

MATH241 Lecture Notes

Edited edition

About this edition. Every page of the original notes has been proofread, corrected and rewritten in complete sentences. Key results sit in coloured boxes: blue boxes for formulas and methods, purple boxes for theorems. The marginal page numbers refer to the pages of the original handwritten notes.

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

Notation used in this edition

This edition uses the standard textbook convention:

- > **Scalars** (real or complex numbers) are plain italic letters: a , α , c , z .
- > **Vectors** are bold lowercase letters: $\mathbf{u}$, $\mathbf{x}$, $\mathbf{e}_1$. The zero vector is $\mathbf{0}$.
- > **Matrices** are italic capital letters, not bold: A , B , Λ . The zero matrix is O and the identity matrix is I .

Entries of a matrix or vector are scalars, so they are plain: a_{ij} is the entry of A , and u_i is the i -th component of $\mathbf{u}$.

Converting between the original notes and this edition

The handwritten notes mark the type of an object by underlining it:

	Original notes	This edition	Example
Scalar	no underline	a	α , c , λ
Complex number	underlined, $\underline{z}$	z (a scalar)	$z = a + jb$
Vector	one underline, $\underline{\mathbf{u}}$	$\mathbf{u}$ (bold)	$\underline{\mathbf{0}} \rightarrow \mathbf{0}$
Matrix	two underlines, $\underline{\underline{A}}$	A (italic capital)	$\underline{\underline{I}} \rightarrow I$, $\underline{\underline{\mathbf{0}}} \rightarrow O$
Cofactor	A_{ik}	C_{ik}	renamed, since A is now the matrix

So $\underline{\underline{A}}\underline{x} = \underline{c}$ in the original notes is written $A\mathbf{x} = \mathbf{c}$ here. Nothing else about the notation has changed: A^T , A^{-1} , $\mathbf{x}^T\mathbf{y}$, $\|\mathbf{u}\|$ and $\det A$ mean what they meant in the notes. Words that are underlined in the notes for emphasis, such as Proof or Def'n, keep their underline.

Source page 1

Introduction to Complex Numbers and Complex Algebra

★ (Chapter 21, Sections up to and including 21.3.2)

Complex numbers were introduced so that every polynomial equation has a complete set of solutions. The real numbers are not enough for this. The equation $x^2 + 1 = 0$ has no real solution at all, and the cubic below has only one real root, $x = 2$, because its quadratic factor $x^2 + x + 1$ has no real roots:

$$x^2 + 1 = 0 \quad \text{or} \quad x^3 - x^2 - x - 2 = (x - 2)(x^2 + x + 1) = 0$$

So the real numbers do not form a large enough set to contain the solutions of every polynomial equation. The complex numbers fix this.

Notation: A complex number is written $z \in \mathbb{C}$, in one of two equivalent ways:

$$\underbrace{z = a + ib}_{\text{mathematicians prefer this notation}} \quad \text{or} \quad \underbrace{z = a + jb}_{\text{electrical engineers prefer this notation}}$$

In both forms a and b are real numbers. We call a the *real part* of z and b its *imaginary part*, and we write

$$a \equiv \operatorname{Re}(a + jb), \quad b \equiv \operatorname{Im}(a + jb).$$

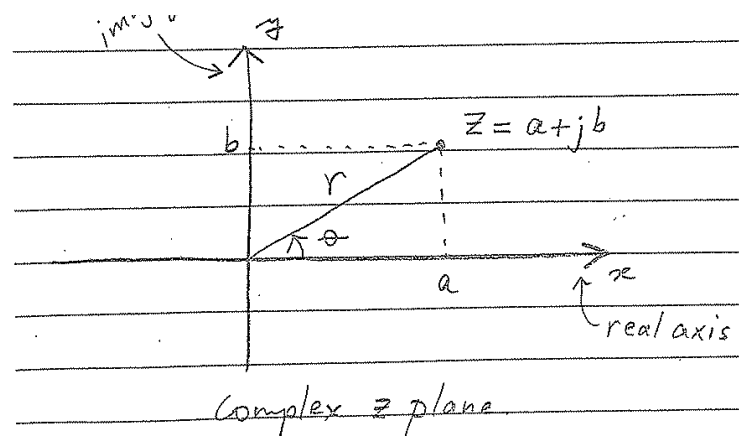
The symbol j stands for a root of $z^2 + 1 = 0$, so that $j^2 = -1$. These notes use j , as electrical engineers do.

Note that: the imaginary part of z is itself a real number, namely b and not jb . Also, every real number is a complex number with imaginary part zero, because $a \in \mathbb{R}$ gives $a + j0 \in \mathbb{C}$. In symbols, $\mathbb{R} \subset \mathbb{C}$.

Definitions:

- (i) A complex number is called *purely real* if $z = \operatorname{Re}(z)$, for example $z = 5 = 5 + j0$.
- (ii) A complex number is called *purely imaginary* if $z = j \operatorname{Im}(z)$, for example $z = j3 = 0 + j3$.
- (iii) Two complex numbers are equal, $z_1 = z_2$, if and only if their real parts are equal and their imaginary parts are equal: $\operatorname{Re}(z_1) = \operatorname{Re}(z_2)$ and $\operatorname{Im}(z_1) = \operatorname{Im}(z_2)$. For example, $z_1 = 2 + j5$ and $z_2 = 2 + j5$ are equal.

A complex number $z = a + jb$ can be drawn as the point (a, b) in the plane. The horizontal axis is the real axis and the vertical axis is the imaginary axis, so z is a point in the *complex plane*.



Source page 2

Algebraic Definitions

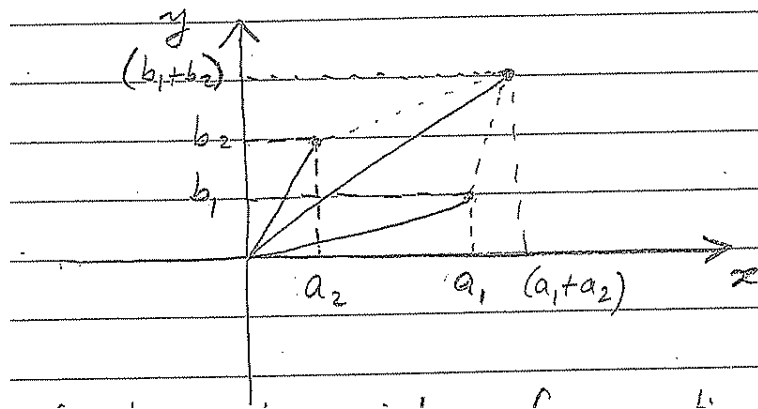
1. Summation Let $z_1 = a_1 + jb_1$ and $z_2 = a_2 + jb_2$. Their sum is the complex number $z = z_1 + z_2 = a + jb$, where the real parts add and the imaginary parts add separately:

Addition

$$(a_1 + jb_1) + (a_2 + jb_2) = (a_1 + a_2) + j(b_1 + b_2).$$

For example, with $z_1 = 2 + j3$ and $z_2 = 5 - j4$ we get $z_1 + z_2 = (2 + 5) + j(3 - 4) = 7 - j$.

In the complex plane, adding two numbers is the same as adding the two arrows from the origin: the sum is the far corner of the parallelogram built on z_1 and z_2 .



Properties of Complex Summation: For any $z_1, z_2 \in \mathbb{C}$:

Addition of complex numbers obeys the same rules as addition of real numbers.

- (i) $z_1 + z_2 = z_2 + z_1$ (commutative).
- (ii) $z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3$ (associative).
- (iii) $z_1 + (0 + j0) = z_1$, so $0 + j0$ is the additive identity.
- (iv) For every $z_1 \in \mathbb{C}$ there is a z_2 with $z_1 + z_2 = 0$.

proof Let $z_1 = a + jb$ and choose $z_2 = -a - jb$. Then

$$z_1 + z_2 = (a - a) + j(b - b) = 0 + j0 = 0.$$

- (v) This additive inverse is unique.

proof Suppose $z_1 = a_1 + jb_1$ had two different additive inverses $z_2 = a_2 + jb_2$ and $z_3 = a_3 + jb_3$, that is,

$$z_1 + z_2 = 0, \quad z_1 + z_3 = 0 \quad \text{but} \quad z_2 \neq z_3.$$

Writing both sums out in terms of real and imaginary parts gives

$$\begin{aligned} z_1 + z_2 &= (a_1 + a_2) + j(b_1 + b_2), \\ z_1 + z_3 &= (a_1 + a_3) + j(b_1 + b_3). \end{aligned}$$

Since both sums equal 0, the real parts agree and the imaginary parts agree:

$$\left\{ \begin{array}{l} a_1 + a_2 = a_1 + a_3 \Rightarrow a_2 = a_3 \\ b_1 + b_2 = b_1 + b_3 \Rightarrow b_2 = b_3 \end{array} \right\} \Rightarrow z_2 = z_3.$$

This contradicts $z_2 \neq z_3$, so the inverse is unique. ■

Source page 3

(vi) The unique additive inverse of $z = a + jb$ is written $-z = -a - jb$.

(vii) *Subtraction* is defined by adding the additive inverse:

$$z_1 - z_2 \equiv z_1 + (-z_2).$$

With $z_1 = a_1 + jb_1$ and $z_2 = a_2 + jb_2$ this gives

$$z_1 - z_2 = (a_1 + jb_1) + (-a_2 - jb_2) = (a_1 - a_2) + j(b_1 - b_2).$$

2. Multiplication Let $z_1 = a_1 + jb_1$ and $z_2 = a_2 + jb_2$. Their product $z = z_1 \cdot z_2$ is defined by

Multiplication

$$z = (a_1 + jb_1)(a_2 + jb_2) = (a_1a_2 - b_1b_2) + j(a_1b_2 + a_2b_1).$$

This is not an arbitrary rule. It is what you get by multiplying out the brackets as usual,

$$(a_1 + jb_1) \cdot (a_2 + jb_2) = a_1a_2 + ja_1b_2 + jb_1a_2 + j^2b_1b_2,$$

and then using $j^2 = -1$ to collect the real part $a_1a_2 - b_1b_2$ and the imaginary part $a_1b_2 + a_2b_1$.

example For $z_1 = 2 + j3$ and $z_2 = 5 - j4$:

$$z = z_1 \cdot z_2 = (2 \cdot 5 - 3 \cdot (-4)) + j(2 \cdot (-4) + 3 \cdot 5) = 22 + j7.$$

Properties of Complex Multiplication

Let $z_1 = a_1 + jb_1$, $z_2 = a_2 + jb_2$ and $z_3 = a_3 + jb_3$. Multiplication is commutative, associative and distributes over addition:

(i) $z_1z_2 = z_2z_1$ (commutative).

(ii) $z_1(z_2z_3) = (z_1z_2)z_3$ (associative).

(iii) $z_1(z_2 + z_3) = z_1z_2 + z_1z_3$ (distributive).

Each of these follows by writing both sides out in terms of real and imaginary parts and comparing them; the calculation is routine.

(iv) $1 \cdot z_1 = z_1$, so 1 is the multiplicative identity.

(v) Every $z_1 \neq 0$ has a multiplicative inverse, a number z with $z_1 \cdot z = 1$.

proof Choose

$$z = \frac{a_1}{a_1^2 + b_1^2} - j \frac{b_1}{a_1^2 + b_1^2}.$$

Multiplying out with the product formula gives real part $\frac{a_1^2 + b_1^2}{a_1^2 + b_1^2} = 1$ and imaginary part

$$\frac{-a_1b_1 + a_1b_1}{a_1^2 + b_1^2} = 0, \text{ so } z_1 \cdot z = 1.$$

(vi) This inverse is unique.

proof Suppose some $z_1 \neq 0$ had two different inverses z and $\hat{z}$.

Source page 4

Multiply $z_1 z = 1$ by $\hat{z}$ and regroup:

$$\begin{aligned} z_1 \cdot z &= 1 \\ \Rightarrow \hat{z}(z_1 z) &= \hat{z} \cdot 1 \\ (\hat{z} z_1) z &= \hat{z} \\ \underbrace{1 \cdot z}_z &= \hat{z} \\ z &= \hat{z}. \end{aligned}$$

This contradicts the assumption that $z \neq \hat{z}$, so the inverse is unique. ■

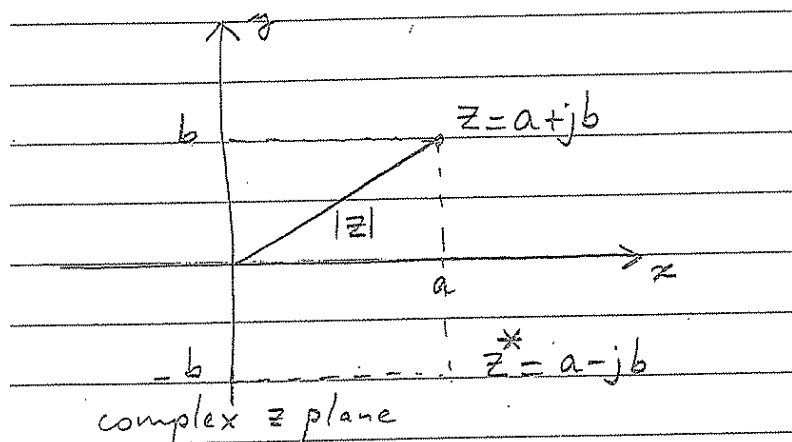
More very important Definitions

Two quantities are attached to every complex number. The *complex conjugate* of $z = a + jb$ flips the sign of the imaginary part, and the *modulus* (or magnitude) is the distance from the origin:

Conjugate and modulus

$$\bar{z} = a - jb = z^*, \quad |z| = \sqrt{a^2 + b^2} = \sqrt{z \cdot \bar{z}} = \sqrt{z \cdot z^*}.$$

Geometrically, z^* is the mirror image of z in the real axis, and $|z|$ is the length of the arrow from the origin to z .



For example, with $z = 2 + j3$ we have $z^* = 2 - j3$ and $|z| = \sqrt{2^2 + 3^2} = \sqrt{13}$.

Properties of complex conjugation:

Conjugation passes through sums and products:

- (i) $(z^*)^* = z$.
- (ii) $(z_1 + z_2)^* = z_1^* + z_2^*$.
- (iii) $(z_1 \cdot z_2)^* = z_1^* z_2^*$.

All three are checked by writing z_1 and z_2 in terms of their real and imaginary parts and comparing both sides.

Triangle Inequality

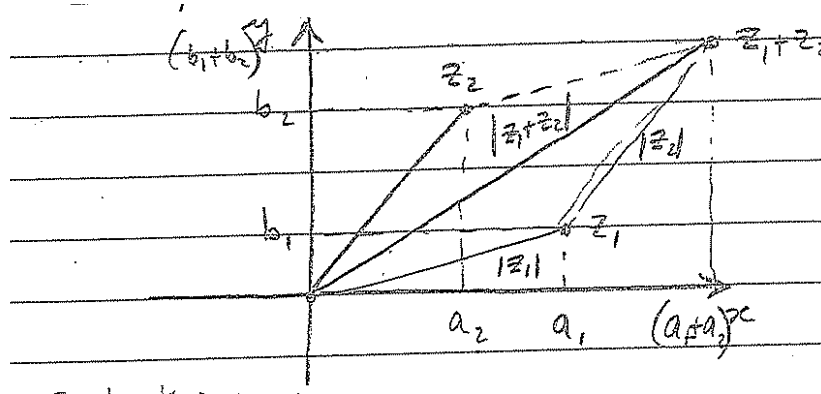
The next result says that going straight to $z_1 + z_2$ is never longer than going to z_1 first and then moving by z_2 .

Triangle inequality

$$\forall z_1 \text{ and } z_2, \quad |z_1 + z_2| \leq |z_1| + |z_2|.$$

Source page 5

Proof The picture explains why the inequality is true. The three numbers $|z_1 + z_2|$, $|z_1|$ and $|z_2|$ are the lengths of the three sides of a triangle, and no side of a triangle is longer than the other two together.



To prove it algebraically, write $z_1 = a_1 + jb_1$ and $z_2 = a_2 + jb_2$, so that $z_1 + z_2 = (a_1 + a_2) + j(b_1 + b_2)$. Squaring the modulus and expanding gives

$$\begin{aligned} |z_1 + z_2|^2 &= (a_1 + a_2)^2 + (b_1 + b_2)^2 \\ &= (a_1^2 + b_1^2) + (a_2^2 + b_2^2) + 2(a_1a_2 + b_1b_2) \\ &\leq (a_1^2 + b_1^2) + (a_2^2 + b_2^2) + 2(|a_1||a_2| + |b_1||b_2|), \end{aligned}$$

where the last step uses $xy \leq |x||y|$. The first two brackets are $|z_1|^2$ and $|z_2|^2$. To handle the last term we claim that

$$|a_1||a_2| + |b_1||b_2| \leq \sqrt{a_1^2 + b_1^2} \sqrt{a_2^2 + b_2^2} = |z_1||z_2|.$$

Both sides are non-negative, so it is enough to compare their squares:

$$(|a_1||a_2| + |b_1||b_2|)^2 \leq (a_1^2 + b_1^2)(a_2^2 + b_2^2).$$

Expanding both sides and cancelling the common terms $a_1^2a_2^2 + b_1^2b_2^2$, this is equivalent to

$$2|a_1||a_2||b_1||b_2| \leq a_1^2b_2^2 + b_1^2a_2^2 \iff 0 \leq (|a_1||b_2| - |b_1||a_2|)^2,$$

which is true because a square is never negative. Putting the pieces together,

$$\begin{aligned} |z_1 + z_2|^2 &\leq |z_1|^2 + |z_2|^2 + 2|z_1||z_2| = (|z_1| + |z_2|)^2 \\ \Rightarrow |z_1 + z_2| &\leq |z_1| + |z_2|. \quad \blacksquare \end{aligned}$$

Def'n: Division is defined through the multiplicative inverse. If $z_2 \neq 0$,

Division

$$\frac{z_1}{z_2} = z_1 \cdot z, \quad \text{where } z \text{ is the unique multiplicative inverse of } z_2.$$

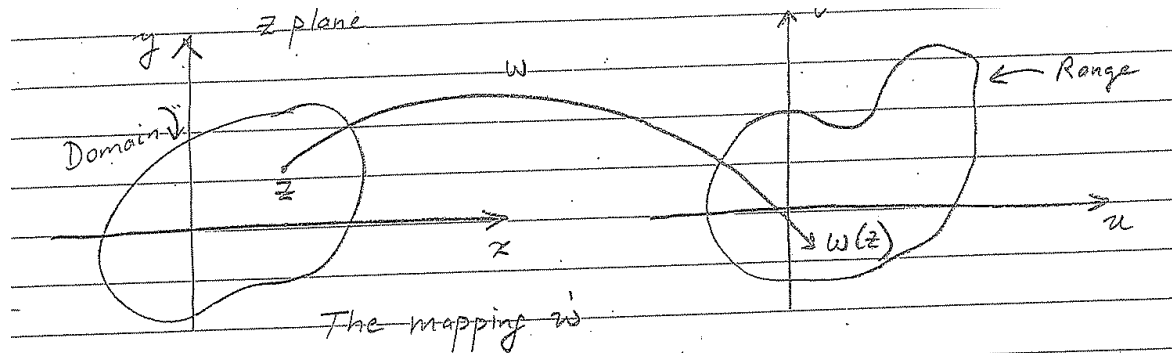
In practice we multiply the top and the bottom by the conjugate of the denominator, which makes the denominator real: ex/ Compute $\frac{2+j3}{3-j4}$.

$$\begin{aligned} \frac{2+j3}{3-j4} &= (2+j3) \cdot \frac{(3+j4)}{(3-j4)(3+j4)} = \frac{(2+j3)(3+j4)}{25} \\ &= \frac{(6-12) + j(8+9)}{25} = -\frac{6}{25} + j\frac{17}{25}. \end{aligned}$$

Source page 6

Elementary Complex Functions

A complex function w takes a complex number z in its *domain* and returns a complex number $w(z)$ in its *range*. Because both z and $w(z)$ live in planes, we picture a function as a map from one plane to another.



If $z = x + jy$, the output has a real part and an imaginary part, each depending on x and y :

$$w(z) = w(x + jy) = u(x, y) + jv(x, y).$$

ex/ *Squaring*. Let $w(z) = z^2$ with the first quadrant of the z plane as its domain, that is, $0 < x < \infty$ and $0 < y < \infty$. Then

$$w(x + jy) = (x + jy)(x + jy) = x^2 - y^2 + j2xy,$$

so $u(x, y) = x^2 - y^2$ and $v(x, y) = 2xy$. Since x and y are both positive, $v > 0$, while u can be any real number. The range is therefore $-\infty < u < \infty$ and $0 < v < \infty$, the upper half plane.

ex/ *Conjugation*. Let $w(z) = z^*$ on the entire complex plane. Then $w(x + jy) = x - jy$, so $u(x, y) = x$ and $v(x, y) = -y$. This reflects the plane in the real axis, and the range is the whole complex plane.

ex/ *The exponential function*. The complex exponential is defined by the same power series as the real one:

$$w(z) = e^z \equiv 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!} + \cdots.$$

For $z = x + jy$ it splits into a real exponential and a pure phase, $e^{x+jy} = e^x \cdot e^{jy}$ (we use this fact without proof here). To understand e^{jy} , put jy into the series. The powers of j repeat, $j^2 = -1$, $j^3 = -j$, $j^4 = 1$, so the terms separate into a real part with the even powers and an imaginary part with the odd powers:

$$\begin{aligned} e^{jy} &= 1 + jy + \frac{(jy)^2}{2!} + \cdots + \frac{(jy)^n}{n!} + \cdots \\ &= \left(1 - \frac{y^2}{2!} + \frac{y^4}{4!} - \cdots + \frac{(-1)^k y^{2k}}{(2k)!} + \cdots \right) \\ &\quad + j \left(y - \frac{y^3}{3!} + \cdots + \frac{(-1)^k y^{2k+1}}{(2k+1)!} + \cdots \right). \end{aligned}$$

The first bracket is the series of $\cos y$ and the second is the series of $\sin y$. This gives one of the most useful formulas in engineering mathematics:

Euler's formula

$$e^{jy} = \cos y + j \sin y.$$

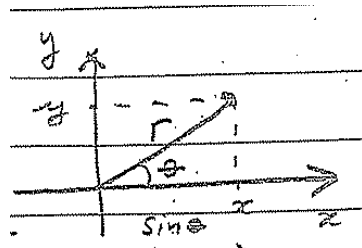
Source page 7

Putting Euler's formula back into $e^z = e^x \cdot e^{jy}$ gives the real and imaginary parts of the exponential:

$$\begin{aligned} w(z) &= e^z = e^x \cdot e^{jy} = e^x (\cos y + j \sin y) \\ &= e^x \cos y + j e^x \sin y \\ &= u(x, y) + j v(x, y), \end{aligned}$$

with $u(x, y) = e^x \cos y$ and $v(x, y) = e^x \sin y$. In particular $|e^z| = e^x > 0$, so e^z is never zero, for any complex z .

Polar representation A point in the plane can be described by its distance $r = |z|$ from the origin and its angle θ from the real axis, instead of by x and y .



Factoring out the length $r = \sqrt{x^2 + y^2}$ leaves a bracket whose two entries are $\cos \theta$ and $\sin \theta$, and by Euler's formula the bracket is $e^{j\theta}$:

$$z = x + jy = \sqrt{x^2 + y^2} \underbrace{\left(\underbrace{\frac{x}{\sqrt{x^2 + y^2}}}_{\cos \theta} + j \underbrace{\frac{y}{\sqrt{x^2 + y^2}}}_{\sin \theta} \right)}_{e^{j\theta}}$$

Polar form

$$z = r e^{j\theta}, \quad r = |z| \text{ (magnitude), } \theta \text{ (phase).}$$

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

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

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*.

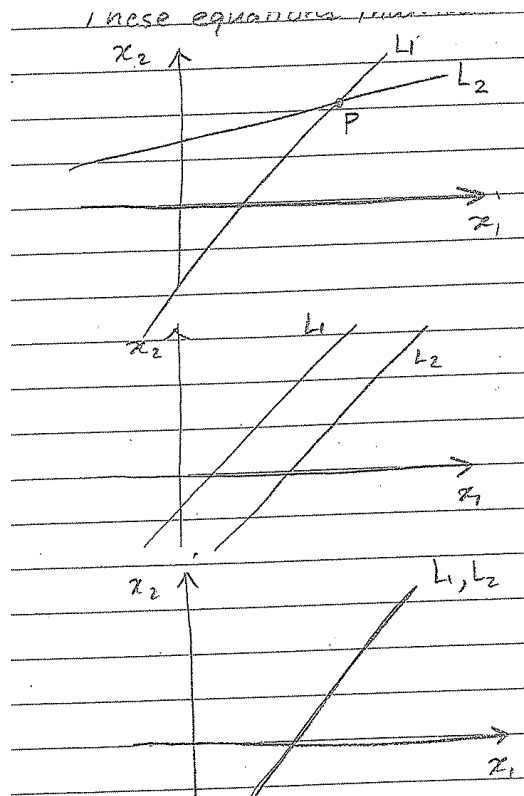
Source page 8

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

$$a_{11}x_1 + a_{12}x_2 = c_1$$

$$a_{21}x_1 + a_{22}x_2 = c_2.$$

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:

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

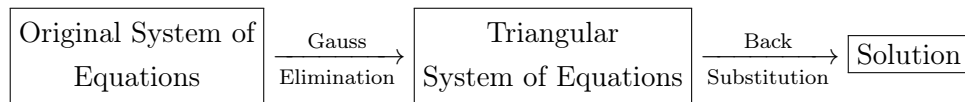
Source page 9

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):

$$\left\{ \begin{array}{l} x_1 + x_2 - x_3 = 1 \\ -2x_2 + 4x_3 = 6 \\ x_3 = 1 \end{array} \right. \quad \left\{ \begin{array}{l} x_1 + x_2 - x_3 = 1 \\ x_2 - 2x_3 = -3 \\ x_3 = 1. \end{array} \right.$$

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 $q = 1$.

Source 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 $S_{LS_1} \subset S_{LS_2}$.

Source 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. ■

Source 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.

$$\begin{cases} 2x_1 + 3x_2 - 2x_3 = 4 \\ x_1 - 2x_2 + x_3 = 3 \\ 7x_1 - x_3 = 2 \end{cases} \quad (\text{eq. 1}) \longrightarrow \frac{1}{2}(\text{eq. 1})$$

$$\begin{cases} x_1 + \frac{3}{2}x_2 - x_3 = 2 \\ x_1 - 2x_2 + x_3 = 3 \\ 7x_1 - x_3 = 2 \end{cases} \quad \begin{array}{l} (\text{eq. 2}) \longrightarrow (\text{eq. 2}) - (\text{eq. 1}) \\ (\text{eq. 3}) \longrightarrow (\text{eq. 3}) - 7(\text{eq. 1}) \end{array}$$

$$\begin{cases} x_1 + \frac{3}{2}x_2 - x_3 = 2 \\ -\frac{7}{2}x_2 + 2x_3 = 1 \\ -\frac{21}{2}x_2 + 6x_3 = -12 \end{cases} \quad \begin{array}{l} (\text{eq. 2}) \longrightarrow -\frac{2}{7}(\text{eq. 2}) \\ (\text{eq. 3}) \longrightarrow (\text{eq. 3}) + \frac{21}{2}(\text{eq. 2}) \end{array}$$

$$\begin{array}{l} x_1 + \frac{3}{2}x_2 - x_3 = 2 \\ x_2 - \frac{4}{7}x_3 = -\frac{2}{7} \\ 0 = -15 \end{array} \quad \leftarrow \begin{array}{l} \text{cannot be satisfied by} \\ \text{any choice of } x_1, x_2 \text{ and } x_3. \end{array}$$

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:

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

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

$$\begin{array}{l} x_3 = \alpha \quad (\text{a free parameter}) \\ x_2 = -\frac{2}{7} + \frac{4}{7}\alpha \\ x_1 = \frac{17}{7} + \frac{1}{7}\alpha. \end{array}$$

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

Solution set

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

Source 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.

Source page 14

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: $x_1 = 5$, $x_2 = -2$, $x_3 = 3$.

Source 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

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

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:

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

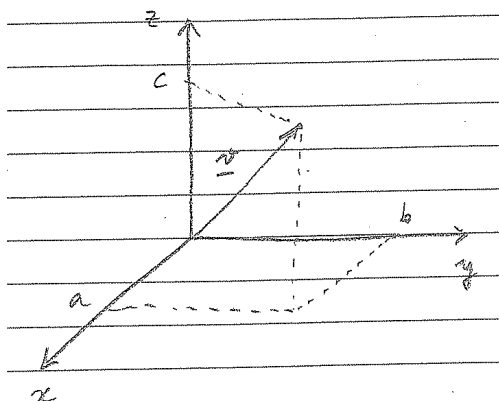
Interchanging rows like this is called *pivoting*. When several rows could be swapped in, we choose the one whose entry has the largest absolute value as the pivot; this is called *partial pivoting*.

Source 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:



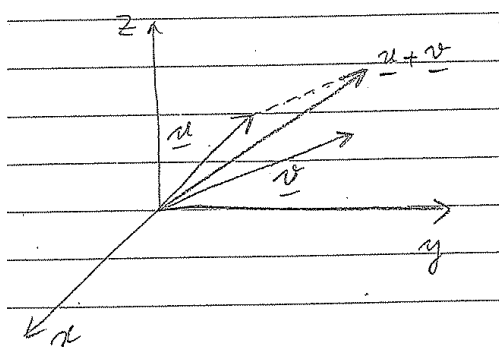
$$\mathbf{v} = (a, b, c) = \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

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).

Source 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 $p \geq 1$: $f(\mathbf{u}) = [\sum_{i=1}^n |u_i|^p]^{1/p} = \|\mathbf{u}\|_p$.

Source 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:

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

Source 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 :

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

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*.

Source 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.

Source 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.

Source page 22

Span and Subspace (Section 9.7 in the textbook)

Def'n Given vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$ in a vector space S , the set of all their linear combinations,

$$\mathbf{u} = \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k, \quad \text{for } \alpha_i \in \mathbb{R}, \quad 1 \leq i \leq k,$$

is called the *span* of $\mathbf{u}_1, \dots, \mathbf{u}_k$, and is written

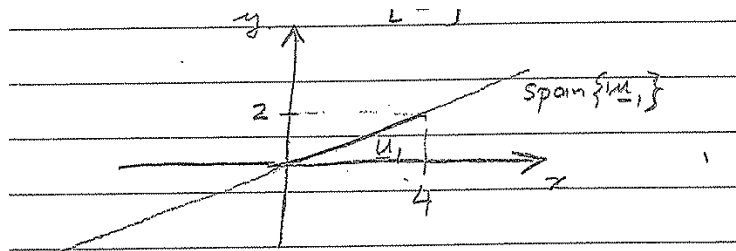
$$\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}.$$

The set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is called a *generating set* of this span.

ex's

- (i) The span of the single vector $\mathbf{u}_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \in \mathbb{R}^2$ is all of its scalar multiples. In the plane this is the straight line through the origin in the direction of $\mathbf{u}_1$:

$$\text{span}\{\mathbf{u}_1\} = \left\{ \mathbf{u} : \mathbf{u} = \alpha \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \alpha \in \mathbb{R} \right\}.$$



- (ii) Adding a second vector does not always enlarge the span. Here $\mathbf{u}_2 = -2\mathbf{u}_1$ points along the same line, so

$$\text{span} \left\{ \underbrace{\begin{bmatrix} 4 \\ 2 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} -8 \\ -4 \end{bmatrix}}_{\mathbf{u}_2} \right\} = \text{span} \left\{ \begin{bmatrix} 4 \\ 2 \end{bmatrix} \right\}.$$

Indeed, every combination of the two vectors is a multiple of $\mathbf{u}_1$:

$$\mathbf{u} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 = \alpha_1 \mathbf{u}_1 - 2\alpha_2 \mathbf{u}_1 = (\alpha_1 - 2\alpha_2) \mathbf{u}_1 = \beta \cdot \mathbf{u}_1.$$

Def'n A non-empty subset T of a vector space S is a *subspace* of S if it satisfies both of the following:

- (i) It is closed under addition: for all $\mathbf{u}$ and $\mathbf{v}$ in T , the sum $\mathbf{u} + \mathbf{v}$ is in T .
- (ii) It is closed under scalar multiplication: for all $\mathbf{u} \in T$ and $\alpha \in \mathbb{R}$, the multiple $\alpha \mathbf{u}$ is in T .

Only these two conditions need to be checked. The other axioms of a vector space are inherited from S , so a subspace is itself a vector space.

Theorem 9.7.1

Every span is a subspace.

Theorem 9.7.1

If $\mathbf{u}_1, \dots, \mathbf{u}_k$ are vectors in a vector space S , then $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a subspace of S .

Proof The span is non-empty, since it contains $\mathbf{0}$ and each $\mathbf{u}_i$. Now take two elements of it. If $\mathbf{u} \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, there are scalars $\alpha_1, \dots, \alpha_k$ with $\mathbf{u} = \alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k$.

Source page 23

Likewise, if $\mathbf{v} \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, there are scalars $\beta_1, \dots, \beta_k$ with $\mathbf{v} = \beta_1\mathbf{u}_1 + \dots + \beta_k\mathbf{u}_k$. Adding gives

$$\mathbf{u} + \mathbf{v} = (\alpha_1 + \beta_1)\mathbf{u}_1 + \dots + (\alpha_k + \beta_k)\mathbf{u}_k \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\},$$

so the span is closed under addition. For any $c \in \mathbb{R}$,

$$c\mathbf{u} = c\alpha_1\mathbf{u}_1 + \dots + c\alpha_k\mathbf{u}_k \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\},$$

so it is closed under scalar multiplication too. Hence $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a subspace of S . ■

Linear Dependence

Def'n A set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is *linearly dependent* (LD) if at least one of the vectors can be written as a linear combination of the others. If none of them can, the set is *linearly independent* (LI). Informally, a dependent set contains a vector that is redundant.

ex's

(i) The set $\left\{ \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\mathbf{u}_2}, \underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{\mathbf{u}_3} \right\}$ is linearly dependent, because $\mathbf{u}_1 = \frac{1}{2}\mathbf{u}_2 + \frac{1}{2}\mathbf{u}_3$.

(ii) The set $\left\{ \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{u}_2} \right\}$ is linearly independent, since neither vector is a multiple of the other.

(iii) The set $\left\{ \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{u}_2}, \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{u}_3} \right\}$ is linearly dependent, because $\mathbf{u}_3 = 0 \cdot \mathbf{u}_1 + 0 \cdot \mathbf{u}_2$. A set containing the zero vector is always dependent.

Theorem 9.8.1

The definition of dependence is hard to test directly, because we would have to try every vector as a combination of the others. The next theorem gives an equivalent test that is easier to use: look for non-trivial ways of combining the vectors to get zero.

Theorem 9.8.1

Let $\mathbf{u}_1, \dots, \mathbf{u}_k$ be vectors in a vector space S . The set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is LD if and only if there exist scalars α_i , not all zero, such that

$$\alpha_1\mathbf{u}_1 + \dots + \alpha_k\mathbf{u}_k = \mathbf{0}.$$

Proof The proof has two parts, one for each direction of “if and only if”.

Source page 24

proof of ($\Rightarrow$) If $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is LD, then for some j we can write

$$\mathbf{u}_j = \sum_{i \neq j} \alpha_i \mathbf{u}_i.$$

Moving $\mathbf{u}_j$ to the other side gives $\mathbf{0} = \sum_{i \neq j} \alpha_i \mathbf{u}_i - \mathbf{u}_j$. This is a combination of all the vectors with coefficients α_i , where the coefficient of $\mathbf{u}_j$ is -1 . So not all coefficients are zero, and the combination equals $\mathbf{0}$. ■

proof of ($\Leftarrow$) Suppose $\sum_{i=1}^k \alpha_i \mathbf{u}_i = \mathbf{0}$ with not all α_i zero. Pick some $\alpha_j \neq 0$. Then

$$\begin{aligned} \sum_{i \neq j} \alpha_i \mathbf{u}_i + \alpha_j \mathbf{u}_j &= \mathbf{0} \\ \Rightarrow \mathbf{u}_j &= \sum_{i \neq j} \left(-\frac{\alpha_i}{\alpha_j} \right) \mathbf{u}_i, \end{aligned}$$

which expresses $\mathbf{u}_j$ as a combination of the others, so the set is LD. ■

ex/ Is $\left\{ \underbrace{\begin{bmatrix} 2 \\ 0 \\ 1 \\ -3 \end{bmatrix}}_{\mathbf{u}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{u}_2}, \underbrace{\begin{bmatrix} 2 \\ 2 \\ 3 \\ 0 \end{bmatrix}}_{\mathbf{u}_3} \right\}$ linearly independent?

By Theorem 9.8.1, the set is LI exactly when $\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 = \mathbf{0}$ has only the trivial solution $\alpha_1 = \alpha_2 = \alpha_3 = 0$. Writing this out component by component gives a homogeneous linear system:

$$\left. \begin{aligned} 2\alpha_1 + 2\alpha_3 &= 0 \\ \alpha_2 + 2\alpha_3 &= 0 \\ \alpha_1 + \alpha_2 + 3\alpha_3 &= 0 \\ -3\alpha_1 + \alpha_2 &= 0 \end{aligned} \right\}$$

We reduce its augmented matrix. The columns are $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ and the zero right-hand side.

$$\begin{array}{cccc|c} \begin{bmatrix} 2 & 0 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 1 & 1 & 3 & 0 \\ -3 & 1 & 0 & 0 \end{bmatrix} & \xrightarrow[\text{Elimination}]{\text{Gauss}} & \begin{bmatrix} \boxed{2} & 0 & 2 & 0 \\ 0 & \boxed{1} & 2 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} & \\ \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{0} & \end{array}$$

The echelon form (reached after interchanging the last two rows) has three pivots and no free variables, so the only solution is the trivial one. Therefore $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is linearly independent.

Source page 25

Theorem 9.8.4

For a linearly independent set, a combination of the vectors can be written in only one way. Two combinations that give the same vector must use the same coefficients.

Theorem 9.8.4

If $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is LI, then

$$\alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k = \beta_1 \mathbf{u}_1 + \dots + \beta_k \mathbf{u}_k \Rightarrow \alpha_i = \beta_i, \quad 1 \leq i \leq k.$$

Proof Subtract the two sides of the equation from each other:

$$\alpha_1 \mathbf{u}_1 + \dots + \alpha_k \mathbf{u}_k = \beta_1 \mathbf{u}_1 + \dots + \beta_k \mathbf{u}_k \Rightarrow (\alpha_1 - \beta_1) \mathbf{u}_1 + \dots + (\alpha_k - \beta_k) \mathbf{u}_k = \mathbf{0}.$$

Because $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is LI, the only combination that gives $\mathbf{0}$ is the trivial one, so $\alpha_i - \beta_i = 0$, that is, $\alpha_i = \beta_i$ for $1 \leq i \leq k$. ■

Theorem 9.8.5

Orthogonal vectors point in completely different directions, so none of them can be built from the others.

Theorem 9.8.5

Every finite set of non-zero orthogonal vectors is LI.

Proof Suppose, for contradiction, that a set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ of non-zero orthogonal vectors were LD. Then for some j ,

$$\mathbf{u}_j = \sum_{i \neq j} \alpha_i \mathbf{u}_i.$$

Take the dot product of both sides with $\mathbf{u}_j$. Every term on the right vanishes, because $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ for $i \neq j$:

$$\mathbf{u}_j \cdot \mathbf{u}_j = \left(\sum_{i \neq j} \alpha_i \mathbf{u}_i \right) \cdot \mathbf{u}_j = \sum_{i \neq j} \alpha_i \underbrace{(\mathbf{u}_i \cdot \mathbf{u}_j)}_0 = 0.$$

So $\mathbf{u}_j \cdot \mathbf{u}_j = 0$, which forces $\mathbf{u}_j = \mathbf{0}$. This contradicts the assumption that all the vectors are non-zero. ■

Important Observation: a set that contains the zero vector is always linearly dependent. For $U = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ with, say, $\mathbf{u}_j = \mathbf{0}$, we can take every coefficient to be 0 except the one in front of $\mathbf{0}$, which can be any non-zero number such as 5:

$$0\mathbf{u}_1 + 0\mathbf{u}_2 + \dots + 5\mathbf{0} + \dots + 0\mathbf{u}_k = \mathbf{0}.$$

This is a combination with a non-zero coefficient that equals $\mathbf{0}$.

Bases, Expansions, Dimension

Def'n A finite set of vectors $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ in a vector space S is a *basis* for S if each vector in S can be written in exactly one way as

$$\mathbf{u} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k = \sum_{i=1}^k \alpha_i \mathbf{e}_i.$$

A basis is a set of building blocks: every vector is made from them, and made in only one way.

Theorem 9.9.1

There is an easier test for a basis than checking uniqueness directly.

Theorem 9.9.1

$\{e_1, \dots, e_k\}$ is a basis for S if $\text{span}\{e_1, \dots, e_k\} = S$ and $\{e_1, \dots, e_k\}$ is LI.

Proof Let $\mathbf{u}$ be any vector in S . Since $\text{span}\{e_1, \dots, e_k\} = S$, we can write $\mathbf{u} = \alpha_1 e_1 + \dots + \alpha_k e_k$ for some scalars α_i .

Source page 26

To conclude that $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is a basis, we still have to show that this expansion of $\mathbf{u}$ is unique. If

$$\mathbf{u} = \beta_1 \mathbf{e}_1 + \dots + \beta_k \mathbf{e}_k = \alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k$$

then $\beta_i = \alpha_i$ for every i , by Theorem 9.8.4, because $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is LI. ■

ex/ In $\mathbb{R}^2$, the vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ form a basis. To see how a vector is expanded in it, take $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and look for the coefficients:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \Rightarrow \begin{aligned} \alpha_1 + \alpha_2 &= u_1 \\ 2\alpha_1 - \alpha_2 &= u_2. \end{aligned}$$

The augmented coefficient matrix reduces as

$$\begin{bmatrix} 1 & 1 & u_1 \\ 2 & -1 & u_2 \end{bmatrix} \xrightarrow[\text{Elimination}]{\text{Gauss}} \begin{bmatrix} 1 & 1 & u_1 \\ 0 & -3 & u_2 - 2u_1 \end{bmatrix}.$$

Back substitution gives

$$\begin{aligned} \alpha_2 &= -\frac{1}{3}u_2 + \frac{2}{3}u_1 \\ \alpha_1 &= u_1 - \alpha_2 = u_1 + \frac{1}{3}u_2 - \frac{2}{3}u_1 = \frac{1}{3}u_1 + \frac{1}{3}u_2. \end{aligned}$$

So every vector of $\mathbb{R}^2$ is expanded in this basis as

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \frac{u_1 + u_2}{3} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{2u_1 - u_2}{3} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The coefficients are unique, as they must be, because $\mathbf{e}_1$ and $\mathbf{e}_2$ are LI.

Def'n The *dimension* of a vector space counts how many independent directions it has. If the greatest number of LI vectors that can be found in S is k , with $1 \leq k < \infty$, then S is k -dimensional, and we write

$$\dim S = k.$$

If $S = \{\mathbf{0}\}$, then $\dim S = 0$. If arbitrarily many LI vectors can be found in S , then S is infinite-dimensional.

Theorem 9.9.2

The dimension can be read off from any basis.

Theorem 9.9.2

If a vector space S has a basis consisting of k vectors, then $\dim S = k$.

Proof Let $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ be the basis. It is LI, because it is a basis, so $\dim S \geq k$.

Source page 27

Now suppose $\dim S > k$. Then there is an LI set $\{\mathbf{u}_1, \dots, \mathbf{u}_{k+1}\}$ of $k+1$ vectors in S . Since $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ is a basis, each $\mathbf{u}_j$ can be written uniquely in terms of it:

$$\mathbf{u}_j = \sum_{i=1}^k \alpha_{ji} \mathbf{e}_i, \quad 1 \leq j \leq k+1.$$

Now consider a combination of the $\mathbf{u}_j$ that equals zero, and rewrite it in terms of the $\mathbf{e}_i$:

$$\begin{aligned} \mathbf{0} &= \beta_1 \mathbf{u}_1 + \dots + \beta_{k+1} \mathbf{u}_{k+1} = \sum_{j=1}^{k+1} \beta_j \mathbf{u}_j \\ &= \sum_{j=1}^{k+1} \beta_j \sum_{i=1}^k \alpha_{ji} \mathbf{e}_i = \sum_{i=1}^k \mathbf{e}_i \left[\sum_{j=1}^{k+1} \alpha_{ji} \beta_j \right]. \end{aligned}$$

Since the $\mathbf{e}_i$ are LI, every bracket must be zero:

$$\Rightarrow \sum_{j=1}^{k+1} \alpha_{ji} \beta_j = 0 \quad \text{for } 1 \leq i \leq k.$$

Written out in full, these are k equations in the $k+1$ unknowns $\beta_1, \dots, \beta_{k+1}$:

$$\left. \begin{aligned} \alpha_{11}\beta_1 + \alpha_{21}\beta_2 + \dots + \alpha_{k+1,1}\beta_{k+1} &= 0 \\ \alpha_{1,2}\beta_1 + \alpha_{2,2}\beta_2 + \dots + \alpha_{k+1,2}\beta_{k+1} &= 0 \\ &\vdots \\ \alpha_{1,k}\beta_1 + \alpha_{2,k}\beta_2 + \dots + \alpha_{k+1,k}\beta_{k+1} &= 0 \end{aligned} \right\} \begin{array}{l} k \text{ equations in} \\ k+1 \text{ unknowns.} \end{array}$$

A homogeneous system with more unknowns than equations always has non-trivial solutions (Theorem 8.3.4). So there are β_j , not all zero, with $\beta_1 \mathbf{u}_1 + \dots + \beta_{k+1} \mathbf{u}_{k+1} = \mathbf{0}$. This contradicts the assumption that $\{\mathbf{u}_1, \dots, \mathbf{u}_{k+1}\}$ is LI. Hence $\dim S = k$. ■

Theorem 9.9.3

Theorem 9.9.3

The dimension of $\mathbb{R}^n$ is n .

Proof Take the standard basis vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, ..., $\mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$. They are LI. Every $\mathbf{u} \in \mathbb{R}^n$ equals $\sum_{i=1}^n \alpha_i \mathbf{e}_i$ with $\alpha_i = u_i$, so $\text{span}\{\mathbf{e}_1, \dots, \mathbf{e}_n\} = \mathbb{R}^n$. Hence $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a basis for $\mathbb{R}^n$, and $\dim \mathbb{R}^n = n$. ■

Source page 28

Theorem 9.9.4

The dimension of a span equals the number of vectors in the generating set that are really needed.

Theorem 9.9.4

For vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$, not all zero, $\dim[\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}]$ equals the greatest number of linearly independent vectors within the generating set $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$.

Proof Let m be the greatest number of linearly independent vectors in $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and number the vectors so that the first m are LI. Each of the remaining vectors $\mathbf{u}_{m+1}, \dots, \mathbf{u}_k$ must then depend linearly on $\mathbf{u}_1, \dots, \mathbf{u}_m$:

$$\mathbf{u}_i = \alpha_{i1}\mathbf{u}_1 + \dots + \alpha_{im}\mathbf{u}_m, \quad m+1 \leq i \leq k.$$

Now take any $\mathbf{v} \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, say $\mathbf{v} = \sum_{i=1}^k \beta_i \mathbf{u}_i$. Replacing each of the last $k-m$ vectors by its expression gives

$$\begin{aligned} \Rightarrow \quad \mathbf{v} &= \sum_{i=1}^m \beta_i \mathbf{u}_i + \sum_{j=m+1}^k \beta_j \mathbf{u}_j \\ \mathbf{v} &= \sum_{i=1}^m \beta_i \mathbf{u}_i + \sum_{j=m+1}^k \beta_j \left[\sum_{i=1}^m \alpha_{ji} \mathbf{u}_i \right] \\ \mathbf{v} &= \sum_{i=1}^m \left[\beta_i + \sum_{j=m+1}^k \beta_j \alpha_{ji} \right] \mathbf{u}_i \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_m\}. \end{aligned}$$

So $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\} = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$. Since $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ is LI and spans this space, it is a basis for $\text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and the dimension is m . ■

Orthogonal Bases

A vector space of dimension two or more has infinitely many bases. In practice we prefer *orthogonal* bases, because the coefficients of a vector are then easy to find.

Let $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ be an orthogonal basis for S . Any $\mathbf{u} \in S$ can be written as $\mathbf{u} = \alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k$. To find α_j , take the dot product with $\mathbf{e}_j$. By orthogonality, every term except the j -th vanishes:

$$\mathbf{u} \cdot \mathbf{e}_j = (\alpha_1 \mathbf{e}_1 + \dots + \alpha_k \mathbf{e}_k) \cdot \mathbf{e}_j = \alpha_j \mathbf{e}_j \cdot \mathbf{e}_j.$$

Solving for α_j gives the coefficient formula, where the denominator $\mathbf{e}_i \cdot \mathbf{e}_i = \|\mathbf{e}_i\|^2$ is the normalization factor:

Expansion in an orthogonal basis

$$\alpha_j = \frac{\mathbf{u} \cdot \mathbf{e}_j}{\mathbf{e}_j \cdot \mathbf{e}_j} \quad \Rightarrow \quad \mathbf{u} = \sum_{i=1}^k \left(\frac{\mathbf{u} \cdot \mathbf{e}_i}{\mathbf{e}_i \cdot \mathbf{e}_i} \right) \mathbf{e}_i.$$

Source page 29

Dividing each basis vector by its length, $\hat{e}_i = \mathbf{e}_i / \|\mathbf{e}_i\|$, turns the formula above into $\mathbf{u} = \sum_{i=1}^k (\mathbf{u} \cdot \hat{e}_i) \hat{e}_i$. If the basis is *orthonormal*, that is, orthogonal with every vector of length one, then the denominators are all 1 and the coefficients are just dot products:

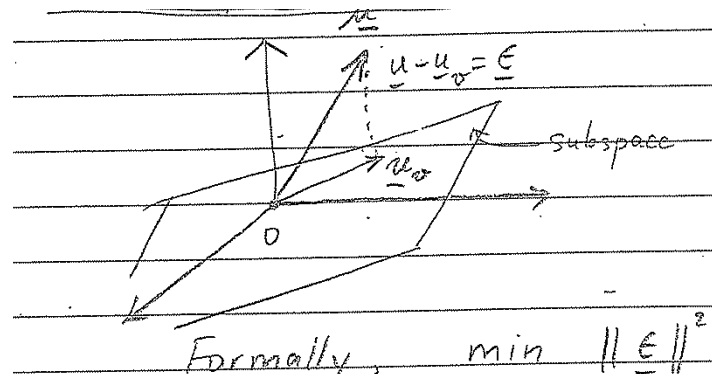
$$\alpha_i = \frac{\mathbf{u} \cdot \mathbf{e}_i}{\underbrace{\mathbf{e}_i \cdot \mathbf{e}_i}_1} = \mathbf{u} \cdot \mathbf{e}_i.$$

Best Approximation

In many engineering applications we are given a vector $\mathbf{u}$ and want its best approximation inside a subspace. Let S be a vector space and V a subspace of S , and let $\{\hat{e}_1, \dots, \hat{e}_N\}$ be an orthonormal basis for V :

$$\hat{e}_i \cdot \hat{e}_j = \delta_{ij}, \quad 1 \leq i, j \leq N, \quad \text{span}\{\hat{e}_1, \dots, \hat{e}_N\} = V.$$

The picture shows the idea. The vector $\mathbf{u}_V$ in V should be as close to $\mathbf{u}$ as possible, so the error $\boldsymbol{\epsilon} = \mathbf{u} - \mathbf{u}_V$ should be as short as possible.



We want to minimize the squared length of the error by choosing $\mathbf{u}_V \in V$:

$$\min_{\mathbf{u}_V} \|\boldsymbol{\epsilon}\|_2^2 = \min_{\mathbf{u}_V} \|\mathbf{u} - \mathbf{u}_V\|_2^2.$$

Every $\mathbf{u}_V \in V$ can be written in the orthonormal basis as $\mathbf{u}_V = \sum_{i=1}^N \alpha_i \hat{e}_i$, so choosing $\mathbf{u}_V$ means choosing the numbers α_i . The problem becomes one of minimizing a cost function J of these numbers:

$$\min_{\alpha_1, \dots, \alpha_N} \left\| \mathbf{u} - \underbrace{\sum_{i=1}^N \alpha_i \hat{e}_i}_{J(\alpha_1, \dots, \alpha_N)} \right\|_2^2.$$

Expanding the squared length as a dot product, and using the fact that the basis is orthonormal, gives

$$\begin{aligned} J(\alpha_1, \dots, \alpha_N) &= \left(\mathbf{u} - \sum_{i=1}^N \alpha_i \hat{e}_i \right) \cdot \left(\mathbf{u} - \sum_{i=1}^N \alpha_i \hat{e}_i \right) \\ &= \mathbf{u} \cdot \mathbf{u} - 2 \sum_{i=1}^N \alpha_i \mathbf{u} \cdot \hat{e}_i + \underbrace{\left(\sum_{i=1}^N \alpha_i \hat{e}_i \right) \cdot \left(\sum_{i=1}^N \alpha_i \hat{e}_i \right)}_{\sum_{i=1}^N \alpha_i^2}. \end{aligned}$$

Source page 30

In short, J is a quadratic function of the α_i :

$$J(\alpha_1, \dots, \alpha_N) = \underbrace{\mathbf{u} \cdot \mathbf{u} - 2 \sum_{i=1}^N \alpha_i \mathbf{u} \cdot \hat{\mathbf{e}}_i + \sum_{i=1}^N \alpha_i^2}_{\text{quadratic function of the } \alpha_i}.$$

Its minimum is where all the partial derivatives are zero:

$$\frac{\partial J}{\partial \alpha_i} = -2\mathbf{u} \cdot \hat{\mathbf{e}}_i + 2\alpha_i = 0 \Rightarrow \alpha_i = \mathbf{u} \cdot \hat{\mathbf{e}}_i.$$

So the best approximation of $\mathbf{u}$ in the subspace V is

$$\mathbf{u}_V = \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i.$$

If $\mathbf{u}$ already lies in V , its best approximation is $\mathbf{u}$ itself.

Now let us look at the error of this best approximation:

$$\boldsymbol{\epsilon} = \mathbf{u} - \mathbf{u}_V = \mathbf{u} - \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i.$$

Claim The error $\boldsymbol{\epsilon}$ is orthogonal to V , that is, orthogonal to every vector in V .

Proof Any vector in V can be written $\mathbf{v} = \sum_{j=1}^N \beta_j \hat{\mathbf{e}}_j$, because $\text{span}\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_N\} = V$. Then

$$\begin{aligned} \boldsymbol{\epsilon} \cdot \mathbf{v} &= \left(\mathbf{u} - \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i \right) \cdot \left(\sum_{j=1}^N \beta_j \hat{\mathbf{e}}_j \right) \\ &= \sum_{j=1}^N \beta_j \mathbf{u} \cdot \hat{\mathbf{e}}_j - \sum_{i=1}^N \sum_{j=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \beta_j \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j \\ &= \sum_{j=1}^N \beta_j \mathbf{u} \cdot \hat{\mathbf{e}}_j - \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \beta_i = 0, \end{aligned}$$

where the last step uses $\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij}$. ■

Because the error is perpendicular to V , we call $\mathbf{u}_V$ the *orthogonal projection* of $\mathbf{u}$ onto V :

Orthogonal projection onto V

$$\text{proj}_V \mathbf{u} = \sum_{i=1}^N (\mathbf{u} \cdot \hat{\mathbf{e}}_i) \hat{\mathbf{e}}_i.$$

This formula needs an orthonormal basis of V . What if we are only given a basis that is not orthonormal?

Source page 31

Answer We can use the Gram–Schmidt orthogonalization process.

The Gram-Schmidt Orthogonalization Process

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_N\}$ be any basis for a subspace V . We want to build from it an orthonormal basis $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_N\}$ for the same subspace. The process works one vector at a time, and each new vector is made orthogonal to all the previous ones.

Step 1 Take the first vector and normalize it: $\hat{\mathbf{e}}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$. Then $\hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_1 = 1$.

Step 2 (i) Project $\mathbf{v}_2$ onto $\text{span}\{\hat{\mathbf{e}}_1\}$ to get $\mathbf{v}_{2p} = (\mathbf{v}_2 \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1$. (ii) Compute the projection error $\boldsymbol{\epsilon}_2 = \mathbf{v}_2 - \mathbf{v}_{2p}$; it is orthogonal to $\hat{\mathbf{e}}_1$. Normalize it:

$$\hat{\mathbf{e}}_2 = \frac{\boldsymbol{\epsilon}_2}{\|\boldsymbol{\epsilon}_2\|}, \quad \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_1 = 0, \quad \|\hat{\mathbf{e}}_2\| = 1.$$

Step 3 Repeat for the remaining vectors, $2 < i \leq N$. Project $\mathbf{v}_i$ onto the span of the vectors found so far, $\{\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_{i-1}\}$:

$$\mathbf{v}_{ip} = \sum_{j=1}^{i-1} (\mathbf{v}_i \cdot \hat{\mathbf{e}}_j) \hat{\mathbf{e}}_j.$$

The projection error is $\boldsymbol{\epsilon}_i = \mathbf{v}_i - \mathbf{v}_{ip}$. Then set

$$\hat{\mathbf{e}}_i = \frac{\boldsymbol{\epsilon}_i}{\|\boldsymbol{\epsilon}_i\|}.$$

The new vector $\hat{\mathbf{e}}_i$ is orthogonal to $\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_{i-1}$, because $\boldsymbol{\epsilon}_i$ is the error of projecting onto their span. It has length one by construction. The division by $\|\boldsymbol{\epsilon}_i\|$ is allowed because $\boldsymbol{\epsilon}_i \neq \mathbf{0}$: the $\mathbf{v}_i$ are LI, so $\mathbf{v}_i$ does not lie in the span of the previous ones. So Gram–Schmidt turns any basis into an orthonormal basis.

Gram–Schmidt formulas

$$\hat{\mathbf{e}}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \quad \hat{\mathbf{e}}_2 = \frac{\mathbf{v}_2 - (\mathbf{v}_2 \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1}{\|\mathbf{v}_2 - (\mathbf{v}_2 \cdot \hat{\mathbf{e}}_1)\hat{\mathbf{e}}_1\|}, \quad \dots, \quad \hat{\mathbf{e}}_i = \frac{\mathbf{v}_i - \sum_{j=1}^{i-1} (\mathbf{v}_i \cdot \hat{\mathbf{e}}_j)\hat{\mathbf{e}}_j}{\left\| \mathbf{v}_i - \sum_{j=1}^{i-1} (\mathbf{v}_i \cdot \hat{\mathbf{e}}_j)\hat{\mathbf{e}}_j \right\|}.$$

Source page 32

Matrices and Linear Equations

(★ Chapter 10 in the textbook)

A *matrix* is a rectangular array of numbers. It has m rows and n columns, and its entries a_{ij} can be real or complex:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \quad n \text{ columns, } m \text{ rows, } a_{ij} \in \mathbb{R} \text{ or } a_{ij} \in \mathbb{C}.$$

ex/ $A = \begin{bmatrix} 2 & 1 \\ 0 & 4 \\ -5 & 7 \end{bmatrix}$ is a real 3×2 matrix, and $B = \begin{bmatrix} 1+j & 2 \\ 3-j & 4+2j \end{bmatrix}$ is a complex 2×2 matrix.

Def'n Two matrices $A = \{a_{ij}\}_{m \times n}$ and $B = \{b_{ij}\}_{k \times l}$ are *equal* if they have the same size, $m = k$ and $n = l$, and all entries agree: $a_{ij} = b_{ij}$ for $1 \leq i \leq m$ and $1 \leq j \leq n$.

Def'n Matrices of the same size are added entry by entry. For $A = \{a_{ij}\}_{m \times n}$ and $B = \{b_{ij}\}_{m \times n}$,

$$C = A + B = \{a_{ij} + b_{ij}\}_{m \times n}.$$

ex/ Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & -5 \\ -1 & 3 & 7 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Then

$$A + B = \begin{bmatrix} 2 & 4 & -2 \\ 3 & 8 & 13 \end{bmatrix},$$

but $A + C$ and $B + C$ are not defined, because C has a different size.

Def'n Multiplying a matrix by a scalar c multiplies every entry by c : for $A = \{a_{ij}\}_{m \times n}$, $cA = \{ca_{ij}\}_{m \times n}$, where $c \in \mathbb{R}$ (or $c \in \mathbb{C}$ for complex matrices).

ex/ For $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ we get $2A = \begin{bmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{bmatrix}$ and $(-1)A = -A = \begin{bmatrix} -1 & -2 & -3 \\ -4 & -5 & -6 \end{bmatrix}$.

Theorem 10.2.1 Properties of matrix addition and scalar multiplication

Addition and scalar multiplication of matrices obey the same rules as for numbers. If A , B and C are $m \times n$ matrices, O is the $m \times n$ zero matrix, and α, β are scalars, then:

- (i) $A + B = B + A$ (commutativity).
- (ii) $(A + B) + C = A + (B + C)$ (associativity).
- (iii) $A + O = A$ (identity element of matrix addition).

Source page 33

The remaining properties are:

- (iv) $A + (-A) = O$ (inverse element of matrix addition).
- (v) $\alpha(\beta A) = (\alpha\beta)A$ (associativity of scalar multiplication).
- (vi) $(\alpha + \beta)A = \alpha A + \beta A$ (distributivity).
- (vii) $\alpha(A + B) = \alpha A + \alpha B$ (distributivity).
- (viii) $1 \cdot A = A$ (identity element of scalar multiplication).
- (ix) $0A = O$.
- (x) $\alpha O = O$.

Def'n Matrix multiplication Matrices can be multiplied when the number of columns of the first equals the number of rows of the second. If $A = \{a_{ij}\}_{m \times n}$ and $B = \{b_{ij}\}_{n \times p}$, their product is the $m \times p$ matrix $C = AB = \{c_{ij}\}_{m \times p}$ whose entry c_{ij} is the dot product of row i of A with column j of B :

Matrix multiplication

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}, \quad 1 \leq i \leq m, \quad 1 \leq j \leq p.$$

ex/ Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & 4 \\ -1 & 0 & 5 \\ 1 & 3 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 2 & 11 & 17 \\ 5 & 26 & 47 \end{bmatrix}, \quad CA = \begin{bmatrix} 13 & 17 & 21 \\ 18 & 24 & 30 \end{bmatrix}$$

but AC , BC , BA and CB are not defined, because the sizes do not match.

Note that even when both AB and BA exist, AB is not necessarily equal to BA . The order matters. If $AB = BA$, we say that A and B *commute* under multiplication.

The definition of matrix multiplication comes from linear systems. A system of m equations in n unknowns,

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

is exactly the statement that a matrix times a column of unknowns equals a column of right-hand sides:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}}_{A \ m \times n} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\mathbf{x} \ n \times 1} = \underbrace{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}}_{\mathbf{c} \ m \times 1}.$$

A linear system in one line

$$A\mathbf{x} = \mathbf{c}.$$

Source page 34

Def'n For a square matrix $A = \{a_{ij}\}_{n \times n}$, the *power* A^p is the product of p copies of A :

$$A^p = \underbrace{A \cdot A \cdots A}_{p \text{ factors}}.$$

Def'n A square matrix $A = \{a_{ij}\}_{n \times n}$ is *diagonal* if $a_{ij} = 0$ for $i \neq j$, that is, if all its non-zero entries lie on the main diagonal.

ex/ $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \end{bmatrix}$ is not diagonal, because it is not square. $B = \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix}$ is a diagonal matrix.

Def'n The *identity matrix* is the diagonal matrix with $a_{ii} = 1$ on the diagonal.

ex/ $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the 2×2 identity matrix, and $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is the 3×3 identity matrix.

The identity matrix leaves any matrix unchanged: $A_{m \times n} I_n = A_{m \times n}$ and $I_m A_{m \times n} = A_{m \times n}$. This holds because in each entry of the product, only one term of the sum is non-zero, and it is multiplied by 1.

Theorem 10.2.2 Exceptional Properties of matrix multiplication:

Matrix multiplication differs from multiplication of numbers in four important ways, so some familiar rules fail:

- (i) $AB \neq BA$ in general. For example, with $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ we get $AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ but $BA = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.
- (ii) Even if $AB = AC$, the matrices B and C need not be equal. For example, with $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$, both products equal $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.
- (iii) $AB = O$ does not imply that $A = O$ or $B = O$. For example, $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = O$.
- (iv) $A^2 = I$ does not imply $A = I$ or $A = -I$. For example, $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ has $A^2 = I$.

Theorem 10.2.3 Ordinary properties of matrix multiplication

The rules that do hold are the ones we need for algebra:

- (i) $(\alpha A)B = \alpha(AB) = A(\alpha B)$ (scalars can be moved around).
- (ii) $(AB)C = A(BC)$ (associativity).
- (iii) $(A + B)C = AC + BC$ (distributivity).
- (iv) $C(A + B) = CA + CB$ (distributivity).
- (v) $A(\alpha B + \beta C) = \alpha AB + \beta AC$ (linearity).

Source page 35

Partitioning in Matrix Operations

A large matrix can be cut by horizontal and vertical lines into smaller matrices called *blocks*, and the blocks can then be treated as if they were entries:

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & \cdots & B_{1q} \\ \vdots & & \vdots \\ B_{p1} & \cdots & B_{pq} \end{bmatrix}.$$

If $m = p$, $n = q$ and each block A_{ij} has the same size as the corresponding block B_{ij} , the sum is computed block by block:

$$A + B = \begin{bmatrix} A_{11} + B_{11} & \cdots & A_{1n} + B_{1n} \\ \vdots & & \vdots \\ A_{m1} + B_{m1} & \cdots & A_{mn} + B_{mn} \end{bmatrix}.$$

If $n = p$, the product $C = AB$ has blocks

$$C_{ij} = \sum_{k=1}^n A_{ik} B_{kj},$$

provided that each product $A_{ik} B_{kj}$ is defined.

ex/ Cut two 3×3 matrices into blocks as shown by the lines:

$$A = \left[\begin{array}{cc|c} 2 & 4 & 1 \\ -1 & 3 & 0 \\ \hline 5 & 4 & 6 \end{array} \right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \left[\begin{array}{cc|c} 0 & 1 & 3 \\ 2 & -4 & -1 \\ \hline 5 & 8 & 2 \end{array} \right] = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

The sum and the product are then computed block by block:

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix} = \left[\begin{array}{cc|c} 2 & 5 & 4 \\ 1 & -1 & -1 \\ \hline 10 & 12 & 8 \end{array} \right].$$

$$\begin{aligned} AB &= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} \\ &= \left[\begin{array}{cc|c} 13 & -6 & 4 \\ 6 & -13 & -6 \\ \hline 38 & 37 & 23 \end{array} \right]. \end{aligned}$$

Partitioning is especially helpful for sparse matrices, where most of the entries are zero and many blocks can be skipped.

Source page 36

Def'n The *transpose* of $A = \{a_{ij}\}_{m \times n}$ swaps its rows and columns:

$$A^T = \{a_{ij}^T\}_{n \times m} \quad \text{where} \quad a_{ij}^T = a_{ji}.$$

ex/ For $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ we get $A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$.

Theorem 10.3.1

Transposition behaves well with the other operations, except that it reverses the order of a product:

- (i) $(A^T)^T = A$.
- (ii) $(A + B)^T = A^T + B^T$.
- (iii) $(\alpha A)^T = \alpha A^T$.
- (iv) $(AB)^T = B^T A^T$, and so $(ABCD)^T = D^T C^T B^T A^T$.

Proof for (iv) Let $C = AB$, so $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$. The (i, j) entry of C^T is the (j, i) entry of C :

$$c_{ij}^T = c_{ji} = \sum_{k=1}^n a_{jk} b_{ki} = \sum_{k=1}^n a_{kj}^T b_{ik}^T = \sum_{k=1}^n b_{ik}^T a_{kj}^T \Rightarrow C^T = B^T A^T. \quad \blacksquare$$

Note that for column vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$, the matrix product of a row with a column is a dot product:

Dot product as a matrix product

$$\mathbf{x}^T \mathbf{y} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i = \mathbf{x} \cdot \mathbf{y}.$$

Def'n

- (i) If $A^T = A$, the matrix A is called *symmetric*.
- (ii) If $A^T = -A$, the matrix A is called *skew-symmetric*.

Every square matrix is the sum of a symmetric part and a skew-symmetric part:

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T).$$

The first term is symmetric, because $(A + A^T)^T = A^T + A$, and the second is skew-symmetric, because $(A - A^T)^T = A^T - A = -(A - A^T)$.

Source page 37

Determinants

The determinant is a single number computed from a square matrix. Among other things, it tells us whether the matrix is invertible, as we will see later in this section. For a square matrix $A = \{a_{ij}\}_{n \times n}$ it is defined by *cofactor expansion* along a row:

Determinant by cofactor expansion

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \sum_{k=1}^n a_{ik} C_{ik} \quad (\text{any fixed row } i),$$

where the *cofactor* is $C_{ik} = (-1)^{i+k} M_{ik}$.

The *minor* M_{ik} is the determinant of the $(n-1) \times (n-1)$ matrix that is left when row i and column k of A are removed. The determinant of a 1×1 or 2×2 matrix is the starting point of the recursion.

ex/ Small determinants.

(i) For a 1×1 matrix, $\det[a_{11}] = a_{11}$.

(ii) For a 2×2 matrix, expanding along the first row gives the familiar formula

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}(-1)^{1+1}|a_{22}| + a_{12}(-1)^{1+2}|a_{21}| = a_{11}a_{22} - a_{12}a_{21}.$$

(iii) For a 3×3 matrix, expanding along the first row reduces the problem to three 2×2 determinants:

$$\begin{aligned} \det \begin{bmatrix} 4 & 7 & -2 \\ 0 & 3 & 2 \\ 1 & 5 & 6 \end{bmatrix} &= 4(-1)^{1+1} \begin{vmatrix} 3 & 2 \\ 5 & 6 \end{vmatrix} + 7(-1)^{1+2} \begin{vmatrix} 0 & 2 \\ 1 & 6 \end{vmatrix} \\ &\quad + (-2)(-1)^{1+3} \begin{vmatrix} 0 & 3 \\ 1 & 5 \end{vmatrix} \\ &= 4 \cdot 8 - 7 \cdot (-2) - 2 \cdot (-3) = 32 + 14 + 6 = 52. \end{aligned}$$

Comments

- (i) The expansion works along any row, so it is best to choose the row with the most zero entries, since zero entries contribute nothing.
- (ii) Cofactor expansion of large matrices is extremely time consuming: the number of operations for an $n \times n$ matrix grows like $N(n) \sim e \cdot n!$, which is enormous even for moderate n .

Source page 38

Properties of Determinants

Row operations change a determinant in a simple, predictable way. This gives a much faster way to compute determinants than cofactor expansion:

- > **D1.** Adding a multiple of one row to another does not change the determinant.
- > **D2.** Interchanging two rows changes the sign of the determinant.
- > **D3.** The determinant of a triangular matrix is the product of its diagonal entries.

example To compute the determinant of the matrix below, reduce it to triangular form using D1 and D2, and then multiply the diagonal. We first interchange the first and third rows (which flips the sign), then use $-2R_1 + R_2 \rightarrow R_2$ and $-\frac{2}{3}R_2 + R_3 \rightarrow R_3$.

$$\begin{aligned} \det \begin{bmatrix} 0 & 2 & -1 \\ 4 & 3 & 5 \\ 2 & 0 & -4 \end{bmatrix} &= - \begin{vmatrix} 2 & 0 & -4 \\ 4 & 3 & 5 \\ 0 & 2 & -1 \end{vmatrix} = - \begin{vmatrix} 2 & 0 & -4 \\ 0 & 3 & 13 \\ 0 & 2 & -1 \end{vmatrix} \\ &= - \begin{vmatrix} 2 & 0 & -4 \\ 0 & 3 & 13 \\ 0 & 0 & -29/3 \end{vmatrix} = -2 \cdot 3 \cdot (-29/3) = 58. \end{aligned}$$

Reducing to triangular form takes about $N_T(n) \sim \frac{2n^3}{3}$ operations, which is far fewer than $e \cdot n!$ when n is large.

Additional Properties of Determinants

- > **D4.** If A has a row or a column of zeros, then $\det A = 0$.
- > **D5.** If two rows (or two columns) are proportional, then $\det A = 0$.
- > **D6.** If a row or column is a linear combination of the other rows or columns, then $\det A = 0$.
- > **D7.** Multiplying every entry of one row or column by α multiplies the determinant by α .
- > **D8.** $\det(\alpha A) = \alpha^n \det A$ for an $n \times n$ matrix, because all n rows are multiplied by α .
- > **D9.** If A , B and C are identical except in one row (or column), and that row of C is the sum of the corresponding rows of A and B , then $\det C = \det A + \det B$.
- > **D10.** $\det(A^T) = \det A$.
- > **D11.** In general $\det(A + B) \neq \det A + \det B$. D9 is only a special case in which the sum of determinants does work.
- > **D12.** $\det(AB) = (\det A)(\det B)$.

Source page 39

Row and Column Spaces

A matrix can be read in two ways: as a stack of rows, or as a row of columns. Write the rows of A as $\mathbf{r}_i^T$ (so each $\mathbf{r}_i \in \mathbb{R}^n$) and the columns as $\mathbf{a}_i$ (so each $\mathbf{a}_i \in \mathbb{R}^m$):

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \vdots \\ \mathbf{r}_m^T \end{bmatrix}, \quad \mathbf{r}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \in \mathbb{R}^n.$$

$$A = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n] \quad \text{where } \mathbf{a}_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{bmatrix} \in \mathbb{R}^m.$$

Def'n The *row space* of A is the span of its rows, and the *column space* is the span of its columns:

$$\text{Row}(A) = \text{span}\{\mathbf{r}_1, \dots, \mathbf{r}_m\}, \quad \text{Col}(A) = \text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}.$$

Def'n The *row rank* is $\dim(\text{Row}(A))$ and the *column rank* is $\dim(\text{Col}(A))$.

ex/ Take the echelon matrix below, with rows $\mathbf{r}_1, \dots, \mathbf{r}_4$.

$$A = \begin{bmatrix} 1 & -3 & 2 & 5 & 3 \\ 0 & 0 & 1 & -4 & 2 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{r}_1 = \begin{bmatrix} 1 \\ -3 \\ 2 \\ 5 \\ 3 \end{bmatrix}, \quad \mathbf{r}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -4 \\ 2 \end{bmatrix}, \quad \mathbf{r}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 7 \end{bmatrix}, \quad \mathbf{r}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The zero row adds nothing, so the row space is spanned by the three non-zero rows, and the row rank is 3:

$$\text{Row}(A) = \text{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\} = \text{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\}.$$

The columns that contain the pivots (the leading or pivot columns) are $\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4$, and they span the column space, so the column rank is also 3:

$$\text{Col}(A) = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5\} = \text{span}\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\}.$$

Row-equivalent matrices have the same row space

If two matrices are row equivalent, they have the same row space and the same row rank.

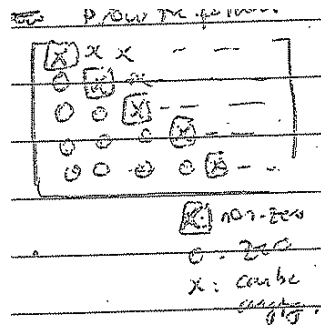
Proof If A is row equivalent to B , then B is obtained from A by a finite sequence of elementary row operations. Every elementary row operation replaces a row by a linear combination of rows, so every row of B is a linear combination of the rows of A . Hence $\text{Row}(B) \subset \text{Row}(A)$. Row operations can be undone by row operations, so A can also be obtained from B , and the same argument gives $\text{Row}(A) \subset \text{Row}(B)$. Therefore $\text{Row}(A) = \text{Row}(B)$. ■

Source page 40

Rows of an echelon matrix

The non-zero row vectors of an echelon matrix are linearly independent, so they form a basis for its row space.

Proof Let $e_1, e_2, \dots, e_k$ be the non-zero rows of an echelon matrix E , and let the first non-zero entry of e_i be in column p_i . Since E is in echelon form, these positions move to the right as we go down: $p_1 < p_2 < \dots < p_k$. The picture shows the pattern, with the first non-zero entry of each row marked.



Suppose, for contradiction, that these rows are linearly dependent. Then there are coefficients $\alpha_1, \dots, \alpha_k$, not all zero, such that

$$\alpha_1 e_1 + \dots + \alpha_k e_k = \mathbf{0}.$$

Look at column p_1 of this equation:

$$\alpha_1 e_{1p_1} + \alpha_2 e_{2p_1} + \dots + \alpha_k e_{kp_1} = 0.$$

All the rows below the first have a zero in column p_1 , that is, $e_{2p_1} = e_{3p_1} = \dots = e_{kp_1} = 0$, while $e_{1p_1} \neq 0$. Hence $\alpha_1 = 0$.

Now look at column p_2 . The first row's coefficient is already zero, and the rows below the second have zeros in this column, so $\alpha_2 e_{2p_2} = 0$ and $\alpha_2 = 0$.

Continuing like this, we find $\alpha_i = 0$ for $1 \leq i \leq k$. This contradicts the assumption that not all the coefficients are zero, so the non-zero rows are linearly independent. ■

So the row space and the row rank of any matrix can be found by reducing it to echelon form with elementary row operations and counting the non-zero rows.

Now we turn to the column space and the column rank.

Column rank is preserved

The column rank of A is the same as the column rank of its row echelon form E .

Proof Consider the solutions of $A\mathbf{x} = \mathbf{0}$. Elementary row operations turn this into $E\mathbf{x} = \mathbf{0}$ without changing the solution set, so any dependence among the columns of A is also a dependence among the columns of E , and vice versa. The proof continues on the next page.

Source page 41

Writing this out in terms of the columns $\mathbf{a}_i$ of A and the columns $\mathbf{e}_i$ of E , we get

$$[\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \mathbf{0} \iff [\mathbf{e}_1 \quad \cdots \quad \mathbf{e}_n] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \mathbf{0}.$$

So the columns of A and of E satisfy exactly the same dependence relations.

The pivot columns of the echelon matrix E span its column space and are linearly independent. (To see independence, suppose a combination of the pivot columns is zero. In the last pivot row only the last pivot column has a non-zero entry, so its coefficient must be zero; then the same argument applies to the previous pivot row, and so on, which is the same argument as in the theorem about the rows of an echelon matrix.) Denote the pivot columns by $\{\mathbf{e}_{p_1}, \dots, \mathbf{e}_{p_r}\}$. Every other column $\mathbf{e}_k$ of E is a combination of these, so for every non-pivot column k ,

$$\mathbf{e}_k = \sum_{i=1}^r c_{ki} \mathbf{e}_{p_i} \Rightarrow \mathbf{e}_k - \sum_{i=1}^r c_{ki} \mathbf{e}_{p_i} = \mathbf{0}.$$

Let $\mathbf{c}_k$ be the vector that has the entry c_{ki} in position p_i and zeros elsewhere. Then the relation above says

$$E \left(\begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} - \mathbf{c}_k \right) = \mathbf{0},$$

where the 1 is in the k^{th} position. By the equivalence of the two solution sets,

$$A \left(\begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} - \mathbf{c}_k \right) = \mathbf{0}.$$

This says that if a column $\mathbf{e}_k$ of E is a combination of the pivot columns of E , then the matching column $\mathbf{a}_k$ of A is the same combination of the matching columns $\{\mathbf{a}_{p_1}, \dots, \mathbf{a}_{p_r}\}$ of A . Therefore

$$\text{span}\{\mathbf{a}_{p_1}, \dots, \mathbf{a}_{p_r}\} = \text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}.$$

These columns are also linearly independent. We show this by contradiction. Suppose some non-zero coefficients solve

$$[\mathbf{a}_{p_1} \quad \mathbf{a}_{p_2} \quad \cdots \quad \mathbf{a}_{p_r}] \begin{bmatrix} \beta_{p_1} \\ \vdots \\ \beta_{p_r} \end{bmatrix} = \mathbf{0}.$$

Here p_i is the index of the i^{th} pivot column. This means $A\mathbf{z} = \mathbf{0}$ for the vector $\mathbf{z}$ with

$$z_i = \begin{cases} 0 & \text{if } i \notin \{p_1, \dots, p_r\} \\ \beta_i & \text{if } i \in \{p_1, \dots, p_r\}. \end{cases}$$

Since $A\mathbf{x} = \mathbf{0}$ and $E\mathbf{x} = \mathbf{0}$ have the same solutions, also $E\mathbf{z} = \mathbf{0}$, that is,

$$[\mathbf{e}_{p_1} \quad \cdots \quad \mathbf{e}_{p_r}] \begin{bmatrix} \beta_{p_1} \\ \vdots \\ \beta_{p_r} \end{bmatrix} = \mathbf{0},$$

so $\{\mathbf{e}_{p_1}, \dots, \mathbf{e}_{p_r}\}$ would be LD. This is a contradiction, because pivot columns are linearly independent. Thus $\{\mathbf{a}_{p_1}, \dots, \mathbf{a}_{p_r}\}$ is LI and spans $\text{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$, so it is a basis for $\text{Col}(A)$.

Source page 42

Hence $\dim(\text{Col}(A)) = r = \dim(\text{Col}(E))$. ■

We can now compare rows and columns. This is one of the central facts of linear algebra.

Row rank equals column rank

The column rank of a matrix A equals its row rank. In particular, for any matrix A the number of linearly independent rows equals the number of linearly independent columns. (Theorem 10.5.2)

Proof Let E be an echelon form of A . We have shown two things:

- (i) $\text{Row}(A) = \text{Row}(E)$, so the row rank of A is the number of pivots of E .
- (ii) The column rank of A equals the column rank of E , which is also the number of pivots of E .

Both ranks equal the number of pivots, so they are equal. ■

We call this common number the *rank* of A and write $r(A)$.

ex/ How many linearly independent vectors are there in

$$\left\{ \begin{bmatrix} 2 \\ 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ -2 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 3 \\ -2 \end{bmatrix} \right\} ?$$

Put the vectors in the columns of a matrix A and reduce it to echelon form:

$$\begin{aligned} \begin{bmatrix} 2 & 1 & -3 & 4 \\ 2 & 4 & -2 & 5 \\ 0 & 3 & 1 & 3 \\ 2 & 1 & -3 & -2 \end{bmatrix} &\xrightarrow{R_1/2 \rightarrow R_1} \begin{bmatrix} 1 & 1/2 & -3/2 & 2 \\ 2 & 4 & -2 & 5 \\ 0 & 3 & 1 & 3 \\ 2 & 1 & -3 & -2 \end{bmatrix} \\ &\xrightarrow{\substack{R_2-2R_1 \rightarrow R_2 \\ R_4-2R_1 \rightarrow R_4}} \begin{bmatrix} 1 & 1/2 & -3/2 & 2 \\ 0 & 3 & 1 & 1 \\ 0 & 3 & 1 & 3 \\ 0 & 0 & 0 & -6 \end{bmatrix} \\ &\xrightarrow{R_3-R_2 \rightarrow R_3} \begin{bmatrix} 1 & 1/2 & -3/2 & 2 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -6 \end{bmatrix} \xrightarrow{R_4+3R_3 \rightarrow R_4} \begin{bmatrix} 1 & 1/2 & -3/2 & 2 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

There are three pivots, in columns 1, 2 and 4, so three of the vectors are linearly independent (the first, second and fourth). The third vector is a combination of the first two: it equals $-\frac{5}{3}$ of the first plus $\frac{1}{3}$ of the second.

ex/ Now consider a system of four equations in six unknowns:

$$\begin{bmatrix} 1 & -1 & 1 & 3 & 0 & 2 \\ 1 & 0 & 3 & 3 & -1 & 6 \\ 2 & -1 & 2 & 1 & -1 & 7 \\ 1 & 0 & 5 & 8 & -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \\ 9 \\ 1 \end{bmatrix}.$$

Reduce the augmented matrix to echelon form:

$$\begin{bmatrix} 1 & -1 & 1 & 3 & 0 & 2 & 4 \\ 1 & 0 & 3 & 3 & -1 & 6 & 3 \\ 2 & -1 & 2 & 1 & -1 & 7 & 9 \\ 1 & 0 & 5 & 8 & -1 & 7 & 1 \end{bmatrix} \xrightarrow[\text{Row Operations}]{\text{Elementary}} \begin{bmatrix} 1 & -1 & 1 & 3 & 0 & 2 & 4 \\ 0 & 1 & 2 & 0 & -1 & 4 & -1 \\ 0 & 0 & 2 & 5 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Here x_1, x_2, x_3 are the pivot variables and x_4, x_5, x_6 are the free variables.

Source page 43

Observations

- (i) The rank of A and the rank of the augmented matrix $[A : \mathbf{c}]$ are the same, so the system is consistent.
- (ii) There are three free variables, so the solutions form a family with three parameters.

We give the free variables the names $x_6 = \alpha$, $x_5 = \beta$ and $x_4 = \gamma$, and solve for the pivot variables by back substitution:

$$\begin{aligned}x_3 &= -1 - \frac{1}{2}\alpha - \frac{5}{2}\gamma \\x_2 &= 1 - 3\alpha + \beta + 5\gamma \\x_1 &= 6 - \frac{9}{2}\alpha + \beta + \frac{9}{2}\gamma.\end{aligned}$$

In vector form this is a constant vector plus a combination of three direction vectors:

$$\mathbf{x} = \underbrace{\begin{bmatrix} 6 \\ 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{x}_0} + \alpha \underbrace{\begin{bmatrix} -9/2 \\ -3 \\ -1/2 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{x}_1} + \beta \underbrace{\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\mathbf{x}_2} + \gamma \underbrace{\begin{bmatrix} 9/2 \\ 5 \\ -5/2 \\ 1 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{x}_3}.$$

Observe that:

- (i) $\mathbf{x}_0$ is a particular solution of $A\mathbf{x} = \mathbf{c}$ (take $\alpha = \beta = \gamma = 0$).
- (ii) $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ solve the homogeneous equation $A\mathbf{x} = \mathbf{0}$. Indeed, for every α, β, γ ,

$$\begin{aligned}A\mathbf{x} &= A(\mathbf{x}_0 + \alpha\mathbf{x}_1 + \beta\mathbf{x}_2 + \gamma\mathbf{x}_3) \\&= \underbrace{A\mathbf{x}_0}_{\mathbf{c}} + \alpha A\mathbf{x}_1 + \beta A\mathbf{x}_2 + \gamma A\mathbf{x}_3 = \mathbf{c},\end{aligned}$$

so $\alpha A\mathbf{x}_1 + \beta A\mathbf{x}_2 + \gamma A\mathbf{x}_3 = \mathbf{0}$ for all α, β, γ , and therefore $A\mathbf{x}_i = \mathbf{0}$ for $1 \leq i \leq 3$.

- (iii) The matrix of these three vectors has rank 3 (its last three rows contain an identity matrix), so $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are linearly independent:

$$[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3] = \begin{bmatrix} -9/2 & 1 & 9/2 \\ -3 & 1 & 5 \\ -1/2 & 0 & -5/2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

This example illustrates a general fact. For $A\mathbf{x} = \mathbf{c}$ with $A \in \mathbb{R}^{m \times n}$ (or $\mathbb{C}^{m \times n}$), writing $r(A)$ for the rank:

How many solutions?

- (i) If $r(A) < r([A : \mathbf{c}])$, there are no solutions.
- (ii) If $r(A) = r([A : \mathbf{c}])$, there is at least one solution. If moreover $n = r(A)$, the solution is unique. If $n > r(A)$, the solution is non-unique, and the solutions form a family with $n - r(A)$ parameters.

Source page 44

(iii) If the system is consistent and $n > r(A)$, every solution can be written as

Particular plus homogeneous solution

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n,$$

where $\mathbf{x}_p$ is one particular solution of $A\mathbf{x} = \mathbf{c}$ and $\mathbf{x}_n$ solves the homogeneous system $A\mathbf{x}_n = \mathbf{0}$.

Def'n The set of all solutions of $A\mathbf{x} = \mathbf{0}$ is called the *null space* of A .

The number of pivots and the size of the null space always add up to the number of unknowns:

Rank-nullity theorem

$$r(A) + \dim(\text{Null}(A)) = n.$$

Proof Consider $A\mathbf{x} = \mathbf{0}$ and reduce A to an echelon matrix E to find the pivots and the free variables. The rank $r(A)$ is the number of pivot variables, and $\dim \text{Null}(A)$ is the number of free variables, since each free variable contributes one independent direction. Every one of the n unknowns is either a pivot variable or a free variable, so

$$\underbrace{\#\text{pivots}}_{r(A)} + \underbrace{\#\text{free variables}}_{\dim(\text{Null}(A))} = n. \quad \blacksquare$$

Inverse Matrix; Cramer's Rule, Factorization

To solve the scalar equation $ax = c$ we divide by a , which is the same as multiplying by a^{-1} , as long as $a \neq 0$:

$$\underbrace{a^{-1}a}_1 x = a^{-1}c \Rightarrow x = a^{-1}c.$$

We would like to do the same for $A\mathbf{x} = \mathbf{c}$: multiply both sides by a suitable matrix A^{-1} such that

$$\underbrace{A^{-1}A}_{\text{we want this to be } I} \mathbf{x} = A^{-1}\mathbf{c}.$$

But we have to be careful. For example, the matrix below has a left inverse, a matrix that multiplies it to give the identity:

$$\begin{bmatrix} -1 & -2 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and note that} \quad \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} -1 & -2 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Source page 45

Here is what goes wrong. We might multiply both sides of the system by the matrix $\begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \end{bmatrix}$:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

and conclude that $x_1 = 3, x_2 = 4$ is a solution. But it is not. The new system is not equivalent to the original one: the original system has no solution at all, and multiplying by a matrix can destroy that information. A matrix that is not square, or not of full rank, cannot serve as an inverse in this way, as the next theorem shows.

Inverses need full-rank square matrices

Consider $A\mathbf{x} = \mathbf{c}$. If A is not a full-rank square matrix, there is no matrix A^{-1} with $A^{-1}A = I$ such that $A\mathbf{x} = \mathbf{c}$ and $(A^{-1}A)\mathbf{x} = A^{-1}\mathbf{c}$ have the same solutions for every $\mathbf{c}$.

Proof Suppose such an A^{-1} existed for a matrix A that is not a full-rank square matrix. We show that this leads to a contradiction. There are two cases to examine.

- (i) Suppose $A\mathbf{x} = \mathbf{c}$ is inconsistent for some $\mathbf{c}$, so its solution set is empty. But $(A^{-1}A)\mathbf{x} = A^{-1}\mathbf{c}$ reads $\mathbf{x} = A^{-1}\mathbf{c}$ and always has exactly one solution. The two systems would then have different solution sets. So $A\mathbf{x} = \mathbf{c}$ must be consistent for all $\mathbf{c}$, which means that A has full row rank, $r(A_{m \times n}) = m$.
- (ii) Suppose $A\mathbf{x} = \mathbf{c}$ has more than one solution for some $\mathbf{c}$. Again $(A^{-1}A)\mathbf{x} = A^{-1}\mathbf{c}$ has the single solution $\mathbf{x} = A^{-1}\mathbf{c}$, so the solution sets would differ. Hence $A\mathbf{x} = \mathbf{c}$ can never have more than one solution, which means that A has full column rank, $r(A_{m \times n}) = n$.

Combining the two cases gives $r(A_{m \times n}) = m = n$, so A is square and of full rank. This contradicts the assumption. ■

Source page 46

The converse is also true: a full-rank square matrix has an inverse.

Existence of the inverse

If A is a square matrix of full rank, there exists a matrix A^{-1} with $A^{-1}A = I$.

Proof If $A^{-1}A = I$, then transposing gives $A^T(A^{-1})^T = I$. We show that this transposed equation has a unique solution. Let $(A^{-1})^T = B$; if we can find B , then $A^{-1} = B^T$. The equation $A^TB = I$ can be written column by column, where $\mathbf{e}_i$ is the i^{th} column of the identity (a 1 in position i , zeros elsewhere):

$$A^T \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}, \quad \mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

This is n separate linear systems:

$$A^T \mathbf{b}_i = \mathbf{e}_i, \quad 1 \leq i \leq n.$$

Since A is square and of full rank, so is A^T . Therefore each system $A^T \mathbf{b}_i = \mathbf{e}_i$ is consistent and has a unique solution. So there is exactly one B with $A^TB = I$. Then $A^{-1} = B^T$ exists, because $B^TA = (A^TB)^T = I$, and it is unique. ■

Def'n A full-rank square matrix is called *invertible* (or non-singular), and its unique inverse is written A^{-1} .

So far we only know $A^{-1}A = I$. The next theorem shows that the inverse also works from the other side.

Two-sided inverse

If A is invertible, then $A^{-1}A = AA^{-1} = I$.

Proof We have shown that $A^{-1}A = I$, so it remains to show that AA^{-1} is also I . Multiply $A^{-1}A = I$ on the left by A :

$$A^{-1}A = I \Rightarrow AA^{-1}A = A \Rightarrow (AA^{-1})A = A.$$

Source page 47

So $(AA^{-1})A\mathbf{x} = A\mathbf{x}$ for every $\mathbf{x}$. Since A has full rank, the vectors $A\mathbf{x}$ fill all of $\mathbb{R}^n$, so

$$(AA^{-1})\mathbf{z} = \mathbf{z} \quad \text{for every } \mathbf{z} \text{ in } \mathbb{R}^n.$$

Hence $(AA^{-1} - I)\mathbf{z} = \mathbf{0}$ for every $\mathbf{z}$. Taking $\mathbf{z} = \mathbf{e}_1, \dots, \mathbf{e}_n$ shows that all the entries of $AA^{-1} - I$ are zero. Thus $AA^{-1} = I$. ■

Def'n Negative powers are powers of the inverse: $A^{-n} \equiv (A^{-1})^n = \underbrace{(A^{-1}) \cdots (A^{-1})}_{n \text{ times}}.$

Theorem Algebra of inverse matrices:

Inverses combine with the other operations as follows:

Rules for inverses

- (a) $(A^{-1})^{-1} = A$.
- (b) $(A^n)^{-1} = (A^{-1})^n$.
- (c) $(AB)^{-1} = B^{-1}A^{-1}$, if both A and B are invertible (note the reversed order).
- (d) A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.

Proof

(a) $A^{-1}A = I$ and $AA^{-1} = I$, so A is the inverse of A^{-1} .

(b) This follows from (c) and the definition of the power:

$$A^n = \underbrace{AA \cdots A}_{n \text{ times}} \Rightarrow (A^n)^{-1} = [AA \cdots A]^{-1} = \underbrace{(A^{-1})(A^{-1}) \cdots (A^{-1})}_{n \text{ times}} \equiv A^{-n} = (A^{-1})^n.$$

(c) We check both products with the claimed inverse $B^{-1}A^{-1}$:

$$\begin{aligned} AB(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I, \\ (B^{-1}A^{-1})AB &= B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I. \end{aligned}$$

(d) $A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$ and $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$.

Theorem Inverses solve linear systems directly.

Solving with the inverse

If A is invertible, the unique solution of $A\mathbf{x} = \mathbf{c}$ is $\mathbf{x} = A^{-1}\mathbf{c}$.

Proof $A(A^{-1}\mathbf{c}) = (AA^{-1})\mathbf{c} = \mathbf{c}$, so $A^{-1}\mathbf{c}$ is a solution. It is the only one, because A has full column rank (it is invertible), so all n columns are pivot columns. ■

Source page 48

Determinants and Invertibility

The determinant gives a quick test for invertibility.

Determinant test

A is invertible if and only if $\det A \neq 0$.

ex/ For $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ we have $\det A = 4 - 6 = -2 \neq 0$, so A is invertible.

Proof Let E be the echelon form obtained from A by Gauss elimination. Row interchanges and adding multiples of one row to another change the determinant at most by a sign, so $\det A = \pm \det E$. In particular, $\det A \neq 0$ exactly when $\det E \neq 0$. Since E is triangular, $\det E$ is the product of its diagonal entries, the pivots. Now we prove both directions.

- (i) If A is invertible, it has full rank, so E has full rank: there are n pivots, all non-zero. Then $\det(E) \neq 0$, and so $\det(A) \neq 0$.
- (ii) If $\det A \neq 0$, then $\det(E) \neq 0$, so all the pivots are non-zero and E has full rank. Hence A has full rank, and A is invertible.

Theorem Cramer's Rule

For a square system with $\det A \neq 0$ there is an explicit formula for each unknown, in terms of determinants. Consider the $n \times n$ system $A\mathbf{x} = \mathbf{b}$ with $A = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n]$ written by its columns.

Cramer's rule

If $\det A \neq 0$, the system is uniquely solvable, and the i^{th} entry of the solution is

$$x_i = \frac{\det [\mathbf{a}_1 \ \cdots \ \mathbf{a}_{i-1} \ \mathbf{b} \ \mathbf{a}_{i+1} \ \cdots \ \mathbf{a}_n]}{\det A},$$

that is, the determinant of A with its i^{th} column replaced by $\mathbf{b}$, divided by $\det A$.

Proof Since $\det A \neq 0$, the matrix A is invertible, so the system has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$. We now find the entries of $\mathbf{x}$.

Source page 49

Let $\mathbf{x}$ be the unique solution. Writing the product $A\mathbf{x}$ as a combination of the columns of A shows that this combination equals $\mathbf{b}$:

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \underbrace{\mathbf{a}_1 x_1 + \cdots + \mathbf{a}_n x_n}_{=\mathbf{b}} = \sum_{k=1}^n \mathbf{a}_k x_k.$$

Now replace the i^{th} column of A by $\mathbf{b}$, and substitute this combination for $\mathbf{b}$. Because the determinant is additive and homogeneous in a single column (D9 and D7), it splits into a sum of n determinants. In every term with $k \neq i$ the column $\mathbf{a}_k$ appears twice, so that determinant is zero (D5). Only the term $k = i$ survives, and it is $x_i \det A$:

$$\begin{aligned} & \det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \cdots & \mathbf{a}_n \end{bmatrix} \\ &= \det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{i-1} & \sum_{k=1}^n \mathbf{a}_k x_k & \mathbf{a}_{i+1} & \cdots & \mathbf{a}_n \end{bmatrix} \\ &= \sum_{k=1}^n x_k \underbrace{\det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{i-1} & \mathbf{a}_k & \mathbf{a}_{i+1} & \cdots & \mathbf{a}_n \end{bmatrix}}_{\substack{0 \text{ if } k \neq i; \det A \text{ if } k=i}}. \end{aligned}$$

Dividing by $\det A$ gives the formula:

$$\Rightarrow \det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \cdots & \mathbf{a}_n \end{bmatrix} = x_i \det A \quad \Rightarrow \quad x_i = \frac{\det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{i-1} & \mathbf{b} & \mathbf{a}_{i+1} & \cdots & \mathbf{a}_n \end{bmatrix}}{\det A}. \quad \blacksquare$$

Cramer's rule gives a closed-form expression for $\mathbf{x}$, which is useful in analytical investigations.

ex/ Use Cramer's rule to solve

$$\begin{aligned} x_1 + 4x_2 + 5x_3 &= 2 \\ 4x_1 + 2x_2 + 5x_3 &= 3 \\ -3x_1 + 3x_2 - x_3 &= 1. \end{aligned}$$

First compute $\det A$ by reducing to triangular form:

$$\det A = \begin{vmatrix} 1 & 4 & 5 \\ 4 & 2 & 5 \\ -3 & 3 & -1 \end{vmatrix} = \det \begin{bmatrix} 1 & 4 & 5 \\ 0 & -14 & -15 \\ 0 & 15 & 14 \end{bmatrix} = -14 \cdot 14 + 15 \cdot 15 = 29.$$

Then each unknown is a determinant, with one column replaced by the right-hand side, divided by 29:

$$x_1 = \frac{\det \begin{bmatrix} 2 & 4 & 5 \\ 3 & 2 & 5 \\ 1 & 3 & -1 \end{bmatrix}}{29} = \frac{33}{29}, \quad x_2 = \frac{\det \begin{bmatrix} 1 & 2 & 5 \\ 4 & 3 & 5 \\ -3 & 1 & -1 \end{bmatrix}}{29} = \frac{35}{29}, \quad x_3 = \frac{\det \begin{bmatrix} 1 & 4 & 2 \\ 4 & 2 & 3 \\ -3 & 3 & 1 \end{bmatrix}}{29} = -\frac{23}{29}.$$

Inverses and Adjoint Matrix

The inverse of a square, full-rank matrix A can be found by solving $AX = I$. Written column by column, this means solving n linear systems, one for each column of the identity:

$$AX = I \quad \text{or} \quad \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix}, \quad A\mathbf{x}_j = \mathbf{e}_j, \quad 1 \leq j \leq n.$$

Source page 50

By Cramer's rule, the i^{th} entry of $\mathbf{x}_j$ is a determinant in which column i is replaced by $\mathbf{e}_j$. Expanding that determinant along the replaced column leaves a single cofactor, C_{ji} :

$$x_{ij} = \frac{|\mathbf{a}_1 \cdots \mathbf{a}_{i-1} \mathbf{e}_j \mathbf{a}_{i+1} \cdots \mathbf{a}_n|}{\det A} = \frac{C_{ji}}{\det A},$$

where C_{ji} is the (j, i) cofactor of A . Collecting these entries gives the inverse as the transposed matrix of cofactors, called the *adjoint* (or adjugate) of A , divided by the determinant:

Inverse by the adjoint

$$X = A^{-1} = \frac{[C_{ij}]^T}{\det A} = \frac{\text{Adj } A}{\det A}.$$

ex/ Take again $A = \begin{bmatrix} 1 & 4 & 5 \\ 4 & 2 & 5 \\ -3 & 3 & -1 \end{bmatrix}$ with $\det A = 29$. We solve $A [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3] = [\mathbf{e}_1 \ \mathbf{e}_2 \ \mathbf{e}_3]$,

where

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} x_{13} \\ x_{23} \\ x_{33} \end{bmatrix}.$$

By Cramer's rule, the entries of the first column $\mathbf{x}_1$ are

$$\begin{aligned} x_{11} &= \frac{\det \begin{bmatrix} 1 & 4 & 5 \\ 0 & 2 & 5 \\ 0 & 3 & -1 \end{bmatrix}}{\det A} = \frac{C_{11}}{\det A}, \\ x_{21} &= \frac{\det \begin{bmatrix} 1 & 1 & 5 \\ 4 & 0 & 5 \\ -3 & 0 & -1 \end{bmatrix}}{\det A} = \frac{C_{12}}{\det A}, \\ x_{31} &= \frac{\det \begin{bmatrix} 1 & 4 & 1 \\ 4 & 2 & 0 \\ -3 & 3 & 0 \end{bmatrix}}{\det A} = \frac{C_{13}}{\det A}, \end{aligned}$$

and the other columns follow in the same way. Together, the inverse is the transposed cofactor matrix over $\det A$:

$$\begin{aligned} X = A^{-1} &= \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} \\ &= \frac{1}{\det A} \underbrace{\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}^T}_{\text{adj } A}. \end{aligned}$$

A more practical technique for manual computation of A^{-1} : the adjoint formula needs many determinants, which is slow by hand. A much faster way is to solve all n systems $AA^{-1} = I$ at once.

Source page 51

Let $A^{-1} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]$, where the columns $\mathbf{x}_i$ are the unknowns. Then

$$AA^{-1} = A[\mathbf{x}_1 \ \cdots \ \mathbf{x}_n] = I = [\mathbf{e}_1 \ \cdots \ \mathbf{e}_n],$$

where $\mathbf{e}_i$ are the standard basis vectors. So $A\mathbf{x}_i = \mathbf{e}_i$ for each i . These could be solved one by one, but they can all be solved simultaneously by putting the identity next to A and row reducing:

Finding A^{-1} by row reduction

Form $[A \ [\mathbf{e}_1 \ \cdots \ \mathbf{e}_n]] = [A \ I]$ and use elementary row operations to reduce it to

$$[I \ [\mathbf{x}_1 \ \cdots \ \mathbf{x}_n]] = [I \ A^{-1}].$$

ex/ Find the inverse of $A = \begin{bmatrix} 1 & 3 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}$.

$$\begin{aligned} [A \ I] &= \left[\begin{array}{ccc|ccc} 1 & 3 & 0 & 1 & 0 & 0 \\ -2 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array} \right] \\ &\xrightarrow{R_2 \rightarrow R_2 + 2R_1} \left[\begin{array}{ccc|ccc} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 9 & 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{R_2 \rightarrow R_2/9} \left[\begin{array}{ccc|ccc} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1/9 & 2/9 & 1/9 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array} \right] \\ &\xrightarrow{R_3 \rightarrow R_3 - R_2} \left[\begin{array}{ccc|ccc} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1/9 & 2/9 & 1/9 & 0 \\ 0 & 0 & 8/9 & -2/9 & -1/9 & 1 \end{array} \right] \xrightarrow{R_3 \rightarrow R_3(9/8)} \left[\begin{array}{ccc|ccc} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1/9 & 2/9 & 1/9 & 0 \\ 0 & 0 & 1 & -1/4 & -1/8 & 9/8 \end{array} \right] \\ &\xrightarrow{R_2 \rightarrow R_2 - \frac{1}{9}R_3} \left[\begin{array}{ccc|ccc} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1/4 & 1/8 & -1/8 \\ 0 & 0 & 1 & -1/4 & -1/8 & 9/8 \end{array} \right] \\ &\xrightarrow{R_1 \rightarrow R_1 - 3R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/4 & -3/8 & 3/8 \\ 0 & 1 & 0 & 1/4 & 1/8 & -1/8 \\ 0 & 0 & 1 & -1/4 & -1/8 & 9/8 \end{array} \right]. \end{aligned}$$

The three columns on the right of the bar are A^{-1} .

LU-Factorization

LU factorization is another way to solve $A\mathbf{x} = \mathbf{c}$ for a square matrix A . The idea is to write A as a product of two simple matrices, one lower triangular and one upper triangular. If Gauss elimination of A needs no row interchanges, then

LU factorization

$$A = \underbrace{L}_{\text{lower triangular}} \underbrace{U}_{\text{upper triangular}}.$$

If row interchanges are needed, one instead factorizes $PA = LU$ with a permutation matrix P .

Source page 52

The factorization is not unique. To see why, count unknowns. A 3×3 matrix A has 9 entries. The two triangular factors have 6 parameters each:

$$A = \underbrace{\begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}}_{6 \text{ parameters}} \underbrace{\begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}}_{6 \text{ parameters}}.$$

That is 12 unknowns for only 9 equations (for an $n \times n$ matrix, $n^2 + n$ unknowns). In *Doolittle's method* we fix the freedom by setting $l_{11} = l_{22} = l_{33} = 1$ and solving for the remaining 9 parameters. For an $n \times n$ matrix, $l_{11} = \dots = l_{nn} = 1$ and there are n^2 remaining parameters.

ex/ Factorize $A = \begin{bmatrix} 2 & -3 & 3 \\ 6 & -8 & 7 \\ -2 & 6 & -1 \end{bmatrix}$ with Doolittle's method. Multiplying out the factors gives

$$\underbrace{\begin{bmatrix} 2 & -3 & 3 \\ 6 & -8 & 7 \\ -2 & 6 & -1 \end{bmatrix}}_A = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} \\ = \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ l_{21}u_{11} & l_{21}u_{12} + u_{22} & l_{21}u_{13} + u_{23} \\ l_{31}u_{11} & l_{31}u_{12} + l_{32}u_{22} & l_{31}u_{13} + l_{32}u_{23} + u_{33} \end{bmatrix}.$$

Comparing entries with A , row by row, each unknown can be found from the ones already known:

$$\begin{aligned} 2 &= u_{11}, & -3 &= u_{12}, & 3 &= u_{13}, \\ 6 &= l_{21}u_{11} \Rightarrow l_{21} = 6/2 = 3; & -8 &= l_{21}u_{12} + u_{22} \\ &\Rightarrow u_{22} = -8 - 3 \cdot (-3) = 1, \\ 7 &= l_{21}u_{13} + u_{23} \Rightarrow u_{23} = 7 - 3 \cdot 3 = -2, \\ -2 &= l_{31}u_{11} \Rightarrow l_{31} = -1; & 6 &= l_{31}u_{12} + l_{32}u_{22} \Rightarrow l_{32} = 3, \\ -1 &= l_{31}u_{13} + l_{32}u_{23} + u_{33} \Rightarrow u_{33} = -1 - (-1) \cdot 3 - 3 \cdot (-2) = 8. \end{aligned}$$

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

Once $A = LU$ is known, the system $A\mathbf{x} = \mathbf{c}$ is solved in two easy steps. Name the product $U\mathbf{x} = \mathbf{y}$. First solve $L\mathbf{y} = \mathbf{c}$ for $\mathbf{y}$ by forward substitution, then solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$ by backward substitution:

$$L \underbrace{U\mathbf{x}}_{\mathbf{y}} = \mathbf{c} \Rightarrow L\mathbf{y} = \mathbf{c} \text{ (forward substitution),} \quad U\mathbf{x} = \mathbf{y} \text{ (backward substitution).}$$

This needs fewer operations than Gauss–Jordan: about $n^3/3$ multiplications for the factorization, versus about $n^3/2$ for Gauss–Jordan, and the factors L and U can be reused for every new right-hand side $\mathbf{c}$.

Source page 53

The Eigenvalue Problem

(Chapter 11 in the textbook, Lecture notes.)

A square matrix A acts on vectors: it transforms a vector $\mathbf{x}$ into the vector $A\mathbf{x}$, a map from $\mathbb{C}^n$ to $\mathbb{C}^n$. Most vectors are turned and stretched by this map. Some special vectors keep their direction and are only scaled. These are the *eigenvectors*, and the eigenvalue problem is to find them: for a square matrix A , find the vectors $\mathbf{x} \neq \mathbf{0}$ and scalars λ with

The eigenvalue equation

$$A\mathbf{x} = \lambda\mathbf{x} \quad \text{for a scalar } \lambda.$$

To solve it, move everything to one side:

$$A\mathbf{x} = \lambda\mathbf{x} \Rightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Rightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}.$$

For a given λ this is a homogeneous system, so it is always consistent, and $\mathbf{x} = \mathbf{0}$ is always a solution. But we are interested in non-zero solutions. A homogeneous system has non-zero solutions exactly when its matrix is not invertible, that is, when

$$\underbrace{\det(A - \lambda I) = 0}_{\text{a polynomial equation in } \lambda}.$$

So non-trivial solutions of $(A - \lambda I)\mathbf{x} = \mathbf{0}$ exist only for those λ that solve this equation.

Def'n A pair $(\lambda, \mathbf{x})$ with $\mathbf{x} \neq \mathbf{0}$ that solves $A\mathbf{x} = \lambda\mathbf{x}$ is an *eigenpair*: λ is an *eigenvalue* of A and $\mathbf{x}$ is a corresponding *eigenvector*.

Def'n The polynomial $\det(A - \lambda I)$ is called the *characteristic polynomial* of A . Its roots are exactly the eigenvalues of A .

Source page 54

ex/ Find the eigenvalues and the corresponding eigenvectors (the eigenspaces) of

$$A = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}.$$

The characteristic polynomial factors as

$$\det(A - \lambda I) = \begin{vmatrix} 2 - \lambda & 2 & 1 \\ 1 & 3 - \lambda & 1 \\ 1 & 2 & 2 - \lambda \end{vmatrix} = -\lambda^3 + 7\lambda^2 - 11\lambda + 5 = -(\lambda - 5)(\lambda - 1)^2.$$

So $\lambda_1 = 5$ is a simple eigenvalue, and $\lambda_2 = 1$ is a repeated eigenvalue with multiplicity 2. For each eigenvalue we solve $(A - \lambda I)\mathbf{x} = \mathbf{0}$.

For $\lambda_1 = 5$:

$$(A - 5I)\mathbf{x} = \begin{bmatrix} -3 & 2 & 1 \\ 1 & -2 & 1 \\ 1 & 2 & -3 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

Gauss elimination gives

$$\begin{bmatrix} -3 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{array}{l} x_1 \text{ \& } x_2 \text{ are pivot variables,} \\ x_3 \text{ is a free variable.} \end{array}$$

Setting $x_3 = \alpha$ gives $x_2 = \alpha$ and then $x_1 = \alpha$, so $\mathbf{x} = \begin{bmatrix} \alpha \\ \alpha \\ \alpha \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. The eigenspace of $\lambda_1 = 5$ is

$$\text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

For $\lambda_2 = 1$:

$$(A - I)\mathbf{x} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix} \mathbf{x} = \mathbf{0}.$$

Gauss elimination leaves a single non-zero row:

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow x_1 \text{ is a pivot variable, and } x_2 \text{ \& } x_3 \text{ are free variables.}$$

Setting $x_3 = \alpha$ and $x_2 = \beta$ gives $x_1 = -2\beta - \alpha$, so

$$\mathbf{x} = \begin{bmatrix} -2\beta - \alpha \\ \beta \\ \alpha \end{bmatrix} = \alpha \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}.$$

The eigenspace associated with $\lambda_2 = 1$ is two-dimensional:

$$\text{span} \left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

Notes

- (i) If $\mathbf{x}$ is an eigenvector, then so is $c\mathbf{x}$ for any $c \neq 0$. Eigenvectors are only determined up to a scale factor.
- (ii) If λ is an eigenvalue with multiplicity k , its eigenspace may have a dimension lower than k . The next example shows this.

Source page 55

ex/ Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Then

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 0 & 1 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^3 = 0.$$

So $\lambda_1 = 1$ is the only eigenvalue, with multiplicity 3. Its eigenspace is the set of solutions of

$$(A - I)\mathbf{x} = \mathbf{0} \Rightarrow \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Here x_1 and x_3 are pivot variables and x_2 is free. The equations force $x_1 = 0$ and $x_3 = 0$, and $x_2 = \alpha$ is arbitrary, so $\mathbf{x} = \begin{bmatrix} 0 \\ \alpha \\ 0 \end{bmatrix} = \alpha \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. The eigenspace is

$$\text{span} \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\},$$

which is only one-dimensional even though the eigenvalue has multiplicity three.

A Special Case: Symmetric Matrices

Symmetric matrices have especially nice eigenvalues and eigenvectors. We collect four facts about them, each building on the one before.

Theorem 11.3.1

If A is real and symmetric, all its eigenvalues are real.

Proof Let $(\lambda, \mathbf{x})$ be an eigenpair of A , possibly with complex entries, and write $\mathbf{x}^H$ for the conjugate transpose of $\mathbf{x}$. Multiply $A\mathbf{x} = \lambda\mathbf{x}$ on the left by $\mathbf{x}^H$. Because A is real and symmetric, $A^H = A$, so $\mathbf{x}^H(A\mathbf{x}) = (A\mathbf{x})^H\mathbf{x}$. Therefore

$$\begin{aligned} A\mathbf{x} = \lambda\mathbf{x} &\Rightarrow \mathbf{x}^H(A\mathbf{x}) = \lambda\mathbf{x}^H\mathbf{x} \\ &= (A\mathbf{x})^H\mathbf{x} = (\lambda\mathbf{x})^H\mathbf{x} = \lambda^*\mathbf{x}^H\mathbf{x}. \end{aligned}$$

Comparing the two ends gives

$$\Rightarrow \lambda\mathbf{x}^H\mathbf{x} = \lambda^*\mathbf{x}^H\mathbf{x} \Rightarrow (\lambda - \lambda^*)\mathbf{x}^H\mathbf{x} = 0.$$

Since $\mathbf{x} \neq \mathbf{0}$, we have $\mathbf{x}^H\mathbf{x} > 0$, so $\lambda = \lambda^*$, and λ is real. ■

Theorem 11.3.2

If λ is an eigenvalue of a real symmetric matrix A with multiplicity k , then its eigenspace is k -dimensional.

Proof This is proved after Theorem 11.4.4.

Source page 56

Theorem 11.3.3

If A is real and symmetric, eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof Let $(\lambda_1, \mathbf{x}_1)$ and $(\lambda_2, \mathbf{x}_2)$ be eigenpairs of A with $\lambda_1 \neq \lambda_2$. As before, move A across the product using $A^H = A$, and recall that λ_2 is real by Theorem 11.3.1, so $\lambda_2^* = \lambda_2$:

$$\begin{aligned} A\mathbf{x}_1 = \lambda_1\mathbf{x}_1 &\Rightarrow \mathbf{x}_2^H(A\mathbf{x}_1) = \lambda_1\mathbf{x}_2^H\mathbf{x}_1 \\ &= (A\mathbf{x}_2)^H\mathbf{x}_1 = \lambda_2^*\mathbf{x}_2^H\mathbf{x}_1 = \lambda_2\mathbf{x}_2^H\mathbf{x}_1. \end{aligned}$$

Comparing the ends,

$$\Rightarrow \lambda_1\mathbf{x}_2^H\mathbf{x}_1 = \lambda_2\mathbf{x}_2^H\mathbf{x}_1 \Rightarrow (\lambda_1 - \lambda_2)\mathbf{x}_2^H\mathbf{x}_1 = 0.$$

Since $\lambda_1 - \lambda_2 \neq 0$, we must have $\mathbf{x}_2^H\mathbf{x}_1 = 0$. ■

Theorem 11.3.4

If $A_{n \times n}$ is real and symmetric, then its eigenvectors provide an orthogonal basis for $\mathbb{R}^n$.

Proof There are two cases.

- (i) If A has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$, their eigenvectors are orthogonal by Theorem 11.3.3, so they form an orthogonal basis.
- (ii) If A has repeated eigenvalues, say distinct eigenvalues $\lambda_1, \dots, \lambda_m$ where λ_i has multiplicity k_i , then each eigenspace is k_i -dimensional by Theorem 11.3.2. So each eigenspace has an orthogonal basis $\{\mathbf{u}_{i1}, \dots, \mathbf{u}_{ik_i}\}$ (for example by Gram–Schmidt). Eigenspaces belonging to different eigenvalues are orthogonal to each other by Theorem 11.3.3, so putting all these bases together gives an orthogonal basis of $\mathbb{R}^n$:

$$\bigcup_{i=1}^m \{\mathbf{u}_{i1}, \dots, \mathbf{u}_{ik_i}\}.$$

There are exactly n vectors in total, because $\sum_{i=1}^m k_i = n$. ■

Source page 57

Diagonalization

A diagonal matrix is the easiest kind of matrix to work with. Diagonalization tries to turn a given matrix into one by changing coordinates. If we can find a matrix Q such that

Diagonalization

$$Q^{-1}AQ = D \quad \leftarrow \text{diagonal}$$

then we say that A is *diagonalizable* and that Q *diagonalizes* A . The next theorem says which matrices are diagonalizable and what Q and D are.

Theorem 11.4.1

Let A be an $n \times n$ matrix.

Theorem 11.4.1

- (i) A is diagonalizable if and only if A has n linearly independent eigenvectors.
- (ii) If A has n linearly independent eigenvectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ and $Q = [\mathbf{e}_1 \ \cdots \ \mathbf{e}_n]$, then $Q^{-1}AQ = D$ is diagonal with $d_i = \lambda_i$, the eigenvalue belonging to $\mathbf{e}_i$.

Proof (i) We show the two directions separately.

($\Rightarrow$) Suppose A is diagonalizable, so $Q^{-1}AQ = D$ for some $Q = [\mathbf{e}_1 \ \cdots \ \mathbf{e}_n]$. Multiply by Q on the left and compare columns:

$$\begin{aligned} Q^{-1}AQ = D &\Rightarrow AQ = QD = [\mathbf{e}_1 \ \cdots \ \mathbf{e}_n] \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \\ &\Rightarrow A[\mathbf{e}_1 \ \cdots \ \mathbf{e}_n] = [d_1\mathbf{e}_1 \ d_2\mathbf{e}_2 \ \cdots \ d_n\mathbf{e}_n]. \end{aligned}$$

So $A\mathbf{e}_i = d_i\mathbf{e}_i$ for every i . No column of an invertible matrix is zero, so $\mathbf{e}_i \neq \mathbf{0}$, and therefore $\mathbf{e}_i$ is an eigenvector with eigenvalue d_i . Since Q is invertible, its columns $\mathbf{e}_i$ are linearly independent. This gives n linearly independent eigenvectors. ■

($\Leftarrow$) Suppose A has n linearly independent eigenvectors $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. Put them in the columns of Q and the eigenvalues on the diagonal of Λ . The statements $A\mathbf{e}_i = \lambda_i\mathbf{e}_i$ for all i together say

$$A[\mathbf{e}_1 \ \cdots \ \mathbf{e}_n] = \underbrace{[\mathbf{e}_1 \ \cdots \ \mathbf{e}_n]}_Q \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_\Lambda.$$

The columns of Q are linearly independent, so Q is invertible, and

$$\Rightarrow Q^{-1}AQ = \Lambda \leftarrow \text{diagonal} \Rightarrow A \text{ is diagonalizable.} \quad \blacksquare$$

Source page 58

Theorem 11.4.2

When are there enough eigenvectors? One easy case is when all the eigenvalues are different.

Theorem 11.4.2

If $A_{n \times n}$ has distinct eigenvalues $\lambda_1, \dots, \lambda_n$, then the corresponding eigenvectors are linearly independent.

Proof Suppose $\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n = \mathbf{0}$. We will show that this forces $\alpha_1 = \dots = \alpha_n = 0$, which means that $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is LI. Written with a column of ones, the assumption says

$$[\alpha_1 \mathbf{e}_1 \ \dots \ \alpha_n \mathbf{e}_n] \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{0}.$$

Applying A to the assumption multiplies each term by its own eigenvalue, $A(\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) = \lambda_1 \alpha_1 \mathbf{e}_1 + \dots + \lambda_n \alpha_n \mathbf{e}_n = \mathbf{0}$, and applying A again and again gives, for every power up to $n-1$,

$$A^{n-1}(\alpha_1 \mathbf{e}_1 + \dots + \alpha_n \mathbf{e}_n) = \lambda_1^{n-1} \alpha_1 \mathbf{e}_1 + \dots + \lambda_n^{n-1} \alpha_n \mathbf{e}_n = \mathbf{0}.$$

Collecting all of these equations into one matrix equation,

$$[\alpha_1 \mathbf{e}_1 \ \dots \ \alpha_n \mathbf{e}_n] \underbrace{\begin{bmatrix} 1 & \lambda_1 & \dots & \lambda_1^{n-1} \\ 1 & \lambda_2 & \dots & \lambda_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{n-1} \end{bmatrix}}_{\text{Vandermonde matrix}} = [\mathbf{0} \ \mathbf{0} \ \dots \ \mathbf{0}] = O_{n \times n}.$$

A Vandermonde matrix has a non-zero determinant when the λ_i are all different (a standard fact about these matrices), so it is invertible. Multiplying on the right by its inverse gives

$$\Rightarrow [\alpha_1 \mathbf{e}_1 \ \dots \ \alpha_n \mathbf{e}_n] = O.$$

Each eigenvector is non-zero, $\mathbf{e}_i \neq \mathbf{0}$, so each $\alpha_i = 0$. Hence $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is LI. ■

Theorem 11.4.3

Theorem 11.4.3

If $A_{n \times n}$ has n distinct eigenvalues, it is diagonalizable.

Proof n distinct eigenvalues give n linearly independent eigenvectors (Theorem 11.4.2), and n linearly independent eigenvectors make A diagonalizable (Theorem 11.4.1). ■

Theorem 11.4.4

Symmetric matrices are always diagonalizable, even when eigenvalues repeat, and moreover the matrix Q can be chosen orthogonal.

Theorem 11.4.4

Every real symmetric matrix A is diagonalizable, by an orthogonal matrix Q (that is, $Q^T Q = I$).

Proof We give a proof that does not use Theorem 11.3.2; this is what makes the proof of Theorem 11.3.2 below legitimate.

Source page 59

The proof is by induction on the size n of the matrix.

For a 1×1 symmetric matrix, $A = [a_{11}]$ is already diagonal, so it is certainly diagonalizable (with $Q = 1$).

Now assume that every symmetric matrix of size $(n-1) \times (n-1)$ is diagonalizable in the following way:

$$Q^T A Q = D \quad \leftarrow \text{diagonal} \quad \text{where } Q^T Q = I \text{ or } Q^{-1} = Q^T.$$

We show that every $n \times n$ symmetric matrix $A_{n \times n}$ then also has an orthogonal $Q_{n \times n}$ that diagonalizes it.

Let $(\lambda, \mathbf{x})$ be an eigenpair of $A_{n \times n}$. The eigenvalue λ is real by Theorem 11.3.1, and we choose the eigenvector real, with length one: $\mathbf{x}^T \mathbf{x} = 1$. Extend $\mathbf{x}$ to an orthonormal basis $\{\mathbf{y}_1, \dots, \mathbf{y}_{n-1}, \mathbf{x}\}$ of $\mathbb{R}^n$, and collect the new vectors in the $n \times (n-1)$ matrix $Y = [\mathbf{y}_1 \ \cdots \ \mathbf{y}_{n-1}]$. It has the following properties:

- (i) $Y^T Y = I_{n-1}$, since $\mathbf{y}_i^T \mathbf{y}_j = \delta_{ij}$.
- (ii) $\mathbf{x}^T Y = \mathbf{0}^T$, since $\mathbf{x}^T \mathbf{y}_i = 0$.
- (iii) $Y^T A Y$ is symmetric, because A is symmetric.

So $Y^T A Y$ is a symmetric matrix of size $(n-1) \times (n-1)$. By the induction hypothesis there is an orthogonal $Q_{(n-1) \times (n-1)}$ that diagonalizes it:

$$Q_{(n-1) \times (n-1)}^T Y^T A Y Q_{(n-1) \times (n-1)} = D. \quad (\star 1)$$

Now define $Q_{n \times n} = [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}]$. We show that it is orthogonal and that it diagonalizes A . First, orthogonality:

Source page 60

$$\begin{aligned}
Q_{n \times n}^T Q_{n \times n} &= [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}]^T [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}] \\
&= \begin{bmatrix} \mathbf{x}^T & \\ Q_{(n-1) \times (n-1)}^T Y^T \end{bmatrix} [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}] \\
&= \begin{bmatrix} \mathbf{x}^T \mathbf{x} & \mathbf{x}^T Y Q_{(n-1) \times (n-1)} \\ Q_{(n-1) \times (n-1)}^T Y^T \mathbf{x} & Q_{(n-1) \times (n-1)}^T Y^T Y Q_{(n-1) \times (n-1)} \end{bmatrix}.
\end{aligned}$$

The four blocks are:

$$\begin{aligned}
> \mathbf{x}^T \mathbf{x} &= 1. \\
> \mathbf{x}^T Y Q_{(n-1) \times (n-1)} &= \mathbf{0}^T, \text{ since } \mathbf{x}^T Y = \mathbf{0}^T. \\
> Q_{(n-1) \times (n-1)}^T Y^T \mathbf{x} &= \mathbf{0}, \text{ since } Y^T \mathbf{x} = \mathbf{0}. \\
> Q_{(n-1) \times (n-1)}^T Y^T Y Q_{(n-1) \times (n-1)} &= Q_{(n-1) \times (n-1)}^T \underbrace{(Y^T Y)}_{I_{n-1}} Q_{(n-1) \times (n-1)} = Q_{(n-1) \times (n-1)}^T Q_{(n-1) \times (n-1)} = \\
&I_{n-1}.
\end{aligned}$$

So

$$\Rightarrow Q_{n \times n}^T Q_{n \times n} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & I_{(n-1) \times (n-1)} \end{bmatrix} = I_{n \times n}.$$

Second, we compute $Q^T A Q$:

$$\begin{aligned}
Q_{n \times n}^T A_{n \times n} Q_{n \times n} &= [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}]^T A_{n \times n} [\mathbf{x} \ Y Q_{(n-1) \times (n-1)}] \\
&= \begin{bmatrix} \mathbf{x}^T A \mathbf{x} & \mathbf{x}^T A Y Q_{(n-1) \times (n-1)} \\ (\mathbf{x}^T A Y Q_{(n-1) \times (n-1)})^T & (Y Q_{(n-1) \times (n-1)})^T A Y Q_{(n-1) \times (n-1)} \end{bmatrix}.
\end{aligned}$$

Its blocks are:

$$\begin{aligned}
> \mathbf{x}^T A \mathbf{x} &= \mathbf{x}^T \lambda \mathbf{x} = \lambda. \\
> \underbrace{\mathbf{x}^T A}_{\lambda \mathbf{x}^T} Y Q_{(n-1) \times (n-1)} &= \lambda \underbrace{\mathbf{x}^T Y}_{\mathbf{0}^T} Q_{(n-1) \times (n-1)} = \mathbf{0}^T. \\
> Q_{(n-1) \times (n-1)}^T (Y^T A Y) Q_{(n-1) \times (n-1)} &= D, \text{ by equation } (\star 1).
\end{aligned}$$

Therefore

$$\Rightarrow Q_{n \times n}^T A_{n \times n} Q_{n \times n} = \begin{bmatrix} \lambda & \mathbf{0}^T \\ \mathbf{0} & D \end{bmatrix} \text{ is diagonal, } Q_{n \times n}^T Q_{n \times n} = I.$$

Thus $A_{n \times n}$ is diagonalizable by an orthogonal matrix. ■

Source page 61

Revisit Theorem 11.3.2

Now that we know every real symmetric matrix is diagonalizable, we can prove the fact that was postponed earlier.

Theorem 11.3.2

If λ is an eigenvalue of a real symmetric matrix A with multiplicity k , then its eigenspace is k -dimensional.

proof By Theorem 11.4.4, a real symmetric $A_{n \times n}$ is diagonalizable by an orthogonal matrix:

$$Q^T A Q = Q^{-1} A Q = D.$$

Since $Q^T(A - \lambda I)Q = D - \lambda I$ and $\det Q \det Q^T = 1$, the characteristic polynomial is $\det(A - \lambda I) = \prod_{i=1}^n (d_i - \lambda)$. So an eigenvalue of multiplicity k appears exactly k times on the diagonal of D . The k matching columns of Q are eigenvectors (because $AQ = QD$) and they are linearly independent, so the eigenspace has dimension at least k . It cannot be larger: eigenspaces of different eigenvalues are linearly independent of each other, and the multiplicities add up to n , so the dimensions cannot add up to more than n . Hence the eigenspace has dimension exactly k . ■

Closing Note

Even if A is not diagonalizable, it can still be brought to its Jordan form (over $\mathbb{C}$), which is upper triangular, with the eigenvalues on the diagonal and only 0's and 1's directly above the diagonal:

$$P^{-1}AP = \underbrace{J}_{\text{Jordan form for } A}.$$

The textbook discusses this on page 582 (exercise 11.4.10).

Cayley–Hamilton Theorem

A matrix satisfies its own characteristic equation. Let $A_{n \times n}$ have the characteristic equation

$$\lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_{n-1} \lambda + \alpha_n = 0.$$

Cayley–Hamilton theorem

Replacing λ by A (and the constant term by $\alpha_n I$) gives the zero matrix:

$$A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I = O.$$

Source page 62

proof We prove the theorem for the case where A has a full set of eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ (the theorem is true in general, but this case shows the idea). Collect the eigenvectors in V and the eigenvalues on the diagonal of Λ :

$$A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}}_V \underbrace{\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}}_\Lambda.$$

Some of the λ_i may be equal, but the columns of V are linearly independent, so V is invertible, and

$$V^{-1}AV = \Lambda.$$

Powers of A are easy to follow through this change of coordinates, because the V and V^{-1} in the middle cancel:

$$(V^{-1}AV)^2 = V^{-1}A \underbrace{VV^{-1}}_I AV = V^{-1}A^2V = \Lambda^2$$

and in general

$$(V^{-1}AV)^k = V^{-1}A^kV = \Lambda^k = \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix}.$$

Each eigenvalue λ_i satisfies the characteristic equation, so the diagonal matrix of those polynomial values is zero:

$$\begin{bmatrix} \lambda_1^n + \alpha_1 \lambda_1^{n-1} + \cdots + \alpha_{n-1} \lambda_1 + \alpha_n & & \\ & \ddots & \\ & & \lambda_n^n + \alpha_1 \lambda_n^{n-1} + \cdots + \alpha_{n-1} \lambda_n + \alpha_n \end{bmatrix} = O.$$

This diagonal matrix is the same as the polynomial evaluated at Λ . Rewriting each power of Λ as $V^{-1}A^kV$ gives

$$\begin{aligned} &= \Lambda^n + \alpha_1 \Lambda^{n-1} + \cdots + \alpha_{n-1} \Lambda + \alpha_n I = O \\ \Rightarrow &V^{-1}A^nV + \alpha_1 V^{-1}A^{n-1}V + \cdots + \alpha_{n-1} V^{-1}AV + \alpha_n V^{-1}IV = O \\ \Rightarrow &V^{-1}(A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I)V = O. \end{aligned}$$

Multiplying on the left by V and on the right by V^{-1} removes the outer factors:

$$A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1} A + \alpha_n I = O. \quad \blacksquare$$

Source page 63

Cayley–Hamilton is useful because it lets us reduce high powers of a matrix to low ones.

ex/ Let $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$. Its characteristic equation is

$$\begin{vmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{vmatrix} = 0 \Rightarrow \lambda^2 + 3\lambda + 2 = 0, \quad \lambda_1 = -1, \quad \lambda_2 = -2.$$

We can check the theorem directly:

$$A^2 = \begin{bmatrix} -2 & -3 \\ 6 & 7 \end{bmatrix}, \quad A^2 + 3A + 2I = \begin{bmatrix} -2 & -3 \\ 6 & 7 \end{bmatrix} + \begin{bmatrix} 0 & 3 \\ -6 & -9 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = O.$$

This says $A^2 = -3A - 2I$. So A^2 is a combination of A and I , and every higher power can be reduced in the same way:

$$A^k = \beta_1 A + \beta_0 I \quad \text{where } \beta_1 \text{ and } \beta_0 \text{ depend on } k.$$

For an $n \times n$ matrix in general, $A^k = \sum_{i=0}^{n-1} \beta_i A^i$.

How do we find the coefficients β_i ? Let λ_j be an eigenvalue of A with eigenvector $\mathbf{v}_j$. Applying both sides of $A^k = \sum_i \beta_i A^i$ to $\mathbf{v}_j$, and using $A^i \mathbf{v}_j = \lambda_j^i \mathbf{v}_j$,

$$\underbrace{A^k \mathbf{v}_j}_{\lambda_j^k \mathbf{v}_j} = \left(\sum_{i=0}^{n-1} \beta_i A^i \right) \mathbf{v}_j \Rightarrow \lambda_j^k \mathbf{v}_j = \left(\sum_{i=0}^{n-1} \beta_i \lambda_j^i \right) \mathbf{v}_j.$$

Since $\mathbf{v}_j \neq \mathbf{0}$, the scalars must agree. So every eigenvalue gives one equation for the unknown β_i :

Finding the coefficients β_i

$$\lambda_j^k = \sum_{i=0}^{n-1} \beta_i \lambda_j^i \quad \text{for each eigenvalue } \lambda_j.$$

ex/ For the same matrix, compute A^5 . Write $A^5 = \beta_1 A + \beta_0 I$ with $\lambda_1 = -1$ and $\lambda_2 = -2$. The two equations for β_0, β_1 are

$$\begin{aligned} \lambda_1^5 &= \beta_1 \lambda_1 + \beta_0 \\ \lambda_2^5 &= \beta_1 \lambda_2 + \beta_0 \end{aligned} \Leftrightarrow \begin{bmatrix} 1 & \lambda_1 \\ 1 & \lambda_2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \lambda_1^5 \\ \lambda_2^5 \end{bmatrix}.$$

Plugging in the numbers and solving,

$$\begin{bmatrix} 1 & -1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} -1 \\ -32 \end{bmatrix} \Rightarrow \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = - \begin{bmatrix} -2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -32 \end{bmatrix} = \begin{bmatrix} 30 \\ 31 \end{bmatrix}.$$

Hence

$$A^5 = 31A + 30I = \begin{bmatrix} 0 & 31 \\ -62 & -93 \end{bmatrix} + \begin{bmatrix} 30 & 0 \\ 0 & 30 \end{bmatrix} = \begin{bmatrix} 30 & 31 \\ -62 & -63 \end{bmatrix}.$$

Source page 64

ex/ A repeated eigenvalue. Let $A = \begin{bmatrix} 0 & 1 \\ -1 & -2 \end{bmatrix}$, which has $\lambda_1 = \lambda_2 = -1$. Find A^{10} . Write $A^{10} = \beta_1 A + \beta_0 I$. The eigenvalue equation gives one condition:

$$\lambda_1^{10} = \beta_1 \lambda_1 + \beta_0 \Rightarrow 1 = -\beta_1 + \beta_0.$$

Since $\lambda_2 = \lambda_1$ this gives no second independent equation for β_1 and β_0 . But λ_1 is a double root of the characteristic polynomial, so differentiating both sides of $\lambda^{10} = \beta_1 \lambda + \beta_0$ with respect to λ also holds at λ_1 , and that gives the missing equation:

$$\left. \frac{d}{d\lambda} \lambda^{10} \right|_{\lambda=\lambda_1} = \left. \frac{d}{d\lambda} (\beta_1 \lambda + \beta_0) \right|_{\lambda=\lambda_1} \Rightarrow 10\lambda_1^9 = \beta_1 \Rightarrow \beta_1 = -10.$$

Together,

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ -10 \end{bmatrix} \Rightarrow \beta_1 = -10, \quad \beta_0 = -9.$$

Hence

$$A^{10} = -10A - 9I = \begin{bmatrix} 0 & -10 \\ 10 & 20 \end{bmatrix} - \begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} -9 & -10 \\ 10 & 11 \end{bmatrix}.$$

The same idea computes the matrix exponential e^{At} , which is important for differential equations later on.

ex/ Find e^{At} for $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$. Since A is 2×2 with $\lambda_1 = \lambda_2 = 1$, write $e^{At} = \beta_1 A + \beta_0 I$. The eigenvalue condition and its derivative with respect to λ at $\lambda = 1$ give

$$e^{\lambda_1 t} = \beta_1 \lambda_1 + \beta_0 \Rightarrow e^t = \beta_1 + \beta_0, \quad \left. \frac{d}{d\lambda} e^{\lambda t} \right|_{\lambda=1} = \left. \frac{d}{d\lambda} (\beta_1 \lambda + \beta_0) \right|_{\lambda=1} \Rightarrow te^t = \beta_1.$$

So $\beta_1 = te^t$ and $\beta_0 = (1 - t)e^t$, and

$$e^{At} = te^t \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} + (1 - t)e^t \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} e^t & -te^t \\ 0 & e^t \end{bmatrix}.$$

Alternatively, if A is diagonalizable, any polynomial in A can be computed from the eigenvalues. Let $f(A) = \sum_{r=0}^p \beta_r A^r$ and $Q^{-1}AQ = \Lambda$, where Λ has the eigenvalues on its diagonal. Then

$$Q^{-1}f(A)Q = \sum_{r=0}^p \beta_r Q^{-1}A^r Q.$$

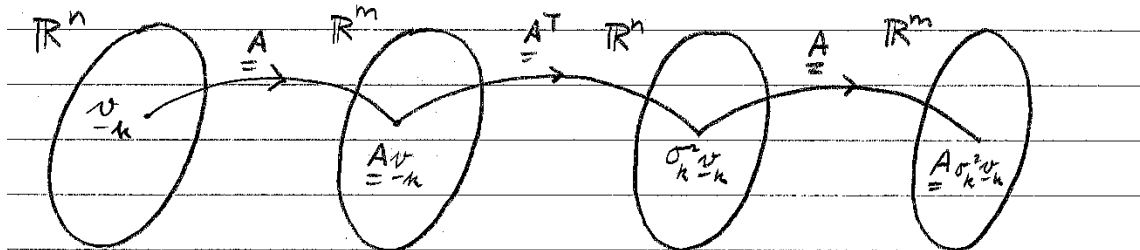
Extra note: SVD Lecture Notes — page 1

Supplementary note, rewritten from the separately supplied sheet.

Singular Value Decomposition (SVD)

For a real $m \times n$ matrix A , the product $A^T A$ is symmetric and positive semi-definite, so it can be diagonalized by an orthogonal matrix:

$$A^T A = V \Lambda V^T \quad \text{where } V^T V = I_n, \quad \Lambda = \text{diag}(\sigma_1^2, \dots, \sigma_n^2), \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0.$$



Now look at what AA^T does to the vector Av_k :

$$(AA^T)Av_k = A(A^T Av_k) = \sigma_k^2 Av_k.$$

Thus Av_k is an eigenvector of the symmetric positive semi-definite matrix AA^T with eigenvalue σ_k^2 , provided $Av_k \neq \mathbf{0}$. Write $Av_k = \alpha_k u_k$ with $\|u_k\| = 1$. Taking norms,

$$\|Av_k\|^2 = |\alpha_k|^2 \Rightarrow v_k^T \underbrace{A^T Av_k}_{\sigma_k^2 v_k} = |\alpha_k|^2 \Rightarrow \alpha_k = \sigma_k \text{ is a solution.}$$

Remark: The product $A^T Av_k$ is replaced by $\sigma_k^2 v_k$, and then $v_k^T v_k = 1$.

So $Av_k = \sigma_k u_k$ for every k . Putting all n of these equations side by side,

$$A[v_1 \ \dots \ v_n] = [u_1 \sigma_1 \ u_2 \sigma_2 \ \dots \ u_n \sigma_n] = [u_1 \ \dots \ u_n] \underbrace{\Sigma}_{\text{diag}(\sigma_1, \dots, \sigma_n)} \\ \Rightarrow AV = U\Sigma \Rightarrow A = U\Sigma V^T.$$

If $\sigma_k = 0$, then $Av_k = \mathbf{0}$ and the formula $u_k = Av_k/\sigma_k$ does not define u_k . Such columns can be completed with any orthonormal vectors, or simply dropped, as in the final form below.

For the non-zero singular values the vectors u_k are automatically orthonormal. Indeed, for $\sigma_k > 0$ and $\sigma_i > 0$,

$$u_k^T u_i = \frac{1}{\sigma_k} v_k^T A^T A \frac{1}{\sigma_i} v_i = \frac{1}{\sigma_k \sigma_i} v_k^T \underbrace{(A^T A) v_i}_{\sigma_i^2 v_i} \\ = \frac{\sigma_i}{\sigma_k} v_k^T v_i = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{if } k \neq i. \end{cases}$$

Final form. Keeping only the r non-zero singular values, the matrix is a sum of r rank-one pieces:

The SVD in its final form

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T, \quad \text{where } r \text{ is the number of non-zero singular values } \sigma_i.$$

$$A = U_r \Sigma_r V_r^T, \quad U_r = [u_1 \ \dots \ u_r], \quad \Sigma_r = \text{diag}(\sigma_1, \dots, \sigma_r), \quad V_r = [v_1, \dots, v_r].$$

Extra note: SVD SUMMARY — page 1

Supplementary note, rewritten from the separately supplied sheet.

SVD (Singular Value Decomposition)

Definition. A real $m \times n$ matrix A of rank r can be written in the “reduced” form

Reduced SVD

$$A_{m \times n} = U \Sigma V^T, \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \sigma_3 & \vdots \\ 0 & \cdots & 0 & \sigma_r \end{bmatrix}_{r \times r}.$$

The numbers σ_i are the *singular values* of A , ordered as $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq \sigma_r > 0$. The columns of U and V are orthonormal bases of the column space and the row space:

$$U_{m \times r} = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r], \quad \text{where } \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\} \text{ is an ON basis for } \text{Col}(A),$$

$$V_{n \times r} = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_r], \quad \text{where } \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\} \text{ is an ON basis for } \text{Row}(A).$$

Scaling Effect of Σ

The matrix of singular values acts as a pure scaling: it stretches or shrinks each coordinate separately. To see this, start from $A\mathbf{x} = \mathbf{y}$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, and multiply by U^T :

$$\begin{aligned} A_{m \times n} \underbrace{\mathbf{x}}_{\in \mathbb{R}^n} &= \underbrace{\mathbf{y}}_{\in \mathbb{R}^m} \xrightarrow{\times U^T} U^T A \mathbf{x} = U^T \mathbf{y} \\ &\Rightarrow \underbrace{U^T U}_{I} \Sigma V^T \mathbf{x} = U^T \mathbf{y} \\ &\Rightarrow \Sigma \underbrace{V^T \mathbf{x}}_{\tilde{\mathbf{x}}} = \underbrace{U^T \mathbf{y}}_{\tilde{\mathbf{y}}} \Rightarrow \Sigma \tilde{\mathbf{x}} = \tilde{\mathbf{y}}. \end{aligned}$$

In the rotated coordinates $\tilde{\mathbf{x}} = V^T \mathbf{x}$ and $\tilde{\mathbf{y}} = U^T \mathbf{y}$ the map is diagonal:

Scaling effect, componentwise

$$\sigma_i \tilde{x}_i = \tilde{y}_i.$$

So the directions $\mathbf{v}_i$ in the input space are sent to the directions $\mathbf{u}_i$ in the output space, and a singular value $\sigma_i > 1$ stretches along that direction while $\sigma_i < 1$ shrinks it.

The Matrix as a Sum of Rank-One Pieces

Multiplying out the three factors,

$$A = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r] \begin{bmatrix} \sigma_1 & \cdots & 0 \\ & \sigma_2 & \\ 0 & \cdots & \sigma_r \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Each term $\sigma_i \mathbf{u}_i \mathbf{v}_i^T$ is a rank-one matrix, and together they add up to a matrix of rank r . Keeping only the terms with the largest σ_i gives the best low-rank approximation of A .

Extra note: SVD SUMMARY — page 2

Supplementary note, rewritten from the separately supplied sheet.

Derivation of the Singular Vectors

Let $W = A^T A$. Then $W \in \mathbb{R}^{n \times n}$ and $W = W^T$.

1) The vectors $\mathbf{v}_i$ (row space)

Start from $A = U\Sigma V^T$, which gives $A^T = V\Sigma^T U^T$. Multiplying,

$$\begin{aligned} A &= U\Sigma V^T \\ A^T &= V\Sigma^T U^T \end{aligned} \Rightarrow A^T A = V \underbrace{\Sigma^T U^T U}_{=I} \Sigma V^T.$$

Remark: The middle product is I because the $\mathbf{u}_i$'s are orthonormal.

$$A^T A = V\Sigma^2 V^T.$$

Multiplying on the right by V and using $V^T V = I$ (the $\mathbf{v}_i$'s are orthonormal),

$$\Rightarrow A^T A V = V \Sigma^2 \underbrace{V^T V}_I = V \Sigma^2.$$

Reading this column by column,

$$(A^T A) \mathbf{v}_i = \mathbf{v}_i \lambda_i, \quad \lambda_i = \sigma_i^2.$$

So the eigenvalues of $A^T A$ are the squares of the singular values, and the columns $\mathbf{v}_i$ of V are the corresponding eigenvectors of $A^T A$. We order them by eigenvalue: if $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n$, then $\mathbf{v}_1$ is the eigenvector for λ_1 , and so on.

2) The vectors $\mathbf{u}_i$ (column space)

From $A = U\Sigma V^T$ we get $AV = U\Sigma$, and column by column $A\mathbf{v}_i = \mathbf{u}_i \sigma_i$. Hence

The left singular vectors

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sigma_i},$$

and these form an orthonormal basis for $\text{Col}(A)$. Each $\mathbf{u}_i$ must be listed in the same order as its partner $\mathbf{v}_i$, so that the pairs $(\mathbf{u}_i, \mathbf{v}_i)$ share the singular value σ_i .

Source page 65

Because $Q^{-1}A^rQ = (Q^{-1}AQ)^r = \Lambda^r$, the change of coordinates carries every power of A to the same power of Λ . So

$$\Rightarrow Q^{-1}f(A)Q = \sum_{r=0}^p \beta_r \Lambda^r, \quad \text{since } \underbrace{(Q^{-1}AQ)^r}_{\Lambda^r} = Q^{-1}A^rQ.$$

The right-hand side is diagonal, and its entries are the polynomial evaluated at each eigenvalue:

$$\Rightarrow Q^{-1}f(A)Q = \begin{bmatrix} \sum_{r=0}^p \beta_r \lambda_1^r & & \\ & \ddots & \\ & & \sum_{r=0}^p \beta_r \lambda_n^r \end{bmatrix} = \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix}.$$

Solving for $f(A)$ gives a recipe for applying a function to a diagonalizable matrix: apply it to the eigenvalues, then change coordinates back.

A function of a matrix

$$f(A) = Q \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} Q^{-1}.$$

ex/ Calculate e^{At} for $A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$, which has $\lambda_1 = -1$ and $\lambda_2 = -2$. This time we also need the eigenvectors. Solving $(A - \lambda_i I)\mathbf{e}_i = \mathbf{0}$ gives

$$\begin{aligned} (A - \lambda_1 I)\mathbf{e}_1 = \mathbf{0} &\Rightarrow \mathbf{e}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ (A - \lambda_2 I)\mathbf{e}_2 = \mathbf{0} &\Rightarrow \mathbf{e}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \end{aligned} \quad \left\{ \begin{aligned} Q = [\mathbf{e}_1 \ \mathbf{e}_2] &= \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \\ Q^{-1} &= \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} \end{aligned} \right.$$

Applying $f(\lambda) = e^{\lambda t}$ to the eigenvalues, we get

$$e^{At} = Q \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{-2t} \end{bmatrix} Q^{-1} = \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-t} + 2e^{-2t} \end{bmatrix}.$$

Quadratic Forms

A *quadratic form* is a polynomial in which every term has degree two. In two variables it looks like

$$f(x_1, x_2) = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2.$$

With n variables, a quadratic form can always be written in matrix form, and the matrix A can be chosen to be symmetric:

Quadratic form

$$f(x_1, \dots, x_n) = \mathbf{x}^T A \mathbf{x}.$$

Def'n A quadratic form is called a *canonical* quadratic form if it has no cross terms, that is, if its matrix is diagonal:

$$f(x_1, \dots, x_n) = \mathbf{x}^T D \mathbf{x}, \quad D \text{ is diagonal.}$$

Source page 66

Theorem 11.6.1

Every quadratic form can be brought to canonical form by a change of variables.

Theorem 11.6.1

Every quadratic form can be transformed into a canonical form.

proof Let $f(x_1, \dots, x_n) = \mathbf{x}^T A \mathbf{x}$ with A symmetric. A symmetric matrix can be diagonalized by an orthogonal matrix Q whose columns are unit-norm eigenvectors:

$$Q^T A Q = \Lambda, \quad \text{so} \quad A = Q \Lambda Q^T.$$

Then, introducing the new variables $\tilde{\mathbf{x}} = Q^T \mathbf{x}$ (equivalently $\mathbf{x} = Q \tilde{\mathbf{x}}$),

$$\begin{aligned} \underbrace{f(x_1, \dots, x_n)}_{f(\mathbf{x})} &= \mathbf{x}^T Q \Lambda Q^T \mathbf{x} \\ &= (Q^T \mathbf{x})^T \Lambda (Q^T \mathbf{x}) \\ &= \tilde{\mathbf{x}}^T \Lambda \tilde{\mathbf{x}}. \end{aligned}$$

Thus $f(Q\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}^T \Lambda \tilde{\mathbf{x}}$, which has no cross terms, so the form is in canonical form in the new variables. ■

Def'n A quadratic form $\mathbf{x}^T A \mathbf{x}$ is *positive definite* if $\mathbf{x}^T A \mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, and *negative definite* if $\mathbf{x}^T A \mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$. A symmetric matrix A is called positive (negative) definite if its quadratic form is.

Theorem 11.6.2

The eigenvalues tell us the sign of the form.

Theorem 11.6.2

For a real symmetric A , A is positive (negative) definite if and only if every eigenvalue of A is positive (negative).

proof Let Q have unit-norm eigenvectors as its columns, so that $A = Q \Lambda Q^T$. Then, with $\tilde{\mathbf{x}} = Q^T \mathbf{x}$,

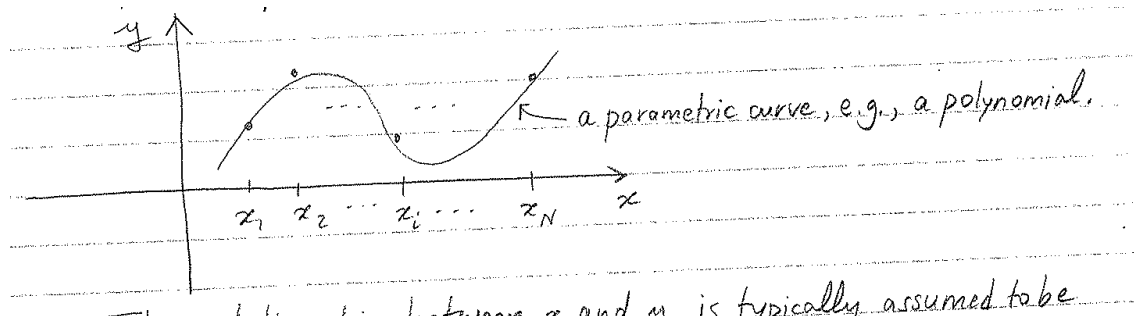
$$\begin{aligned} \mathbf{x}^T A \mathbf{x} &= \mathbf{x}^T (Q \Lambda Q^T) \mathbf{x} \\ &= (Q^T \mathbf{x})^T \Lambda (Q^T \mathbf{x}) \\ &= \tilde{\mathbf{x}}^T \Lambda \tilde{\mathbf{x}} \\ &= \sum_{i=1}^n \lambda_i \tilde{x}_i^2 \quad \begin{cases} > 0 & \text{if all } \lambda_i > 0 \\ < 0 & \text{if all } \lambda_i < 0. \end{cases} \end{aligned}$$

Here $\tilde{\mathbf{x}} \neq \mathbf{0}$ whenever $\mathbf{x} \neq \mathbf{0}$, so at least one term is non-zero and the sign of the sum is the common sign of the eigenvalues. Conversely, if A is positive definite and $(\lambda, \mathbf{x})$ is an eigenpair with $\mathbf{x}^T \mathbf{x} = 1$, then $\lambda = \lambda \mathbf{x}^T \mathbf{x} = \mathbf{x}^T A \mathbf{x} > 0$; the negative definite case is the same. ■

Source page 67

Curve Fitting

In many applications we are given a set of measurements (x_i, y_i) , $1 \leq i \leq N$, where x_i may be the time or the position of the i^{th} measurement.



The relationship between x and y is usually assumed to be a parametric function, with a vector θ of parameters:

$$y = f(x; \underbrace{\theta}_{\text{the parameters, in vector form}}).$$

We then choose the parameters that fit the measurements best. A common choice is the parameters that minimize the sum of the squared errors:

$$\sum_{i=1}^N (y_i - f(x_i; \theta))^2, \quad \text{or} \quad \theta^* = \arg \min_{\theta} \sum_{i=1}^N (y_i - f(x_i; \theta))^2.$$

In general this is solved with numerical optimization tools, such as those available in MATLAB. But in some cases it reduces to linear algebra. This happens when f is a polynomial in x :

$$f(x; \theta) = \theta_n x^n + \theta_{n-1} x^{n-1} + \cdots + \theta_1 x + \theta_0.$$

Let e_i be the error at the point x_i :

$$e_i = y_i - (\theta_n x_i^n + \theta_{n-1} x_i^{n-1} + \cdots + \theta_1 x_i + \theta_0).$$

Collecting all the errors in a vector gives a matrix equation, where X has one row $[1, x_i, x_i^2, \dots, x_i^n]$ per measurement:

$$\underbrace{\begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}}_e = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_y - \underbrace{\begin{bmatrix} 1 & x_1 & \cdots & x_1^n \\ \vdots & & & \vdots \\ 1 & x_N & \cdots & x_N^n \end{bmatrix}}_X \underbrace{\begin{bmatrix} \theta_0 \\ \vdots \\ \theta_n \end{bmatrix}}_{\theta}.$$

Source page 68

So the optimization problem is to choose $\boldsymbol{\theta}$ to make the error vector as short as possible:

$$\min_{\boldsymbol{\theta}} \|\mathbf{e}\|^2 = \min_{\boldsymbol{\theta}} \|\mathbf{y} - X\boldsymbol{\theta}\|^2.$$

This is a best-approximation problem like the one we solved earlier: we look for the vector $X\boldsymbol{\theta}$ in the column space of X that is closest to $\mathbf{y}$. The answer is characterized by orthogonality: at the optimal $\boldsymbol{\theta}$ the error $\mathbf{e}$ must be orthogonal to the column space of X , that is, $X^T \mathbf{e} = \mathbf{0}$. Substituting $\mathbf{e} = \mathbf{y} - X\boldsymbol{\theta}$:

The normal equations

$$X^T \mathbf{e} = \mathbf{0} \Rightarrow X^T(\mathbf{y} - X\boldsymbol{\theta}) = \mathbf{0} \Rightarrow X^T X \boldsymbol{\theta} = X^T \mathbf{y}.$$

The normal equations are always consistent, because the closest point to $\mathbf{y}$ in the column space of X always exists (it is the orthogonal projection of $\mathbf{y}$), and it has the form $X\boldsymbol{\theta}$ for some $\boldsymbol{\theta}$. So they can be solved for the optimal parameters.

ex/ Fit a second-order polynomial to the data $\mathbf{y} = \begin{bmatrix} 1 \\ -1 \\ 3 \\ 2 \end{bmatrix}$ measured at $x_1 = 0, x_2 = 1, x_3 = 2, x_4 = 3$. There are three parameters $\theta_0, \theta_1, \theta_2$, and the normal equations are

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}}_{X^T} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}}_X \underbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix}}_{\boldsymbol{\theta}} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}}_{X^T} \underbrace{\begin{bmatrix} 1 \\ -1 \\ 3 \\ 2 \end{bmatrix}}_{\mathbf{y}}$$

Multiplying out,

$$\begin{bmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \\ 29 \end{bmatrix}.$$

Solving this system gives the best-fit parameters

$$\boldsymbol{\theta} = \begin{bmatrix} 9/20 \\ -1/20 \\ 1/4 \end{bmatrix}.$$

Source page 69

Linear Constant Coefficient Difference Equations

A difference equation defines a sequence by a rule that gives each term from the previous ones. A linear constant-coefficient difference equation of order k has the form

$$x[n] = \underbrace{a_1 x[n-1] + \cdots + a_k x[n-k]}_{k^{\text{th}}\text{-order difference equation}}$$

Such equations appear very often in discrete-time systems and in discretized models of continuous-time systems. A good way to study them is to turn the single order- k equation into a first-order equation for a vector. Stack the k most recent values into a vector $\mathbf{x}_n$. Then the next vector is obtained from the current one by a fixed matrix A , the *companion matrix*, whose first row holds the coefficients and whose other rows just shift the older values down:

$$\mathbf{x}_n = \begin{bmatrix} x[n] \\ \vdots \\ x[n-k+1] \end{bmatrix}, \quad A = \begin{bmatrix} a_1 & \cdots & a_k \\ 1 & & 0 \\ & \ddots & \vdots \\ & & 1 & 0 \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_k \\ I_{k-1} & \mathbf{0} \end{bmatrix}.$$

Vector form

$$\mathbf{x}_n = A\mathbf{x}_{n-1}.$$

Applying this rule again and again, starting from the initial conditions $\mathbf{x}_0$, gives

$$\mathbf{x}_n = A(A(\cdots(A\mathbf{x}_0)\cdots)) = A^n \mathbf{x}_0, \quad \mathbf{x}_0 = \underbrace{\begin{bmatrix} x[0] \\ \vdots \\ x[-k+1] \end{bmatrix}}_{\text{initial conditions}}.$$

So, in principle, $\mathbf{x}_n$ for any n can be found by multiplying $\mathbf{x}_0$ by A a total of n times. This is not efficient, and it does not give a closed-form answer. Eigenvectors give more insight.

Theorem DE.1**Theorem DE.1**

If $(\lambda, \mathbf{x}_0)$ is an eigenpair of A , then the solution of $\mathbf{x}_n = A\mathbf{x}_{n-1}$ that starts at $\mathbf{x}_0$ is $\mathbf{x}_n = \lambda^n \mathbf{x}_0$.

proof $\mathbf{x}_n = A^n \mathbf{x}_0 = \lambda^n \mathbf{x}_0$, since each multiplication by A just multiplies the eigenvector by λ .

Source page 70

Theorem DE.2

If the starting vector is not an eigenvector, we can still handle it when A has enough eigenvectors: write $\mathbf{x}_0$ as a combination of eigenvectors and let each one evolve by itself.

Theorem DE.2

If A has a full set of eigenvectors $\mathbf{v}_i$ that span $\mathbb{R}^k$, then $\mathbf{x}_n = A^n \mathbf{x}_0$ is given by

$$\mathbf{x}_n = \sum_{i=1}^k \alpha_i \lambda_i^n \mathbf{v}_i \quad \text{where} \quad \mathbf{x}_0 = \sum_{i=1}^k \alpha_i \mathbf{v}_i.$$

proof Multiply the expansion of $\mathbf{x}_0$ by A^n , and use $A^n \mathbf{v}_i = \lambda_i^n \mathbf{v}_i$:

$$A^n \mathbf{x}_0 = A^n \left(\sum_{i=1}^k \alpha_i \mathbf{v}_i \right) = \sum_{i=1}^k \alpha_i A^n \mathbf{v}_i = \underbrace{\sum_{i=1}^k \alpha_i \lambda_i^n \mathbf{v}_i}_{\mathbf{x}_n}. \quad \blacksquare$$

Example Solve $x[n] = 2x[n-1] + x[n-2]$ with initial conditions $x[0] = 3$ and $x[-1] = 1$. Here $k = 2$ and the companion matrix has eigenvalues $1 \pm \sqrt{2}$:

$$\mathbf{x}_n = \begin{bmatrix} x[n] \\ x[n-1] \end{bmatrix}, \quad A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \Rightarrow \underbrace{\lambda_1 = 1 + \sqrt{2}; \quad \lambda_2 = 1 - \sqrt{2}}_{\text{eigenvalues}}.$$

We find the eigenvectors by solving $(A - \lambda I)\mathbf{v} = \mathbf{0}$. For λ_1 :

$$(A - (1 + \sqrt{2})I)\mathbf{v} = \mathbf{0} \Rightarrow \begin{bmatrix} 1 - \sqrt{2} & 1 \\ 1 & -1 - \sqrt{2} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{v}_1 = \begin{bmatrix} 1 + \sqrt{2} \\ 1 \end{bmatrix}.$$

For λ_2 :

$$(A - (1 - \sqrt{2})I)\mathbf{v} = \mathbf{0} \Rightarrow \begin{bmatrix} 1 + \sqrt{2} & 1 \\ 1 & -1 + \sqrt{2} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{v}_2 = \begin{bmatrix} -1 + \sqrt{2} \\ -1 \end{bmatrix}.$$

Next we express the initial vector $\mathbf{x}_0 = \begin{bmatrix} x[0] \\ x[-1] \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ in the eigenvectors:

$$\begin{aligned} \mathbf{x}_0 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} &= \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 = \begin{bmatrix} 1 + \sqrt{2} & -1 + \sqrt{2} \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \\ \Rightarrow \alpha_2 &= \frac{2 - \sqrt{2}}{2\sqrt{2}} = \frac{\sqrt{2} - 1}{2} = -\frac{1 - \sqrt{2}}{2}, \quad \alpha_1 = 1 + \alpha_2 = \frac{\sqrt{2} + 1}{2}. \end{aligned}$$

By Theorem DE.2,

$$\begin{aligned} \mathbf{x}_n &= \alpha_1 \lambda_1^n \mathbf{v}_1 + \alpha_2 \lambda_2^n \mathbf{v}_2 \\ &= \frac{\sqrt{2} + 1}{2} (1 + \sqrt{2})^n \begin{bmatrix} 1 + \sqrt{2} \\ 1 \end{bmatrix} - \frac{1 - \sqrt{2}}{2} (1 - \sqrt{2})^n \begin{bmatrix} \sqrt{2} - 1 \\ -1 \end{bmatrix} \\ \Rightarrow \mathbf{x}_n &= \begin{bmatrix} \frac{(1 + \sqrt{2})^{n+2}}{2} + \frac{(1 - \sqrt{2})^{n+2}}{2} \\ \frac{(1 + \sqrt{2})^{n+1}}{2} + \frac{(1 - \sqrt{2})^{n+1}}{2} \end{bmatrix}. \end{aligned}$$

The first component of $\mathbf{x}_n$ is $x[n]$:

Closed form

$$x[n] = \frac{1}{2} \left[(1 + \sqrt{2})^{n+2} + (1 - \sqrt{2})^{n+2} \right], \quad n \geq 0.$$

As a check, this gives $x[0] = 3$, and the recursion gives $x[1] = 2 \cdot 3 + 1 = 7$, which the formula also gives.

Source page 71

There is a quicker way to the same answer, based on one observation. Given the difference equation

$$x[n] = a_1x[n-1] + a_2x[n-2] + \cdots + a_kx[n-k],$$

try a solution of the form $x[n] = r^n$. It works exactly when

$$r^n = a_1r^{n-1} + a_2r^{n-2} + \cdots + a_kr^{n-k}$$

or, moving everything to one side,

$$r^n - a_1r^{n-1} - a_2r^{n-2} - \cdots - a_kr^{n-k} = 0 \quad \forall n.$$

Factoring out r^{n-k} leaves a polynomial in r :

$$\Rightarrow r^{n-k} \underbrace{(r^k - a_1r^{k-1} - a_2r^{k-2} - \cdots - a_k)}_{\text{characteristic equation of the difference equation}} = 0.$$

So r^n is a solution exactly when r is a root of the *characteristic equation* of the difference equation. This is the same polynomial (up to a sign) as $\det(A - rI)$ for the companion matrix A above, so the roots r are exactly the eigenvalues of

$$A = \begin{bmatrix} a_1 & \cdots & a_k \\ & & 0 \\ & I_{(k-1)} & \vdots \\ & & 0 \end{bmatrix}.$$

There are at most k distinct roots. If they are all distinct, a general solution is any combination of the corresponding powers:

General solution, distinct roots

$$x[n] = \alpha_1 r_1^n + \cdots + \alpha_k r_k^n = \sum_{i=1}^k \alpha_i r_i^n, \quad n \geq 0.$$

The constants α_i are fixed by the initial conditions, given either as $x[0], \dots, x[-k+1]$ or as $x[-1], \dots, x[-k]$. For the second form, evaluating the formula at $n = -1, \dots, -k$ gives a linear system for the α_i with an invertible coefficient matrix:

$$\underbrace{\begin{bmatrix} r_1^{-1} & r_2^{-1} & \cdots & r_k^{-1} \\ r_1^{-2} & r_2^{-2} & \cdots & r_k^{-2} \\ \vdots & & & \vdots \\ r_1^{-k} & r_2^{-k} & \cdots & r_k^{-k} \end{bmatrix}}_{\text{invertible}} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{bmatrix} = \begin{bmatrix} x[-1] \\ \vdots \\ x[-k] \end{bmatrix}.$$

ex/ Apply this approach to the previous example.

Source page 72

The difference equation is $x[n] = 2x[n-1] + x[n-2]$, with characteristic equation $r^2 - 2r - 1 = 0$ and roots $r_1 = 1 + \sqrt{2}$ and $r_2 = 1 - \sqrt{2}$. So $x[n] = \alpha_1 r_1^n + \alpha_2 r_2^n$. The initial conditions $x[0] = 3$ and $x[-1] = 1$ give two equations:

$$\begin{aligned} 3 &= \alpha_1 + \alpha_2 \\ 1 &= \underbrace{\frac{\alpha_1}{r_1} + \frac{\alpha_2}{r_2}}_{r_1 r_2 = r_2 \alpha_1 + r_1 \alpha_2} \Rightarrow \begin{aligned} \alpha_1 &= \frac{r_1 r_2 - 3r_1}{r_2 - r_1} = \frac{4 + 3\sqrt{2}}{2\sqrt{2}} \\ \alpha_2 &= 3 - \alpha_1 = \frac{3\sqrt{2} - 4}{2\sqrt{2}}. \end{aligned} \end{aligned}$$

Hence

$$x[n] = \left(\frac{4 + 3\sqrt{2}}{2\sqrt{2}} \right) (1 + \sqrt{2})^n + \left(\frac{3\sqrt{2} - 4}{2\sqrt{2}} \right) (1 - \sqrt{2})^n, \quad n \geq 0.$$

This looks different from the first answer, but it is the same sequence: both give $x[0] = 3$, $x[1] = 7$, $x[2] = 17$, $x[3] = 41, \dots$

ex/ The second approach also handles repeated roots, with one change. Consider

$$x[n] = 2x[n-1] - x[n-2], \quad x[-1] = 3, \quad x[-2] = 4.$$

The characteristic equation is $r^2 - 2r + 1 = 0$, so $r_1 = r_2 = 1$ is a double root. When a root is repeated, the solution takes the form

$$x[n] = \alpha_1 r_1^n + \alpha_2 n r_1^n.$$

(A proof for the similar situation of differential equations is given later.) The initial conditions give

$$\begin{aligned} x[-1] = 3 &= \alpha_1 r_1^{-1} - \alpha_2 r_1^{-1} = (\alpha_1 - \alpha_2) r_1^{-1} \\ x[-2] = 4 &= \alpha_1 r_1^{-2} - 2\alpha_2 r_1^{-2} = (\alpha_1 - 2\alpha_2) r_1^{-2}. \end{aligned}$$

With $r_1 = 1$ these are two simple equations:

$$\Rightarrow \left. \begin{aligned} \alpha_1 - \alpha_2 &= 3r_1 = 3 \\ \alpha_1 - 2\alpha_2 &= 4r_1^2 = 4 \end{aligned} \right\} \alpha_2 = -1, \quad \alpha_1 = 2.$$

Therefore

$$\Rightarrow x[n] = 2 - n, \quad n \geq 0.$$

Source page 73

The Z-Transform

The z-transform is a tool for analyzing and designing discrete-time systems. It is also very useful for solving difference equations, because it turns the difference equation into an algebraic equation.

Def'n Given a sequence $x[n]$, its *z-transform* $\bar{X}(z)$ is the sum

The z-transform

$$\bar{X}(z) = \sum_{n=0}^{\infty} x[n]z^{-n},$$

defined for those values of z for which the sum converges. These values of z form the *region of convergence*, or ROC.

Two useful facts about convergence:

- (i) For a finite-length sequence, the z-transform converges for every z , except possibly $z = 0$.
- (ii) If the z-transform of a sequence converges at all, it converges for every $|z| > r_0$, where r_0 depends on the sequence.

Properties:

(i) *Linearity*. Suppose $x_1[n] \xleftrightarrow{Z} \bar{X}_1(z)$ with region of convergence ROC_1 , and $x_2[n] \xleftrightarrow{Z} \bar{X}_2(z)$ with region of convergence ROC_2 . Then, for any constants α_1 and α_2 ,

$$\alpha_1 x_1[n] + \alpha_2 x_2[n] \longleftrightarrow \alpha_1 \bar{X}_1(z) + \alpha_2 \bar{X}_2(z),$$

and the region of convergence contains the overlap, $\text{ROC} \supset \text{ROC}_1 \cap \text{ROC}_2$.

proof Split the sum into two sums:

$$\begin{aligned} \sum_{n=0}^{\infty} (\alpha_1 x_1[n] + \alpha_2 x_2[n]) z^{-n} &= \sum_{n=0}^{\infty} (\alpha_1 x_1[n] z^{-n} + \alpha_2 x_2[n] z^{-n}) \\ &= \alpha_1 \underbrace{\sum_{n=0}^{\infty} x_1[n] z^{-n}}_{\bar{X}_1(z)} + \alpha_2 \underbrace{\sum_{n=0}^{\infty} x_2[n] z^{-n}}_{\bar{X}_2(z)} \\ &= \alpha_1 \bar{X}_1(z) + \alpha_2 \bar{X}_2(z). \quad \blacksquare \end{aligned}$$

Source page 74

(ii) *Delay by one step.* If $x[n] \xleftrightarrow{\mathcal{Z}} \overline{X}(z)$, then

$$x[n-1] \xleftrightarrow{\mathcal{Z}} z^{-1}\overline{X}(z) + x[-1].$$

The extra term $x[-1]$ is how the initial condition enters. proof Substitute $n' = n - 1$ and pull one factor z^{-1} out of the sum:

$$\begin{aligned} \sum_{n=0}^{\infty} x[\underbrace{n-1}_{n'}]z^{-n} &= \sum_{n'=-1}^{\infty} x[n']z^{-(n'+1)} \\ &= \sum_{n'=0}^{\infty} x[n']z^{-1}z^{-n'} + x[-1] \\ &= z^{-1} \underbrace{\sum_{n'=0}^{\infty} x[n']z^{-n'}}_{\overline{X}(z)} + x[-1] \\ &= z^{-1}\overline{X}(z) + x[-1]. \quad \blacksquare \end{aligned}$$

(iii) *Delay by k steps.* If $x[n] \xleftrightarrow{\mathcal{Z}} \overline{X}(z)$, then for $k \geq 1$

$$x[n-k] \xleftrightarrow{\mathcal{Z}} z^{-k}\overline{X}(z) + z^{-k+1}x[-1] + z^{-k+2}x[-2] + \cdots + x[-k].$$

proof The same substitution $n' = n - k$ gives

$$\begin{aligned} \sum_{n=0}^{\infty} x[\underbrace{n-k}_{n'}]z^{-n} &= \sum_{n'=-k}^{\infty} x[n']z^{-(n'+k)} \\ &= z^{-k} \sum_{n'=0}^{\infty} x[n']z^{-n'} + z^{-k}(x[-1]z + x[-2]z^2 + \cdots + x[-k]z^k) \\ &= z^{-k}\overline{X}(z) + z^{-k+1}x[-1] + \cdots + x[-k]. \end{aligned}$$

(iv) *The exponential sequence.* For $x[n] = a^n$, $n \geq 0$, the sum is a geometric series, which converges when $|az^{-1}| < 1$:

$$\overline{X}(z) = \sum_{n=0}^{\infty} a^n z^{-n} = \sum_{n=0}^{\infty} (az^{-1})^n = \frac{1}{1 - az^{-1}}, \quad |az^{-1}| < 1 \Rightarrow \underbrace{|a| < |z|}_{\text{ROC}}.$$

(v) *The sequence na^n .* For $x[n] = na^n$, $n \geq 0$, differentiate the geometric series with respect to z :

$$\begin{aligned} \overline{X}(z) &= \sum_{n=0}^{\infty} na^n z^{-n} = \left[\sum_{n=0}^{\infty} na^n z^{-(n+1)} \right] z = - \left(\frac{d}{dz} \sum_{n=0}^{\infty} a^n z^{-n} \right) z \\ &= - \left[\frac{d}{dz} \frac{1}{1 - az^{-1}} \right] z \\ \overline{X}(z) &= \frac{az^{-1}}{(1 - az^{-1})^2}. \quad \blacksquare \end{aligned}$$

Source page 75

We can now solve difference equations by transforming them. The method is: apply the z-transform to both sides (the delay rules bring in the initial conditions), solve for $\bar{X}(z)$ algebraically, expand it in partial fractions, and invert each term.

ex/ Solve $x[n] = \frac{5}{6}x[n-1] - \frac{1}{6}x[n-2]$ with $x[-1] = -4$ and $x[-2] = -14$.

Transform both sides to the z-domain, using properties (ii) and (iii):

$$\begin{aligned}\bar{X}(z) &= \frac{5}{6}(z^{-1}\bar{X}(z) + x[-1]) - \frac{1}{6}(z^{-2}\bar{X}(z) + z^{-1}x[-1] + x[-2]) \\ \bar{X}(z) \left(1 - \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2}\right) &= -\frac{20}{6} + \frac{4}{6}z^{-1} + \frac{14}{6} \\ &= \frac{4}{6}z^{-1} - 1.\end{aligned}$$

Solving for $\bar{X}(z)$ and factoring the denominator, we write it as a sum of simple fractions:

$$\begin{aligned}\bar{X}(z) &= \frac{\frac{4}{6}z^{-1} - 1}{1 - \frac{5}{6}z^{-1} + \frac{1}{6}z^{-2}} = \frac{\frac{4}{6}z^{-1} - 1}{(1 - \frac{1}{2}z^{-1})(1 - \frac{1}{3}z^{-1})} \\ &= \underbrace{\frac{A}{1 - \frac{1}{2}z^{-1}} + \frac{B}{1 - \frac{1}{3}z^{-1}}}_{\text{partial fraction expansion}}.\end{aligned}$$

To find A , multiply by $(1 - \frac{1}{2}z^{-1})$ and set $z = \frac{1}{2}$; to find B , multiply by $(1 - \frac{1}{3}z^{-1})$ and set $z = \frac{1}{3}$:

$$A = (1 - \frac{1}{2}z^{-1})\bar{X}(z)|_{z=1/2} = \frac{\frac{4}{6} \cdot 2 - 1}{1 - \frac{2}{3}} = \frac{1/3}{1/3} = 1$$

$$B = (1 - \frac{1}{3}z^{-1})\bar{X}(z)|_{z=1/3} = \frac{\frac{4}{6} \cdot 3 - 1}{1 - \frac{1}{2} \cdot 3} = \frac{1}{-1/2} = -2.$$

Each fraction is of the form in property (iv), so we can invert it:

$$\Rightarrow \bar{X}(z) = \frac{1}{1 - \frac{1}{2}z^{-1}} - \frac{2}{1 - \frac{1}{3}z^{-1}} \xrightarrow{\mathcal{Z}^{-1}} x[n] = \left(\frac{1}{2}\right)^n - 2\left(\frac{1}{3}\right)^n, \quad n \geq 0.$$

ex/ Solve $x[n] = x[n-1] - \frac{1}{4}x[n-2]$ with $x[-1] = 1$ and $x[-2] = -1$.

Transform both sides to the z-domain:

$$\begin{aligned}\bar{X}(z) &= (z^{-1}\bar{X}(z) + x[-1]) - \frac{1}{4}(z^{-2}\bar{X}(z) + z^{-1}x[-1] + x[-2]) \\ &= z^{-1}\bar{X}(z) + 1 - \frac{1}{4}z^{-2}\bar{X}(z) - \frac{1}{4}z^{-1} + \frac{1}{4}. \\ \Rightarrow \bar{X}(z) \left(1 - z^{-1} + \frac{1}{4}z^{-2}\right) &= \frac{5}{4} - \frac{1}{4}z^{-1}.\end{aligned}$$

This time the denominator is a perfect square, $(1 - \frac{1}{2}z^{-1})^2$, so the partial fraction expansion contains a repeated factor:

$$\Rightarrow \bar{X}(z) = \frac{\frac{5}{4} - \frac{1}{4}z^{-1}}{1 - z^{-1} + \frac{1}{4}z^{-2}} = \frac{\frac{5}{4} - \frac{1}{4}z^{-1}}{(1 - \frac{1}{2}z^{-1})^2} = \frac{A}{1 - \frac{1}{2}z^{-1}} + \frac{B}{(1 - \frac{1}{2}z^{-1})^2}.$$

Source page 76

Clearing denominators and comparing the constant terms and the coefficients of z^{-1} :

$$\Rightarrow \frac{5}{4} - \frac{1}{4}z^{-1} = A(1 - \frac{1}{2}z^{-1}) + B = A + B - \frac{A}{2}z^{-1}.$$

$$\begin{aligned} & A + B = 5/4 \\ \Rightarrow -A/2 = -1/4 & \Rightarrow A = 1/2 \Rightarrow B = 5/4 - A = 5/4 - 1/2 = 3/4. \end{aligned}$$

Hence

$$\overline{X}(z) = \frac{1}{2} \frac{1}{1 - \frac{1}{2}z^{-1}} + \frac{3}{4} \frac{1}{(1 - \frac{1}{2}z^{-1})^2}.$$

To invert these we use two transform pairs. The first is property (iv). The second follows by multiplying the geometric series $\sum_n w^n = \frac{1}{1-w}$ by itself with $w = az^{-1}$, since $(\sum_n w^n)^2 = \sum_n (n+1)w^n$:

$$\frac{1}{1 - az^{-1}} \xleftrightarrow{\mathcal{Z}^{-1}} a^n, \quad \frac{1}{(1 - az^{-1})^2} \xleftrightarrow{\mathcal{Z}^{-1}} (n+1)a^n, \quad n \geq 0.$$

Therefore

$$\overline{X}(z) \xleftrightarrow{\mathcal{Z}^{-1}} x[n] = \frac{1}{2} \left(\frac{1}{2}\right)^n + \frac{3}{4}(n+1) \left(\frac{1}{2}\right)^n, \quad n \geq 0,$$

which simplifies to

Solution

$$x[n] = \frac{5}{4} \left(\frac{1}{2}\right)^n + \frac{3}{4}n \left(\frac{1}{2}\right)^n, \quad n \geq 0.$$

Linear Constant Coefficient Differential Equations

(Chapter 3, sections 4 and 5 in the textbook.)

Differential equations describe continuous-time systems in the way that difference equations describe discrete-time ones. A homogeneous linear differential equation with constant coefficients of order n has the form

$$\underbrace{\frac{d^n y}{dx^n} + a_1 \frac{d^{n-1} y}{dx^{n-1}} + \cdots + a_{n-1} \frac{dy}{dx} + a_n y = 0}_{\text{homogeneous } n^{\text{th}}\text{-order linear constant coefficient DE}}$$

The simplest case is first order: $\frac{dy}{dx} + a_1 y = 0$. Try a solution of the form $y(x) = e^{\lambda x}$. Then $\frac{dy}{dx} = \lambda e^{\lambda x}$, so

$$\lambda e^{\lambda x} + a_1 e^{\lambda x} = 0 \Rightarrow \lambda = -a_1.$$

Hence the general solution is $Ce^{-a_1 x}$. To determine the constant C we need an initial condition such as $y(0)$:

$$y(0) = Ce^{-a_1 \cdot 0} = C \Rightarrow y(x) = y(0)e^{-a_1 x}.$$

Source page 77

ex/ A second-order equation. Consider

$$\frac{d^2y}{dx^2} + a_1 \frac{dy}{dx} + a_2 y = 0 \quad \text{which is also written } y'' + a_1 y' + a_2 y = 0.$$

We again try solutions of the form $y(x) = e^{\lambda x}$. Each derivative just brings down another factor λ :

$$\Rightarrow \lambda^2 e^{\lambda x} + a_1 \lambda e^{\lambda x} + a_2 e^{\lambda x} = 0$$

$$\Rightarrow (\lambda^2 + a_1 \lambda + a_2) e^{\lambda x} = 0 \Rightarrow \underbrace{\lambda^2 + a_1 \lambda + a_2 = 0}_{\text{characteristic equation}}.$$

Since $e^{\lambda x}$ is never zero, λ has to be a root of the characteristic equation:

$$\lambda_1 = \frac{-a_1 + \sqrt{a_1^2 - 4a_2}}{2}, \quad \lambda_2 = \frac{-a_1 - \sqrt{a_1^2 - 4a_2}}{2}.$$

For real a_1 and a_2 these are either two real roots or a pair of complex conjugate roots. The general solution is a combination of the two:

General solution, distinct roots

$$y(x) = C_1 e^{\lambda_1 x} + C_2 e^{\lambda_2 x}.$$

Given initial conditions $y(0)$ and $y'(0)$, the constants C_1 and C_2 are determined uniquely.

ex/ Real roots. Solve $y'' - 9y = 0$ with $y(0) = 4$ and $y'(0) = -5$. The characteristic equation $\lambda^2 - 9 = 0$ gives $\lambda_1 = 3$ and $\lambda_2 = -3$, so

$$y(x) = C_1 e^{3x} + C_2 e^{-3x}.$$

The initial conditions give two equations for C_1 and C_2 :

$$\left. \begin{array}{l} 4 = C_1 + C_2 \\ -5 = 3C_1 - 3C_2 \end{array} \right\} \Rightarrow 6C_1 = 7 \Rightarrow C_1 = \frac{7}{6}, \quad C_2 = 4 - C_1 = 4 - \frac{7}{6} = \frac{17}{6}.$$

$$\Rightarrow y(x) = \frac{7}{6} e^{3x} + \frac{17}{6} e^{-3x}.$$

ex/ Imaginary roots. Solve $y'' + 9y = 0$. The characteristic equation $\lambda^2 + 9 = 0$ gives $\lambda_1 = j3$ and $\lambda_2 = -j3$, so $y(x) = C_1 e^{j3x} + C_2 e^{-j3x}$. Complex exponentials are not convenient to read, so we rewrite the solution with sines and cosines, using Euler's formula in the form $\cos 3x = \frac{1}{2}(e^{j3x} + e^{-j3x})$ and $\sin 3x = \frac{1}{2j}(e^{j3x} - e^{-j3x})$:

$$\begin{aligned} \Rightarrow y(x) &= C_1 e^{j3x} + C_2 e^{-j3x} \\ &= \left(\frac{C_1 + C_2}{2} + \frac{C_1 - C_2}{2} \right) e^{j3x} + \left(\frac{C_2 + C_1}{2} + \frac{C_2 - C_1}{2} \right) e^{-j3x} \\ \Rightarrow y(x) &= \left(\frac{C_1 + C_2}{2} \right) (e^{j3x} + e^{-j3x}) + j \left(\frac{C_1 - C_2}{2j} \right) (e^{j3x} - e^{-j3x}) \\ &= (C_1 + C_2) \cos 3x + j(C_1 - C_2) \sin 3x \\ &= A \cos 3x + B \sin 3x \quad (A \text{ and } B \text{ can be real or complex}). \end{aligned}$$

Source page 78

ex/ Complex roots with a real part. Solve $y'' + 4y' + 7y = 0$ with $y(0) = 1$ and $y'(0) = 0$. The characteristic equation is $\lambda^2 + 4\lambda + 7 = 0$, so

$$\Rightarrow \lambda = \frac{-4 \pm \sqrt{16 - 28}}{2} = \frac{-4 \pm \sqrt{-12}}{2} = -2 \pm j\sqrt{3}.$$

The real part -2 gives a decaying factor e^{-2x} , and the imaginary part $\pm j\sqrt{3}$ gives oscillation:

$$\begin{aligned} \Rightarrow y(x) &= C_1 e^{(-2+j\sqrt{3})x} + C_2 e^{(-2-j\sqrt{3})x} \\ &= e^{-2x} (C_1 e^{j\sqrt{3}x} + C_2 e^{-j\sqrt{3}x}) \\ &= e^{-2x} (A \cos \sqrt{3}x + B \sin \sqrt{3}x). \end{aligned}$$

Now apply the initial conditions. First, $y(0) = 1$ gives $A = 1$. Differentiating y and evaluating at $x = 0$,

$$y'(0) = \left[-2e^{-2x} (A \cos \sqrt{3}x + B \sin \sqrt{3}x) + e^{-2x} (-\sqrt{3}A \sin \sqrt{3}x + \sqrt{3}B \cos \sqrt{3}x) \right] \Big|_{x=0} = -2A + \sqrt{3}B = 0,$$

so $B = \frac{2}{\sqrt{3}}$ and

$$\Rightarrow y(x) = e^{-2x} \left(\cos \sqrt{3}x + \frac{2}{\sqrt{3}} \sin \sqrt{3}x \right).$$

Theorem 3.4.1

Combining solutions as above only works if the exponentials $e^{\lambda_i x}$ really are independent of each other. This is guaranteed when the roots are distinct.

Theorem 3.4.1

If the characteristic equation of an n^{th} -order linear constant coefficient DE has n distinct roots $\lambda_1, \dots, \lambda_n$, then the set $\{e^{\lambda_1 x}, \dots, e^{\lambda_n x}\}$ is linearly independent on any interval $I \subset \mathbb{R}$.

proof Suppose, for contradiction, that they are linearly dependent. Then there are coefficients, not all zero, with

$$\alpha_1 e^{\lambda_1 x} + \alpha_2 e^{\lambda_2 x} + \dots + \alpha_n e^{\lambda_n x} = 0, \quad \forall x \in I.$$

Differentiating this identity repeatedly, each derivative multiplies the i^{th} term by another λ_i :

$$\alpha_1 \lambda_1 e^{\lambda_1 x} + \alpha_2 \lambda_2 e^{\lambda_2 x} + \dots + \alpha_n \lambda_n e^{\lambda_n x} = 0$$

$$\vdots$$

$$\alpha_1 \lambda_1^{n-1} e^{\lambda_1 x} + \alpha_2 \lambda_2^{n-1} e^{\lambda_2 x} + \dots + \alpha_n \lambda_n^{n-1} e^{\lambda_n x} = 0.$$

These n equations can be written as one matrix equation:

$$\Rightarrow \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_n \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \dots & \lambda_n^{n-1} \end{bmatrix}}_{\text{a Vandermonde matrix, invertible because the } \lambda_i \text{ are distinct}} \begin{bmatrix} \alpha_1 e^{\lambda_1 x} \\ \alpha_2 e^{\lambda_2 x} \\ \vdots \\ \alpha_n e^{\lambda_n x} \end{bmatrix} = \mathbf{0}$$

$$\Rightarrow \begin{bmatrix} \alpha_1 e^{\lambda_1 x} \\ \vdots \\ \alpha_n e^{\lambda_n x} \end{bmatrix} = \mathbf{0} \Rightarrow \alpha_i e^{\lambda_i x} = 0 \Rightarrow \alpha_i = 0.$$

All the coefficients are zero, which contradicts the assumption. ■

Source page 79

Theorem 3.4.2 — Repeated Roots Case

If a root is repeated, we have fewer than n exponentials, so we need more solutions. The missing ones are obtained by multiplying by powers of x .

Theorem 3.4.2

If λ_1 is a root of order k of the characteristic equation, then $e^{\lambda_1 x}, xe^{\lambda_1 x}, \dots, x^{k-1}e^{\lambda_1 x}$ are linearly independent solutions of the differential equation.

proof Write the equation in operator form as $L[y] = 0$, where

$$L = \frac{d^n}{dx^n} + a_1 \frac{d^{n-1}}{dx^{n-1}} + \cdots + a_{n-1} \frac{d}{dx} + a_n.$$

Applying L to an exponential just multiplies it by the characteristic polynomial. Because λ_1 is a root of multiplicity k , the polynomial factors, with a leftover polynomial $p(\lambda)$ of degree $n - k$:

$$\begin{aligned} L[e^{\lambda x}] &= (\lambda^n + a_1 \lambda^{n-1} + \cdots + a_{n-1} \lambda + a_n) e^{\lambda x} \\ &= (\lambda - \lambda_1)^k p(\lambda) e^{\lambda x}. \end{aligned}$$

At $\lambda = \lambda_1$ this is zero, so $L[e^{\lambda_1 x}] = 0$ and $e^{\lambda_1 x}$ is a solution.

To show that $xe^{\lambda_1 x}$ is also a solution, differentiate $L[e^{\lambda x}]$ with respect to λ :

$$\begin{aligned} \frac{d}{d\lambda} L[e^{\lambda x}] &= \frac{d}{d\lambda} [(\lambda - \lambda_1)^k p(\lambda) e^{\lambda x}] \\ &= k(\lambda - \lambda_1)^{k-1} p(\lambda) e^{\lambda x} + (\lambda - \lambda_1)^k \frac{d}{d\lambda} (p(\lambda) e^{\lambda x}). \end{aligned}$$

On the other hand, the derivative with respect to λ can be moved inside L , which acts on x :

$$\frac{d}{d\lambda} L[e^{\lambda x}] = L \left[\frac{d}{d\lambda} e^{\lambda x} \right] = L[xe^{\lambda x}].$$

Evaluate both expressions at $\lambda = \lambda_1$. Both terms of the right-hand side contain a positive power of $(\lambda - \lambda_1)$ (this needs $k \geq 2$), so they vanish:

$$\begin{aligned} L[xe^{\lambda x}] \Big|_{\lambda=\lambda_1} &= \underbrace{\left[k(\lambda - \lambda_1)^{k-1} p(\lambda) e^{\lambda x} + (\lambda - \lambda_1)^k \frac{d}{d\lambda} (p(\lambda) e^{\lambda x}) \right]}_0 \Big|_{\lambda=\lambda_1} \\ &\Rightarrow L[xe^{\lambda_1 x}] = 0 \Rightarrow xe^{\lambda_1 x} \text{ is also a solution.} \end{aligned}$$

Differentiating more times with respect to λ in the same way shows that $x^2 e^{\lambda_1 x}, \dots, x^{k-1} e^{\lambda_1 x}$ are solutions too. ■

example Solve $y^{(iv)} - 8y''' + 26y'' - 40y' + 25y = 0$. The characteristic equation is $\lambda^4 - 8\lambda^3 + 26\lambda^2 - 40\lambda + 25 = 0$. It has repeated complex roots: $\lambda_1 = 2 + j$ with multiplicity 2 and $\lambda_2 = 2 - j$ with multiplicity 2. Each root contributes its exponential and x times its exponential, so the general solution is

$$\begin{aligned} y(x) &= (C_1 + C_2 x) e^{(2+j)x} + (C_3 + C_4 x) e^{(2-j)x} \\ &= e^{2x} (C_1 e^{jx} + C_3 e^{-jx} + x(C_2 e^{jx} + C_4 e^{-jx})) \\ &= e^{2x} ((A \cos x + B \sin x) + x(C \cos x + D \sin x)). \end{aligned}$$

Source page 80

Def'n — Stability: consider the homogeneous differential equation

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_{n-1} \frac{dy}{dt} + a_n y = 0 \quad (\text{DE}_n)$$

with general solution $y(t) = C_1 y_1(t) + \cdots + C_n y_n(t)$. It is called *stable* if all of its solutions stay bounded for $t > 0$, that is,

$$|y_i(t)| < M_i < \infty, \quad 1 \leq i \leq n, \quad t > 0.$$

Theorem 3.4.3

Stability is decided by where the roots of the characteristic equation lie in the complex plane.

Theorem 3.4.3

(DE_n) is stable if and only if its characteristic equation has no roots to the right of the imaginary axis, and any roots on the imaginary axis are non-repeated.

proof (i) *Non-repeated roots.* A non-repeated root $\lambda = a + jb$ contributes a solution $e^{\lambda t} = e^{at} e^{jbt}$. If $a \leq 0$ this stays bounded:

$$|e^{\lambda t}| = |e^{at} e^{jbt}| = |e^{at}| \underbrace{|e^{jbt}|}_1 = e^{at} \leq 1, \quad t > 0.$$

(ii) *Repeated roots.* A repeated root $\lambda = a + jb$ of order k contributes the k solutions $e^{\lambda t}, t e^{\lambda t}, \dots, t^{k-1} e^{\lambda t}$. For $a < 0$ we show that all of them are bounded. For $t > 0$,

$$|t^i e^{\lambda t}| = |t^i| \cdot |e^{at} \cdot e^{jbt}| = t^i e^{at}.$$

Since $t^i e^{at} \rightarrow 0$ as $t \rightarrow \infty$ when $a < 0$, and $t^i e^{at}$ is finite on every bounded interval $0 < t < T$, the function $t^i e^{at}$ is bounded for $t > 0$.

(iii) *Instability.* If a simple root has $a > 0$, or a repeated root has $a \geq 0$, the solutions it contributes grow without bound, so the system is unstable. ■

Theorem 3.4.4 — Coefficients of mixed sign

There is a quick necessary check for stability that needs no root finding: look at the signs of the coefficients.

Theorem 3.4.4

If the coefficients of the characteristic equation are real and of mixed sign (at least one positive and at least one negative), then there is at least one root λ with $\Re\{\lambda\} > 0$, so the system is unstable.

Source page 81

Note that the theorem does not say that every unstable system has coefficients of mixed sign. It says only that mixed signs are enough to guarantee instability. (An example appears below.)

Proof We prove the contrapositive: if a system is stable, its characteristic polynomial cannot have coefficients of mixed sign. A stable system has all its roots in the closed left half of the complex plane, $\Re\{\lambda\} \leq 0$. Since the coefficients are real, the complex roots come in conjugate pairs, so the monic characteristic polynomial factors into two kinds of factors:

- (i) a factor $(\lambda - \lambda_r) = \lambda + c$ for each real root $\lambda_r = -c \leq 0$, which has non-negative coefficients 1 and c ;
- (ii) a factor for each conjugate pair $a \pm jb$ with $a \leq 0$,

$$(\lambda - (a + jb))(\lambda - (a - jb)) = \lambda^2 - 2a\lambda + (a^2 + b^2),$$

which also has non-negative coefficients, because $-2a \geq 0$ and $a^2 + b^2 \geq 0$.

When two polynomials with non-negative coefficients are multiplied, every coefficient of the product is a sum of products of non-negative numbers, so it is non-negative too. Hence the characteristic polynomial of a stable system has only non-negative coefficients, and in particular its coefficients cannot be of mixed sign. ■

Source page 82

Conversely, a polynomial of mixed sign must have a root with a positive real part, which makes the system unstable. This is the statement of the theorem.

Now we see why the converse of the theorem fails: unstable roots do not always produce mixed signs. For an unstable real root $\lambda_r > 0$, multiplying a polynomial $q(\lambda)$ with non-negative coefficients by $(\lambda - \lambda_r)$ gives a polynomial whose last non-zero coefficient is negative, so the signs do mix. But for an unstable pair of complex roots, $a > 0$, the factor $\lambda^2 - 2a\lambda + (a^2 + b^2)$ itself has mixed signs, and multiplying it by another polynomial may or may not leave the result with mixed signs.

Note that even if the added roots are unstable, $p(\lambda)$ may have no sign mix. For example, take

$$q(\lambda) = (\lambda + 1), \quad \lambda^2 - 2a\lambda + (a^2 + b^2) = \lambda^2 - \epsilon\lambda + 1, \quad \epsilon > 0,$$

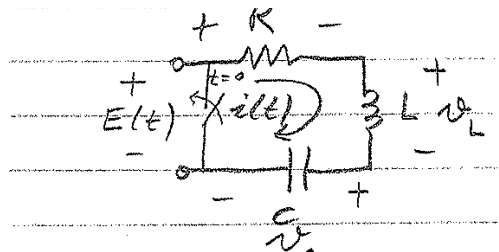
which has the unstable pair $a = \epsilon/2 > 0$. Then

$$\begin{aligned} p(\lambda) &= q(\lambda) \cdot (\lambda^2 - 2a\lambda + (a^2 + b^2)) = (\lambda + 1)(\lambda^2 - \epsilon\lambda + 1) \\ &= \lambda^3 + (1 - \epsilon)\lambda^2 + (1 - \epsilon)\lambda + 1. \end{aligned}$$

If $0 < \epsilon < 1$, all the coefficients of $p(\lambda)$ are positive, so the coefficients do not have mixed signs, even though the system is unstable.

Application Example: Oscillator Circuit

Here is a physical system described by a second-order differential equation: a resistor R , an inductor L and a capacitor C connected in a loop to a voltage source $E(t)$.



By Kirchhoff's voltage law, the source voltage equals the sum of the voltages across the three elements. Each element has its own voltage-current law:

$$\begin{aligned} E(t) &= v_R(t) + v_L(t) + v_C(t) \\ v_R(t) &= Ri(t) \\ v_L(t) &= L \frac{di(t)}{dt} \\ i(t) &= C \frac{dv_C(t)}{dt} \Rightarrow v_C(t) = \frac{1}{C} \int_{-\infty}^t i(\tau) d\tau. \end{aligned}$$

Substituting these laws into the first one, and then differentiating once to remove the integral, gives a second-order differential equation for the current:

$$\begin{aligned} \Rightarrow E(t) &= Ri(t) + L \frac{di(t)}{dt} + \frac{1}{C} \int_{-\infty}^t i(\tau) d\tau \\ \Rightarrow E'(t) &= L \frac{d^2i(t)}{dt^2} + R \frac{di(t)}{dt} + \frac{1}{C} i(t). \end{aligned}$$

Source page 83

Consider the free oscillator, with no source, $E(t) = 0$. Dividing by L ,

$$\frac{d^2 i}{dt^2} + \frac{R}{L} \frac{di}{dt} + \frac{1}{LC} i = 0, \quad \text{characteristic equation: } \lambda^2 + \frac{R}{L} \lambda + \frac{1}{LC} = 0.$$

$$\Rightarrow \lambda_1 = \frac{-R/L + \sqrt{R^2/L^2 - 4/LC}}{2}, \quad \lambda_2 = \frac{-R/L - \sqrt{R^2/L^2 - 4/LC}}{2}.$$

The sign of the number under the square root decides the behavior, and it depends on the resistance R .

(i) *Large resistance.* If $\frac{R^2}{L^2} - \frac{4}{LC} > 0$, which is equivalent to $R^2 C - 4L > 0$, that is,

$$R > 2\sqrt{\frac{L}{C}},$$

then there are two real roots, both negative (the square root is smaller than R/L). The current is

$$i(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}, \quad \lambda_1 < 0, \quad \lambda_2 < 0.$$

There is no oscillation: this is the *overdamped* case.

(ii) *The borderline.* If $R = 2\sqrt{L/C}$, then $\lambda_1 = \lambda_2 = -\frac{R}{2L}$ and

$$i(t) = C_1 e^{\lambda_1 t} + C_2 t e^{\lambda_1 t}.$$

There is again no oscillation: this is the *critically damped* case.

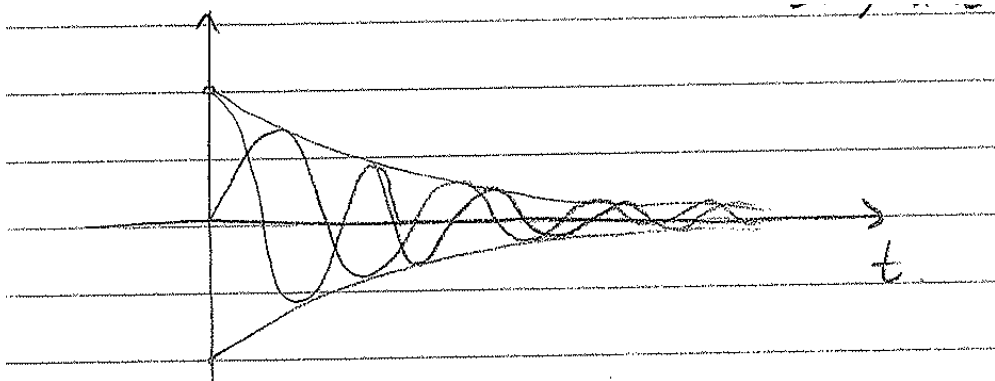
(iii) *Small resistance.* If $R < 2\sqrt{L/C}$, the roots are a complex conjugate pair,

$$\lambda_1 = \frac{-R/L + \sqrt{R^2/L^2 - 4/LC}}{2} = \frac{-R/L + j\sqrt{4/LC - R^2/L^2}}{2} = \lambda_2^*.$$

The real part makes the oscillation decay, and the imaginary part sets its frequency:

$$i(t) = \underbrace{e^{-\frac{R}{2L}t}}_{\text{decay envelope (time constant } 2L/R)} (A \cos(\omega t) + B \sin(\omega t)), \quad \underbrace{\omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}}_{\text{damped natural frequency}}.$$

These are *underdamped* oscillations. The constants A and B are determined from $i(0)$ and $i'(0)$.



Source page 84

Laplace Transform*

(* Chapter 5 in the textbook.)

A general linear integral transform has the form

$$F(s) = \int_a^b K(t, s) f(t) dt,$$

where $K(t, s)$ is called the *transformation kernel*. The Laplace transform is the case with kernel e^{-st} on the interval $[0, \infty)$, where s is a complex-valued variable of the transform domain:

The Laplace transform

$$F(s) = \int_0^\infty f(t) e^{-st} dt.$$

Just as the z-transform reduces difference equations to linear algebraic equations, the Laplace transform reduces linear constant-coefficient differential equations to linear algebraic equations.

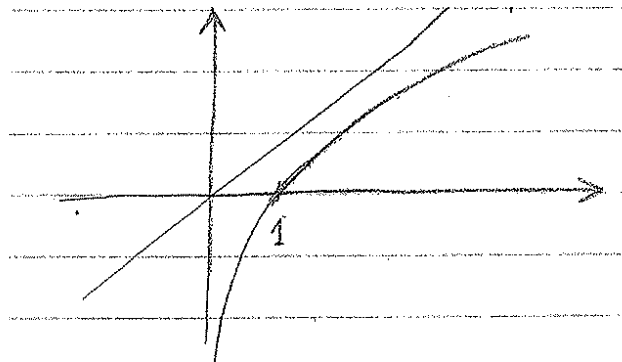
Calculation of the transform

The integral $F(s) = \int_0^\infty f(t) e^{-st} dt$ does not converge for every function. Convergence is assured if f is of *exponential order*, that is, if there are constants K and c with

$$|f(t)| \leq K e^{ct} \quad \text{for } t > T.$$

ex/ $f(t) = \sin(\omega t)$ is of exponential order, because $|f(t)| \leq e^{0t} = 1$ for $t > 0$.

ex/ $f(t) = t^2$ is of exponential order, because $t^2 \leq e^t$ for $t > T$. The condition $T^2 \leq e^T$ is the same as $2 \ln T \leq T$, which holds for every $T > 0$, as the picture shows: the line $y = T$ stays above the curve $y = 2 \ln T$.



We can also check this with l'Hôpital's rule:

$$\lim_{t \rightarrow \infty} \frac{t^2}{e^t} = \lim_{t \rightarrow \infty} \frac{2t}{e^t} = \lim_{t \rightarrow \infty} \frac{2}{e^t} = 0.$$

l'Hôpital's rule

Source page 85

ex/ A function that is not of exponential order: $f(t) = e^{t^2}$. It grows faster than any exponential e^{ct} , whatever c we choose:

$$\lim_{t \rightarrow \infty} \frac{e^{t^2}}{e^{ct}} = \lim_{t \rightarrow \infty} e^{(t^2 - ct)} = \infty, \quad \text{for any } c.$$

Theorem 5.2.1

The next theorem says when the Laplace transform exists.

Theorem 5.2.1: existence of the Laplace transform

Let $f(t)$ satisfy

- (i) $f(t)$ is piecewise continuous on every interval $0 \leq t \leq A$, with a finite number of discontinuities, and
- (ii) $f(t)$ is of exponential order: $|f(t)| \leq Ke^{ct}$ for $t \geq T$.

Then the Laplace transform $F(s)$ exists for $\Re\{s\} > c$.

proof Split the integral at T . The first part is over a finite interval, and it exists for every s because of condition (i):

$$F(s) = \int_0^\infty f(t)e^{-st} dt = \underbrace{\int_0^T f(t)e^{-st} dt}_{\text{exists for all } s \text{ by (i)}} + \int_T^\infty f(t)e^{-st} dt.$$

For the second part, we bound the integrand using condition (ii) and $|e^{-st}| = e^{-\Re\{s\}t}$:

$$\begin{aligned} \left| \int_T^\infty f(t)e^{-st} dt \right| &\leq \int_T^\infty |f(t)e^{-st}| dt \leq \int_T^\infty |f(t)| |e^{-st}| dt \\ &\leq \int_T^\infty Ke^{ct} e^{-\Re\{s\}t} dt = K \int_T^\infty e^{-(\Re\{s\}-c)t} dt \\ &= K \frac{e^{-(\Re\{s\}-c)T}}{\Re\{s\} - c} < \infty \end{aligned}$$

for $\Re\{s\} > c$, so the second part converges absolutely. ■

ex/ The exponential $f(t) = e^{at}$:

$$F(s) = \int_0^\infty e^{at} e^{-st} dt = \int_0^\infty e^{-(s-a)t} dt = \frac{1}{s-a}, \quad \Re\{s\} > a.$$

ex/ The sine $f(t) = \sin at$. Write the sine with complex exponentials and use the previous example twice:

$$\begin{aligned} F(s) &= \int_0^\infty \sin at e^{-st} dt = \int_0^\infty \left(\frac{e^{jat} - e^{-jat}}{2j} \right) e^{-st} dt \\ &= \frac{1}{2j} \int_0^\infty e^{-(s-ja)t} dt - \frac{1}{2j} \int_0^\infty e^{-(s+ja)t} dt \\ &= \frac{1}{2j} \frac{1}{s-ja} - \frac{1}{2j} \frac{1}{s+ja} = \frac{1}{2j} \frac{(s+ja) - (s-ja)}{s^2 + a^2} = \frac{a}{s^2 + a^2}, \quad \Re\{s\} > 0. \end{aligned}$$

Source page 86

Notes

(i) The inverse Laplace transform is also an integral transform, but computing it directly requires techniques from complex analysis. In these notes we therefore invert by inspection, using known transform pairs, as we did for the z-transform.

(ii) The inverse is unique: if

$$f_1(t) \xrightarrow{\mathcal{L}} F_1(s), \quad f_2(t) \xrightarrow{\mathcal{L}} F_2(s)$$

and $\int_0^\infty |f_1(t) - f_2(t)| dt > 0$, then $F_1(s) \neq F_2(s)$. (Functions that differ only at isolated points have the same transform, which is why the condition is stated with an integral.)

Properties of the Laplace Transform

These properties are what make the Laplace transform useful for differential equations.

(i) *Linearity*. For all constants α, β and all functions $u(t)$ and $v(t)$ of exponential order,

$$\mathcal{L}\{\alpha u(t) + \beta v(t)\} = \alpha \mathcal{L}\{u(t)\} + \beta \mathcal{L}\{v(t)\}.$$

(ii) *Linearity of the inverse*. For all constants α, β and transforms $U(s)$ and $V(s)$,

$$\mathcal{L}^{-1}\{\alpha U(s) + \beta V(s)\} = \alpha \mathcal{L}^{-1}\{U(s)\} + \beta \mathcal{L}^{-1}\{V(s)\}.$$

(iii) *Derivative*. Differentiation in time becomes multiplication by s , with the initial value subtracted:

Derivative property

$$\mathcal{L}\{f'(t)\} = s\mathcal{L}\{f(t)\} - f(0).$$

proof Integrate by parts:

$$\begin{aligned} \mathcal{L}\{f'(t)\} &= \int_0^\infty f'(t)e^{-st} dt = f(t)e^{-st} \Big|_0^\infty + s \int_0^\infty f(t)e^{-st} dt \\ &= -f(0) + sF(s). \end{aligned}$$

(The property remains valid even if $f'(t)$ is only piecewise continuous.)

(iv) *Higher derivatives*. Applying the rule twice gives the second derivative:

$$\begin{aligned} \mathcal{L}\{f''(t)\} &= s\mathcal{L}\{f'(t)\} - f'(0) \\ &= s^2\mathcal{L}\{f(t)\} - sf(0) - f'(0). \end{aligned}$$

In general,

$$\Rightarrow \mathcal{L}\{f^{(k)}(t)\} = s^k \mathcal{L}\{f(t)\} - \sum_{i=1}^k s^{k-i} f^{(i-1)}(0).$$

This is useful for solving differential equations with given initial conditions, since the initial conditions appear automatically.

Source page 87

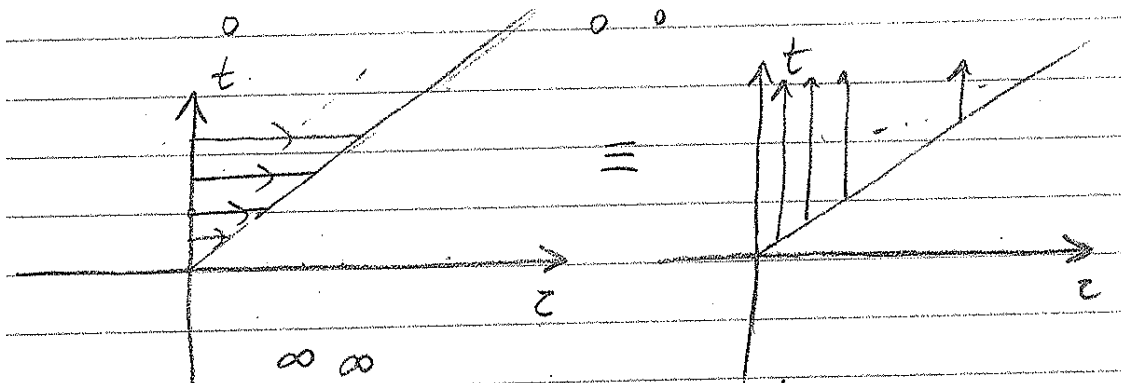
(v) *Convolution*. For functions $f(t)$ and $g(t)$ that are zero for $t < 0$, the *convolution* is

$$h(t) = f(t) * g(t) = (f * g)(t) = \int_0^t f(\tau)g(t - \tau) d\tau.$$

Its Laplace transform is the double integral

$$H(s) = \int_0^\infty h(t)e^{-st} dt = \int_0^\infty \int_0^t f(\tau)g(t - \tau) d\tau e^{-st} dt.$$

The key step is to swap the order of integration. The region of integration, $0 \leq \tau \leq t < \infty$, is the same whether we sum along horizontal slices (τ from 0 to t) or along vertical slices (t from τ to ∞):



After the swap, we substitute $\mu = t - \tau$ in the inner integral, and the double integral splits into a product of two single integrals:

$$\begin{aligned} \Rightarrow H(s) &= \int_0^\infty \int_\tau^\infty f(\tau)g(t - \tau)e^{-st} dt d\tau \\ &= \int_0^\infty f(\tau) \left[\int_\tau^\infty \underbrace{g(t - \tau)}_\mu e^{-st} dt \right] d\tau \\ &= \int_0^\infty f(\tau) \left[\int_0^\infty g(\mu)e^{-s(\tau + \mu)} d\mu \right] d\tau \\ &= \int_0^\infty f(\tau) \left[\int_0^\infty g(\mu)e^{-s\mu} d\mu \right] e^{-s\tau} d\tau \\ &= \left[\int_0^\infty f(\tau)e^{-s\tau} d\tau \right] \left[\int_0^\infty g(\mu)e^{-s\mu} d\mu \right]. \end{aligned}$$

Convolution theorem

$$H(s) = F(s) \cdot G(s).$$

So convolution in time becomes ordinary multiplication in the s -domain.

Source page 88

ex/ Find the inverse transform of $F(s) = \frac{3}{s^2 + 3s - 10} = \frac{3}{(s+5)(s-2)}$. There are two ways.

(i) *Partial fractions.*

$$\frac{3}{(s+5)(s-2)} = \frac{3}{7} \left[\frac{1}{s-2} - \frac{1}{s+5} \right] \Rightarrow f(t) = \frac{3}{7}(e^{2t} - e^{-5t}), \quad t \geq 0.$$

(ii) *Convolution.* Write $F(s)$ as a product of two simple transforms, and use the convolution property:

$$\begin{aligned} \frac{3}{(s+5)(s-2)} &= 3 \cdot \frac{1}{s+5} \cdot \frac{1}{s-2} \\ \Rightarrow f(t) &= 3 \cdot (e^{-5t} * e^{2t}) \\ &= 3 \cdot \int_0^t e^{2\tau} e^{-5(t-\tau)} d\tau \\ &= 3e^{-5t} \int_0^t e^{7\tau} d\tau = 3e^{-5t} \frac{1}{7} (e^{7\tau} \Big|_0^t) \\ &= \frac{3}{7} e^{-5t} (e^{7t} - 1) \\ &= \frac{3}{7} (e^{2t} - e^{-5t}), \quad t \geq 0. \end{aligned}$$

Both methods give the same answer.

Solution to Linear Constant Coefficient Differential Equations by Laplace Transform

Transform the differential equation, solve the resulting algebraic equation, then invert. As an example take $x'' + ax' + bx = f(t)$ and transform both sides with the derivative property, writing $\bar{X}(s)$ for the transform of x :

$$(s^2 \bar{X}(s) - sx(0) - x'(0)) + a(s\bar{X}(s) - x(0)) + b\bar{X}(s) = F(s).$$

$$\Rightarrow (s^2 + as + b)\bar{X}(s) = sx(0) + ax(0) + x'(0) + F(s).$$

Dividing by $s^2 + as + b$ shows that the solution has two parts: one due to the initial conditions, and one due to the input $f(t)$:

Solution in the s -domain

$$\bar{X}(s) = \underbrace{\frac{(s+a)x(0) + x'(0)}{s^2 + as + b}}_{\text{invert by partial fraction expansion}} + \underbrace{\frac{F(s)}{s^2 + as + b}}_{\text{invert by partial fractions or by the convolution property}}.$$

Source page 89

ex/ A first-order equation with an input. Solve $x' + px = q(t)$ with $x(0) = x_0$. In the Laplace domain the derivative property turns it into an algebraic equation, with $Q(s)$ the transform of $q(t)$:

$$s\bar{X}(s) - x(0) + p\bar{X}(s) = Q(s)$$

$$\Rightarrow \bar{X}(s) = \frac{x_0}{s+p} + \frac{Q(s)}{s+p}.$$

Inverting, the first term gives $x_0 e^{-pt}$ and the second term is a convolution of $q(t)$ with e^{-pt} :

$$\Rightarrow x(t) = x_0 e^{-pt} + \int_0^t q(\tau) e^{-p(t-\tau)} d\tau$$

Pulling the factor e^{-pt} out of the integral gives the form that is often called the integrating-factor solution:

Solution of $x' + px = q(t)$

$$x(t) = x_0 e^{-pt} + e^{-pt} \int_0^t q(\tau) e^{p\tau} d\tau = \left[x_0 + \int_0^t q(\tau) e^{p\tau} d\tau \right] e^{-pt}.$$

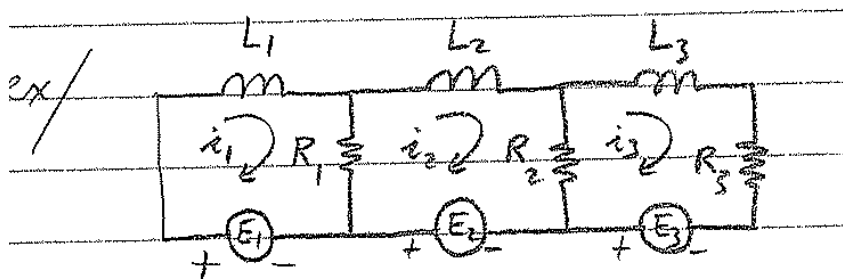
Systems of Linear Differential Equations

Many systems have several unknown functions that influence each other. Their equations form a *system of linear differential equations*:

$$\left. \begin{aligned} a_{11}(t)x_1' + \cdots + a_{1n}(t)x_n' + b_{11}(t)x_1 + \cdots + b_{1n}(t)x_n &= f_1(t) \\ &\vdots \\ a_{n1}(t)x_1' + \cdots + a_{nn}(t)x_n' + b_{n1}(t)x_1 + \cdots + b_{nn}(t)x_n &= f_n(t) \end{aligned} \right\} \text{forcing functions.}$$

The $a_{ij}(t)$ and $b_{ij}(t)$ are known coefficients, and the $f_i(t)$ are given *forcing functions* (inputs).

ex/ An electrical circuit with three loops gives such a system. Applying Kirchhoff's voltage law to each loop gives the three equations on the left, and rearranging them to collect the unknown currents i_1, i_2, i_3 gives the equations on the right:



$$\left. \begin{aligned} L_1 i_1' + R_1(i_1 - i_2) &= E_1(t) \\ L_2 i_2' + R_2(i_2 - i_3) + R_1(i_2 - i_1) &= E_2(t) \\ L_3 i_3' + R_3 i_3 + R_2(i_3 - i_2) &= E_3(t) \end{aligned} \right\} \quad \begin{aligned} L_1 i_1' + R_1 i_1 - R_1 i_2 &= E_1(t) \\ L_2 i_2' - R_1 i_1 + (R_2 + R_1) i_2 - R_2 i_3 &= E_2(t) \\ L_3 i_3' - R_2 i_2 + (R_3 + R_2) i_3 &= E_3(t). \end{aligned}$$

Theorem 3.9.1

In the standard form, each equation has one derivative on the left, and solutions exist and are unique.

Theorem 3.9.1: existence and uniqueness

Let $a_{ij}(t)$, $1 \leq i, j \leq n$, and $f_i(t)$, $1 \leq i \leq n$, be continuous on a closed interval I , and let $x_i(t_0) = b_i$, $1 \leq i \leq n$, for some $t_0 \in I$. Then the system

$$\begin{aligned}x_1' &= a_{11}(t)x_1 + \cdots + a_{1n}(t)x_n + f_1(t) \\&\vdots \\x_n' &= a_{n1}(t)x_1 + \cdots + a_{nn}(t)x_n + f_n(t)\end{aligned}$$

has a unique solution on the entire interval I .

Source page 90

Note that

- (i) In this standard form the left-hand side contains each derivative on its own. A general system can be brought to this form by Gauss–Jordan elimination on the equations (possible when the matrix $[a_{ij}(t)]$ of derivative coefficients is invertible).
- (ii) The system can be written in vector form, which is how we will work with it from now on:

$$\mathbf{x}' = A(t)\mathbf{x} + \mathbf{f}$$

$$\text{where } \mathbf{x} = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, A(t) = [a_{ij}(t)]_{n \times n} \text{ and } \mathbf{f} = \begin{bmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{bmatrix}.$$

ex/ Solve a 2×2 system by the Laplace transform. First we bring it to standard form by subtracting the first equation from the second:

$$\left. \begin{aligned} x_1' &= x_1 + x_2 + 3t \\ x_1' + x_2' &= 5x_1 + 2x_2 + 5 \end{aligned} \right\} \Rightarrow \begin{aligned} x_1' &= x_1 + x_2 + 3t \\ x_2' &= 4x_1 + x_2 + 5 - 3t. \end{aligned}$$

In vector form,

$$\begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 3t \\ 5 - 3t \end{bmatrix}.$$

In the Laplace transform domain this becomes an algebraic matrix equation,

$$(s\bar{\mathbf{x}} - \mathbf{x}_0) = A\bar{\mathbf{x}} + \mathbf{F}, \quad \bar{\mathbf{x}} = \begin{bmatrix} \bar{X}_1(s) \\ \bar{X}_2(s) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} F_1(s) \\ F_2(s) \end{bmatrix}$$

where the transforms of the forcing terms are

$$\begin{aligned} F_1(s) &= \int_0^\infty 3te^{-st} dt \\ &= \underbrace{-\frac{3te^{-st}}{s} \Big|_0^\infty}_0 + \frac{3}{s} \underbrace{\int_0^\infty e^{-st} dt}_{-\frac{1}{s}e^{-st} \Big|_0^\infty = \frac{1}{s}} = \frac{3}{s^2}, \\ F_2(s) &= \int_0^\infty (5 - 3t)e^{-st} dt = \frac{5}{s} - \frac{3}{s^2}. \end{aligned}$$

Collecting the unknowns on the left and solving,

$$\Rightarrow (sI - A)\bar{\mathbf{x}} = \mathbf{x}_0 + \mathbf{F} \quad \Rightarrow \quad \bar{\mathbf{x}} = (sI - A)^{-1}(\mathbf{x}_0 + \mathbf{F}),$$

where, for our matrix, the 2×2 inverse formula gives

$$(sI - A) = \begin{bmatrix} s-1 & -1 \\ -4 & s-1 \end{bmatrix}, \quad (sI - A)^{-1} = \frac{1}{s^2 - 2s - 3} \begin{bmatrix} s-1 & 1 \\ 4 & s-1 \end{bmatrix}.$$

Source page 91

Note that $s^2 - 2s - 3 = (s - 3)(s + 1)$. Multiplying out gives two parts, one that depends on the initial conditions and one that depends on the forcing:

$$\begin{aligned}\Rightarrow \bar{\mathbf{x}} &= \frac{1}{(s-3)(s+1)} \begin{bmatrix} s-1 & 1 \\ 4 & s-1 \end{bmatrix} \left(\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} + \begin{bmatrix} 3/s^2 \\ 5/s - 3/s^2 \end{bmatrix} \right) \\ &= \begin{bmatrix} \frac{(s-1)x_1(0) + x_2(0)}{(s-3)(s+1)} \\ \frac{4x_1(0) + (s-1)x_2(0)}{(s-3)(s+1)} \end{bmatrix} + \begin{bmatrix} \frac{(s-1)3}{s^2(s-3)(s+1)} + \frac{5s-3}{s^2(s-3)(s+1)} \\ \frac{12}{s^2(s-3)(s+1)} + \frac{(5s-3)(s-1)}{s^2(s-3)(s+1)} \end{bmatrix}\end{aligned}$$

Each entry is a ratio of polynomials. Expanding each in partial fractions gives

$$= \begin{bmatrix} \frac{2x_1(0) + x_2(0)}{4(s-3)} + \frac{2x_1(0) - x_2(0)}{4(s+1)} \\ \frac{x_2(0) + 2x_1(0)}{2(s-3)} + \frac{x_2(0) - 2x_1(0)}{2(s+1)} \end{bmatrix} + \begin{bmatrix} -\frac{4}{6s} + \frac{2}{5s^2} + \frac{1/2}{s-3} + \frac{7/2}{s+1} \\ \frac{1}{s} - \frac{1}{s^2} + \frac{1}{s-3} - \frac{1}{s+1} \end{bmatrix}.$$

Inverting term by term, using $\frac{1}{s-a} \leftrightarrow e^{at}$, $\frac{1}{s} \leftrightarrow 1$ and $\frac{1}{s^2} \leftrightarrow t$, gives the solution:

$$\begin{aligned}\mathbf{x}(t) &= \underbrace{\begin{bmatrix} \left(\frac{2x_1(0) + x_2(0)}{4} \right) e^{+3t} + \left(\frac{2x_1(0) - x_2(0)}{4} \right) e^{-t} \\ \frac{x_2(0) + 2x_1(0)}{2} e^{3t} + \frac{x_2(0) - 2x_1(0)}{2} e^{-t} \end{bmatrix}}_{\text{determined by the initial conditions } x_1(0), x_2(0)} \\ &\quad + \underbrace{\begin{bmatrix} -4 + 2t + \frac{1}{2}e^{3t} + \frac{7}{2}e^{-t} \\ 6 - 5t + e^{3t} - 7e^{-t} \end{bmatrix}}_{\text{particular solution for zero initial conditions}}.\end{aligned}$$

There is another approach, based on *state-transition matrices*. We want to solve

$$\mathbf{x}' = A(t)\mathbf{x} + \mathbf{f}(t) \quad \text{with initial value } \mathbf{x}(0) = \mathbf{x}_0.$$

Observations

We start with the homogeneous system $\mathbf{x}' = A(t)\mathbf{x}$.

(i) The set of solutions of the homogeneous system

$$\mathbf{x}' = A(t)\mathbf{x}$$

is an n -dimensional vector space of functions. It is closed under the vector space operations, because if $\mathbf{x}_1(t)$ is a solution then so is $\alpha\mathbf{x}_1(t)$, and if $\mathbf{x}_2(t)$ is also a solution then so is $\mathbf{x}_1(t) + \mathbf{x}_2(t)$.

Source page 92

(ii) Suppose $\{\phi_1(t), \dots, \phi_n(t)\}$ are n linearly independent solutions of $\mathbf{x}' = A(t)\mathbf{x}$. Then every solution has the form

$$\mathbf{x}(t) = \alpha_1 \phi_1(t) + \dots + \alpha_n \phi_n(t)$$

for some constants α_i . In matrix form, with the solutions as the columns of a *fundamental matrix* $\Phi(t)$,

$$\mathbf{x}(t) = \Phi(t)\boldsymbol{\alpha}$$

where $\Phi(t) = [\phi_1(t) \ \dots \ \phi_n(t)]$ and $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$.

(iii) *Notation.* We write

$$\Phi(t; t_0, X_0) = [\phi(t; t_0, \mathbf{x}_1) \ \dots \ \phi(t; t_0, \mathbf{x}_n)],$$

where $\phi(t; t_0, \mathbf{x}_i)$ is the solution of

$$\mathbf{x}' = A\mathbf{x}, \quad \text{with } \mathbf{x}(t_0) = \mathbf{x}_i.$$

(iv) The special case $X_0 = I$, that is, starting each solution from a standard basis vector, is called the state-transition matrix:

$$\Phi(t; t_0) = [\phi(t; t_0, \mathbf{e}_1) \ \dots \ \phi(t; t_0, \mathbf{e}_n)].$$

(v) The unique solution of $\mathbf{x}' = A(t)\mathbf{x}$ with $\mathbf{x}(t_0) = \mathbf{x}_0$ can be written with either matrix. Write the initial vector in the starting vectors $\mathbf{x}_i$, with coefficients $\boldsymbol{\alpha}$, and use the fact that a combination of solutions is a solution:

$$\phi(t; t_0, \mathbf{x}_0) = \sum_{i=1}^n \phi(t; t_0, \mathbf{x}_i) \alpha_i = \Phi(t; t_0, X_0) \boldsymbol{\alpha},$$

where

$$\mathbf{x}_0 = \alpha_1 \mathbf{x}_1 + \dots + \alpha_n \mathbf{x}_n = \underbrace{[\mathbf{x}_1 \ \dots \ \mathbf{x}_n]}_{X_0} \boldsymbol{\alpha} \Rightarrow \boldsymbol{\alpha} = X_0^{-1} \mathbf{x}_0.$$

So

$$\Rightarrow \phi(t; t_0, \mathbf{x}_0) = \Phi(t; t_0, X_0) X_0^{-1} \mathbf{x}_0.$$

The choice $X_0 = I$ gives the state-transition matrix:

The state-transition matrix

$$\phi(t; t_0, \mathbf{x}_0) = \Phi(t; t_0) \mathbf{x}_0 \quad \Longrightarrow \quad \Phi(t; t_0, X_0) = \Phi(t; t_0) X_0.$$

In words, the state-transition matrix carries the state at time t_0 to the state at time t .

Source page 93

Properties of the State-Transition Matrix

Three properties of $\Phi(t; t_0)$ make it easy to work with. Throughout, I is the interval on which the solutions exist.

(i) *Non-singularity.* $\Phi(t; t_0, X_0)$ is non-singular for all t and t_0 in I , whenever X_0 is non-singular.

proof We argue by contradiction. Assume that $\Phi(t_1; t_0, X_0)$ is singular for some t_1 and t_0 in I . Then there is a vector $\mathbf{c} \neq \mathbf{0}$ with

$$\Phi(t_1; t_0, X_0)\mathbf{c} = \mathbf{0}.$$

Define $\boldsymbol{\psi}(t) = \Phi(t; t_0, X_0)\mathbf{c}$. This is a combination of solutions, hence a solution of $\mathbf{x}' = A(t)\mathbf{x}$, and it satisfies $\boldsymbol{\psi}(t_1) = \mathbf{0}$. The zero function is also a solution with this initial value, and the solution is unique (Theorem 3.9.1), so $\boldsymbol{\psi}(t) = \mathbf{0}$ for all t , the zero function. In particular

$$\boldsymbol{\psi}(t_0) = X_0\mathbf{c} = \mathbf{0},$$

which says X_0 is singular, contradicting our assumption. ■

(ii) *The composition rule.* For all t, t_1, t_0 in I ,

$$\Phi(t; t_1)\Phi(t_1; t_0) = \Phi(t; t_0).$$

In words: going from t_0 to t_1 and then on to t is the same as going from t_0 to t directly.

proof Let $\Psi(t) = \Phi(t; t_1)\Phi(t_1; t_0)$. Differentiating, and using that $\Phi(t; t_1)$ solves the matrix equation,

$$\begin{aligned}\Psi' &= \Phi'(t; t_1)\Phi(t_1; t_0) = A(t) \underbrace{\Phi(t; t_1)\Phi(t_1; t_0)}_{\Psi(t)} \\ &\Rightarrow \Psi' = A(t)\Psi(t),\end{aligned}$$

and at $t = t_1$ the initial value is

$$\Psi(t_1) = \underbrace{\Phi(t_1; t_1)}_I \Phi(t_1; t_0) = \Phi(t_1; t_0).$$

Thus $\Psi(t)$ is the unique solution of $X' = A(t)X$ with $X(t_1) = \Phi(t_1; t_0)$. But $\Phi(t; t_0)$ solves the same problem, so by uniqueness

$$\Psi(t) = \Phi(t; t_0) \Rightarrow \Phi(t; t_0) = \Phi(t; t_1)\Phi(t_1; t_0). \quad \blacksquare$$

(iii) *The inverse.* Taking $t = t_0$ in rule (ii) gives

$$\Phi(t_0; t_0) = I = \Phi(t_0; t_1)\Phi(t_1; t_0) \quad \forall t_0, t_1 \in I,$$

so $\Phi(t_0; t_1)$ is the inverse of $\Phi(t_1; t_0)$. Renaming the variables:

Inverse of the state-transition matrix

$$\Phi^{-1}(t; t_0) = \Phi(t_0; t).$$

Running the system backward in time is the same as inverting the forward map.

Source page 94

The Non-Homogeneous Case

Now we add a forcing term and solve

$$\mathbf{x}' = A(t)\mathbf{x} + \mathbf{f}(t).$$

Let $\Phi(t; t_0, X_0)$ be any fundamental matrix of the homogeneous problem. Then $\mathbf{x} = \Phi(t; t_0, X_0)\mathbf{c}$ solves the homogeneous problem for every constant vector $\mathbf{c}$.

Idea. Use the *variation of parameters* technique: replace the constant vector $\mathbf{c}$ by an unknown function $\mathbf{v}(t)$, so we look for a solution of the form

$$\mathbf{x} = \Phi(t; t_0, X_0)\mathbf{v}(t).$$

Substituting this guess into the non-homogeneous equation and using the product rule,

$$\begin{aligned}\mathbf{x}' &= \Phi'(t; t_0, X_0)\mathbf{v}(t) + \Phi(t; t_0, X_0)\mathbf{v}'(t) \\ &= A(t) \underbrace{\Phi(t; t_0, X_0)\mathbf{v}(t)}_{\mathbf{x}} + \Phi(t; t_0, X_0)\mathbf{v}'(t) \\ &= A(t)\mathbf{x} + \Phi(t; t_0, X_0)\mathbf{v}'(t).\end{aligned}$$

Comparing with the original equation, the extra term must equal the forcing:

$$\begin{aligned}\Rightarrow \Phi(t; t_0, X_0)\mathbf{v}'(t) &= \mathbf{f}(t) \\ \Rightarrow \mathbf{v}'(t) &= \Phi^{-1}(t; t_0, X_0)\mathbf{f}(t) \\ \Rightarrow \mathbf{v}(t) &= \underbrace{\int \Phi^{-1}(t; t_0, X_0)\mathbf{f}(t) dt}_{\mathbf{V}(t)} + \mathbf{c} = \mathbf{V}(t) + \mathbf{c}.\end{aligned}$$

Hence the general solution is the sum of a particular part and a complementary part:

$$\mathbf{x} = \Phi(t; t_0, X_0) \cdot (\mathbf{V}(t) + \mathbf{c}) = \underbrace{\Phi_p(t)}_{\text{particular solution}} + \underbrace{\Phi_c(t)}_{\text{complementary solution}}.$$

Source page 95

If an initial condition $\mathbf{x}(t_0) = \mathbf{x}_0$ is given, we choose $\Phi(t; t_0, X_0)$ to be the state-transition matrix and integrate from t_0 :

$$\mathbf{V}(t) = \int_{t_0}^t \Phi^{-1}(\tau; t_0) \mathbf{f}(\tau) d\tau = \int_{t_0}^t \Phi(t_0; \tau) \mathbf{f}(\tau) d\tau.$$

Then $\mathbf{V}(t_0) = \mathbf{0}$, so $\mathbf{c}$ must be $\mathbf{x}_0$. Using $\Phi(t; t_0)\Phi(t_0; \tau) = \Phi(t; \tau)$ from rule (ii), the solution becomes:

Solution of the non-homogeneous problem

$$\mathbf{x}(t) = \underbrace{\Phi(t; t_0)\mathbf{x}_0}_{\substack{\phi_0(t) \\ \text{due to initial condition}}} + \underbrace{\int_{t_0}^t \Phi(t; \tau) \mathbf{f}(\tau) d\tau}_{\substack{\phi_f(t) \\ \text{due to input}}}.$$

The first term is the response of the system to its initial state alone. The second adds up the response to the input: the input applied at time τ is carried to time t by $\Phi(t; \tau)$.

ex/ Find the solution of the initial value problem

$$\mathbf{x}' = \begin{bmatrix} -1 & 1 \\ -2 & 1 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} u(t), \quad \mathbf{x}(0) = \mathbf{0}, \quad u(t) = \begin{cases} 1 & t > 0 \\ 0 & t \leq 0. \end{cases}$$

First we need a fundamental matrix for the homogeneous problem. The two functions

$$\phi_1 = \begin{bmatrix} \cos t \\ \cos t - \sin t \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} \sin t \\ \cos t + \sin t \end{bmatrix}$$

are linearly independent solutions, so a fundamental matrix is

$$\Phi(t) = [\phi_1(t) \ \phi_2(t)] = \begin{bmatrix} \cos t & \sin t \\ \cos t - \sin t & \cos t + \sin t \end{bmatrix}.$$

The state-transition matrix is then $\Phi(t; t_0) = \Phi(t)\Phi^{-1}(t_0)$, which evaluates to

$$\begin{aligned} \Phi(t; t_0) &= \begin{bmatrix} \cos t & \sin t \\ \cos t - \sin t & \cos t + \sin t \end{bmatrix} \begin{bmatrix} \cos t_0 + \sin t_0 & -\sin t_0 \\ -\cos t_0 + \sin t_0 & \cos t_0 \end{bmatrix} \\ &= \begin{bmatrix} \cos(t - t_0) - \sin(t - t_0) & \sin(t - t_0) \\ -2\sin(t - t_0) & \cos(t - t_0) + \sin(t - t_0) \end{bmatrix}. \end{aligned}$$

As expected for a time-invariant system, it depends only on $t - t_0$.

Source page 96

For $t > 0$ we have $u(t) = 1$, so $\mathbf{f}(t) = (1, 2)^T$. We solve by variation of parameters, with the fundamental matrix $\Phi(t)$ found above (any fundamental matrix works):

$$\mathbf{x} = \Phi(t)\mathbf{v}(t).$$

$$\begin{aligned}\mathbf{v}'(t) &= \Phi^{-1}(t)\mathbf{f}(t) = \begin{bmatrix} \cos t + \sin t & -\sin t \\ -\cos t + \sin t & \cos t \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad t > 0 \\ &= \begin{bmatrix} \cos t - \sin t \\ \cos t + \sin t \end{bmatrix}.\end{aligned}$$

$$\Rightarrow \mathbf{v}(t) = \begin{bmatrix} \sin t + \cos t + c_1 \\ \sin t - \cos t + c_2 \end{bmatrix}.$$

Multiplying by $\Phi(t)$ gives the general solution:

$$\mathbf{x} = \Phi(t)\mathbf{v}(t) = \begin{bmatrix} 1 + c_1 \cos t + c_2 \sin t \\ (c_1 + c_2) \cos t + (c_2 - c_1) \sin t \end{bmatrix}.$$

We impose the initial condition to find c_1 and c_2 :

$$\mathbf{x}(0) = \mathbf{0} \Rightarrow \begin{bmatrix} 1 + c_1 \\ c_1 + c_2 \end{bmatrix} = \mathbf{0} \Rightarrow c_1 = -1, \quad c_2 = 1.$$

Hence the solution is

Solution of the example

$$\mathbf{x}(t) = \begin{bmatrix} 1 - \cos t + \sin t \\ 2 \sin t \end{bmatrix}, \quad t \geq 0.$$

A Second Way: the Formula Directly

Since $\mathbf{x}(0) = \mathbf{0}$, the initial-condition term vanishes and the input term gives the answer in one step, with no unknown constants to fix:

$$\begin{aligned}\mathbf{x}(t) &= \int_0^t \Phi(t; \tau) \mathbf{f}(\tau) d\tau \\ &= \int_0^t \begin{bmatrix} \cos(t - \tau) - \sin(t - \tau) & \sin(t - \tau) \\ -2 \sin(t - \tau) & \cos(t - \tau) + \sin(t - \tau) \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} d\tau \\ &= \int_0^t \begin{bmatrix} \cos(t - \tau) + \sin(t - \tau) \\ 2 \cos(t - \tau) \end{bmatrix} d\tau = \begin{bmatrix} 1 - \cos t + \sin t \\ 2 \sin t \end{bmatrix},\end{aligned}$$

which agrees with the result above.

Extra note: Linear Dynamical Systems — page 1 (printed 97)

Supplementary note, rewritten from the separately supplied sheet.

ex/ A time-varying system with fixed eigenvectors. Consider the homogeneous, time-varying linear system of differential equations

$$\mathbf{x}' = A(t)\mathbf{x} \quad (\text{Eq.1})$$

where $A(t)$ is $n \times n$ and has the form $A(t) = V \underbrace{\Lambda(t)}_{\text{diagonal}} V^{-1}$. That is, for every t the matrix $A(t)$

has the same eigenvectors, but the corresponding eigenvalues are functions of time.

In this special case we can decouple the system. Define the new unknown $\mathbf{z} = V^{-1}\mathbf{x}$, and substitute $\mathbf{x} = V\mathbf{z}$ into (Eq.1). Because V is constant, $\mathbf{x}' = V\mathbf{z}'$, so

$$\begin{aligned} V\mathbf{z}' &= V\Lambda(t)V^{-1}V\mathbf{z} \\ \Rightarrow \mathbf{z}' &= \Lambda(t)\mathbf{z} \quad (\text{Eqn.2}). \end{aligned}$$

Since $\Lambda(t)$ is diagonal, (Eqn.2) is a set of n separate first-order equations:

$$z'_k = \lambda_k(t)z_k, \quad 1 \leq k \leq n.$$

Each one is solved by separating variables:

$$\begin{aligned} \frac{dz_k}{dt} &= \lambda_k(t)z_k \Rightarrow \int_{z_k(0)}^{z_k(t)} \frac{dz_k}{z_k} = \int_0^t \lambda_k(\tau) d\tau \\ \Rightarrow \ln |z_k| &= \int_0^t \lambda_k(\tau) d\tau + d_k. \end{aligned}$$

Exponentiating gives

Decoupled solution

$$z_k(t) = C_k \cdot e^{\int_0^t \lambda_k(\tau) d\tau}.$$

Returning to the original unknown, with $\mathbf{v}_k$ the columns of V ,

$$\mathbf{x}(t) = V\mathbf{z}(t) = \sum_{k=1}^n \mathbf{v}_k z_k(t) = \sum_{k=1}^n \mathbf{v}_k C_k e^{\int_0^t \lambda_k(\tau) d\tau}.$$

Note that if the system starts at an eigenvector, $\mathbf{x}(0) = \mathbf{v}_m$, then $C_k = 0$ for $k \neq m$ and $C_m = 1$, so the solution stays on the line through $\mathbf{v}_m$. Collecting these n solutions as columns gives a fundamental matrix:

$$\underbrace{\Phi(t; 0, V)}_{\text{fundamental matrix}} = [\mathbf{v}_1 e^{\int_0^t \lambda_1(\tau) d\tau} \quad \dots \quad \mathbf{v}_n e^{\int_0^t \lambda_n(\tau) d\tau}].$$

Extra note: Linear Dynamical Systems — page 2 (printed 98)

Supplementary note, rewritten from the separately supplied sheet.

The same fundamental matrix can be written as a product with a diagonal matrix:

$$\Phi(t; 0, V) = V \begin{bmatrix} e^{\int_0^t \lambda_1(\tau) d\tau} & & \\ & \ddots & \\ & & e^{\int_0^t \lambda_n(\tau) d\tau} \end{bmatrix}.$$

The state-transition matrix is obtained by normalizing with V^{-1} :

$$\Phi(t; 0) = \Phi(t; 0, V)V^{-1} = V \begin{bmatrix} e^{\int_0^t \lambda_1(\tau) d\tau} & & \\ & \ddots & \\ & & e^{\int_0^t \lambda_n(\tau) d\tau} \end{bmatrix} V^{-1}.$$

To start at a general time t_0 we use the composition and inverse rules of the previous section, $\Phi(t; t_0) = \Phi(t; 0)\Phi^{-1}(t_0; 0)$. The middle factors $V^{-1}V$ cancel:

$$\begin{aligned} \Phi(t; t_0) &= \Phi(t; 0)\Phi^{-1}(t_0; 0) \\ &= V \begin{bmatrix} e^{\int_0^t \lambda_1(\tau) d\tau} & & \\ & \ddots & \\ & & e^{\int_0^t \lambda_n(\tau) d\tau} \end{bmatrix} V^{-1} V \begin{bmatrix} e^{-\int_0^{t_0} \lambda_1(\tau) d\tau} & & \\ & \ddots & \\ & & e^{-\int_0^{t_0} \lambda_n(\tau) d\tau} \end{bmatrix} V^{-1}. \end{aligned}$$

Combining the exponentials, the two integrals merge into one from t_0 to t :

STM for constant eigenvectors

$$\Phi(t; t_0) = V \begin{bmatrix} e^{\int_{t_0}^t \lambda_1(\tau) d\tau} & & \\ & \ddots & \\ & & e^{\int_{t_0}^t \lambda_n(\tau) d\tau} \end{bmatrix} V^{-1} \quad (\text{Eqn.3}).$$

ex/ Find the state-transition matrix of

$$\mathbf{x}' = \underbrace{\begin{bmatrix} 1-t & t/2-1 \\ 0 & -t/2 \end{bmatrix}}_{A(t)} \mathbf{x}.$$

Since $A(t)$ is upper triangular, its eigenvalues are the diagonal entries, $\lambda_1 = 1-t$ and $\lambda_2 = -\frac{t}{2}$. We find the eigenvectors from $(A - \lambda I)\mathbf{v} = \mathbf{0}$:

$$(A - \lambda_1 I)\mathbf{v}_1 = \mathbf{0} \Rightarrow \begin{bmatrix} 0 & t/2-1 \\ 0 & t/2-1 \end{bmatrix} \underbrace{\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}}_{\mathbf{v}_1} = \mathbf{0} \Rightarrow \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$(A - \lambda_2 I)\mathbf{v}_2 = \mathbf{0} \Rightarrow \begin{bmatrix} 1-t/2 & t/2-1 \\ 0 & 0 \end{bmatrix} \mathbf{v}_2 = \mathbf{0} \Rightarrow \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Both eigenvectors are constants, which is what (Eqn.3) requires.

Extra note: Linear Dynamical Systems — page 3 (printed 99)

Supplementary note, rewritten from the separately supplied sheet.

Therefore $A(t)$ factors as

$$\begin{aligned} A(t) &= [\mathbf{v}_1 \ \mathbf{v}_2] \begin{bmatrix} \lambda_1(t) & \\ & \lambda_2(t) \end{bmatrix} [\mathbf{v}_1 \ \mathbf{v}_2]^{-1} \\ &= \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_V \underbrace{\begin{bmatrix} 1-t & 0 \\ 0 & -t/2 \end{bmatrix}}_{\Lambda(t)} \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}}_{V^{-1}}. \end{aligned}$$

By (Eqn.3) of the previous page, the state-transition matrix is

$$\Phi(t; t_0) = V \begin{bmatrix} e^{\int_{t_0}^t \lambda_1(\tau) d\tau} & 0 \\ 0 & e^{\int_{t_0}^t \lambda_2(\tau) d\tau} \end{bmatrix} V^{-1}.$$

The two integrals are

$$\int_{t_0}^t \lambda_1(\tau) d\tau = \int_{t_0}^t (1 - \tau) d\tau = \tau - \frac{1}{2}\tau^2 \Big|_{t_0}^t = (t - t_0) - \frac{1}{2}(t^2 - t_0^2),$$

$$\int_{t_0}^t \lambda_2(\tau) d\tau = \int_{t_0}^t -\frac{\tau}{2} d\tau = -\frac{\tau^2}{4} \Big|_{t_0}^t = -\frac{1}{4}(t^2 - t_0^2).$$

Thus

$$\begin{aligned} \Phi(t; t_0) &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^{(t-t_0)-\frac{1}{2}(t^2-t_0^2)} & 0 \\ 0 & e^{-\frac{1}{4}(t^2-t_0^2)} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e^{(t-t_0)-\frac{1}{2}(t^2-t_0^2)} & -e^{(t-t_0)-\frac{1}{2}(t^2-t_0^2)} \\ 0 & e^{-\frac{1}{4}(t^2-t_0^2)} \end{bmatrix}. \end{aligned}$$

Carrying out the last product gives the final answer:

State-transition matrix of the example

$$\Phi(t; t_0) = \begin{bmatrix} e^{(t-t_0)-\frac{1}{2}(t^2-t_0^2)} & -e^{(t-t_0)-\frac{1}{2}(t^2-t_0^2)} + e^{-\frac{1}{4}(t^2-t_0^2)} \\ 0 & e^{-\frac{1}{4}(t^2-t_0^2)} \end{bmatrix}.$$

Extra note: STATE TRANS MATRIX — page 1

Supplementary note, rewritten from the separately supplied sheet.

How to Find the State-Transition Matrix

This page collects the standard ways to find the state-transition matrix (STM) $\Phi(t; t_0)$, first for time-invariant systems $\mathbf{x}' = A\mathbf{x}$, then for a time-varying case.

Time-Invariant Homogeneous Equations

Method 1: diagonalization (or the Cayley–Hamilton method). If $A = Q\Lambda Q^{-1}$, the STM is the matrix exponential, which is computed on the eigenvalues:

STM by diagonalization

$$\Phi(t; 0) = e^{At} = Q \begin{bmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{\lambda_n t} \end{bmatrix} Q^{-1}.$$

Remark: For a time-invariant system, $\Phi(t; t_0) = \Phi(t - t_0; 0) = e^{A(t-t_0)}$.

Method 2: inverse Laplace transform. Transforming $\mathbf{x}' = A\mathbf{x}$ gives $\bar{\mathbf{x}} = (sI - A)^{-1}\mathbf{x}_0$, so

STM by the inverse Laplace transform

$$\Phi(t; 0) = \mathcal{L}^{-1} \{ (sI - A)^{-1} \}.$$

Method 3: from a fundamental matrix. Find

$$\Phi(t; t_0, X_0) = [\phi(t; t_0, \mathbf{x}_1) \cdots \phi(t; t_0, \mathbf{x}_n)]$$

by building the solutions from the eigenpairs $(\lambda_k, \mathbf{u}_k)$; for $n = 2$ these are combinations $C_1 \mathbf{u}_1 e^{\lambda_1 t} + C_2 \mathbf{u}_2 e^{\lambda_2 t}$. Then normalize with the inverse of the starting matrix:

$$\Phi(t; t_0) = \Phi(t; t_0, X_0) \cdot X_0^{-1}.$$

Time-Varying Homogeneous Equations

Diagonalization with constant eigenvectors. If $A(t) = Q\Lambda(t)Q^{-1}$ with a constant eigenvector matrix Q , then each eigenvalue is integrated separately:

$$\Phi(t; t_0) = Q \begin{bmatrix} e^{\int_{t_0}^t \lambda_1(\tau) d\tau} & & \\ & \ddots & \\ & & e^{\int_{t_0}^t \lambda_n(\tau) d\tau} \end{bmatrix} Q^{-1}.$$

Remark: This is valid only when the eigenvector matrix Q does not change with time; see the Linear Dynamical Systems notes for the derivation.

Time-Invariant Non-Homogeneous Equations

For $\mathbf{x}' = A\mathbf{x} + B\mathbf{u}(t)$ with $\mathbf{x}(t_0) = \mathbf{x}_0$, the solution $\phi(t; t_0, \mathbf{x}_0)$ combines the response to the initial state with the response to the input:

Time-invariant forced response

$$\phi(t; t_0, \mathbf{x}_0) = e^{A(t-t_0)}\mathbf{x}_0 + e^{At} \int_{t_0}^t e^{-A\tau} B\mathbf{u}(\tau) d\tau.$$