n Two matrices are called *row-equivalent* if one can be obtained from the other by finitely many elementary row operations.

Gauss-Jordan Reduction or Elimination

Gauss-Jordan reduction is another way to find the solution set. Instead of stopping at a triangular form, it keeps going until every pivot is 1 and is the only non-zero entry in its column. This is well suited to hand computation. (For large problems solved on a computer, Gauss elimination followed by back substitution needs fewer operations.)

ex/ Solve

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 4 \\ 3x_1 + 8x_2 + 7x_3 &= 20 \\ 2x_1 + 7x_2 + 9x_3 &= 23. \end{aligned}$$

We form the augmented coefficient matrix and reduce it. The first three steps below are Gauss elimination and end in the echelon form; the remaining steps clear the entries above the pivots.

$$\begin{aligned} \begin{bmatrix} 1 & 2 & 1 & 4 \\ 3 & 8 & 7 & 20 \\ 2 & 7 & 9 & 23 \end{bmatrix} &\xrightarrow{\substack{-3R_1+R_2 \rightarrow R_2 \\ -2R_1+R_3 \rightarrow R_3}} \begin{bmatrix} 1 & 2 & 1 & 4 \\ 0 & 2 & 4 & 8 \\ 0 & 3 & 7 & 15 \end{bmatrix} \xrightarrow{\frac{1}{2}R_2 \rightarrow R_2} \begin{bmatrix} 1 & 2 & 1 & 4 \\ 0 & 1 & 2 & 4 \\ 0 & 3 & 7 & 15 \end{bmatrix} \\ &\xrightarrow{-3R_2+R_3 \rightarrow R_3} \underbrace{\begin{bmatrix} 1 & 2 & 1 & 4 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix}}_{\text{Result of Gauss elimination (row echelon form)}} \xrightarrow{-2R_2+R_1 \rightarrow R_1} \begin{bmatrix} 1 & 0 & -3 & -4 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{bmatrix} \\ &\xrightarrow{\substack{3R_3+R_1 \rightarrow R_1 \\ -2R_3+R_2 \rightarrow R_2}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{bmatrix}}_{\text{Result of Jordan elimination (reduced row echelon form)}}. \end{aligned}$$

The solution can now be read directly from the last column: $x_1 = 5$, $x_2 = -2$, $x_3 = 3$.

— notes page 15 —

ex/ A system with free variables. Solve

$$\begin{aligned}x_1 + x_2 + x_3 + x_4 &= 12 \\x_1 + 2x_2 + 5x_4 &= 17 \\3x_1 + 2x_2 + 4x_3 - x_4 &= 31.\end{aligned}$$

Reducing the augmented matrix gives

$$\begin{aligned}&\begin{bmatrix} 1 & 1 & 1 & 1 & 12 \\ 1 & 2 & 0 & 5 & 17 \\ 3 & 2 & 4 & -1 & 31 \end{bmatrix} \xrightarrow[\begin{smallmatrix} -R_1+R_2 \rightarrow R_2 \\ -3R_1+R_3 \rightarrow R_3 \end{smallmatrix}]{\begin{smallmatrix} -R_1+R_2 \rightarrow R_2 \\ -3R_1+R_3 \rightarrow R_3 \end{smallmatrix}} \begin{bmatrix} 1 & 1 & 1 & 1 & 12 \\ 0 & 1 & -1 & 4 & 5 \\ 0 & -1 & 1 & -4 & -5 \end{bmatrix} \\&\xrightarrow{R_2+R_3 \rightarrow R_3} \begin{bmatrix} 1 & 1 & 1 & 1 & 12 \\ 0 & 1 & -1 & 4 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{-R_2+R_1 \rightarrow R_1} \underbrace{\begin{bmatrix} 1 & 0 & 2 & -3 & 7 \\ 0 & 1 & -1 & 4 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{\text{reduced row echelon form}}.\end{aligned}$$

The pivots are in the first two columns, so x_1 and x_2 are the pivot variables and x_3 and x_4 are the free variables. The row of zeros means the system is consistent. We give each free variable a parameter, $x_3 = \alpha$ and $x_4 = \beta$, and solve the two non-zero rows for the pivot variables:

$$\begin{aligned}x_1 &= -2\alpha + 3\beta + 7 \\x_2 &= \alpha - 4\beta + 5.\end{aligned}$$

Collecting the constants and the coefficients of each parameter shows the structure of the solution set: a fixed particular solution plus combinations of two direction vectors.

Solution set in vector form

$$S = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 7 \\ 5 \\ 0 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -2 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \end{bmatrix} : \alpha \in \mathbb{R}, \beta \in \mathbb{R} \right\}.$$

Partial Pivoting

Sometimes the entry we want as the next pivot is zero.

ex/ Suppose Gauss elimination has reached

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

The entry in the second row and second column is 0, so it cannot serve as a pivot. We swap the second and third rows and carry on as usual:

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 1 & 2 & 4 & 7 \\ 0 & 3 & 6 & 9 \\ 0 & 0 & 5 & 8 \end{bmatrix}.$$

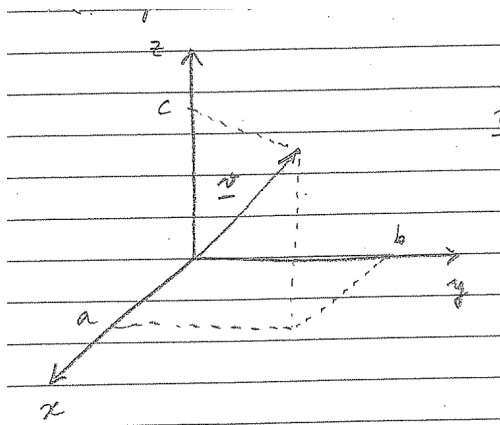
Interchanging rows like this is called *pivoting*. When several rows could be swapped in, we choose the one whose entry has the largest absolute value as the pivot; this is called *partial pivoting*.

— notes page 16 —

Vector Spaces

★ (Chapters 9 and 12 in the textbook.)

So far the unknowns have been numbers. We now describe a more general setting, in which the unknowns are *vectors*. A vector in three dimensions can be written either as an ordered triple or as a column of its components:



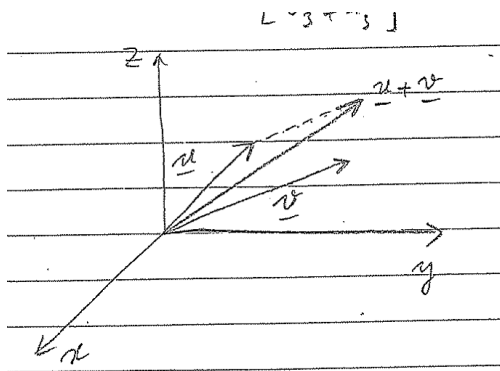
$$\mathbf{v} = (a, b, c) = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

We will use the column notation, and call a , b and c the components of $\mathbf{v}$.

Def'n *Vector addition* adds the vectors component by component:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \quad \mathbf{v} + \mathbf{u} = \begin{bmatrix} v_1 + u_1 \\ v_2 + u_2 \\ v_3 + u_3 \end{bmatrix}.$$

Geometrically, $\mathbf{u} + \mathbf{v}$ is the diagonal of the parallelogram spanned by $\mathbf{u}$ and $\mathbf{v}$.



Def'n *Scalar multiplication* multiplies every component by the same real number c :

$$\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \quad c \cdot \mathbf{v} = \begin{bmatrix} cv_1 \\ \vdots \\ cv_n \end{bmatrix}, \quad c \in \mathbb{R}.$$

Def'n *Vector space*. A set V of vectors, with vector addition and scalar multiplication defined on it, is a vector space if, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $r, s \in \mathbb{R}$, the following axioms hold (and V is closed under both operations: $\mathbf{u} + \mathbf{v} \in V$ and $r\mathbf{u} \in V$):

- (a) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (commutativity of vector addition).
- (b) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (associativity of vector addition).

— notes page 17 —

The remaining axioms are:

- (c) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$ (existence of a zero element).
- (d) $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ (existence of an additive inverse).
- (e) $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$ (scalar multiplication distributes over vector addition).
- (f) $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$ (a sum of scalars distributes over a vector).
- (g) $r(s\mathbf{u}) = (rs)\mathbf{u}$ (associativity of scalar multiplication).
- (h) $1 \cdot \mathbf{u} = \mathbf{u}$ (multiplicative identity).

These are the familiar rules of arithmetic, written for vectors. The most important example is the space of all columns of n real numbers.

Theorem: $\mathbb{R}^n$ is a vector space

$$\mathbb{R}^n = \left\{ \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} : v_i \in \mathbb{R}, 1 \leq i \leq n \right\}$$

is a vector space with the vector addition and scalar multiplication defined earlier.

Proof Each of the axioms (a) to (h) holds for $\mathbb{R}^n$ because the operations act on one component at a time, and each component is an ordinary real number, which obeys the same rules. ■

Three more facts follow from the axioms, for any $\alpha \in \mathbb{R}$ and any $\mathbf{u} \in \mathbb{R}^n$:

- (i) $\alpha \cdot \mathbf{0} = \mathbf{0}$. Indeed, $\alpha\mathbf{0} = \alpha(\mathbf{0} + \mathbf{0}) = \alpha\mathbf{0} + \alpha\mathbf{0}$, and subtracting $\alpha\mathbf{0}$ from both sides leaves $\mathbf{0} = \alpha\mathbf{0}$.
- (ii) $0 \cdot \mathbf{u} = \mathbf{0}$. The same argument works: $0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u}$.
- (iii) $-\mathbf{u} = (-1) \cdot \mathbf{u}$. Indeed, $\mathbf{u} + (-1)\mathbf{u} = (1 + (-1))\mathbf{u} = 0 \cdot \mathbf{u} = \mathbf{0}$, so $(-1)\mathbf{u}$ is the additive inverse of $\mathbf{u}$.

Def'n A *norm* measures the length of a vector. A function f from a vector space V to the real numbers is a norm if, for all $\mathbf{u}, \mathbf{v} \in V$ and $\alpha \in \mathbb{R}$:

- (i) $f(\mathbf{u}) > 0$ if $\mathbf{u} \neq \mathbf{0}$, and $f(\mathbf{u}) = 0$ if $\mathbf{u} = \mathbf{0}$ (only the zero vector has length zero).
- (ii) $f(\alpha\mathbf{u}) = |\alpha|f(\mathbf{u})$ (scaling a vector scales its length).
- (iii) $f(\mathbf{u} + \mathbf{v}) \leq f(\mathbf{u}) + f(\mathbf{v})$ (the triangle inequality).

examples In the vector space $\mathbb{R}^n$, the following functions are norms:

1. The Euclidean norm, the usual length of a vector: $f(\mathbf{u}) = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = [\sum_{i=1}^n u_i^2]^{1/2} = \|\mathbf{u}\|_2$.

2. The p -norm, for any $p \geq 1$: $f(\mathbf{u}) = [\sum_{i=1}^n |u_i|^p]^{1/p} = \|\mathbf{u}\|_p$.

— notes page 18 —

Def'n The *dot product* of two vectors in $\mathbb{R}^n$ multiplies matching components and adds the results:

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u_i v_i.$$

Note that the dot product of a vector with itself is its squared Euclidean length, $\|\mathbf{u}\|_2 = \sqrt{\mathbf{u} \cdot \mathbf{u}}$.

The properties of the dot product

- (i) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ (commutativity).
- (ii) $\mathbf{u} \cdot \mathbf{u} > 0$ for all $\mathbf{u} \neq \mathbf{0}$, and $\mathbf{u} \cdot \mathbf{u} = 0$ for $\mathbf{u} = \mathbf{0}$.
- (iii) $(\alpha\mathbf{u} + \beta\mathbf{v}) \cdot \mathbf{w} = \alpha(\mathbf{u} \cdot \mathbf{w}) + \beta(\mathbf{v} \cdot \mathbf{w})$ (linearity).

example Linearity lets us expand a dot product just like a product of brackets. To expand $(3\mathbf{u} - \mathbf{v}) \cdot (5\mathbf{w} + \mathbf{t})$:

$$\begin{aligned} (3\mathbf{u} - \mathbf{v}) \cdot (5\mathbf{w} + \mathbf{t}) &= (3\mathbf{u} - \mathbf{v}) \cdot 5\mathbf{w} + (3\mathbf{u} - \mathbf{v}) \cdot \mathbf{t} \\ &= 15(\mathbf{u} \cdot \mathbf{w}) - 5(\mathbf{v} \cdot \mathbf{w}) + 3(\mathbf{u} \cdot \mathbf{t}) - (\mathbf{v} \cdot \mathbf{t}). \end{aligned}$$

Theorem: The Cauchy–Schwarz inequality

The dot product of two vectors can never be larger in size than the product of their lengths.

Cauchy–Schwarz inequality

$$\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^n, \quad |\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2.$$

Proof If $\mathbf{v} = \mathbf{0}$, both sides are zero, so assume $\mathbf{v} \neq \mathbf{0}$. The idea is to look at the squared length of $\mathbf{u} + \alpha\mathbf{v}$ as α varies. A squared length is never negative:

$$J(\alpha) = (\mathbf{u} + \alpha\mathbf{v}) \cdot (\mathbf{u} + \alpha\mathbf{v}) \geq 0 \quad \text{for any } \alpha \in \mathbb{R}.$$

— notes page 19 —

Expanding with the properties of the dot product shows that $J(\alpha)$ is a quadratic function of α , whose graph is a parabola that never goes below zero:

$$J(\alpha) = (\mathbf{u} + \alpha\mathbf{v}) \cdot (\mathbf{u} + \alpha\mathbf{v}) = \underbrace{(\mathbf{u} \cdot \mathbf{u}) + 2\alpha(\mathbf{u} \cdot \mathbf{v}) + \alpha^2(\mathbf{v} \cdot \mathbf{v})}_{\text{a quadratic function of } \alpha}.$$

Its lowest point is where the derivative is zero:

$$\frac{d}{d\alpha} J(\alpha) = 2\mathbf{u} \cdot \mathbf{v} + 2\alpha \underbrace{(\mathbf{v} \cdot \mathbf{v})}_{\|\mathbf{v}\|_2^2} = 0 \quad \Rightarrow \quad \alpha = -\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|_2^2}.$$

This is a minimum, not a maximum, because the second derivative is $\frac{d^2}{d\alpha^2} J(\alpha) = 2\|\mathbf{v}\|_2^2 > 0$ for $\mathbf{v} \neq \mathbf{0}$. Substituting this α back into J gives the smallest value of the parabola, and that value must still be ≥ 0 :

$$\left(\mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|_2^2} \mathbf{v} \right) \cdot \left(\mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{v}\|_2^2} \mathbf{v} \right)$$

$$\begin{aligned}
 &= \|\mathbf{u}\|^2 - \frac{2(\mathbf{u} \cdot \mathbf{v})^2}{\|\mathbf{v}\|^2} + \frac{(\mathbf{u} \cdot \mathbf{v})^2}{\|\mathbf{v}\|^2} \\
 &= \|\mathbf{u}\|^2 - \frac{(\mathbf{u} \cdot \mathbf{v})^2}{\|\mathbf{v}\|^2} \geq 0. \\
 \Rightarrow \quad &\|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \geq (\mathbf{u} \cdot \mathbf{v})^2 \\
 \Rightarrow \quad &|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|. \quad \blacksquare
 \end{aligned}$$

Def'n The *angle* between two non-zero vectors in $\mathbb{R}^n$ is

Angle between vectors

$$\theta = \cos^{-1} \left[\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2} \right], \quad \text{equivalently} \quad \mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|_2 \cdot \|\mathbf{v}\|_2 \cdot \cos \theta.$$

By the Cauchy–Schwarz inequality the ratio inside $\cos^{-1}$ lies in $[-1, 1]$, so the angle is well defined.

Def'n

- (i) Two vectors are *orthogonal* (perpendicular) if $\mathbf{u} \cdot \mathbf{v} = 0$.
- (ii) A set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is an *orthogonal set* if every two different vectors in it are orthogonal: $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $1 \leq i \neq j \leq k$.

Question Given a vector $\mathbf{u} \neq \mathbf{0}$, how can we find a vector $\hat{\mathbf{u}}$ of length one that points in the same direction as $\mathbf{u}$, so that the angle between them is zero?

Answer Only vectors of the form $\alpha \mathbf{u}$ with $\alpha > 0$ make a zero angle with $\mathbf{u}$. So we look for the α for which $\|\alpha \mathbf{u}\| = 1$:

$$1 = \|\alpha \mathbf{u}\|^2 = (\alpha \mathbf{u}) \cdot (\alpha \mathbf{u}) = \alpha^2 (\mathbf{u} \cdot \mathbf{u}) = \alpha^2 \|\mathbf{u}\|^2 \quad \Rightarrow \quad \alpha = \frac{1}{\|\mathbf{u}\|}.$$

So $\hat{\mathbf{u}} = \frac{\mathbf{u}}{\|\mathbf{u}\|}$. This process is called *vector normalization*.

— notes page 20 —

Def'n A set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a set of *orthonormal* vectors if the vectors are mutually orthogonal and each has length one:

$$\mathbf{u}_i \cdot \mathbf{u}_j = 0 \quad \text{for } 1 \leq i \neq j \leq k \quad \text{and} \quad \|\mathbf{u}_i\| = 1 \quad \text{for } 1 \leq i \leq k.$$

Both conditions can be written in one line with the Kronecker delta, which is 1 when the two indices are equal and 0 otherwise:

$$\mathbf{u}_i \cdot \mathbf{u}_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

More Generalizations

The ideas above reach well beyond columns of numbers.

- > Not every vector space is finite dimensional. For example, the real-valued functions on the interval $[0, 1]$ form a vector space, with addition and scalar multiplication done pointwise:

$$(\mathbf{u} + \mathbf{v})(x) \equiv u(x) + v(x), \quad (\alpha \mathbf{u})(x) \equiv \alpha u(x).$$

- > Not every vector space comes with a dot (“inner”) product and a norm; these are extra structure.
- > If a vector space has a dot product, we can use it to define a norm, $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$, called the *induced norm*.
- > A common alternative to the ordinary dot product is the *weighted dot product*, in which each term gets a positive weight w_i :

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u_i v_i w_i, \quad \text{where } w_i > 0, 1 \leq i \leq n.$$

Extension to Complex Case

- > Vectors and matrices can have complex entries, with complex numbers as scalars instead of real ones.
- > This is very useful in electrical engineering, where linear systems with complex coefficients are common.

— notes page 21 —

With complex vectors, the dot product has to be defined with care. If it were defined as in the real case, a non-zero complex vector could have a squared length that is zero or even negative. The fix is to conjugate the entries of the second vector. For $\mathbf{u} \in \mathbb{C}^n$ and $\mathbf{v} \in \mathbb{C}^n$,

Dot product of complex vectors

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u_i v_i^*, \quad \|\mathbf{u}\|_2 = \sqrt{\sum_{i=1}^n u_i u_i^*} = \sqrt{\sum_{i=1}^n |u_i|^2}.$$

Now the squared length is a sum of the non-negative numbers $|u_i|^2$, as it should be.

example Linear systems with complex numbers are solved with the same row operations, now using complex arithmetic. Find the solution set of

$$\begin{aligned} (1+j2)x_1 + (2-j)x_2 &= 3+j8 \\ (2-j)x_1 + (1+j)x_2 &= 4+j3. \end{aligned}$$

We reduce the augmented coefficient matrix. The first step divides the first row by $1+j2$; the two divisions needed are worked out below the matrices.

$$\begin{aligned} & \begin{bmatrix} 1+j2 & 2-j & 3+j8 \\ 2-j & 1+j & 4+j3 \end{bmatrix} \xrightarrow{R_1/(1+j2) \rightarrow R_1} \begin{bmatrix} 1 & -j & \frac{19+j2}{5} \\ 2-j & 1+j & 4+j3 \end{bmatrix} \\ & \xrightarrow{R_2-(2-j)R_1 \rightarrow R_2} \begin{bmatrix} 1 & -j & \frac{19+j2}{5} \\ 0 & 2+j3 & -4+j6 \end{bmatrix} \xrightarrow{R_2/(2+j3) \rightarrow R_2} \begin{bmatrix} 1 & -j & \frac{19+j2}{5} \\ 0 & 1 & \frac{10+j24}{13} \end{bmatrix} \\ & \xrightarrow{R_1+jR_2 \rightarrow R_1} \begin{bmatrix} 1 & 0 & \frac{127+j76}{65} \\ 0 & 1 & \frac{10+j24}{13} \end{bmatrix}. \end{aligned}$$

Here we used $\frac{2-j}{1+j2} = \frac{(2-j)(1-j2)}{5} = -j$ and $\frac{3+j8}{1+j2} = \frac{(3+j8)(1-j2)}{5} = \frac{19+j2}{5}$. The solution is read from the last column:

$$x_1 = \frac{127+j76}{65}, \quad x_2 = \frac{10+j24}{13}.$$

Substituting back confirms it: $(1 + j2)x_1 + (2 - j)x_2 = 3 + j8$ and $(2 - j)x_1 + (1 + j)x_2 = 4 + j3$. As this example shows, arithmetic with complex matrices takes considerably more work than with real ones.